Meike Köhler

Flexible Bayesian joint models for longitudinal biomarkers and time-to-event outcomes with applications to type 1 diabetes research

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

Eingereicht am 22. August 2017

Meike Köhler

Flexible Bayesian joint models for longitudinal biomarkers and time-to-event outcomes with applications to type 1 diabetes research

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

Eingereicht am 22. August 2017

Erster Berichterstatter: Prof. Dr. Sonja Greven
Zweiter Berichterstatter: Prof. Dr. Thomas Kneib

Tag der Disputation: 11. Oktober 2017

"You know more than you think you know, just as you know less than you want to know."

Oscar Wilde, The Picture of Dorian Gray

Acknowledgments

I am deeply grateful for all the help and support I received with this thesis. My special thanks go to ...

... Sonja Greven for her statistical advice throughout the projects, fruitful discussions on joint modeling, thorough feedback in writing and continuous support during the whole dissertation project.

... Andreas Beyerlein for support and advice in all projects, especially regarding the applied analyses.

... Anette-Gabriele Ziegler for all support provided by the Institute of Diabetes Research and many learning opportunities beyond the focus of this work.

... Thomas Kneib as the external examiner of the thesis.

... Nikolaus Umlauf for invaluable advice and help in implementing the joint model and for his unbreakable optimistic view on the subject.

... HELENA, the graduate school of the Helmholtz Zentrum München, for their support in all organizational issues and allowing me to focus completely on my work.

... my colleagues in the working group, especially Jona, for support, feedback and a great atmosphere.

... my office mate David for technical support in all situations, fruitful discussions on my topic as well as leisure and laughter (together with Henry).

... my colleagues at the Institute of Statistics, the Institute of Diabetes Research and my fellows at the doctoral student initiative DINI for enriching my daily life beyond modeling biomarkers.

... Sarah, David, Almond and Kristin for proofreading parts of the thesis.

... my family and friends for their continuous support.

... and Manu - for everything.

Zusammenfassung

Die Analyse des komplexen Zusammenhangs zwischen längsschnittlich erfassten Biomarkern und der Zeit bis zum Eintreten eines (Krankheits-)Ereignisses ist für viele biomedizinische Anwendungen von Interesse. Joint Models zur Modellierung von Längsschnitt- und Überlebenszeitdaten finden in diesem Bereich zunehmend Verbreitung und können potentiell neue Erkenntnisse in der Erforschung des Typ 1 Diabetes (T1D) liefern, dem Ausgangspunkt der vorliegenden Methodenentwicklung.

T1D, eine der häufigsten chronischen Erkrankungen im Kindesalter, geht mit einer lebenslangen Abhängigkeit von künstlichem Insulin einher. Schon vor dem Einsetzen klinischer Symptome ist eine Diagnose der Krankheit durch T1D-spezifische Autoantikörper möglich. Die hohe Variabilität in der Dauer dieser präklinischen Phase kann potentiell durch Autoantikörper als längsschnittliche Biomarker erklärt werden.
Die Analyse längsschnittlicher Biomarker ist statistisch herausfordernd, da die Biomarker nur an diskreten Messzeitpunkten bis zum Eintreten des Ereignisses mit einem gewissen Messfehler beobachtet werden. Um diese fehlerbehafteten, informativ zensierten Daten ohne Verzerrung mit der Überlebenszeit in Verbindung zu setzen, schätzen Joint Models ein längsschnittliches Modell und ein Überlebenszeitmodell in einer gemeinsamen Likelihoodfunktion. In bestehenden Joint Models fehlt allerdings die für den Anwendungsfall T1D nötige Flexiblität in der Modellierung von personenspezifischen nichtlinearen Biomarkertrajektorien sowie in der zeitabhängigen Modellierung der Assoziation zwischen Marker und Überlebenszeit.

Um komplexe Zusammenhänge zwischen longitudinalen Biomarkern und dem Eintreten eines Ereignisses abbilden zu können, wird ein flexibles Bayesianisches Joint Model entwickelt. Das Joint Model ist hierbei als struktiertes additives Regressionsmodell formuliert, wodurch eine Vielzahl von verschiedenen Effekten wie glatte nichtlineare, zeitvariierende oder zufällige Effekte integriert werden können. Durch einen Bayesianischen Ansatz wird eine stabile Schätzung dieses komplexen Modells, insbesondere hochdimensionaler zufälliger Effekte, erreicht. Diese Flexibilität geht über bisherige Entwicklungen hinaus und ermöglicht einen detaillierten Einblick in den Zusammenhang zwischen personenspezifischen nichtlinearen Antikörperverläufen und dem Einsetzen klinischer Symptome am Beispiel zweier T1D-Risikokohorten. Darüber hinaus wird dieses Rahmenmodell weiter generalisiert um auch nichtlineare Zusammenhänge zwischen Biomarkern und der Überlebenszeit abbilden zu können. Diese Erweiterung erlaubt die Überprüfung einer üblicherweise nicht getesteten Linearitätsannahme zwischen Biomarker und logarithmiertem Hazard, wie in der Analyse eines weiteren biomedizinischen Datensatzes über eine seltene Lebererkrankung veranschaulicht. Um Anwendern die Nutzung dieses flexiblen Modells zu ermöglichen wird die Implementation im R Paket **bamlss** in der vorliegenden Arbeit vorgestellt.

Das neu entwickelte Bayesianische additive Joint Model erlaubt eine flexible Modellierung längsschnittlicher Biomarker und deren potentiell komplexen Zusammenhang mit der Überlebenszeit, weit über bisherige Joint Model Ansätze hinaus. Damit können neue Einblicke in den zeitabhängigen Zusammenhang zwischen T1D-spezifischen Biomarkern und dem Einsetzen klinischer Symptome gewonnen werden.

Summary

The potentially complex association between a longitudinal biomarker and a time-to-event process, often called survival process, is of large interest in various biomedical domains. Joint models of longitudinal and survival data are a tool of growing popularity for this analysis and can potentially enable new insights into the disease progression of type 1 diabetes (T1D), the motivating application of the presented work.

T1D, one of the most common chronic diseases in childhood, is marked by the lifelong need of insulin substitution. A preclinical period of the disease can be diagnosed by the emergence of the T1D-specific autoantibodies which can potentially also serve as longitudinal biomarkers to explain the high variability in the time to clinical T1D.
The analysis of longitudinal biomarkers and survival processes is complicated by observing the longitudinal marker at discrete time points, subject to measurement error, and informatively censored at the occurrence of an event. Joint models allow an unbiased analysis of this data by estimating a longitudinal and a survival submodel within a joint likelihood function. Previously developed joint models, however, lack the flexibility to capture the highly nonlinear individual T1D-biomarker trajectories and a potentially time-varying association between the marker and the onset of clinical symptoms of T1D.

In order to provide additional insights into the complex relationship between longitudinal biomarkers and survival processes a flexible Bayesian joint model framework is developed. By formulating the model using structured additive predictors, a variety of effects can be incorporated, such as smooth nonlinear, time-varying and random effects, thereby allowing flexibility beyond established joint model implementations. The adoption of a Bayesian framework enables a stable estimation of this complex model, especially regarding potentially high-dimensional random effect structures. The resulting flexibility is used to explore the association between highly heterogeneous autoantibody profiles and the progression to T1D in two different diabetes risk cohorts. This framework is further generalized by incorporating nonlinear association structures between the longitudinal and the survival process. The extension allows to assess the usually unchecked assumption of linearity between the marker and the log-hazard of a survival process as illustrated in the analysis of biomedical data on a rare fatal liver disease. To facilitate its application the flexible Bayesian joint model is implemented in the R package **bamlss** and an introduction to its usage is given.

The newly developed additive joint model offers flexibility in modeling longitudinal biomarkers and complex associations structures beyond established joint model approaches, thereby allowing new insights into the time-varying association between T1D-specific autoantibodies and the onset of clinical symptoms.

Contents

Chapter 1

Introduction

1.1 Modeling longitudinal biomarkers and time-to-event processes

The potentially complex association between a longitudinal biomarker and a time-to-event process, often also called survival process, is of large interest in various biomedical domains, not only with regard to better understanding disease processes but also in the growing field of personalized medicine (Rizopoulos et al., 2014). The usual data structure in this context, which is, the observed longitudinal marker measurements at discrete time points, potentially subject to measurement error, and informatively censored at the occurrence of an event, poses challenges to the analysis of this association. Joint models of longitudinal and survival data, which allow for unbiased estimations of the longitudinal and time-to-event processes in this data structure, gained larger popularity in the last decade and continue to be an active area of research (Gould et al., 2015). By estimating the longitudinal model and the survival model within a joint likelihood function, this model class can account for the informative censoring in the longitudinal model and incorporate the longitudinal information as a continuous-time covariate without measurement error in the survival model. This approach also has the potential to offer new insights into the disease progression of type 1 diabetes (T1D).

1.2 Applied research question and data sources

T1D is one of the most common chronic diseases in childhood, with worldwide increasing incidence (Patterson et al., 2009). Disease progression starts before the onset of clinical symptoms, i.e. the lifelong need of insulin substitution, when insulin-producing β-cells in the pancreas are gradually destroyed by the body's own immune system. This preclinical period can be diagnosed by the emergence of the T1D-specific autoantibodies to insulin (IAA), glutamic acid decarboxylase (GADA), and insulinoma-associated protein 2 (IA2A). The time from the first emergence of autoantibodies, called seroconversion, to the onset of clinical symptoms varies considerably between individuals, ranging from weeks to decades (Ziegler et al., 2013). It is known that the combination of different autoantibodies as well as the autoantibody levels, especially of IAA, are associated with progression time (Achenbach et al., 2004; Parikka et al., 2012; Steck et al., 2011, 2015). However, it remains an open

question how the longitudinal patterns of the autoantibodies are associated with the progression to T1D.

In order to gain insights into this research question two different data sets from T1D risk cohorts were analyzed. The first data set as presented in Chapter 3 is the combined data set of two ongoing German T1D risk cohorts, the BABYDIAB and the BABYDIET study. The goal of these prospective birth cohorts is to investigate the natural history of T1D development. For this, 2441 children with familial increased risk of T1D were followed from birth to the development of T1D or loss to follow-up for up to 21 years (Ziegler et al., 1993, 1999; Hummel et al., 2004, 2011). Autoantibody measurements were taken from the children at age 9 months and 2, 5, 8, 11, 14 and 17 years and additionally every 6 months after islet autoantibodies had emerged. In Chapter 3 the data of 127 children who developed IAA during follow-up is analyzed of which 69 (54%) progressed to T1D.

The second data set is from the multinational study named The Environmental Determinants of Diabetes in the Young (TEDDY) study (The TEDDY Study Group, 2007; Krischer et al., 2015), an ongoing prospective cohort study funded by the National Institutes of Health with the primary goal to identify environmental causes of T1D. In six clinical research centers located in the USA, Finland, Germany, and Sweden the TEDDY study enrolled 8,676 children with increased genetic risk for T1D between 2004 and 2010 and autoantibody measurements were taken every 3 months. In Chapter 4 we present analyses of all 613 children who had developed one or more islet autoantibodies in the course of follow-up of which 175 (29%) progressed to T1D.

In order to exemplify the different approaches to longitudinal and survival analysis, data from a mouse model are presented and further analyzed in Chapter 2. Here 71 mice who are highly susceptible to diabetes (NOD mice) and 40 mice without an increased susceptibility (NOR mice) were followed from 4 weeks until 36 weeks of age or the onset of diabetes. Blood measurements were drawn biweekly to assess autoantibodies and different cell populations in order to gain insights into the immunologic characteristics of these mice at risk.

To further illustrate the development of our method and its implementation in the R package **bamlss** (Umlauf et al., 2017) a further biomedical data set is analyzed in Chapter 5, which is widely studied in the context of joint models, a data set on patients of the rare fatal liver disease primary biliary cirrhosis (PBC). In a study conducted between 1974 and 1984 by the Mayo clinic (Murtaugh et al., 1994), patients were followed to study the influence of the drug D-penicillamine on the survival of the patients. Visits were scheduled at six months, 12 months and annually thereafter. We analyze data of 312 patients of which 140 (45 %) died during follow-up.

1.3 Scope of this work

The focus of this dissertation lies on the development of a flexible Bayesian joint model framework in order to allow for additional insights into the complex relationships between longitudinal biomarkers and event processes. By formulating the joint model using structured additive predictors this framework allows to incorporate a variety of effects, such as smooth nonlinear, time-varying and random effects. The Bayesian framework thereby enables a stable estimation of these complex

structures, especially regarding potentially high-dimensional random effects. This flexibility is used to explore the association between highly heterogeneous autoantibody profiles and the progression to T1D. The presented framework is further enlarged by the incorporation of nonlinear association structures between longitudinal and event processes, thereby allowing the exploration of highly complex relationships beyond the initially motivating research question.

The outline of this dissertation is as follows. Chapter 2 gives a brief overview of modeling approaches for longitudinal data and time-to-event data, both separately and within joint models with illustrations using the data on the T1D mouse model. Special attention is paid to extensions beyond linear, parametric models and on the Bayesian estimation of these approaches in order to lay the groundwork for the methods development presented in Chapter 3. In this chapter, the developed model is tested in simulations and used to tackle the motivating research question by analyzing data of the BABYDIAB/BABYDIET study. The study of disease progression to clinical T1D by means of our flexible Bayesian joint model is further extended in Chapter 4 when data of the TEDDY study are analyzed in detail. While the presented framework allows for high flexibility in all parts of the joint model, the special emphasis in these two chapters lies on the flexible modeling of nonlinear, individual trajectories of T1D-specific biomarkers and a time-varying association between the biomarker and the onset of the disease. The presented framework is further generalized in Chapter 5 to also allow for nonlinear associations between the longitudinal biomarkers and the log-hazard of an event. This extension is tested in a simulation study and illustrated in modeling the PBC data. In order to facilitate the application of the presented flexible framework the model is implemented within the R package **bamlss**. Chapter 6 presents the central concepts and functionalities of this software implementation. After a summary, this dissertation closes with a discussion of the achievements within this work and an outlook to future research directions in Chapter 7. Technical details of the presented framework can be found in Appendix A and further modeling results from the applied analyses in Appendix B.

1.4 Contributing manuscripts

This dissertation is based on the manuscripts:

> Tanja Telieps*, Meike Köhler*, Irina Treise, Katharina Foertsch, Thure Adler, Dirk H. Busch, Martin Hrabě de Angelis, Admar Verschoor, Kerstin Adler, Ezio Bonifacio, and Anette-Gabriele Ziegler (2016). Longitudinal frequencies of blood leukocyte subpopulations differ between NOD and NOR mice but do not predict diabetes in NOD Mice. *Journal of Diabetes Research*, 2016: e420815. `https://doi.org/10.1155/2016/4208156`.

* *shared first authorship*

Extracts of this manuscript are presented in Section 2.1. I conducted all analyses and wrote the statistical part of the manuscript.

Meike Köhler, Nikolaus Umlauf, Andreas Beyerlein, Christiane Winkler, Anette-Gabriele Ziegler, and Sonja Greven (2017). Flexible Bayesian additive joint models with an application to type 1 diabetes research. *Biometrical Journal*. https://doi.org/10.1002/bimj.201600224. (online first)

This manuscript is presented in Chapter 3. I developed the framework in collaboration with Prof. Dr. Sonja Greven and Dr. Nikolaus Umlauf, jointly implemented the model in the existing R package **bamlss** together with Dr. Umlauf, who contributed significantly, conducted all simulations and applied analyses and wrote the manuscript with input from all coauthors.

Meike Köhler*, Andreas Beyerlein*, Kendra Vehik, Sonja Greven, Nikolaus Umlauf, Åke Lernmark, William A. Hagopian, Marian Rewers, Jin-Xiong She, Jorma Toppari, Beena Akolkar, Jeffrey P. Krischer, Ezio Bonifacio, Anette-G. Ziegler, and the TEDDY study group (2017). Joint modeling of longitudinal autoantibody patterns and progression to type 1 diabetes: results from the TEDDY study. *Acta Diabetologica*. https://doi.org/10.1007/s00592-017-1033-7. (online first)

* *shared first authorship*

This manuscript forms Chapter 4. I conducted all analyses and wrote the paper jointly with Dr. Andreas Beyerlein with input from all coauthors.

Meike Köhler, Nikolaus Umlauf, Sonja Greven (2017). Nonlinear association structures in flexible Bayesian additive joint models. *arXiv e-prints [stat.ME]*. 1708.06337. https://arxiv.org/abs/1708.06337.

This manuscript is the basis of Chapter 5. I developed the generalized framework in collaboration with Prof. Dr. Sonja Greven and Dr. Nikolaus Umlauf, implemented the model in the existing R package **bamlss** with support from Dr. Nikolaus Umlauf, conducted all simulations and applied analyses and wrote the manuscript with input from all coauthors.

In order to allow Chapters 3 to 5 to be self-sufficient readings of the methods development and application of the flexible additive joint model, the contributing manuscripts were only modified minimally with regard to formatting and referencing. We believe that despite some redundancy in the further chapters this enhances the overall readability of this dissertation.

Chapter 2

From heuristic approaches to joint models

The analysis of longitudinal biomarkers in combination with time-to-event data connects two branches of statistics that are usually treated separately: longitudinal data analysis and survival analysis. Over the last two decades a variety of models have been proposed to analyze these outcomes and their associations, ranging from naive, heuristic models via sophisticated correction methods to the growing field of joint models. These approaches differ with regard to the focus of the analysis which can be placed on the longitudinal outcome, the survival outcome, their association or more complex combinations (Rizopoulos & Lesaffre, 2014). Joint models allow the unbiased estimation of all these components within one model. In the following a simple approach for the analysis of longitudinal biomarkers in the context of a T1D mouse model will be presented. This example will be further extended to illustrate central challenges and approaches in the separate and joint analysis of longitudinal biomarkers and time-to-event processes.

2.1 Example: Finding biomarkers in a type 1 diabetes mouse model

The research of complex human diseases often utilizes model organisms in which, by means of breeding or gene alteration, human diseases are mimicked and can therefore be studied in more detail. The NOD mouse is an important model of spontaneous autoimmune diabetes that has been used to understand T1D pathogenesis and develop therapeutics to prevent β-cell destruction (Anderson and Bluestone, 2005). Similar to humans who have genetic susceptibility to type 1 diabetes and islet autoimmunity, not all NOD mice progress to overt diabetes, and progression to diabetes occurs at a variable time point after the start of autoimmunity (Atkinson and Leiter, 1999; Pozzilli et al., 1993). NOD mice have alterations in certain immune-related cell populations as compared to other mouse

strains (Mariño et al., 2008; Marrero et al., 2012). It is therefore possible that the extent of change may associate with the likelihood and rate of diabetes development in individual mice and that this may reveal markers that may help stratify the rate of progression to diabetes in man. In order to potentially find further biomarkers for T1D progression, certain cell populations in the blood from NOD and NOR mice, a control group without T1D susceptibility, were quantified in biweekly intervals from 4 weeks to 36 weeks of age or until diabetes onset. We compared trajectories of different immune cell subpopulations as well as the autoantibody IAA between NOD and NOR mice and within NOD mice that progressed and did not progress to diabetes within 36 weeks. In total 71 NOD mice and 40 NOR mice were studied of which 41 NOD mice (59%) developed diabetes within the follow-up of 36 weeks. In the experimental group, that is analyzed in the following, of 58 NOD mice and 22 NOR mice blood was drawn biweekly to assess IAA, cell subpopulations as well as further physiological measures.

Although all cell populations were of interest in the study, we focus in the following on the analysis of the longitudinal trajectories of IAA. These were modeled using first-, second-, and third-order polynomial growth models for each measure (Singer and Willett, 2003) with random intercepts for each mouse to account for the longitudinal correlation structure. Model selection was based on the Akaike information criterion (Akaike, 1974). Trajectories were compared between (a) NOD and NOR mice as well as (b) NOD mice that developed overt diabetes and NOD mice that did not develop diabetes. For the comparisons only data until the age of 30 weeks was used in order to ensure sufficient data in all groups. For the group comparisons of the 14 cell subpopulations we accounted for multiple testing resulting in the Bonferroni corrected significance level of 0.004. Longitudinal trajectories were visualized by time pointwise means and confidence intervals (CI) and smoothed locally via LOESS (Cleveland et al., 1992).

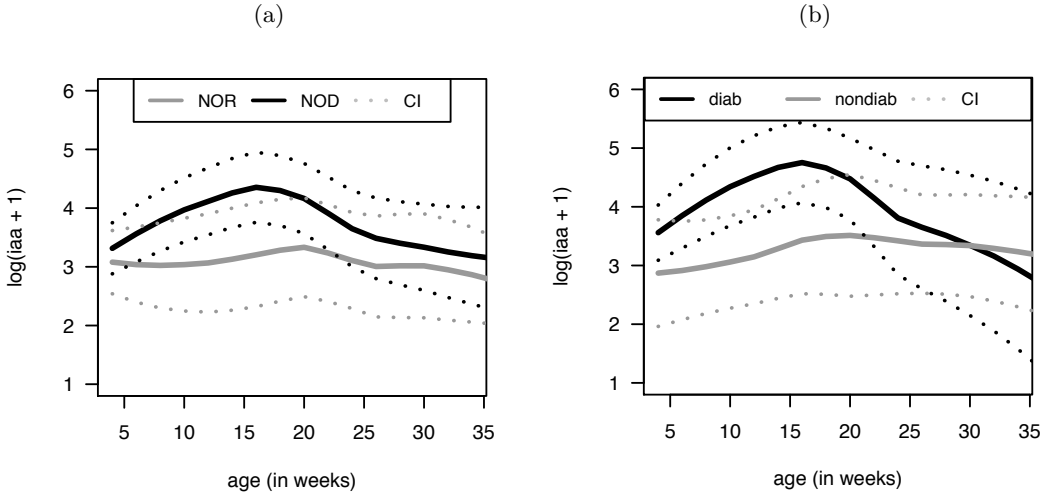

Figure 2.1: Trajectories of IAA in (a) NOD mice versus NOR mice and (b) NOD mice that developed diabetes (diab) versus NOD mice that were diabetes-free at age of 36 weeks (nondiab).[1]

In these analyses trajectories of IAA were higher in NOD mice than in NOR mice with a mean difference between the groups of 0.92 [95% confidence interval: 0.39, 1.44] and higher in NOD mice that developed diabetes as compared to NOD mice that did not develop diabetes with mean difference of 1.05 [95% confidence interval: 0.43, 1.67] (Figure 2.1). By means of this straightforward modeling approach differences between the experimental mice and the control subjects as well as between progressors and non-progressors were observed. Potential issues with and refinements of this analysis are further explored in the subsequent sections.

2.2 A Bayesian primer

The flexible joint model that is developed and presented in this dissertation is estimated in a Bayesian framework. In the following, basic concepts of the Bayesian view on statistical data analysis are presented. These concepts are the groundwork to understand not only the Bayesian view on modeling approaches in the remainder of this chapter but also for the setup and estimation of the flexible additive joint model, which is the core of this work. The following section is meant as a brief overview of relevant concepts, based on Gelman et al. (2013). We refer to this and further literature for more detailed insights into Bayesian statistics (Robert, 2007; Carlin and Louis, 2008).

As Gelman et al. (2013, p. 3) states, the "essential characteristic of Bayesian methods is their explicit use of probability for quantifying uncertainty in inferences based on statistical data analysis", thereby pointing out key features of Bayesian statistics. In statistical data analysis *observed data* $\boldsymbol{y} = [y_1, \cdots, y_n]^\top \in \mathbb{R}^n$ is seen as a random sample from an underlying probability distribution $p(\boldsymbol{y}|\boldsymbol{\theta})$ depending on some unknown parameter (vector) $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The main goal lies in deriving *inferential conclusions* for $\boldsymbol{\theta}$, i.e. in providing parameter estimates $\hat{\boldsymbol{\theta}}$ and information on the *uncertainty* of the parameter estimation. In contrast to a frequentist view, where the parameters of interest are assumed to have a fixed true value in the population, in the Bayesian view all parameters of interest are random variables with a certain prior distribution $p(\boldsymbol{\theta})$. The aim of the Bayesian analysis is to obtain information of the *probability distribution of these parameters* based on the observed data, $p(\boldsymbol{\theta}|\boldsymbol{y})$, which can be derived by making use of the Bayes theorem.
In line with Gelman et al. (2013) we denote in the following conditional probability densities with $p(\cdot|\cdot)$, marginal distributions with $p(\cdot)$ and use the terms distribution and density interchangeably.

2.2.1 Bayes theorem

The first step for obtaining probability statements about $p(\boldsymbol{\theta}|\boldsymbol{y})$ is to develop the joint probability distribution

$$p(\boldsymbol{\theta}, \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

[1]Figure modfied from Telieps et al. (2016).

from the prior distribution $p(\boldsymbol{\theta})$, representing knowledge on $\boldsymbol{\theta}$, and the sampling distribution $p(\boldsymbol{y}|\boldsymbol{\theta})$, which is the distributional assumption on the data or the statistical model, also called the likelihood. Bayes theorem allows then to derive the quantity of interest, the posterior distribution

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{\theta}, \boldsymbol{y})}{p(\boldsymbol{y})}$$

with the marginal distribution $p(\boldsymbol{y}) = \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta})d\boldsymbol{\theta}$ for continuous $\boldsymbol{y}$. The posterior density is at the center of Bayesian inference and represents the probability distribution of the parameter of interest after observing the data. Whereas $p(\boldsymbol{y})$ is the normalizing constant that turns $p(\boldsymbol{\theta}|\boldsymbol{y})$ into a proper distribution, it is a constant with respect to $\boldsymbol{\theta}$ and for computational ease usually the unnormalized posterior density

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

is used. Statistics of interest can be derived from the posterior density such as the posterior mean $\hat{\boldsymbol{\theta}} = \int_{\boldsymbol{\Theta}} \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}$ or the posterior mode $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \; p(\boldsymbol{\theta}|\boldsymbol{y})$. Furthermore, regions of confidence can be derived such as the $(1-\alpha)$-credibility interval $[\boldsymbol{\theta}_l, \boldsymbol{\theta}_u]$ for which it holds that $\int_{\boldsymbol{\theta}_l}^{\boldsymbol{\theta}_u} p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} = 1-\alpha$. Credibility intervals are not unique and different intervals of interest can be computed. We use the 2.5th and 97.5th percentiles of the empirical posterior distribution (see following subsection) in the upcoming chapters.

2.2.2 Bayesian estimation approaches

For simple models closed-form solutions for the posterior distribution can be found, often using conjugate priors. These are priors that, multiplied with the sampling distribution, result in closed-form posterior distributions of the same distribution family as the prior distribution. However, this is usually not possible for statistical models with more complicated distributional assumptions such as the flexible additive joint model, that is presented in the subsequent chapters. For models to which no closed-form solution for the posterior exists, the target posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ is approximated by an empirical distribution of *samples* from the posterior. In some cases direct sampling from $p(\boldsymbol{\theta}|\boldsymbol{y})$ is possible, however, more often elaborate sampling techniques, so called Markov Chain Monte Carlo Methods (MCMC) are used. The main idea behind MCMC methods is to create a Markov process with stationary distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$.

Two widely used MCMC samplers are the Gibbs sampler (or short Gibbs) and the Metropolis-Hastings algorithm. Gibbs sampling is a popular approach for applications in which the full parameter vector $\boldsymbol{\theta}$ can be divided into subvectors $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_s, \cdots, \boldsymbol{\theta}_S]$ and the conditional distributions $p(\boldsymbol{\theta}_s|\boldsymbol{\theta}_{-s}, \boldsymbol{y})$ are known or samples can be drawn from them. Here $\boldsymbol{\theta}_{-s}$ denotes all components of $\boldsymbol{\theta}$ except of $\boldsymbol{\theta}_s$. In every iteration $l = 1, \ldots, L$ this sampler cycles through all S subvectors and draws subsets $\boldsymbol{\theta}_s^{[l]}$ conditional on the most recent sample of $\tilde{\boldsymbol{\theta}}_{-s}$. The Gibbs sampler is a special case of the Metropolis-Hastings algorithm (see Gelman et al., 2013, p.281). In the Metropolis-Hastings algorithm, a random walk is generated to draw a new sample $\boldsymbol{\theta}^*$ based on the previous sample such that $\boldsymbol{\theta}^* = \boldsymbol{\theta}^{[l-1]} + \epsilon$ where ϵ is drawn from a specific distribution. In more detail, in every iteration l a

sample is drawn from the proposal distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{[l-1]})$ and accepted as a sample of the posterior $\boldsymbol{\theta}^{[l]} = \boldsymbol{\theta}^*$ with probability

$$\min\left(1, \frac{p(\boldsymbol{\theta}^*|\boldsymbol{y})/q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{[l-1]})}{p(\boldsymbol{\theta}^{[l-1]}|\boldsymbol{y})/q(\boldsymbol{\theta}^{[l-1]}|\boldsymbol{\theta}^*)}\right).$$

Otherwise the previous sample is kept as $\boldsymbol{\theta}^{[l]} = \boldsymbol{\theta}^{[l-1]}$. In order for the algorithm to work, the Markov chain must have a stationary distribution and this stationary distribution has to equal the target distribution.

Different diagnostic measures allow to judge the convergence to a stationary distribution, e.g. the graphical inspection of traceplots or starting multiple chains from different starting values and assessing the posterior distributions from the different chains. In order to reduce the influence of the starting values, a certain amount of samples, called burnin, is discarded in the beginning. For most statistics of interest that are derived from the empirical posterior distribution, a moderate amount of independent samples usually suffices. However, often samples exhibit a certain amount of autocorrelation which increases the needed amount of samples for the same precision in the statistics. Thinning the chain by keeping only every kth sample ($k \in \mathbb{N}, k > 1$) can alleviate the issue of autocorrelation. For the Metropolis-Hastings algorithm it holds that the closer the proposal distribution approximates the posterior distribution, the higher the acceptance rate, i.e., the more new samples are kept and consequently a shorter chain suffices to estimate quantities of interest. For this reason, the flexible additive joint model uses a derivative-based approximation of the posterior based on a second-order Taylor approximation as proposal distribution, which results in faster convergence and good mixing (see Chapter 3).

A further sampling method that is also used for certain model parameters in the flexible additive joint model is slice sampling (Neal, 2003). In slice sampling the empirical posterior distribution is obtained by sampling from the area under the target distribution. In more detail, a starting value h is drawn from $U(0, p(\boldsymbol{\theta}_j^{[l-1]}|\cdot)$. From this starting value a horizontal "slice" $S = \{\boldsymbol{\theta}_j : h < p(\boldsymbol{\theta}_j|\cdot)\}$ of the target distribution is obtained and a sample $\boldsymbol{\theta}_j^{[l]} \sim U(S)$ is drawn (see also the implementation in **bamlss** in Umlauf et al., 2017).

In practice, different sampling approaches are often combined in order to estimate posterior distributions of complex models. For the combination of Gibbs and Metropolis-Hastings, for example, subvectors with given closed-form conditional posterior distributions are sampled using Gibbs and other subvectors of $\boldsymbol{\theta}$ are sampled using the Metropolis-Hastings algorithm. Examples of the Bayesian estimation of longitudinal, survival and joint models can be found in the following subsections.

2.3 Longitudinal models

In order to gain insights into the natural history of diseases and to derive causal conclusions on treatments, cohort studies with potential case-control grouping are a central instrument. Usually subjects are followed over time and measurements are taken at fixed or subject-specific time points. This longitudinal design induces a temporal correlation structure in the data, as multiple measurements are

observed for every subject, which has to be accounted for in further investigations. In the following a standard approach for longitudinal data analysis, linear mixed effects models, and its extension to structured additive regression models is briefly presented. This general framework represents the basis for the flexible additive joint model developed in later chapters. Additionally, different missing data mechanisms are briefly explored with the focus on the special data structure in our context, which is longitudinal markers that are observed until the occurrence of an event. This missing data mechanism motivates the use of joint models and generic modeling approaches for this scenario are presented.

2.3.1 Linear mixed effects models

Linear mixed effects models are a standard approach for modeling correlated data with a grouping structure and thus also suitable when n_i longitudinal measurements $\boldsymbol{y}_i = [y_{i1}, \cdots, y_{ij}, \cdots, y_{in_i}]^\top$ are observed on the same subject $i = 1, \ldots, n$ at potentially subject-specific time points $t_{ij} = t_{i1}, \ldots, t_{in_i}$. The linear mixed effects model (Verbeke and Molenberghs, 2000) is

$$\boldsymbol{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \tag{2.1}$$

where $\mathbf{X}_i$ and $\mathbf{Z}_i$ are the design matrices of observed covariates, $\boldsymbol{\beta}$ is a vector of length p of population-specific fixed parameters, $\boldsymbol{b}_i$ is a vector of length q of subject-specific random effects and $\boldsymbol{\epsilon}_i$ is a vector of residuals. The random effects and the errors are assumed to be independent and $\boldsymbol{b}_i \sim N(\mathbf{0}, \tilde{\mathbf{D}})$ and $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$ with $\tilde{\mathbf{D}}$ being the covariance matrix of random effects and $\mathbf{R}_i = \sigma^2\boldsymbol{I}_{n_i}$ as the covariance matrix of the residuals. The *conditional* representation of this model in eq. (2.1) may be also written as

$$\boldsymbol{y}_i|\boldsymbol{b}_i \sim N(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{b}_i, \mathbf{R}_i)$$

and allows for a subject-specific interpretation in which the expected value is the sum of the population average and subject-specific deviations from it. The *marginal* representation of the model can be derived from the above as

$$\boldsymbol{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{R}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top). \tag{2.2}$$

The representation in eq. (2.2) illustrates how the random effects induce an additional correlation structure in the data. Furthermore, the marginal model is the basis for the estimation of the linear mixed effects model. For a quick overview of this estimation we consider the full design matrices $\mathbf{X} = [\mathbf{X}_1^\top, \cdots, \mathbf{X}_n^\top]^\top$ and $\mathbf{Z} = \text{blockdiag}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$, the random effects $\boldsymbol{b} = [\boldsymbol{b}_1^\top, \cdots, \boldsymbol{b}_n^\top]^\top$ as well as the covariance matrix for the random effects $\mathbf{D} = \text{blockdiag}(\tilde{\mathbf{D}}, \ldots, \tilde{\mathbf{D}})$ and the covariance matrix for the residuals $\mathbf{R} = \text{blockdiag}(\mathbf{R}_1, \ldots, \mathbf{R}_n)$ where $\mathbf{D} =: \mathbf{D}(\boldsymbol{\vartheta})$ and $\mathbf{R} =: \mathbf{R}(\boldsymbol{\vartheta})$ depend on the unknown covariance parameters $\boldsymbol{\vartheta}$. The covariance of the response is then $V := \text{Cov}(\boldsymbol{y}) = \mathbf{R}+\mathbf{Z}\mathbf{D}\mathbf{Z}^\top$. For given $\boldsymbol{\vartheta}$, estimates of the fixed effects can be obtained via least squares as $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{V}(\boldsymbol{\vartheta})^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}(\boldsymbol{\vartheta})^{-1}\boldsymbol{y}$. It can be shown that $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$. The best linear unbiased predictor (BLUP) for the random effects $\boldsymbol{b}$ can be obtained from $\hat{\boldsymbol{b}} = \mathbf{D}(\boldsymbol{\vartheta})\mathbf{Z}^\top\mathbf{V}(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$. Estimates for

$\boldsymbol{\vartheta}$ are also based on the log-likelihood of the marginal model which is, omitting additive constants, (Fahrmeir et al., 2013)

$$\ell(\boldsymbol{\beta}, \boldsymbol{\vartheta}) = -\frac{1}{2}\left[\log(\det(\boldsymbol{\vartheta}) + (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})\right]. \tag{2.3}$$

Maximizing this likelihood with respect to $\boldsymbol{\beta}$ and inserting this estimate $\hat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})$ into eq. (2.3) results in the profile likelihood $\ell_P(\boldsymbol{\vartheta})$. When maximizing this profile likelihood for $\boldsymbol{\vartheta}$ the maximum likelihood (ML) estimator $\hat{\boldsymbol{\vartheta}}_{ML}$ is obtained. As this estimate is biased downwards, often the estimation is based on the restricted likelihood $\ell_R(\boldsymbol{\vartheta}) = \log\left(\int L(\boldsymbol{\beta}, \boldsymbol{\vartheta}) d\boldsymbol{\beta}\right)$ by integrating out $\boldsymbol{\beta}$ (Fahrmeir et al., 2013). By maximizing $\ell_R(\boldsymbol{\vartheta})$ for $\boldsymbol{\vartheta}$ the restricted ML estimator $\hat{\boldsymbol{\vartheta}}_{REML}$ is obtained. As these likelihoods have no closed-form solutions the estimators are computed numerically for example using the computationally more demanding Newton-Raphson algorithm or Expectation-Maximization (EM) algorithms (Pinheiro and Bates, 2000). A combination of both was also used in the analysis of the longitudinal antibody trajectories in mice, as presented in the previous section.

From a Bayesian view not only the random effects $\boldsymbol{b}_i$ but also the population effects $\boldsymbol{\beta}$ are considered random variables with a prior distribution $p(\boldsymbol{\beta})$. For the population effects noninformative priors $p(\boldsymbol{\beta}) \propto const$ or weakly informative priors $p(\boldsymbol{\beta}) \propto N(\boldsymbol{m}, \mathbf{M})$ can be used where the expectation $\boldsymbol{m}$ and the covariance $\mathbf{M}$ are assumed to be known. The normal random effects distribution is now seen as the respective prior of $\boldsymbol{b}$. In an empirical Bayes approach the covariance parameters $\boldsymbol{\vartheta}$ are considered unknown but fixed whereas in a fully Bayesian approach they are considered random variables with a suitable prior $p(\boldsymbol{\vartheta})$, also called hyperprior. In the latter usually MCMC based approaches are used for obtaining the posterior such as block-wise Gibbs sampling (see, e.g., Fahrmeir et al., 2013).

2.3.2 Nonparametric regression

While the presented mixed effects model allows to accurately analyze correlated longitudinal data, linear relationships are assumed between covariates and the response analogous to a simple linear regression where

$$E(\boldsymbol{y}) = \boldsymbol{x}\beta \tag{2.4}$$

with the observed response $\boldsymbol{y} = [y_1, \cdots, y_n]^\top$, the observed covariate $\boldsymbol{x} = [x_1, \cdots, x_n]^\top$ and the unknown regression coefficient β. Note that in eq. (2.4) and in the following presentations the conditional expectation $E(\boldsymbol{y}|\boldsymbol{x})$ is implied.
To relax the linearity assumption covariates can be transformed, as for example the polynomial transformation in the described mouse example. However such a polynomial modeling lacks flexibility and can lead to stability issues. Instead, the necessary flexibility, as in the given cohort data, can be achieved by nonparametric regression which allows estimating

$$E(\boldsymbol{y}) = f(\boldsymbol{x}) \tag{2.5}$$

where few restrictions are placed on the functional form of $f(\cdot)$, but often it is of interest to achieve a smooth function. Many approaches have been developed to estimate such smooth unspecified functions, such as LOESS smoothing (Cleveland et al., 1992), kernel smoothers (Wand and Jones, 1994), fractional polynomials (Royston and Altman, 1994) and several spline approaches (Fahrmeir et al., 2013; Wood, 2006).

We focus on penalized B-spline representations, P-splines (Eilers and Marx, 1996), in the following due to their excellent properties, which are also used in the flexible additive joint model. B-splines (d. Boor, 1978) are piecewise polynomials functions of degree l which are fused smoothly at knots $\kappa_1 < \cdots < \kappa_m$, the borders of the intervals, over which the piece-wise functions are defined. B-splines can be represented as linear combination (Fahrmeir et al., 2013)

$$f(x_i) = \sum_{d=1}^{D} \gamma_d B_d(x_i)$$

where γ_d represents the coefficient for the respective polynomial spline basis function $B_d(x_i)$ at the respective covariate value x_i. The B-spline basis of degree l is given by $D = m + l - 1$ basis functions which are defined recursively as

$$B_d^l(z) = \frac{z - \kappa_{d-l}}{\kappa_d - \kappa_{d-l}} B_{d-1}^{l-1}(z) + \frac{\kappa_{d+l} - z}{\kappa_{d+1} - \kappa_{d+1-l}} B_d^{l-1}(z).$$

B-splines have certain advantages (Fahrmeir et al., 2013) such as their local definition, i.e. the basis functions are positive only on the interval over $l+2$ knots thereby avoiding numerical problems, they have a straightforward derivative that makes use of the same basis coefficients γ_d (see eq. (A.1) in the Appendix) and for every z it holds $\sum_{d=1}^{D} B_d(z) = 1$. Despite the possibility to model smooth nonlinear functions $f(\cdot)$, B-splines can be represented within the linear model making use of the design matrix

$$\mathbf{B} = \begin{pmatrix} B_1(x_1) & \cdots & B_D(x_1) \\ \vdots & & \vdots \\ B_1(x_n) & \cdots & B_D(x_n) \end{pmatrix}$$

and the least squares criterion can be used to estimate the coefficients $\boldsymbol{\gamma} = [\gamma_1, \cdots, \gamma_D]^\top$. One disadvantage of B-splines is their dependency on the number of knots m determining the amount of flexibility in the estimation.

Depending on the number of knots, the function $f(\cdot)$ can become to wiggly, resulting in overfitting or too flat. To overcome this problem regularization techniques are used, such as for example shrinking the coefficients $\boldsymbol{\gamma}$ towards $\mathbf{0}$. In the framework of P-Splines (Eilers and Marx, 1996) an r-th order difference penalty is used, which penalizes differences in adjacent spline coefficients. The smoothing parameter λ controls the trade-off between flexibility and smoothness while typically many knots are used. For the equidistant knots a 2nd order penalty also corresponds to a roughness penalty on the second derivative of the spline $\int (f(z))^2 dz$, which penalizes a strong curvature of a function, thereby

inducing smoothness. The penalization of adjacent spline coefficients, results in the penalized least squares (PLS) criterion

$$PLS(\lambda) = \sum_{i=1}^{n} \left(y_i - \sum_{d=1}^{D} \gamma_d B_d(x_i) \right)^2 + \lambda \sum_{d=r+1}^{D} (\Delta_r \gamma_d)^2.$$

with recursively defined difference penalties $\Delta_r \gamma_d = \Delta_r \gamma_d - \Delta_{r-1}\gamma_{d-1}$ (Fahrmeir et al., 2013). This penalty can also be expressed in matrix notation by making use of difference matrices $\mathbf{D}_r$. Consider for example the first order difference matrix and the second order difference matrix, of which the latter is widely used

$$\mathbf{D}_1 = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{pmatrix} \qquad \mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{pmatrix}$$

which induce the penalty above as

$$\lambda \sum_{d=r+1}^{D} (\Delta_r \gamma_j)^2 = \lambda \boldsymbol{\gamma}^\top \mathbf{D}_r^\top \mathbf{D}_r \boldsymbol{\gamma} = \lambda \boldsymbol{\gamma}^\top \mathbf{K}_r \boldsymbol{\gamma}.$$

This penalty ensures smoothness across the spline with the amount of smoothness determined by the parameter λ, which can be found, for example, by minimizing the AIC criterion or cross-validation. Note that for a penalty of order r a large value of the smoothing parameter, i.e. $\lambda \to \infty$, results in a polynomial of degree $k-1$ (Eilers and Marx, 1996) such that a large penalty on the commonly used second order differences results in a straight line under strong penalization.

P-splines can also be estimated in a Bayesian framework as Bayesian P-splines (Lang and Brezger, 2004) where the coefficients γ_d also considered random variables and smoothness is induced by the choice of a corresponding prior. In more detail the difference penalties are replaced by their stochastic analogues, i.e. Gaussian random walk priors of the respective order as for example first and second order random walks:

$$\gamma_{dp} = \gamma_{d,p-1} + u_{dp} \quad \text{or} \quad \gamma_{dp} = 2\gamma_{d,p-1} - \gamma_{d,p-2} + u_{dp}$$

where $u_{dp} \sim N(0, \tau_d^2)$ are Gaussian errors and usually $\gamma_{d1} \propto const$ is used as a prior for the starting value γ_{d1} for first order random walks or for γ_{d1} and γ_{d2} for second order walks. When looking at the conditional distribution of γ_d in a first order random walk $\gamma_d|\gamma_{d-1}, \cdots, \gamma_1 \sim N(\gamma_{d-1}, \tau^2)$ we see that the conditional expectation for γ_d is γ_{d-1} and that deviations from this conditional expectation are larger for larger τ^2 (Fahrmeir et al., 2013). By this relationship τ^2 determines the amount of smoothness and corresponds to the inverse of the smoothing parameter λ in the previously presented

approach. An equivalent representation of the smoothness prior using random walks is therefore (Lang and Brezger, 2004)

$$\boldsymbol{\gamma}|\tau^2 \propto \exp -\frac{1}{2}\boldsymbol{\gamma}^\top \mathbf{K}\boldsymbol{\gamma} \tag{2.6}$$

where $\mathbf{K}$ is the appropriate penalty matrix as previously defined. As $\mathbf{K}$ is not of full rank, this prior is partially improper. For detail on the Bayesian inference of Bayesian P-splines we refer to chapters 3 and 5

2.3.3 Structured additive regression

In cohort data, as in our given applications, it is often necessary to account not only for multiple measurements per subject using random effects but also to model influences of covariates in a nonlinear way as presented in the previous subsection. Structured additive models allow to incorporate random and nonlinear effects, as well as other effects such as spatial effects or time-varying coefficients, in a unified framework. We briefly illustrate this general framework for modeling longitudinal data and revisit it in generalized form in the context of time-to-event models.

In general we model the conditional expectation $E(\boldsymbol{y}_i) = \eta(\boldsymbol{x}_i)$ with the potentially time-varying covariate $\boldsymbol{x}_i$ and the structured additive predictor $\eta(\boldsymbol{x}_i)$ as

$$\eta(\boldsymbol{x}_i) = \sum_{k=1}^{M_k} f_k(\boldsymbol{x}_i)$$

where the functions $f(\cdot)$ can represent linear parametric terms as in eq. (2.4), smooth effects as in eq. (2.5) or random effects. Random effects can be incorporated into the framework by an appropriate specification of the design matrix and a penalized estimation of the respective coefficients. We have already illustrated the penalized least squares view on smoothing splines. Linear mixed effects models can also be seen from this angle as the joint likelihood

$$\ell(\boldsymbol{\beta}, \boldsymbol{b}) = -\frac{1}{2}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b})^\top \mathbf{R}(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b}) - \frac{1}{2}\boldsymbol{b}^\top \mathbf{D}^{-1}\boldsymbol{b} \tag{2.7}$$

can be maximized using PLS with $\boldsymbol{b}^\top \mathbf{D}^{-1}\boldsymbol{b}$ penalizing deviations from $E(\boldsymbol{b}) = \mathbf{0}$:

$$PLS(\boldsymbol{\beta}, \boldsymbol{b}) = (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b})^\top \mathbf{R}(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{b}) + \boldsymbol{b}^\top \mathbf{D}^{-1}\boldsymbol{b}.$$

resulting in coefficient estimates which are shrinked towards 0 due to the penalty.

As an illustration consider random intercepts $\boldsymbol{b}_0 = [b_{01}, \cdots, b_{0n}]^\top$ with $b_{0i} \sim N(0, \tau_0^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$ in which the shrinkage of the random intercept coefficients can be expressed as

$$PLS(\boldsymbol{\beta}, \boldsymbol{b}) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{x}_{ij}^\top\boldsymbol{\beta} - b_{0i})^2 + \lambda\sum_{i=1}^{n} b_{0i}^2$$

with $\boldsymbol{y} = [\boldsymbol{y}_1^\top, \cdots, \boldsymbol{y}_n^\top]^\top$ and, importantly, $\lambda = \sigma^2/\tau_0^2$ (Fahrmeir et al., 2013). Here the amount of shrinkage λ is determined by the trade-off between the between-subject variance σ^2 and the within-subject variance τ_0^2. As previously explained smoothing and random effects structures can also be seen from a Bayesian viewpoint. In the given example of random intercepts the distributional assumption for the random effects $b_{0i} \sim N(0, \tau_0^2)$ translates to a prior distribution such that the posterior of the model in eq. (2.1) is

$$p(\boldsymbol{\beta}, \boldsymbol{b}|\boldsymbol{y}) \propto L(\boldsymbol{\beta}, \boldsymbol{b}) \prod_{i=1}^{n} \frac{1}{\sqrt{\tau_0^2}} \exp(-\frac{1}{2\tau_0^2} b_{0i}^2) \tag{2.8}$$

with $L(\boldsymbol{\beta}, \boldsymbol{b})$ as the Gaussian likelihood of the conditional model (see Fahrmeir et al., 2013).

This similarity in structure, i.e. the modeling of random effects and smoothing splines by penalized least squares or respective prior distributions, can be exploited for fitting penalized smooth terms within the mixed effects model framework. From a Bayesian perspective, however, the priors in eq. (2.6) are improper as $\text{rk}(\mathbf{K}) < \dim(\boldsymbol{\gamma})$ and consequently the inverse $\mathbf{K}^{-1}$ does not always exist. A reparameterization of the models such that the penalized smooth term is split into an unpenalized term and a penalized term with i.i.d random effects solves this problem (Kneib, 2005) and allows to use mixed effects model estimation and inference procedures. Consider for example the function $f_k(\boldsymbol{x}) = \mathbf{B}\boldsymbol{\gamma}$ which is a B-spline for a function of $\boldsymbol{x}$ to be smoothed as a Bayesian P-spline as in eq. (2.6) where $\text{rk}(\mathbf{K}) = h$ and $h < \dim(\boldsymbol{\gamma})$. The vector $\boldsymbol{\gamma}$ can be decomposed into an unpenalized and a penalized part as

$$\boldsymbol{\gamma} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{Z}}\boldsymbol{b}$$

so that $\mathbf{K}\tilde{\mathbf{X}} = 0$, meaning this part of the decomposition remains unpenalized with $\tilde{\boldsymbol{\beta}}$ as fixed effects, and $\tilde{\mathbf{Z}}^\top \mathbf{K} \tilde{\mathbf{Z}} = \boldsymbol{I}$, resulting in $\boldsymbol{b} \sim N(0, \tau^2 \boldsymbol{I}_h)$. For technical details on this reparameterization we refer for example to Wand (2003) and Kneib (2005).

Besides smooth and random effects terms this framework also allows the straightforward inclusion of varying coefficients terms which are an interaction between a continuous variable and a categorical variable x_i. In our model development especially time-varying coefficient terms $f(t) \cdot x_i$ are of interest, where the nonlinear effect of time t is allowed to vary for different levels of x_i. This effect term can be easily included in the basis function approach by modeling the term as

$$f(t) \cdot x_i = \sum_{d=1}^{D} \gamma_d B_d(t) x_i$$

and setting up the spline basis design matrix of the interaction $\mathbf{B}_{int}$ to contain the product, that is $\mathbf{B}_{int}[i, d] = B_d(t) \cdot x_i$ (see Fahrmeir et al., 2013).

The structured additive regression framework can be further generalized for responses which are not normally distributed by introducing the response function $h(\cdot)$ such that $E(\boldsymbol{y}_i) = h(\eta(\boldsymbol{x}_i))$ and adjusting the joint likelihood in eq. (2.7) and eq. (2.8) accordingly. These generalized structured additive regression models are the basic framework for the flexible survival models in Section 2.4.4 and the additive joint model that is developed in the subsequent chapters.

2.3.4 Missing data structure in longitudinal cohorts

Besides the necessity of appropriately modeling the correlated data structure and nonlinear effects in the longitudinal cohort data, a special missing data structure is the main complication in modeling longitudinal biomarkers. First, despite a rigid measurement schedule, subjects skip, miss or delay study visits and thus intermittent missing values occur as well as subject-specific observation times. Second, measurements of a subject are only observed until the event takes place and not afterwards. For example in modeling the autoantibody markers until the onset of T1D, autoantibody measurements are not sensible after the onset of the disease as the body's immune system reacts to the externally given insulin. Through this, the number of measurements and the length of the observations is systematically related to the disease onset.

In order to assess potential issues with missing data in cohort studies some notation is helpful to describe the missing data mechanism. Reconsider our observed longitudinal response $\boldsymbol{y}$, where we now acknowledge that only some measurements $\boldsymbol{y}_{obs}$ can be observed and others are missing $\boldsymbol{y}_{mis}$ such that the full data vector $\boldsymbol{y} = [\boldsymbol{y}_{obs}, \boldsymbol{y}_{mis}]$. The missing data pattern is described by the response indicator $\boldsymbol{r}$ for which $r_{ij} = 0$ if the random variable y_{ij} is observed and 1 if it is missing. The missing data mechanism is then characterized by the conditional distribution $f(\boldsymbol{r}|\boldsymbol{y}, \boldsymbol{\phi})$, where $\boldsymbol{\phi}$ denotes relevant model parameters, with three different missing data mechanisms (Rubin, 1976; Little and Rubin, 2002).

- Missing completely at random (MCAR) where the missingness does not depend on any missing or observed data of $\boldsymbol{y}$ such that $f(\boldsymbol{r}|\boldsymbol{y}, \boldsymbol{\phi}) = f(\boldsymbol{r}|\boldsymbol{\phi})$ for all $\boldsymbol{y}, \boldsymbol{\phi}$,
- Missing at random (MAR) where the missingness depends only on the observed components $\boldsymbol{y}_{obs}$ of the response such that $f(\boldsymbol{r}|\boldsymbol{y}, \boldsymbol{\phi}) = f(\boldsymbol{r}|\boldsymbol{y}_{obs}, \boldsymbol{\phi})$ for all $\boldsymbol{y}_{mis}, \boldsymbol{\phi}$,
- Missing Not at Random (MNAR) where the missingness depends on the missing values $\boldsymbol{y}_{mis}$ of the response such that $f(\boldsymbol{r}|\boldsymbol{y}, \boldsymbol{\phi}) = f(\boldsymbol{r}|\boldsymbol{y}_{obs}, \boldsymbol{y}_{mis}, \boldsymbol{\phi})$

Under MCAR the observed data $\boldsymbol{y}_{obs}$ can be regarded as random subsample of the full data and thus statistical procedures on the observed data are valid (Little and Rubin, 2002; Rizopoulos, 2012). Under MAR likelihood inference and Bayes inference is valid even if only the observed data $\boldsymbol{y}_{obs}$ is used (and the missingness is ignored) if the subvectors in $\boldsymbol{\phi} = [\boldsymbol{\phi}_r, \boldsymbol{\phi}_y]$, namely the parameter vector of the missingness model, $\boldsymbol{\phi}_r$, and the parameter vector of the measurement model, $\boldsymbol{\phi}_y$, are independent (see Little and Rubin, 2002, chapter 6). Ignoring MNAR, however, can distort the model estimations.

In the given longitudinal cohort data the presence of intermittent missingness can be seen as MCAR, if no systematic reasons for skipping or delaying visits is present, and as MAR, if for example the previous measurement determines if subjects will go to the next visit on time. As the mixed effects model approach does not require balanced data but can handle subject-specific measurement times this missingness mechanism does not further complicate the model estimation. In contrast, the informative censoring of longitudinal observations, i.e., that observations are only made until the event occurs, is a MNAR missing mechanism which has to be accounted for in the statistical analysis.

In order to derive valid inference statements from this data the joint distribution of measurement and the missingness process has to be used for an analysis. Different model-based approaches have been developed to make use of this joint distribution such as pattern mixture models, selection models and shared parameter models which factorize the joint distribution of the response and the missingness mechanism (Little and Rubin, 2002; Rizopoulos, 2012). For a closer look consider the probability distribution of the response $f(\boldsymbol{y}|\boldsymbol{\phi}_y)$ and the conditional probability distribution of the missingness $f(\boldsymbol{r}|\boldsymbol{y}, \boldsymbol{\phi}_r)$ for models where the observations are modeled as independent such that $f(\boldsymbol{r}, \boldsymbol{y}|\boldsymbol{\phi}) = \prod_{i=1}^{n} f(\boldsymbol{r}_i, \boldsymbol{y}_i|\boldsymbol{\phi})$. *Selection models* (Diggle and Kenward, 1994) factorize the joint distribution of response and missingness as

$$f(\boldsymbol{r}_i, \boldsymbol{y}_i|\boldsymbol{\phi}) = f(\boldsymbol{y}_i|\boldsymbol{\phi}_y) f(\boldsymbol{r}_i|\boldsymbol{y}_i, \boldsymbol{\phi}_r)$$

into a marginal distribution of $\boldsymbol{y}_i$ in the population and the incidence of missing data conditional on the data. *Pattern-mixture* models (Little, 1993, 1994) alternatively factorize the joint distribution as

$$f(\boldsymbol{r}_i, \boldsymbol{y}_i|\boldsymbol{\phi}) = f(\boldsymbol{y}_i|\boldsymbol{r}_i, \boldsymbol{\phi}_y) f(\boldsymbol{r}_i|\boldsymbol{\phi}_r),$$

where the first factor represents the distribution of $\boldsymbol{y}_i$ conditioned on the missingness and a marginal model of the missingness process. These models allow for a mixture of missingness patterns and different distributions of the response for each pattern. *Shared parameter models* (Wu and Carroll, 1988; Wu and Bailey, 1989) factorize and introduce a random effect to account for the association between the measurement and the missingness process

$$f(\boldsymbol{r}_i, \boldsymbol{y}_i|\boldsymbol{\phi}) = \int f(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\phi}_y) f(\boldsymbol{r}_i|\boldsymbol{b}_i, \boldsymbol{\phi}_r) f(\boldsymbol{b}_i|\boldsymbol{\phi}_b) d\boldsymbol{b}_i. \tag{2.9}$$

This approach is the basis of the shared parameter joint model of longitudinal and survival data that is explored in the remainder of this work. In these models a joint likelihood of a longitudinal model, here the measurement process, and a survival model, here the missingness process, with conditional independence given latent parameters, here $\boldsymbol{b}$, is used for estimation. These shared parameters are in most developments shared random effects (Rizopoulos, 2012) but can also be other latent parameters (see Section 2.5).

Mouse example. When reviewing the aforementioned mouse data example, a potential issue with the analysis is the informative censoring in the longitudinal trajectories. We therefore reanalyze the longitudinal model and corrected the estimation by means of a shared parameter joint model approach. In this alternative analysis the group difference between progressors and non-progressors is estimated as 1.29 (95% credibility interval: 0.74, 1.89), and is therefore slightly higher than in the previous analysis. However the results do not vary strongly, potentially due to the censoring of observations at 30 weeks which may alleviate the issue of informative missingness with the price of discarding some of the observed data.

2.4 Survival models

Often, a major focus in the analysis of cohort data is the modeling and/or prediction of an event of interest, such as disease onset or death. Within the field of survival analysis a variety of different models and estimations exist. In the following we will briefly present important concepts and models for estimating the time to a single, potentially right censored event. We present both simple models to include time-varying predictors in survival models and a structured additive regression approach to survival analysis.

2.4.1 Basic concepts

Survival analysis is concerned with the modeling and estimation of survival or, more precisely, the time until a certain event of interest occurs. In cohort studies sample units can be humans under treatment, but also mechanical units, as in the survival analysis in engineering. For every subject i the nonnegative time-to-event T_i is observed as response. The analysis of survival is often complicated by potential right-censoring, i.e., the dropout of a subject before the event occurred, which is indicated by the censoring indicator δ_i. This indicator is 1 if a subject experienced an event and 0 if it is right-censored. We denote the true uncensored event time as T_i^*, which is only observed if $\delta_i = 1$ and assume independence between the censoring and the event times. Important concepts that describe the distribution of this type of response are the probability density $f(t)$ of T_i^*, the distribution function $F(t)$ with

$$F(t) = Pr(T^* \leq t) = \int_0^t f(u)du,$$

and the survival (also called survivor) function $S(t)$, which is

$$S(t) = Pr(T^* > t) = 1 - F(t).$$

The survivor function is usually assumed to be continuous and differentiable. Furthermore, the hazard $h(t)$ with

$$h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq T^* < t + \Delta t | T^* \geq t)$$

describes the instantaneous risk of an event given that a subject survived until time t. From this the cumulative hazard

$$\Lambda(t) = \int_0^t h(u)du$$

can be computed which is closely related to the survival function as

$$S(t) = \exp(-\Lambda(t)). \tag{2.10}$$

2.4.2 Proportional hazards model

The Cox proportional hazards (PH) model (Cox, 1972) is a popular model to fit survival data due to its simplicity in setup and estimation. The hazard in the Cox PH model is specified as

$$h_i(t) = h_0(t) \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})$$

where $h_0(t)$ is the baseline hazard rate, identical for all subjects, and $\boldsymbol{\beta}$ is a vector of regression coefficients linearly relating the subject-specific covariates $\boldsymbol{x}_i$ to the log-hazard. In consequence, individual risk differences are completely determined by the predictor $\boldsymbol{x}_i^\top \boldsymbol{\beta}$, which is linked through the exponential function to ensure a nonnegative hazard. In more detail, the hazard ratio of two individuals with covariates $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ and $i \neq j$

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t) \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{h_0(t) \exp(\boldsymbol{x}_j^\top \boldsymbol{\beta})} = \frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{\exp(\boldsymbol{x}_j^\top \boldsymbol{\beta})} = \exp((\boldsymbol{x}_i - \boldsymbol{x}_j)^\top \boldsymbol{\beta})$$

is constant over time and independent from the mutual baseline hazard which results in proportional hazard rates. This property facilitates the use of a partial likelihood for estimation and inference

$$pl(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{\sum_{j \in R_i} \exp(\boldsymbol{x}_j^\top \boldsymbol{\beta})} \right)^{\delta_i}$$

leaving the baseline hazard completely unspecified with $R_i = \{j : T_j \geq T_i\}$ denoting the respective risk set for individual i. Coefficients can then be estimated using ML. As Andersen and Gill (1982) show, the partial likelihood estimates $\hat{\boldsymbol{\beta}}$ are asymptotically normal with mean $\boldsymbol{\beta}$ and covariance matrix given by the inverse observed Fisher information.

2.4.3 Time-varying covariates in survival models

Using the counting process notation (see Therneau and Grambsch, 2000, for an in-depth introduction) the Cox PH model was extended by Andersen and Gill (1982) to also include time-varying covariates $\boldsymbol{x}_i(t)$ in the hazard function of the *extended* Cox PH model

$$h_i(t) = h_0(t) \exp(\boldsymbol{x}_i^\top(t) \boldsymbol{\beta})$$

within the partial likelihood setting. This model can be used as a simple approach to incorporate a longitudinal biomarker into a survival model. However, all observed marker values that enter the model are assumed to be measured without error and time-constant between measurements.

Mouse example. Reconsidering the mouse study, the association between the longitudinal marker and the disease process was modeled as a marker level difference between mice progressing to the disease and those remaining event free. This question can, however, also be tackled by focusing

on the risk of progression in a survival analysis. Instead of dichotomizing subjects into progressors and non-progressors at a certain time point, the influence of the longitudinal biomarker on the progression risk over time can be assessed. Using this approach the full information of subjects, also after 30 weeks, can be taken into account. When using this simple approach and fitting the time-dependent Cox PH model on the full data, the hazard ratio of progression to diabetes between two subjects with one unit difference in the log-transformed IAA is 1.20 [95% confidence interval: 1.02, 1.41].

This approach, however, may contain severe bias as the special data structure of biomarkers as time-varying covariates is not taken into account. To look at this more closely the concepts of *external* and *internal* time-varying covariates have to be explored. Consider a single covariate $x(t)$ which is time-varying and can be observed at time t. As Kalbfleisch and Prentice (2002, Chapter 6) explain, external time-varying covariates are variables for which it holds that

$$f(x(t)|x(u), T^* \geq u) = f(x(t)|x(u), T^* = u) \tag{2.11}$$

for all u, t with $0 < u \leq t$, where $f(\cdot)$ denotes the density of the value of the longitudinal process at a certain time point. This means that the path of the covariate from u to t is independent of an event occurrence at time u. Examples of covariates which fulfill this equation are the medication in a case-control study of which the dose is planned in advance or air pollution measurements in a study of health. An equivalent expression of the previous formula is

$$Pr(T^* \in [u, t)|x(u), T^* > u) = Pr(T^* \in [u, t)|x(t), T^* > u) \tag{2.12}$$

meaning the probability of an event in the interval $[u, t)$ should be independent of a marker observation in the same interval. The two equivalent expressions eq. (2.11) and eq. (2.12) illustrate the two views on external covariates, namely a marker measurement has to be obtainable at time t despite an occurred event at time u and the presence of a marker measurement at time t should not allow statements on the probability of an event at this time. Internal time-varying covariates are those, for which the conditions eq. (2.11) and eq. (2.12) are not fulfilled.

In many cohort studies longitudinal biomarkers are only measured until an event occurs as described in Section 2.3.4. As the mere presence of a biomarker measurement is informative for the disease, i.e.

$$Pr(T^* > t|x(t)) = 1, \tag{2.13}$$

the above conditions are not fulfilled and the respective biomarkers are therefore *internal* time-varying covariates. As a result, the direct relationship between the hazard and the survivor function as stated in eq. (2.10) does not hold for internal covariates (see also Rizopoulos, 2012, Chapter 3).

In addition, internal covariates are often measured with error due to biological variation (Rizopoulos, 2012). Ignoring this and fitting an extended Cox PH model using time-dependent covariates may result in estimates of the association which are biased towards 0 (Prentice, 1982). Furthermore, in-

ternal covariates are only observed at discrete time points and fitted as if they were constant between study visits ("last observation carried forward") while their true latent process is continuous in time. To alleviate the last two issues, two-stage models have been used prior to the development of joint models. In these, a mixed effects model is estimated first for the longitudinal biomarker and its predictions are then inserted as a time-varying covariate in a survival model. While being computationally less demanding than the subsequently presented joint model, this approach is still biased (see for example Dafni and Tsiatis, 1998; Tsiatis and Davidian, 2001; Ye et al., 2008; Sweeting and Thompson, 2011).

2.4.4 Survival modeling using structured additive regression

Similar to the extensions to linear regression models in Section 2.3.3 also in the context of survival modeling further methods have been developed to allow relaxing certain assumptions. In the following we focus on developments allowing for relaxing the proportionality assumption by including time-varying effects and the inclusion of nonlinear effects of covariates (Kauermann, 2005; Kauermann et al., 2008; Kneib and Fahrmeir, 2007). Generalized structured additive regression provides a powerful framework to extend beyond the presented Cox-type models. The linear predictor of the Cox PH model is replaced by a flexible structured additive predictor (Kneib and Fahrmeir, 2007)

$$h_i(t) = \exp(\eta_i(t)) \text{ with } \eta_i(t) = \boldsymbol{v}_i^\top \boldsymbol{\gamma} + \sum_{l=1}^{L_g} g_l(t) u_{il} + \sum_{l=1}^{L_f} f_l(x_{il})$$

with linear effects $\boldsymbol{\gamma}$ of covariates $\boldsymbol{v}_i$, time-varying effects $g_l(t)$ of covariates u_{il}, also including the baseline hazard and non-linear effects $f_l(x_{il})$ of continuous covariates x_{il}. All these effects can be reformulated using a basis function representation such that all covariate values or spline representations of nonlinear effects can be represented as linear terms with design matrix $\mathbf{X}_l$ and coefficient vector $\tilde{\boldsymbol{\beta}}_l$ in the generic notation

$$\boldsymbol{\eta} = \mathbf{X}_1 \tilde{\boldsymbol{\beta}}_1 + \cdots + \mathbf{X}_p \tilde{\boldsymbol{\beta}}_p$$

Furthermore, a respective model mixed effects model formulation can be exploited to estimate the model. As demonstrated in 2.3.3 every coefficient $\tilde{\boldsymbol{\beta}}_l$ for the nonlinear and time-varying effects, can be decomposed into an unpenalized part $\boldsymbol{\beta}_l$ and a penalized part $\boldsymbol{b}_l$ by a reparameterization of the respective penalty matrix and design matrix.

As the proportionality assumption does not hold anymore under existence of time-varying effects the full likelihood is used for estimating this model instead of the partial likelihood. Under the assumption of non-informative censoring the likelihood of this model is

$$L(\boldsymbol{\beta}, \boldsymbol{b}) = \prod_{i=1}^{n} h_i(T_i)^{\delta_i} \exp\left(- \int_0^{T_i} h_i(t) dt \right),$$

where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^\top, \cdots, \boldsymbol{\beta}_p^\top]^\top$ and $\boldsymbol{b} = [\boldsymbol{b}_1^\top, \cdots, \boldsymbol{b}_q^\top]^\top$. After reparameterization of the predictor into an unpenalized and a penalized part both a frequentist penalized likelihood approach for fitting the model

$$l_{pen}(\boldsymbol{\beta}, \boldsymbol{b}) = l(\boldsymbol{\beta}, \boldsymbol{b}) - \sum_{l=1}^{p} \frac{1}{2\tau_l^2} \boldsymbol{b}_l^\top \mathbf{K}_l \boldsymbol{b}_l,$$

with penalty matrix $\mathbf{K}$, and a Bayesian estimation by specifying appropriate priors and hyperpriors as presented in 2.3.3 is feasible. Model selection is not easy in such a complex and flexible model and we refer to Hofner et al. (2011) for a model building strategy in this setup.

2.5 Joint models and beyond

The analysis of longitudinal covariates and time-to-event outcomes has been of considerate interest in many biomedical fields in the past 30 years. Early work on joint models and their precursors was motivated, for example, by the study of HIV progression where it was of interest to model the association between CD4 and the clinical outcome AIDS (Wulfsohn and Tsiatis, 1997; Dafni and Tsiatis, 1998; Tsiatis and Davidian, 2001). The challenges to correctly estimate informatively censored longitudinal data and to appropriately include this internal covariate, measured intermittently with error, in a survival model led to the development of joint models for longitudinal and survival data. Since then joint models have been developed to allow detailed insights in various biomedical research areas such as for example prostate cancer and PSA values (Taylor et al., 2013), breast cancer (Chi and Ibrahim, 2006), aneurysms (Sweeting and Thompson, 2011) or cardiovascular problems (Andrinopoulou et al., 2014). Currently, they increasingly find entry into applied research (see for an overview Sudell et al., 2016).

Joint models provide unbiased estimations by using the shared parameter approach eq. (2.9) in which a joint likelihood for the survival and longitudinal submodels is built under the assumption of conditional independence given the latent parameter that links the two models. Different approaches exist on how to specify this latent structure and the association between the submodels. One possible specification of a shared parameter model is the joint latent class model (see Proust-Lima et al., 2014, for a review) in which a heterogeneous population of different classes is assumed. The classes have different, class-specific marker trajectories and different risks, with the latent class membership linking the two submodels. Whereas this models allows for heterogeneous subpopulations and nonlinear association structures it is of limited value in assessing the detailed nature of the association between the longitudinal marker and the hazard. In the remainder of this work we therefore focus on shared random effects models, where random effects link the survival and longitudinal submodels.

In the following the standard joint model setup and frequentist estimation approaches are presented, followed by a brief overview of Bayesian joint model approaches. Finally, some light is shed on important areas of research in joint models before our newly developed structured additive joint model is presented in the subsequent chapters.

2.5.1 Joint model setup and estimation

The basic joint model using shared random effects, closely following Rizopoulos (2012), consists of a survival submodel where the hazard

$$h_i(t) = h_0(t) \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \alpha m_i(t))$$

is modeled by the baseline survival covariates $\boldsymbol{x}_i$ and the "true" longitudinal marker $m_i(t)$ where α denotes the association between the longitudinal and the time-to-event process. This marker is derived from the observations y_{ij} as

$$y_{ij} = m_i(t_{ij}) + \epsilon_{ij}$$

using a linear mixed effects model as in eq. (2.1) with $m_i(t_{ij}) = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{z}_{ij}\boldsymbol{b}_i$ and $\epsilon_{ij} \sim N(0, \sigma^2)$. In consequence, the marker enters the survival model as a covariate defined continuously in time without measurement error thereby alleviating some of the previously mentioned issues. The baseline hazard $h_0(t)$ is fully modeled as a positive function in time using, for example, parametric distributions such as a Weibull or exponential, piece-wise constant functions or spline-based approaches to avoid stability issues and bias in the estimation of standard errors (as illustrated in more detail in Rizopoulos, 2012, Chapter 4).

Adequate modeling of the MNAR missing mechanism in the longitudinal submodel and appropriate incorporation of the longitudinal information in the survival model is achieved by estimating the submodels in a shared parameter framework with the joint likelihood

$$p(\boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} \int p(T_i, \delta_i|\boldsymbol{b}_i, \boldsymbol{\theta}_t) \left[\prod_{j=1}^{n_i} p(y_{ij})|\boldsymbol{b}_i; \boldsymbol{\theta}_y) \right] p(\boldsymbol{b}_i, \boldsymbol{\theta}_b) d\boldsymbol{b}_i$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}_t, \boldsymbol{\theta}_y, \boldsymbol{\theta}_b]$ is the full parameter vector of the parameters from the survival model $\boldsymbol{\theta}_t$, the longitudinal model $\boldsymbol{\theta}_y$, and the random effects $\boldsymbol{\theta}_b$. This likelihood is based on assuming conditional independence of both, the two submodels and the multiple observations per subject, given the random effects. Furthermore, independent censoring and visiting times are assumed. The likelihood function for the survival model is

$$p(\boldsymbol{T}, \boldsymbol{\delta}|\boldsymbol{\theta}_t, \boldsymbol{b}_i) = h_i(T_i|\boldsymbol{\theta}_t, \boldsymbol{b}_i)^{\delta_i} \cdot \exp\left[- \int_0^{T_i} h_i(s|\boldsymbol{\theta}_t, \boldsymbol{b}_i) ds \right]$$

and for the longitudinal model

$$p(\boldsymbol{y}|\boldsymbol{\theta}_y, \boldsymbol{b}_i) = \prod_{j=1}^{n_i} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(- \frac{(y_{ij} - m_i(t_{ij}))^2}{2\sigma^2} \right)$$

and a multivariate normal density for $p(\boldsymbol{b}_i, \boldsymbol{\theta}_b)$.

In the frequentist setting estimation is based on maximizing this likelihood function. The initial estimation approach using EM algorithms as in Wulfsohn and Tsiatis (1997) and Henderson et al. (2000) continues to be used (Elashoff et al., 2007; Ding and Wang, 2008). As the algorithm shows slow convergence especially near the maximum Rizopoulos (2012) uses a hybrid approache in combination with Newton-Raphson to achieve faster convergence. A direct maximization using a Newton-Raphson-like algorithm, a modified Marquardt algorithm, is employed for estimating latent class joint models (Proust-Lima et al., 2007, 2009).

The estimation of the maximum likelihood is challenged by the necessary numerical integration for both the integral with respect to time in the survival likelihood and the integral with respect to the random effects in the score vector of the likelihood of which no analytical expressions exist. Especially the integration over the random effects becomes problematic when highly flexible subject-specific structures and therefore many random effects are necessary. This can lead to unstable or even unfeasible estimations in a frequentist setting (Gould et al., 2015). In this context Gaussian quadrature has received attention for approximating the integrals (Crowther et al., 2016).

Different software implementations have been developed for fitting joint models in a frequentist setting such as the R packages **JM** (Rizopoulos, 2010) and **joineR** (Philipson et al., 2017), as well as the stata module **stjm** (Crowther, 2012) and the SAS macro **JMFit** (Zhang et al., 2016). The packages differ largely in the flexibility that is allowed for specifying the baseline hazard, often only a piece-wise constant or simple parametric terms, the flexibility in the longitudinal trajectories, often only random slope random intercept specifications, and the association structures that are allowed.

2.5.2 Bayesian analysis for joint models

Many developments of joint models are made in a Bayesian framework, as Bayesian joint model approaches allow a straightforward model assessment, no need for asymptotic approximations for inference and the potential integration of prior knowledge or historical data (Gould et al., 2015). Importantly also, this approach allows for more flexibility when complex random effects structures are used. A look at the posterior of the joint model (Rizopoulos, 2016b)

$$p(\boldsymbol{\theta}, \boldsymbol{b}|\boldsymbol{y}, \boldsymbol{T}, \boldsymbol{\delta}) \propto \prod_{i=1}^{n} p(T_i, \delta_i|\boldsymbol{b}_i, \boldsymbol{\theta}_t) \left[\prod_{j=1}^{n_i} p(y_i j)|\boldsymbol{b}_i; \boldsymbol{\theta}_y) \right] p(\boldsymbol{b}_i, \boldsymbol{\theta}_b) p(\boldsymbol{\theta})$$

illustrates this where now also the parameters $\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}_y$ are assumed random variables with the prior distribution $p(\boldsymbol{\theta})$. By omitting the normalization constant, no integration over potentially high-dimensional random effects is necessary. This can easily be shown using Bayes theorem

$$
\begin{aligned}
p(\boldsymbol{\theta}, \boldsymbol{b} | \boldsymbol{y}, \boldsymbol{T}, \boldsymbol{\delta}) &= \frac{p(\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{y}, \boldsymbol{T}, \boldsymbol{\delta})}{p(\boldsymbol{y}, \boldsymbol{T}, \boldsymbol{\delta})} \\
&= \frac{p(\boldsymbol{y}, \boldsymbol{T}, \boldsymbol{\delta} | \boldsymbol{\theta}, \boldsymbol{b}) p(\boldsymbol{\theta} | \boldsymbol{b}) p(\boldsymbol{b})}{\int_{\boldsymbol{\theta}} \int_{\boldsymbol{b}} p(\boldsymbol{y}, \boldsymbol{T}, \boldsymbol{\delta} | \boldsymbol{\theta}, \boldsymbol{b}) p(\boldsymbol{\theta} | \boldsymbol{b}) p(\boldsymbol{b}) d\boldsymbol{b} d\boldsymbol{\theta}} \\
&\propto p(\boldsymbol{y}, \boldsymbol{T}, \boldsymbol{\delta} | \boldsymbol{\theta}, \boldsymbol{b}) p(\boldsymbol{\theta} | \boldsymbol{b}) p(\boldsymbol{b})
\end{aligned}
$$

Note that the differentiation between random parameters $\boldsymbol{\theta}$ and random effects $\boldsymbol{b}$ is used only for illustration of the differences in the frequentist and Bayesian approach.

For this posterior, however, no closed-form solution exists especially for the marginal model so that MCMC sampling approaches are employed. Starting from early work (Faucett and Thomas, 1996) up to current approaches (Jiang et al., 2015), Gibbs sampling has been a central approach in the Bayesian estimation of joint models. Often Gibbs sampling is combined with adaptive rejection sampling for parameters without closed-form but with log-concave full conditional distributions (Chi and Ibrahim, 2006; Brown et al., 2005), additional Metropolis-Hastings steps (Tang and Tang, 2015; Huang et al., 2010) or slice sampling (Rizopoulos, 2016b). Much work utilizes WinBUGS or OpenBUGS software (Gilks et al., 1994) to estimate newly developed approaches with code examples obtainable from authors (see, for example, Guo and Carlin, 2004; Sweeting and Thompson, 2011; He and Luo, 2016) thereby limiting, however, the transfer of the these developments into applied research. So far only one software implementation is available for the Bayesian estimation, the R package **JMbayes** (Rizopoulos, 2016b) which allows much flexibility in the specification of the baseline hazard, the individual longitudinal trajectories and the association. However, the package has some limitations with regard to the flexible specification of subject-specific trajectories and nonlinear effects. Further information on existing and newly developed software for the estimation of joint models can be found in Chapter 6.

Mouse example. Reviewing the mouse example again, we now model the association between longitudinal IAA and the progression to diabetes in a Bayesian joint model. In this joint model the estimated hazard ratio of progression to diabetes between two subjects with one unit difference in the log-transformed IAA is 2.43 [95% credibility interval: 1.74, 3.62] per transformed unit of IAA. This estimated effect is clearly larger than in the naive approach using a time-dependent Cox PH model. This result illustrates the aforementioned bias towards 0 that can occur by not accounting for the potential measurement error in the longitudinal biomarker and using the marker in a last observation carried forward approach.

2.5.3 Areas of research in shared random effects models

In the past two decades shared random effects joint models have been an area of intensive research and methods development. For detailed overviews on the topic we refer to the work of Tsiatis and Davidian (2004), Diggle et al. (2008), Rizopoulos (2012) and Gould et al. (2015) and present in the following a short overview of interesting areas of research.

Firstly, the above presented model where the current value of the modeled marker trajectory is associated linearly with the log-hazard is only one possibility for the parameterization of the association. Many different structures are possible such as, for example, the association with

- the current slope of the marker as $h_0(t)\exp(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \alpha m_i(t)')$
- the marker with a time lag Δt as $h_0(t)\exp(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \alpha m_i(t-\Delta t))$
- cumulative effects of the marker as $h_0(t)\exp(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \alpha\int_{t_l}^{t_u} m_i(t)')$ with t_l and t_u as lower and upper limits of the integration over time
- a marker and its interaction with baseline covariates $h_0(t)\exp(\boldsymbol{x}_{1i}^\top\boldsymbol{\beta} + \alpha(\boldsymbol{x}_{2i}\cdot m_i(t))$
- random effects only $h_0(t)\exp(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \boldsymbol{\alpha}^\top\boldsymbol{b}_i)$.

Further potential association structures are presented in Hickey et al. (2016). The last parameterization, including only the random effects, results in a time-constant survival model, except for the baseline hazard, and is therefore often used to facilitate the estimation (for example in Barrett et al., 2015; Martins et al., 2016; Waldmann et al., 2017). This paramterization, however, only has straightforward interpretations for random intercept and slope models. If spline representations of the subject-specific trajectories are used, the interpretation of the association is problematic. This renders this parameterization approach unfavorable in our case, when flexibility in the individual trajectories is needed. The choice of an appropriate association structure can be based on prior knowledge or based on model choice criteria such as AIC and BIC (Zhang et al., 2014), but also more elaborate approaches have been developed such as Bayesian model averaging to include combinations of different associations (Rizopoulos et al., 2014) and a Bayesian shrinkage approach (Andrinopoulou and Rizopoulos, 2016) in which different association structures are included in a model and shrinkage priors are placed on the association parameters to achieve a more parsimonious model.

With regard to the random effects structure many model developments focus on random intercept/random slope models (e.g. Sweeting and Thompson, 2011; Gueorguieva et al., 2012; Philipson et al., 2017) or parametric trajectories based on prior knowledge (Ibrahim et al., 2004; Taylor et al., 2013) while only some more recent approaches also allow more flexibility in the longitudinal model. For an overview of joint models which allow for subject-specific nonlinear trajectories we refer to the following chapter. Additionally, other approaches were developed to include further longitudinal response distributions in the joint model such as mixtures of distributions, e.g. for zero-inflated longitudinal data (Rizopoulos et al., 2008; Liu, 2009; Hatfield et al., 2012), or relaxing the normality assumption of the random effects (Tang and Tang, 2015; Tang et al., 2017). Furthermore, quantile regression joint models are developed to assess the association between quantiles of interest of the longitudinal model with the hazard (Farcomeni and Viviani, 2015). Many extensions have also been developed to account for competing risks (Huang et al., 2010), multiple events per subject (Henderson et al., 2000) or interval-censored outcomes (Gueorguieva et al., 2012). However, in our biomedical research setting the focus lies on a single right-censored event.

Along with current trends towards personalized medicine, joint models can gain insights beyond the unbiased and efficient estimation of longitudinal and survival models by allowing (potentially dynamic) prediction of the event of interest. In different joint model specifications, tools for dynamic predictions of patient outcomes have been developed which can be updated as soon as new longitudinal information is obtained. Examples of these approaches are from a frequentist shared random effects perspective Rizopoulos (2011) and Rizopoulos (2012), from the respective Bayesian perspective Rizopoulos et al. (2014) and for the latent class joint model developments by Proust-Lima and Taylor (2009). The online risk calculator on prostate cancer basing on PSA measurements is an interesting example how joint model developments can find entry into clinical practice and be used to communicate risk for patients (Taylor et al., 2013). While dynamic predictions are an interesting topic of ongoing research, focus in the present development of the flexible additive joint model lies in the exploration of association structures between T1D-specific autoantibodies and T1D onset to obtain more detailed insights into the mechanisms of disease progression. The prediction of T1D onset, however, might be an interesting field of future research.

2.6 Challenges and demands in modeling type 1 diabetes progression

Whereas the trajectories of autoantibodies have been studied (Parikka et al., 2012; Steck et al., 2011), no clear consensus exists regarding a parametric description of autoantibody trajectories over time. Instead, subjects' trajectories show highly nonlinear patterns over time and differ strongly between individuals as shown in Figure 2.2. For the adequate modeling of autoantibodies and their association with the onset of T1D consequently a highly flexible specification of the subject-specific trajectories is necessary.

Besides the flexibility in the longitudinal model also a generalization of the association is necessary to include a time-varying relationship between the biomarker and the time-to-event. T1D-specific autoantibodies are indicators for an ongoing immune process in which insulin-producing beta cells are gradually destroyed. It is plausible that the association between the autoantibodies and the hazard of T1D varies over time, as the immune system is constantly being regulated.

These necessary extensions for the detailed modeling of autoantibodies in the context of T1D have not been appropriately addressed in previous joint model developments and were therefore the main drivers for developing the joint modeling framework that is presented in the following chapters.

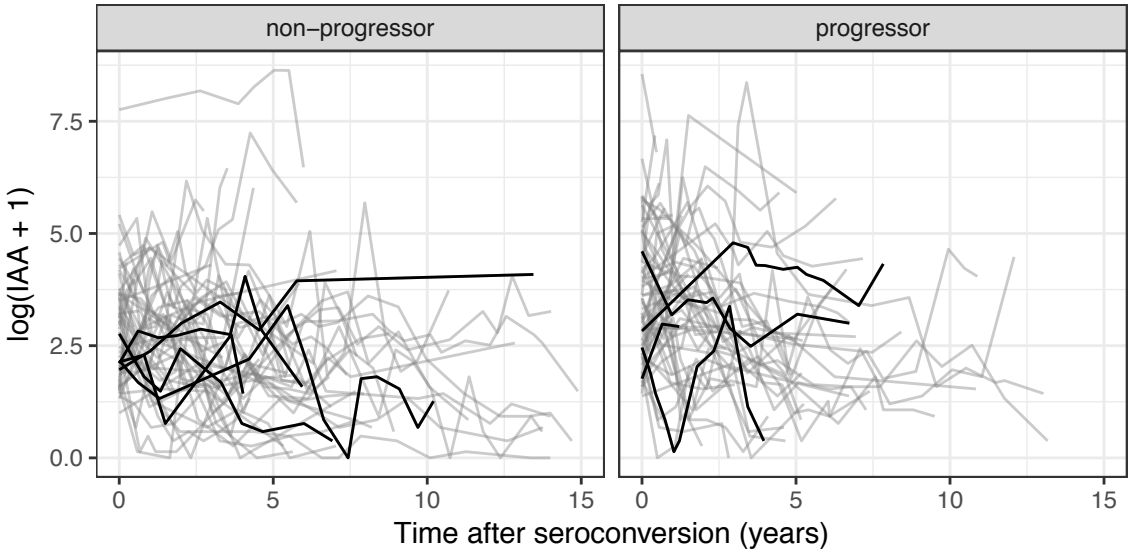

Figure 2.2: Observed linearly interpolated trajectories of the longitudinal marker values of $\log(IAA+1)$ in the BABYDIAB/BABYDIET study of subjects developing clinical T1D (progressor) and those remaining event free during follow-up (non-progressor) with a random selection of subjects highlighted.

Chapter 3

A general framework: Flexible Bayesian additive joint models

The joint modeling of longitudinal biomarkers and the time to disease onset or death offers unique insights into disease progression in various medical domains (Taylor et al., 2013; Gras et al., 2013; Daher Abdi et al., 2013). Depending on the disease and the respective biomarker different challenges have to be faced in joint modeling. In the following, a general framework for the flexible joint modeling of longitudinal data and time-to-event is presented, which was motivated by unique cohort data from studies exploring the development of T1D. The research on T1D underwent a paradigm shift in the past decade, when disease-specific autoantibodies were shown to be diagnostic for the disease before the onset of clinical symptoms and thus paving the way for a preclinical diagnosis of T1D (Ziegler et al., 2013; Bonifacio, 2015; Insel et al., 2015). Prior to the onset of clinical symptoms, i.e. the need of insulin substitution, the disease is already progressing and insulin-producing beta-cells in the pancreas are gradually destroyed by the body's own immune system. This immune process, leading to an onset of clinical symptoms within months up to more than a decade, can be diagnosed by the emergence of T1D-specific autoantibodies. However, it remains an open question whether the longitudinal patterns of these autoantibodies might be associated with the rate of progression to T1D.

In recent years joint models gained larger popularity in the modeling of associations between time-varying biomarkers and time-to-event. By estimating a submodel for a longitudinal biomarker, usually a mixed effects model, jointly with the survival submodel of a time-to-event process, one can account for the informative censoring and the within-subject errors in the longitudinal model and can incorporate the longitudinal information, observed only at person-specific discrete time points, as a continuous-time covariate in the survival model. Comprehensive overviews on the topic are given in Tsiatis and Davidian (2004), Rizopoulos (2012) and Gould et al. (2015). In our work we focus on

extensions of so-called shared parameter models. These assume that a set of parameters influences both the longitudinal and the survival model, and that there is conditional independence given those parameters.

In T1D research little is known concerning typical trajectories of autoantibodies as biomarkers. At the same time the observed trajectories show highly nonlinear patterns over time and differ strongly between subjects, see Figure 3.1a. In consequence, a flexible specification of individual trajectories in the longitudinal model is needed in our application.

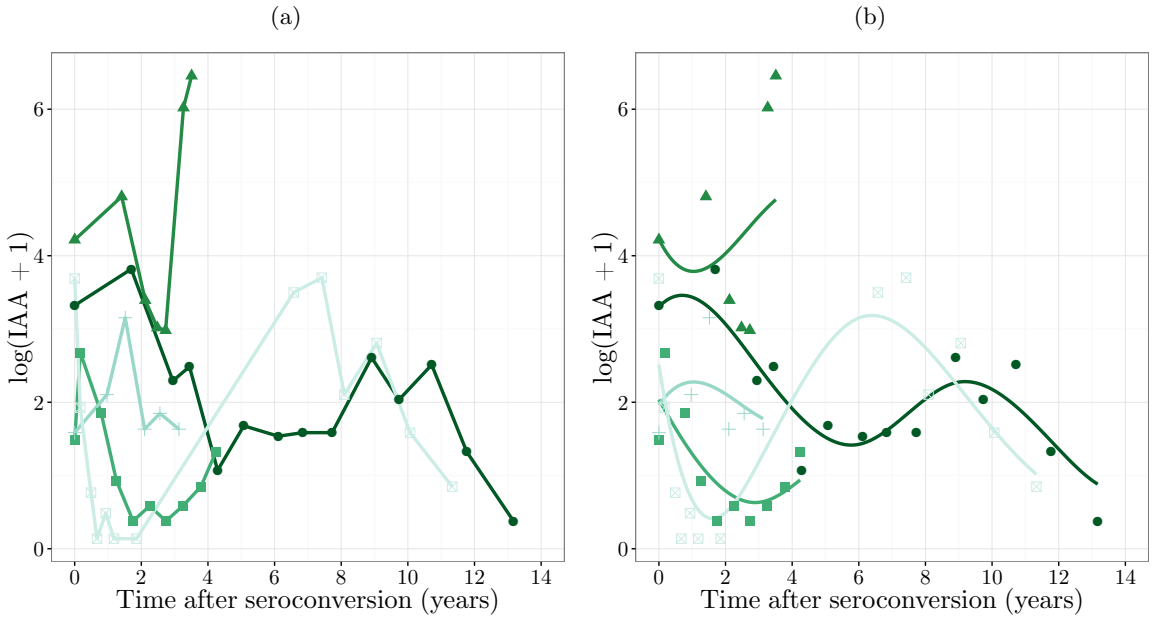

Figure 3.1: Longitudinal marker values of $\log(IAA + 1)$ for five randomly selected subjects in the BABYDIAB/BABYDIET study. (a) Observed values (points) and linear interpolation (lines); (b) Observed values (points) and estimated trajectories (lines).

Much work on joint models has focused on simple parametric longitudinal trajectories, while only few approaches allow for more flexible, potentially non-parametric longitudinal models. Ding and Wang (2008) model mean trajectories by B-splines and allow for one multiplicative random effect per subject. For our application however it remains questionable if such a model is flexible enough to capture the highly different trajectories. Spline based approaches, that allow also the random effects to be non-linear functions in time, are mentioned by Song and Wang (2008) and were employed by Rizopoulos and Ghosh (2011) and Rizopoulos et al. (2014) as well as Brown et al. (2005) and Brown (2009). While allowing for flexibility, a disadvantage of all these approaches is finding an optimal number of knots to specify the flexible longitudinal model, e.g. by AIC or DIC. As the number of random effects increases with the number of knots, this number is limited in practice. We aim to avoid the explicit choice of knots and number of basis functions by using a penalized spline approach, where a larger number of knots is specified and smoothness penalties are employed (Lang and Brezger,

2004). Tang and Tang (2015) also make use of P-Splines in modeling longitudinal trajectories, but do so only in estimating the mean function, whereas we model also the individual trajectories as smooth functions of time. This is similar in spirit to the specification of individual trajectories in Jiang et al. (2015), however we do not assume an underlying class membership for the random effects.

The estimation of joint models with complex subject-specific trajectories poses a challenge to frequentist estimation approaches due to the necessary integration over potentially high-dimensional random effects distributions. Due to this drawback and further advantages of the Bayesian approach in joint modeling, such as straightforward model assessment and the potential integration of previous knowledge via priors (Gould et al., 2015), many complex joint models, like e.g. the aforementioned models, are specified within a Bayesian framework. The most widely used sampling approach for the parameter distributions in Bayesian joint models is Gibbs sampling, e.g. Faucett and Thomas (1996); Guo and Carlin (2004); Brown and Ibrahim (2003), also in conjunction with Metropolis-Hastings algorithms (Tang and Tang, 2015). In addition, the well-established R package **JMbayes** (Rizopoulos, 2016a,b) implementing Rizopoulos et al. (2014) employs a random walk Metropolis-Hastings algorithm. Our Bayesian estimation approach is different as we employ a derivative-based Metropolis-Hastings algorithm, where we draw samples from approximations of the full conditionals using score vectors and Hessians of the parameters. Despite being computationally demanding this algorithm shows a high stability in the model estimation, as we also show in our simulations.

In addition to the need for a flexible longitudinal model, a further generalization of existing joint models seems necessary in our application, namely a time-varying association between the biomarker and the time-to-event. Here, the biomarker indicates an ongoing immune process eventually leading to the destruction of the insulin-producing beta cells. As the activity of the immune system is constantly regulated, it is plausible that the association between a biomarker and the hazard of T1D varies over time. For example a recent paper by Meyer et al. (2016) indicated that patients with an autoimmune disease can also present unique disease-ameliorating autoantibodies. Such a time-varying association has rarely been studied in the context of joint models. Using a discretized time-scale and a probit model for the discrete hazard function, Barrett et al. (2015) allow for the association to vary over the discrete time points in their model. However, this flexible specification is not considered in their simulations, the applied examples or the code provided to fit the models. A time-varying coefficient to associate the marker and the event process is the focus of the conditional score estimation approach in Song and Wang (2008). This approach can be seen as a weighted local partial likelihood without any assumptions on the distribution of the random effects. While this approach accounts for measurement error and short-term biological fluctuations in the longitudinal marker when modeling the hazard, it only permits inference on the survival parameters and not on the longitudinal model.

In order to allow for these two extensions, the flexible longitudinal trajectories and a potentially nonlinear time-varying association, both modeled by penalized splines, we develop and implement a highly flexible framework for joint models available within the R package **bamlss**. As we represent all parts of this flexible joint model as structured additive predictors, which can include linear, parametric but also nonparametric penalized terms, we are able to allow potentially nonlinear, smooth, random, and time-varying effects in both submodels. In consequence the possibilities of this implementation

go way beyond the two extensions that originally triggered the development. By applying this flexible model to the combined data set from two German high-risk T1D birth cohorts we aim to shed further light on the complex relationship between T1D-associated autoantibodies and the onset of clinical disease.

The remainder of this paper is structured as follows: The general model structure and potential extensions are outlined in Section 3.1. In Section 3.2, details on the Bayesian estimation procedure are given. A thorough testing of the model estimation through simulations is presented in Section 3.3 and the application to our T1D research question in Section 3.4. Concluding remarks are given in Section 3.5 and technical details can be found in the Appendix A.1. The presented model is implemented in the R package **bamlss** (Umlauf et al., 2017). Source code to reproduce the simulation results is available as Supporting Information of the published manuscript.

3.1 Model

In the following, the general setup for additive joint models is presented with a special focus on two extensions of existing approaches: the flexible specification of longitudinal trajectories as well as the time-varying association between the longitudinal marker and the event. An overview of potential further model specifications illustrates the flexibility of the presented model family.

3.1.1 General setup

For every subject $i = 1, \ldots, n$ we observe a potentially right-censored follow-up time T_i and the event indicator δ_i (1 if subject i experiences the event, 0 if it is censored). We model the hazard of an event at time t as

$$h_i(t) = \exp\{\eta_i(t)\} = \exp\{\eta_{\lambda i}(t) + \eta_{\gamma i} + \eta_{\alpha i}(t) \cdot \eta_{\mu i}(t)\} \tag{3.1}$$

including in the full predictor η a predictor η_λ for all survival covariates that are time-varying or have a time-varying coefficient including the log baseline hazard, a predictor for baseline survival covariates η_γ as well as a predictor η_α representing the potentially time-varying association between the longitudinal marker η_μ and the hazard.
We also observe a longitudinal response $\boldsymbol{y}_i = [y_{i1}, \cdots, y_{in_i}]^\top$ at the potentially subject-specific ordered time points $\mathbf{t}_i = [t_{i1}, \cdots, t_{in_i}]^\top$ with $t_{i1} \leq \cdots \leq t_{in_i} \leq T_i$. $\mathbf{t} = [\mathbf{t}_1^\top, \cdots, \mathbf{t}_n^\top]^\top$ denotes the vector of the $N = \sum_{i=1}^n n_i$ longitudinal measurement time points of all subjects. The longitudinal response at t_{ij} with $j = 1, \ldots, n_i$ is modeled as

$$y_{ij} = \eta_{\mu i}(t_{ij}) + \varepsilon_{ij} \tag{3.2}$$

with independent errors $\varepsilon_{ij} \sim N(0, \exp[\eta_{\sigma i}(t_{ij})]^2)$ allowing to also model the error variance. Thus, $\eta_{\mu i}(t_{ij})$ represents the longitudinally observed marker value without error at time point t_{ij}. This "true" marker value serves as a continuous-time covariate in the hazard in eq. (3.1) and links the

two model equations.

Each predictor η_{ki} with $k \in \{\lambda, \gamma, \alpha, \mu, \sigma\}$ is a structured additive predictor, i.e. a sum of M_k functions of covariates $\tilde{\boldsymbol{x}}_i$,[2]

$$\eta_{ki} = \sum_{m=1}^{M_k} f_{km}(\tilde{\boldsymbol{x}}_{ki}).$$

Different subsets $\tilde{\boldsymbol{x}}_{ki}$ of $\tilde{\boldsymbol{x}}_i$ can serve as covariates for the different predictors, with each f_{km} typically depending on one or two covariates. For time-varying predictors the functions can also depend on time $\eta_{ki}(t) = \sum_{m=1}^{M_k} f_{km}(\tilde{\boldsymbol{x}}_{ki}(t), t)$ with a potentially time-varying covariate vector $\tilde{\boldsymbol{x}}_{ki}(t)$. We express the vector of predictors for all subjects as $\boldsymbol{\eta}_k = [\eta_{k1}, \cdots, \eta_{kn}]^\top$. In the survival part of the model, see eq. (3.1), the vectors are of length n and potentially time-varying where $\boldsymbol{\eta}_k(t)$ denotes the evaluation at time t. In the longitudinal part of the model, see eq. (3.2), the vector $\boldsymbol{\eta}_k(\mathbf{t})$ is of length N, containing entries $\eta_{ki}(t_{ij})$ for all i and j, i.e. evaluations at all observed time points $\mathbf{t}$ for the corresponding subjects. A general overview and details on the setup of the predictor vectors in the submodels can be found in table 3.1.

Table 3.1: Overview of the predictor vectors, function evaluations and design matrices in the survival and longitudinal submodel.

	predictor vector	function evaluation	design matrix
survival model			
$k \in \{\gamma\}$	$\boldsymbol{\eta}_k = [\eta_{ki}]^\top$	$\boldsymbol{f}_k = [f_k(\tilde{\boldsymbol{x}}_{ki})]^\top$	$\mathbf{X}_k$
	$n \times 1$	$n \times 1$	$n \times p_k$
$k \in \{\lambda, \alpha, \mu\}$	$\boldsymbol{\eta}_k(t) = [\eta_{ki}(t)]^\top$	$\boldsymbol{f}_k(t) = [f_k(\tilde{\boldsymbol{x}}_{ki}(t), t)]^\top$	$\mathbf{X}_k(t)$
	$n \times 1$	$n \times 1$	$n \times p_k$
longitudinal model			
$k \in \{\mu, \sigma\}$	$\boldsymbol{\eta}_k(\mathbf{t}) = [\boldsymbol{\eta}_{ki}(\mathbf{t}_i)^\top]^\top$	$\boldsymbol{f}_k(\mathbf{t}) = [f_k(\tilde{\boldsymbol{x}}_{ki}(\mathbf{t}_i), \mathbf{t}_i)^\top]^\top$	$\mathbf{X}_k(\mathbf{t})$
	$N \times 1$	$N \times 1$	$N \times p_k$

For ease of notation we denote the vector $\mathbf{a}^\top = [a_1, \cdots, a_n]$ as $[a_i]$ for $i = 1, \ldots, n$ and drop the subscript m for the different terms per predictor in this illustration.

The functions $f_{km}(\tilde{\boldsymbol{x}}_{ki})$ can model a variety of effects, such as smooth, spatial, time-varying or random effects terms which can be expressed in a straightforward notation for every term m of predictor k by using suitable basis function expansions and corresponding penalties $\mathbf{P}_{km}$. In a generic setup we let

$$\boldsymbol{f}_{km} = \mathbf{X}_{km}\boldsymbol{\beta}_{km} \qquad \text{and} \qquad \mathbf{P}_{km} = \frac{1}{\tau_{km}^2}\boldsymbol{\beta}_{km}^\top \mathbf{K}_{km}\boldsymbol{\beta}_{km}, \tag{3.3}$$

with the vector of function evaluations $\boldsymbol{f}_{km}$ stacked over subjects, the design matrix $\mathbf{X}_{km}$, the coefficient vector $\boldsymbol{\beta}_{km}$, the penalty matrix $\mathbf{K}_{km}$ and the variance parameter τ_{km}^2 that controls the amount

[2]Note that in contrast to the original publication and in line with notation in subsequent chapters we denote the observed covariate vectors as $\tilde{\boldsymbol{x}}_i$ and the resulting design vectors as $\boldsymbol{x}_i$. The vectors can coincide for parametric terms.

of penalization of the respective term. In the Bayesian setting a penalization is imposed by specifying an appropriate prior distribution for the parameters, $\boldsymbol{\beta}_{km} \sim N(\mathbf{0}, [\frac{1}{\tau^2_{km}}\mathbf{K}_{km}]^-)$ with $\boldsymbol{A}^-$ denoting the generalized inverse of $\boldsymbol{A}$, as presented in more detail in section 3.2.3. Note that these basic penalties can be extended further as shown in more detail in the next subsection. In analogy to the differences in form in the generic vector of predictors, i.e. $\boldsymbol{\eta}_k$, $\boldsymbol{\eta}_k(t)$ and $\boldsymbol{\eta}_k(\mathbf{t})$, the form of the generic vectors of function evaluations $\boldsymbol{f}_{km}$ and the generic design matrices $\mathbf{X}_{km}$ also differs between predictors and submodels. We refer to Table 3.1 for further details.

We illustrate this setup by two important examples and the exemplary specification of a standard shared parameter joint model. First, smooth functions in time can be modeled using P-splines with a B-spline basis, $f_{km}(t) = \sum_{d=1}^{D} \beta_d B_d(t) =: \boldsymbol{x}_{km}^\top(t)\boldsymbol{\beta}_{km}$, and corresponding penalty matrix $\mathbf{K}_{km} = \mathbf{D}_r^\top \mathbf{D}_r$ with the r-th difference matrix $\mathbf{D}_r$ of appropriate dimension (Eilers and Marx, 1996). For Bayesian P-Splines, smoothing is induced by appropriate prior specification, where the difference penalties are replaced by their stochastic analogues, i.e. random walks (Lang and Brezger, 2004). Second, random intercepts in the longitudinal part are incorporated by specifying $\mathbf{X}_{km}$ as an $N \times n$ indicator matrix, where the ith column indicates which longitudinal measurements belong to subject i, $\boldsymbol{\beta}_{km} = [\beta_{km1}, \cdots, \beta_{kmn}]$ denotes the coefficient vector and an $n \times n$ identity matrix as penalty $\mathbf{K}_{km} = \boldsymbol{I}_n$ ensures $\beta_{kmi} \sim N(0, \tau^2_{km})$ independently. Further, although the model allows for highly flexible joint models, a simple shared parameter joint model is contained as a special case with

- P-spline log-baseline hazard $\boldsymbol{\eta}_\lambda(t) = \boldsymbol{f}_\lambda(t) = \mathbf{X}_\lambda(t)\boldsymbol{\beta}_\lambda$ with $\mathbf{K}_\lambda = \mathbf{D}_r^\top \mathbf{D}_r$: $\mathbf{X}_\lambda(t)$ contains n stacked replications of $\mathbf{x}_\lambda(t)$, the vector of D B-spline basis functions $B_d(t)$ at t

- parametric effects of baseline survival covariates $\boldsymbol{\eta}_\gamma = \boldsymbol{f}_\gamma = \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma$ with $\mathbf{K}_\gamma = \mathbf{0}$: $\mathbf{X}_\gamma$ contains the stacked subject-specific covariate vectors $\mathbf{x}_{\gamma i}$

- time-constant association between longitudinal and survival model $\boldsymbol{\eta}_\alpha = \boldsymbol{f}_\alpha = \mathbf{1}_n\beta_\alpha$ with $\mathbf{K}_\alpha = \mathbf{0}$

- longitudinal model with a random intercept $\boldsymbol{\eta}_\mu(\mathbf{t}) = \boldsymbol{f}_{\mu 1}(\boldsymbol{t}) + \boldsymbol{f}_{\mu 2}(\boldsymbol{t}) = \mathbf{X}_{\mu 1}(\boldsymbol{t})\boldsymbol{\beta}_{\mu 1} + \mathbf{X}_{\mu 2}(\boldsymbol{t})\boldsymbol{\beta}_{\mu 2}$ with $\mathbf{K}_{\mu 1} = \mathbf{0}$, $\mathbf{K}_{\mu 1} = \boldsymbol{I}_n$: $\mathbf{X}_{\mu 1}(\boldsymbol{t})$ is the fixed effects design matrix potentially including a parametric effect of time and $\mathbf{X}_{\mu 2}(\boldsymbol{t})$ the indicator matrix for a random intercept

- constant error variance $\boldsymbol{\eta}_\sigma(\mathbf{t}) = \boldsymbol{f}_\sigma(\mathbf{t}) = \mathbf{1}_N\beta_\sigma$ with $\mathbf{K}_\sigma = \mathbf{0}$

where $\mathbf{1}_n$ and $\mathbf{1}_N$ are vectors of ones of length n and N, respectively, and $\mathbf{0}$ is a zero matrix. Note that we drop the index m for predictors which consist of only one term.

3.1.2 Important extensions of current models

A special focus in our joint model approach lies on the flexibility of the longitudinal predictor η_μ. We model the trajectory for every subject as the sum of fixed covariate effects, a smooth function of time, a random intercept as well as smooth subject-specific deviations from this function over time,

$$\eta_{\mu i}(t) = f_{\mu 1}(t) + f_{\mu 2}(i) + f_{\mu 3}(t, i) + \sum_{m=4}^{M_\mu} f_{\mu m}(\tilde{\boldsymbol{x}}_{\mu m}). \tag{3.4}$$

In this parameterization $f_{\mu 1}(t)$ is a smooth effect of time and $f_{\mu 2}(i)$ is a random intercept. The term $f_{\mu 3}(t, i)$ denotes the smooth subject-specific deviations from the global time effect using functional random intercepts (Scheipl et al., 2015). Additionally, linear or parametric effects, including a global intercept, as well as further smooth effects of covariates can be represented by an extra term in $\sum_{m=4}^{M_\mu} f_{\mu m}(\tilde{\boldsymbol{x}}_{\mu m})$. The basis for the functional random intercepts can be specified within the basis function approach as row tensor products of the marginal basis of a random intercept, marked by the subscript s, and the marginal basis for a smooth effect of time, marked by the subscript t. We denote the vector of function evaluations at every observed longitudinal time point in $\mathbf{t}$ for the corresponding subjects in $\boldsymbol{i} = [1, \cdots, n]^\top$ as

$$f_{\mu 3}(\mathbf{t}, \boldsymbol{i}) = (\mathbf{X}_{\mu 3s} \odot \mathbf{X}_{\mu 3t})\boldsymbol{\beta}_{\mu 3} = \mathbf{X}_{\mu 3}\boldsymbol{\beta}_{\mu 3},$$

where $\mathbf{X}_{\mu 3s}$ is an $N \times n$ indicator matrix as the basis for a random intercept as specified for $\mathbf{X}_{\mu 2}$ in the previous sub-section, $\mathbf{X}_{\mu 3t}$ is an $N \times D$ matrix of evaluations of a marginal spline basis at $\mathbf{t}$ and $\mathbf{X}_{\mu 3}$ is the $N \times nD$ basis matrix resulting from the row tensor product. The row tensor product $\odot$ of a $p \times a$ matrix $\boldsymbol{A}$ and a $p \times b$ matrix $\boldsymbol{B}$ is defined as the $p \times ab$ matrix $\boldsymbol{A} \odot \boldsymbol{B} = (\boldsymbol{A} \otimes \mathbf{1}_b^\top) \cdot (\mathbf{1}_a^\top \otimes \boldsymbol{B})$ with $\cdot$ denoting element-wise multiplication and $\otimes$ denoting the Kronecker product.

The corresponding penalty term is constructed from the marginal penalty matrices:

$$\mathbf{P}_{\mu 3} = \boldsymbol{\beta}_{\mu 3}^\top \left(\frac{1}{\tau_{\mu 3s}^2} \mathbf{K}_{\mu 3s} \otimes \boldsymbol{I}_t + \frac{1}{\tau_{\mu 3t}^2} \boldsymbol{I}_s \otimes \mathbf{K}_{\mu 3t} \right) \boldsymbol{\beta}_{\mu 3} = \boldsymbol{\beta}_{\mu 3}^\top \left(\frac{1}{\tau_{\mu 3s}^2} \tilde{\boldsymbol{K}}_{\mu 3s} + \frac{1}{\tau_{\mu 3t}^2} \tilde{\boldsymbol{K}}_{\mu 3t} \right) \boldsymbol{\beta}_{\mu 3}, \tag{3.5}$$

where $\mathbf{K}_{\mu 3s} = \boldsymbol{I}_n$ is the penalty matrix for the random effect and $\mathbf{K}_{\mu 3t}$ is an appropriate penalty matrix for the smooth effect of time such as a difference penalty for B-splines. The enlarged penalty matrices $\tilde{\boldsymbol{K}}_{\mu 3s}$ and $\tilde{\boldsymbol{K}}_{\mu 3t}$ yield a penalization for every subject, resulting in a random effects structure and a smoothness penalization across time for each subject. Note that by specifying two variance parameters, $\tau_{\mu 3s}^2$ and $\tau_{\mu 3t}^2$, the amount of penalization can differ in the direction of time and across subjects, resulting in an anisotropic penalty. This specification allows for a highly flexible modeling of individual trajectories over time.

Given the specification of a separate global intercept and subject-specific random intercepts, the constraints $\int f_{\mu 1}(t)dt = 0$ and $\int f_{\mu 3}(t, i)dt = 0$ for every i are set in order to ensure identifiability. The necessary linear constraint $\int f_{\mu 1}(t)dt = 0$ is implemented for B-splines by transforming the

marginal basis $\mathbf{X}_{\mu 3t}$ into an $N \times (D-1)$ matrix $\dot{\boldsymbol{X}}_{\mu 3t}$ for which it holds that $\dot{\boldsymbol{X}}_{\mu 3t}\mathbf{1}_{D-1} = \mathbf{0}$ as shown in Wood (2006, chapter 1.8), and adjusting the penalty accordingly. Constructing the row tensor product $\mathbf{X}_{\mu 3}$ using the transformed marginal basis matrix $\dot{\boldsymbol{X}}_{\mu 3t}$ with correspondingly adjusted marginal penalty ensures that the identification constraint $\int f_{\mu 3}(t,i)dt = 0$ for every i is also fulfilled.

As a second extension to existing shared-parameter models we also specify the association between the longitudinal and the survival model as a structured additive predictor η_α. In consequence, this predictor can be modeled as a function of time and/or other covariates. Motivated by our applied research questions we model $\eta_\alpha(t) = f_\alpha(t)$ as a smooth function of time by using penalized splines, as specified for the baseline hazard. This allows us to find patterns beyond the standard joint model specification to explain the relationship between the longitudinal marker and the survival process. These patterns could for example be critical time windows in which a non-zero effect of η_α is present or a potential change in the direction of the association η_α over time.

3.1.3 Further potential specifications

The presented general framework of structured additive joint models allows for a variety of different effect specifications by making use of the flexibility of Bayesian structured additive regression models as well as adding functional extensions. Besides the presented smooth, time-varying, random effects and functional random intercept terms, a variety of further effects can be incorporated. Table 3.2 gives an overview of possible terms. All these terms can be specified by formulating the desired effect in a basis function representation with an appropriate penalty term. For details on the specification of such effects please refer to Fahrmeir et al. (2004); Scheipl et al. (2015); Wood (2006). Further details on the practical aspects within our implementation are given in section 3.2.4.

Table 3.2: Effects $f_{km}(\tilde{\boldsymbol{x}}_{ki})$ that can be specified within a predictor η_k in structured additive joint models; modified from a similar table in Scheipl et al. (2015).

covariate (subset of $\tilde{\boldsymbol{x}}$)	$f_{km}(\tilde{\boldsymbol{x}}_k)$ constant over t	$f_{km}(\tilde{\boldsymbol{x}}_k)$ varying over t
no covariate	scalar intercept $1 \cdot \beta$	smooth effect of time $f(t)$
scalar covariate z	linear effect $z \cdot \beta$	linear effect varying over time $z \cdot f(t)$
	smooth effect $f(z)$	smooth effect over time $f(z,t)$
spatial covariate(s) s	spatial effect $f(s)$	spatial effect over time $f(s,t)$
grouping variable g	random intercept β_g	functional random intercept $f_g(t)$
scalar and grouping variable	random slope $z \cdot \beta_g$	functional random slope $z \cdot f_g(t)$
vector of scalars $[z_1, z_2]$	linear interaction $z_1 \cdot z_2 \cdot \beta$	linear interaction over time $z_1 \cdot z_2 \cdot f(t)$
	varying coefficient $z_1 \cdot f(z_2)$	
	smooth effect $f(z_1, z_2)$	

3.2 Estimation

We estimate the model in a Bayesian framework using Newton-Raphson and MCMC algorithms.

3.2.1 Likelihood

Under the assumption of conditional independence of the survival outcomes $[T_i, \delta_i]$ and the longitudinal outcome $\boldsymbol{y}_i$, given the random effects, the likelihood of the specified joint model is the product of the two submodel likelihoods L^{surv} and L^{long} for the survival and the longitudinal model.

The log-likelihood of the survival part is

$$\ell^{\text{surv}}\left[\boldsymbol{\eta}_\lambda(\mathbf{T}), \boldsymbol{\eta}_\gamma, \boldsymbol{\eta}_\alpha\left(\mathbf{T}\right), \boldsymbol{\eta}_\mu\left(\mathbf{T}\right)|\mathbf{T}, \boldsymbol{\delta}\right] = \boldsymbol{\delta}^\top \boldsymbol{\eta}(\mathbf{T}) - \mathbf{1}_n^\top \boldsymbol{\Lambda}\left(\mathbf{T}\right), \tag{3.6}$$

where $\mathbf{T} = [T_1, \cdots, T_n]^\top$ and $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_n]^\top$. Here $\boldsymbol{\Lambda}(\mathbf{T}) = [\Lambda_1(T_1), \ldots, \Lambda_n(T_n)]^\top$ is the vector of the cumulative hazard rates $\Lambda_i(T_i) = \exp(\eta_{\gamma i}) \int_0^{T_i} \exp[\eta_{\lambda i}(u) + \eta_{\alpha i}(u) \cdot \eta_{\mu i}(u)]du$ and $\boldsymbol{\eta}(\mathbf{T}) = [\eta_1(T_1), \cdots, \eta_n(T_n)]$ denotes the vector of the full predictors evaluated at the subject-specific survival times. The additive predictors implicitely also depend on covariates and model parameters. The log-likelihood of the longitudinal part of the model is

$$\ell^{\text{long}}\left[\boldsymbol{\eta}_\mu\left(\mathbf{t}\right), \boldsymbol{\eta}_\sigma\left(\mathbf{t}\right)|\boldsymbol{y}\right] = -\frac{N}{2}\log(2\pi) - \mathbf{1}_N^\top \boldsymbol{\eta}_\sigma\left(\mathbf{t}\right) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{\eta}_\mu\left(\mathbf{t}\right))^\top \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{\eta}_\mu\left(\mathbf{t}\right)). \tag{3.7}$$

$\boldsymbol{\eta}_\mu(\mathbf{t})$ and $\boldsymbol{\eta}_\sigma(\mathbf{t})$ are the predictor vectors of length N corresponding to the longitudinal response $\boldsymbol{y} = [\boldsymbol{y}_1^\top, \cdots, \boldsymbol{y}_n^\top]^\top$ and $\boldsymbol{R} = \text{blockdiag}(\boldsymbol{R}_1, \cdots, \boldsymbol{R}_n)$, where $\boldsymbol{R}_i$ can reflect the error structure of interest. In our case, we assume $\boldsymbol{R}_i = \text{diag}(\exp[\eta_{\sigma i}(t_{i1})]^2, \cdots, \exp[\eta_{\sigma i}(t_{in_i})]^2)$ so that $\boldsymbol{R}$ reduces to a diagonal matrix.

3.2.2 Priors and posterior

In this general framework above, a variety of terms (cf. Table 3.2) can be specified using corresponding priors. For linear or parametric terms we use vague normal priors on the vectors of the regression coefficients, e.g. $\boldsymbol{\beta}_{km} \sim N(0, 1000^2)$, approximately corresponding to the precision matrices $\mathbf{K}_{km} = \mathbf{0}$ as presented above. Smooth and random effect terms are regularized by placing suitable multivariate normal priors on the coefficients

$$p(\boldsymbol{\beta}_{km}|\tau_{km}^2) \propto \left(\frac{1}{\tau_{km}^2}\right)^{\frac{\text{rank}(\mathbf{K}_{km})}{2}} \exp\left(-\frac{1}{2\tau_{km}^2}\boldsymbol{\beta}_{km}^\top \mathbf{K}_{km}\boldsymbol{\beta}_{km}\right)$$

with precision matrix $\mathbf{K}_{km}$ as specified in the penalty eq. (3.3). We use independent inverse Gamma hyperpriors $\tau_{km}^2 \sim IG(0.001, 0.001)$ to obtain an inverse Gamma full conditional for the variance parameters. In addition to the inverse gamma distribution, different priors are possible for the variance parameters in our implementation, such as half-Cauchy and half-normal distributions. The variance parameters τ_{km}^2 control the trade-off between flexibility and smoothness in the nonlinear

modeling of effects. As such they can be interpreted analogous to inverse smoothing parameters in a frequentist approach.

For anisotropic smooths, when multiple variance parameters $\boldsymbol{\tau}^2_{km} = (\tau^2_{kms}, \tau^2_{kmt})$ are involved as in eq. (3.5), we use the prior

$$p(\boldsymbol{\beta}_{km}|\boldsymbol{\tau}^2_{km}) \propto \left|\frac{1}{\tau^2_{kms}}\tilde{\boldsymbol{K}}_{kms} + \frac{1}{\tau^2_{kmt}}\tilde{\boldsymbol{K}}_{kmt}\right|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\beta}^\top_{km}\left[\frac{1}{\tau^2_{kms}}\tilde{\boldsymbol{K}}_{kms} + \frac{1}{\tau^2_{kmt}}\tilde{\boldsymbol{K}}_{kmt}\right]\boldsymbol{\beta}_{km}\right). \tag{3.8}$$

The resulting posterior of the model is

$$p(\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{\delta}, \boldsymbol{y}) \propto L^{\text{surv}}\left[\boldsymbol{\eta}_\lambda(\mathbf{T}), \boldsymbol{\eta}_\gamma, \boldsymbol{\eta}_\alpha(\mathbf{T}), \boldsymbol{\eta}_\mu(\mathbf{T})|\mathbf{T}, \boldsymbol{\delta}\right] \cdot L^{\text{long}}\left[\boldsymbol{\eta}_\mu(\mathbf{t}), \boldsymbol{\eta}_\sigma(\mathbf{t})|\boldsymbol{y}\right] \cdot \prod_{k\in\{\lambda,\gamma,\alpha,\mu,\sigma\}}\prod_{m=1}^{M_k}\left[p(\boldsymbol{\beta}_{km}|\boldsymbol{\tau}^2_{km})p(\boldsymbol{\tau}^2_{km})\right],$$

where $\boldsymbol{\theta}$ is the vector of all parameters in the model and $p(\boldsymbol{\beta}_{km}|\boldsymbol{\tau}^2_{km})$ and $p(\boldsymbol{\tau}^2_{km})$ are the priors for the regression coefficients and variance parameters for each term m and predictor k, respectively.

3.2.3 Bayesian estimation

Point estimates of $\boldsymbol{\theta}$ can be obtained by posterior mode and posterior mean estimation. We estimate the posterior mode by maximizing the log-posterior of the model using a Newton-Raphson procedure, the posterior mean is obtained via derivative-based Metropolis-Hastings sampling and thus computationally demanding. We therefore recommend to use posterior mode estimates for a first quick assessment of the model and in order to obtain starting values for the posterior mean sampling.

In the maximization of the log-posterior to obtain the posterior mode, we update blockwise each term m of predictor k in each iteration l as

$$\boldsymbol{\beta}^{[l+1]}_{km} = \boldsymbol{\beta}^{[l]}_{km} - \nu^{[l]}_{km}\boldsymbol{H}\left(\boldsymbol{\beta}^{[l]}_{km}\right)^{-1}\boldsymbol{s}\left(\boldsymbol{\beta}^{[l]}_{km}\right)$$

with potentially varying steplength $\nu^{[l]}_{km}$ and with the score vector $\boldsymbol{s}(\boldsymbol{\beta}_{km})$ and the Hessian $\boldsymbol{H}(\boldsymbol{\beta}_{km})$, which can be found in the Appendix. We optimize the variance parameters in each updating step to minimize the corrected AIC (AICc, Hurvich et al., 1998), which showed good performance in smoothing parameter estimation in Belitz and Lang (2008). Additionally, we optimize the steplength $\nu^{[l]}_{km}$ over $(0, 1]$ in each step to maximize the log-posterior. We assume the coefficients to have an approximately normal posterior distribution and derive credibility intervals from $N(\hat{\boldsymbol{\beta}}_{km}, [-\boldsymbol{H}(\hat{\boldsymbol{\beta}}_{km})]^{-1})$ for quick approximate inference.

For the posterior mean sampling we construct approximate full conditionals $\pi(\boldsymbol{\beta}_{km}|\cdot)$ based on a second order Taylor expansion of the log-posterior centered at the last state $\beta^{[l]}_{km}$, similar to Fahrmeir et al. (2004), Klein et al. (2015a) and Klein et al. (2015b) and as shown in more detail in Umlauf et al. (2017). The proposal density from this approximate full conditional is proportional to a multivariate normal distribution with the precision matrix $(\boldsymbol{\Sigma}^{[l]}_{km})^{-1} = -\boldsymbol{H}(\boldsymbol{\beta}^{[l]}_{km})$ and the mean $\boldsymbol{\mu}^{[l]}_{km} = \boldsymbol{\beta}^{[l]}_{km} -$

$\boldsymbol{H}(\boldsymbol{\beta}_{km}^{[l]})^{-1}\boldsymbol{s}(\boldsymbol{\beta}_{km}^{[l]})$. In each iteration l of the Metropolis-Hastings sampler and for updating block km a candidate $\boldsymbol{\beta}_{km}^{*}$ is drawn from the proposal density $q(\boldsymbol{\beta}_{km}^{*}|\boldsymbol{\beta}_{km}^{[l]}) = N(\boldsymbol{\mu}_{km}^{[l]}, \boldsymbol{\Sigma}_{km}^{[l]})$. By drawing candidates from a close approximation of the full conditional, we approximate a Gibbs sampler and achieve high acceptance rates and good mixing.

For the sampling of the variance parameters τ_{km}^2 Gibbs sampling is employed, as the full conditionals $\pi(\tau_{km}^2|\cdot)$ follow an inverse Gamma distribution, if inverse Gamma hyperpriors are used. Slice sampling is employed when no simple closed-form full conditional can be obtained as for example in the sampling of variance parameters for anisotropic smooths, see eq. (3.8), or for other hyperpriors.

Model selection can be conducted via DIC which is readily available within our implementation.

3.2.4 Implementation details

The model estimation is implemented within R (R Core Team, 2016) in the package **bamlss** (Umlauf et al., 2016, 2017) that allows the Bayesian estimation of a variety of models within the framework of Bayesian Additive Models for Location, Scale and Shape. The specification of appropriate design matrices and penalties for the desired effects is conducted internally via the R package **mgcv** (Wood, 2011). In consequence the full range of implemented smoothing approaches, such as P-splines, thin-plate regression splines, random effects, and Markov Random Fields, can be used within our implementation. We refer to Wood (2006) and Wood et al. (2016) for further information on model terms, bases and penalties. In our model specification in the simulations and the application we make use of Bayesian P-splines (Lang and Brezger, 2004) to model smooth effects. As the integrals in the survival likelihood as well as in the respective scores and Hessians have no analytical solution, they are approximated numerically using the trapezoidal rule and a fixed number of 25 integration points. Starting values for the posterior mean sampling are obtained by estimating the posterior mode of the joint model. For estimating the highly flexible functional random intercepts, the implementation makes use of the block diagonal structure of the design matrices to increase computation speed. The posterior mean sampling is implemented to potentially run in parallel on a number of specified cores on Linux systems. More details can be found in the documentation of the **bamlss** R package.

3.3 Simulation

We assess the estimation of our model by means of a simulation study with focus on two aspects: First, comparing our results with the established joint model implementation in **JMbayes** (Rizopoulos, 2016a) for models with time-constant η_α. Second, we want to assess the ability to model highly complex longitudinal trajectories as well as a time-varying effect of $\eta_\alpha(t)$, the two important new extensions within our framework. With this simulation we also aim to gain insights into the estimation quality of the model when applied to real data sets of T1D cohorts that motivated our methods development. Therefore, we simulate two differing data situations, mimicking real cohort data. The first simulated data setting, corresponding to the German cohort data presented in the Application Section, has fewer subjects, at more variably spaced time points but with a longer follow-up, than the

other. Finally, we aim to assess how well the posterior mode estimation can approximate the effects in comparison with the posterior mean estimates.

3.3.1 Simulation design

For every setting we generate longitudinal measurements for n subjects at a fixed grid of time points $\mathcal{P}$ based on a true longitudinal model $\eta_\mu(t)$ as specified in eq. (3.4) with the time effect $f_{\mu 1}(t) = 0.1(t+2)\exp(-0.075t)$, the random intercepts $f_{\mu 2}(i) = r_i$ where $r_i \sim N(0, 0.25)$, the functional random intercepts $f_{\mu 3}(t,i) = \mathbf{X}_{\mu 3}\boldsymbol{\beta}_{\mu 3}$, and the global intercept and covariate effect $f_{\mu 4}(\tilde{\boldsymbol{x}}_{\mu i}) = 0.5$ and $f_{\mu 5}(\tilde{\boldsymbol{x}}_{\mu i}) = 0.6\sin(\tilde{x}_{2i})$ with $\tilde{x}_{2i} \sim unif(-3,3)$. We simulate the functional random intercepts flexibly by P-Splines where we draw the true vector of spline-coefficients for all subjects from $\boldsymbol{\beta}_{\mu 3} \sim N(\mathbf{0}, [(1/\tau_s^2)\tilde{\boldsymbol{K}}_s + (1/\tau_t^2)\tilde{\boldsymbol{K}}_t]^{-1})$ as in eq. (3.5) with $\mathbf{K}_t = \mathbf{D}_2^\top\mathbf{D}_2$, $\tau_s^2 = 1$ and $\tau_t^2 = 0.2$. The hazard $h_i(t)$ for every subject is calculated according to eq. (3.1) using the true survival predictor functions $\eta_\lambda(t) = 1.4\log((t+10)/1000)$, $\eta_{\gamma i} = 0.5\sin(\tilde{x}_{1i})$, with $\tilde{x}_{1i} \sim unif(-3,3)$ and $\eta_\alpha(t)$ varying for the two simulation settings. Based on $h_i(t)$, survival times are generated for every subject as described in Bender et al. (2005) and Crowther and Lambert (2013). Every subject is censored after $\max(\mathcal{P})$ and we additionally apply uniform censoring $U(0, 1.5 \cdot \max(\mathcal{P}))$ to the survival times. In order to mimic missing measurements in the real data, p% of the remaining longitudinal data are randomly set to missing. Longitudinal observations are obtained from $\eta_{\mu i}(t)$ by adding independent errors $\epsilon_{ij} \sim N(0, 0.3^2)$ for each t_{ij} in $\mathbf{t}$. The influence of different data structures on the estimation is assessed by simulating two different data settings in each of the two simulations settings. In the smaller data setting, a, observations for $n_a = 150$ subjects are generated at the measurements points $\mathcal{P}_a = [0, 1, \ldots, 120]$ where $p_a = 75\%$ of the longitudinal measurements are missing and on average 108 (72 %) events occur, compared to $n_b = 300$ subjects at the time points $P_b = [0, 3, \ldots, 72]$ with $p_b = 10\%$ missings and 165 (55 %) events in the larger data setting, b.

In each data and simulation setting we draw $Q = 200$ samples. To ensure convergence, we run the model estimation with 23000 samples, a burn-in of 3000 and a thinning of 20, yielding 1000 samples, as assessed in preliminary simulations. For each estimated model q within a simulation setting we assess bias, mean-squared error (MSE) and frequentist coverage of the 95% credibility intervals, defined by the 2.5th and the 97.5th percentiles of the MCMC samples for the posterior mean and the approximate normal intervals for the posterior mode. We evaluate bias, MSE and coverage both averaged over all time points and averaged per time point. For the predictors in the longitudinal model, i.e. $k \in \{\mu, \sigma\}$, the average bias in each sample q is $B_k^q = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}[\hat{\eta}_{ki}^q(t_{ij}) - \eta_{ki}^q(t_{ij})]$ where $\hat{\eta}_{ki}$ denotes the estimate. To assess the model fit over time we also evaluate the bias per time point $B_k^q(t) = \frac{1}{n}\sum_{i=1}^{n}[\hat{\eta}_{ki}^q(t) - \eta_{ki}^q(t)]$ for all t in $\mathcal{P}$. The computations for MSE and coverage are analogous. For the survival predictors, i.e. $k \in \{\gamma, \lambda, \alpha\}$, the average bias is $B_k^q = \frac{1}{n}\sum_{i=1}^{n}[\hat{\eta}_{ki}^q(T_i) - \eta_{ki}^q(T_i)]$ using evaluations at the subject's event times. The bias of the time-varying survival predictor η_λ, and for setting 2 also η_α, is additionally evaluated at the fixed grid of time points t in $\mathcal{P}$ as above with MSE and coverage computed accordingly. These error measures are then averaged over all Q samples per setting.

For the comparison with the joint model implementation in **JMbayes** in settings 1a and 1b, data is generated with $\eta_\alpha(t) = 1$ as time-constant. In our implementation we model the longitudinal submodel by P-splines with cubic B-splines, a second order difference penalty and 12 knots (4 internal knots), for both the overall mean and the individual trajectories. After application of the constraints this yields $7 \cdot n$ basis functions. For the time-varying effect of the baseline hazard, η_λ, as well as the nonlinear effect in η_γ we use 10 knots (2 internal knots) resulting in 5 basis functions per effect after application of the constraints. In order to achieve a comparable model in the package **JMbayes** we model nonlinear effects in the longitudinal submodel and survival covariate effects by B-splines and determine the number of knots to minimize the DIC in preliminary simulations. Details on the inclusion of nonlinear effects in both submodels can be found in the source code of the Supporting Information of the published manuscript. As a result we model the longitudinal part by cubic B-splines for both the fixed and random effects with 1 internal knot for the larger data setting and without internal knots for the smaller data setting, resulting in 4 and 3 basis functions for both the fixed and random effects of time, respectively. As prior simulations had shown convergence issues using an unrestricted covariance matrix of the random effects, we restrict it to be diagonal, resulting in independent random effects. Also based on DIC from preliminary simulations we specify the effect in η_γ in the survival part with cubic B-splines with 3 internal knots using 5 basis functions. We model the baseline hazard with P-splines using the default settings from **JMbayes**, i.e. a cubic B-spline basis with 17 basis functions and a second order difference penalty. For the MCMC procedure we also use the default settings of 20000 iterations, including a burn-in of 3000 and a thinning such that 2000 samples are kept.

In our second simulation, i.e. settings 2a and 2b, we specify the longitudinal trajectories as before but generate data using a time-varying association predictor $\eta_\alpha(t) = \cos((t - 33)/33)$ for data in a and $\eta_\alpha(t) = \cos((t-20)/20)$ for b in order to achieve a similar shape despite a differing time scale. We fit the model using the same specification as in setting 1. Additionally, η_α is modeled as a P-spline with 10 knots (2 internal knots) resulting in 5 basis functions after application of the constraints.

3.3.2 Simulation results

The focus of the first simulation is the comparison with the package **JMbayes** regarding the accuracy of the modeling of the longitudinal trajectories and the time-constant association parameter η_α in settings 1a and 1b. Table 3.3 shows the MSE, bias and coverage for the estimation of η_α.

For both methods η_α is estimated more precisely and with a higher coverage in the larger data setting b compared to a. In both data settings **bamlss** achieves lower MSE, less bias and a higher coverage in the estimation of the association compared to **JMbayes**. For **JMbayes** the coverage for η_α is not satisfactory in both settings (0.840 and 0.890). As the further survival predictors, η_γ and η_λ, are parameterized differently in the two estimation methods with regard to the intercept term and sum-to-zero constraints, we assess only the prediction quality of $\eta_\lambda + \eta_\gamma$. We observe that **JMbayes** shows a higher bias in the estimation of the sum of these two predictors. Regarding the longitudinal submodel for η_μ both methods are fairly equal regarding the average MSE over the larger data setting (**bamlss**: 0.028 vs. **JMbayes**: 0.029), but our approach seems to be more precise

Table 3.3: Posterior mean simulation results from **bamlss** and results from **JMbayes** from setting 1 (time-constant η_α) for small (a) and large (b) data sets.

		MSE		bias		coverage	
		a	b	a	b	a	b
η_α	**bamlss**	0.032	0.016	0.003	−0.009	0.925	0.970
	JMbayes	0.049	0.021	0.100	0.048	0.840	0.890
$\eta_\gamma + \eta_\lambda$	**bamlss**	0.127	0.077	−0.007	0.011	0.935	0.946
	JMbayes	0.155	0.101	−0.095	−0.048	0.743	0.742
η_μ	**bamlss**	0.022	0.028	0.001	0.000	0.944	0.942
	JMbayes	0.031	0.029	−0.001	0.008	∗	∗
η_σ	**bamlss**	0.001	0.001	0.009	0.014	0.940	0.875
	JMbayes	0.007	0.002	0.080	0.039	∗	∗

∗ No credibility intervals and thus no coverage could be calculated for these predictors.

in the smaller data setting (**bamlss**: 0.022 vs. **JMbayes**: 0.031). To further understand the cause of this difference we look at the bias in the estimation of η_μ over the whole observed time course for the smaller data setting. As shown in Figure 3.2, **JMbayes** seems to underestimate some nonlinearity of the true predictor. Both methods show higher uncertainty for later time points when, due to

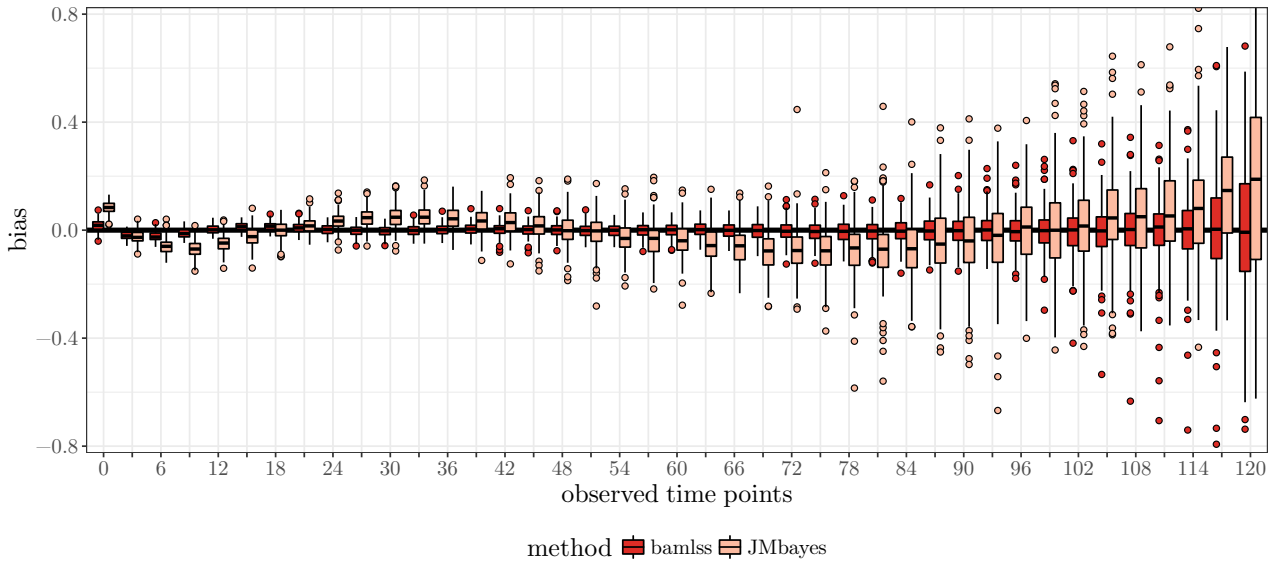

Figure 3.2: Comparison of the bias over time for $\eta_\mu(t)$ estimates from **bamlss** and **JMbayes** in setting 1a.

censoring and the occurrence of events, less information is available. For the longitudinal predictors we were not able to calculate credibility intervals in **JMbayes**. Finally, the estimation of the error variance is more precise in **bamlss**.

There are large runtime differences where **JMbayes** models took on average 9 minutes and 13 minutes for data setting a and b, respectively, and the implementation **bamlss**, due to the more flexible

functional random effects specification, took on average 2.5 hours and 5.1 hours to run on a single core of a 2.60 GHz Intel Xeon Processor E5-2650.

The aim of the second simulation setting is to shed light on the precision of the estimation of all predictors in the model with a special focus on the estimation of η_α, which is nonlinear in time. Additionally, we also compare the precision of the posterior mode to the posterior mean estimation. Table 3.4 gives an overview of the estimation precision of all predictors.

Similarly to setting 1 we observe an effect of sample size: All survival predictors $(\eta_\lambda, \eta_\gamma, \eta_\alpha)$ show

Table 3.4: Posterior mode and posterior mean simulation results for setting 2 (time-varying $\eta_\alpha(t)$) for small (a) and large (b) data sets.

		MSE		bias		coverage	
		a	b	a	b	a	b
η_α	mean	0.171	0.078	0.007	0.002	0.940	0.961
	mode	0.177	0.117	0.058	0.069	0.608	0.593
η_γ	mean	0.097	0.062	−0.035	−0.032	0.931	0.948
	mode	0.089	0.059	0.022	−0.001	0.804	0.795
η_λ	mean	0.083	0.065	0.000	0.000	0.945	0.957
	mode	0.101	0.082	0.000	0.000	0.592	0.549
η_μ	mean	0.022	0.028	0.000	0.000	0.943	0.942
	mode	0.025	0.031	0.000	0.000	0.882	0.865
η_σ	mean	0.001	0.001	0.009	0.015	0.905	0.855
	mode	0.004	0.004	−0.057	−0.057	0.175	0.045

a smaller MSE for data setting b compared to a probably due to the higher number of events. In contrast, the MSE is smaller for the estimation of η_μ in data setting a compared to b potentially due to the longer follow-up and a slightly higher number of longitudinal observations per subject. Whereas the precision of the point estimates is overall similar or only slightly worse for the posterior mode compared to the posterior mean estimation, the coverage is not acceptable for the posterior mode but close to 95% for the posterior mean. The only exception is the estimation of η_σ, where the coverage is somewhat lower for the posterior mean. As η_σ is very precisely estimated and formal inference is usually not of interest for this predictor, we do not rate this under-coverage as too problematic. In order to illustrate the precision in the time-varying effect estimates and to assess the cause of differences in MSE, Figure 3.3 displays the true and estimated predictors $\eta_\lambda(t)$, $\eta_\alpha(t)$ and their joint effect $\eta_\lambda(t) + \eta_\alpha(t) \cdot \frac{1}{n}\sum_{i=1}^{n} \eta_{\mu i}(t)$ evaluated at the mean of the estimated effect of $\eta_\mu(t)$ for all time-points. Overall the estimated predictors match the true functions quite well. For the smaller data sets there is more uncertainty in the estimation, especially at later time points, when less subjects are still observed.For example in setting a at a late time point ($t = 110$) on average only 6 subjects remain in the risk set as compared to 37 subjects at a late time point ($t = 65$) in setting b. Therefore, the estimates for later time points are highly dependent on the remaining subjects. This uncertainty is not only visible in the single predictors but also in their combined effect, see the right

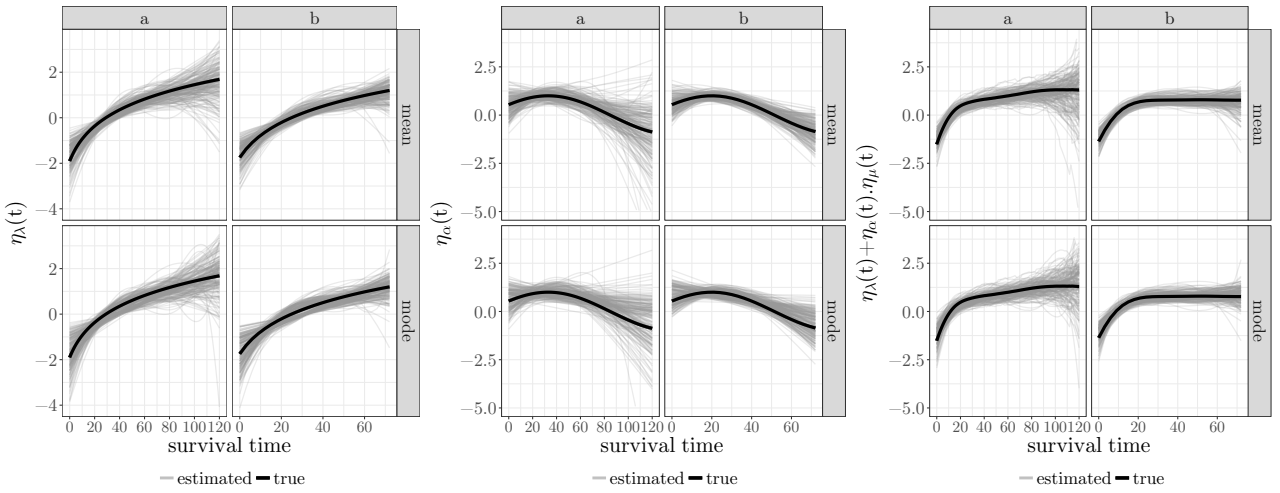

Figure 3.3: True (black) and estimated (grey) predictors from posterior mean and posterior mode estimation for small (a) and large (b) data sets in simulation setting 2. Left: $\eta_\lambda(t)$; middle: $\eta_\alpha(t)$; right: $\eta_\lambda(t) + \eta_\alpha(t) \cdot \frac{1}{n}\sum_{i=1}^{n} \eta_{\mu i}(t)$.

panel of Figure 3.3 (later time points in column a).
With on average only 10 and 18 minutes to run for data setting a and b respectively, the posterior mode estimation has clear advantages in computation time over the more precise posterior mean estimation with 2.6 and 5.2 hours on average in this setting.

In conclusion, our simulations show that the estimation of models with constant associations between marker and event performs well, even outperforming the implementation in **JMbayes** in some aspects. The estimation of more flexible models that are newly covered by our approach in contrast to existing implementations, i.e. with a time-varying association parameter and the specification of flexible trajectories, is equally satisfactory. While the more precise posterior mean estimation is time-consuming, the posterior mode offers a computationally efficient way to quickly assess the point estimates in a given model specification, even though credibility bands are only approximate.

3.4 Application

In order to gain insights into our motivating research question we apply the model to a combined data set of two ongoing German T1D risk cohorts to investigate whether longitudinal trajectories of insulin autoantibodies (IAA) are associated with the rate of progression to T1D. Whereas different autoantibodies are diagnostic for a preclinical stage of the disease, our focus lies on the analysis of the levels of IAA as a marker from the time when it first exceeded a specific threshold, called seroconversion, to the onset of T1D or loss to follow-up. The marker IAA is most often the first autoantibody to appear (Ziegler et al., 1993, 1999; Hummel et al., 2004). Both its initial value at seroconversion and its mean over time have been shown to be positively associated with the emergence

of T1D and negatively related to the age at T1D diagnosis (Steck et al., 2011, 2015). As seroconversion defines the beginning of a preclinical, symptom-free stage of T1D (Insel et al., 2015) we chose this time point as the starting point for our analysis instead of modeling the trajectories from birth and accounting for zeros up to the time point of seroconversion (Liu, 2009; Hatfield et al., 2012; Rizopoulos et al., 2008).

The BABYDIAB and BABYDIET studies, both prospective birth cohorts with a joint study protocol, aim to investigate the natural history of T1D development, i.e. explore environmental and genetic factors associated with the progression to T1D. In the BABYDIAB study 1650 offspring of patients with T1D were followed from birth to the development of T1D or loss to follow-up for up to 21 years (median: 15.7 years) (Ziegler et al., 1993, 1999; Hummel et al., 2004). Using the BABYDIAB study protocol, the BABYDIET study subsequently recruited 791 additional offspring or siblings of patients with T1D (Hummel et al., 2011). In both studies, autoantibody measurements were taken at age 9 months and 2, 5, 8, 11, 14 and 17 years and additionally every 6 months after positive islet autoantibodies had emerged. The exact age at the emergence of clinical T1D was assessed also between study visits.

In our joint model we use data of $n = 127$ children who developed IAA during follow-up, censored at 15 years after seroconversion due to the extremely low sample size at later time points, of which 69 (54%) subjects progressed to T1D (see Figure B.1 in the Appendix). The median progression time after seroconversion for IAA was 4.7 years (range: 0, 14.5 years; 25th and 75th percentile: 1.9, 6.7) with a median event-free follow-up of 5.9 years (range: 0, 15 years; 25th and 75th percentile: 3.4, 8.6). Subjects seroconverted for IAA at a median age of 2.1 years (range: 0.6, 18.1 years) with 86% of the subjects having IAA as the first autoantibody to appear. In total $N = 894$ longitudinal measurements of IAA after seroconversion (median number of measurements per subject: 6, range: 1, 24) were used and log-transformed as $\log(IAA + 1)$ for the analysis. We model subject's transformed autoantibody levels using functional random intercepts and two further covariates which were chosen a priori based on medical knowledge to fully capture the pattern of autoantibody levels. First, the age at seroconversion is included as a linear effect and second a binary variable indicates whether the autoantibody was among the first autoantibodies to appear. We model the association between marker and event, $\eta_\alpha(t)$, to be a non-linear function of time. Further we allow the covariates in the longitudinal model to also influence the survival process directly and expect a positive association between the age at seroconversion and the time to T1D (Steck et al., 2011; Ziegler et al., 2013). In our Bayesian model estimation we sample for 33000 iterations with a burnin of 3000 and thinning of 15 to obtain 2000 samples, with starting values for the posterior mean estimation obtained from the posterior mode estimates. Convergence is assessed by the inspection of traceplots, of which a subset is presented in the Appendix B.1. In order to assess the sensitivity of the results to the number of knots we specify three models with differing numbers of knots. We specify two models using either 12 (i.e. 4 internal) knots or 20 (i.e. 8 internal) knots for the overall mean and the individual trajectories in the functional random intercepts and 10 (i.e. 2 internal) knots in the survival submodel. Additionally, we specify a model with 20 (i.e. 8 internal) knots for nonlinear terms in both, the longitudinal and the survival submodels.

The results from the three specified models in our sensitivity analysis are highly similar for all predictors with regard to mean estimates and the credibility intervals. However, we observe lower DIC for the models with more knots in the functional random intercepts along with a closer fit of the individual trajectories and more narrow credibility intervals for the estimated association $\eta_\alpha(t)$ (cf. Figure B.2). Using more knots in the survival submodel results in a better mixing in the traceplots and a slightly lower DIC. Hence, we assume the results to be robust regarding the exact number of knots and present results of the model with the lowest DIC in the following.

As shown in Figure 3.1b for 5 randomly selected subjects, we are able to closely approximate the individual non-linear trajectories of IAA. The association between the marker and the onset of clinical T1D is estimated as stable over time with an average slope of -0.01 [95% credibility interval: -0.09, 0.07]. The average slope was defined as the mean over the first derivative of the association $\eta_\alpha(t)$ evaluated at all observed event and follow-up times $\mathbf{T}$, and its posterior distribution can be easily obtained by numerically deriving $\eta'_\alpha(t)$ in every sample. The credibility interval for the estimated association is above 0 for the first 2 years after seroconversion and there is more uncertainty at later time points, i.e. when less event and follow-up times are observed and when less subjects remain in the risk set, as indicated by the credibility intervals (Figure 3.4). In the longitudinal submodel we observe that trajectories have a lower level, if subjects seroconverted at an older age (in years, $\beta_{\mu 4} = -0.07$; 95% credibility interval: [-0.14, -0.004]) and a higher level if IAA was amongst the first markers to appear ($\beta_{\mu 5} = 0.89$; [0.23, 1.53]). In the survival submodel the log-hazard is decreased if IAA was amongst the first markers to appear ($\beta_{\gamma 2} = -0.94$; [-1.73, -0.12]). In sum if IAA is amongst the first markers to appear the log-hazard is reduced by 0.7. This net effect can be derived as the sum of the direct effect in η_γ and the indirect effect in $\eta_\alpha \cdot \eta_\mu$ with an average association of $\frac{1}{n}\sum_i \eta_\alpha(T_i) = 0.28$ Additionally, we do not observe a direct effect of the age at seroconversion ($\beta_{\gamma 3} = -0.09$; [-0.19, 0.01]).

As the estimated association $\eta_\alpha(t)$ showed no clear time trend we further compared this main model with a simpler model assuming a time-constant association using both our joint model implementation in **bamlss** and the implementation in **JMbayes**. In **bamlss** the same model was specified as in our main analysis with the exception that only an intercept was allowed in the predictor η_α. In **JMbayes** individual trajectories were modeled by cubic B-splines (see section 3.3). The models resulted in highly similar effect estimates for the covariate effects and the intercept of the association effect (cf. Figure 3.4). The effects on the log-hazard of the association and the covariate indicating that IAA was the first marker to appear were estimated slightly smaller in the **JMbayes** model. Despite more flexibility in the modeling, our main model achieved a slightly lower DIC than the time-constant **bamlss** fit (DIC: 2645.6 vs. 2648.9). As the DIC values are random and show small variations depending on the posterior sample this small difference in DIC should not be over interpreted. The DIC of the **JMbayes** estimation was not directly comparable due to a different likelihood formulation. We further conducted a sensitivity analysis regarding the prior specification of the variance parameters as well as the parametric terms. Basing on our main model we estimated 3 additional models (a) using $\tau^2_{km} \sim IG(0.001, 1)$, i.e. a distribution closer to a uniform for the variance parameters, (b) a

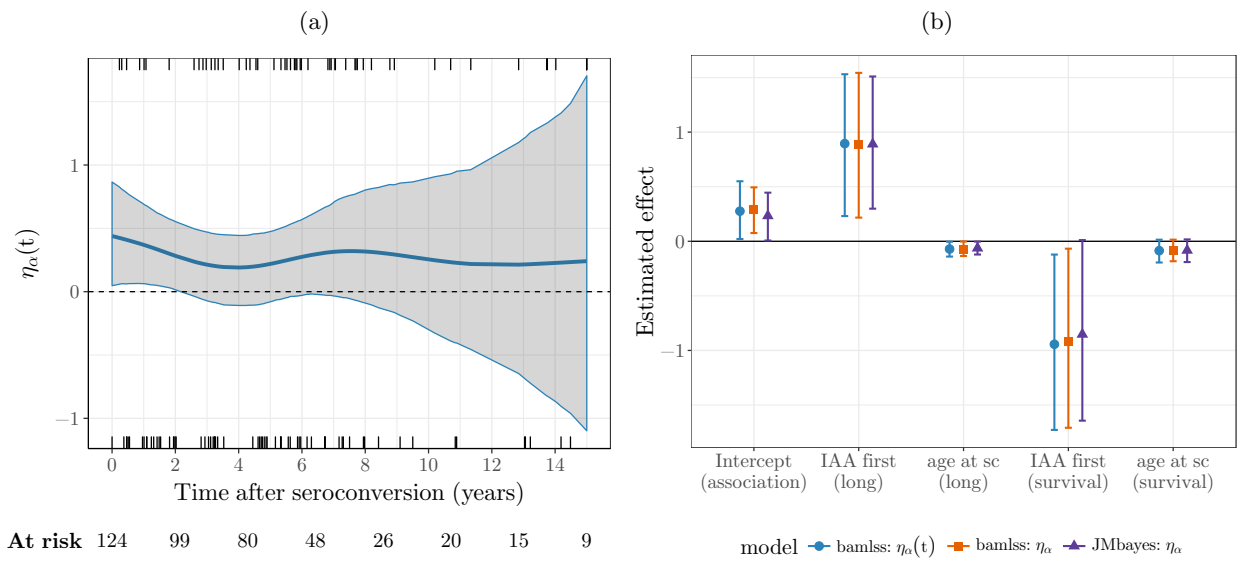

Figure 3.4: Estimated effects in the BABYDIAB/BABYDIET data. (a) Estimated posterior mean of $\eta_\alpha(t)$ with 95% pointwise credibility bands (shaded area), observed event times (rugs bottom), and censoring times (rugs top). (b) linear effect estimates from the survival and longitudinal submodel from model fits based on (i) **bamlss** with time-varying $\eta_\alpha(t)$, (ii) **bamlss** with time-constant η_α and (iii) **JMbayes**. *association* stands for effects in η_α, *long* for the longitudinal submodel, *sc* indicates seroconversion.

half-Cauchy distribution for the variance parameters, and (c) using a slightly more informative prior $\boldsymbol{\beta}_{km} \sim N(0, 50^2)$ for all linear terms. Again all estimated effects were highly similar (cf. Figure B.6).

In line with previous findings (Steck et al., 2011, 2015) these results indicate that the quantitative levels of the marker IAA are informative for the rate of progression to T1D in the first years after seroconversion with higher levels increasing the hazard of T1D. The direct relationship between the hazard and age at seroconversion is not supported by the model, suggesting that the previously established influence of this baseline covariate on T1D progression may be mediated by the marker levels, i.e. the effect in the respective log-hazard is reduced if the marker levels over time are taken into account as in our flexible parameterization. The association between IAA and the hazard of T1D over time was not found to be strongly time-varying in this relatively small study despite a slightly smaller DIC for the time-varying model. There is much uncertainty around the nonlinear time-varying estimate of the association $\eta_\alpha(t)$. This uncertainty is potentially a result of the flexibility in the estimation in combination with the amount of data in the survival part. Furthermore, the precision in the assessment of the exact age at seroconversion and therefore the time point 0 per subject is limited by the timing of the follow-up visits, potentially inducing further noise in the assessment of the time-varying association between the biomarker and the time to clinical T1D. All estimated effects were robust regarding different model specifications, i.e. varying number of knots, differently specified associations, as well as different prior distributions.

3.5 Discussion and outlook

We presented a flexible joint model that allows to fit a broad range of joint model specifications using structured additive predictors for all model components. The approach is fully implemented in the R package **bamlss**. While the framework is very flexible, as illustrated by Table 3.2, the focus in this work lies on the flexible modeling of individual trajectories and the specification of a time-varying association between marker and event. The proposed model shows satisfactory performance in various simulation settings and has the potential to offer new insights into complex relationships between biomarkers and time-to-event processes.
Our methods development was motivated by a specific research question from T1D studies and data from two cohorts, the presented combined German BABYDIAB/BABYDIET cohort as well as a multinational cohort. We aimed at assessing the potentially nonlinear time-varying association between a highly variable disease-specific marker and the time to T1D onset. We saw that even by using the combined BABYDIAB/BABYDIET cohort, the sample size of the data set considered in the application in Section 3.4 is at the lower limit for the complexity of our model, as indicated by our simulation study and by the width of the credibility intervals in the applied results. Nevertheless we found a positive association between a disease-related biomarker and the occurrence of clinical T1D. Although our model allows for a time-varying association between the biomarker and the event process, at least in this small data set it was estimated to be roughly constant with a DIC comparable to that of a simpler constant model and a slight larger effect early after seroconversion. In consequence our flexible model can also be used to check the modeling assumptions of simpler models that are commonly used. We aim to further explore the relationship between T1D-specific autoantibodies and the progression to T1D in the larger data set from the multinational T1D cohort (with sample size exceeding data setting b in our simulations) as in Steck et al. (2015). We note that these results, as the majority of findings on preclinical T1D, can only be generalized to subjects with increased risk of T1D and not to the general population.

Due to the complexity of the model and its estimation, the computation speed is still a drawback in our implementation. Hence, we are constantly working on speeding up the computations further. As shown in simulation 2, the posterior mode estimation offers a computationally efficient way to obtain point estimates from a flexible joint model before starting the full MCMC sampling. These posterior mode estimates show a precision similar to that of the posterior mean estimates. However, the credibility intervals obtained from posterior modes are not wide enough, potentially due to the fact that the uncertainty around the variance parameters τ^2_{km} is not included in the credibility intervals. In consequence, only the credibility intervals of the posterior mean estimates should be used for inference.
As is well-known in the survival context, the number of potential parameters in the model is limited by the number of observed events (Harrell et al., 1996). This also holds in our approach for the predictors in the survival part of the model, η_λ, η_γ, and η_α. We achieve to alleviate this issue to some extent by the penalized approach, which decreases the effective number of degrees of freedom and thus

allows for a richer model than would be possible without a penalty. Still, we recommend modeling only those effects as non-linear functions, where a strong indication for non-linearity is given.

Within the framework of the presented additive joint model several further extensions are possible. So far model selection is conducted via DIC. We note that more advanced model selection techniques such as Bayesian Lasso selection (Tang et al., 2017) or boosting (Waldmann et al., 2017) have been developed. Including these techniques into the presented framework are topics for future work. Regarding potential effects to be specified we aim to extend the model by including the derivative of the longitudinal trajectories to model the event process similar to Ye et al. (2008), Brown (2009) and Rizopoulos et al. (2014), allowing to model the potentially time-varying association between changes in the marker and the hazard. Further, functional historical effects of the trajectories, including information on the history of the marker (Malfait and Ramsay, 2003; Gellar et al., 2014), could potentially offer additional insights into complex relationships between markers and event processes.

Chapter 4

Joint modeling of longitudinal autoantibody patterns and progression to type 1 diabetes

T1D is one of the most common chronic diseases in childhood, with worldwide increasing incidence (Patterson et al., 2009). The disease is preceded by a preclinical period of islet autoimmunity, which most commonly develops in early infancy (Parikka et al., 2012; Ziegler et al., 2012). The presence of islet autoantibodies is associated with the progression to clinical diabetes (Ziegler et al., 2013). However, the time from the first emergence of autoantibodies, called seroconversion, to the onset of clinical symptoms varies considerably between individuals, ranging from weeks to decades (Ziegler et al., 2013). It is also known that the combination of different autoantibodies as well as the autoantibody titer is associated with progression time (Achenbach et al., 2004). For insulin autoantibodies (IAA), both their titers around seroconversion and their mean levels over time have been found to be associated with progression to T1D (Parikka et al., 2012; Steck et al., 2011), and similar findings have been recently reported for other islet autoantibodies (Steck et al., 2015, 2016; Endesfelder et al., 2016). Nevertheless, detailed analyses of autoantibody titers over time are lacking. Here, we investigated data of more than 600 islet-autoantibody positive children followed up within the prospective The Environmental Determinants of Diabetes in the Young (TEDDY) study (The TEDDY Study Group, 2007; Krischer et al., 2015). In contrast to previous analyses, we used joint models of longitudinal and survival data. This class of models has the advantage to avoid potential bias due to characteristics of the longitudinal markers (here autoantibodies), such as random biological fluctuations, informative censoring and discrete measurement time points (Asar et al., 2015). By applying a novel approach of joint modeling, we gained further insights into the potentially complex relationship between lon-

gitudinal islet autoantibody measures and the time to T1D progression, particularly with respect to time-varying associations of both.

4.1 Methods

TEDDY is an ongoing prospective cohort study funded by the National Institutes of Health with the primary goal to identify environmental causes of T1D. The TEDDY study enrolled 8,676 children with increased genetic risk for T1D who were recruited in six clinical research centers located in the USA, Finland, Germany, and Sweden between 2004 and 2010 shortly after birth. Detailed information on study design, eligibility and methods has been previously published (The TEDDY Study Group, 2007, 2008; Hagopian et al., 2011). Written informed consents were obtained for all participants from a parent or primary caretaker, separately, for genetic screening and for participation in prospective follow-up before inclusion in the study. The study was approved by local Institutional Review or Ethics Boards and is monitored by the External Advisory Board formed by the National Institutes of Health. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008 (5). For this analysis, we used the data of all children who had developed one or more persistent islet autoantibodies by the time of our data access (31 December 2014). At that time point the median age of the children analyzed at their last visit was 6.5 years with a range from 0.75 to 10.2 years.

4.1.1 Definition of islet autoimmunity

Development of persistent islet autoimmunity was assessed every 3 months and defined by the presence of at least one islet autoantibody among autoantibodies to insulin (IAA), glutamic acid decarboxylase (GADA), and insulinoma-associated protein 2 (IA2A) on two or more consecutive visits confirmed by two laboratories. Date of persistent autoimmunity to an autoantibody was defined as the draw date of the first sample of the two consecutive samples which deemed the child persistent confirmed positive for this autoantibody. As described in more detail elsewhere (Steck et al., 2011), the respective autoantibody titers were standardized to be comparable across study laboratories (University of Bristol, UK; and University of Colorado, Denver, US) by subtracting the laboratory- and antibody-specific threshold and dividing by the laboratory- and autoantibody-specific standard deviation, and were log-transformed afterwards.

4.1.2 Study outcome

The main outcome of this analysis was the time to development of T1D after seroconversion in months. T1D diagnosis was based on American Diabetes Association criteria (American Diabetes Association, 2011).

4.1.3 Statistical analyses

Of the 8,676 children enrolled, 613 had developed one or more autoantibodies at the time of our data access. We created three subsets of the data where we restricted the data to children who had seroconverted to IAA ($n = 442$), GADA ($n = 466$) or IA2A ($n = 288$), respectively. These subsets were not mutually exclusive, as children had potentially seroconverted to multiple autoantibodies. Children were assigned to each subset irrespectively of whether the specific autoantibody was amongst the first islet autoantibodies to appear or appeared at a later time during follow-up. For example, if a child developed autoantibodies to IAA first and autoantibodies to GADA later, the child would be assigned to both the IAA and GADA subset.

We used a novel shared parameter joint model approach to assess the association between the longitudinal autoantibody titers from seroconversion with the time to T1D. Joint models allow the incorporation of longitudinal titers as time-varying covariates into the survival model of progression to T1D by estimating a longitudinal model and a proportional hazards model, using a joint likelihood for both submodels (Rizopoulos, 2012). We further extended this model to a more flexible joint model, where we were able to assess heterogeneous and nonlinear individual biomarker trajectories and to explore complex associations between the biomarkers and the time-to-event (Köhler et al., 2017b). We refer to the previous chapter for further details.[3]

Using this novel approach we specified the autoantibody titers over time as smooth, nonlinear, subject-specific trajectories in the longitudinal model. Furthermore, we allowed the association between the modeled trajectories and the time to T1D to be time-varying in our main analysis. In additional explorative analyses we allowed the association to differ between subjects with different characteristics, and to differ over time between subjects with different characteristics. We fitted these models for each of the three autoantibodies IAA, GADA and IA2A, separately, within each autoantibody-specific subset. In the longitudinal submodels, we assessed the associations of each autoantibody titer with a) age at seroconversion of the respective autoantibody, b) a binary variable indicating whether the autoantibody was among the first autoantibodies to appear, and c) two time-varying binary variables indicating which of the other two autoantibodies were present at each observed time point. In each proportional hazards submodel, we assessed the associations of the smooth subject-specific autoantibody trajectories from the longitudinal model with progression time from seroconversion of the respective autoantibody to T1D. Baseline covariates were a) the age at seroconversion of the respective autoantibody and b) whether the autoantibody was among the first autoantibodies to appear. We further assessed whether the association between the autoantibody trajectories and the time to T1D differed over time between subjects with and without a first-degree relative with T1D or between girls and boys. Additionally, we checked for differences in the association between HLA genotypes. Due to the limited size of certain HLA subgroups we modeled this association as time-constant.

[3]The published manuscript encompasses model details in an Appendix which is omitted here as the respective flexible additive joint model is described in detail in the previous chapter.

All models were estimated within a Bayesian framework using the R package **bamlss** (Umlauf et al., 2017). Weakly informative normal priors were used for all coefficients. We report the posterior mean estimates / hazard ratios (HR) and 95% credibility intervals (CI) for all modeled parameters. Bayesian CIs can be interpreted as the interval in which the population parameter lies with a given probability (here 95%). We assessed convergence of the Markov chains by visual inspection of traceplots and conducted sensitivity analyses with regard to prior specification. All calculations were carried out with R version 3.2.5 (R Core Team, 2016).

4.2 Results

Table B.1[4] shows the study characteristics in each subset, i.e. the subsets of children who developed IAA, GADA or IA2A autoantibodies, respectively, at any time during follow-up. In most cases, either IAA, GADA, or both, were present at the time of the first seroconversion, whereas IA2A occurred at a later time point. The children seroconverted to the different autoantibodies at different median ages ($p < 0.001$, Kruskal-Wallis Test) with IAA seroconversion taking place at a lower median age. Apart from that, children with different autoantibodies were similar regarding the progression time to T1D and other variables. The individual autoantibody patterns over time after seroconversion were heterogeneous, but on average IAA titers declined after an initial increase, and GADA and IA2A titers increased shortly after seroconversion and remained relatively stable thereafter (Figure B.7 in the Appendix).

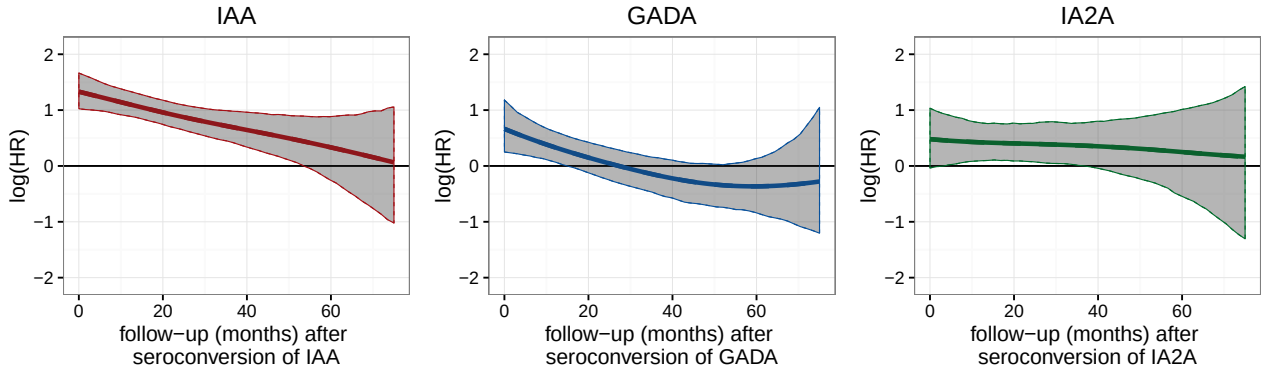

Figure 4.1: Posterior mean estimates (lines) and 95% credibility intervals (shaded areas) of $\eta_\alpha(t)$, the time-varying log hazard ratio (HR) of the association between longitudinal autoantibody trajectories and T1D.

In the joint modeling of autoantibody titers over time and the time to T1D, we observed for all autoantibodies a positive association between the titer and the risk of progression to T1D. Titers over time were lower for subjects who seroconverted at an older age for the respective autoantibody, and higher if the respective autoantibody appeared at the initial seroconversion, and if other autoantibodies were present (Table 4.1). For each autoantibody, a higher age at the respective seroconversion

[4]To enhance readability the table was moved to the Appendix in contrast to the published manuscript.

Table 4.1: Posterior mean estimates of coefficients (β) and hazard ratios with corresponding 95% credibility intervals from joint models of autoantibody trajectories (IAA, GADA and IA2A, as estimated in longitudinal submodels) and progression to T1D (survival submodels).

Autoantibodies	Covariate	Longitudinal models		Survival models	
		β	95% CI	HR	95% CI
IAA	IAA present at first sc	**0.35**	0.20, 0.51	0.66	0.42, 1.02
	IAA sc age (years)	**-0.10**	-0.13, -0.07	**0.84**	0.72, 0.98
	GADA positive (time-varying)	**0.31**	0.24, 0.37	[a]	[a]
	IA2A positive (time-varying)	**0.17**	0.12, 0.22	[a]	[a]
GADA	GADA present at first sc	**0.34**	0.19, 0.47	0.72	0.51, 1.03
	GADA sc age (years)	**-0.06**	-0.09, -0.03	**0.61**	0.52, 0.72
	IAA positive (time-varying)	**0.25**	0.20, 0.32	[a]	[a]
	IA2A positive (time-varying)	**0.06**	0.02, 0.11	[a]	[a]
IA2A	IA2A present at first sc	**0.31**	0.11, 0.50	1.07	0.61, 1.80
	IA2A sc age (years)	**-0.05**	-0.08, -0.01	**0.66**	0.56, 0.78
	IAA positive (time-varying)	**0.22**	0.09, 0.36	[a]	[a]
	GADA positive (time-varying)	**0.29**	0.18, 0.41	[a]	[a]

[a] Covariate only included in the longitudinal submodel.
Bold font indicates that the 95% CI does not include 0 (for β) or 1 (for HR).
seroconversion (sc).

was also associated with lower risk of progression to clinical T1D. For example, children had a hazard ratio [95% CI] of 0.84 [0.72, 0.98] if they seroconverted one year later for IAA. We further investigated whether the association between the estimated trajectories of autoantibodies and the progression to T1D was time-varying or constant.

By using our approach, we observed that the association was time-varying for IAA and GADA with the association being highest early after seroconversion and decreasing over time (Figure 4.1) and stronger for IAA than GADA: The hazard ratio for IAA (per transformed unit) was 3.38 [2.66; 4.38] at 6 months after seroconversion, 3.02 [2.44, 3.81] at 12 months after seroconversion, and 2.02 [1.55, 2.68] at 36 months after seroconversion (Table 3) with an average decrease in the hazard ratio of 10% [95% CI; 2%, 18%] every 6 months. The hazard ratio for GADA (per transformed unit) was 1.63 [1.20, 2.30] at 6 months after seroconversion, 1.40 [1.07, 1.85] at 12 months after seroconversion, and 0.85 [0.61, 1.17] at 36 months after seroconversion with an average decrease of 9% [1%, 15%] every 6 months. For IA2A, the positive association between autoantibody titer and T1D progression was estimated as time-constant: The hazard ratios for IA2A (per transformed unit) were 1.56 [1.04, 2.42] at 6 months after seroconversion, 1.53 [1.10, 2.16] at 12 months after seroconversion, and 1.44 [1.005, 2.16] at 36 months after seroconversion with a negligible average decrease of 2% [-8%, 13%] every 6 months.

Table 4.2: Posterior mean hazard ratios with corresponding 95% credibility intervals at different time-points after seroconversion of each autoantibody for the association between autoantibody trajectories from the longitudinal model and progression time to T1D.

Time point	IAA		GADA		IA2A	
	HR	95% CI	HR	95% CI	HR	95% CI
0 months	**3.78**	2.78, 5.28	**1.94**	1.28, 3.25	1.62	0.96, 2.81
6 months	**3.38**	2.66, 4.38	**1.63**	1.20, 2.30	**1.56**	1.04, 2.42
12 months	**3.02**	2.44, 3.81	**1.40**	1.07, 1.85	**1.53**	1.10, 2.16
24 months	**2.43**	1.94, 3.02	1.07	0.80, 1.41	**1.49**	1.08, 2.13
36 months	**2.02**	1.55, 2.68	0.85	0.61, 1.17	**1.44**	1.005, 2.16
48 months	**1.69**	1.17, 2.50	0.73	0.50, 1.04	1.37	0.82, 2.33
60 months	1.39	0.77, 2.42	0.69	0.43, 1.14	1.28	0.58, 2.74

Bold font indicates that the 95% CI does not include 0 (for β) or 1 (for HR).

As indicated by the credibility intervals in Figure 4.1, positive associations with T1D progression were observed for IAA up to 54 months after seroconversion, for GADA up to 18 months after seroconversion and for IA2A between 6 and 36 months after seroconversion. The traceplots indicated satisfactory convergence of the Markov chains (Figure B.8, B.9, and B.10) and sensitivity analyses showed robustness against different prior specifications (Figure B.11).

We further observed differences in the time-varying association of autoantibodies with progression to T1D between children with and without a first-degree relative with T1D. For all autoantibodies the associations were higher amongst children with a first degree relative at early time points and decreased more strongly within this group (Figure 4.2, upper panel). For IAA, the associations between the two groups differed from seroconversion until about 12 months thereafter, as indicated by the credibility bands of the differences (Figure 4.2, lower panel), but only from 4 to 6 months after seroconversion for GADA and from 1 to 16 months after seroconversion for IA2A.

For all autoantibodies HLA subgroups were similar in the association between autoantibody trajectories and the time to T1D (Figure 4.3). An exception was a higher association for subjects with IAA autoantibodies and the DR3/3 genotype, a genotype which is less prevalent among IAA positive children (n = 30, 7%). In accordance with the difference in the hazard, the mean titer levels between progressors and non-progressors differed more strongly within the small subgroup of DR3/3 than within other HLA genotypes with non-progressors showing an especially low level (Figure B.12). We did not observe consistent differences in the association over time between girls and boys (Figure B.13).

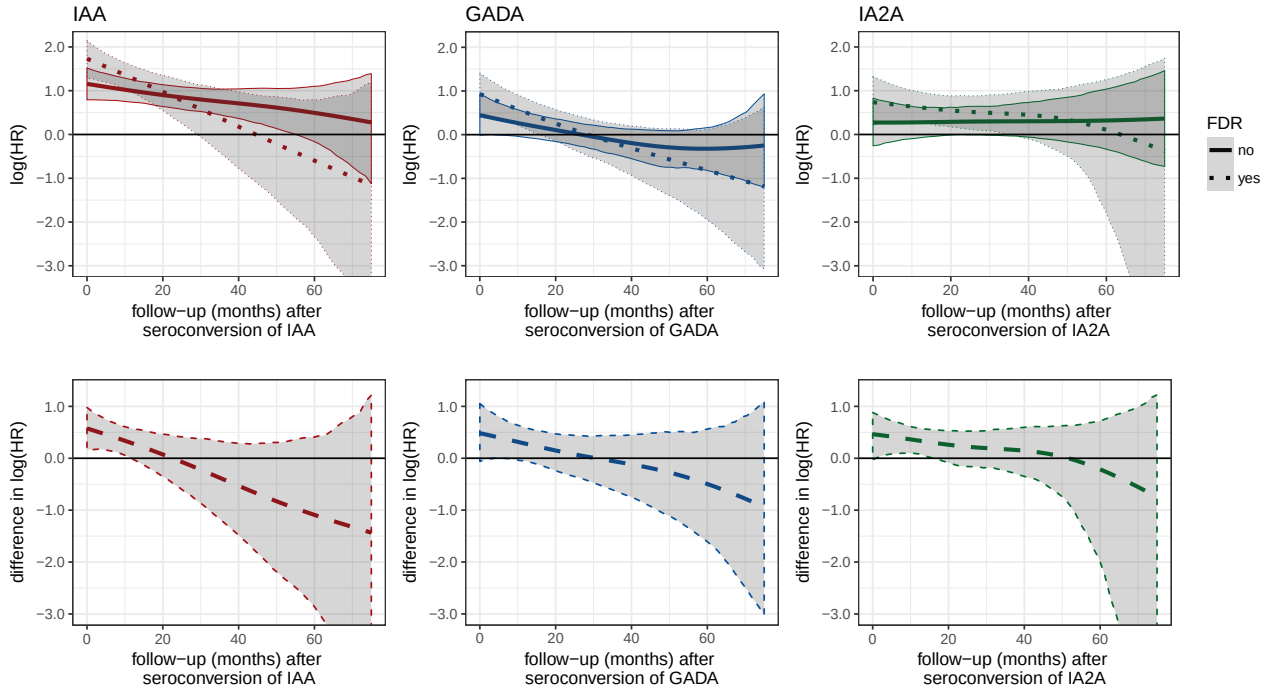

Figure 4.2: Posterior mean estimates (lines / dots) and 95% credibility intervals (shaded areas) of $\eta_\alpha(t, \text{FDR})$, the time-varying log hazard ratio (HR) of the association between longitudinal autoantibody trajectories and T1D progression stratified for children that had a first-degree relative (FDR) with T1D or not (upper panel) and of the difference of the association between the groups over time, $\eta_\alpha(t, \text{FDR} = 1) - \eta_\alpha(t, \text{FDR} = 0)$ (lower panel).

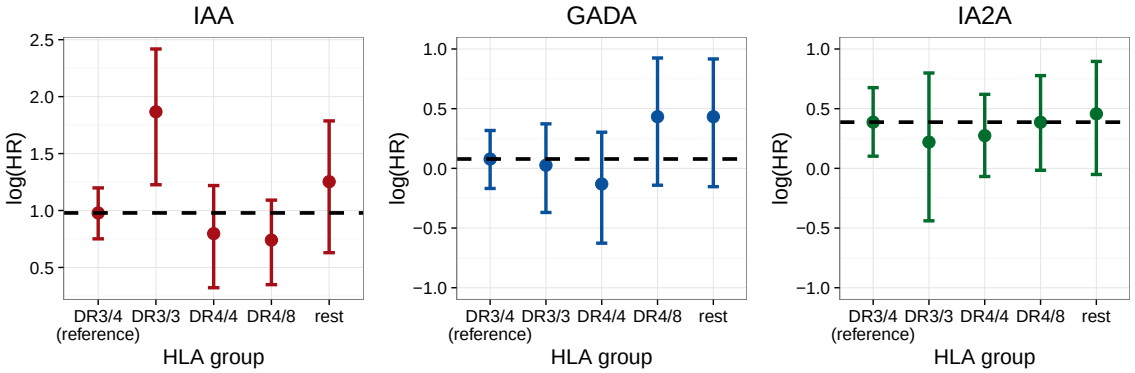

Figure 4.3: Posterior mean estimates and 95% credibility intervals of , the time-constant log hazard ratio (HR) of $\eta_\alpha(t, \text{HLA})$, the association between longitudinal autoantibody trajectories and T1D progression, per HLA genotype. The dashed line represents the estimated log hazard ratio of the reference group.

4.3 Discussion

In the present study the complex relationship between longitudinally measured autoantibodies and the risk of progression to T1D diabetes was explored using a novel joint modeling approach. We observed potentially time-varying positive associations between the autoantibody titers of IAA and GADA, and the risk of T1D progression, indicating that the T1D progression risk associated with autoantibody titers was highest shortly after seroconversion of the respective autoantibody. The hazard ratio was highest for IAA, especially at early time points. Additionally, we observed that the associations of the autoantibody titer and the T1D risk early after seroconversion were more pronounced in children with first-degree relatives with T1D.

These results were in line with earlier results from other cohorts, where initial and mean IAA and IA2A titers were shown to be associated with the risk of progression (Mrena et al., 2006; Parikka et al., 2012; Steck et al., 2011) as well as from a more recent and methodologically advanced study based on the TEDDY data. In this study the relationship between titers of the same autoantibodies over time and the risk of progression to T1D was modeled assuming a time constant association (Steck et al., 2015). By using mean levels of the respective autoantibodies as time-varying predictors in a Cox model, the authors could show a positive association between autoantibody titers and the time to T1D progression for IAA and IA2A in their analyses. Potential limitations of this previous approach are however that (a) only subjects' mean titers until a certain time point are taken into account and not all observed values over time, (b) in a time-varying Cox model the time-varying predictor is assumed constant between observations, and (c) the association between autoantibodies and the risk of progression is assumed to be time-constant.

These limitations were addressed by our joint modeling approach. Here, we flexibly modeled the trajectories of all three autoantibodies in each subject as a smooth function of time, i.e. obtaining predictions for the autoantibody titers between the measurements at discrete time points, and could use all this information as a time-varying covariate in the survival model. Additionally, we allowed their association with the risk of T1D progression to vary over time and between groups of subjects (children with and without first degree relatives with T1D as well as boys and girls). In consequence, we were able to explore the association between autoantibodies and the risk of T1D beyond the previous results. For example, we observed that increased GADA titers may predict T1D progression within the first 1.5 years after seroconversion, but not thereafter. As this association averages to 0 over the whole time range this association was potentially not captured in the simpler modeling from the previous analysis. Furthermore, our modeling approach revealed that the time-varying associations appear to be more pronounced in children with first degree relatives with T1D compared to children without.

The modeling of autoantibodies as longitudinal biomarkers and the time to clinical T1D poses a challenge due to the nature of the data beyond the aspects mentioned above. Longitudinal biomarkers usually contain potential random variation both due to the laboratory measurement process as well

as short and long term biological fluctuations, and are only observed until an event occurs. Whereas not accounting for the random fluctuations in a time-varying Cox model might result in an underestimation of the hazard ratio (Asar et al., 2015), ignoring the latter might distort the estimation of covariate effects in the longitudinal model. By jointly analyzing the longitudinal and survival model we could address these issues and gained further insights as to how covariates affected both the autoantibody titers over time, and the risk of T1D progression. We found that earlier seroconversion for the respective autoantibody, if the respective seroconversion was the initial one, as well as the presence of other autoantibodies was associated with higher autoantibody titers. The age at the respective seroconversion was also inversely related to the risk of T1D progression for every autoantibody.

While joint modeling approaches allow for detailed and unbiased estimations, they demand a high number of subjects, especially when complex associations are modeled in the survival part. TEDDY is the largest prospective study on the determinants of T1D worldwide and thus offers a unique opportunity to explore the application of joint modeling techniques on these complex relationships due to the high number of subjects and the detailed measurement schedule. Currently, the presented flexible joint model only allows the assessment of one longitudinal biomarker at a time. In consequence, one limitation is that we were not able to combine all three markers into one joint model. We partly addressed this issue by including information on the presence of other autoantibodies and the order of their occurrence in our model. While they provide insights into the mechanisms of disease progression, a drawback of our results is that they cannot easily be translated from a cohort setting with frequent measurements into clinical practice, as the age at the respective seroconversion plays a crucial role in the prediction of T1D progression risk, but is not readily available in practice.

In conclusion, by using state of the art joint modeling techniques we were able to give insights into the complex relationship between longitudinal autoantibody titers and the risk of progression to clinical T1D. Risk stratification basing on autoantibody titers should focus on time points early after seroconversion.

Chapter 5

Nonlinear associations in the flexible additive joint model

The joint modeling of longitudinal and survival processes has gained large attention in the last decade and has seen a broad range of developments. In this work we present a flexible framework for Bayesian additive joint models that allows for a highly flexible specification of the association between longitudinal biomarkers and a survival process to gain further insights into complex diseases. A special focus is placed on potentially nonlinear associations between a longitudinal biomarker and the log-hazard of an event.

The research into joint models has largely been motivated by biomedical applications such as modeling of CD4 counts and HIV progression (Wulfsohn and Tsiatis, 1997; Tsiatis and Davidian, 2001), PSA values and prostate cancer (Taylor et al., 2013) or breast cancer (Chi and Ibrahim, 2006) and receives growing attention in applied research (Sudell et al., 2016). In all these applications there is a need for unbiased modeling of a longitudinal covariate, often a biomarker, and its association to the hazard of an event. This situation demands a special treatment as the longitudinal covariate is potentially subject to measurement error, measured at individual-specific time points and observed only until the occurrence of the event. Joint models take all these complications into account by formulating a joint likelihood for the longitudinal and the survival submodel and thereby achieve an unbiased modeling of both. As a detailed overview of the field of joint models for longitudinal and time-to-event data is beyond the scope of this work, we refer to the excellent reviews on the topic from Tsiatis and Davidian (2004), Rizopoulos (2012) and Gould et al. (2015). The main idea of this modeling framework is that a set of parameters is assumed to influence both the longitudinal and the survival submodel with conditional independence between the two models, given those parameters. This shared parameter linking the two submodels can be a latent class structure, as in joint latent class models (Proust-Lima et al., 2014), or random effects, as is the case in most developments in joint modeling. The associations between longitudinal marker and log-hazard in this class of shared random

This chapter is based on Köhler et al. (2017c) published as an arXiv e-print [stat.ME]. For more information on the contributions of the authors see Section 1.4. Modifications to the original version are indexed with footnotes.

effects models can be parameterized differently such that only the random effects are associated, the current value of the marker at a certain time point or further transformations of this current value (see Hickey et al. (2016) for an overview of associations structures in multivariate joint modeling). Focus in this work is placed on the current value association.

Existing shared random effects models include the linearity assumption that the effect of the modeled marker trajectories on the logarithm of the hazard is linear. In the context of survival analysis checking the linearity assumption as well as the modeling of an appropriate functional form has been under study (Buchholz and Sauerbrei, 2011; Holländer and Schumacher, 2006). In different biomedical applications it was shown that appropriate modeling of the functional form of continuous covariate effects reduces bias and allows for additional insights into prognostic factors, for example in the study of breast cancer (Gray, 1992; Sauerbrei et al., 1999), lung cancer (Gagnon et al., 2010) and leukemia (Inaba et al., 2012). For the accurate specification of nonlinear effects of continuous covariates in the time-to-event model different strategies have been applied, such as fractional polynomials (Royston and Altman, 1994; Sauerbrei et al., 2007) as well as unpenalized (Sleeper and Harrington, 1990; Wynant and Abrahamowicz, 2016) and penalized spline approaches (Hastie and Tibshirani, 1995; Hofner et al., 2011).

The results from survival modeling suggest that the linearity assumption may also not always be met when modeling the effects of a longitudinal marker in a joint model. To our knowledge, to date no shared random effects joint model approach extends or even tests this assumption. The user of a joint model can only assume that, given an appropriate transformation of the raw marker values such as a log-transformation, the association is indeed linear. The present work aims to fill this gap by allowing greater flexibility in the specification of the association between marker and event. Note that joint latent class models (Proust-Lima et al., 2014), where the latent class is associated with the log-hazard and the association between marker and event is only implicit, also allow for a nonlinear relationship between marker and hazard. However, our interest lies in gaining insights in the detailed nature of this association, and therefore an explicit modeling of this association is necessary.

In Chapter 3 a general framework for flexible structured additive joint models was presented with the focus on modeling highly subject-specific nonlinear individual longitudinal trajectories as well as a time-varying association. This flexibility is achieved by formulating the joint model as a structured additive regression (Fahrmeir et al., 2004) in which all model parts, which are the baseline hazard, baseline covariates effects, mean and variance of the modeled longitudinal marker as well as the association are structured additive predictors. These predictors can encompass nonlinear, smooth and time-varying effects by making use of P-splines (Eilers and Marx, 1996) and capture highly flexible nonlinear individual trajectories by modeling them as functional random intercepts (Scheipl et al., 2015). The model is estimated in a Bayesian framework with smoothness and random effects structures induced by appropriate prior specifications. In the present work this framework is generalized further to allow for nonlinear associations between a marker and the event process as well as to allow this nonlinear association to vary with covariates.

In order to facilitate the application of this flexible joint model it is fully implemented in the R package **bamlss**, thereby adding to the available range of joint model packages. Software packages in

the shared random effects approach are **JM** (Rizopoulos, 2010) and its Bayesian counterpart **JMbayes** (Rizopoulos, 2016a), **joineR** (Philipson et al., 2017), **frailtypack** (Rondeau et al., 2012) as well as the stata package **stjm** (Crowther, 2012) and the SAS macro **JMFit** (Zhang et al., 2016) of which many are rather restricted in the amount of flexibility they allow in modeling nonlinear individual trajectories and the association itself. Out of these packages up to date the R package **JMbayes** offers the most flexibility in modeling individual trajectories and different association structures while, however, also assuming linearity in the association between the marker and the log-hazard. We therefore compare our implementation with this established package in our simulation study.

The paper is structured as follows: In Section 5.1 the general framework is presented with details on the Bayesian estimation in Section 5.2. An extensive simulation study in Section 5.3 and a practical application of the model on the well-known data on primary biliary cirrhosis (PBC) (Murtaugh et al., 1994), which is included in the R package **JMbayes**, in Section 5.4 aim to give further insights into the performance of this flexible model. Finally, concluding remarks are given in Section 5.5 and further technical details can be found in the Appendix B.3. The presented model is implemented in the R package **bamlss** and source code to fully reproduce the results of the simulations and the application is given in the Supporting Information of the original contribution.

5.1 Methods

In the following we further generalize the previously formulated flexible additive joint model (see Chapter 3) to allow for complex nonlinear association structures between a longitudinal marker and the time-to-event process.

5.1.1 General model

For each subject $i = 1, \dots, n$ we observe the longitudinal response $\boldsymbol{y}_i = [y_{i1}, \cdots, y_{in_i}]^\top$ at the potentially subject-specific time points $\mathbf{t}_i = [t_{i1}, \cdots, t_{in_i}]^\top$ with $t_{i1} \leq \cdots \leq t_{in_i} \leq T_i$, modeled by

$$y_{ij} = \eta_{\mu i}(t_{ij}) + \varepsilon_{ij} \text{ with } \varepsilon_{ij} \sim N(0, \exp[\eta_{\sigma i}(t_{ij})]^2). \tag{5.1}$$

The predictor η_μ denotes the "true" longitudinal marker that serves as a time-varying covariate in the time-to-event model. Additionally we observe for every subject $i = 1, \dots, n$ a potentially right-censored follow-up time T_i and the event indicator δ_i, which is 1 if subject i experiences the event and 0 if it is censored. The hazard of an event at time t is modeled by structured additive predictors η_k, $k \in \{\lambda, \gamma, \alpha, \mu\}$ as

$$h_i(t) = \exp\{\eta_i(t)\} = \exp\{\eta_{\lambda i}(t) + \eta_{\gamma i} + \eta_{\alpha i}(\eta_{\mu i}(t), t)\} \tag{5.2}$$

with η_λ the predictor for all time-varying survival covariates and effects including the log baseline hazard, η_γ representing all baseline survival covariates, the longitudinal marker η_μ and the potentially nonlinear association between the longitudinal marker and the hazard η_α. Note that by modeling the

latter as a function of η_μ and time t, a variety of association structures can be specified.
In general, the vector of predictors for all subjects is expressed as $\boldsymbol{\eta}_k = [\eta_{k1}, \cdots, \eta_{kn}]^\top$, $k \in \{\lambda, \gamma, \alpha, \mu, \sigma\}$. In the longitudinal part of the model, the predictor vector is $\boldsymbol{\eta}_\mu(\mathbf{t})$ of length $N = \sum_i n_i$ containing entries $\eta_{\mu i}(t_{ij})$ for all $j = 1, \ldots, n_i$ per subject i, i.e. corresponding evaluations at all observed time points $\mathbf{t} = [\mathbf{t}_1^\top, \cdots, \mathbf{t}_n^\top]^\top$. In the survival part of the model, the predictor vector $\boldsymbol{\eta}_\mu(t)$ is of length n containing one observation per subject at time t. This setup in the survival part is analogous to the setup for the other predictors in the survival submodel and additionally, $\boldsymbol{\eta}_k(\mathbf{T})$ denotes the evaluation of the respective predictor at the vector of follow-up times for all subjects $\mathbf{T} = [T_i, \cdots, T_n]^\top$.

Each predictor η_{ki} with $k \in \{\lambda, \gamma, \mu, \sigma\}$ is a structured additive predictor $\eta_{ki} = \sum_{m=1}^{M_k} f_{km}(\tilde{\boldsymbol{x}}_{kmi})$ of M_k functions of covariates $\tilde{\boldsymbol{x}}_i$. Each function f_{km} depends on one or two covariates, i.e. different subsets $\tilde{\boldsymbol{x}}_{kmi}$ of $\tilde{\boldsymbol{x}}_i$. For time-varying predictors the functions can also depend on time $\eta_{ki}(t) = \sum_{m=1}^{M_k} f_{km}(\tilde{\boldsymbol{x}}_{kmi}(t), t)$. By using suitable (e.g. spline) basis matrices $\mathbf{X}_{km}$ for every term m of predictor k and corresponding penalty $\mathbf{P}_{km}$ a variety of effects such as nonlinear, spatial, time-varying or random effects can be modeled under the generic structure

$$\boldsymbol{f}_{km} = \mathbf{X}_{km}\boldsymbol{\beta}_{km} \qquad \text{and} \qquad \mathbf{P}_{km} = \frac{1}{\tau_{km}^2}\boldsymbol{\beta}_{km}^\top \mathbf{K}_{km}\boldsymbol{\beta}_{km}. \tag{5.3}$$

Here, $\boldsymbol{f}_{km}$ denotes the vector of function evaluations stacked over subjects, $\mathbf{X}_{km}$ are the design matrices of size $n \times p_{km}$ or $N \times p_{km}$ for the survival and longitudinal submodel, respectively, and $\boldsymbol{\beta}_{km} = [\beta_{k1}, \cdots, \beta_{kp_{km}}]^\top$ denotes the coefficient vector of length p_{km}. Note that for the survival part $\boldsymbol{x}_{kmi}$ denotes the i-th row of the design matrix $\mathbf{X}_{km}$ whereas $\tilde{\boldsymbol{x}}_{kmi}$ denotes the respective covariate vector. For parametric terms these two often coincide, whereas for spline representations of smooth covariate effects or random effects terms $\boldsymbol{x}_{kmi}$ represents the respective basis evaluation vector of $\tilde{\boldsymbol{x}}_{kmi}$. For example, random intercepts are modeled using the design matrix $\mathbf{X}_{km}$, an $N \times n$ indicator matrix with the ith column indicating which longitudinal measurements belong to subject i, the coefficient vector $\boldsymbol{\beta}_{km} = [\beta_{km1}, \cdots, \beta_{kmn}]$ and the penalty matrix $\mathbf{K}_{km} = \boldsymbol{I}_n$, which is an $n \times n$ identity matrix. This penalty ensures $\beta_{kmi} \sim N(0, \tau_{km}^2)$ independently. For the setup of smooth effects using P-splines we refer to the next subsection and details on the setup of the predictors, function evaluations and design matrices for the submodels can be found in Table 3.1.

All effects are modeled within a Bayesian framework by specifying appropriate prior distributions for the coefficient vectors, as presented in more detail in Section 5.2.

5.1.2 Flexible associations

The special focus in the generalization of the hazard in eq. (5.2) lies on the flexible specification of the predictor η_α to incorporate not only time-varying and covariate-dependent associations as presented in Chapter 3, but also nonlinear associations between the predicted longitudinal marker and the time-to-event process.

The general predictor is formulated as $\eta_{\alpha i}(\eta_{\mu i}(t), t) = f_\alpha\left(\eta_{\mu i}(t), \tilde{\boldsymbol{x}}_{\alpha i}, t\right)$, that is a function of the potentially smooth time-varying predicted marker trajectories $\eta_{\mu i}(t)$ from eq. (5.1), further covariates

$\tilde{\boldsymbol{x}}_{\alpha i}$ as well as time t. Note that we drop the subscript m whenever this is simpler. We make use of a suitable basis representation to incorporate this flexible specification into our framework as

$$f_\alpha\left(\eta_{\mu i}(t), \tilde{\boldsymbol{x}}_{\alpha i}, t\right) = \left[\boldsymbol{g}_1(\eta_{\mu i}(t)) \odot \boldsymbol{g}_2(\tilde{\boldsymbol{x}}_{\alpha i}, t)\right] \boldsymbol{\beta}_\alpha = \boldsymbol{x}_{\alpha i}^\top \boldsymbol{\beta}_\alpha \tag{5.4}$$

with $\odot$ denoting the row tensor product. The row tensor product of a $p \times a$ matrix $\boldsymbol{A}$ and a $p \times b$ matrix $\boldsymbol{B}$ is defined as the $p \times ab$ matrix $\boldsymbol{A} \odot \boldsymbol{B} = (\boldsymbol{A} \otimes \mathbf{1}_b^\top) \cdot (\mathbf{1}_a^\top \otimes \boldsymbol{B})$ with $\cdot$ denoting element-wise multiplication and $\otimes$ the Kronecker product. In this notation $\boldsymbol{g}_1(\eta_{\mu i}(t))$ represents the basis vector of the potentially nonlinear effect of the longitudinal predictor $\eta_{\mu i}(t)$ and $\boldsymbol{g}_2(\tilde{\boldsymbol{x}}_{\alpha i}, t)$ represents the basis vector for the effects of relevant covariates and/or a smooth function of time t. The resulting design vector $\boldsymbol{x}_{\alpha i}$ and parameter vector $\boldsymbol{\beta}_\alpha$ are of length $p_\alpha = p_{\alpha 1} \cdot p_{\alpha 2}$.

The standard linear association between the longitudinal predictor and the log-hazard can be formulated as $\boldsymbol{g}_1(\eta_{\mu i}(t)) = I(\eta_{\mu i}(t)) = \eta_{\mu i}(t)$, where $I(\cdot)$ denotes the identity, and $\boldsymbol{g}_2 \equiv 1$. For nonlinear associations we use P-splines (Eilers and Marx, 1996) by specifying a B-spline representation of the longitudinal predictor effect $\boldsymbol{g}_1(\eta_{\mu i}(t)) = \mathbf{B}\left(\eta_{\mu i}(t)\right) = \left[B_1\left(\eta_{\mu i}(t)\right), \cdots, B_{p_{\alpha 1}}\left(\eta_{\mu i}(t)\right)\right]$. Here, B_m denotes the m-th basis function over the observed range of $\eta_{\mu i}(t)$, with $\boldsymbol{g}_1(\eta_{\mu i}(t))$ being the corresponding design vector of length $p_{\alpha 1}$ of the spline evaluations at $\eta_{\mu i}(t)$. The corresponding penalty matrix of the effect of $\eta_\mu(t)$ is a zero matrix $\mathbf{K}_{\alpha 1} = \mathbf{0}$ for $\boldsymbol{g}_1(\eta_{\mu i}(t)) = I(\eta_{\mu i}(t))$ and a P-spline penalty matrix $\mathbf{K}_{\alpha 1} = \mathbf{D}_r^\top \mathbf{D}_r$ with the r-th difference matrix $\mathbf{D}_r$ for $\boldsymbol{g}_1(\eta_{\mu i}(t)) = \mathbf{B}(\eta_{\mu i}(t))$. For simplicity, we denote the function transforming any covariate values $\boldsymbol{z}$ into a matrix of evaluations of a spline basis generally as $\mathbf{B}(z)$. This function returns the matrix of respective basis evaluations with $p_{\alpha 1}$ columns and $length(\boldsymbol{z})$ rows.

In order to model simple parametric, nonlinear or time-varying effects, $\boldsymbol{g}_2(\tilde{\boldsymbol{x}}_{\alpha i}, t)$ can be specified accordingly as a constant, a spline representation of a continuous covariate or as spline representation of an effect of time t with appropriate penalty matrix $\mathbf{K}_{\alpha 2}$. To further illustrate the notation, consider the following effect specifications

- time-constant, linear association $f_\alpha\left(\eta_{\mu i}(t)\right) = \left[I\left(\eta_{\mu i}(t)\right) \odot 1\right] \beta_\alpha$ where $p_\alpha = 1 \cdot 1$,
- (linearly) covariate-dependent, linear association $f_\alpha\left(\eta_{\mu i}(t), \tilde{\boldsymbol{x}}_{\alpha i}\right) = \left[I\left(\eta_{\mu i}(t)\right) \odot \tilde{\boldsymbol{x}}_{\alpha i}^\top\right] \boldsymbol{\beta}_\alpha$ where $p_\alpha = 1 \cdot p_{\alpha 2}$ with $p_{\alpha 2}$ the length of $\tilde{\boldsymbol{x}}_{\alpha i}$,
- time-varying, linear association $f_\alpha\left(\eta_{\mu i}(t), t\right) = \left[I\left(\eta_{\mu i}(t)\right) \odot \mathbf{B}(t)\right] \boldsymbol{\beta}_\alpha$ where $p_\alpha = 1 \cdot p_{\alpha 2}$ with $p_{\alpha 2}$ the number of spline basis functions in $\mathbf{B}(t)$,
- time-constant, nonlinear association $f_\alpha\left(\eta_{\mu i}(t)\right) = \left[\mathbf{B}\left(\eta_{\mu i}(t)\right)^\top \odot 1\right] \boldsymbol{\beta}_\alpha$ where $p_\alpha = p_{\alpha 1} \cdot 1$ with $p_{\alpha 1}$ the number of spline basis functions in $\mathbf{B}\left(\eta_{\mu i}(t)\right)$,
- covariate-dependent, nonlinear association $f_\alpha\left(\eta_{\mu i}(t), \tilde{\boldsymbol{x}}_{\alpha i}\right) = \left[\mathbf{B}\left(\eta_{\mu i}(t)\right)^\top \odot \tilde{\boldsymbol{x}}_{\alpha i}^\top\right] \boldsymbol{\beta}_\alpha$ where $p_\alpha = p_{\alpha 1} \cdot p_{\alpha 2}$ with $p_{\alpha 1}$ the number of spline basis functions in $\mathbf{B}\left(\eta_{\mu i}(t)\right)$ and $p_{\alpha 2}$ the length of $\tilde{\boldsymbol{x}}_{\alpha i}$,
- time-varying, nonlinear association $f_\alpha\left(\eta_{\mu i}(t), t\right) = \left[\mathbf{B}\left(\eta_{\mu i}(t)\right)^\top \odot \mathbf{B}(t)\right] \boldsymbol{\beta}_\alpha$ where $p_\alpha = p_{\alpha 1} \cdot p_{\alpha 2}$ with $p_{\alpha 1}$ and $p_{\alpha 2}$ the number of spline basis functions in $\mathbf{B}\left(\eta_{\mu i}(t)\right)$ and $\mathbf{B}(t)$, respectively.

For both, time-varying effects and nonlinear associations, Bayesian P-Splines (Lang and Brezger, 2004) are employed where smoothing is induced by appropriate prior specification. In more detail the difference penalties are replaced by their stochastic analogues, i.e. random walks. The full penalty $\mathbf{P}_\alpha$ allows for different amounts of smoothing across both $\eta_{\mu i}(t)$ and the covariate or time effects by using an anisotropic smooth with

$$\mathbf{P}_\alpha = \boldsymbol{\beta}_\alpha^\top \left(\frac{1}{\tau_{\alpha 1}^2} \mathbf{K}_{\alpha 1} \otimes \boldsymbol{I}_{p_{\alpha 2}} + \frac{1}{\tau_{\alpha 2}^2} \boldsymbol{I}_{p_{\alpha 1}} \otimes \mathbf{K}_{\alpha 2} \right) \boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_\alpha^\top \left(\frac{1}{\tau_{\alpha 1}^2} \tilde{\boldsymbol{K}}_{\alpha 1} + \frac{1}{\tau_{\alpha 2}^2} \tilde{\boldsymbol{K}}_{\alpha 2} \right) \boldsymbol{\beta}_\alpha, \tag{5.5}$$

where $\boldsymbol{I}_a$ is an $a \times a$ identity matrix. Within the R package **bamlss** currently all above mentioned linear associations as well as constant and group-specific nonlinear associations are implemented. Further nonlinear associations are under construction.

5.1.3 Identifiability

Given the additive structure of the model and the fact that all model parts always contain an intercept in our construction, constraints on certain predictors are necessary to obtain an identifiable model. The general constraint for all nonlinear terms in the model is a sum-to-zero constraint over all n or N observations for predictors in the survival and longitudinal submodel, respectively, e.g. $\sum_i f_{\lambda m}(T_i) = 0$ or $\sum_i f_{\gamma m}(\tilde{\boldsymbol{x}}_{\gamma mi}) = 0$. These constraints are implemented for B-splines by transforming the $n \times p_{km}$ basis matrix $\mathbf{X}_{km}$ into an $n \times (p_{km} - 1)$ matrix $\dot{\boldsymbol{X}}_{km}$ for which it holds that $\dot{\boldsymbol{X}}_{km} \mathbf{1}_{p_{km}-1} = \mathbf{0}$ as shown in Wood (2006, chapter 1.8), and adjusting the penalty accordingly. For tensor product smooth terms the constraint is achieved by transforming the marginal basis matrix of the continuous covariate and the corresponding marginal penalty.

In the case of a nonlinear specification of $\eta_{\alpha i}(\eta_{\mu i}(t))$, the predictor also needs to be constrained. As the predictor $\eta_{\mu i}(t)$ and therefore also its spline basis evaluation is estimated within the model, we choose a constraint based on the observed marker. In more detail, we constrain the term to sum to zero on a fixed grid $\boldsymbol{y}^*$ from the 2.5th to the 97.5th quantile of the observed longitudinal response, i.e. $\mathbf{1}^\top \boldsymbol{\eta}_\alpha(\boldsymbol{y}^*) = 0$ with $\mathbf{1}$ a vector of ones. For nonlinear effects per factor level g the same constraint is enforced for every level g and one intercept per factor level except the reference level is included in the model.

5.2 Estimation

We estimate the model in a Bayesian framework using a Newton-Raphson procedure and a derivative-based MCMC algorithm to estimate the mode and the mean of the posterior distribution of the vector $\boldsymbol{\theta}$ of all parameters, respectively.

Assuming conditional independence of the survival outcomes $[T_i, \delta_i]$ and the longitudinal outcomes $\boldsymbol{y}_i$, given the parameters $\boldsymbol{\theta}$, the posterior of the full model is

$$p(\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{\delta}, \boldsymbol{y}) \propto L^{\text{long}}\left[\boldsymbol{\theta}|\boldsymbol{y}\right] \cdot L^{\text{surv}}\left[\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{\delta}\right] \prod_{k \in \{\lambda,\gamma,\alpha,\mu,\sigma\}} \prod_{m=1}^{M_k} \left[p(\boldsymbol{\beta}_{km}|\boldsymbol{\tau}_{km}^2) p(\boldsymbol{\tau}_{km}^2)\right],$$

with the likelihood of the longitudinal submodel L^{long} (cf. eq. (5.1)) and the survival submodel L^{surv} (cf. eq. (5.2)), and the response vectors $\boldsymbol{y} = [\boldsymbol{y}_1^\top, \cdots, \boldsymbol{y}_n^\top]^\top$ and $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_n]^\top$. Further, $p(\boldsymbol{\beta}_{km}|\boldsymbol{\tau}_{km}^2)$ and $p(\boldsymbol{\tau}_{km}^2)$ denote the priors of the vectors of regression parameters and variance parameters for each term m and predictor k. Note that for anisotropic smooths, multiple variance parameters are used resulting in the vector $\boldsymbol{\tau}_{km}^2$.

5.2.1 Likelihood

The log-likelihood of the longitudinal part is

$$\ell^{\text{long}}\left[\boldsymbol{\theta}|\boldsymbol{y}\right] = -\frac{N}{2}\log(2\pi) - \mathbf{1}_N^\top \boldsymbol{\eta}_\sigma\left(\mathbf{t}\right) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{\eta}_\mu\left(\mathbf{t}\right))^\top \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{\eta}_\mu\left(\mathbf{t}\right))$$

where $\boldsymbol{R} = \text{blockdiag}(\boldsymbol{R}_1, \cdots, \boldsymbol{R}_n)$. $\boldsymbol{R}$ simplifies to a diagonal matrix, as we assume $\boldsymbol{R}_i = \text{diag}(\exp[\eta_{\sigma i}(t_{i1})]^2, \cdots, \exp[\eta_{\sigma i}(t_{in_i})]^2)$.

The log-likelihood of the survival part of the model is

$$\ell^{\text{surv}}\left[\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{\delta}\right] = \boldsymbol{\delta}^\top \boldsymbol{\eta}(\mathbf{T}) - \mathbf{1}_n^\top \boldsymbol{\Lambda}\left(\mathbf{T}\right)$$

where $\boldsymbol{\Lambda}(\mathbf{T}) = [\Lambda_1(T_1), \cdots, \Lambda_n(T_n)]^\top$ denotes the vector of cumulative hazard rates with $\Lambda_i(T_i) = \exp(\eta_{\gamma i}) \int_0^{T_i} \exp[\eta_{\lambda i}(u) + \eta_{\alpha i}(\eta_{\mu i}(u), u)]du$.

5.2.2 Priors

In our setup different terms, such as smooth, time-varying or random effects, are specified by the choice of corresponding design matrices and priors. For linear or parametric terms we use vague normal priors on the vectors of the regression coefficients, e.g. $\boldsymbol{\beta}_{km} \sim N(\mathbf{0}, 1000^2\boldsymbol{I})$, to approximate a precision matrix $\mathbf{K}_{km} = \mathbf{0}$. Multivariate normal priors

$$p(\boldsymbol{\beta}_{km}|\tau_{km}^2) \propto \left(\frac{1}{\tau_{km}^2}\right)^{\frac{\text{rank}(\mathbf{K}_{km})}{2}} \exp\left(-\frac{1}{2\tau_{km}^2}\boldsymbol{\beta}_{km}^\top \mathbf{K}_{km}\boldsymbol{\beta}_{km}\right)$$

are used to regularize smooth and random effect terms with precision matrix $\mathbf{K}_{km}$ as specified in the penalty eq. (5.3). For anisotropic smooths as in the flexible association η_α in eq. (5.5), when multiple variance parameters are involved, e.g. $\boldsymbol{\tau}_\alpha^2 = (\tau_{\alpha 1}^2, \tau_{\alpha 2}^2)$, we use the prior

$$p(\boldsymbol{\beta}_{km}|\boldsymbol{\tau}_{km}^2) \propto \left|\frac{1}{\tau_{km1}^2}\tilde{\boldsymbol{K}}_{km1} + \frac{1}{\tau_{km2}^2}\tilde{\boldsymbol{K}}_{km2}\right|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\beta}_{km}^\top \left[\frac{1}{\tau_{km1}^2}\tilde{\boldsymbol{K}}_{km1} + \frac{1}{\tau_{km2}^2}\tilde{\boldsymbol{K}}_{km2}\right]\boldsymbol{\beta}_{km}\right).$$

As priors for the variance parameters τ^2_{km}, which control the trade-off between flexibility and smoothness in the nonlinear modeling of effects, we use independent inverse Gamma hyperpriors $\tau^2_{km} \sim IG(0.001, 0.001)$ to obtain an inverse Gamma full conditional (component-wise in the case of variance vectors). Further priors for the variance parameters, such as half-Cauchy, are possible.

5.2.3 Posterior mode and posterior mean

To obtain starting values for the posterior mean estimation and to gain a quick model assessment we estimate the mode of the posterior using a Newton-Raphson procedure. In more detail, we maximize the log-posterior by updating blockwise each term m of predictor k in each iteration l as

$$\boldsymbol{\beta}^{[l+1]}_{km} = \boldsymbol{\beta}^{[l]}_{km} - \nu^{[l]}_{km} \boldsymbol{H} \left(\boldsymbol{\beta}^{[l]}_{km}\right)^{-1} \boldsymbol{s}\left(\boldsymbol{\beta}^{[l]}_{km}\right)$$

with steplength $\nu^{[l]}_{km}$, the score vector $\boldsymbol{s}(\boldsymbol{\beta}_{km})$ and the Hessian $\boldsymbol{H}(\boldsymbol{\beta}_{km})$. In each updating step we optimize the steplength $\nu^{[l]}_{km}$ over $(0, 1]$ to maximize the log-posterior and the variance parameters to minimize the corrected AIC (AICc, Hurvich et al., 1998). The block-wise score vectors and Hessians can be found in Appendix A. For quick approximate inference we derive credibility intervals from $N(\hat{\boldsymbol{\beta}}_{km}, [-\boldsymbol{H}(\hat{\boldsymbol{\beta}}_{km})]^{-1})$ assuming an approximately normal posterior distribution for the coefficients $\boldsymbol{\beta}_{km}$. Note, however, that as these credibility intervals do not take into account the optimization of the variance parameters, they tend to underestimate the variability and posterior mean sampling should be used for exact inference.

The focus of our model estimation lies on the derivative-based Metropolis-Hastings posterior mean sampling. We construct approximate full conditionals $\pi(\boldsymbol{\beta}_{km}|\cdot)$ based on a second order Taylor expansion of the log-posterior centered at the last state $\beta^{[l]}_{km}$ as shown in Umlauf et al. (2017). This approximate full conditional results in a multivariate normal proposal density with the precision matrix $(\boldsymbol{\Sigma}^{[l]}_{km})^{-1} = -\boldsymbol{H}(\boldsymbol{\beta}^{[l]}_{km})$ and the mean $\boldsymbol{\mu}^{[l]}_{km} = \boldsymbol{\beta}^{[l]}_{km} - \boldsymbol{H}(\boldsymbol{\beta}^{[l]}_{km})^{-1}\boldsymbol{s}(\boldsymbol{\beta}^{[l]}_{km})$. We draw a candidate $\boldsymbol{\beta}^*_{km}$ from the proposal density $q(\boldsymbol{\beta}^*_{km}|\boldsymbol{\beta}^{[l]}_{km}) = N(\boldsymbol{\mu}^{[l]}_{km}, \boldsymbol{\Sigma}^{[l]}_{km})$ in each iteration l of the Metropolis-Hastings sampler for updating block km. Despite being computationally demanding, drawing candidates from a close derivative-based approximation of the full conditional results in high acceptance rates and good mixing as we approximate a Gibbs sampler. Samples for the variance parameters τ^2_{km} are either obtained via Gibbs sampling, if inverse Gamma hyperpriors are used and the full conditionals $\pi(\tau^2_{km}|\cdot)$ in consequence follow an inverse Gamma distribution, or via slice sampling when no simple closed-form full conditional can be obtained. This is the case in the sampling of variance parameters for anisotropic smooths or when other hyperpriors than the inverse Gamma are used. We suggest to use DIC for model selection.

5.3 Simulation

The performance of the presented framework is tested in extensive simulations of which a subset is shown in the following. Three main questions motivated the simulations: First, we aim to assess how

well the flexible joint model can estimate truly linear associations, also in comparison to established implementations as in the R package **JMbayes**. Second, we explore how well the model can capture truly nonlinear associations and assess if any bias occurs if the nonlinear association is falsely modeled as linear in the log-hazard in **JMbayes**. Third, the performance of fitting a nonlinear effect per subgroup is assessed. As previous work has shown a strong dependence of the estimation precision on the number of subjects, we test data sets of two different sizes in all three simulation settings.

5.3.1 Simulation design

We simulate data according to eq. (5.1) and eq. (5.2) where we use in setting 1 the linear association $\eta_{\alpha i}(\eta_{\mu i}(t)) = 1 \cdot \eta_{\mu i}(t)$ between the longitudinal marker and the log-hazard, in setting 2 the nonlinear association $\eta_{\alpha i}(\eta_{\mu i}(t)) = -0.1(\eta_{\mu i}(t) + 3)^2 + \eta_{\mu i}(t) + 1.8$ and in setting 3 a group-specific nonlinear association $\eta_{\alpha i}(\eta_{\mu i}(t), g_i = 1) = -0.1(\eta_{\mu i}(t) + 3)^2 + \eta_{\mu i}(t) + 1.8$ and $\eta_{\alpha i}(\eta_{\mu i}(t), g_i = 0) = 0.1(\eta_{\mu i}(t) - 3)^2 + 0.75\eta_{\mu i}(t) - 0.8$. In all settings we generate $Q = 200$ data sets with $n = 300, 600$, respectively, to assess the influence of sample size on the precision of the estimates.

In more detail we generate longitudinal marker values $\eta_{\mu i}(t) = \sum_{m=1}^{5} f_{\mu m}(\tilde{\boldsymbol{x}}_{\mu mi}, t)$ at a fixed grid of time points $t^* = 1, \ldots, 120$ with the time effect $f_{\mu 1}(t) = 0.1(t + 2)\exp(-0.075t)$, random intercepts $f_{\mu 2}(i) = r_i$ where $r_i \sim N(0, 0.25)$, functional random intercepts (i.e. smooth subject-specific trajectories) $f_{\mu 3}(t, i) = \mathbf{X}_{\mu 3}\boldsymbol{\beta}_{\mu 3} = (\mathbf{X}_{\mu 3s} \odot \mathbf{X}_{\mu 3t})\boldsymbol{\beta}_{\mu 3}$ where $\mathbf{X}_{\mu 3s}$ and $\mathbf{X}_{\mu 3t}$ are the basis representations of a random intercept and a spline over t, respectively, as well as a global intercept $f_{\mu 4}(\tilde{\boldsymbol{x}}_{\mu i}) = 0.5$ and covariate effect $f_{\mu 5}(\tilde{\boldsymbol{x}}_{\mu i}) = 0.6\sin(\tilde{x}_{2i})$ with $\tilde{x}_{2i} \sim unif(-3, 3)$. The functional random intercepts are simulated using P-Splines based on cubic B-splines where the true vector of spline-coefficients with 4 basis functions per subject is drawn from $\boldsymbol{\beta}_{\mu 3} \sim N(\mathbf{0}, [(1/\tau^2_{\mu 3s})\tilde{\boldsymbol{K}}_{\mu 3s} + (1/\tau^2_{\mu 3t})\tilde{\boldsymbol{K}}_{\mu 3t}]^{-1})$ where $\tilde{\boldsymbol{K}}_{\mu 3s} = \mathbf{K}_{\mu 3s} \otimes \boldsymbol{I}_4$ with $\mathbf{K}_{\mu 3s} = \boldsymbol{I}_n$ as the penalty matrix for the random effect and $\tilde{\boldsymbol{K}}_t = \boldsymbol{I}_n \otimes \mathbf{K}_{\mu 3t}$ with $\mathbf{K}_{\mu 3t}$ as an appropriate penalty matrix for the smooth effect of time with $\mathbf{K}_t = \mathbf{D}_2^\top \mathbf{D}_2$, $\tau_s^2 = 1$ and $\tau_t^2 = 0.2$. Similar to eq. (3.5) the two marginal penalties are extended by the Kronecker product $\otimes$ with suitable identity matrices.

We calculate the hazard $h_i(t)$ for every subject using $\eta_\lambda(t) = 1.4\log((t + 10)/1000)$, $\eta_{\gamma i} = 0.3\tilde{x}_{1i}$, with $\tilde{x}_{1i} \sim U(-3, 3)$ and η_α as described above. Survival times for every subject are derived using survival probabilities obtained by numerical integration as described in Bender et al. (2005) and Crowther and Lambert (2013) and censored at $t = 120$. We additionally censor all survival times uniformly using $U(0, 1.5 \cdot 120)$. In order to mimic the irregular measurement times we randomly delete 90% of the generated longitudinal measurements resulting in a median of 6 measurements per subject (interquartile range (IQR): 3, 10) for every setting. Finally we obtain longitudinal observations y_{ij} from $\eta_{\mu i}(t)$ by adding independent errors $\epsilon_{ij} \sim N(0, 0.3^2)$ for each t_{ij} in $\mathbf{t}$. As the estimation showed stability issues in the most complex model in setting 3 for small samples we fit setting 3 also leaving more longitudinal observations by deleting only 80% of the simulated observations resulting in a median of 12 measurements per subject (IQR: 6, 18).

We fit the 1600 generated data sets ($(3+1)$ settings $\times$ 2 sample sizes $\times$ 200 replications) with our model implementation in **bamlss**. Additionally we compare our results in setting 1 and 2 with the linear estimation in **JMbayes**. For **bamlss** we estimate the longitudinal trajectories using P-splines

(Eilers and Marx, 1996) with cubic B-Splines, a second order difference penalty and 10 knots (2 internal knots) for the overall mean and the individual trajectories resulting in $5 \cdot n$ basis functions. The association $\eta_{\alpha i}(\eta_{\mu i}(t))$ is also modeled using P-Splines with 5 basis functions after imposing the constraint in setting 1 and 2, and for each of both groups in setting 3. In a few cases the posterior mode estimation led to extreme predictions in $\eta_{\mu i}(t)$ for single subjects. In these cases we reduced the number of coefficients for $\eta_{\alpha i}(\eta_{\mu i}(t))$ by 2 to stabilize the estimation. This occurred 2 and 4 times in setting 1, for small and large data sets, respectively, 3 and 2 times in setting 2, and 5 and 1 time in setting 3 with a median of 6 observations per person as well as 2 times each with a median of 12 observations per person. Further, the baseline hazard η_λ is estimated using P-Splines with 9 resulting basis functions. For setting 3 we allow the nonlinear association to vary between the two subgroups $\eta_{\alpha i}(\eta_{\mu i}(t), g_i)$. For comparison we also fit the data sets assuming a linear association with the log-hazard using **JMbayes** in setting 1 and 2 and try to achieve otherwise comparable models by modeling the nonlinear effects in the longitudinal submodel by unpenalized B-splines and the baseline hazard by P-splines. The number of knots were assessed in preliminary simulations to minimize the AIC resulting in 3 basis functions per subject with diagonal covariance matrix of the random effects for $n = 300$ and 4 basis functions per subject for $n = 600$. For the posterior mean estimation we sample for 13000 iterations, discard 3000 samples as burnin and keep 5000 samples per model after thinning.

In every estimated model we calculate mean-squared error (MSE), bias, and frequentist coverage of the 95% credibility interval both averaged over all time points and averaged per time point. For the predictors in the longitudinal model, i.e. $k \in \{\mu, \sigma\}$, the average MSE in each sample q is $MSE_k^q = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}[\hat{\eta}_{ki}^q(t_{ij}) - \eta_{ki}^q(t_{ij})]^2$ with the estimate $\hat{\eta}_{ki}$, and the MSE per time point is $MSE_k^q(t) = \frac{1}{n}\sum_{i=1}^{n}[\hat{\eta}_{ki}^q(t) - \eta_{ki}^q(t)]^2$ for all t in t^*. For the survival predictors η_γ and η_λ, the average MSE is $MSE_k^q = \frac{1}{n}\sum_{i=1}^{n}[\hat{\eta}_{ki}^q(T_i) - \eta_{ki}^q(T_i)]^2$ using evaluations at the subject's event times for η_λ and for the time-constant η_γ. For η_λ the error is additionally evaluated at the fixed grid of time points t^* as above. For the potentially nonlinear association $\eta_{\alpha i}(\eta_{\mu i}(t))$ a variety of different evaluations are possible. As the association is a survival predictor we compute the average error as $MSE_\alpha^q = \frac{1}{n}\sum_{i=1}^{n}[\hat{\eta}_{\alpha i}^q(\eta_{\mu i}(T_i)) - \eta_{\alpha i}^q(\eta_{\mu i}(T_i))]^2$. To assess the performance over the full range of the marker values and to assess deviations from a linear fit we also compute $MSE_\alpha^q(\eta_\mu^*) = [\hat{\eta}_\alpha^q(\eta_\mu^*) - \eta_\alpha^q(\eta_\mu^*)]^2$ where η_μ^* is from a fixed grid from -0.5 to 2 in 120 steps. This fixed grid was chosen as the maximum range of true values η_μ that were simulated in all settings. For setting 3, this measure is computed per group and then averaged over groups. All these error measures are then averaged over all Q samples per setting. Additionally we compute a point estimate of the average slope of the association as the averaged first derivative $\frac{1}{n}\sum_{i=1}^{n}\eta'_{\alpha i}(\eta_{\mu i}(T_i))$ of the estimated association in setting 1.

5.3.2 Simulation results

In setting 1 **bamlss** allows for an unbiased modeling of the linear association with satisfactory frequentist coverage of the credibility bands. All survival predictors show systematically less error when more information is available as for $n = 600$ (cf. Table 5.1). Only the predictor η_σ has a coverage

clearly below 0.95. However, as inference for this predictor is rarely of interest, this deviation is not deemed problematic. **JMbayes** achieves similar performance in setting 1 for most predictors, however the coverage for η_λ and also for η_α in the smaller data sets is below the nominal 0.95.

Table 5.1: Posterior mean estimation results from **bamlss** and **JMbayes** from setting 1 (linear η_α) for small and large data sets.

		MSE		bias		coverage	
		$n = 300$	$n = 600$	$n = 300$	$n = 600$	$n = 300$	$n = 600$
η_α	**bamlss**	0.025	0.016	−0.005	−0.006	0.976	0.958
	JMbayes	0.016	0.007	0.002	0.000	0.930	0.944
η_γ	**bamlss**	0.020	0.010	−0.003	0.018	0.953	0.951
	JMbayes	0.021	0.010	−0.016	0.015	0.950	0.954
η_λ	**bamlss**	0.042	0.024	−0.000	0.000	0.948	0.951
	JMbayes	0.043	0.024	0.000	0.000	0.915	0.933
η_μ	**bamlss**	0.031	0.031	−0.001	0.000	0.946	0.946
	JMbayes	0.039	0.030	−0.000	0.010	∗	∗
η_σ	**bamlss**	0.001	0.001	0.013	0.014	0.898	0.859
	JMbayes	0.009	0.000	0.093	0.008	∗	∗

∗ No credibility intervals and thus no coverage could be calculated for these predictors. Results are based on 186 estimates for $n = 300$ and 198 estimates for $n = 600$.

Despite a flexible nonlinear specification, our model captures the linearity well with a mean over all calculated average slopes of 0.99 [average 2.5% and 97.5% quantile of the posterior: 0.68; 1.32] for $n = 300$ and 0.96 [0.75; 1.19] for $n = 600$. The estimates of different iterations show less variability when more data is available, both when more subjects are observed and in areas where more observations of η_μ are made (see left panel of Figure 5.1). These results are highly similar to the respective linear estimates of **JMbayes** of 1.02 [0.74; 1.31] and 0.99 [0.79; 1.19], respectively. Note also that the difference of the average quantiles is not noticeably larger for **bamlss** despite a more flexible model formulation.

The estimation of the nonlinear model in **bamlss** shows some stability issues when less data is available such that initially 10% of the estimations for $n = 300$ and 4% of the estimations for $n = 600$ failed as they got stuck in areas of the parameter space where the Hessian for $\boldsymbol{\beta}_\mu$ was no longer negative definite. When restarting the algorithm is such cases with a different seed, these error rates decreased to 7% and 1%, respectively. Due to the flexibility in the model, especially in the random functional intercepts, the estimation of **bamlss** takes on average 3.6 and 7.3 hours for $n = 300$ and $n = 600$, respectively, compared to 4 and 7 minutes for **JMbayes** on a single core of a 2.6GHz Intel Xeon Processor E5-2650. This computation time can be reduced by using more than one core in the MCMC sampling in **bamlss** as implemented in the package for Linux systems.

A similar overall pattern is seen in setting 2 (cf. Table 5.2) for the estimation of **bamlss**: All estimates of the survival submodel are better with more data, and the coverage is satisfactory except for η_σ. The nonlinearity is captured in the estimation, as shown in Figure 5.1, although there is more uncertainty for very high and very low values of η_μ, where few observations are available. For the

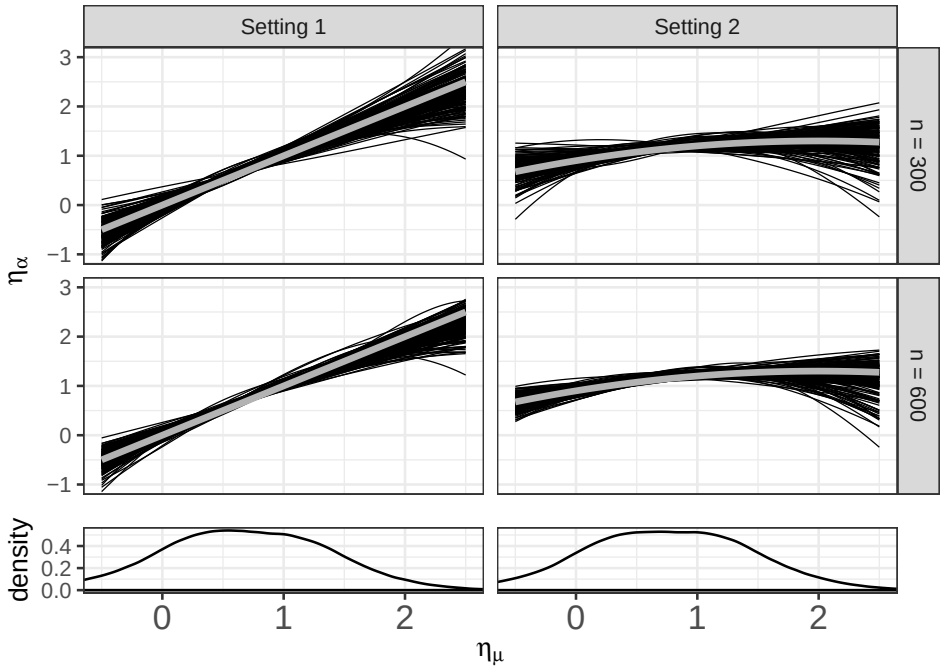

Figure 5.1: True (grey) and estimated (black) predictors from posterior mean estimation of setting 1 (true linear) and setting 2 (true nonlinear) for $n = 300$ and $n = 600$ as well as respective densities of the simulated η_μ. Displayed effects are subject to centering constraints as explained in 5.1.3

estimation in **JMbayes**, assuming linearity, the point estimates for the association are good, at least in this setting with small curvature of the association, however the coverage is very low under this misspecification with 0.70 and 0.66 for $n = 300$ and $n = 600$.
Again some stability issues emerge for **bamlss** in the smaller data setting where initially 15.5% of the estimations fail which was reduced to 7.5% by restarting the estimation with a different seed. For $n = 600$ initially 5% of the estimations produced an error which was reduced to 3%. Similarly to setting 1 the estimation takes on average 3.9 and 7.2 hours for $n = 300$ and $n = 600$, respectively.

In the most complex model in setting 3 where the association is nonlinear and group-specific, $\eta_{\alpha i}(\eta_{\mu i}(t), g_i)$, the estimation of this association is less precise than in setting 2 (cf. Table 5.3 as well as Figure 5.2). The estimates are mainly unbiased with only η_γ showing a slightly stronger negative bias for $n = 300$ and show a satisfactory coverage, except for η_σ. The estimates of the association $\eta_\alpha(\eta_{\mu i}(t), g_i)$ show higher amounts of variability as seen in Table 5.3 and Figure 5.2. The precision of the association estimate is generally higher for more subjects, with more longitudinal observations per subject and especially in the areas where η_μ is more densely observed. As in the previous simulations, less information about η_μ was available for the lower and higher values.

The most complex estimation of a nonlinear association also suffers most from stability issues such that 43.5% of the estimations in setting 3 for $n = 300$ as well as 18.5% of those for $n = 600$ fail for a median of 6 longitudinal observations per subject. These numbers reduced to 30% and 13% after

Table 5.2: Posterior mean simulation results from **bamlss** and results from **JMbayes** from setting 2 (nonlinear η_α) for small and large data sets.

		MSE		bias		coverage	
		$n = 300$	$n = 600$	$n = 300$	$n = 600$	$n = 300$	$n = 600$
η_α	**bamlss**	0.018	0.011	0.004	0.001	0.963	0.963
	JMbayes	0.016	0.010	0.008	0.007	0.702	0.656
η_γ	**bamlss**	0.017	0.009	−0.022	−0.017	0.941	0.955
	JMbayes	0.014	0.007	−0.006	−0.002	0.949	0.955
η_λ	**bamlss**	0.037	0.037	0.000	0.000	0.943	0.944
	JMbayes	0.031	0.020	0.000	0.000	0.914	0.922
η_μ	**bamlss**	0.032	0.032	−0.001	0.000	0.947	0.947
	JMbayes	0.039	0.031	−0.006	0.003	$*$	$*$
η_σ	**bamlss**	0.002	0.001	0.014	0.012	0.914	0.897
	JMbayes	0.008	0.000	0.085	0.007	$*$	$*$

$*$ No credibility intervals and thus no coverage could be calculated for these predictors. Results are based on 185 estimates for $n = 300$ and 194 estimates for $n = 600$.

restarting the algorithm with a different seed. Included in these problematic estimations are also 2 and 1 estimation, respectively, in which a low acceptance rate ($< 30\%$) in η_μ indicated sampling issues. In comparison, using more observations per subject results in error rates of only 13% and 4.5% which reduced to 5% and 1% after restarting with a different seed. Simulations took on average 4.4 and 8.4 hours for $n = 300$ and $n = 600$, respectively, for a median of 6 observations per subject and 4.6 and 9 hours for a median of 12 observations.

In conclusion, the simulations show that both truly linear associations and truly nonlinear associations can be modeled precisely and unbiasedly with the flexible additive joint model. Estimates are comparable between **bamlss** and **JMbayes**, however, the latter shows coverage issues, especially when truly nonlinear associations are present. The model is further able to distinguish between nonlinear associations of different subgroups. This estimation is however only feasible with enough data, both regarding the total number of subjects and the number of observations per subject and is more stable in areas where much longitudinal information is available. Stability issues in the estimation can be alleviated by restarting the algorithm with a different seed.

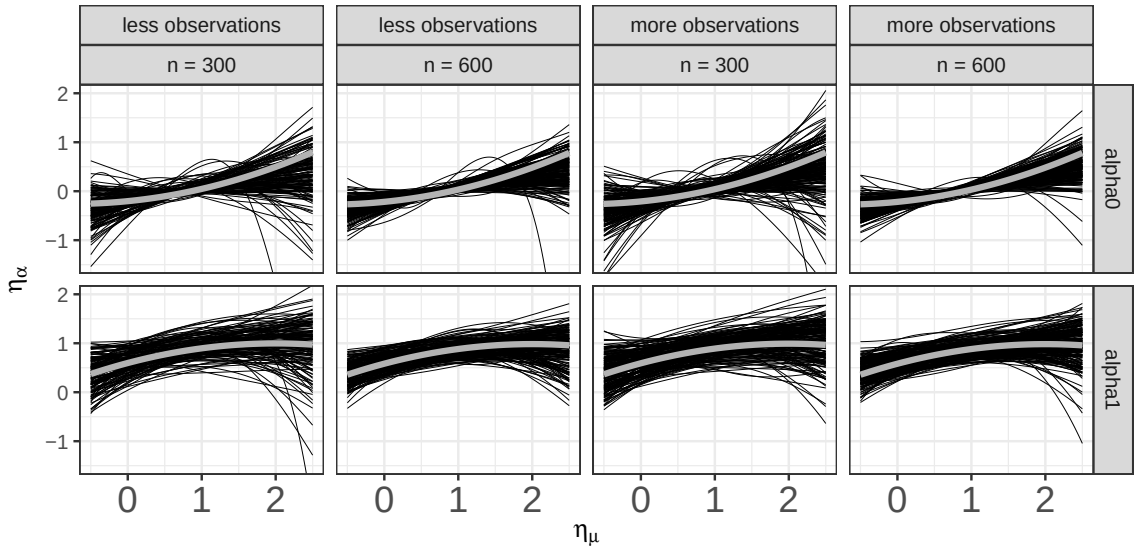

Figure 5.2: True (grey) and estimated (black) predictors from posterior mean estimates of $\eta_{\alpha i}(\eta_{\mu i}(t), g_i)$ in setting 3 for $g_i = 0$ and $g_i = 1$, $n = 300$ and $n = 600$ as well as for a median of 6 longitudinal observations per subject (less obs) and 12 observations per subject (more obs); displayed effects are subject to centering constraints as explained in Section 5.1.3

Table 5.3: Posterior mean simulation results from **bamlss** from setting 3 (nonlinear, group-specific η_α) using a median of 6 observations per subject or 12 observations per subject for small and large data sets.

		MSE		bias		coverage	
		$n = 300$	$n = 600$	$n = 300$	$n = 600$	$n = 300$	$n = 600$
η_α	less observations	0.082	0.062	0.013	−0.005	0.960	0.946
	more observations	0.058	0.028	0.018	0.004	0.953	0.945
η_γ	less observations	0.034	0.017	−0.072	−0.020	0.963	0.933
	more observations	0.030	0.017	−0.053	−0.023	0.969	0.938
η_λ	less observations	0.057	0.028	−0.000	0.000	0.942	0.937
	more observations	0.038	0.023	−0.000	0.000	0.955	0.946
η_μ	less observations	0.042	0.032	−0.002	0.000	0.946	0.944
	more observations	0.021	0.020	−0.000	0.000	0.945	0.945
η_σ	less observations	0.003	0.012	0.018	0.022	0.914	0.892
	more observations	0.002	0.004	0.010	0.020	0.921	0.817

Results are based on 140 and 176 estimates using a median of 6 observations per subject for $n = 300$ and $n = 600$, respectively, and 190 and 197 estimates using more observations per subject.

5.4 Application

We illustrate the flexible modeling approach on the widely used PBC biomedical data (Murtaugh et al., 1994), included in the R package **JMbayes**, which is concerned with the study of survival in subjects with a rare fatal liver disease. By reanalyzing this data set with the flexible additive joint model, assumptions and modeling alternatives can be tested. In more detail we aim to assess the adequacy of the linearity assumption of the association between marker and log-hazard and are interested in the best transformation of the marker. Our framework allows us to check several transformations and base a decision on the DIC and/or residual diagnostics without having worry about a potentially resulting nonlinear association between the transformed marker and the log-hazard. Lastly, the analysis of subgroups regarding their association between marker and log-hazard is of interest.

In this study 312 subjects were followed in the Mayo Clinic from 1974 to 1984 to study the influence of the drug D-penicillamine on the survival of the patients. Visits were scheduled at six months, 12 months and annually thereafter. In the dataset 140 subjects died during follow-up with a median survival time of 3.72 years (IQR: 2.08, 6.66) and 172 survived of which 29 received a transplant with a median censoring time of 7.77 (IQR: 5.73, 9.91). In total there are 1945 longitudinal observations with a median number of visits per subject of 5 (IQR: 3, 9).

To illustrate the general framework we model the survival of PBC-patients as a function of the baseline covariates medication (drug vs. placebo), age at study entry in years and the presence of an enlarged liver at baseline. We chose these baseline covariates based on previous joint model analyses of the data (Rizopoulos, 2012, 2016a). The focus of the analysis is the association between the levels of serum bilirubin, a biomarker expected to be a strong indicator of disease progression, and the log-hazard of death. To account for individual nonlinear marker trajectories we model the levels of serum bilirubin using functional random intercepts with 5 basis functions per subject.

To further explore the influence of the marker parameterization on the association we fit three models, differing in their association between serum bilirubin and survival. First, we model serum bilirubin using the log-transformed marker $\log(Bilirubin)$, as previously used in (Rizopoulos, 2012, 2016a) and allow the association to be nonlinear. Second, we use a square-root transformation of the raw marker values $\sqrt{Bilirubin}$ and again allow the association to be nonlinear. Third, we allow the non-linear association between $\log(Bilirubin)$ and the log-hazard to also vary between the patients with an enlarged liver at baseline and those without. This predictor η_α is parameterized as potentially nonlinear effect for both groups, subject to the sum-to-zero constraint as explained in Section 5.1.3, with an additional intercept for the group of subjects with an enlarged liver to allow not only for differences in the nonlinearity of the biomarker effect but also in the overall level. As the group difference for the hazard is already included in η_α, the baseline effect of an enlarged liver not included η_γ in model 3 to avoid redundancy. As our focus lies primarily on the association between the biomarker and survival, and to avoid instabilities in the estimation, we censor subjects 1 year after their last longitudinal measurement.

In all three models, and in line with previous analyses, the treatment is not associated with survival (log-hazard effect estimate [95% credibility interval]: model 1: -0.03 [-0.42; 0.34]; model 2: -0.02 [-0.42; 0.36]; model 3: -0.01 [-0.39; 0.39]) whereas age at baseline is positively associated with the hazard of death (model 1: 0.05 [0.03; 0.07]; model 2: 0.05 [0.04; 0.07]; model 3: 0.05 [0.03; 0.07]). Additionally subjects with an enlarged liver at baseline have a higher risk of dying in model 1 (0.76 [0.29; 1.21]) as well as in model 2 (0.77 [0.32; 1.21]). In model 3 this effect is included in the group-specific intercept for the association with the marker. Irrespective of their marker value, subjects with an enlarged liver at baseline have a higher log-hazard for the event (0.49 [-0.36; 1.45]) though this effect has a wider credibility band in model 3.

The focus of interest is the nonlinearly modeled association predictor η_α. As Figure 5.3 shows, the association between marker and the log-hazard for the event is linear when using the log-transformed marker $\log(Bilirubin)$ and nonlinear when transformed differently as $\sqrt{Bilirubin}$. In model 3 the groups differ in their overall level, although the credibility interval of the intercept coefficient covers 0. Additionally the slope of the association is highly similar in both groups. When comparing the models via DIC, model 1 achieves the lowest DIC (1876.76) followed by model 3 (1889.67) and 2 (2194.58).

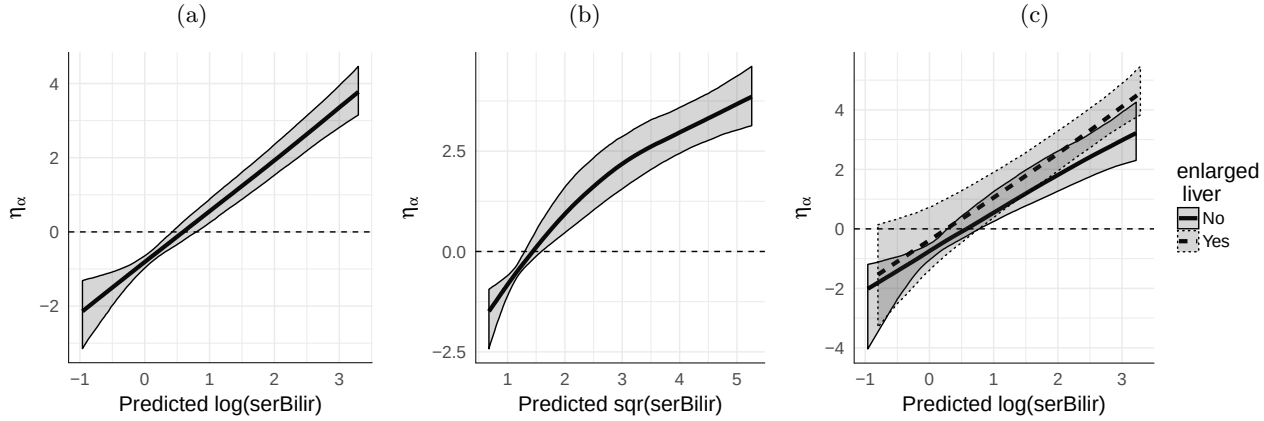

Figure 5.3: Estimated posterior mean of the association $\eta_\alpha(\eta_\mu(t))$ in the PBC data. (a) model 1: nonlinear estimation of $\log(Bilirubin)$ (b) model 2: nonlinear estimation of $\sqrt{Bilirubin}$ (c) model 3: nonlinear estimation of $\log(Bilirubin)$ of patients with and without enlarged liver at baseline.

Traceplots of the estimated coefficients $\boldsymbol{\beta}_\alpha$ as well as results from sensitivity analyses for the variance parameters of these results using different priors (differently specified IG and Half-Cauchy hyperpriors), showing robustness of the results, can be found in the Appendix B.3.

Our flexible joint model thus allowed us to check previously made model assumptions for this data set and to conclude that in this particular case, a linear association that is not covariate-dependent is sufficient to model the relationship between the log-marker and the log-hazard. Additionally, nonlinear associations can also be captured in real data if necessary, as shown for the square-root transformation in model 2. The model potentially further allows to observe group-specific nonlinear

association structures for subgroups of subjects, even though no strong group structure was present in this data set.

5.5 Discussion and outlook

In this work a highly flexible additive joint model is presented, which allows for nonlinear, potentially covariate-dependent association structures between marker values and the log-hazard of an event. The benefits and challenges of this flexibility were shown based not only on simulated data but also on the well-known PBC data set.

Using this new model the generally unchecked linearity assumption as well as the appropriateness of transformations of marker values can be assessed in the context of joint models. This is particularly important if marker values need to be transformed to better fulfill the normality assumption in the longitudinal submodel and different transformations are compared. It is clear that several transformations cannot fulfill the linearity and normality assumption simultaneously and relaxing the linearity assumption allows to choose the most appropriate model in terms of residual normality and/or DIC. The modeling of nonlinear associations between a longitudinal marker and the log-hazard does not only avoid bias but also allows further insights into underlying disease mechanisms. Additionally, subgroups of subjects with different marker associations can be identified.
The simulation results show that our model can identify truly linear as well as truly nonlinear associations. We used the model to check the linearity assumption when using transformed Bilirubin values in the PBC data set and could confirm that the association is linear if $\log(Bilirubin)$ is used, while using $\sqrt{Bilirubin}$ would necessitate estimating a nonlinear association structure.

This flexible modeling, however, also comes at a price. When modeling longitudinal trajectories using flexible functional random intercepts and allowing for nonlinear association structures, many subjects and a dense grid of measurements until the event time are necessary in order to achieve a stable estimation. Further there should not be large gaps between the latest longitudinal measurements and the event time to allow for a stable estimation. If these gaps are present in real data, censoring as in Section 5.4 can alleviate the stability issue. Additionally the estimation takes more time than standard joint models but can be parallelized if corresponding computing facilities are available.

As a next step we aim at stabilizing the estimations further. One potential approach is the joint updating and sampling of coefficients in η_α and η_μ instead of the current separate block-wise procedure. Besides, the updating and sampling algorithm could be modified by using a transformed Hessian matrix to ensure that an inverse exists, similar to the modified Marquardt algorithm used by Proust-Lima et al. (2007). In addition, we plan to implement additional nonlinear association structures within the R package **bamlss** and to speed up the computations further in order to allow for a broader usage of this flexible additive joint model framework.

Chapter 6

Model implementation in the R package bamlss

Joint models of longitudinal and survival data are important tools for the unbiased analysis of longitudinal cohort data. In order for joint models to find entry into the tool set of applied researchers, the availability of ready-to-use software is crucial (Gould et al., 2015; Yuen and Mackinnon, 2016). For this reason the presented flexible Bayesian additive joint model was incorporated in the R **bamlss** (Bayesian Additive Models for Location Scale and Shape, Umlauf et al., 2017). In the following the main functionalities for fitting joint models in **bamlss** are presented.

Joint models allow the unbiased analysis of longitudinal covariates, measured at discrete subject-specific time points, potentially subject to measurement error and their association with a time-to-event process. The main idea behind this modeling approach is to formulate a joint likelihood of a longitudinal submodel, usually a mixed effects model, and a survival model. The two submodels are assumed to be independent given latent parameters which influence both, the longitudinal and the survival model. In many recent joint model developments these shared latent parameters are random effects, as is the case in **bamlss**.

Multiple software packages have been developed for estimating joint models with shared random effects structure in the frequentist setting such as the R package **JM** (Rizopoulos, 2010) and **joineR** (Philipson et al., 2017), as well as the stata module **stjm** (Crowther, 2012) and the SAS macro **JMFit** (Zhang et al., 2016). All implementations differ in their specification of the association between the longitudinal marker and the time-to-event process and in the flexibility for modeling the observed marker trajectories. For example, **JMfit** and **joineR** only allow a random effects association and restrict the random effects in the longitudinal model to a random intercept, slope and a possible quadratic random effects term while the baseline hazard is unspecified or modeled piece-wise constant, respectively, for the two implementations. The packages **JM** and **stjm** allow more flexibility in the longitudinal trajectories, which can be modeled also using splines (both) or fractional polynomials (**stjm**), by allowing parametric (both) and nonparametric (**JM**) baseline hazard functions and different association structures such as current value, current slope and random effects only. However, as presented in Section 2.5, the frequentist estimation can become unstable or unfeasible when flexible

random effects structures are necessary. Many flexible joint models are therefore developed in a Bayesian framework with estimation procedures making use of the BUGS language (WinBUGS and OpenBUGS, Gilks et al., 1994) or JAGS (Plummer et al., 2003) and respective code available from authors (see for example overviews in Sudell et al., 2016; Hickey et al., 2016). To our knowledge the R package **JMbayes** (Rizopoulos, 2016b) is the only ready-to-use software implementation besides **bamlss** to allow the fitting of joint models in a Bayesian framework. Notably, **JMbayes** allows for much flexibility in the modeling of parametric and nonparametric baseline hazards, spline-based nonlinear longitudinal trajectories, different association parameterizations such as current value, slope or cumulative effects and dynamic predictions.
The implementation of the flexible additive joint model in **bamlss** contributes to the existing software as it offers even greater flexibility in the specification of subject-specific longitudinal trajectories, different association structures as well as a variety of possible effect specifications in the baseline survival and time-varying survival effects.

In the remainder of this chapter central aspects of the flexible additive joint model are presented and the core structure of the package **bamlss** is explained. The main functionalities for fitting the joint model are then illustrated using three different applications. The chapter closes with an outlook of planned future developments. The models were fitted using **bamlss version 0.1-2** which makes strong use of **mgcv version 1.18** (Wood, 2011). Examples are based on analysis of simulated data and the PBC data set (Murtaugh et al., 1994) which is included in the R package **JMbayes** (Rizopoulos, 2016b).

6.1 Theoretical background

As the framework for the flexible additive joint models was presented extensively in the previous chapters, we restrict ourselves to a quick overview of central concepts of the framework.

6.1.1 Model

We observe for each subject $i = 1, \ldots, n$ a longitudinal response $\boldsymbol{y}_i = [y_{i1}, \cdots, y_{in_i}]^\top$ at potentially subject-specific time points $\mathbf{t}_i = [t_{i1}, \cdots, t_{in_i}]^\top$ where $t_{i1} \leq \cdots \leq t_{in_i} \leq T_i$. This observed longitudinal response is modeled by

$$y_{ij} = \eta_{\mu i}(t_{ij}) + \varepsilon_{ij} \text{ with } \varepsilon_{ij} \sim N(0, \exp[\eta_{\sigma i}(t_{ij})]^2), \tag{6.1}$$

where η_μ denotes the "true" longitudinal marker linking the longitudinal and the survival submodel. We further observe for every subject i a potentially right-censored follow-up time T_i and the event indicator δ_i, with 1 indicating an event. We use structured additive predictors η_k, $k \in \{\lambda, \gamma, \alpha, \mu\}$ to model the hazard of an event at time t as

$$h_i(t) = \exp\{\eta_i(t)\} = \exp\{\eta_{\lambda i}(t) + \eta_{\gamma i} + \eta_{\alpha i}(\eta_{\mu i}(t), t)\}. \tag{6.2}$$

These structured additive predictors are the building blocks of the implementation in **bamlss** and are in more detail

- η_λ denoting the predictor for all time-varying survival covariates and effects including the log baseline hazard.
- η_γ representing all baseline survival covariates, potentially including smooth or spatial effects.
- η_μ as the potentially nonlinear, subject-specific modeled marker.
- η_σ representing the variance of the longitudinal marker with an exponential link function.
- η_α denoting the association function between longitudinal marker and time-to-event which can be linear, nonlinear, covariate- or time-dependent.

Each of these predictors η_{ki} with $k \in \{\lambda, \gamma, \mu, \sigma\}$ has an additive structure $\eta_{ki} = \sum_{m=1}^{M_k} f_{km}(\tilde{\boldsymbol{x}}_{kmi})$ of M_k functions of observed covariates $\tilde{\boldsymbol{x}}_i$. Each function f_{km} typically depends on one or two covariates $\tilde{\boldsymbol{x}}_{kmi}$ and can further be time-dependent such that $\eta_{ki}(t) = \sum_{m=1}^{M_k} f_{km}(\tilde{\boldsymbol{x}}_{kmi}(t), t)$. By using suitable (e.g. spline) basis matrices $\mathbf{X}_{km}$ and corresponding penalties $\mathbf{P}_{km}$ every term m of predictor k is set up as a linear predictor

$$\mathbf{X}_{km}\boldsymbol{\beta}_{km} \text{ with penalty } \mathbf{P}_{km} = \frac{1}{\tau_{km}^2}\boldsymbol{\beta}_{km}^\top \mathbf{K}_{km}\boldsymbol{\beta}_{km},$$

where $\boldsymbol{\beta}_{km}$ is the coefficient vector of length p_{km} that is estimated. The association between the longitudinal marker and the log-hazard is also a structured additive predictor, however with the special structure $\eta_{\alpha i}(\eta_{\mu i}(t), t) = f_\alpha\left(\eta_{\mu i}(t), \tilde{\boldsymbol{x}}_{\alpha i}, t\right)$ with

$$f_\alpha\left(\eta_{\mu i}(t), \tilde{\boldsymbol{x}}_{\alpha i}, t\right) = \left[\boldsymbol{g}_1(\eta_{\mu i}(t)) \odot \boldsymbol{g}_2(\tilde{\boldsymbol{x}}_{\alpha i}, t)\right]\boldsymbol{\beta}_\alpha = \boldsymbol{x}_{\alpha i}^\top \boldsymbol{\beta}_\alpha. \tag{6.3}$$

This predictor, where we drop the subscript m for simplicity, is the row tensor product of a function of η_μ and further covariates or time. We refer to Chapter 5 for details on the specification of this association but stress at this point the distinction between linear and nonlinear associations which is relevant for the model setup in **bamlss**. Whereas for nonlinear associations the full row-tensor product in eq. (6.3) is used, the association simplifies to $\eta_{\alpha i}(\eta_{\mu i}(t), t) = \eta_{\alpha i}(t)\eta_{\mu i}(t)$ with $\eta_{\alpha i}(t)$ having the same structure as the other predictors for linear associations.
As all of these predictors initially contain an intercept, constraints are implemented to obtain an identifiable model as explained in more detail in the Sections 3.1.2 and 5.1.3. These force the predictors to sum to zero over all n or N observations for predictors in the survival and longitudinal submodel, respectively, and for a fixed grid of observed marker values for a nonlinearly specified association $\eta_{\alpha i}(\eta_{\mu i}(t))$. The global intercept of the survival model in eq. (6.2) is included in η_γ.

6.1.2 Estimation

Assuming conditional independence of the survival outcomes $[T_i, \delta_i]$ and the longitudinal outcomes $\boldsymbol{y}_i$, given the vector of all parameters $\boldsymbol{\theta}$, the log-likelihood of this joint model is the sum of the log-likelihoods of the two submodels, i.e.,

$$\begin{aligned}\ell\left[\boldsymbol{\theta}|\boldsymbol{y}, \mathbf{T}, \boldsymbol{\delta}\right] = &- \frac{N}{2}\log(2\pi) - \mathbf{1}_N^\top \boldsymbol{\eta}_\sigma\left(\mathbf{t}\right) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{\eta}_\mu\left(\mathbf{t}\right))^\top \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{\eta}_\mu\left(\mathbf{t}\right)) \\ &+ \boldsymbol{\delta}^\top \boldsymbol{\eta}(\mathbf{T}) - \mathbf{1}_n^\top \boldsymbol{\Lambda}\left(\mathbf{T}\right).\end{aligned}$$

Here $\boldsymbol{R} = \text{blockdiag}(\boldsymbol{R}_1, \cdots, \boldsymbol{R}_n)$ is assumed diagonal, $\boldsymbol{y} = [\boldsymbol{y}_1^\top, \cdots, \boldsymbol{y}_n^\top]^\top$, $\mathbf{T} = [T_1, \cdots, T_n]^\top$ and $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_n]^\top$ are the response vectors and $\boldsymbol{\Lambda}(\mathbf{T}) = [\Lambda_1(T_1), \ldots, \Lambda_n(T_n)]^\top$ denotes the vector of cumulative hazard rates with $\Lambda_i(T_i) = \exp(\eta_{\gamma i}) \int_0^{T_i} \exp[\eta_{\lambda i}(u) + \eta_{\alpha i}(\eta_{\mu i}(u), u)]du$.

In this general setup different effects, such as random effects per subject or smooth nonlinear effects of time and/or covariates are induced by the specification of appropriate design matrices and respective priors. For linear or parametric terms a precision matrix $\mathbf{K}_{km} = \mathbf{0}$ is approximated by using vague normal priors $\boldsymbol{\beta}_{km} \sim N(\mathbf{0}, 1000^2\boldsymbol{I})$ on the vectors of the regression coefficients. Smooth and random effect terms with precision matrix $\mathbf{K}_{km}$ are regularized with multivariate normal priors

$$p(\boldsymbol{\beta}_{km}|\tau_{km}^2) \propto \left(\frac{1}{\tau_{km}^2}\right)^{\frac{\text{rank}(\mathbf{K}_{km})}{2}} \exp\left(-\frac{1}{2\tau_{km}^2}\boldsymbol{\beta}_{km}^\top \mathbf{K}_{km}\boldsymbol{\beta}_{km}\right)$$

with variance parameter τ_{km}^2 which control the trade-off between flexibility and smoothness in the nonlinear modeling of effects. For anisotropic smooths, which induce a different amount of smoothness across different dimensions by using different variance priors $\boldsymbol{\tau}_{km}^2$ for the different precision matrices, as in the nonlinear association η_α or in fitting flexible longitudinal trajectories as in Section 6.4.2, we use the prior

$$p(\boldsymbol{\beta}_{km}|\boldsymbol{\tau}_{km}^2) \propto \left|\frac{1}{\tau_{km1}^2}\tilde{\boldsymbol{K}}_{km1} + \frac{1}{\tau_{km2}^2}\tilde{\boldsymbol{K}}_{km2}\right|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\beta}_{km}^\top\left[\frac{1}{\tau_{km1}^2}\tilde{\boldsymbol{K}}_{km1} + \frac{1}{\tau_{km2}^2}\tilde{\boldsymbol{K}}_{km2}\right]\boldsymbol{\beta}_{km}\right),$$

where $\tilde{\boldsymbol{K}}_{km1} = \mathbf{K}_{km1} \otimes \boldsymbol{I}_{p_{km2}}$ and $\tilde{\boldsymbol{K}}_{km2} = \mathbf{K}_{km2} \otimes \boldsymbol{I}_{p_{km1}}$. For the variance parameters independent inverse Gamma hyperpriors $\tau_{km}^2 \sim IG(0.001, 0.001)$ are used, resulting in an inverse Gamma full conditional. Further priors for the variance parameters, such as half-Cauchy, are possible.

The resulting posterior of the full model is

$$p(\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{\delta}, \boldsymbol{y}) \propto L^{\text{long}}\left[\boldsymbol{\theta}|\boldsymbol{y}\right] \cdot L^{\text{surv}}\left[\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{\delta}\right] \prod_{k\in\{\lambda,\gamma,\alpha,\mu,\sigma\}} \prod_{m=1}^{M_k} \left[p(\boldsymbol{\beta}_{km}|\boldsymbol{\tau}_{km}^2)p(\boldsymbol{\tau}_{km}^2)\right],$$

with the likelihood of the longitudinal submodel L^{long} and the survival submodel L^{surv} and the priors of the vectors of regression parameters and variance parameters, $p(\boldsymbol{\beta}_{km}|\boldsymbol{\tau}_{km}^2)$ and $p(\boldsymbol{\tau}_{km}^2)$, respectively.

Estimates of this posterior distribution are the posterior mode and the posterior mean, where the posterior mode is mainly used for a quick model assessment and to obtain starting values for the MCMC posterior mean sampling. The posterior mode estimate is obtained by maximizing the log-posterior using a Newton-Raphson algorithm. Every term m in each predictor k is updated block-wise in each iteration l as

$$\boldsymbol{\beta}_{km}^{[l+1]} = \boldsymbol{\beta}_{km}^{[l]} - \nu_{km}^{[l]} \boldsymbol{H} \left(\boldsymbol{\beta}_{km}^{[l]}\right)^{-1} \boldsymbol{s}\left(\boldsymbol{\beta}_{km}^{[l]}\right)$$

with steplength $\nu_{km}^{[l]}$, the score vector $\boldsymbol{s}(\boldsymbol{\beta}_{km})$ and the Hessian $\boldsymbol{H}(\boldsymbol{\beta}_{km})$. For detailed expressions of the block-wise score and the Hessian we refer to the Appendix A. The steplength $\nu_{km}^{[l]}$ is optimized in each step to maximize the log-posterior. Furthermore, the variance parameters are optimized to minimize the corrected AIC (AICc, Hurvich et al., 1998). Approximate credibility intervals for the coefficients are derived from $N(\hat{\boldsymbol{\beta}}_{km}, [-\boldsymbol{H}(\hat{\boldsymbol{\beta}}_{km})]^{-1})$ assuming an approximately normal posterior distribution. These credibility intervals, however, tend to underestimate the variability as they do not take into account the optimization of the variance parameters. To obtain posterior mean estimates and derive exact inference a derivative-based Metropolis-Hastings algorithm is used. Here the full conditional $\pi(\boldsymbol{\beta}_{km}|\cdot)$ is approximated by a second order Taylor expansion of the log-posterior centered at the last state $\beta_{km}^{[l]}$ (see Umlauf et al., 2017) resulting in a multivariate normal proposal density with the precision matrix $(\boldsymbol{\Sigma}_{km}^{[l]})^{-1} = -\boldsymbol{H}(\boldsymbol{\beta}_{km}^{[l]})$ and the mean $\boldsymbol{\mu}_{km}^{[l]} = \boldsymbol{\beta}_{km}^{[l]} - \boldsymbol{H}(\boldsymbol{\beta}_{km}^{[l]})^{-1}\boldsymbol{s}(\boldsymbol{\beta}_{km}^{[l]})$. Candidates $\boldsymbol{\beta}_{km}^{*}$ from the proposal density $q(\boldsymbol{\beta}_{km}^{*}|\boldsymbol{\beta}_{km}^{[l]}) = N(\boldsymbol{\mu}_{km}^{[l]}, \boldsymbol{\Sigma}_{km}^{[l]})$ are drawn in each iteration l of the Metropolis-Hastings sampler for updating block km. By drawing candidates from a close derivative-based approximation of the full conditional we achieve high acceptance rates and good mixing. Samples for the variance parameters τ_{km}^2 are obtained via Gibbs sampling, if inverse Gamma hyperpriors are used and consequently the full conditionals $\pi(\tau_{km}^2|\cdot)$ follow an inverse Gamma distribution. Slice sampling (Neal, 2003) is used if no simple closed-form full conditional can be obtained, when sampling variance parameters for anisotropic smooths or when other hyperpriors than the inverse Gamma are used.

6.1.3 Diagnostics

We suggest to use DIC for model choice which can be directly computed from the MCMC output $\boldsymbol{\beta}^{[1]}, \ldots, \boldsymbol{\beta}^{[l]}, \ldots, \boldsymbol{\beta}^{[L]}$ of the full parameter vector $\boldsymbol{\beta}$ as

$$DIC = \overline{D(\boldsymbol{\beta})} + pd = 2\overline{D(\boldsymbol{\beta})} - D(\overline{\boldsymbol{\beta}})$$

where $\overline{D(\boldsymbol{\beta})} = \frac{1}{L}\sum_{l=1}^{L} D(\boldsymbol{\beta}^{[l]})$ with $D(\boldsymbol{\beta}) = -2\ell\left[\boldsymbol{\theta}|\boldsymbol{y}, \mathbf{T}, \boldsymbol{\delta}\right]$, $D(\overline{\boldsymbol{\beta}}) = D\left(\frac{1}{L}\sum_{l=1}^{L}\boldsymbol{\beta}^{[l]}\right)$ and with the effective number of parameters $pd = \overline{D(\boldsymbol{\beta})} - D(\overline{\boldsymbol{\beta}})$. Note that this computation is similar in spirit to Brown et al. (2005) but differs from the computation in **JMbayes**. DICs of similar models from **bamlss** and **JMbayes** are therefore not directly comparable.

Furthermore, conditional survival probabilities can be derived from the posterior samples. As shown in eq. (2.10) the probability of survival at time u is

$$S_i(u) = \exp\left\{-\int_0^u \exp(\eta_i(s))ds\right\}.$$

As the observed marker $\boldsymbol{y}_i$ is an internal time-varying covariate only observed until the occurrence of an event, the measurement of a marker value $y_i(t)$ at time u, $0 < t < u$ implies $Pr(T_i^* \geq t|y_i(t)) = 1$, see also eq. (2.13). Therefore, a survival probability $\rho_i(u|t)$ conditioned on the survival up to a certain time point t, e.g. the time point where a marker was last observed, is more relevant in the joint model context (Rizopoulos, 2012). Following Taylor et al. (2013) this conditional probability can be described as

$$\rho_i(u|t) = Pr(T_i^* \geq u|T_i^* > t, \boldsymbol{\theta}, \tilde{\boldsymbol{x}}_i) = \exp\left\{-\int_t^u \exp(\eta_i(s))ds\right\}.$$

Estimates of this conditional survival probability can be derived from the posterior distribution, as for example the mean estimate

$$\hat{\rho}_i(u|t) = \frac{1}{L}\sum_{l=1}^{L} \exp\left\{-\int_t^u \exp\left(\eta_i(s)^{[l]}\right) ds\right\} \tag{6.4}$$

where $\eta_i(s)^{[l]}$ is the estimated predictor of sample l and the integral is approximated numerically. Credibility intervals can be derived analogously.

6.2 The R package bamlss

The R package **bamlss** (Umlauf et al., 2017) is a highly flexible package to fit complex distributional regression models (Klein et al., 2015a). Different model families can be fitted within that framework using the following basic building blocks:

- structured additive predictors using a basis function representation
- priors to induce the effects structures such as smooth, random effects and time-varying
- likelihood functions, scores and Hessians for the estimation of the posterior

In order to fit joint models using the family `"jm"` the corresponding predictors, priors, likelihood function and derivatives were implemented in the package. The main fitting function is

```
bamlss(formula, family = "jm", data, timevar, idvar)
```

in which the predictors can be specified in `formula` and necessary covariates `timevar` for the longitudinally observed time variable and `idvar` for the subject identifier have to be supplied. For every predictor, a variety of effects can be specified using the functionalities of the R package

mgcv. With a simple formula interface, **mgcv** sets up the design matrices and corresponding penalty matrices for every term m of predictor k allowing for high flexibility in the estimation.
While the general model setup is implemented in the joint model in **bamlss** many aspects of the estimation can be user-specified. Though we mainly use Bayesian P-splines in the model estimation due to their favorable properties, all spline options that are available in **mgcv** can be used within **bamlss**. Additionally, different prior specifications for the variance parameters can be given as input (see Umlauf et al., 2017, Table 1).

Even with the fast convergence and good mixing performance of the algorithm, allowing this flexibility and using the derivative-based Metropolis-Hastings algorithm is a high computational burden. Hence, much effort was placed in speeding up computations. A special challenge in the fitting of joint models is the setup of time-varying predictors and the necessary integration over the subject-specific hazard function to obtain the cumulative hazard $\Lambda_i(T_i)$. We calculate this by numerical integration using the trapezoidal rule over a subject-specific grid of integration points and migrated this computation to C. Furthermore, the computation of the highly flexible functional random intercepts is computationally challenging as many parameters per subject have to be estimated. The design matrix of this term, however, has a sparse block-structure and the Hessian of this term is a block diagonal. We achieved large gains in computational speed by taking advantage of this structure in the posterior mean sampling.

6.3 Simulating joint model data

The package includes an elaborate simulation setup to generate user-specified data in the joint model framework. The function `simJM()` is a wrapper, specifying all predictors η_k for $k \in \{\lambda, \gamma, \alpha, \mu, \sigma\}$ for the function `rJM()`, which simulates survival times T_i based on Bender et al. (2005).

The starting point for this approach for simulating survival times is the distribution function $F(t) = 1 - \exp(-\Lambda(t))$ (see Section 2.4.1). If we let Y be a random variable with distribution function F, then the random variable $R := F(Y)$ follows a uniform distribution on $[0, 1]$, i.e., $R \sim U(0, 1)$ and the same holds for $(1 - R)$. In consequence, for the survival time T it holds that

$$R = \exp(-\Lambda(T)) \sim U(0, 1) \quad \text{and} \quad T = \Lambda^{-1}[-\log(R)],$$

assuming $h(t) > 0$ for all t. Whereas closed-form distributions for this inverse function exist for time-constant survival models with parametric baseline hazard (see Bender et al., 2005), this is not the case in our setup. Here survival times are generated by numerical integration of $\Lambda(t)$ and solving the equation $\Lambda(t) = -\log(R)$ where R is simulated using a random number generator over $[0, 1]$. The numerical integration and the solving of this equation by finding the root of $\Lambda(t) + \log(R) = 0$ is the core functionality of `rJM()`.

The wrapper `simJM()` sets up all true predictors k. Certain specifications to model true predictors are predefined, but writing a specific wrapper for simulating other settings is straightforward. The

function `rJM()` merely needs to be supplied with the hazard function `hazard()`, functions defining the predictors (`lambda()`, `gamma()`, `alpha()`, `mu()`, `sigma()`) as well as relevant simulated covariates (`x` for fixed covariates and `r` for random effects). For most situations, however, it should suffice to use the default predictor functions of `simJM()` for the simulations. The baseline hazard `lambda`, $\eta_\lambda(t) = 1.4\log((t+10)/1000)$, is fixed for all settings whereas the baseline survival predictor `gamma` differs between linear and nonlinear association settings with $\eta_{\gamma i} = \sin(\tilde{x}_{\gamma i})$ and $\eta_{\gamma i} = 0.3\tilde{x}_{\gamma 1}$, respectively, and $\tilde{x}_{\gamma 1} \sim U(-3, 3)$ for both. The error variance of the longitudinal marker `sigma` can be supplied in the function call of `simJM()` and defaults to $\exp(\eta_\sigma)^2 = 0.3^2$. For the further predictors different parameterizations can be supplied in the function call. For η_μ three different settings are possible and supplied via the argument `long_setting` from

- `linear`, for a random slope random intercept model with $\eta_{\mu i}(t) = 1.25 + b_{0i} - 0.01t + 0.02t \cdot b_{1i} + 0.6\sin(\tilde{x}_{\mu i})$,
- `nonlinear`, for a model with nonlinear population effect and random intercept with $\eta_{\mu i}(t) = 0.55 + b_{0i} - 0.1(t+2) \cdot \exp(-0.075t) + 0.6\sin(\tilde{x}_{\mu i})$,
- `functional`, for subject-specific nonlinear trajectories with $\eta_{\mu i}(t) = 0.55 + b_{0i} - 0.1(t+2) \cdot \exp(-0.075t) + 0.6\sin(\tilde{x}_{\mu i}) + f_i(t)$,

where $f_i(t)$ denotes subject-specific P-splines for which the number of knots can be supplied via `long_df`. Similarly, η_α is also pre-specified in different settings via `alpha_setting` and further differs for the specification of linear and nonlinear associations between η_μ and the log-hazard, respectively. Table 6.1 gives an overview of the different possible settings for η_α where $\tilde{x}_{\alpha i} \sim binom(0.5)$. Note that the settings for η_α and η_μ are modular and can be combined freely.

Table 6.1: Possible effect specifications for $\eta_{\alpha i}(\eta_{\mu i}(t), t, \tilde{\boldsymbol{x}}_{\alpha i})$ in the simulation function `simJM()` as specified in `alpha_setting`

	linear association $\eta_{\alpha i}(t) \cdot \eta_{\mu i}(t)$	nonlinear association $\eta_{\alpha i}(\eta_{\mu i}(t), \tilde{\boldsymbol{x}}_{\alpha i})$
`zero`	0	0
`constant`	1	1
`linear`	$(1 - 0.015t) \cdot \eta_{\mu i}(t)$	$1 \cdot \eta_{\mu i}(t)$
`nonlinear`	$(\cos((t-60)/20) + 1) \cdot \eta_{\mu i}(t)$	$-0.1(\eta_{\mu i}(t) + 3)^2 + \eta_{\mu i}(t) + 1.8$
`nonlinear2`	$(\cos((t-33)/33) + 1) \cdot \eta_{\mu i}(t)$	$\tilde{x}_{\alpha i} \cdot [-0.1(\eta_{\mu i}(t) + 3)^2 + \eta_{\mu i}(t) + 1.8] +$ $(1 - \tilde{x}_{\alpha i}) \cdot [0.1(\eta_{\mu i}(t) - 3)^2 + 0.75\eta_{\mu i}(t) - 0.7]$

To illustrate the usage of this simulation function, data for a simple joint model with fixed and random intercepts and slopes in the longitudinal model and a time-constant linear association between current marker values and the time-to-event is simulated for 150 subjects.

```
library("bamlss")
d <- simJM(nsub = 150, times = seq(0, 120, 1), probmiss = 0.75,
```

```
          long_setting = "linear", alpha_setting = "constant",
          dalpha_setting = "zero", full = FALSE, seed = 5)
```

This call generates the covariates, random effects and true predictors on a fixed grid of time points (`times`) based on the hazard as given in eq. (6.2) for 150 subjects. After the simulation of survival times, the longitudinal observations are censored at the event times and additional uniform censoring is induced. In order to generate subject-specific observation times `probmiss` percent of the generated longitudinal information are discarded randomly. The generated data contains for every subject in `id` the longitudinal marker `y`, observed at a grid of measurement time points (`obstime`), and the survival time `survtime`, generated with an according event indicator (`event`). Additionally, the baseline survival covariate $\tilde{x}_{\gamma i}$ (`x1`), the longitudinal covariate $\tilde{x}_{\mu i}$ (`x2`) and the grouping indicator $\tilde{x}_{\alpha i}$ (`x3`) are observed for every subject.

```
'data.frame': 1393 obs. of  23 variables:
 $ id      : Factor w/ 150 levels "1","2","3","4",..: 1 1 1 1 1 2 2 2 2 2 ...
 $ survtime: num  30.6 30.6 30.6 30.6 30.6 ...
 $ event   : num  1 1 1 1 1 1 1 1 1 1 ...
 $ x1      : num  2.32 2.32 2.32 2.32 2.32 ...
 $ x2      : num  2.32 2.32 2.32 2.32 2.32 ...
 $ r1      : num  -0.21 -0.21 -0.21 -0.21 -0.21 ...
 $ r2      : num  0.378 0.378 0.378 0.378 0.378 ...
 $ b1      : num  0.412 0.412 0.412 0.412 0.412 ...
 $ b2      : num  -0.278 -0.278 -0.278 -0.278 -0.278 ...
 $ b3      : num  0.48 0.48 0.48 0.48 0.48 ...
 $ b4      : num  0.0901 0.0901 0.0901 0.0901 0.0901 ...
 $ b5      : num  -0.293 -0.293 -0.293 -0.293 -0.293 ...
 $ b6      : num  0.166 0.166 0.166 0.166 0.166 ...
 $ cumhaz  : num  1.6 1.6 1.6 1.6 1.6 ...
 $ obstime : num  0 5 10 11 19 0 2 7 8 12 ...
 $ alpha   : num  1 1 1 1 1 1 1 1 1 1 ...
 $ dalpha  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ lambda  : num  0.207 0.207 0.207 0.207 0.207 ...
 $ gamma   : num  -3.96 -3.96 -3.96 -3.96 -3.96 ...
 $ mu      : num  1.48 1.47 1.45 1.45 1.43 ...
 $ dmu     : num  -0.00243 -0.00243 -0.00243 -0.00243 -0.00243 ...
 $ sigma   : num  -1.2 -1.2 -1.2 -1.2 -1.2 ...
 $ y       : num  1.78 1.15 1.8 1.48 2 ...
```

The sum-to-zero constraints which are set for identifiability in the structured additive model are applied to the generated data to ensure a direct comparability between true and estimated predictors. Hence, $\sum_{i=1}^{n} \eta_{\lambda i}(T_i) = 0$, $\sum_{i=1}^{n} \eta_{\gamma i} = c$ with c as the global intercept, and for nonlinear associations also $\mathbf{1}^\top \boldsymbol{\eta}_\alpha(\boldsymbol{y}^*) = 0$ where $\boldsymbol{y}^*$ is a fixed grid from the 2.5th to the 97.5th percentiles of the observed

values of `y`. The simulated nonlinear effects of $\eta_\lambda(t)$, η_γ and the nonlinear covariate effect $\sin(\tilde{x}_{\mu i})$ in η_μ are shown in Figure 6.1.

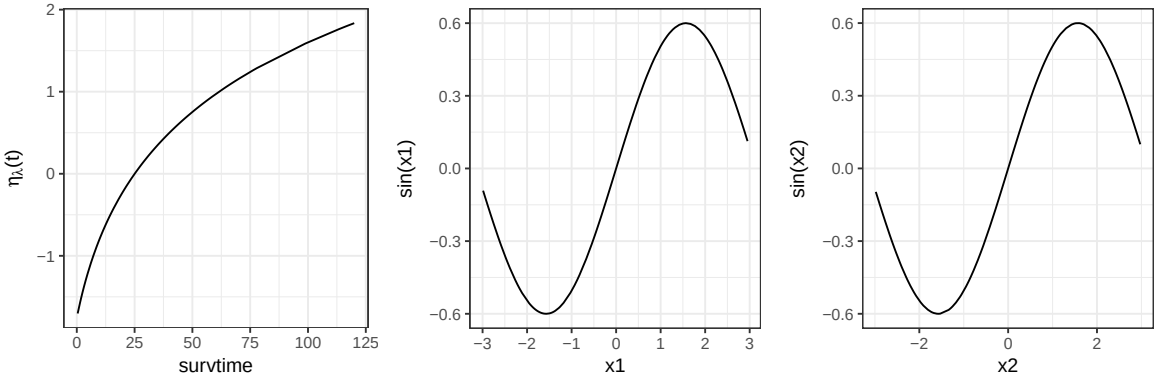

Figure 6.1: Exemplary simulated effects from `simJM()` for terms in $\eta_\lambda(t)$, η_γ and η_μ.

6.4 Practical use

In the following, we introduce the most important functionalities of the joint model fit in **bamlss** using examples of increasing complexity.

6.4.1 Basic functionalities (Example: standard joint model)

We start the illustration by fitting a simple joint model using random slopes and random intercepts in the longitudinal model and a time-constant linear association between current marker values and the time-to-event. Data for 150 subjects is generated accordingly as shown in the previous subsection. In order to fit the joint model according to the structured additive setup as shown in eq. (6.1) and eq. (6.2) all predictors are given to **bamlss** in a list of formulas specified using the utilities of `s()` for smooth terms in `mgcv`.

```
f <- list(Surv2(survtime, event, obs = y) ~ s(survtime, bs = "ps"),
          gamma ~ s(x1, bs = "ps"),
          mu ~ obstime + s(id, bs = "re") +
               s(id, obstime, bs = "re") + s(x2, bs = "ps"),
          sigma ~ 1,
          alpha ~ 1,
          dalpha ~ -1)
```

The time-varying survival predictor η_λ is specified using a modified survival formula interface. In this model fit in the predictor η_λ only the baseline hazard is specified using a P-spline over the survival times with the `mgcv` P-spline smoothing basis `bs = "ps"`. The baseline covariate `x1` is also formulated

as a smooth spline in the predictor η_γ. Furthermore, the longitudinal marker η_μ is modeled using a global intercept, a fixed slope for time, a random intercept `s(id, bs = "re")`, a random slope `s(id, obstime, bs = "re")` as well as a smooth nonlinear effect of the covariate `x2`. The variance of the longitudinal marker is modeled as constant over all subjects and so is the association η_α. The predictor `dalpha` represents a predictor relating the slope of the modeled marker to the log-hazard. As this predictor is currently under construction, it is specified as `-1` in order to be omitted from the estimation. If the number of basis functions is left unspecified in a smooth term `s()` it defaults to `k = 10` which results, due to the sum-to-zero constraints, in 9 spline basis functions and coefficients to be estimated.

As explained in the previous section both posterior mode and posterior mean estimates can be obtained. In practice, posterior mode estimates are used as starting values for the posterior mean, which is computationally more demanding but allows for correct inference. The sequence of posterior mode and mean estimation is started by specifying

```
b <- bamlss(f, data = d, family = "jm",
            timevar = "obstime", idvar = "id")
```

where the list of formulas `f`, the data `d`, the joint model family `"jm"`, the names of the longitudinal time variable as well as the subject identifier variable are given to the wrapper function. Note that the subject identifier must be a factor. This function calls internally the function `jm.transform()` to set up all design matrices for the joint model and the mode estimation function `jm.mode()` using the resulting estimates as starting values for the subsequent mean estimation in `jm.mcmc()` which returns the posterior mean estimates. In order to obtain also posterior mode estimates this sequence of model estimations can also be called manually by first estimating the posterior mode.

```
b_mode <- bamlss(f, data = d, family = "jm",
                 timevar = "obstime", idvar = "id",
                 sampler = MVNORM)
```

By specifying `sampler = MVNORM` samples are drawn to get approximate credibility intervals assuming a normal posterior distribution for all coefficients. Note however, that these estimates tend to underestimate the variability. If only point estimates are of interest, the sampling can also be switched off with `sampler = FALSE`. The posterior mode estimates can then serve as starting values for the posterior mean estimation.

```
set.seed(55)
b_mean <- bamlss(f, data = d, family = "jm",
                 timevar = "obstime", idvar = "id",
                 optimizer = FALSE, start = parameters(b_mode),
                 n.iter = 12000, burnin = 2000, thin = 5)
```

By switching off the optimizer, no posterior mode estimation is started and instead the parameters supplied are used as starting values for the MCMC. In this example 12000 sampling iterations are

specified in `n.iter` of which the first 2000 are discarded as burnin (`burnin`) and every 5th resulting sample is kept after thinning (`thin = 5`). We run the posterior mode and the posterior mean estimation and have a look at the results. Calling `summary(b_mean)` allows inspecting effect estimates for the parametric terms and the smoothing parameters. For illustration, we show the summary output of η_γ.

```
---
Formula gamma:
---
gamma ~ s(x1, bs = "ps")
-
Parametric coefficients:
              Mean   2.5%    50% 97.5%
(Intercept) -4.678 -5.106 -4.672 -4.27
-
Smooth terms:
                  Mean      2.5%       50% 97.5%
s(x1).tau21 0.0388032 0.0002914 0.0174951 0.211
s(x1).edf   3.2277031 1.3206949 3.2481793 5.198
s(x1).alpha 0.8540236 0.1004269 0.9733803 1.000
```

The predictor `gamma` is modeled using an intercept and a smooth effect of `x1`. Note that this is the joint intercept for η_γ and η_λ (and in the case of a nonlinear modeling of the association also of the predictor $\eta_\alpha(\eta_\mu)$) and hence should not be interpreted from a substantial point of view. For the smooth term we see the posterior mean as well as quantiles of the sampled posterior distribution for the variance parameter τ^2_{km}, the estimated degrees of freedom, computed as $\text{trace}(\mathbf{H}^*(\boldsymbol{\beta}_{km})\mathbf{H}(\boldsymbol{\beta}_{km})^{-1})$ with $\mathbf{H}^*(\boldsymbol{\beta}_{km})$ as the Hessian of the log-likelihood, as well as the acceptance rate (`.alpha`) of this term.

As a second example we assess the summary output for the parametric terms in η_μ.

```
---
Formula mu:
---
mu ~ obstime + s(id, bs = "re") + s(id, obstime, bs = "re") +
    s(x2, bs = "ps")
-
Parametric coefficients:
                 Mean      2.5%       50%  97.5%
(Intercept)  1.234798  1.180911  1.234339  1.291
obstime     -0.009679 -0.011554 -0.009718 -0.007
```

The global intercept of the modeled marker is 1.23 with a fixed slope of -0.01 which matches the simulated true intercept of 1.25 and the true fixed slope of -0.01 very well. Additionally, the DIC for model comparison can be obtained from the summary.

```
Sampler summary:
-
DIC = 1821.97 logLik = -819.6845 logPost = -810.2169
pd = 182.6014
---
```

Regarding the nonlinear terms in the different predictors the inbuilt plot-function can give a first overview of the estimated effects. We call `plot(b_mean)` to obtain the effect estimates for all smooth terms from this model fit.

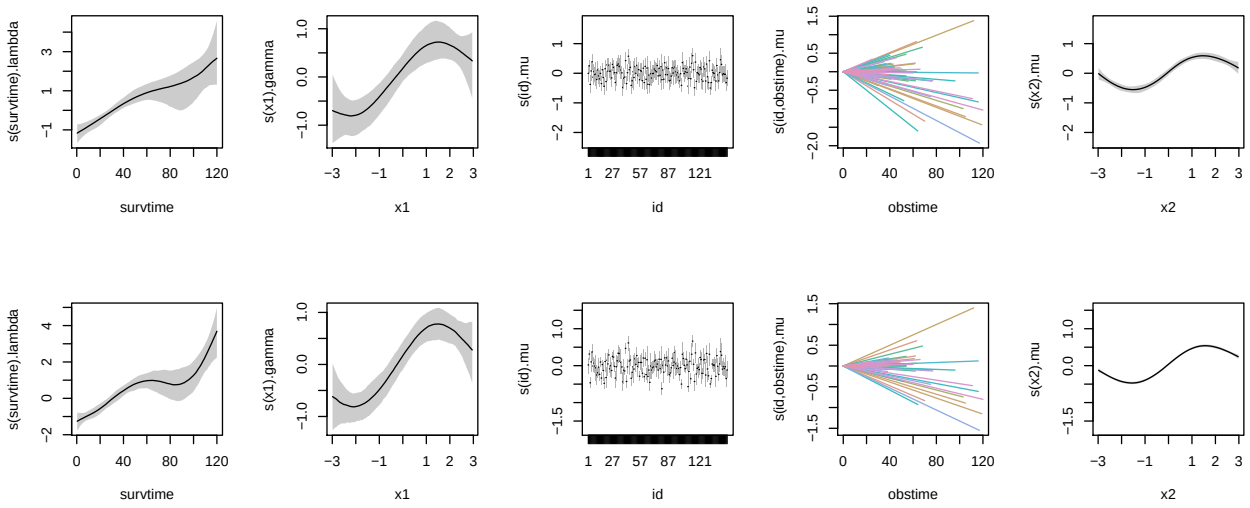

Figure 6.2: Point estimates and credibility bands for all predictors using smooth terms from 150 simulated subjects from posterior mean (upper panel) and posterior mode (lower panel).

As the upper panel of Figure 6.2 shows, the nonlinear effects for the baseline hazard, the survival covariates and the longitudinal covariates are well captured in comparison with the true simulated effects in Figure 6.1. Furthermore, the random intercept and slope estimations of the subjects are displayed. Note that the credibility bands of the nonlinear effects in the survival submodel are much wider compared to the effects in the longitudinal submodel, as in the survival part only n response values are given compared to N response values in the longitudinal submodel. We can further compare these effect estimates to the posterior mode estimation (cf. lower panel of Figure 6.2) and see that the estimates are highly similar, but have more narrow credibility bands, for example for `s(x1)` and `s(x2)`.

6.4.2 Model diagnostics (Example: flexible additive joint model)

The major motivation for the development of the additive joint model was achieving more flexibility in modeling subject-specific nonlinear trajectories using functional random intercepts and a time-varying association between the longitudinal marker and the time-to-event as developed in Chapter 3 and applied in Chapter 4. The specification of such a model is presented in the following and important functionalities for the model assessment are shown. For this example we fit the biomedical data of the previous chapter, the PBC data that are included in the R package **JMbayes** (Rizopoulos, 2016b). We first load the data and inspect the trajectories of the longitudinal marker $\log(Bilirubin)$ (cf. Figure 6.3) suggesting a nonlinear modeling of the trajectories).

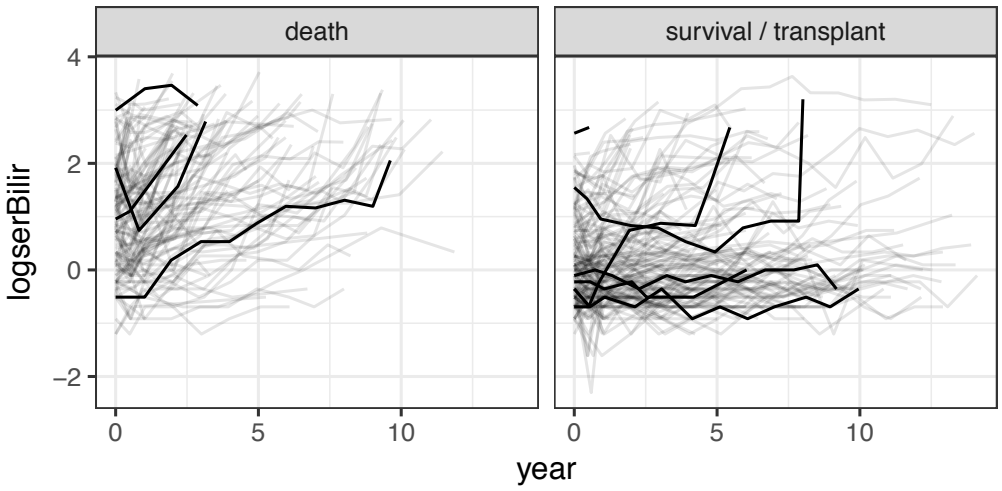

Figure 6.3: Trajectories of the $\log(Bilirubin)$ per status.

We fit the respective model in **bamlss** using as baseline covariates in the survival the drug treatment (D-penicillamine vs. placebo), the age in years at baseline and hepatomegaly, an enlarged liver, at the baseline measurement. We allow the association and the baseline hazard to be nonlinearly time-varying using P-splines with 9 basis functions after implementation of the sum-to-zero constraint and model the functional random intercepts with 7 resulting basis functions per subject.

```
library(JMbayes)
pbc2$logserBilir <- log(pbc2$serBilir)
pbc2$base_hepatomegaly <- pbc2.id$hepatomegaly[pbc2$id]
nsub <- nrow(pbc2.id)
long_df <- 8

f <- list(Surv2(years, status2, obs = logserBilir) ~ s(years, k = 10, bs = "ps"),
          gamma ~ drug + age + base_hepatomegaly,
          mu ~ ti(id, bs = "re") + ti(year, bs = "ps", k = long_df) +
```

```
              ti(id, year, bs=c("re", "ps"), k=c(nsub, long_df)),
          sigma ~ 1,
          alpha ~ s(years, k=10, bs="ps"),
          dalpha ~ -1)
b_mode <- bamlss(f, data = pbc2, family = "jm", timevar = "year", idvar = "id",
                 sampler = MVNORM, maxit = 200)
b_mean <- bamlss(f, data = pbc2, family = "jm", timevar = "year", idvar = "id",
                  sampler = jm.mcmc, start = parameters(b_mode), optimizer = FALSE,
                  n.iter = 4000, burnin = 2000, thin = 5, cores = 10)
```

For the specification of functional random intercepts we use the tensor product smooth setup of **mgcv** to induce different amounts of penalization across time and per subject. As suggested by **mgcv** we supply "main" and "interaction" effects separately using `ti()`. In this model fit we make use of the multi-core setup of bamlss for Linux machines, allowing us to start multiple chains on different cores to reduce computation time. Note that the specification of the number of iterations, the burnin and the thinning applies to each chain, for example in this model $(4000 - 2000)/5 = 400$ samples are obtained from every chain resulting in 4000 samples of the posterior. We have a look at the summary of the effects of the baseline survival covariates, which are similar to the effect estimates in the previous chapter and inspect convergence and mixing of the sampler using traceplots. For the sake of space only the first 4 coefficients of $\boldsymbol{\beta}_\alpha$ are presented (see Figure 6.4).

```
# Formula gamma:
# ---
# gamma ~ drug + age + base_hepatomegaly
# -
# Parametric coefficients:
#                            Mean     2.5%      50%  97.5%
# (Intercept)           -7.94350 -9.13449 -7.92430 -6.839
# drugD-penicil         -0.00748 -0.37032 -0.01105  0.350
# age                    0.06172  0.04462  0.06154  0.079
# base_hepatomegalyYes   0.47743  0.07960  0.47386  0.900

plot(b_mean, model = "alpha", which = "samples")
```

This plot method displays the traceplot for each estimated coefficient as well as the autocorrelation function.

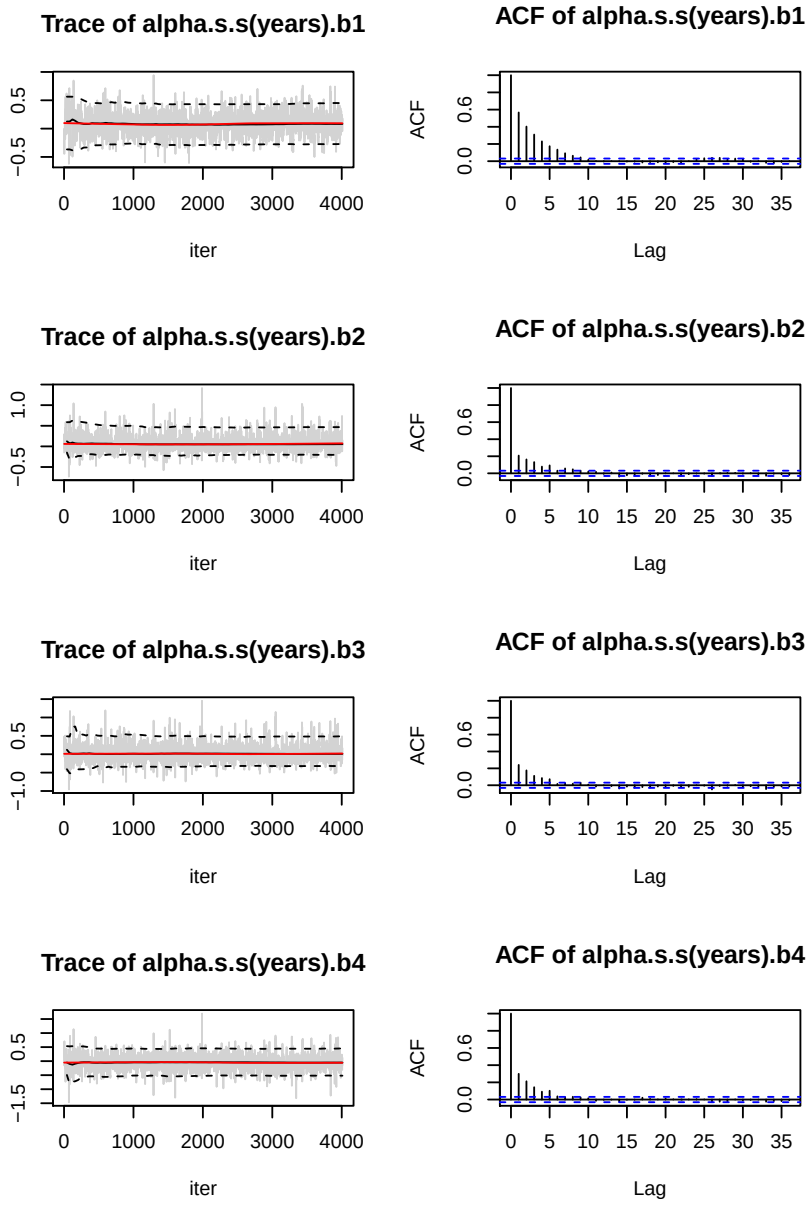

Figure 6.4: Traceplot and autocorrelation function for $\beta_{\alpha 1}, \ldots, \beta_{\alpha 4}$.

For a more detailed view on the estimated model we can make use of the `predict` function to obtain a variety of predictions from a fitted joint model object, such as

- mean estimates of the fitted predictors of the model (default)
- mean estimates of the predictors fitted on a subset of the data by specifying `newdata`

- quantiles of interest or other summaries from the posterior, e.g., credibility bands by giving the respective summary function to `FUN`
- the cumulative hazard up to the survival time or for a time window after the last longitudinal measurement by specifying `type = "cumhaz"`
- survival probabilities conditional on survival up to the last longitudinal measurement by specifying `type = "probabilities"`

We have a closer look at two of these possible specifications in the following. If we want to get a detailed view on the estimated association over the whole time range we can do so by predicting values and plotting them.

```
ND <- pbc2.id[order(pbc2.id[["years"]]),]
pred <- predict(b_mean, model="alpha", newdata=ND, FUN=c95)
# c95 obtains the 2.5th percentile, the mean and the 97.5th percentile
#   from the respective posterior samples
pdata <- data.frame(time = ND[,"years"],
                    status = ND[,"status2"],
                    effect = pred[,2],
                    effect_l = pred[,1],
                    effect_h = pred[,3])

palpha <-  ggplot(data=pdata) +
  theme_bw() +
  geom_line(aes(x = time, y = effect), size = 1.5) +
  geom_ribbon(aes(x = time, ymin = effect_l, ymax = effect_h), alpha = 0.2) +
  geom_hline(yintercept = 0, linetype = 2) +
  labs(y = expression(eta[alpha]*"(t)")) +
  scale_x_continuous("Time (years)", limits=c(0, 15),
  breaks=seq(0, 15, 2), labels=seq(0, 15, 2))
```

We see in Figure 6.5 that there is a positive association between the log-transformed biomarker Bilirubin and the hazard which is slightly decreasing over time. We can further obtain conditional survival predictions, see eq. (6.4), for the subjects basing on this model fit. For example, the conditional survival probability of subject 20 two years after the last measurement can be obtained with

```
predict(b_mean, type = "probabilities", dt = 2, FUN = c95, id = 25)
jm.survplot(bb.pbc, id = 25)
```

The conditional survival probabilities as well as the predicted longitudinal trajectories for a range of time points after the last longitudinal observation can be displayed using `jm.survplot()`, as shown in Figure 6.6.

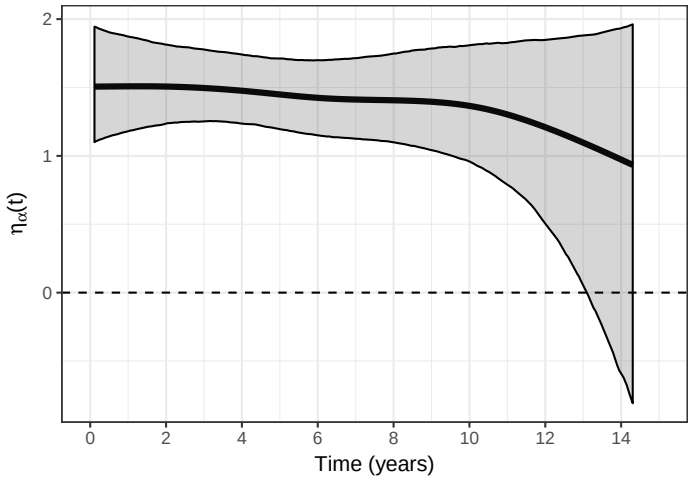

Figure 6.5: Posterior mean estimation and credibility bands of $\eta_\alpha(t)$ from the PBC fit.

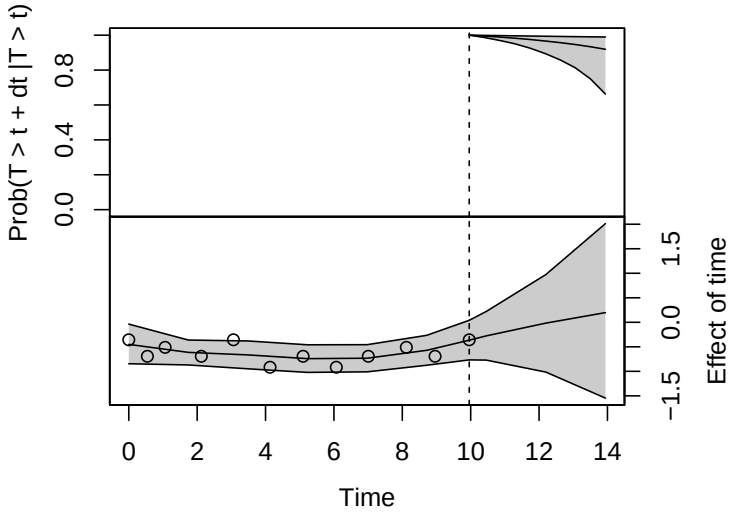

Figure 6.6: Observed and fitted longitudinal trajectory as well as predicted longitudinal trajectory and conditional survival probability for one subject.

6.4.3 Important structures (Example: nonlinear associations in the joint model)

In a last example important structures of **bamlss** joint model fits and the setup of nonlinear models are illustrated. We generate data for a joint model with functional random intercepts, a nonlinear association between marker and the log-hazard and a parametric baseline covariate by the respective call to **simJM()**. The following code illustrates the data generation for $n = 300$ and the model fit of this nonlinear association $\eta_\alpha(\eta_\mu)$, which is modeled internally as a P-spline.

```
set.seed(123)
d <- simJM(nsub=300, long_setting="functional", alpha_setting="nonlinear",
           interaction=TRUE,full=FALSE)

long_df <- 7
f_start <- y ~ ti(id, bs="re") + ti(obstime, bs="ps", k=long_df) +
               ti(id, obstime, bs=c("re", "ps"), k=c(nlevels(d$id), long_df)) +
               s(x2, bs="ps")
b_start <- bamlss(f_start, data=d, sampler=FALSE)
mu <- predict(b_start)$mu
```

In this case we fit a separate longitudinal model before starting the joint model estimation. The results from the longitudinal model are used as starting parameters for the posterior mode estimation leading to a faster convergence. More importantly, however, they provide a good starting estimation for the predictor $\eta_\mu(t)$ to base the placement of the knots for the spline representation of $\eta_\alpha(\eta_\mu(t))$ on. These knots are set up initially and have to remain fixed in the course of the model estimation. If no starting values for `start_mu` are supplied, knots are based on the observed longitudinal marker values.

```
f <- list(
  Surv2(survtime, event, obs = y) ~ s(survtime, bs="ps"),
  gamma ~ x1,
  mu ~ ti(id, bs="re") + ti(obstime, bs="ps", k=long_df) +
       ti(id, obstime, bs=c("re", "ps"), k=c(nlevels(d$id), long_df)) +
       s(x2, bs="ps"),
  sigma ~ 1,
  alpha ~ 1,
  dalpha ~ -1
)

b <- bamlss(f, data=d, family="jm", timevar="obstime", idvar="id",
            interaction=TRUE, start_mu=mu,
            n.iter=12000, burnin=2000, thin=5)
```

Note that for the nonlinear estimation of the model the user supplies only the setup for $\boldsymbol{g}_2(\tilde{\boldsymbol{x}}_{\alpha i}, t)$. By specifying `interaction=TRUE` in the call the spline representation $\boldsymbol{g_1}(\eta_{\mu i}(t))$ and the respective row-tensor product from eq. (6.3) is set up. Whereas for a simple nonlinear effect $\boldsymbol{g}_2(\tilde{\boldsymbol{x}}_{\alpha i}, t)$ is reduced to an intercept, as in the code above, a covariate dependent nonlinear association can be modeled by specifying the respective covariate in the predictor, e.g. `alpha ~ x3` for the group-specific nonlinear association. Currently only simple nonlinear, as shown here, and group-specific nonlinear interactions are implemented in the package.

Two structures of the model fit are especially interesting for diving further into the setup of the model. First, samples from the MCMC estimation are kept in the fitted **bamlss** object and can be used for further inference. If the combined predictor of $\boldsymbol{\eta}_\gamma + \boldsymbol{\eta}_\lambda$ is needed, as for example for the comparison with **JMbayes** in Chapter 3, this can be obtained easily from the respective samples. This can be either done using the `predict` function which can return all samples

```
mcmc.lambda <- predict(b_mean, model="lambda", FUN=function(x) {x})
mcmc.gamma <- predict(b_mean, model="gamma", FUN=function(x) {x})
mcmc.gamma_lambda <- mcmc.gamma + mcmc.lambda
gamma_lambda <- apply(mcmc.gamma_lambda, 1, c95)
```

or using the list of all posterior mean samples in `b_mean$samples` from which samples of single parameters can be obtained by call to the respective column name.

Second, the setup of all smooth terms in the model can be obtained from the model fit if needed, for example, to visualize the spline or obtain the knots. The list `b_mean$x` contains the smooth setup from **mgcv** of every predictor k and for every fitted term m. For example the setup of the smooth effect of $f_{\mu 4}(\tilde{x}_{\mu i})$ can be found in

```
b_mean$x$mu$smooth.construct$‘s(x2)‘
```

The presented model setup, utility functions and further information provide the starting point to fit a variety of joint models within the package **bamlss**. We refer to the respective help pages and the code supplements of the published manuscripts for further details.

6.5 Future plans

The implementation of the structured additive joint model in **bamlss** allows users to fit the presented flexible joint model within R. While the main functionalities are set up and tested, the implementation can be extended in various ways. Firstly, as also mentioned in the previous chapter, more nonlinear association structures can be included such as a nonlinear time-varying association between marker and time-to-event process $f_\alpha(\eta_{\mu i}(t), t)$ or a nonlinear covariate-dependent association $f_\alpha(\eta_{\mu i}(t), \tilde{\boldsymbol{x}}_{\alpha i})$ beyond the implemented dummy covariate. Basing on previous experience, however, this extension is only sensible with a high number of subjects observed longitudinally on a dense grid. Second, in certain situations it could be of interest to associate not only the current value of the modeled

marker $\eta_{\mu i}(t)$ but also the current slope of this trajectory $\eta'_{\mu i}(t)$ with the hazard. By adding a further predictor in the hazard as

$$h_i(t) = \exp\left\{\eta_{\lambda i}(t) + \eta_{\gamma i} + \eta_{\alpha i}\left(\eta_{\mu i}(t), t\right) + \eta_{\alpha' i}\left(\eta'_{\mu i}(t), t\right)\right\}$$

all flexibility in the association with the current value such as nonlinear, time-varying and covariate-dependent associations, could also be extended to the association with the current slope. This additional association structure, which makes use of the straightforward analytical derivative of B-splines (see Appendix A.2), is already under development but still needs thorough testing.

Besides these additional structures and associations further utility functions regarding model diagnostics should be implemented. For assessing the accuracy of survival models calibration measures, indicating how well the model can predict observed event rates, as well as discrimination measures, assessing the model's ability to differentiate between subjects with high and low risk, are crucial. There are different adaptations of calibration (Henderson et al., 2002) and discrimination (Rizopoulos, 2012) measures in the context of joint models which, incorporated into the joint model family in **bamlss** could enhance the assessment of estimated flexible additive joint models.

The usability of the present model and all of its extensions is crucially dependent on the computational ease and speed. Although we have made large gains in computation speed by transferring the numerical integration into C and by exploiting the block-structure of the design matrix and Hessian of the functional random intercepts, we are continuously working on speeding up the computations further.

Chapter 7

Discussion

In this concluding chapter a main summary with a special focus on the contributions to T1D research and, in particular, to the field of joint modeling is given. Modeling aspects discussed in the outlook of the previous Chapters 3, 4 and 5 are summarized and extended in order to achieve the main conclusions for the whole research project. An outlook on potential future steps concludes this work.

7.1 Concluding summary

Joint models of longitudinal and survival data are an important tool for the analysis of longitudinal biomarkers and their association with an event of interest. They increasingly find entry into applied analyses of longitudinal cohort data (Sudell et al., 2016), in which longitudinal markers are usually observed at subject-specific discrete measurement times, subject to measurement error and informatively censored at event occurrence. By the joint estimation of a longitudinal and survival submodel within one likelihood, joint models allow the unbiased estimation of this data. Aim of this work is the extension of established joint model approaches to provide more flexibility, as needed in the research on disease progression of T1D.

In order to achieve this aim, a general Bayesian joint model framework is developed in Chapter 3, which allows the flexible incorporation of smooth nonlinear, time-varying and random effects terms into shared random effects joint models beyond previous approaches. The developed model is applied to two different T1D risk cohorts in Chapters 3 and 4, allowing deeper insights into the time-varying nature of the association between disease-specific biomarkers and the onset of clinical symptoms of T1D.

Existing joint model approaches further assume linearity between the longitudinal marker and the log-hazard. This assumption has not been tested and might lead to biased results if strong nonlinearity is present. The developed framework is therefore further extended in Chapter 5 to test and relax this assumption. The ability to capture truly linear and truly nonlinear associations is shown not only in simulations but also in modeling further biomedical data.

Notably the developed framework allows for a more complex specification of the association between longitudinal markers and a time-to-event process such as nonlinear, subgroup-specific or time-

varying specifications and thereby has the potential to help to understand complex relationships between longitudinal biomarkers and disease progression. To facilitate applications of the developed framework, the model is readily available within the R package **bamlss** (Umlauf et al., 2017), and an introduction to fitting flexible additive joint models is given in Chapter 6.

7.1.1 Contributions to type 1 diabetes research

We analyzed the associations between T1D-specific autoantibody markers and the progression to T1D in a mouse model and two data sets from unique T1D birth cohorts. In line with previous results higher autoantibody levels of IAA were associated with a higher progression risk in the BABYDIAB/BABYDIET data. The association was estimated as roughly constant with a slightly higher association at earlier time points but also with wide credibility bands around the effect estimate. This uncertainty is likely due to the complexity of the specified nonlinear time-varying association, given the limited sample size of this data set.

By modeling the larger, densely measured data set from the large multinational TEDDY cohort the detailed analysis of associations was feasible, such as potentially time-varying or subgroup-specific association structures. As a result the association between autoantibody markers and the time to T1D progression was estimated as declining over time for IAA and GADA and as time-constant positive for IA2A. These more detailed results refine previous findings from TEDDY data and other cohorts. For the first time, the time-varying nature of the estimated associations could be quantified using our flexible joint model. The findings indicate that risk stratification based on autoantibody measurements is most informative early after seroconversion for the respective autoantibody. This trend was even more pronounced in children with first-degree relatives with T1D.

7.1.2 Contributions to the field of joint modeling

Research on joint models is a fast growing field with developments in many areas such as the analysis of multivariate longitudinal measurements, relaxing distributional assumptions and also the flexible modeling of all model parts and their associations. The focus of our contribution lies in the latter where we greatly extended the flexibility of shared random effect models by incorporating the joint model in a structured additive framework. Whereas many developments in this area use a random effects link to incorporate the marker in the hazard for computational ease, we include the full, flexibly modeled trajectory, thereby allowing for a straightforward assessment and interpretation of the association also under the specification of complex random effects structures.

The special focus in the previous chapters lies on extending given models to enable a flexible modeling of subject-specific trajectories and to allow the associations between marker and hazard to be potentially covariate-dependent, time-varying and nonlinear. By modeling the longitudinal marker to closely describe the subject-specific nonlinear trajectories using functional random intercepts, the association between the marker and the hazard can be estimated more precisely. This is of special importance if the exploration of potentially complex association structures is of interest.
Although many different parameterizations for the association between the marker and the event exist,

modeled associations were previously restricted to be time-constant in the log-hazard. By resolving this restriction and allowing associations to differ covariate-dependently, further insights into disease mechanisms are possible as shown in Chapter 4. Simultaneously with the development of the flexible additive joint model also Andrinopoulou et al. (2016) extended an existing joint model to allow for a time-varying association in the modeling of the heart valve function after a cardio-thoracic surgery. While also making use of Bayesian P-splines their model allows to specify a time-varying association of both a current marker and its current slope and to derive improved dynamic predictions from the model. However, it offers less flexibility in the specification of individual trajectories and further nonlinear covariate effects than the presented additive joint model. Still, this similar extension and the biomedical research question motivating this development stresses the importance of including time-varying associations in joint models.
Furthermore, most models assume a linear association between marker and log-hazard. Often parametric transformations are used to achieve normally distributed markers, which are at the same time (hopefully) linear in the log-hazard. Our model extension allows, for the first time, to check this assumption in joint models and to relax it in modeling the influence of longitudinal markers.

An important prerequisite for joint models to find broader entry into applied research in various biomedical fields is the availability of ready-to-use software implementations. Therefore, the developed framework is implemented in the R package **bamlss** to facilitate the usage for researchers interested in flexibly modeling this kind of data. The package thereby enlarges the options of estimating flexible joint models beyond previous implementations.

Although these extensions are the focus of the present work, far more modeling options are possible in this framework and have already been implemented in **bamlss**. Such further effects are, for example, time-varying covariates in the survival part in predictor η_λ or spatial effects in the time-constant survival part in predictor η_γ as in Martins et al. (2016). Also other error structures, such as autocorrelated errors or group-specific error structures are already available within our implementation.

7.2 Outlook

Areas of further development were already mentioned in the previous chapters and are therefore summarized and extended in the following. We first consider useful extensions of the software before discussing further model developments.

As discussed in the previous chapter the current implementation in the software could be extended by adding further utility functions and model diagnostics such as residuals for the survival part of the model as well as further statistics to assess the calibration and discrimination of a fitted model. Furthermore, only a subset of the presented nonlinear association structures is currently implemented and could be extended further. Finally, it is of great interest to speed up computations of the current implementation.

Besides these straightforward implementation options, the developed general framework allows for many extensions. Before presenting potential further developments, however, a note of caution is

necessary. In applying the flexible additive joint model to different biomedical data sets and simulated data we observed that a certain amount of information is necessary to achieve a stable and precise estimation. For modeling flexible linear associations the number of subjects and events times is the limiting factor, e.g. in Chapter 3 modeling $n = 127$ subjects with 69 event times was the lower limit to achieve tolerable credibility intervals. For modeling flexible nonlinear associations also the number and spacing of longitudinal observations is important to achieve stable estimations, e.g. in Chapter 5 for group-specific nonlinear associations a median of 6 observations per subject was a lower limit. Basing on this experience the implementation within R to achieve (a) a stable estimation procedure, (b) basing on a realistic amount of data, i.e. 150 to 500 subjects and (c) with tolerable computation times are major challenges in all further developments and should be kept in mind in the following.

One area of possible extensions is the inclusion of further association structures such as the influence of the current slope of a marker which could also be nonlinear and time-varying. In addition, elaborate cumulative effects, similar to historical functional effects as discussed in Chapter 3, could be included within the given framework. Similar to the incorporation of functional random intercepts in the longitudinal trajectories this could be another approach originating from functional data analysis to be incorporated in the joint modeling framework. It has to be noted, however, that due to the informatively censored observations in the longitudinal model which are observed at subject-specific time points and not on a fixed grid, functional data approaches cannot directly be incorporated in the joint model framework but have to be modified. Furthermore, as historical functional effects would model the association even more flexibly than previous extensions, a high amount of data, observed on a dense grid, would be necessary to achieve stable estimations.
Further possible extensions concern the distributional assumptions for the longitudinal response. For example, the assumption of normal errors can be too strict in the presence of heavy-tail distributed errors and could be relaxed to t-distributed errors (Huang et al., 2010; Baghfalaki et al., 2014). Besides, mixtures of distributions could be allowed (Rizopoulos, 2016b) to account for censoring of longitudinal measurements at certain thresholds. The longitudinal model part could further be allowed to model binary or count data by making full use of the generalized additive regression framework.
Currently, many approaches are being developed to include multiple longitudinal marker trajectories in the joint model (see Hickey et al. (2016) for an extensive overview) and also the flexible additive joint model could be extended in this direction. Challenges of multivariate longitudinal approaches are, however, the appropriate modeling of the correlation structure between multiple markers and their joint or separate association with the hazard.

Dynamic prediction, which is the risk prediction for a subject that can be dynamically updated when additional longitudinal information is given, is an area of great interest in current joint model developments (see, for example, Andrinopoulou et al., 2016; Ferrer et al., 2017; Barrett and Su, 2017). Developing dynamic predictions in our flexible additive joint model could improve the precision in the prediction compared to existing approaches, as, for example, Andrinopoulou et al. (2016) could show an improved prediction based on the inclusion of a time-varying association and Barrett and

Su (2017), using a discretized time-scale and a random effects association, showed improvements by including highly flexible individual trajectories in the predictions.

Model selection within joint models is another potential direction of future research. Many developed joint models, as also our model, use model based statistics to compare models such as DIC, conditional predictive ordinate (CPO) or Bayes Factors (see, for example, Brown, 2009; Tang and Tang, 2015). Recently, first approaches for more elaborate model selection techniques in the joint model framework have been developed. For example, Waldmann et al. (2017) use gradient-based boosting to select covariates in the survival and the longitudinal model, so far restricting the longitudinal trajectories to random intercept and slope models and a random effects association. Different Bayesian shrinkage priors are tested by Andrinopoulou and Rizopoulos (2016) to determine appropriate association structures and a Bayesian LASSO approach is used by Tang et al. (2017) to select variables in the longitudinal and survival submodel as well as their association. Both approaches, boosting and shrinkage, could be interesting developments for the flexible additive joint model to enable an automatic variable selection in the different predictors.

Finally, although joint models were developed for and are mainly used in biomedical research to associate a longitudinal biomarker and the time-to-event, the application of joint models is not restricted to this research field. Also in another applied areas of survival analysis, such as engineering, joint models could be used. For example, joint models could help to improve explaining mechanical failures based on internal longitudinal covariates such as sensor information. This generalizes to statistical analyses in all applied fields in which longitudinal, informatively censored information is potentially associated with an event of interest.

In conclusion the developed model framework is a valuable contribution to more flexible joint models that can be extended in various ways to meet the further demands of applied researchers.

Appendices

Appendix A

Technical details

A.1 Technical details of Chapter 3

We derive score vectors and Hessians for the regression coefficients of every predictor. We introduce some further notation to formulate these derivatives. For the time-varying predictors of the survival part $k \in \{\lambda, \alpha, \mu\}$ the design matrix $\mathbf{X}_k(\mathbf{T})$ denotes the $n \times p_k$ matrix of evaluations at the vector of survival times $\mathbf{T}$. For the time-varying predictors of the longitudinal part $k \in \{\mu, \sigma\}$ the $N \times p_k$ design matrix $\mathbf{X}_k(\mathbf{t})$ contains the evaluations at all observed subject-specific time points $\mathbf{t}$. The column vector $\boldsymbol{x}_{ki}$ is the i-th row of the respective design matrix $\mathbf{X}_k$ for $k \in \{\gamma, \lambda, \alpha, \mu, \sigma\}$ and $\boldsymbol{\beta}_k = [\beta_{k1}, \cdots, \beta_{kM_k}]^\top$. Let ℓ denote the log-likelihood, i.e. the sum of the contributions of the longitudinal and survival submodels defined in eq. (3.6) and eq. (3.7). In more detail, the full log-likelihood is

$$\begin{aligned}\ell\left[\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{\delta}, \boldsymbol{y}\right] =& \boldsymbol{\delta}^\top \left[\mathbf{X}_\lambda(\mathbf{T})\boldsymbol{\beta}_\lambda + \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma + \mathbf{X}_\alpha(\mathbf{T})\boldsymbol{\beta}_\alpha \cdot \mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu\right] \\ & - \sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \int_0^{T_i} \exp\left[\boldsymbol{x}_{\lambda i}^\top(u)\,\boldsymbol{\beta}_\lambda + \boldsymbol{x}_{\alpha i}^\top(u)\,\boldsymbol{\beta}_\alpha \left(\boldsymbol{x}_{\mu i}^\top(u)\,\boldsymbol{\beta}_\mu\right)\right] du \\ & - \frac{N}{2}\log(2\pi) - \mathbf{1}_N^\top \mathbf{X}_\sigma(\mathbf{t})\,\boldsymbol{\beta}_\sigma - \frac{1}{2}(\boldsymbol{y} - \mathbf{X}_\mu(\mathbf{t})\,\boldsymbol{\beta}_\mu)^\top \boldsymbol{R}^{-1}(\boldsymbol{y} - \mathbf{X}_\mu(\mathbf{t})\,\boldsymbol{\beta}_\mu)\end{aligned}$$

resulting in the log-posterior

$$\log p(\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{\delta}, \boldsymbol{y}) \propto \ell\left[\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{\delta}, \boldsymbol{y}\right] + \sum_{k \in \{\lambda,\gamma,\alpha,\mu,\sigma\}} \sum_{m=1}^{M_k} \left[\log p(\boldsymbol{\beta}_{km}|\boldsymbol{\tau}_{km}^2) + \log p(\boldsymbol{\tau}_{km}^2)\right].$$

The full score vectors $\boldsymbol{s}(\boldsymbol{\beta}_k)$ and Hessians $\boldsymbol{H}(\boldsymbol{\beta}_k)$ of the respective predictors are computed as the sum of the score $\boldsymbol{s}^*(\boldsymbol{\beta}_k)$ and Hessians $\boldsymbol{H}^*(\boldsymbol{\beta}_k)$ based on the log-likelihood function and the score and Hessian of the respective log-prior densities. For the multivariate normal prior as specified in 3.2.2 these are $-\frac{1}{\tau_{km}^2}\mathbf{K}_{km}\boldsymbol{\beta}_{km}$ and $-\frac{1}{\tau_{km}^2}\mathbf{K}_{km}$, respectively.

Score Vectors

$$
\begin{aligned}
\boldsymbol{s}^*(\boldsymbol{\beta}_\mu) = \frac{\partial \ell}{\partial \boldsymbol{\beta}_\mu} =& \mathbf{X}_\mu(\mathbf{t})^\top \boldsymbol{R}^{-1}(\boldsymbol{y} - \mathbf{X}_\mu(\mathbf{t})\boldsymbol{\beta}_\mu) + \mathbf{X}_\mu^\top(\mathbf{T})\,\mathrm{diag}(\boldsymbol{\delta})\,[\mathbf{X}_\alpha(\mathbf{T})\boldsymbol{\beta}_\alpha] \\
& - \sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \int_0^{T_i} \omega_i(u)\ \boldsymbol{x}_{\alpha i}^\top(u)\,\boldsymbol{\beta}_\alpha \boldsymbol{x}_{\mu i}(u)\,du \\
\boldsymbol{s}^*(\boldsymbol{\beta}_\gamma) = \frac{\partial \ell}{\partial \boldsymbol{\beta}_\gamma} =& \boldsymbol{\delta}^\top \mathbf{X}_\gamma - \sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \boldsymbol{x}_{\gamma i} \int_0^{T_i} \omega_i(u)\ du \\
\boldsymbol{s}^*(\boldsymbol{\beta}_\alpha) = \frac{\partial \ell}{\partial \boldsymbol{\beta}_\alpha} =& \mathbf{X}_\alpha^\top(\mathbf{T})\,\mathrm{diag}(\boldsymbol{\delta})\,[\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu] - \sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \int_0^{T_i} \omega_i(u)\ \boldsymbol{x}_{\alpha i}(u)\left(\boldsymbol{x}_{\mu i}^\top(u)\,\boldsymbol{\beta}_\mu\right) du \\
\boldsymbol{s}^*(\boldsymbol{\beta}_\lambda) = \frac{\partial \ell}{\partial \boldsymbol{\beta}_\lambda} =& \boldsymbol{\delta}^\top \mathbf{X}_\lambda(\mathbf{T}) - \sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \int_0^{T_i} \omega_i(u)\ \boldsymbol{x}_{\lambda i}(u)\,du \\
\boldsymbol{s}^*(\boldsymbol{\beta}_\sigma) = \frac{\partial \ell}{\partial \boldsymbol{\beta}_\sigma} =& -\mathbf{X}_\sigma(\mathbf{t})^\top \mathbf{1}_N + [\mathbf{X}_\sigma(\mathbf{t}) \odot (\boldsymbol{y} - \mathbf{X}_\mu(\mathbf{t})\boldsymbol{\beta}_\mu)]^\top \boldsymbol{R}^{-1}(\boldsymbol{y} - \mathbf{X}_\mu(\mathbf{t})\boldsymbol{\beta}_\mu)
\end{aligned}
$$

Hessian

$$
\begin{aligned}
\boldsymbol{H}^*(\boldsymbol{\beta}_\mu) = \frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_\mu \partial \boldsymbol{\beta}_{\mu\top}} =& -\mathbf{X}_\mu(\mathbf{t})^\top \boldsymbol{R}^{-1} \mathbf{X}_\mu(\mathbf{t}) \\
& - \sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \int_0^{T_i} \omega_i(u)\ \left[\boldsymbol{x}_{\alpha i}^\top(u)\,\boldsymbol{\beta}_\alpha\right]^2 \boldsymbol{x}_{\mu i}(u)\,\boldsymbol{x}_{\mu i}^\top(u)\,du \\
\boldsymbol{H}^*(\boldsymbol{\beta}_\gamma) = \frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_\gamma \partial \boldsymbol{\beta}_\gamma^\top} =& -\sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \boldsymbol{x}_{\gamma i} \boldsymbol{x}_{\gamma i}^\top \int_0^{T_i} \omega_i(u)\ du \\
\boldsymbol{H}^*(\boldsymbol{\beta}_\alpha) = \frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_\alpha \partial \boldsymbol{\beta}_\alpha^\top} =& -\sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \int_0^{T_i} \omega_i(u)\ \left[\boldsymbol{x}_{\mu i}^\top(u)\,\boldsymbol{\beta}_\mu\right]^2 \boldsymbol{x}_{\alpha i}(u)\,\boldsymbol{x}_{\alpha i}^\top(u)\,du \\
\boldsymbol{H}^*(\boldsymbol{\beta}_\lambda) = \frac{\partial^2 \ell_i}{\partial \boldsymbol{\beta}_\lambda \partial \boldsymbol{\beta}_\lambda^\top} =& -\sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \int_0^{T_i} \omega_i(u)\ \boldsymbol{x}_{\lambda i}(u)\,\boldsymbol{x}_{\lambda i}^\top(u)\,du \\
\boldsymbol{H}^*(\boldsymbol{\beta}_\sigma) = \frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_\sigma \partial \boldsymbol{\beta}_\sigma^\top} =& -2\,[\mathbf{X}_\sigma(\mathbf{t}) \odot (\boldsymbol{y} - \mathbf{X}_\mu(\mathbf{t})\boldsymbol{\beta}_\mu)]^\top \boldsymbol{R}^{-1} [\mathbf{X}_\sigma(\mathbf{t}) \odot (\boldsymbol{y} - \mathbf{X}_\mu(\mathbf{t})\boldsymbol{\beta}_\mu)]
\end{aligned}
$$

where $\omega_i(u) = \exp\left[\boldsymbol{x}_{\lambda i}^\top(u)\,\boldsymbol{\beta}_\lambda + \boldsymbol{x}_{\alpha i}^\top(u)\,\boldsymbol{\beta}_\alpha \left(\boldsymbol{x}_{\mu i}^\top(u)\,\boldsymbol{\beta}_\mu\right)\right]$ and $\boldsymbol{R} = \mathrm{diag}\left(\exp\left[\mathbf{X}_\sigma(\mathbf{t})\,\boldsymbol{\beta}_\sigma\right]^2\right)$.

A.2 Technical details of Chapter 5

In the following score vectors and Hessians for the regression coefficients of every predictor are presented. For their computation the predictors are also necessary at the survival times $\mathbf{T}$. Here, $\mathbf{X}_k(\mathbf{T})$ denotes the respective $n \times p_k$ design matrix of evaluations of the time-varying predictors of the survival part $k \in \{\lambda, \alpha, \mu\}$ at time points $\mathbf{T}$. Please note that in comparison with the previously presented flexible additive joint model in Chapter 3 and in (Köhler et al., 2017b) only the score and Hessians for the predictors η_α and η_μ have changed relevantly for the nonlinear specification and are presented in the following. For all other predictors we refer to Section A.1

The full log-likelihood is

$$\begin{aligned}\ell\left[\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{\delta}, \boldsymbol{y}\right] =& \boldsymbol{\delta}^\top \left[\mathbf{X}_\lambda(\mathbf{T})\boldsymbol{\beta}_\lambda + \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma + \left[\boldsymbol{g}_1\left(\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu\right) \odot \boldsymbol{g}_2(\tilde{\mathbf{X}}_\alpha(\mathbf{T}))\right]\boldsymbol{\beta}_\alpha\right] \\ & - \sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \int_0^{T_i} \exp\left[\boldsymbol{x}_{\lambda i}^\top(u)\,\boldsymbol{\beta}_\lambda + \left[\boldsymbol{g}_1\left(\boldsymbol{x}_{\mu i}^\top(u)\,\boldsymbol{\beta}_\mu\right) \odot \boldsymbol{g}_2(\tilde{\boldsymbol{x}}_{\alpha i}^\top(u))\right]\boldsymbol{\beta}_\alpha\right] du \\ & - \frac{N}{2}\log(2\pi) - \mathbf{1}_N^\top \mathbf{X}_\sigma(\mathbf{t})\,\boldsymbol{\beta}_\sigma - \frac{1}{2}(\boldsymbol{y} - \mathbf{X}_\mu(\mathbf{t})\,\boldsymbol{\beta}_\mu)^\top \boldsymbol{R}^{-1}(\boldsymbol{y} - \mathbf{X}_\mu(\mathbf{t})\,\boldsymbol{\beta}_\mu)\end{aligned}$$

For the flexible association in eq. (5.4) the term $\boldsymbol{g}_1\left(\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu\right)$ reduces to $\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu$ for a linear association and is $\mathbf{B}\left(\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu\right)$ for a nonlinear association. Likewise, the term $\boldsymbol{g}_2(\tilde{\mathbf{X}}_\alpha(\mathbf{T}))$ reduces to $\mathbf{1_n}$ for a simple constant association, is the covariate vector or design matrix of the parametric input for covariate-dependent associations and is the evaluation of a spline basis matrix for a time-varying association. We denote this term in the following as $\mathbf{X}_{\alpha 2}$ to represent all three possible forms.
The resulting log-posterior is

$$\log p(\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{\delta}, \boldsymbol{y}) \propto \ell\left[\boldsymbol{\theta}|\mathbf{T}, \boldsymbol{\delta}, \boldsymbol{y}\right] + \sum_{k \in \{\lambda,\gamma,\alpha,\mu,\sigma\}} \sum_{m=1}^{M_k} \left[\log p(\boldsymbol{\beta}_{km}|\boldsymbol{\tau}_{km}^2) + \log p(\boldsymbol{\tau}_{km}^2)\right].$$

The scores $\boldsymbol{s}(\boldsymbol{\beta}_k)$ and Hessians $\boldsymbol{H}(\boldsymbol{\beta}_k)$ are computed as the sum of the respective derivatives of the log-likelihood and of the log-prior densities. The latter are for example $-\frac{1}{\tau_{km}^2}\mathbf{K}_{km}\boldsymbol{\beta}_{km}$ and $-\frac{1}{\tau_{km}^2}\mathbf{K}_{km}$ for the multivariate normal prior as specified in Section **??**. The score vectors $\boldsymbol{s}^*(\boldsymbol{\beta}_k)$ and Hessians $\boldsymbol{H}^*(\boldsymbol{\beta}_k)$ of the log-likelihood function are presented in the following.

Score Vectors

$$\begin{aligned}\boldsymbol{s}^*(\boldsymbol{\beta}_\mu) = \frac{\partial \ell}{\partial \boldsymbol{\beta}_\mu} =& \mathbf{X}_\mu(\mathbf{t})^\top \boldsymbol{R}^{-1}\left(\boldsymbol{y} - \mathbf{X}_\mu(\mathbf{t})\,\boldsymbol{\beta}_\mu\right) + \mathbf{X}_\mu^\top(\mathbf{T})\,\mathrm{diag}(\boldsymbol{\delta})\left[\boldsymbol{g}_1'\left(\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu\right) \odot \mathbf{X}_{\alpha 2}(\mathbf{T})\right]\boldsymbol{\beta}_\alpha \\ & - \sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \int_0^{T_i} \psi_i(u)\ \left[\boldsymbol{g}_1'\left(\boldsymbol{x}_{\mu i}^\top(u)\,\boldsymbol{\beta}_\mu\right) \odot \mathbf{x}_{\alpha 2i}^\top(u)\right]\boldsymbol{\beta}_\alpha \boldsymbol{x}_{\mu i}(u) du\end{aligned}$$

$$\begin{aligned}\boldsymbol{s}^*(\boldsymbol{\beta}_\alpha) = \frac{\partial \ell}{\partial \boldsymbol{\beta}_\alpha} =& \boldsymbol{\delta}^\top \left[\boldsymbol{g}_1\left(\mathbf{X}_\mu\left(\mathbf{T}\right)\boldsymbol{\beta}_\mu\right) \odot \mathbf{X}_{\alpha 2}(\mathbf{T})\right] \\ & - \sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \int_0^{T_i} \psi_i(u) \left[\boldsymbol{g}_1\left(\boldsymbol{x}_{\mu i}^\top(u)\,\boldsymbol{\beta}_\mu\right) \odot \mathbf{x}_{\alpha 2 i}^\top(u)\right]^\top du\end{aligned}$$

with $\psi_i(u) = \exp\left[\boldsymbol{x}_{\lambda i}^\top(u)\,\boldsymbol{\beta}_\lambda + \left[\boldsymbol{g}_1\left(\boldsymbol{x}_{\mu i}^\top(u)\,\boldsymbol{\beta}_\mu\right)^\top \odot \mathbf{x}_{\alpha 2 i}^\top(u)\right]\boldsymbol{\beta}_\alpha\right]$ and the diagonal matrix $\boldsymbol{R} = \operatorname{diag}\left(\exp\left[\mathbf{X}_\sigma\left(\mathbf{t}\right)\boldsymbol{\beta}_\sigma\right]^2\right)$. For the score vector $\boldsymbol{s}^*(\boldsymbol{\beta}_\mu)$ the derivative of $\boldsymbol{g}_1\left(\boldsymbol{x}_{\mu i}^\top(u)\boldsymbol{\beta}_\mu\right)$ with respect to $\boldsymbol{\beta}_\mu$ is needed which can be derived by chain rule

$$\frac{\partial \boldsymbol{g}_1(\boldsymbol{x}_{\mu i}^\top(u)\boldsymbol{\beta}_\mu)}{\partial \boldsymbol{\beta}_\mu} = \frac{\partial \boldsymbol{g}_1(\boldsymbol{\eta}_\mu(u))}{\partial \boldsymbol{\eta}_\mu(u)} \cdot \frac{\partial \boldsymbol{\eta}_\mu(u)}{\partial \boldsymbol{\beta}_\mu}.$$

The derivative of $\boldsymbol{g}_1\left(\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu\right)$ follows analogously. Whereas the inner derivative $\frac{\partial \boldsymbol{\eta}_\mu(u)}{\partial \boldsymbol{\beta}_\mu} = \boldsymbol{x}(u)$ is the same for both linear and nonlinear associations, the outer derivative, which we denote by $\boldsymbol{g}_1'\left(\boldsymbol{x}_{\mu i}^\top(u)\boldsymbol{\beta}_\mu\right)$, differs between the parameterizations. For linear associations it holds that $\boldsymbol{g}_1'\left(\boldsymbol{x}_{\mu i}^\top(u)\boldsymbol{\beta}_\mu\right) = 1$ and $\boldsymbol{g}_1'\left(\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu\right) = \mathbf{1}_\mathbf{n}^\top$. Nonlinear associations are implemented as penalized B-splines in **bamlss**, $\boldsymbol{g}_1'\left(\boldsymbol{x}_{\mu i}^\top(u)\boldsymbol{\beta}_\mu\right) = \mathbf{B}'(\boldsymbol{x}_{\mu i}^\top(u)\boldsymbol{\beta}_\mu)$ and $\boldsymbol{g}_1'\left(\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu\right) = \mathbf{B}'(\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu)$, which have a straightforward analytical solution for the derivative (Fahrmeir et al., 2013)

$$\frac{\partial}{\partial z}\sum_d B_d^l(z) = l\left(\frac{1}{\kappa_d - \kappa_{d-1}} B_{d-1}^{l-1}(z) - \frac{1}{\kappa_{d+1} - \kappa_{d+1-l}} B_d^{l-1}(z)\right). \tag{A.1}$$

where l denotes the degree of the spline, d is the index for the basis functions and κ denotes the knots with the interior knots $\kappa_1, \ldots, \kappa_m$ and $2l$ outer knots.

Hessian

$$\begin{aligned}\boldsymbol{H}^*(\boldsymbol{\beta}_\mu) = \frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_\mu \partial \boldsymbol{\beta}_{\mu\top}} =& -\mathbf{X}_\mu\left(\mathbf{t}\right)^\top \boldsymbol{R}^{-1} \mathbf{X}_\mu\left(\mathbf{t}\right) + \mathbf{X}_\mu^\top\left(\mathbf{T}\right) \operatorname{diag}(\boldsymbol{\delta}) \left[\boldsymbol{g}_1''\left(\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu\right) \odot \mathbf{X}_{\alpha 2}(\mathbf{T})\right] \boldsymbol{\beta}_\alpha \mathbf{X}_\mu\left(\mathbf{T}\right) \\ & - \sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \int_0^{T_i} \psi_i(u) \cdot \\ & \left[\left(\left[\boldsymbol{g}_1'\left(\boldsymbol{x}_{\mu i}^\top\left(u\right)\boldsymbol{\beta}_\mu\right) \odot \mathbf{x}_{\alpha 2 i}^\top(u)\right]\boldsymbol{\beta}_\alpha\right)^2 + \left[\boldsymbol{g}_1''\left(\boldsymbol{x}_{\mu i}^\top\left(u\right)\boldsymbol{\beta}_\mu\right) \odot \mathbf{x}_{\alpha 2 i}^\top(u)\right]\boldsymbol{\beta}_\alpha\right] \cdot \\ & \boldsymbol{x}_{\mu i}(u)\boldsymbol{x}_{\mu i}^\top(u) du\end{aligned}$$

$$\boldsymbol{H}^*(\boldsymbol{\beta}_\alpha) = \frac{\partial^2 \ell}{\partial \boldsymbol{\beta}_\alpha \partial \boldsymbol{\beta}_\alpha^\top} = -\sum_{i=1}^{n} \exp\left(\boldsymbol{x}_{\gamma i}^\top \boldsymbol{\beta}_\gamma\right) \int_0^{T_i} \psi_i(u) \left[\boldsymbol{g}_1\left(\boldsymbol{x}_{\mu i}^\top(u)\,\boldsymbol{\beta}_\mu\right) \odot \mathbf{x}_{\alpha 2i}^\top(u)\right] \cdot \left[\boldsymbol{g}_1\left(\boldsymbol{x}_{\mu i}^\top(u)\,\boldsymbol{\beta}_\mu\right) \odot \mathbf{x}_{\alpha 2i}^\top(u)\right]^\top du$$

Here $\boldsymbol{g}_1''\left(\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu\right)$ denote the second derivatives with respect to $\boldsymbol{\eta}_\mu(\mathbf{T})$, i.e. the second outer derivative which is $\mathbf{0_n}$ for a linear association and $B''(\mathbf{X}_\mu(\mathbf{T})\boldsymbol{\beta}_\mu)$ for a nonlinear association, for which again an analytical formula exists. The same setup holds for $\boldsymbol{g}_1''\left(\boldsymbol{x}_{\mu i}^\top(u)\boldsymbol{\beta}_\mu\right)$.

Appendix B

Additional modeling results

B.1 Additional modeling results from Chapter 3

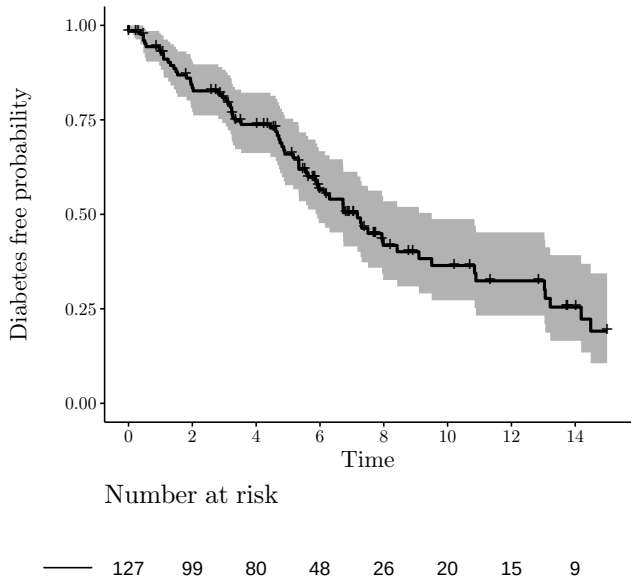

Figure B.1: Kaplan-Meier diabetes free probability estimates for the BABYDIAB/BABYDIET data and number of subjects at risk per time point.

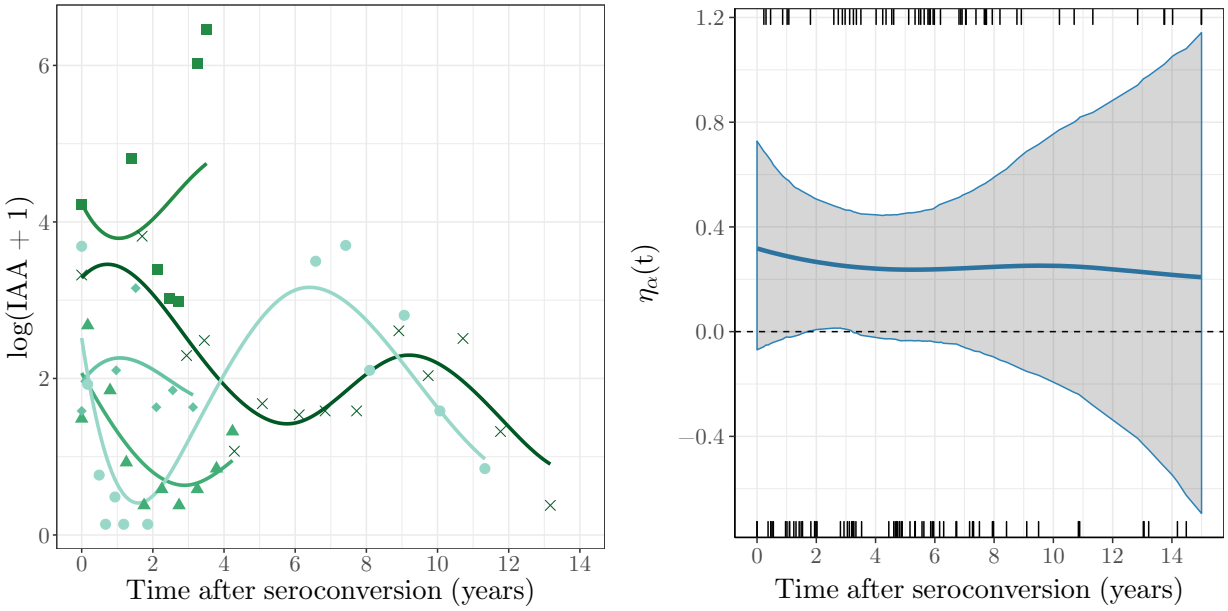

Figure B.2: Results from the sensitivity analysis for the T1D data using 12 (i.e. 4 internal) knots for both the overall mean and the functional random intercepts in the longitudinal submodel; (a) Observed values (points) and estimated trajectories (lines) of the longitudinal marker values of $log(IAA + 1)$ for five randomly selected subjects; (b) Estimated posterior mean of $\eta_\alpha(t)$ with 95% pointwise credibility bands (shaded area), observed event times (rugs bottom) and censoring times (rugs top).

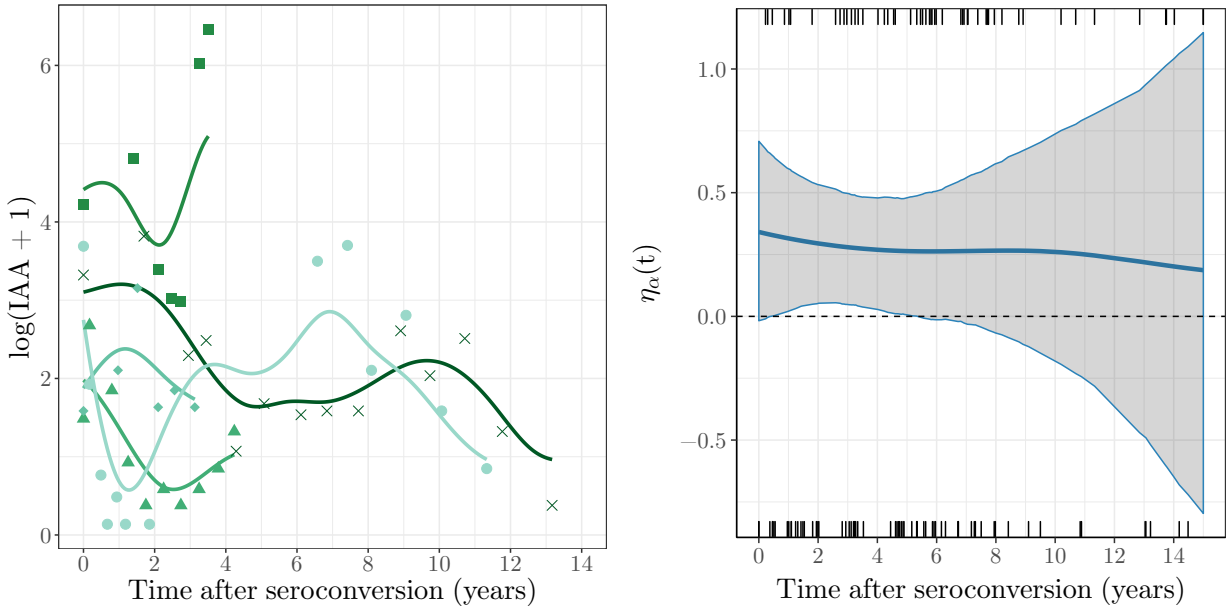

Figure B.3: Results from the sensitivity analysis for the T1D data using 20 (i.e. 8 internal) knots for both the overall mean and the functional random intercepts in the longitudinal submodel and 10 (i.e.2 internal knots) in the survival submodel; (a) Observed values (points) and estimated trajectories (lines) of the longitudinal marker values of $log(IAA + 1)$ for five randomly selected subjects; (b) Estimated posterior mean of $\eta_\alpha(t)$ with 95% pointwise credibility bands (shaded area), observed event times (rugs bottom) and censoring times (rugs top), and number of subjects at risk per time point (bottom).

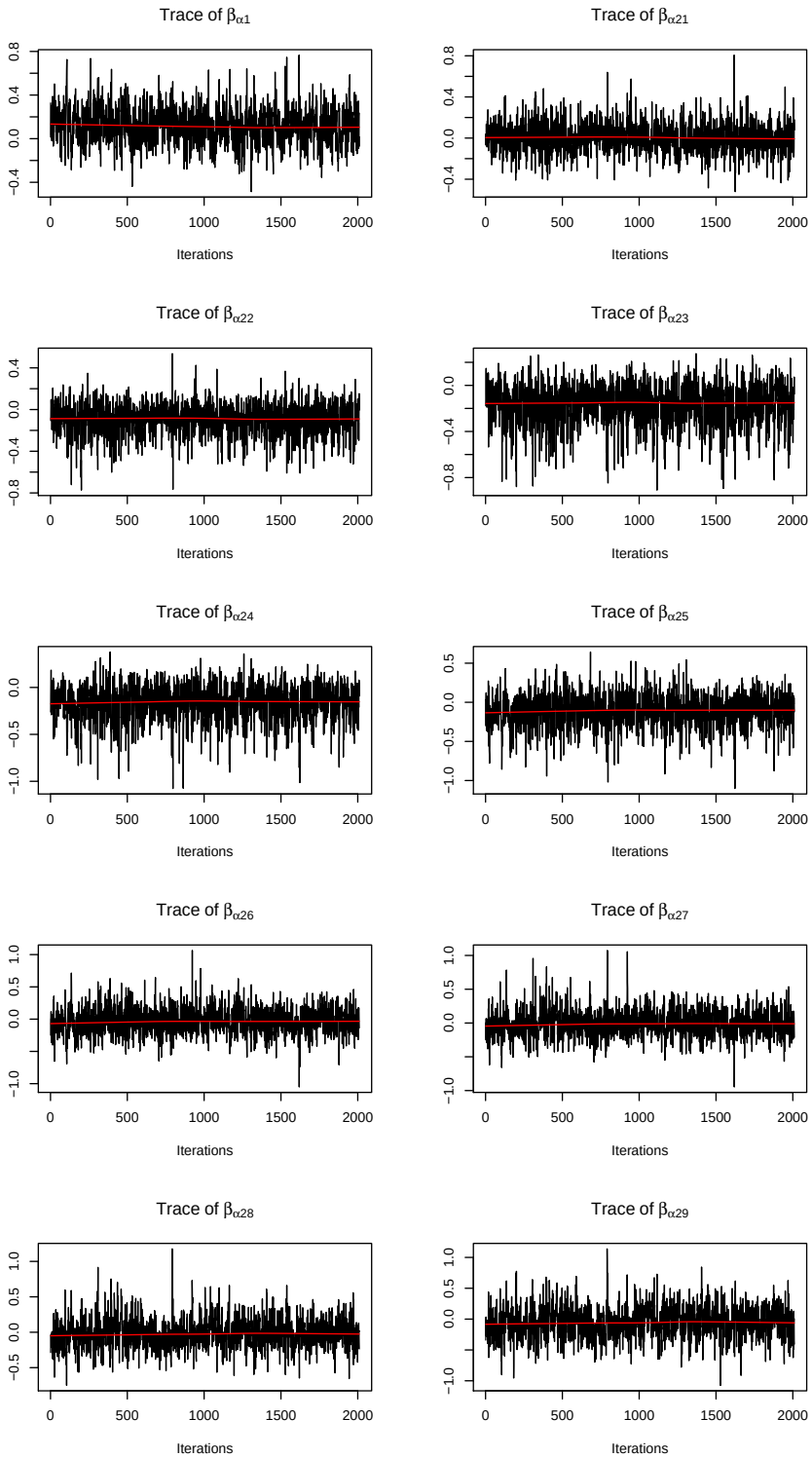

Figure B.4: Traceplots of the posterior samples for the intercept $\beta_{\alpha 1}$ and the coefficient vector $\boldsymbol{\beta}_{\alpha 1}$ in $\eta_\alpha(t)$

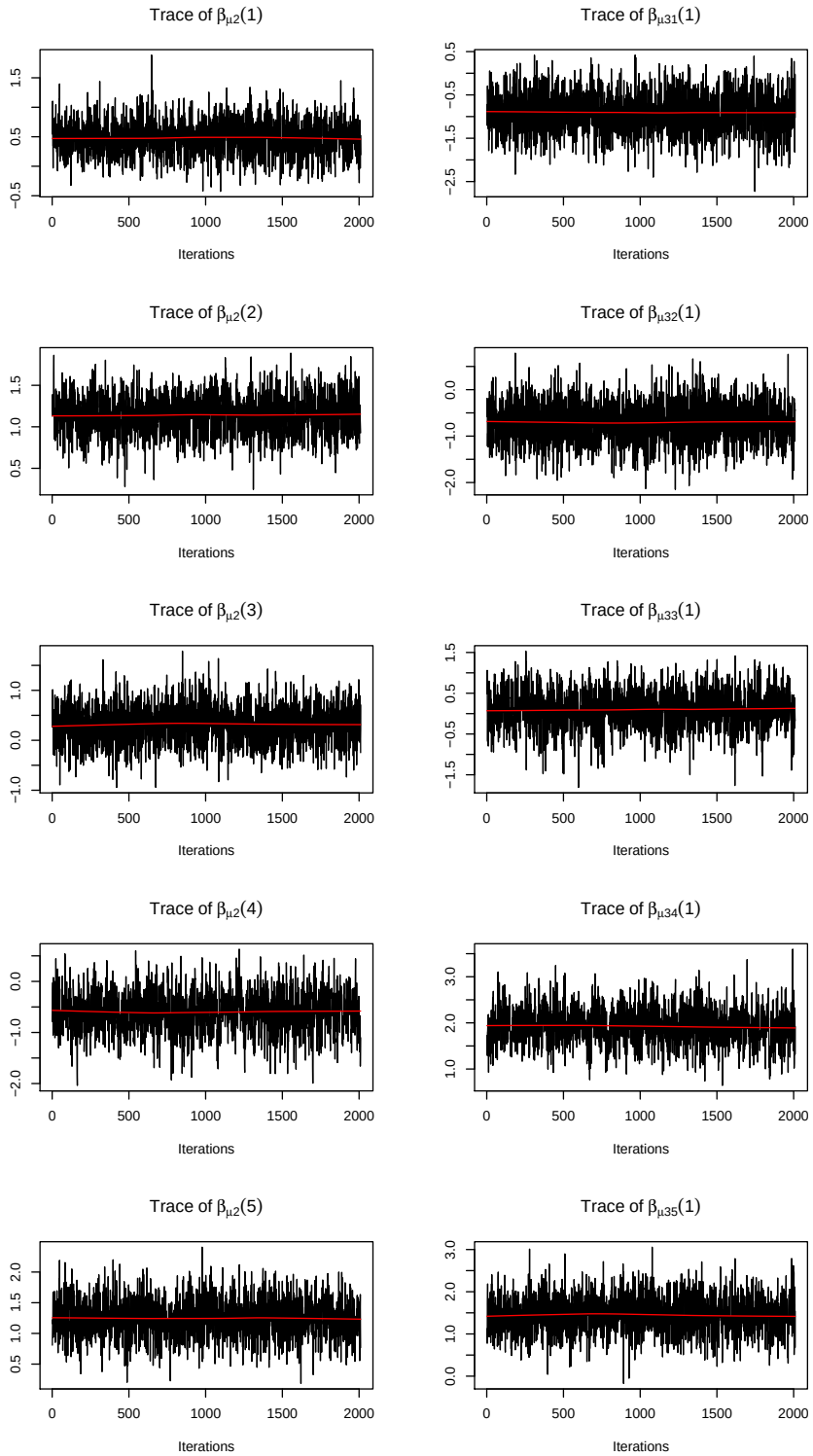

Figure B.5: Traceplots of the posterior samples for (left) the random intercepts $\beta_{\mu 2}(i)$ of subjects $i = 1, \ldots, 5$, and (right) the coefficient vector $\boldsymbol{\beta}_{\mu 3}(t, i)$ for subject $i = 1$ in $\eta_\mu(t)$

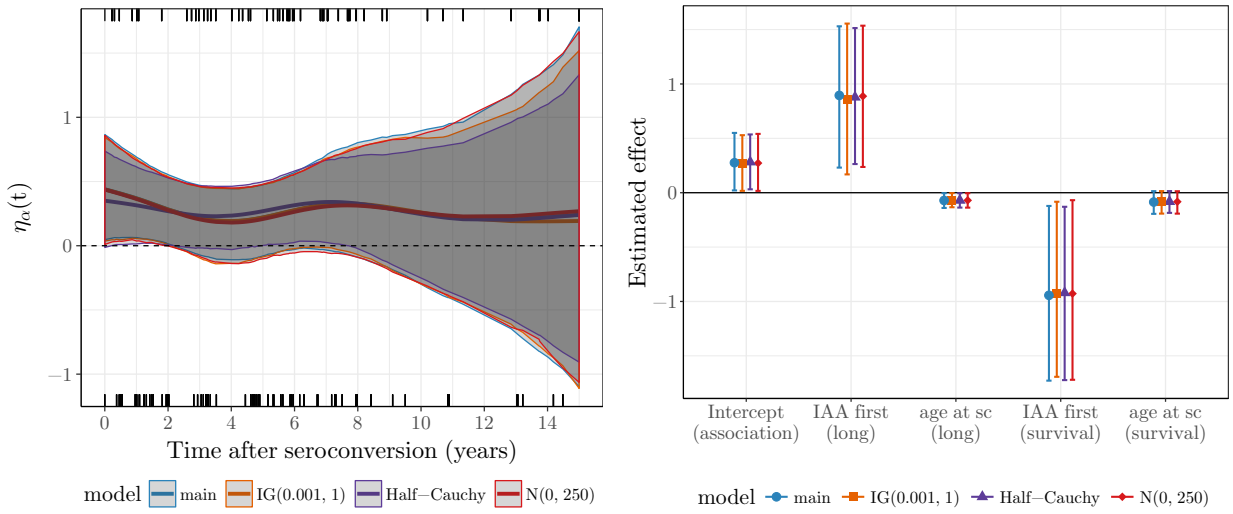

Figure B.6: Results from the sensitivity analysis for the T1D data. Estimated effects from model fits based on (i) the main model as presented in the Application section, (ii) $IG(0.001, 1)$ as prior distribution for the variance parameters, (iii) a Half-Cauchy distribution for the variance parameters and (iv) $N(0, 50^2)$ as prior for the parametric terms; (a) Estimated posterior mean of $\eta_\alpha(t)$ with 95% pointwise credibility bands (shaded area), observed event times (rugs bottom) and censoring times (rugs top); (b) linear effect estimates from the survival and longitudinal submodel; *association* stands for effects in η_α, *long* for the longitudinal submodel and *sc* indicates seroconversion.

B.2 Additional modeling results from Chapter 4

Table B.1: Description of the study population by type of persistent autoantibody. Values are reported as n (% of non-missing observations) for categorical variables and median (interquartile range) for continuous variables.

		Type of persistent autoantibody		
Variable	Total	IAA	GADA	IA2A
Total number of children	613	442	466	288
Age at respective seroconversion (years)	2.2 (1.2, 3.8)	2.0 (1.1, 3.5)	2.7 (1.6, 4.2)	2.8 (1.9, 4.5)
Girls	268 (44%)	200 (45%)	212 (45%)	112 (39%)
Country				
US	206 (34%)	136 (31%)	166 (36%)	94 (33%)
Finland	153 (25%)	125 (28%)	109 (23%)	85 (30%)
Germany	47 (8%)	40 (9%)	32 (7%)	22 (8%)
Sweden	207 (34%)	141 (32%)	159 (34%)	87 (30%)
Child having a first degree relative with T1D	128 (21%)	105 (24%)	97 (21%)	71 (25%)
HLA-DR genotype				
DR3/4	311 (51%)	241 (55%)	251 (54%)	148 (51%)
DR4/4	106 (17%)	74 (17%)	81 (17%)	64 (22%)
DR4/8	92 (15%)	71 (16%)	51 (11%)	46 (16%)
DR3/3	76 (12%)	30 (7%)	64 (14%)	16 (6%)
other	28 (5%)	26 (6%)	19 (4%)	14 (5%)
Additionally autoantibody positive for				
IAA			302 (65%)	252 (88%)
GADA		302 (68%)		237 (83%)
IA2A		252 (57%)	237 (51%)	
Autoantibody present at first seroconversion		353 (80%)	344 (74%)	40 (14%)
Number of children who developed T1D	175 (29%)	162 (37%)	134 (29%)	127 (44%)

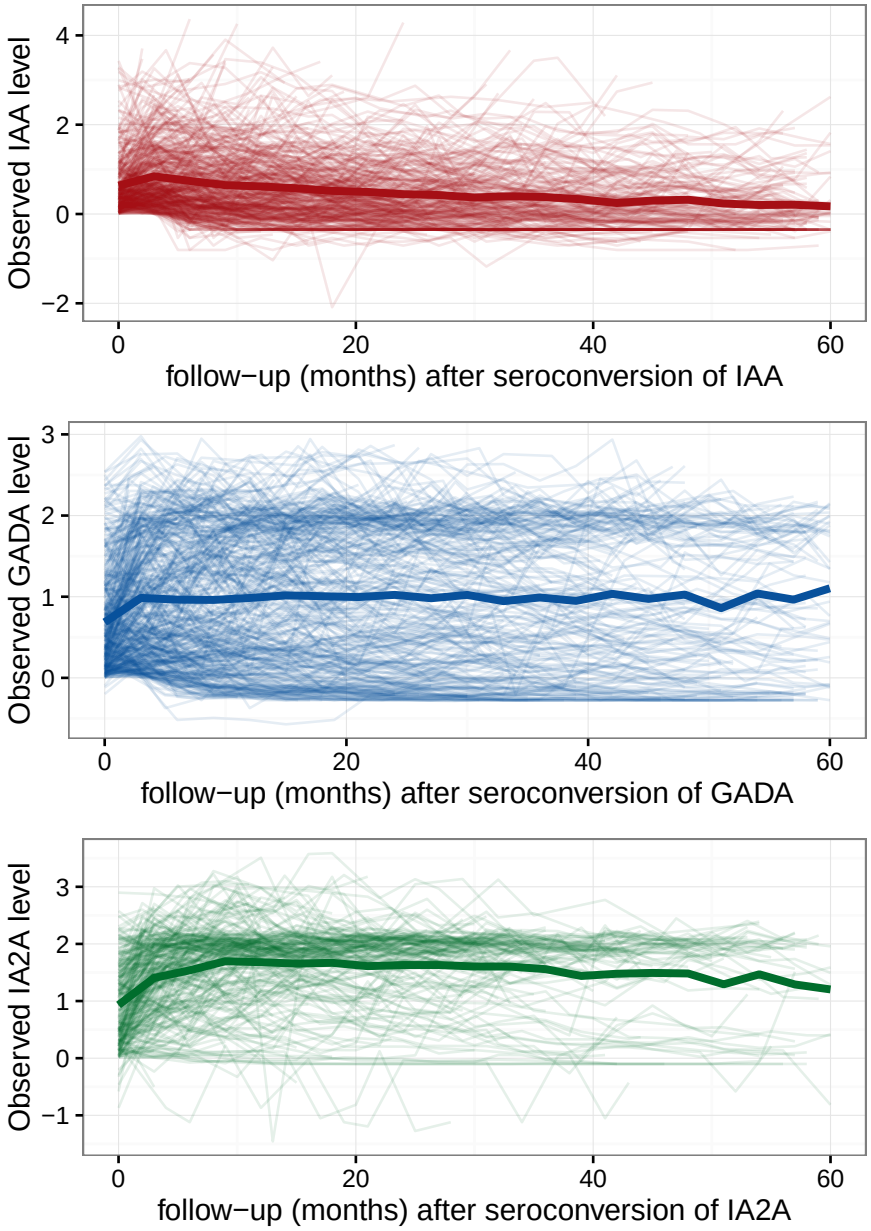

Figure B.7: Individual and mean transformed titers of IAA, GADA and IA2A autoantibodies after seroconversion to the respective autoantibody.

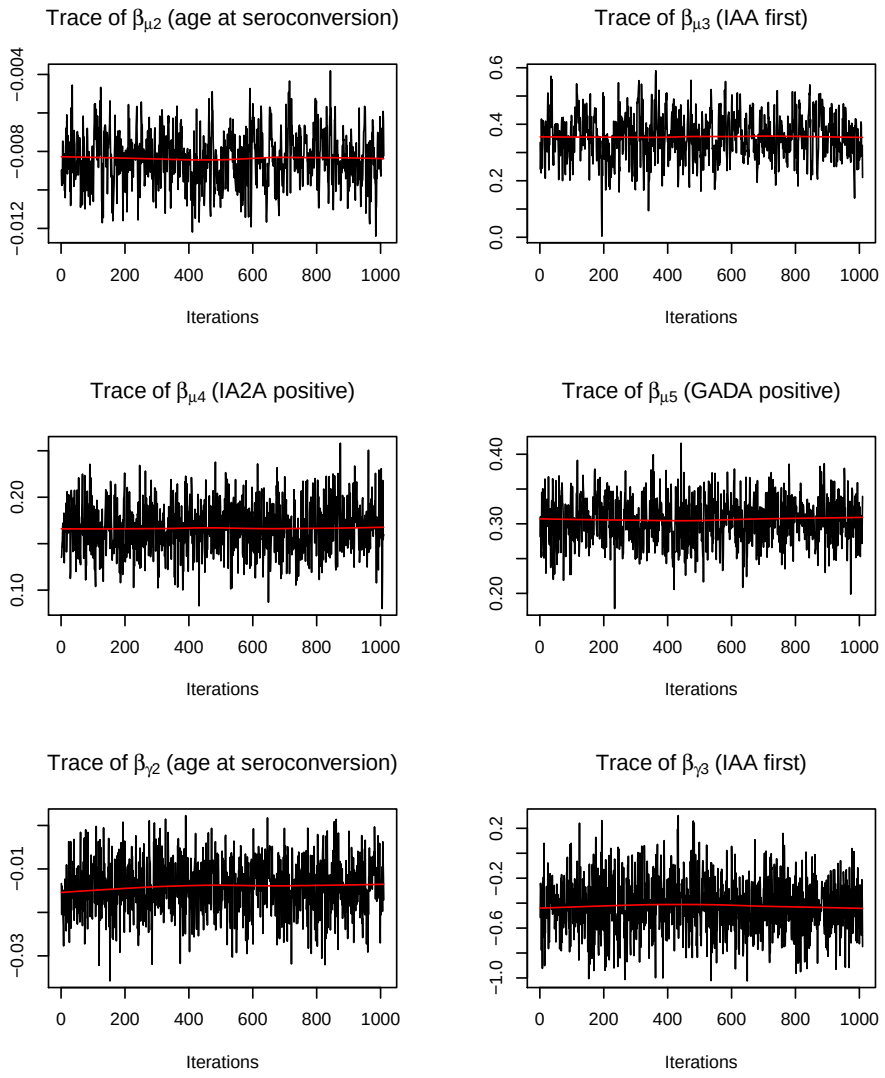

Figure B.8: Convergence plots for exemplary coefficients from the joint model of IAA trajectories and progression to T1D. Coefficients with index μ represent the longitudinal submodel and γ the survival submodel.

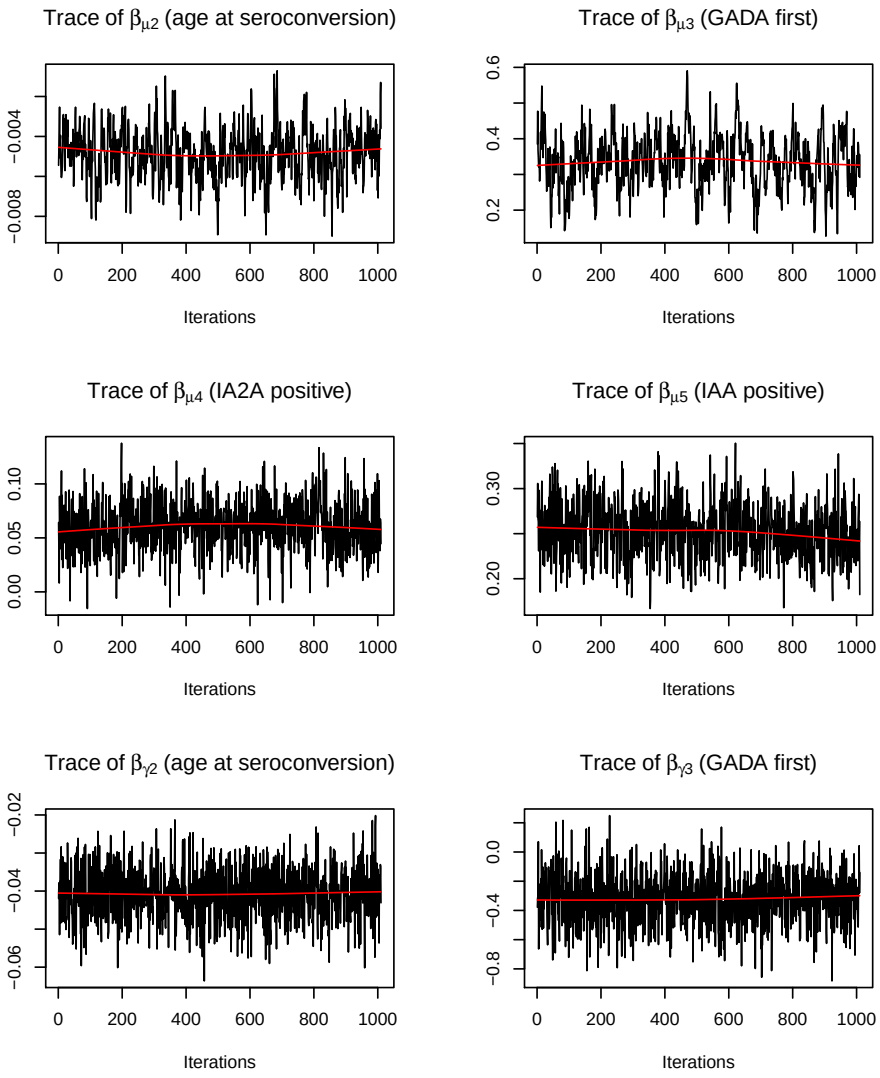

Figure B.9: Convergence plots for exemplary coefficients from the joint model of GADA trajectories and progression to T1D. Coefficients with index μ represent the longitudinal submodel and γ the survival submodel (see also Appendix).

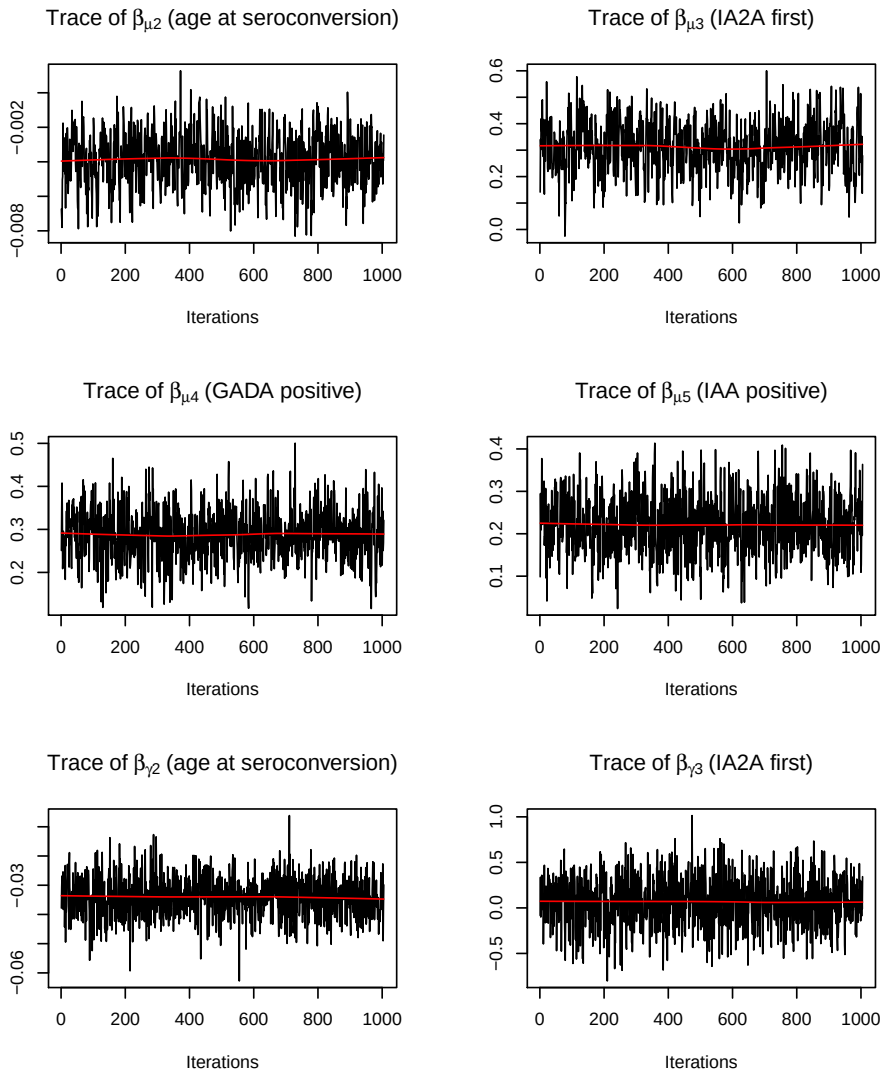

Figure B.10: Convergence plots for exemplary coefficients from the joint model of IA2A trajectories and progression to T1D. Coefficients with index μ represent the longitudinal submodel and γ the survival submodel (see also Appendix).

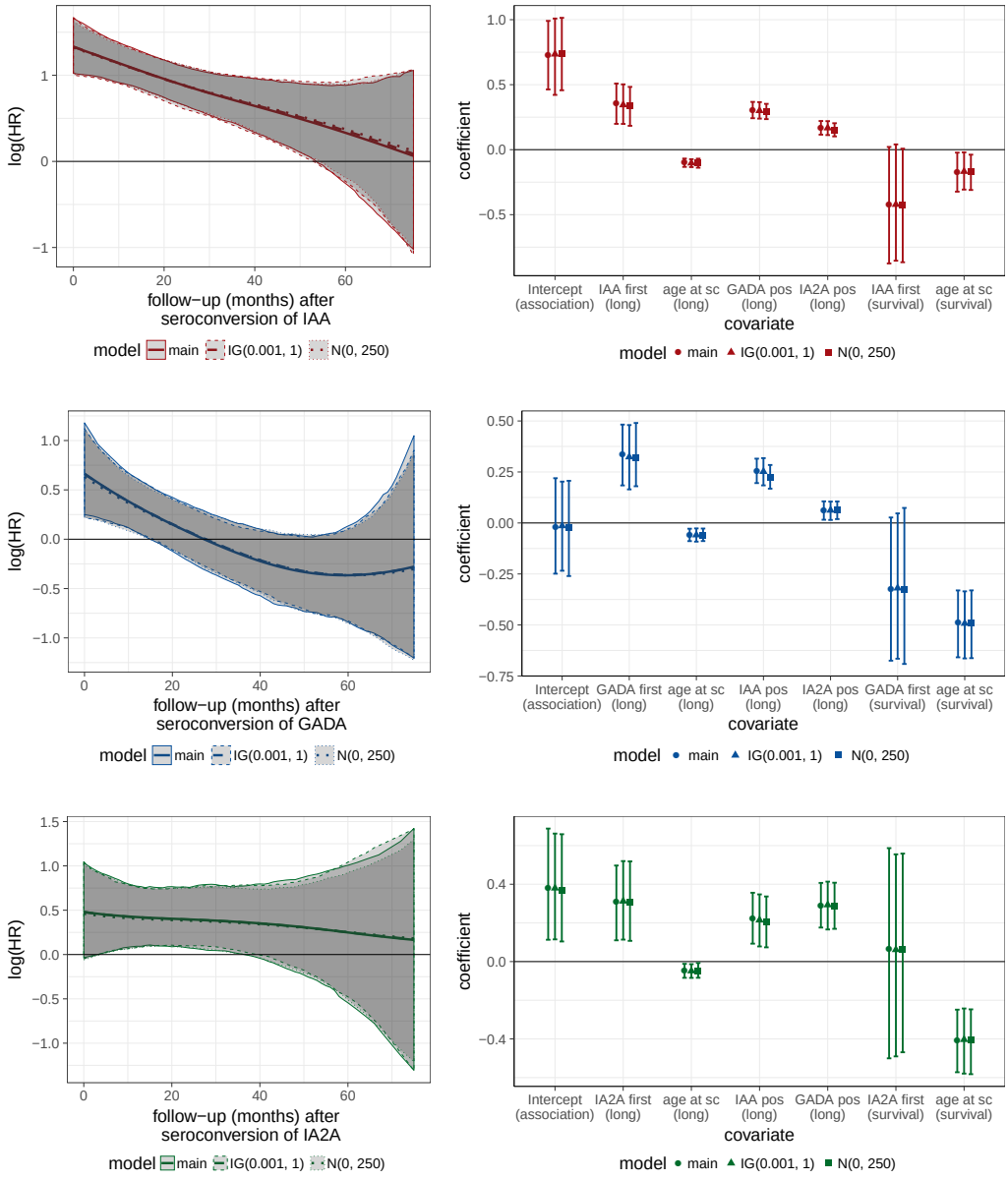

Figure B.11: Results from the sensitivity analysis for the joint models of autoantibody trajectories (IAA, GADA and IA2A) and progression to T1D data. Estimated coefficients from model fits based on (i) the main model as presented in the Application section with $IG(0.0001, 0.0001)$ as prior distribution for the variance parameters and $N(0, 1000^2)$ as weakly informative prior for the parametric terms, (ii) with $IG(0.001, 1)$ as prior distribution for the variance parameters, and (iii) $N(0, 50^2)$ as prior for the parametric terms. Left panel: Posterior mean estimates (lines) and 95% pointwise credibility intervals (shaded areas) of $\eta_\alpha(t)$. Right panel: Posterior mean estimates of coefficients and hazard ratios with corresponding 95% credibility intervals. *Association* denotes the intercept of $\eta_\alpha(t)$, *long* the coefficients of the longitudinal submodel and *survival* the coefficients (i.e. the log hazard ratio) of the survival model.

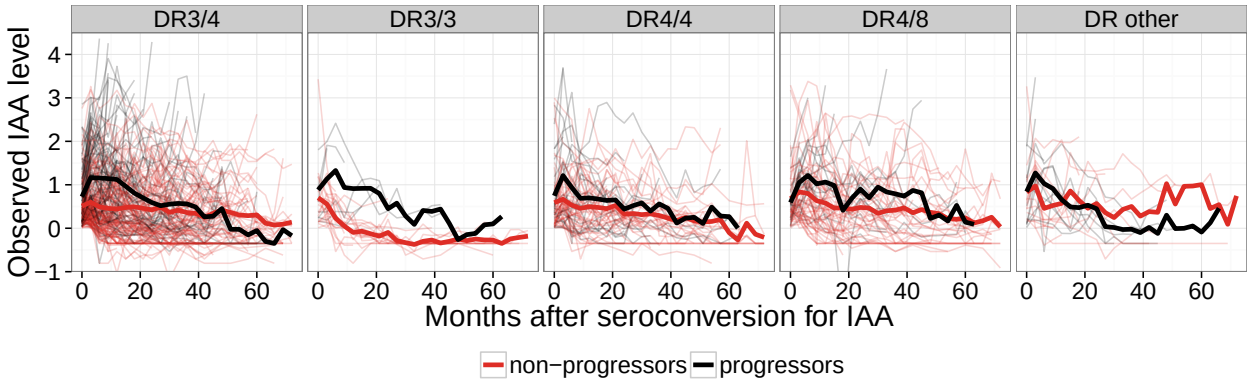

Figure B.12: Individual and mean transformed titers of IAA autoantibodies after seroconversion per HLA genotype, stratified for progression to T1D.

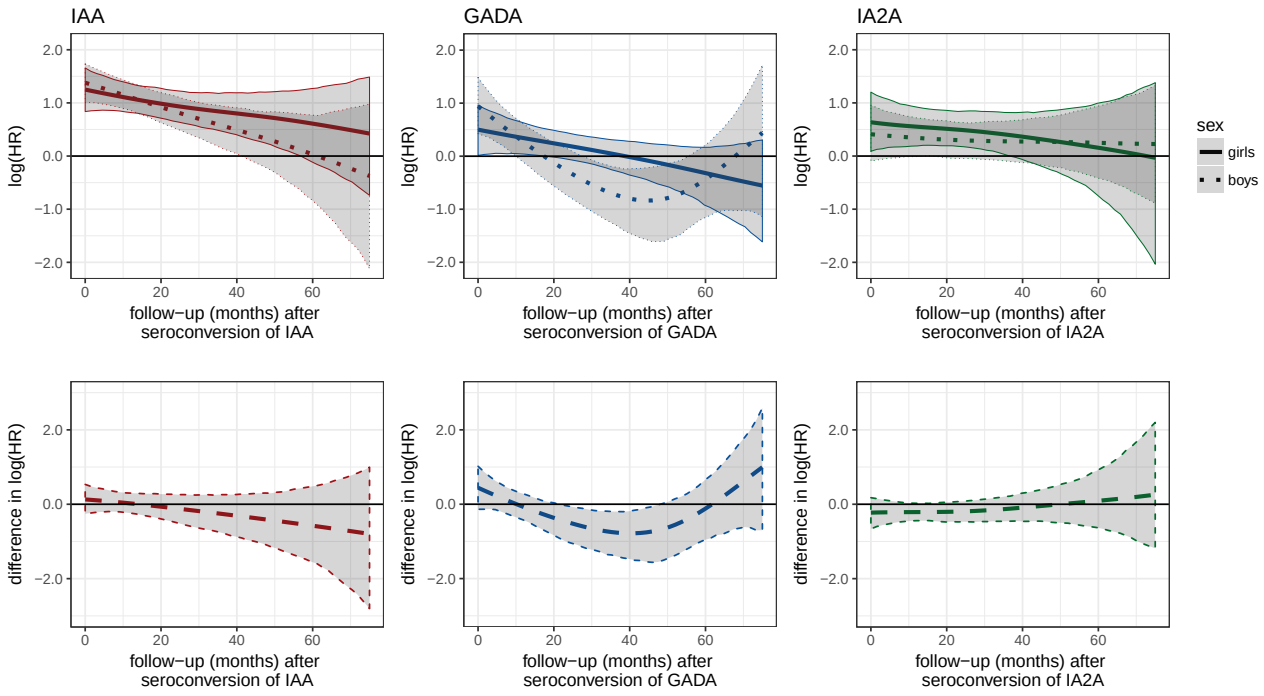

Figure B.13: Posterior mean estimates (lines / dots) and 95% credibility intervals (shaded areas) of $\eta_\alpha(t, \text{sex})$, the time-varying log hazard ratio (HR) of the association between longitudinal autoantibody trajectories and T1D progression stratified for girls and boys (upper panel) and of the difference of the association between the groups over time, $\eta_\alpha(t, \text{sex} = \text{boys}) - \eta_\alpha(t, \text{sex} = \text{girls})$ (lower panel).

B.3 Additional modeling results from Chapter 5

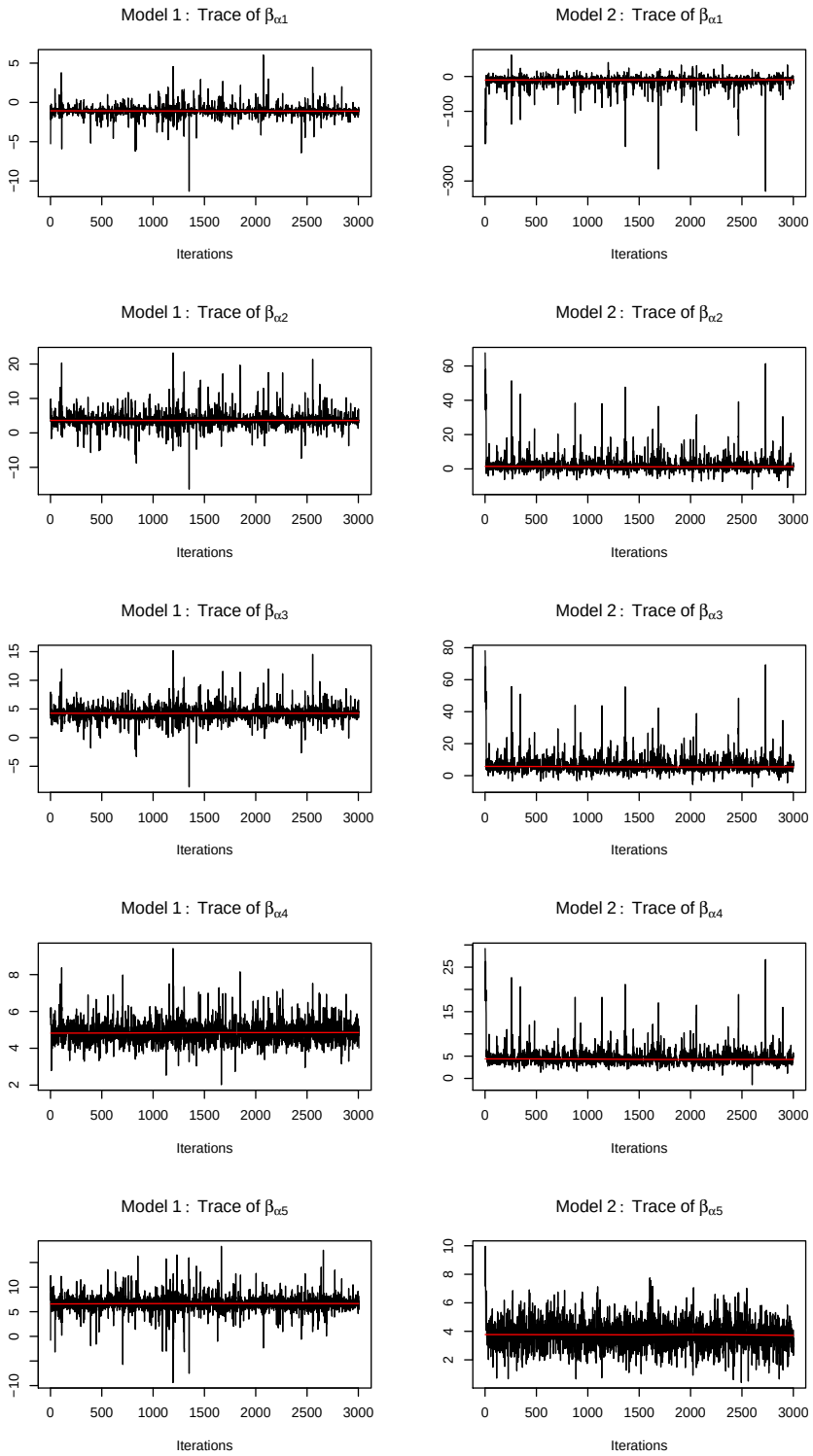

Figure B.14: Traceplots of the posterior samples for $\boldsymbol{\beta}_\alpha$ in $\eta_\alpha(\eta_\mu(t))$ from model 1 (left) and model 2 (right).

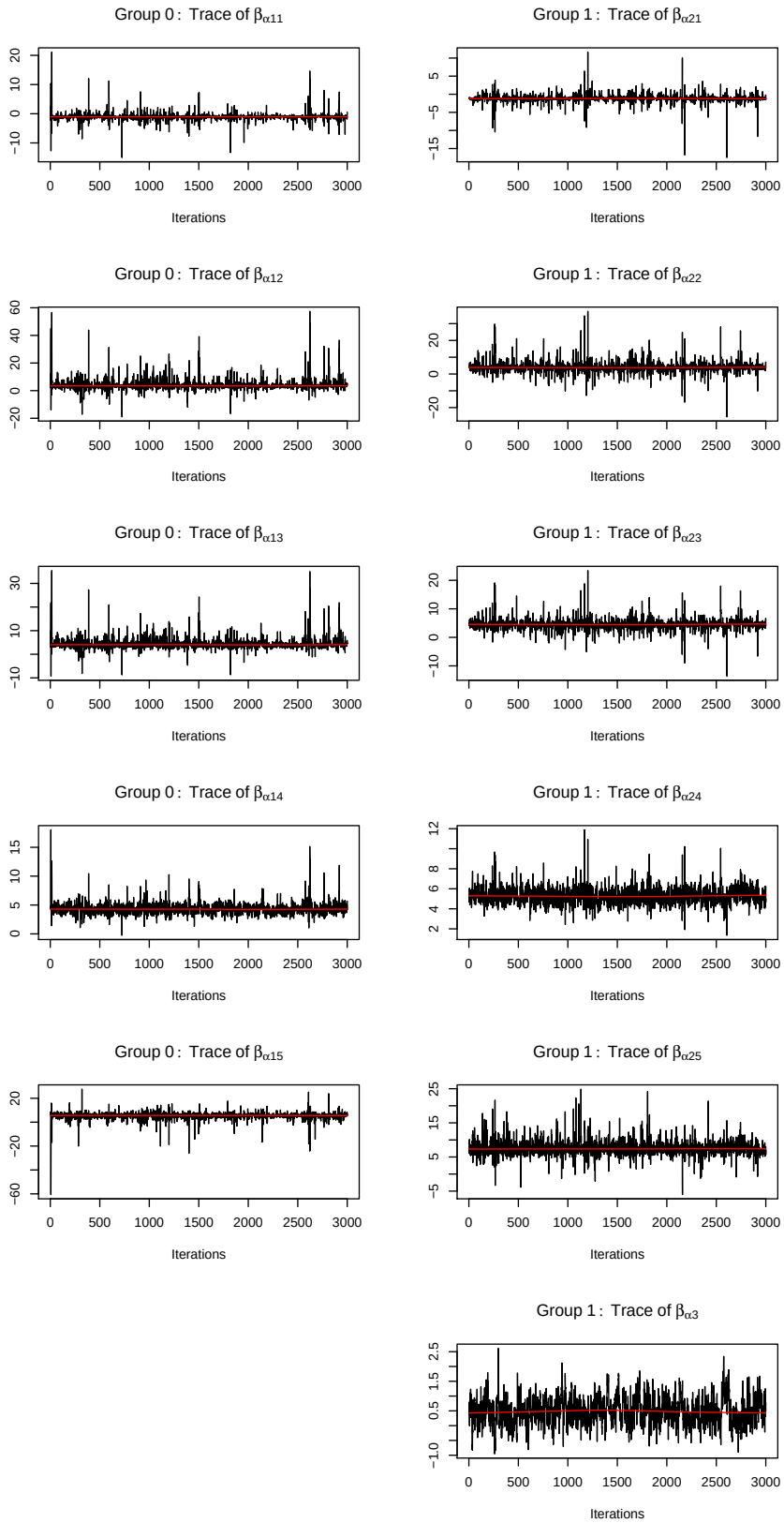

Figure B.15: Traceplots of the posterior samples of model 3 for $\boldsymbol{\beta}_\alpha$ in $\eta_\alpha(\eta_\mu(t), g)$ from subjects without an enlarged liver at baseline (left) and with this condition (right). For the latter also a group-specific intercept $\beta_{\alpha 3}$ is estimated.

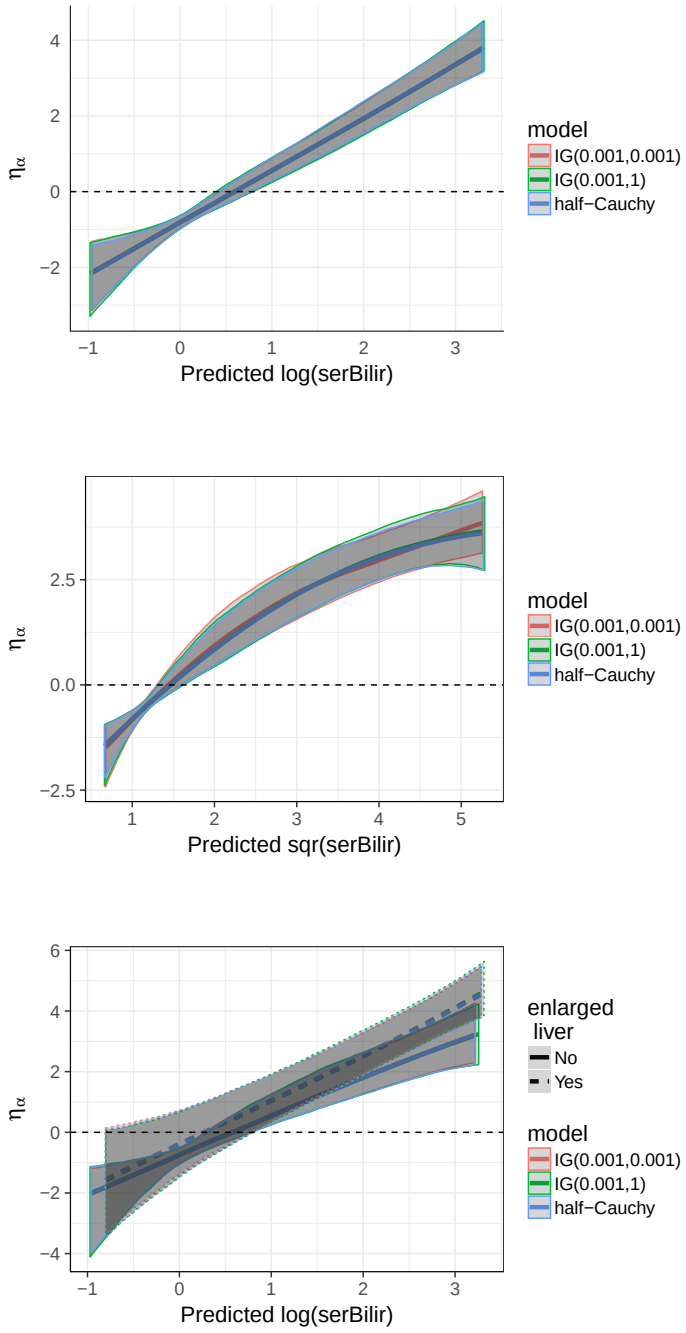

Figure B.16: Results from the sensitivity analysis for the PBC data. Estimated posterior mean of the association $\eta_\alpha(\eta_\mu(t))$ of (a) model 1 (nonlinear estimation of $\log(Bilirubin)$), (b)model 2 (nonlinear estimation of $\sqrt{Bilirubin}$) and (c) model 3 (nonlinear estimation of $\log(Bilirubin)$ of patients with and without enlarged liver at baseline) using as prior distribution for the variance parameters (i) $IG(0.001, 0.001)$, (ii) $IG(0.001, 1)$, and (iii) a half-Cauchy distribution for the variance parameters.

Bibliography

Achenbach, P., Warncke, K., Reiter, J., Naserke, H. E., Williams, A. J., Bingley, P. J., Bonifacio, E., and Ziegler, A. G. (2004). Stratification of type 1 diabetes risk on the basis of islet autoantibody characteristics. *Diabetes*, 53(2):384–92.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

American Diabetes Association (2011). Executive summary: standards of medical care in diabetes–2011. *Diabetes Care*, 34(Suppl 1):S11–61.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10(4):1100–1120.

Anderson, M. S. and Bluestone, J. A. (2005). The NOD mouse: a model of immune dysregulation. *Annual Review of Immunology*, 23:447–485.

Andrinopoulou, E.-R. and Rizopoulos, D. (2016). Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. *Statistics in Medicine*, 35(26):4813–4823.

Andrinopoulou, E.-R., Rizopoulos, D., Eilers, P. H., and Takkenberg, J. J. (2016). Improved dynamic predictions from joint models of longitudinal and survival data with time-varying effects using P-splines. *arXiv preprint*. arXiv:1609.03439.

Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J. J. M., and Lesaffre, E. (2014). Joint modeling of two longitudinal outcomes and competing risk data. *Statistics in Medicine*, 33(18):3167–3178.

Asar, O., Ritchie, J., Kalra, P. A., and Diggle, P. J. (2015). Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *International Journal of Epidemiology*, 44(1):334–344.

Atkinson, M. A. and Leiter, E. H. (1999). The NOD mouse model of type 1 diabetes: as good as it gets? *Nature Medicine*, 5(6):601–604.

Baghfalaki, T., Ganjali, M., and Hashemi, R. (2014). Bayesian joint modeling of longitudinal measurements and time-to-event data using robust distributions. *Journal of Biopharmaceutical Statistics*, 24(4):834–855.

Barrett, J., Diggle, P., Henderson, R., and Taylor-Robinson, D. (2015). Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):131–148.

Barrett, J. and Su, L. (2017). Dynamic predictions using flexible joint models of longitudinal and time-to-event data. *Statistics in Medicine*, 36(9):1447–1460.

Belitz, C. and Lang, S. (2008). Simultaneous selection of variables and smoothing parameters in structured additive regression models. *Computational Statistics & Data Analysis*, 53(1):61–81.

Bender, R., Augustin, T., and Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, 24(11):1713–1723.

Bonifacio, E. (2015). Predicting type 1 diabetes using biomarkers. *Diabetes Care*, 38(6):989–996.

Brown, E. R. (2009). Assessing the association between trends in a biomarker and risk of event with an application in pediatric HIV/AIDS. *The Annals of Applied Statistics*, 3(3):1163–1182.

Brown, E. R. and Ibrahim, J. G. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59(2):221–228.

Brown, E. R., Ibrahim, J. G., and DeGruttola, V. (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73.

Buchholz, A. and Sauerbrei, W. (2011). Comparison of procedures to assess non-linear and time-varying effects in multivariable models for survival data. *Biometrical Journal*, 53(2):308–331.

Carlin, B. P. and Louis, T. A. (2008). *Bayesian methods for data analysis*. Chapman and Hall/CRC, Boca Raton.

Chi, Y.-Y. and Ibrahim, J. G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2):432–445.

Cleveland, W. S., Grosse, E., and Shyu, W. M. (1992). Local regression models. In Chambers, J. M. and Hastie, T. J., editors, *Statistical Models in S*, volume 2, pages 309–376. Wadsworth, Pacific Grove.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

Crowther, M. J. (2012). STJM: Stata module to fit shared parameter joint models of longitudinal and survival data. Statistical Software Components, Boston College Department of Economics.

Crowther, M. J., Andersson, T. M.-L., Lambert, P. C., Abrams, K. R., and Humphreys, K. (2016). Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification. *Statistics in Medicine*, 35(7):1193–1209.

Crowther, M. J. and Lambert, P. C. (2013). Simulating biologically plausible complex survival data. *Statistics in Medicine*, 32(23):4118–4134.

d. Boor, C. (1978). *A Practical Guide to Splines*. Springer, New York.

Dafni, U. G. and Tsiatis, A. A. (1998). Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics*, 54(4):1445–1462.

Daher Abdi, Z., Essig, M., Rizopoulos, D., Le Meur, Y., Prémaud, A., Woillard, J. B., Rérolle, J. P., Marquet, P., and Rousseau, A. (2013). Impact of longitudinal exposure to mycophenolic acid on acute rejection in renal-transplant recipients using a joint modeling approach. *Pharmacological Research*, 72:52–60.

Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(1):49–93.

Diggle, P. J., Sousa, I., and Chetwynd, A. G. (2008). Joint modelling of repeated measurements and time-to-event outcomes: The fourth Armitage lecture. *Statistics in Medicine*, 27(16):2981–2998.

Ding, J. and Wang, J.-L. (2008). Modeling longitudinal data with monparametric multiplicative random effects jointly with survival data. *Biometrics*, 64(2):546–556.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–102.

Elashoff, R. M., Li, G., and Li, N. (2007). An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in Medicine*, 26(14):2813–2835.

Endesfelder, D., Hagen, M., Winkler, C., Haupt, F., Zillmer, S., Knopff, A., Bonifacio, E., Ziegler, A.-G., Zu Castell, W., and Achenbach, P. (2016). A novel approach for the analysis of longitudinal profiles reveals delayed progression to type 1 diabetes in a subgroup of multiple-islet-autoantibody-positive children. *Diabetologia*, 59(10):2172–2180.

Fahrmeir, L., Kneib, T., and Lang, S. (2004). Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica*, 14(3):731–761.

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

Farcomeni, A. and Viviani, S. (2015). Longitudinal quantile regression in the presence of informative dropout through longitudinal–survival joint modeling. *Statistics in Medicine*, 34(7):1199–1213.

Faucett, C. L. and Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine*, 15(15):1663–1685.

Ferrer, L., Putter, H., and Proust-Lima, C. (2017). Individual dynamic predictions using landmarking and joint modelling: validation of estimators and robustness assessment. *arXiv preprint*. arXiv:1707.03706.

Gagnon, B., Abrahamowicz, M., Xiao, Y., Beauchamp, M.-E., MacDonald, N., Kasymjanova, G., Kreisman, H., and Small, D. (2010). Flexible modeling improves assessment of prognostic value of C-reactive protein in advanced non-small cell lung cancer. *British Journal of Cancer*, 102(7):1113–1122.

Gellar, J. E., Colantuoni, E., Needham, D. M., and Crainiceanu, C. M. (2014). Variable-domain functional regression for modeling ICU data. *Journal of the American Statistical Association*, 109(508):1425–1439.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton.

Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994). A language and program for complex Bayesian modelling. *The Statistician*, 43(1):169–178.

Gould, A. L., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S., and Bois, F. Y. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Statistics in Medicine*, 34(14):2181–2195.

Gras, L., Geskus, R. B., Jurriaans, S., Bakker, M., van Sighem, A., Bezemer, D., Fraser, C., Prins, J. M., and Berkhout, B. (2013). Has the rate of CD4 cell count decline before initiation of antiretroviral therapy changed over the course of the dutch HIV epidemic among MSM? *PLoS ONE*, 8(5):1–8.

Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, 87(420):942–951.

Gueorguieva, R., Rosenheck, R., and Lin, H. (2012). Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, 175(2):417–433.

Guo, X. and Carlin, B. P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58(1):16–24.

Hagopian, W. A., Erlich, H., Lernmark, Å., Rewers, M., Ziegler, A. G., Simell, O., Akolkar, B., Vogt, R., J., Blair, A., Ilonen, J., Krischer, J., She, J., and the TEDDY Study Group (2011). The Environmental Determinants of Diabetes in the Young (TEDDY): genetic criteria and international diabetes risk screening of 421 000 infants. *Pediatric Diabetes*, 12(8):733–43.

Harrell, F. E., Lee, K. L., and Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15(4):361–387.

Hastie, T. and Tibshirani, R. (1995). Generalized additive models for medical research. *Statistical Methods in Medical Research*, 4(3):187–196.

Hatfield, L. A., Boye, M. E., Hackshaw, M. D., and Carlin, B. P. (2012). Multilevel Bayesian models for survival times and longitudinal patient-reported outcomes with many zeros. *Journal of the American Statistical Association*, 107(499):875–885.

He, B. and Luo, S. (2016). Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson's disease. *Statistical Methods in Medical Research*, 25(4):1346–1358.

Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480.

Henderson, R., Diggle, P., and Dobson, A. (2002). Identification and efficacy of longitudinal markers for survival. *Biostatistics*, 3(1):33–50.

Hickey, G. L., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology*, 16(1):117.

Hofner, B., Kneib, T., Hartl, W., and Kuchenhoff, H. (2011). Building Cox-type structured hazard regression models with time-varying effects. *Statistical Modelling*, 11(1):3–24.

Holländer, N. and Schumacher, M. (2006). Estimating the functional form of a continuous covariate's effect on survival time. *Computational Statistics & Data Analysis*, 50(4):1131–1151.

Huang, X., Li, G., and Elashoff, R. M. (2010). A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements. *Statistics and Its Interface*, 3(2):185–195.

Hummel, M., Bonifacio, E., Schmid, S., Walter, M., Knopff, A., and Ziegler, A.-G. (2004). Brief communication: Early appearance of islet autoantibodies predicts childhood type 1 diabetes in offspring of diabetic parents. *Annals of Internal Medicine*, 140(11):882–886.

Hummel, S., Pflüger, M., Hummel, M., Bonifacio, E., and Ziegler, A.-G. (2011). Primary dietary intervention study to reduce the risk of islet autoimmunity in children at increased risk for type 1 diabetes. *Diabetes Care*, 34(6):1301–1305.

Hurvich, C. M., Simonoff, J. S., and Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):271–293.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2004). Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica*, 14(3):863–883.

Inaba, H., Surprise, H. C., Pounds, S., Cao, X., Howard, S. C., Ringwald-Smith, K., Buaboonnam, J., Dahl, G., Bowman, W. P., Taub, J. W., Campana, D., Pui, C.-H., Ribeiro, R. C., and Rubnitz, J. E. (2012). Effect of body mass index on the outcome of children with acute myeloid leukemia. *Cancer*, 118(23):5989–5996.

Insel, R. A., Dunne, J. L., Atkinson, M. A., Chiang, J. L., Dabelea, D., Gottlieb, P. A., Greenbaum, C. J., Herold, K. C., Krischer, J. P., Lernmark, Å., Ratner, R. E., Rewers, M. J., Schatz, D. A., Skyler, J. S., Sosenko, J. M., and Ziegler, A.-G. (2015). Staging presymptomatic type 1 diabetes: A scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. *Diabetes Care*, 38(10):1964–1974.

Jiang, B., Wang, N., Sammel, M. D., and Elliott, M. R. (2015). Modelling short- and long-term characteristics of follicle stimulating hormone as predictors of severe hot flashes in the Penn Ovarian Aging Study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(5):731–753.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, Hoboken, 2nd edition.

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

Kauermann, G., Xu, R., and Vaida, F. (2008). Stacked Laplace-EM algorithm for duration models with time-varying and random effects. *Computational Statistics & Data Analysis*, 52(5):2514–2528.

Klein, N., Kneib, T., Klasen, S., and Lang, S. (2015a). Bayesian structured additive distributional regression for multivariate responses. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(4):569–591.

Klein, N., Kneib, T., Lang, S., and Sohn, A. (2015b). Bayesian structured additive distributional regression with an application to regional income inequality in Germany. *Annals of Applied Statistics*, 9(2):1024–1052.

Kneib, T. (2005). *Mixed model based inference in structured additive regression*. PhD thesis, Ludwigs-Maximilians-Universität München.

Kneib, T. and Fahrmeir, L. (2007). A Mixed Model Approach for Geoadditive Hazard Regression. *Scandinavian Journal of Statistics*, 34(1):207–228.

Krischer, J. P., Lynch, K. F., Schatz, D. A., Ilonen, J., Lernmark, Å., Hagopian, W. A., Rewers, M. J., She, J.-X., Simell, O. G., Toppari, J., Ziegler, A.-G., Akolkar, B., Bonifacio, E., and the TEDDY Study Group (2015). The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. *Diabetologia*, 58(5):980–987.

Köhler, M., Beyerlein, A., Vehik, K., Greven, S., Umlauf, N., Lernmark, Å., Hagopian, W. A., Rewers, M., She, J.-X., Toppari, J., Akolkar, B., Krischer, J. P., Bonifacio, E., and Ziegler, A.-G. (2017a). Joint modeling of longitudinal autoantibody patterns and progression to type 1 diabetes: results from the TEDDY study. *Acta Diabetologica*. (online first).

Köhler, M., Umlauf, N., Beyerlein, A., Winkler, C., Ziegler, A.-G., and Greven, S. (2017b). Flexible bayesian additive joint models with an application to type 1 diabetes research. *Biometrical Journal*. (online first).

Köhler, M., Umlauf, N., and Greven, S. (2017c). Nonlinear association structures in flexible Bayesian additive joint models. *arXiv preprint*. arXiv: 1708.06337.

Lang, S. and Brezger, A. (2004). Bayesian P-Splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212.

Little, R. and Rubin, D. (2002). *Statistical analysis with missing data*. Wiley, Hoboken.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134.

Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika*, 81(3):471–483.

Liu, L. (2009). Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statistics in Medicine*, 28(6):972–986.

Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *Canadian Journal of Statistics*, 31(2):115–128.

Mariño, E., Batten, M., Groom, J., Walters, S., Liuwantara, D., Mackay, F., and Grey, S. T. (2008). Marginal-zone B-cells of nonobese diabetic mice expand with diabetes onset, invade the pancreatic lymph nodes, and present autoantigen to diabetogenic T-cells. *Diabetes*, 57(2):395–404.

Marrero, I., Vong, A., Dai, Y., and Davies, J. D. (2012). T cell populations in the pancreatic lymph node naturally and consistently expand and contract in NOD mice as disease progresses. *Molecular Immunology*, 52(1):9–18.

Martins, R., Silva, G. L., and Andreozzi, V. (2016). Bayesian joint modeling of longitudinal and spatial survival AIDS data. *Statistics in Medicine*, 35(19):3368–3384.

Meyer, S., Woodward, M., Hertel, C., Vlaicu, P., Haque, Y., Kärner, J., Macagno, A., Onuoha, S., Fishman, D., Peterson, H., Metsküla, K., Uibo, R., Jäntti, K., Hokynar, K., Wolff, A., Krohn, K., Ranki, A., Peterson, P., Kisand, K., and Hayday, A. (2016). AIRE-deficient patients harbor unique high-affinity disease-ameliorating autoantibodies. *Cell*, 166(3):582–595.

Mrena, S., Virtanen, S. M., Laippala, P., Kulmala, P., Hannila, M.-L., Åkerblom, H. K., Knip, M., and the Childhood Diabetes in Finland Study Group (2006). Models for predicting type 1 diabetes in siblings of affected children. *Diabetes Care*, 29(3):662–667.

Murtaugh, P. A., Dickson, E. R., Van Dam, G. M., Malinchoc, M., Grambsch, P. M., Langworthy, A. L., and Gips, C. H. (1994). Primary biliary cirrhosis: Prediction of short-term survival based on repeated patient visits. *Hepatology*, 20(1):126–134.

Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, 31(3):705–741.

Parikka, V., Näntö-Salonen, K., Saarinen, M., Simell, T., Ilonen, J., Hyöty, H., Veijola, R., Knip, M., and Simell, O. (2012). Early seroconversion and rapidly increasing autoantibody concentrations predict prepubertal manifestation of type 1 diabetes in children at genetic risk. *Diabetologia*, 55(7):1926–1936.

Patterson, C. C., Dahlquist, G. G., Gyurus, E., Green, A., Soltesz, G., and Group, E. S. (2009). Incidence trends for childhood type 1 diabetes in europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. *Lancet*, 373(9680):2027–33.

Philipson, P., Sousa, I., Diggle, P. J., Williamson, P., Kolamunnage-Dona, R., Henderson, R., and Hickey, G. L. (2017). *joineR: Joint Modelling of Repeated Measurements and Time-to-Event Data*. R package version 1.2.2, available at `https://github.com/graemeleehickey/joineR/`.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer, New York.

Plummer, M. et al. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, page 125. Vienna, Austria.

Pozzilli, P., Signore, A., Williams, A. J., and Beales, P. E. (1993). NOD mouse colonies around the world–recent facts and figures. *Immunology Today*, 14(5):193–196.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69(2):331–342.

Proust-Lima, C., Joly, P., Dartigues, J.-F., and Jacqmin-Gadda, H. (2009). Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach. *Computational Statistics & Data Analysis*, 53(4):1142–1154.

Proust-Lima, C., Letenneur, L., and Jacqmin-Gadda, H. (2007). A nonlinear latent class model for joint analysis of multivariate longitudinal data and a binary outcome. *Statistics in Medicine*, 26(10):2229–2245.

Proust-Lima, C., Séne, M., Taylor, J. M., and Jacqmin-Gadda, H. (2014). Joint latent class models for longitudinal and time-to-event data: a review. *Statistical Methods in Medical Research*, 23(1):74–90.

Proust-Lima, C. and Taylor, J. M. G. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics*, 10(3):535–549.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at `http://www.R-project.org/`.

Rizopoulos, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–829.

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Chapman and Hall/CRC, Boca Raton.

Rizopoulos, D. (2016a). *JMbayes: Joint Modeling of Longitudinal and Time-to-Event Data under a Bayesian Approach*. R package version 0.7-9, available at `https://cran.r-project.org/package=JMbayes`.

Rizopoulos, D. (2016b). The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *Journal of Statistical Software*, 72(7):1–46.

Rizopoulos, D. and Ghosh, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*, 30(12):1366–1380.

Rizopoulos, D., Hatfield, L. A., Carlin, B. P., and Takkenberg, J. J. M. (2014). Combining dynamic predictions from joint models for longitudinal and time-to-event data using Bayesian model averaging. *Journal of the American Statistical Association*, 109(508):1385–1397.

Rizopoulos, D., Verbeke, G., Lesaffre, E., and Vanrenterghem, Y. (2008). A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics*, 64(2):611–619.

Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer, New York.

Rondeau, V., Mazroui, Y., and Gonzalez, J. R. (2012). frailtypack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software*, 47(4):1–28.

Royston, P. and Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Applied Statistics*, 43(3):429–467.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Sauerbrei, W., Royston, P., Bojar, H., Schmoor, C., and Schumacher, M. (1999). Modelling the effects of standard prognostic factors in node-positive breast cancer. *British Journal of Cancer*, 79(11-12):1752–1760.

Sauerbrei, W., Royston, P., and Look, M. (2007). A new proposal for multivariable modelling of time-varying effects in survival data based on fractional polynomial time-transformation. *Biometrical Journal*, 49(3):453–473.

Scheipl, F., Staicu, A.-M., and Greven, S. (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics*, 24(2):477–501.

Singer, J. D. and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press, New York.

Sleeper, L. A. and Harrington, D. P. (1990). Regression splines in the cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association*, 85(412):941–949.

Song, X. and Wang, C. Y. (2008). Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. *Biometrics*, 64(2):557–566.

Steck, A. K., Dong, F., Waugh, K., Frohnert, B. I., Yu, L., Norris, J. M., and Rewers, M. J. (2016). Predictors of slow progression to diabetes in children with multiple islet autoantibodies. *Journal of Autoimmunity*, 72(Supplement C):113 – 117.

Steck, A. K., Johnson, K., Barriga, K. J., Miao, D., Yu, L., Hutton, J. C., Eisenbarth, G. S., and Rewers, M. J. (2011). Age of islet autoantibody appearance and mean levels of insulin, but not GAD or IA-2 autoantibodies, predict age of diagnosis of type 1 diabetes: Diabetes Autoimmunity Study in the Young. *Diabetes Care*, 34(6):1397–1399.

Steck, A. K., Vehik, K., Bonifacio, E., Lernmark, Å., Ziegler, A.-G., Hagopian, W. A., She, J., Simell, O., Akolkar, B., Krischer, J., Schatz, D., and Rewers, M. J. (2015). Predictors of progression from the appearance of islet autoantibodies to early childhood diabetes: The Environmental Determinants of Diabetes in the Young (TEDDY). *Diabetes Care*, 38(5):808–813.

Sudell, M., Kolamunnage-Dona, R., and Tudur-Smith, C. (2016). Joint models for longitudinal and time-to-event data: a review of reporting quality with a view to meta-analysis. *BMC Medical Research Methodology*, 16:168.

Sweeting, M. J. and Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5):750–763.

Tang, A.-M. and Tang, N.-S. (2015). Semiparametric Bayesian inference on skew–normal joint modeling of multivariate longitudinal and survival data. *Statistics in Medicine*, 34(5):824–843.

Tang, A.-M., Zhao, X., and Tang, N.-S. (2017). Bayesian variable selection and estimation in semiparametric joint models of multivariate longitudinal and survival data. *Biometrical Journal*, 59(1):57–78.

Taylor, J. M. G., Park, Y., Ankerst, D. P., Proust-Lima, C., Williams, S., Kestin, L., Bae, K., Pickles, T., and Sandler, H. (2013). Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics*, 69(1):206–213.

Telieps, T., Köhler, M., Treise, I., Foertsch, K., Adler, T., Busch, D. H., Hrabě de Angelis, M., Verschoor, A., Adler, K., Bonifacio, E., and Ziegler, A.-G. (2016). Longitudinal frequencies of blood leukocyte subpopulations differ between NOD and NOR mice but do not predict diabetes in NOD mice. *Journal of Diabetes Research*, 2016:e4208156.

The TEDDY Study Group (2007). The Environmental Determinants of Diabetes in the Young (TEDDY) study: study design. *Pediatric Diabetes*, 8(5):286–298.

The TEDDY Study Group (2008). The Environmental Determinants of Diabetes in the Young (TEDDY) study. *Annals of the New York Academy of Sciences*, 1150:1–13.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York.

Tsiatis, A. A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88(2):447–458.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(3):809–834.

Umlauf, N., Klein, N., and Zeileis, A. (2017). BAMLSS: Bayesian Additive Models for Location, Scale and Shape (and beyond). Working Paper, Faculty of Economics and Statistics, University of Innsbruck.

Umlauf, N., Klein, N., Zeileis, A., and Koehler, M. (2016). *bamlss: Bayesian Additive Models for Location Scale and Shape (and Beyond)*. R package version 0.1-1, available at `https://cran.r-project.org/package=bamlss`.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.

Waldmann, E., Taylor-Robinson, D., Klein, N., Kneib, T., Pressler, T., Schmid, M., and Mayr, A. (2017). Boosting joint models for longitudinal and time-to-event data. *Biometrical Journal*. (online first).

Wand, M. P. (2003). Smoothing and mixed models. *Computational Statistics*, 18(2):223–249.

Wand, M. P. and Jones, M. C. (1994). *Kernel smoothing*. Chapman and Hall/CRC, Boca Raton.

Wood, S. N. (2006). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC, Boca Raton.

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., Pya, N., and Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516):1548–1563.

Wu, M. C. and Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics*, 45(3):939–955.

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44(1):175–188.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330–339.

Wynant, W. and Abrahamowicz, M. (2016). Flexible estimation of survival curves conditional on non-linear and time-dependent predictor effects. *Statistics in Medicine*, 35(4):553–565.

Ye, W., Lin, X., and Taylor, J. M. G. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data: A two-stage regression calibration approach. *Biometrics*, 64(4):1238–1246.

Yuen, H. P. and Mackinnon, A. (2016). Performance of joint modelling of time-to-event data with time-dependent predictors: an assessment based on transition to psychosis data. *PeerJ*, 4:e2582.

Zhang, D., Chen, M.-H., Ibrahim, J., Boye, M., and Shen, W. (2016). JMFit: A SAS macro for joint models of longitudinal and survival data. *Journal of Statistical Software*, 71(3):1–24.

Zhang, D., Chen, M.-H., Ibrahim, J. G., Boye, M. E., Wang, P., and Shen, W. (2014). Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials. *Statistics in Medicine*, 33(27):4715–4733.

Ziegler, A. G., Bonifacio, E., and the BABYDIAB-BABYDIET Study Group (2012). Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. *Diabetologia*, 55(7):1937–1943.

Ziegler, A. G., Hillebrand, B., Rabl, W., Mayrhofer, M., Hummel, M., Mollenhauer, U., Vordemann, J., Lenz, A., and Standl, E. (1993). On the appearance of islet associated autoimmunity in offspring of diabetic mothers: a prospective study from birth. *Diabetologia*, 36(5):402–408.

Ziegler, A.-G., Hummel, M., Schenker, M., and Bonifacio, E. (1999). Autoantibody appearance and risk for development of childhood diabetes in offspring of parents with type 1 diabetes: the 2-year analysis of the German BABYDIAB Study. *Diabetes*, 48(3):460–468.

Ziegler, A. G., Rewers, M., Simell, O., Simell, T., Lempainen, J., Steck, A., Winkler, C., Ilonen, J., Veijola, R., Knip, M., Bonifacio, E., and Eisenbarth, G. S. (2013). Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA*, 309(23):2473–2479.

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 13.10.2017 Meike Köhler