Simon Wiegrebe

Statistical Frameworks for Modeling Longitudinal and Time-to-Event Outcomes

With Applications to Epidemiology and Genetics

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

Eingereicht am 05.11.2025

Erster Berichterstatter: Prof. Dr. Helmut Küchenhoff
Zweite Berichterstatterin: Prof. Dr. Iris M. Heid
Dritter Berichterstatter: Prof. Dr. Matthias Schmid

Tag der Disputation: 22.01.2026

Acknowledgments

Throughout my Ph.D. journey, I have had the steady support and companionship of many wonderful people. I would like to express my sincere gratitude to... ...

... *my Ph.D. supervisors Prof. Dr. Helmut Küchenhoff and Prof. Dr. Iris Heid, for the opportunity to do my Ph.D. under your supervision and for your continuous support and assistance throughout. Thank you, Helmut, for these wonderful years at the StaBLab; you have created a place where one can truly grow professionally while always feeling at home. And thank you, Iris, for the amazing time as part of GenEpi Regensburg (even though I missed out on quite a few cakes and bottles of sparkling wine via Zoom...!);*

... *Prof. Dr. Matthias Schmid for his willingness to act as the third reviewer for my Ph.D. thesis;*

... *Prof Dr. Christian Heumann and Prof. Dr. Göran Kauermann for their availability to be part of the examination panel at my Ph.D. defense;*

... *Dr. Andreas Bender for the excellent supervision of and collaboration on so many of my research projects;*

... *my office mates Alex, Max, Henri, Giovanni, and Christina for making office life so enriching and enjoyable, for many delicious Chinese noodle soups, and for making me see that Spezi bottles are the perfect way to save for retirement;*

... *all members of the StaBLab family and all colleagues at the Department of Statistics for the familiar and cooperative atmosphere, great football tournaments, and many more activities;*

... *all members of the GenEpi Regensburg team, in particular Gianina, Mathias, and Thomas, for the great collaboration, the Italian lessons, and for caring about my well-balanced diet at the many conferences;*

... *my family, in particular my parents Justine and Eckard for their unwavering support and my wife Marjory for always being at my side (I know it has not always been easy, mi amor);*

... *and finally, my daughter Sophie, who has made this entire Ph.D. journey so unique and beautiful in her very own way.*

Summary

Time-dependent data, such as longitudinal and time-to-event data, are particularly informative because they enable both between- and within-subject analyses. Yet analyzing this type of data introduces new challenges beyond those inherent in cross-sectional data. While numerous methods exist to model time-dependent data, their application to complex, high-dimensional settings and their combination with machine learning techniques remain underexplored. This dissertation presents statistical frameworks for analyzing longitudinal and time-to-event outcomes, specifically tailored to high-dimensional data and the incorporation of machine learning techniques, with a focus on their applications in epidemiology and genetics.

The first part of this dissertation presents approaches for modeling longitudinal data in genetics, where the predictor space is high-dimensional. For many (disease) traits, progression – that is, trait change – is of primary interest but difficult to investigate based solely on between-subject comparisons from cross-sectional data. The first contributing article identifies linear mixed models (LMMs) as a well-calibrated and scalable statistical method with type I error control and high power for modeling genetic effects on trait change. The article further demonstrates that modeling genetic effects on trait change as interaction with time or age is advantageous compared to directly modeling the effect on previously computed trait change outcomes. This is because trajectories of arbitrary length can be incorporated and effect size estimates are unbiased. LMMs are subsequently used to identify novel genetic variants associated with kidney function decline in a large-scale UK Biobank dataset. The second contributing article shows that, under certain assumptions, genetic-by-age interactions from cross-sectional data can be indicative of genetic associations with longitudinal trait change and proposes a two-stage approach: genome-wide pre-screening for genetic-by-age interaction in (abundant) cross-sectional data, followed by testing identified variants for longitudinal change in (scarce) independent longitudinal data.

The second part of this dissertation focuses on analyzing time-to-event data by integrating machine and deep learning techniques. The third contributing article provides a comprehensive overview of deep learning-based methods for survival analysis according to both deep learning- and survival-specific aspects. The fourth contributing article presents a methodological comparison of different reduction techniques for time-to-event data, which transform survival tasks into standard regression or classification tasks. This allows for the use of a broad variety of estimation techniques, in particular facilitating the use of machine learning algorithms. The fifth contributing article combines these two topics by developing a concrete time-to-event method based on the piecewise exponential additive model (PAM), which is both deep learning- and reduction-based.

The third part of this dissertation revisits the task of modeling longitudinal data, but now from the angle of multi-stage disease histories, which are increasingly being derived from longitudinal data. One example is chronic kidney disease, whose multiple stages are defined by clinically meaningful thresholds of a quantitative trait (estimated glomerular filtration rate). While multi-state models are natural candidates for analyzing such multi-stage disease history data, this type of analysis comes with new challenges: dependent left-truncation, multiple time scales, index event bias, and interval-censoring. The final contributing article shows via simulations how a modeling framework based on multi-state PAMs is capable of addressing most of these challenges. This framework is then applied to model transition probabilities of and genetic variant associations with chronic kidney disease onset and progression, using the same UK Biobank dataset as in the first contributing article.

Zusammenfassung

Zeitabhängige Daten, wie beispielsweise Längsschnitt- und Ereigniszeitdaten, sind besonders aussagekräftig, da sie Analysen sowohl zwischen als auch innerhalb von Individuen ermöglichen. Die Analyse dieser Art von Daten bringt jedoch neue Herausforderungen mit sich, die über diejenigen bei Querschnittsdaten hinausgehen. Zwar gibt es zahlreiche Methoden zur Modellierung zeitabhängiger Daten, doch ihre Anwendung in komplexen, hochdimensionalen Kontexten und ihre Kombination mit Techniken des maschinellen Lernens sind noch wenig erforscht. Diese Dissertation stellt statistische Rahmenwerke für die Analyse von Längsschnitt- und Ereigniszeit-Zielgrößen vor, die speziell auf hochdimensionale Daten und die Einarbeitung von Techniken des maschinellen Lernens zugeschnitten sind, wobei der Schwerpunkt auf ihren Anwendungen in der Epidemiologie und Genetik liegt.

Im ersten Teil dieser Dissertation werden Ansätze zur Modellierung von Längsschnittdaten in der Genetik vorgestellt, wobei der Prädiktorraum hochdimensional ist. Bei vielen (Krankheits-)Merkmalen ist der Verlauf – also die Merkmalsveränderung – von primärem Interesse, der allein anhand von Vergleichen zwischen Individuen aus Querschnittsdaten jedoch schwer zu untersuchen ist. Der erste Beitrag identifiziert lineare gemischte Modelle (linear mixed models; LMMs) als gut kalibrierte, skalierbare statistische Methode mit Typ-I-Fehlerkontrolle und hoher Teststärke für die Modellierung genetischer Effekte auf Merkmalsveränderungen. Der Artikel zeigt außerdem, dass die Modellierung genetischer Effekte auf Merkmalsveränderungen als Interaktion mit Zeit oder Alter, im Vergleich zur direkten Modellierung des Effekts auf zuvor berechnete Zielgrößen der Merkmalsveränderung, vorteilhaft ist. Dies liegt daran, dass Trajektorien beliebiger Länge einbezogen werden können und die Schätzungen der Effektgrößen unverzerrt sind. LMMs werden anschließend verwendet, um neue genetische Varianten zu identifizieren, die mit einer Abnahme der Nierenfunktion in einem groß angelegten Datensatz der UK Biobank assoziiert sind. Der zweite Beitrag zeigt, dass unter bestimmten Annahmen Genetik-Alter-Interaktionen aus Querschnittsdaten auf genetische Assoziationen mit longitudinalen Merkmalsveränderungen hinweisen können, und es wird ein zweistufiger Ansatz vorgeschlagen: genomweites Vorab-Screening auf Genetik-Alter Interaktionen in (reichlich vorhandenen) Querschnittsdaten, gefolgt von Tests identifizierter Varianten auf longitudinale Veränderungen in (begrenzt vorhandenen) unabhängigen Längsschnittdaten.

Der zweite Teil dieser Dissertation konzentriert sich auf die Analyse von Ereigniszeitdaten unter Integration von Machine- und Deep-Learning-Techniken. Der dritte Beitrag bietet einen umfassenden Überblick über Deep-Learning-basierte Methoden für die Ereigniszeitanalyse im Hinblick auf Deep-Learning- sowie Ereigniszeitanalyse-spezifische Aspekte. Der vierte Beitrag präsentiert einen methodischen Vergleich verschiedener Reduktionstechniken für Ereigniszeitdaten, die Ereigniszeitaufgaben in Standard Regressions- oder Klassifikationsaufgaben umwandeln. Dadurch wird der Einsatz einer Vielzahl von Schätzverfahren ermöglicht und insbesondere die Verwendung von Algorithmen des maschinellen Lernens erleichtert. Der fünfte Beitrag kombiniert diese beiden Themen, indem er eine konkrete Ereigniszeitmethode auf der Grundlage des stückweisen exponentiellen additiven Modells (PAM) entwickelt, die sowohl auf Deep Learning als auch auf Reduktion basiert.

Der dritte Teil dieser Dissertation befasst sich erneut mit der Modellierung von Längsschnittdaten, diesmal jedoch unter dem Gesichtspunkt mehrstufiger Krankheitsgeschichten, die zunehmend aus Längsschnittdaten abgeleitet werden. Ein Beispiel hierfür ist die chronische Nierenerkrankung,

deren verschiedene Stadien durch klinisch relevante Schwellenwerte eines quantitativen Merkmals (geschätzte glomeruläre Filtrationsrate) definiert sind. Mehrstadienmodelle sind natürliche Kandidaten für die Analyse solcher mehrstufiger Krankheitsgeschichten-Daten, doch diese Art der Analyse bringt neue Herausforderungen mit sich: abhängige Linkstrunkierung, mehrere Zeitskalen, Indexereignis-Bias und Intervallzensierung. Der letzte Beitrag zeigt anhand von Simulationen, wie ein auf Mehrstadien-PAMs basierender Modellierungsrahmen mit den meisten dieser Herausforderungen umgehen kann. Anschließend wird dieser Modellierungsrahmen angewendet, um Übergangswahrscheinlichkeiten und Assoziationen genetischer Varianten für das Einsetzen und Fortschreiten chronischer Nierenerkrankungen auf demselben UK Biobank Datensatz zu modellieren, der auch für den ersten Beitrag verwendet wurde.

Contents

Part I.

Introduction and Background

1. Introduction

1.1. Outline

This dissertation addresses the analysis of time-dependent data, in particular longitudinal and time-to-event data, with motivation from and application to epidemiology and genetics.

It contains six contributing articles: Two contributions focus on modeling trait changes and their genetic determinants in longitudinal data, specifically in the context of genome-wide association studies. Three contributions focus on modeling time-to-event data, where one article provides an overview of deep learning-based survival methods, one article presents a framework of different reduction techniques for survival analysis, and one article introduces a concrete time-to-event method based on piecewise-exponential additive models (PAMs), which incorporates both deep learning and reduction techniques. The last contribution investigates how a PAM-based multi-state modeling framework can be applied to multi-stage disease histories derived from longitudinal data and how it tackles the statistical challenges that inevitably come with this type of data.

This dissertation is organized as follows. Chapter I is an introductory chapter. This Section 1 provides an overview and motivation of the topics addressed by this dissertation. Sections 2, 3, and 4 give context and statistical background on these topics: Section 2 discusses the modeling of longitudinal data, first in general and then specifically in the context of genome-wide association studies. Section 3 presents key concepts of survival analysis as well as its generalization to multi-state models, and introduces reduction techniques. Section 4 provides background on machine learning and deep learning. Chapters II (Sections 5 and 6), III (Sections 7, 8, and 9) and IV (Section 10) contain the contributing articles along with detailed author contributions and links to supplementary material.

1.2. Motivation and Scope

The fields of epidemiology and genetics, among others, are currently experiencing a vast expansion of data availability. This is a consequence of the growing adoption of registry data (Slawomirski et al., 2023; Rau et al., 2024) and the increasing popularity of biobanks (Sudlow et al., 2015; All of Us Research Program Investigators, 2019; Greiser et al., 2023) but also of the vast amount of data collected since the onset of the COVID-19 pandemic (Lammi et al., 2025). Much of this newly generated data is time-dependent, because interest typically lies in longitudinal outcomes (e.g., evolution of biomarkers over time; Gorski et al., 2022; Peterhoff et al., 2023) or in time-to-event outcomes (e.g., time until disease onset; Hagar et al., 2014; Coens et al., 2024). Analyzing such data poses new challenges, such as high dimensionality and intra-subject dependency. In addition, accommodating machine learning techniques – especially in the time-to-event setting – is often not straightforward. This dissertation is motivated by these challenges.

Chapter II of this dissertation focuses on genome-wide association studies (GWASs) of trait change using longitudinal data. This is of high clinical relevance, as the rate of change in biomarkers is often directly linked to disease progression. Statistical approaches for analyzing longitudinal data, such as mixed models, are well established (Laird and Ware, 1982; Cheng et al., 2010; Fahrmeir et al., 2022). Yet, their application in such high-dimensional settings – here due to millions of genetic variants – necessitates scalability and resource efficiency, owing to the large computational cost of accounting for intra-subject dependency structures. In addition, the abovementioned variety of data sources often makes the longitudinal datasets highly unbalanced; in the UK Biobank (Sudlow et al., 2015), for instance, most subjects only have a single measurement of a given trait from their study center assessment, while for some (usually less healthy) individuals trajectories of hundreds of measurements are available after incorporating electronic health records (see, e.g., Gorski et al., 2025). Statistical methods for the analysis of genetic effects on trait change must take into account intra-subject correlation, scalability, and the potential unbalancedness of longitudinal datasets – while also being powerful and well-calibrated and providing type I error control. These considerations motivate the contributions presented in Chapter II. The article in Section 5, Wiegrebe et al. (2024a), compares multiple approaches for modeling genetic effects on trait change via simulations, followed by a GWAS on kidney function decline in UK Biobank data using linear mixed models. The article in Section 6, Winkler et al. (2024), develops a two-stage approach combining GWAS in abundant cross-sectional data with validation in sparse longitudinal data, subsequently applying it to multiple phenotypes in the UK Biobank.

Chapter III of this dissertation investigates how survival analysis methods, which are used to analyze time-to-event data, can incorporate machine learning and deep learning techniques. The advancement of machine and deep learning since the beginning of this century have had a profound impact on almost all scientific fields. The wealth of deep learning-based survival analysis methods being developed in recent years provides the motivation for the work in Section 7: The article Wiegrebe et al. (2024b) offers an overview of the rapidly growing field of deep learning-based survival analysis methods in terms of both deep learning- and survival-related aspects, catering to researchers and practitioners from deep learning and survival analysis alike. The article in Section 8, Piller et al. (2025), is motivated by the fact that the idiosyncrasies of survival analysis tasks – in particular, the presence of censoring – prevent (standard and novel) machine learning techniques from directly being applied. The article provides a framework of different reduction techniques which transform (complex) survival tasks into standard regression or classification tasks via specific data transformations; subsequently, any machine learning algorithm can be employed on these transformed tasks. Finally, the article in Section 9, Kopper et al. (2022), develops a novel survival analysis method, which is both reduction- and DL-based, using piecewise exponential additive models (PAMs).

Chapter IV seeks to combine longitudinal data modeling (Chapter II) and survival analysis (Chapter III) by framing longitudinal data analysis as a multi-state modeling problem. The article in Section 10, Wiegrebe et al. (2025), is motivated by the increasing availability of multi-stage disease histories, derived from longitudinal data, as a result of the expansion of registry and biobank data discussed above. The article discusses statistical challenges that necessarily arise when analyzing multi-stage disease histories using multi-state models, subsequently demonstrating how a PAM-based multi-state framework can address most of these challenges on simulated data and on the UK Biobank kidney function dataset from Section 5.

2. Modeling Longitudinal Data

Longitudinal datasets, characterized by repeated observations of subjects over time, typically stem from cohort studies with predetermined assessment times, from registry data with subject-specific visit histories, or from a mix of both (Diggle, 2002; Cheung et al., 2017; Fitzmaurice et al., 2012; Hyun et al., 2017; Lokku et al., 2020; Gorski et al., 2025). Such datasets allow not only for between-subject comparisons but also for the investigation of subject-specific trajectories, for example in terms of trajectory changes or variability. At the same time, longitudinal data is more complex due to its additional intra-subject correlation structure. The key focus of statistical methods for analyzing longitudinal data is modeling this structure.

In the following, we index subjects by $i \in \{1, \dots, n\}$ and the measurements for subject i by $t \in \{1, \dots, n_i\}$, implying that the longitudinal dataset contains n subjects with a total of $m = \sum_{i=1}^{n} n_i$ measurements. Longitudinal data for subject i thus consists of outcomes $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,t}, \dots, y_{i,n_i})^\top$ and covariates $(\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,t}, \dots, \mathbf{x}_{i,n_i})$, where $\mathbf{x}_{i,t} = (1, x_{i,t,1}, \dots, x_{i,t,p})^\top$ is the covariate (column) vector of subject i at time t and p is the number of covariates. Due to the nature of the outcomes considered in the contributing articles of this dissertation (see Sections 5, 6 and 10), here we explicitly discuss quantitative (i.e., continuous) outcomes $y_{i,t}$.

Likely the most widely used statistical approach for modeling longitudinal data (and correlated data more generally) with a quantitative outcome variable is the linear mixed model (LMM; Laird and Ware, 1982; Cheng et al., 2010; Fahrmeir et al., 2022); for marginal models based on generalized estimating equations, see, e.g., Liang and Zeger (1986) or Hardin and Hilbe (2002). LMMs extend standard linear regression models by incorporating subject-specific deviations from global parameters (e.g., intercept or slope) via random effects. The combination of global and random effects allows for modeling both the population mean and the heterogeneity of individual trajectories around that mean. LMMs can naturally handle unbalanced data, for instance in terms of varying trajectory lengths and irregular measurement times – a common feature of biobank or registry data (UK Biobank, 2023; Garrett et al., 2024). Compared to explicitly estimating person-specific effects as fixed effects, which is highly parameter-intensive, random effects models are much more parsimonious because random effects are assumed to be independently and identically distributed (i.i.d.), conventionally according to a normal distribution with mean $\mathbf{0}$ and unknown variance $\mathbf{Q}$, which is to be estimated. The model equation for an LMM is

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \mathbf{u}_{i,t}\boldsymbol{\gamma}_i + \varepsilon_{i,t}, \tag{2.1}$$

where $(\mathbf{u}_{i,1}, \dots, \mathbf{u}_{i,t}, \dots, \mathbf{u}_{i,n_i})$ are the design vectors for the random effects, with $\mathbf{u}_{i,t} = (1, u_{i,t,1}, \dots, u_{i,t,q})^\top$ and $q \leq p$. For example, setting $\mathbf{u}_{i,t} = 1$ in Equation (2.1) characterizes a random intercept model, while $\mathbf{u}_{i,t} = (1, time_{i,t})$ represents a random slope model with slopes varying over (covariate) *time*. $\boldsymbol{\gamma}_i$ is the vector of random effects, which are assumed to be distributed according to $\boldsymbol{\gamma}_i \sim N(\mathbf{0}, \mathbf{Q})$, with unknown $(q+1) \times (q+1)$-covariance matrix $\mathbf{Q}$ (Fahrmeir et al., 2022). The errors $\varepsilon_{i,t}$ are usually assumed to be i.i.d. $N(0, \sigma^2)$. This implies

a multivariate normal distribution of the subject-specific errors, $\boldsymbol{\varepsilon}_i := (\varepsilon_{i,1}, \ldots, \varepsilon_{i,t}, \ldots, \varepsilon_{i,n_i})^\top \sim N(\mathbf{0}, \mathbf{R}_i)$, $\forall i$, with $\mathbf{R}_i = \sigma^2 \mathbf{I}_{n_i}$; inter-subject independence between errors, $\boldsymbol{\varepsilon}_i \perp\!\!\!\perp \boldsymbol{\varepsilon}_j$, $\forall i \neq j$; as well as independence between errors and random effects, $\boldsymbol{\varepsilon}_i \perp\!\!\!\perp \boldsymbol{\gamma}_j$, $\forall i, \forall j$.

The main challenge of estimating LMMs lies in the estimation of the unknown parameters $\boldsymbol{\vartheta}$ that characterize the variance-covariance structures of the random effects $\boldsymbol{\gamma}_i$ and the error terms $\boldsymbol{\varepsilon}_i$, that is, $\mathbf{Q}$ and $\mathbf{R}_i$. Since maximum likelihood (ML) estimation of $\boldsymbol{\vartheta}$ is biased, $\boldsymbol{\vartheta}$ is usually estimated via the less biased (and sometimes even unbiased) restricted maximum likelihood (REML) estimator. The REML estimator is based on the restricted (or marginal) log-likelihood $l_R(\vartheta) = \log(\int L(\boldsymbol{\beta}, \boldsymbol{\vartheta}) d\boldsymbol{\beta})$, which integrates out $\boldsymbol{\beta}$ from the likelihood $L(\boldsymbol{\beta}, \boldsymbol{\vartheta})$ (Harville, 1977; Fahrmeir et al., 2022). The REML estimator requires numerical computation through iterative algorithms, such as Newton-Raphson or Fisher scoring. Once $\boldsymbol{\vartheta}$ is estimated, fixed effect estimates $\hat{\boldsymbol{\beta}}$ and concrete predictions of the random effects $\hat{\boldsymbol{\gamma}}_i$ can be derived as the solution to Henderson's mixed model equations (Henderson et al., 1959). $\hat{\boldsymbol{\beta}}$ is estimated by weighted least squares conditional on the variance estimates $\hat{\boldsymbol{\vartheta}}$, while $\hat{\boldsymbol{\gamma}}_i$ is usually estimated as the best linear unbiased predictor (BLUP), which is the conditional expected value $\mathbb{E}[\boldsymbol{\gamma}_i \mid \mathbf{y}_i]$ (Fahrmeir et al., 2022).

The concrete specification of the random effects structure is an additional challenge: Too complex structures often cause singularity issues (e.g., convergence to boundary values or singular fits), typically reflecting model overparametrization relative to the data (Bates et al., 2015a,b). By contrast, overly simplistic structures – in particular, omitting random slopes – can lead to type I error inflation (Barr et al., 2013; Matuschek et al., 2017). As recommended by Barr et al. (2013), a prudent approach is to start with the most complex random effects structure still justified by the study design; this structure can subsequently be simplified if necessary (e.g., in case of singularity issues) via iterative model reduction techniques (Bates et al., 2015a; Matuschek et al., 2017).

LMMs can be extended to linear additive mixed models, which also allow for non-linear effect estimation. Upon inclusion of K covariates $(z_{i,t,1}, \ldots, z_{i,t,k}, \ldots, z_{i,t,K})$ with non-linear effects, the model equation becomes

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \sum_{k=1}^{K} f_k(z_{i,t,k}) + \mathbf{u}_{i,t}\boldsymbol{\gamma}_i + \varepsilon_{i,t}, \tag{2.2}$$

where $f_k()$ is a smooth function modeling the effect of covariate $z_{i,t,k}$. For estimation, these smooth functions are parametrized as

$$f_k() = \sum_{m=1}^{M_k} \delta_{m,k} B_{m,k}(z_{i,t,k}), \tag{2.3}$$

with $B_{m,k}$ being suitable basis functions (e.g., B-splines (Eilers and Marx, 1996) or thin-plate splines (Wood, 2003)), $\delta_{m,k}$ the corresponding basis coefficients, and M_k the basis dimension of the k-th smooth. The basis expansion representation of smooth functions in Equation (2.3) implies that the model in Equation (2.2) remains linear in the coefficients and can be estimated accordingly.

In order to address the arbitrariness of the choice of M_k and to avoid overfitting, smooths are usually estimated via penalized splines (P-splines). This approach implies choosing M_k to be relatively large so as to ensure sufficiently flexible modeling of the non-linear effect, but simultaneously introducing an additional, usually quadratic penalty term to prevent overfitting (Bender, 2018; Fahrmeir et al., 2022). For each smooth function $f_k()$, the corresponding

penalty term, which is then added to the log-likelihood, can be written as $\lambda_k \boldsymbol{\delta}_k^\top \mathbf{S}_k \boldsymbol{\delta}_k$, where $\boldsymbol{\delta}_k = (\delta_{1,k}, \ldots, \delta_{m,k}, \ldots, \delta_{M_k,k})^\top$ and $\mathbf{S}_k$ is a suitable penalty matrix which depends on the chosen type of smoother. A very popular smoother for cubic B-splines is the second-order difference smoother proposed by Eilers and Marx (1996), $\Delta_k^2 \delta_m^k := (\delta_{m,k} - \delta_{m-1,k}) - (\delta_{m-1,k} - \delta_{m-2,k})$, which penalizes the difference between neighboring differences of coefficients. λ_k is the smoothing parameter, a hyperparameter controlling the degree of smoothness of $f_k()$. In the extreme cases, $\lambda_k \to 0$ is equivalent to unpenalized estimation of $f_k()$, while $\lambda_k \to \infty$ implies parameters will be estimated so that $\boldsymbol{\delta}_k^\top \mathbf{S}_k \boldsymbol{\delta}_k \to 0$; for the second-order difference smoother, for example, $\lambda_k \to \infty$ means that $f_k()$ converges to a straight line. The hyperparameters λ_k can be optimized via multiple criteria, such as the UBRE score or (generalized) cross-validation (Wood, 2017). The recommended approach by Wood (2011), however, is the REML criterion: here, this means viewing the basis coefficients $\delta_{m,k}$ as random effects from an assumed prior distribution with mean equal to zero and a variance that is to be estimated. Using the REML criterion, estimation of non-linear effects via P-splines in an additive model is thus analogous to random effects estimation in a mixed model.

2.1. Genome-wide association studies

Genome-wide association studies (GWASs) have established themselves as the standard approach for identifying associations of genetic variants with trait levels (McCarthy et al., 2008; Visscher et al., 2017; Abdellaoui et al., 2023) and, more recently, have also been employed to study trait change (Gorski et al., 2022; Robinson-Cohen et al., 2023; Wiegrebe et al., 2024a) or trait variability (Ko et al., 2022). Common GWAS traits are, e.g., body mass index (e.g., Speliotes et al., 2010; Locke et al., 2015), blood pressure (e.g., CKDGen Consortium et al., 2011; Warren et al., 2017), lipids (e.g., Teslovich et al., 2010; Global Lipids Genetics Consortium, 2013), lung function (e.g., Artigas et al., 2011; Wain et al., 2017), or kidney function (e.g., Köttgen et al., 2009; Stanzick et al., 2021). Genetic variants found to be associated with a trait via GWAS provide immediate insights into the trait's underlying biology and heritability (Uffelmann et al., 2021). Beyond that, GWAS results are also clinically relevant as they contribute to disease risk prediction (Torkamani et al., 2018; Liu et al., 2019; Tam et al., 2019) and help to identify potential novel drug targets (Nelson et al., 2015). The establishment of international research consortia and the use of standardized GWAS evaluation procedures further facilitate the comparison and meta-analysis of GWAS results – and, as a consequence, contribute to their replicability (NCI-NHGRI Working Group on Replication in Association Studies, 2007). Moreover, due to generally small genetic effect sizes of common alleles, meta-analyses are often crucial to obtain the large sample sizes required for the detection of genetic signals (Evangelou and Ioannidis, 2013).

As suggested by its name, a GWAS aims to study associations between genetic variants (usually single-nucleotide polymorphisms, SNPs) and a given trait (e.g., biomarker levels) in a genome-wide manner. Therefore, GWASs do not require strong prior hypotheses regarding potential genes or regions of interest – as opposed to candidate-gene approaches, which consequently suffer from low replicability (NCI-NHGRI Working Group on Replication in Association Studies, 2007). SNPs, the most common type of interhuman genetic variation, are variations of a single nucleotide – adenine (A), guanine (G), cytosine (C), or thymine (T) – at a specific position of the genome. The position of a SNP within the genome thus corresponds to a location of genetic differences between humans. The large majority of SNPs are bi-allelic (Casci, 2010), meaning that each of the two base

pair nucleotides can be one of two alleles (e.g., A or T). As standard GWAS workflows typically focus on bi-allelic SNPs and disregard parent-of-origin effects of heterozygous SNPs (though this is beginning to change; see, e.g., Hofmeister et al., 2022), SNPs can be considered random variables that can take on one of three possible allele combinations (e.g., AA, AT, and TT). SNPs are usually coded numerically, by defining an effect allele (e.g., A) as opposed to the corresponding other allele (T) and then counting the number of effect alleles (i.e., SNP $\in \{0, 1, 2\}$).

The total number of SNPs in the human genome has been cataloged to be 261.9 million based on the Genome Aggregation Database (Karczewski et al., 2020). However, as SNPs exhibit regionally restricted correlation structures (*linkage disequilibrium*), not all SNPs are independent from each other. As a consequence, characterizing the human genome does not require directly assaying every single SNP. Instead, genotyping arrays typically measure between 200.000 and 2 million SNPs (e.g., $\sim$ 800.000 in the UK Biobank; Bycroft et al., 2018), and linkage disequilibrium–based imputation from reference panels yields genotypes at millions of unobserved variants (Visscher et al., 2017). For the purpose of GWASs, only non-rare variants with sufficient imputation quality are considered (e.g., minor allele frequency (MAF) $\geq 0.5\%$ and imputation quality score INFO ≥ 0.6 in Wiegrebe et al. (2024a), yielding 11.3 million SNPs; or MAF $\geq 0.1\%$ and INFO ≥ 0.8 in Winkler et al. (2024), yielding 13.2 million SNPs). For imputed SNPs, the so-called dosage is continuous, i.e., SNP $\in [0, 2]$.

Following imputation and filtering of SNPs, the GWAS performs a genome-wide screening of associations between all available SNPs and the trait of interest. Since SNPs are screened independently, multiple testing is corrected for, typically via Bonferroni correction (Dunn, 1961) which controls the family-wise error rate; a common Bonferroni-corrected significance threshold for GWAS, assuming a nominal significance level of 0.05 and 1 million independent SNPs, is then $5\text{x}10^{-8}$ (Chen et al., 2021). SNPs are said to be *identified* if their corresponding P-value is smaller than this threshold. However, the Bonferroni correction is known to be very conservative and entail low power (Perneger, 1998; Tam et al., 2019). Several authors have proposed to instead control the false discovery rate, which can substantially increase power (Benjamini and Hochberg, 1995; Storey and Tibshirani, 2003), yet this is only rarely done in GWAS workflows (Sham and Purcell, 2014). Therefore, a general limitation of the GWAS approach is that the multiple testing burden (aggravated by the use of the Bonferroni correction), in combination with small genetic effect sizes, necessitate very large sample sizes – which is a particular challenge for rare or complex traits (Tam et al., 2019).

The output of a GWAS is typically a list of SNPs, each with its estimated effect size, standard error, and P-value. This output can then be directly evaluated, meta-analyzed (Willer et al., 2010), used to derive polygenic scores (Dudbridge, 2013), or employed for causal inference (e.g., via Mendelian randomization; Smith and Ebrahim, 2003; Hemani et al., 2018). The P-values of all evaluated SNPs are routinely visualized via Manhattan plots and quantile-quantile plots (see, e.g., Wiegrebe et al., 2024a, Figure 4). Moreover, as SNPs in the same genetic region are often correlated due to linkage disequilibrium, significant SNP associations with the phenotype of interest usually form regional clusters, which is visualized by region plots (see, e.g., Wiegrebe et al., 2024a, Supplementary Figure 12). Therefore, to identify the most likely causal variant(s) among a set of (genome-wide) significant signals within a genetic region and to subsequently determine the functional pathway, a multitude of statistical and biological follow-up analyses need to be conducted, such as statistical fine-mapping and pathway analyses (Gallagher and Chen-Plotkin, 2018; Tam et al., 2019; Uffelmann et al., 2021).

Subsections 2.1.1 and 2.1.2 below provide details on the statistical methods used for GWASs. This is directly connected with the choice of the trait of interest, because investigating trait levels (as done by traditional GWASs) may require a different phenotype or model specification than, e.g., a GWAS investigating trait change.

2.1.1. Statistical methods for cross-sectional GWAS

In this subsection, we only consider cross-sectional data (i.e., $n_i = 1$, $\forall i \in \{1, \dots, n\}$, and $m = n$), and assume the phenotype to be a quantitative trait level. Since GWASs screen the entire genome one variant at a time, the statistical GWAS model for SNP g is a standard linear regression model

$$y_i = \beta_0 + \beta_g \cdot g_i + \boldsymbol{\beta}_{\mathbf{c}}^\top \mathbf{c}_i + \varepsilon_i, \tag{2.4}$$

where $\mathbf{c}$ is a vector of further (control) variables.

The vector $\mathbf{c}$ can contain known and, importantly, broadly available risk factors, mostly either demographic (e.g., age or sex) or clinical (e.g., body mass index, BMI, or diabetes status). In addition, the control variables usually also include a number of principal components due to population stratification in GWAS data. This arises because SNP distributions tend to differ across population subgroups and leads to confounder bias in genetic association estimates β_g if the trait distribution also differs across population subgroups. Accounting for population structure by including principal components into the GWAS model has been shown to adjust for this potential source of confounding bias (Price et al., 2006).

In case of a binary trait, the GWAS model is a logistic regression; see, e.g., Balding (2006) or Uffelmann et al. (2021) for details on binary-trait GWASs. GWASs are usually restricted to either quantitative or binary phenotypes, because computationally efficient and scalable estimation algorithms exist for these settings (e.g., Chang et al., 2015; Mbatchou et al., 2021). Interestingly, while only few specialized GWAS software tools beyond quantitative and binary outcomes exist (e.g., `gwasurvivr` (Rizvi et al., 2019) for survival analysis), the reduction techniques presented in Section 8 offer a practical approach for applying existing GWAS tools, originally designed for quantitative or binary traits, to time-to-event phenotypes.

The association of g with y can then be tested for via standard statistical tests (e.g., Wald tests or likelihood ratio tests; Günther, 2021). From Equation (2.4) it follows that, whenever the phenotype is a trait level y, β_g is the *marginal* genetic effect of SNP g on the trait level, which is the association of interest. In recent years, however, and building upon the success of a myriad of standard GWASs (Visscher et al., 2017; Buniello et al., 2019), many recent GWAS workflows have moved beyond merely estimating marginal SNP effects. An important extension of the GWAS model is the genome-wide interaction study (GWIS) model (see, e.g., Thomas, 2010), which investigates the interaction between g and an interaction variable e:

$$y_i = \beta_0 + \beta_g \cdot g_i + \beta_e \cdot e_i + \beta_{g\times e} \cdot g_i \cdot e_i + \boldsymbol{\beta}_{\mathbf{c}}^\top \mathbf{c}_i + \varepsilon_i. \tag{2.5}$$

The variable e can, for example, be an environmental or lifestyle factor (e.g., smoking; Bentley et al., 2019), a demographic factor (e.g., sex; Liu et al., 2012), or a clinical risk factor (e.g., BMI; Manning et al., 2012). The GWIS model thus contains two genetic effects: a main effect β_g and an interaction effect $\beta_{g\times e}$. Power for GWISs is typically lower than for marginal GWASs, due to

the inherent difficulty of estimating interaction effects and because genetic interaction effects tend to be even smaller than marginal effects (McClelland and Judd, 1993; Aschard, 2016).

In the next subsection, we introduce GWAS approaches based on longitudinal data, in particular for identifying SNP associations with trait change.

2.1.2. Statistical methods for longitudinal GWAS

We now consider longitudinal GWAS (longGWAS) data in combination with a quantitative trait. In this setting, two additional considerations arise: First, the time scale must be chosen, usually chronological age or time since the beginning of the study; indeed, the problem of choosing the most appropriate time scale is not specific to GWASs or mixed models (cf. contributing article Wiegrebe et al. (2025) in Section 10). Here, we simply refer to the chosen time scale as *time*. Second, the linearity of trait trajectories over *time* must be explored (e.g., via exploratory additive modeling). In case of substantial non-linearities, transformations of the trait variable can be applied, as done routinely in many GWAS papers; see, e.g., Robinson-Cohen et al. (2023).

A longGWAS LMM with random intercepts and random slopes for estimating marginal SNP effects follows directly from Equations (2.1) and (2.4):

$$y_{i,t} = \beta_0 + \beta_{time} \cdot time_{i,t} + \beta_g \cdot g_i + \boldsymbol{\beta}_{\mathbf{c}}^\top \mathbf{c}_{i,t} + \gamma_{0,i} + \gamma_{1,i} \cdot time_{i,t} + \varepsilon_{i,t}. \tag{2.6}$$

Here, γ_0 denotes the random intercept and γ_1 the random slope varying over *time*; furthermore, the control variables $\mathbf{c}$ can now also vary over time. This model can be directly extended to also include a genetic interaction term (cf. Equation (2.5)):

$$y_{i,t} = \beta_0 + \beta_{time} \cdot time_{i,t} + \beta_g \cdot g_i + \beta_{g\times time} \cdot g_i \cdot time_{i,t} + \boldsymbol{\beta}_{\mathbf{c}}^\top \mathbf{c}_{i,t} + \gamma_{0,i} + \gamma_{1,i} \cdot time_{i,t} + \varepsilon_{i,t}. \tag{2.7}$$

In these two models (Equations (2.6) and (2.7)), the usage of longitudinal data increases power by increasing the sample size from n to m. However, the random effects structure must be estimated for every single SNP anew. This makes a genome-wide screening with millions of SNPs computationally infeasible, even when using efficient mixed model software such as the R packages `lme4` (Bates et al., 2015c) or `mgcv` (Wood, 2017).

One way to substantially reduce the computational burden of such longGWAS models consists in, first, estimating a null model, which is identical to Equation (2.6) but excluding the SNP effect:

$$y_{i,t} = \beta_0 + \beta_{time} \cdot time_{i,t} + \boldsymbol{\beta}_{\mathbf{c}}^\top \mathbf{c}_{i,t} + \gamma_{0,i} + \gamma_{1,i} \cdot time_{i,t} + \varepsilon_{i,t}. \tag{2.8}$$

Subsequently, score vectors are constructed from the null model in Equation (2.8), which are then used along with matrix projection approaches to approximate test statistics (Chen et al., 2016; Wang et al., 2020). This approach is implemented in the R packages `GMMAT` (Chen et al., 2023) and `MAGEE` (Wang et al., 2025).

Longitudinal data does not only increase power, but specifically enables the investigation of trait *change*. Within the GWAS literature, trait change has been mostly studied via two-stage approaches: In the first stage, an explicit trait change phenotype Δy is derived. Two common

approaches to construct Δy are: (i) a direct difference approach, where Δy is the average trait change during the observation period per unit of *time* (see, e.g., Gorski et al., 2022); or (ii) an LMM-based approach, which uses as Δy the random slopes of the null model in Equation (2.8), estimated via BLUPs (see, e.g., Robinson-Cohen et al., 2023). In the second stage, Δy is used as the outcome within a standard cross-sectional GWAS model (cf. Equation (2.4)):

$$\Delta y_i = \beta_0 + \beta_g \cdot g_i + \boldsymbol{\beta}_{\mathbf{c}}^{\top} \mathbf{c}_i + \varepsilon_i. \tag{2.9}$$

Two-stage approaches are computationally very efficient, because the construction of trait change phenotypes reduces the longitudinal data to cross-sectional data (see Equation (2.9)), for which specialized GWAS software is readily available (see Section 2.1.1). However, these approaches also have evident drawbacks, such as their inability to incorporate trajectories of arbitrary lengths or biased effect size estimates stemming from the inherent L_2-regularization bias of random slopes (Gelman and Hill, 2007).

To address these drawbacks, the longGWAS LMM with genetic interaction from Equation (2.7) can be exploited, by using the fact that a SNP-*time* interaction represents the change in the genetic effect on the trait level per 1-unit change in *time*. This way, the SNP effect on trait *change* can be directly modeled via a SNP-*time* interaction within a one-stage LMM-based approach. The contributing article Wiegrebe et al. (2024a) (Section 5) implements this one-stage LMM-based approach by employing the abovementioned software `GMMAT/MAGEE`, which was originally developed for gene-environment interaction analysis, in order to conduct a longGWAS on kidney function decline. The contributing article (Winkler et al., 2024) (Section 6) further investigates the connection and combinability of interaction-based and two-stage modeling of trait change, in particular when SNP-by-*time* interactions can be estimated from abundant cross-sectional data whereas longitudinal data is scarce.

Finally, longitudinal data can also be used to conduct a GWAS on trait *variability*. For instance, using a generalized additive model for location, scale and shape (GAMLSS; Rigby and Stasinopoulos, 2005), SNP effects on both trait levels and trait variance can be estimated (Ko et al., 2022).

3. Survival Analysis

Survival analysis denotes a branch of statistics concerned with (partially) censored and/or truncated time-to-event outcomes T, e.g., death or disease onset. Assuming continuous T and letting $f(\tau)$ and $F(\tau) := P(T \leq \tau)$ be density and cumulative distribution function, respectively, the survival function of T is defined as

$$S(\tau) := P(T > \tau) = 1 - F(\tau), \tag{3.1}$$

which is the probability of surviving beyond some time point τ.

The hazard rate

$$h(\tau) := \lim_{\Delta \searrow 0} \frac{P(\tau < T \leq \tau + \Delta | T \geq \tau)}{\Delta} \tag{3.2}$$

represents the instantaneous risk of observing an event, given the event has not yet occurred at time τ. Integrating the hazard rate from 0 to τ yields the cumulative hazard

$$H(\tau) := \int_0^\tau h(u)du = -\log(S(\tau)). \tag{3.3}$$

The survival function, the hazard rate, and the cumulative hazard all constitute common quantities of interest in survival analysis.

In the case of discrete time $\tilde{T} \in \{1, \ldots, j, \ldots, J\}$, the discrete-time hazard

$$h_d(j) = P(\tilde{T} = j \mid \tilde{T} \geq j) \tag{3.4}$$

is the probability of event occurrence during the j-th interval, conditional on survival until the beginning of that interval. The discrete-time survival function is

$$S_d(j) = P(\tilde{T} > j) = \prod_{l=1}^{j} (1 - h_d(l)). \tag{3.5}$$

Discrete-time survival analysis is naturally useful whenever event times are *intrinsically* discrete (Tutz et al., 2016), but can also be used to approximate continuous-time distributions after event time discretization (cf. Section 3.2): In this case, the follow-up time is partitioned into J intervals $(a_0 = 0, a_1], \ldots, (a_{j-1}, a_j], \ldots, (a_{J-1}, a_J]$, where I_j denotes the j-th interval, i.e., $(a_{j-1}, a_j]$, and $P(T \in I_j) \Leftrightarrow P(\tilde{T} = j)$. The remainder of this section mostly deals with continuous-time survival analysis, which is why event times are referred to by T (as opposed to $\tilde{T}$).

The most common survival setting, which also gave the field its name, contains the two states 0 and 1 (e.g., alive and dead), where only the transition $0 \rightarrow 1$ is possible, and only once (i.e., state 1 is absorbing). In this standard survival setting, we denote the right-censoring time of subject

$i \in \{1, \ldots, n\}$ by C_i, the observed time by $Y_i := \min(T_i, C_i)$ with realization y_i, and the right-censoring status indicator by $d_i := I(T_i \leq C_i)$. We further assume random censoring conditional on covariates $\mathbf{x}_i$ ($\forall i : T_i \perp\!\!\!\perp C_i \mid \mathbf{x}_i$). The resulting right-censored, single-risk, single-event survival data is typically represented via tuples $(y_i, d_i, \mathbf{x}_i)$. However, the field of survival analysis is by no means restricted to this standard survival settings.

For subject i, let C_i^L and C_i^R denote left- and right-censoring times, and L_i and R_i censoring interval endpoints. Interval-censoring implies $T_i \in (L_i, R_i]$, as only the event *interval* is known. Both right-censoring $T_i \in (L_i = C_i^R, \infty)$ and left-censoring $T_i \in (L_i = 0, R_i = C_i^L)$ represent special cases of interval-censoring. Truncation means that some subjects are excluded from the risk set for a specific event at certain time points (or even from the entire dataset). With T_i^L and T_i^R denoting left- and right-truncation times, left-truncation implies that subjects with $T_i < T_i^L$ are excluded from the study (while $T_i^R = \infty$). Analogously, right-truncation occurs when subjects with $T_i > T_i^R$ are excluded. In general, the objective of survival analysis is to estimate the distribution of T_i, given a p-dimensional covariate vector $\mathbf{x}_i$ and taking into account relevant censoring and truncation present in the data.

Expansions of survival analysis techniques beyond single-risk, single-event scenarios include *recurrent-event*, *competing-risk*, and, most generally, *multi-state* analysis (see Figure 2 of Piller et al., 2025). *Recurrent-event* analysis describes a setting where a single, non-terminal event (such as epilepsy or malaria infections) may be experienced repeatedly over time, inducing intra-person correlation. *Competing-risk* analysis, on the other hand, assumes a single event occurrence along with $k \in \{1, \ldots, q\}$ distinct, mutually exclusive risks (e.g., death in hospital versus hospital discharge). A common quantity of interest in competing-risk analysis is the cumulative incidence function (CIF) for risk k,

$$CIF_k(\tau) = P(T \leq \tau, K = k) = \int_0^\tau h_k(u) S(u)\, du, \tag{3.6}$$

with cause-specific transition hazard

$$h_k(\tau) = \lim_{\Delta \searrow 0} \frac{P(\tau \leq T \leq \tau + \Delta, K = k | T \geq \tau)}{\Delta} \tag{3.7}$$

and $S(u)$ being the all-cause survival probability derived using Equation (3.1) and the all-cause hazard $h(\tau) = \sum_{k=1}^q h_k(\tau)$ (Piller et al., 2025).

Most approaches for estimating the quantities of interest in survival analysis have been developed for the standard survival setting. Common approaches include the non-parametric Kaplan-Meier (KM) estimator for the survival function (Kaplan and Meier, 1958); the non-parametric Nelson-Aalen (NA) estimator for the cumulative hazard (Equation (3.3); Aalen, 1978) along with the Breslow estimator (Breslow, 1974) to derive the survival function (Equation (3.1)) from the cumulative hazards; the semi-parametric Cox Proportional Hazards (PH) model (Cox, 1972), the piecewise exponential model (PEM; Friedman, 1982), and discrete-time models (see, e.g., Tutz et al., 2016), all of them estimating the hazard (Equations (3.2) or (3.4)); and the fully parametric accelerated failure time (AFT) models (Kalbfleisch and Prentice, 2002), which estimate the entire event time distribution (that is, the parameters of, e.g., a Weibull or log-normal distribution). Recurrent-event data can be modeled by including frailty terms, which are essentially subject-level random effects (cf. Section 2), into the linear predictor of a survival model and/or by stratifying the

linear predictor by the event count. For the modeling of competing-risk data, various approaches have been introduced, such as the Fine-Gray model (Fine and Gray, 1999) for direct estimation of the CIF (Equation (3.6)) or cause-specific Cox PH models for estimation of cause-specific hazards (Equation (3.7)). The standard survival setting, recurrent events, and competing risks can all be viewed as special cases of the multi-state setting. This setting, along with multi-state modeling, is described in detail below.

3.1. Multi-state modeling

Multi-state settings allow for multiple states, (back- and forth-) transitions between them, as well as event recurrence via multiple episodes. For instance, a recurrent-event analysis with an additional terminal event or a competing-risk setting with further progression after transitioning into some intermediate state can be directly cast as multi-state problems. Multi-state problems are commonly represented via state diagrams, which visualize all possible transitions $o \to \ell$ (where o and ℓ are the from- and to-state, respectively), indexed by $k = 1, \ldots, q$; see, e.g., Figures 1 and 3 in Wiegrebe et al. (2025). Within a state diagram, transient states (out of which transitions to other states are possible) are typically indicated by circles, absorbing states (e.g., death) by squares.

For multi-state modeling, we extend the single-risk, single-event hazard rate from Equation (3.2) to transition-specific hazards

$$h_{k,e}(\tau \mid \mathbf{x}_{i,k,e}) := \lim_{\Delta \searrow 0} \frac{1}{\Delta} P(\tau \leq T < \tau + \Delta, o, e \mid T \geq \tau, \ell, \mathbf{x}_{i,k,e}), \tag{3.8}$$

where $\mathbf{x}_{i,k,e}$ is the subject-, transition-, and episode-specific covariate vector. The risk-specific hazard from the competing-risk setting in Equation (3.7) can be obtained from Equation (3.8) by dropping episode e (single-event), setting from-state $l = 0$ (as all subjects start in the same initial state 0), and letting k be the transition from 0 to what was denoted as risk (i.e., state) k in Equation (3.7).

Multi-state data stores transition information via tuples of the form $(y^{entry}_{i,k,e}, y^{exit}_{i,k,e}, d_{i,k,e}, \mathbf{x}_{i,k,e})$, where $y^{entry}_{i,k,e}$ and $y^{exit}_{i,k,e}$ are risk set entry and exit times of subject i for transition $k = 1, \ldots, q$ in episode $e = 1, \ldots, m$, due to either transitioning or censoring; $d_{i,k,e}$ is the subject-, transition-, and episode-specific binary status indicator. This data structure enables the calculation of the transition probability matrix (Aalen and Johansen, 1978; Beyersmann et al., 2011), whose (ℓ, o)-th element represents the probability of transition k. The dependence of transitions at time τ on past transitions in fact requires the computation of the matrix product across prior transition probability matrices. Transition probabilities may be time-inhomogeneous, hence we define the instantaneous probability $dP_{k,e}(\tau) = (\mathbf{dP}_e(\tau))_{\ell,o} = P(\tau \leq T \leq \tau + \Delta, o, e \mid T \geq \tau, \ell), \ell \neq o$. By defining the probability of staying in state ℓ as $1 + (\mathbf{dP}_e(\tau))_{\ell,\ell} = 1 - \sum_k dP_{k,e}(\tau)$, we can denote the transition probability matrix as

$$\mathbf{P}_e(s, \tau) = \prod_{u \in [s,\tau)} \left(\mathbf{I} + \mathbf{dP}_e(u)\right), \tag{3.9}$$

with $\prod_{u\in[s,\tau)}$ being the product integral over the interval $[s, \tau)$ (Piller et al., 2024, 2025; Wiegrebe et al., 2025).

Transition probabilities $dP_{k,e}(\tau)$ are commonly derived using the Aalen-Johansen (AJ) estimator (Aalen and Johansen, 1978), which employs count processes and has also been shown to work in non-Markov scenarios under the assumption of random censoring (Nießl et al., 2023). Another approach consists in approximating $dP_{k,e}(\tau)$ by $H_{k,e}(\tau + \Delta \mid \mathbf{x}_{i,k,e}) - H_{k,e}(\tau \mid \mathbf{x}_{i,k,e})$, that is, by the incremental of the transition-specific cumulative hazard $H_{k,e}(\tau \mid \mathbf{x}_{i,k,e}) = \int_0^\tau h_{k,e}(u \mid \mathbf{x}_{i,k,e})du$. With this approach, it suffices to estimate transition-specific hazards $h_{k,e}(\tau \mid \mathbf{x}_{i,k,e})$ (Equation (3.8)) in order to compute the transition probability matrix. This approach further allows flexible specification of covariate-dependent hazards, so that the resulting transition probabilities depend on the underlying covariate structure (Piller et al., 2024). As covariates may also be history-dependent (e.g., multiple time scales or state-entry times), non-Markov scenarios can also be dealt with by Equation (3.9) (Piller et al., 2025; Wiegrebe et al., 2025). In practice, PEMs can be used for the purpose of estimating transition-specific hazards, as we show in the contributing article Piller et al. (2025) (Section 8). The article further shows how discrete-time methods can be used to estimate competing-risk hazards in a discrete-time setting, and Tutz et al. (2016) describe the extension to hazard-based discrete-time multi-state modeling. Finally, there are also deep learning-based approaches that are capable of computing transition probabilities $dP_{k,e}(\tau)$ of a multi-state model, in particular the contributing article Kopper et al. (2022) (Section 9) but also others (e.g., Groha et al., 2020; Cottin et al., 2022). The contributing article Wiegrebe et al. (2024b) (Section 7) provides an overview of deep learning-based survival methods, considering, among other things, their capability of handling competing-risk and multi-state problems.

As a result of the expansion of registry data usage and biobanks, multi-stage disease histories derived from longitudinal data constitute a type of multi-state data that is becoming increasingly available. Examples include Chronic Kidney Disease (CKD), whose disease stages are defined by thresholds of the quantitative biomarker estimated glomerular filtration rate (eGFR; Levin et al., 2013); or age-related macular degeneration (AMD), whose disease stages are characterized by the accumulation of subretinal drusenoid deposits (early/intermediate AMD) and cell atrophy (late AMD). When modeling multi-stage disease history data, the focus lies, e.g., on disease onset and progression (Wiegrebe et al., 2025) or on disease relapse and death (Iacobelli and Carstensen, 2013). Multi-state modeling of such data, while naturally suitable, introduces additional statistical challenges; in particular, dependent left-truncation, multiple time scales, index event bias, and interval-censoring. These challenges, and how a flexible PEM-based multi-state approach can handle most of them, are addressed in the last contributing article, Wiegrebe et al. (2025) (Section 10).

3.2. Reduction techniques

Direct adoption of machine learning techniques to the field of survival analysis is hampered – or, at the least, delayed – by the fact that most standard machine learning algorithms – e.g., Random Forests (Breiman, 2001) or XGBoost (Chen and Guestrin, 2016) – are designed for regression or classification tasks, not for censored and/or truncated time-to-event data. Indeed, maximizing the Cox partial likelihood is more complex than maximizing the likelihood of a standard generalized linear model (e.g., with Gaussian, Bernoulli, or Poisson distributional assumption) because the

former one requires iterative recomputation of risk sets and tie handling is computationally expensive (Cox, 1972; Breslow, 1974; Efron, 1977). Currently, for the field of survival analysis, survival-specific adoptions have been developed – e.g., Random Survival Forests (Ishwaran et al., 2008) or XGBoost for AFT models (Barnwal et al., 2022) – but with delays of multiple years; see Figure 1 of Piller et al. (2025).

The concept of reduction techniques originally derives from the field of computer science, describing approaches that transform a problem into a simpler one for which a solution is already established and then inferring the solution to the original problem from the solution of the reduced problem (Armoni, 2009). In the context of survival analysis, reduction techniques reduce the complexity of the underlying estimation problem, as a complicated survival task is transformed into a standard regression or classification task for which machine learning algorithms are readily available. We note, however, that these reduction techniques for survival analysis do not reduce the size of the underlying survival data; in fact, for some reduction techniques (e.g., the PEM reduction; see below) the dataset size is even substantially increased as part of necessary data preprocessing.

The PEM (Friedman, 1982) is one of the most popular reduction techniques. It starts from a general proportional hazards model

$$h(\tau \mid \mathbf{x}_i) = h_0(\tau) \cdot \exp(\mathbf{x}_i^\top \boldsymbol{\beta}), \tag{3.10}$$

with baseline hazard $h_0(\tau)$ and linear predictor $\eta_i := \mathbf{x}_i^\top \boldsymbol{\beta}$. Whereas the Cox PH model (Cox, 1972) estimates $h_0(\tau)$ in Equation (3.10) non-parametrically and $\boldsymbol{\beta}$ via partial likelihood maximization, the PEM is fully parametric in terms of parameter estimation. This is possible due to a partition of the follow-up into J intervals, identical to the discretization for discrete-time survival analysis described at the beginning of Section 3. On these intervals, the PEM assumes hazards to be piecewise constant, $h(\tau \mid \mathbf{x}_i) = \exp(\beta_{0,j} + \mathbf{x}_i^\top \boldsymbol{\beta}) = \exp(\eta_{i,j}) =: h_{i,j}, \forall \tau \in (a_{j-1}, a_j]$. Next, the PEM reduction defines (i) an interval-specific event indicator $d_{i,j}$, which is equal to 1 if $y_i \in I_j \wedge d_i = 1$, and 0 otherwise; (ii) the time-at-risk $y_{i,j}$, which is equal to $a_j - a_{j-1}$ if $a_j < y_i$, and $y_i - a_{j-1}$ if $a_{j-1} < y_i \leq a_j$; and (iii) an offset $o_{i,j} = \log(y_{i,j})$. This newly defined pseudo-data thus represents a long-form dataset containing one row per subject and interval-at-risk $j \in \{1, \ldots, J_i\}$, where $J_i \in \{1, \ldots, J\}$ is the interval containing y_i; see Table 2 of Piller et al. (2025). With this data transformation, the piecewise constant hazards assumption, and $\boldsymbol{\beta}_0 = (\beta_{0,1}, \ldots, \beta_{0,J})$, the log-likelihood can be rewritten as

$$\begin{aligned}
l(\boldsymbol{\beta}, \boldsymbol{\beta}_0) &= \log\left(\prod_{i=1}^{n} h(y_i \mid \mathbf{x}_i)^{d_i} S(y_i \mid \mathbf{x}_i)\right) \\
&= \log\left(\prod_{i=1}^{n} h(y_i \mid \mathbf{x}_i)^{d_i} \exp\Big(-\int_0^{y_i} h(s \mid \mathbf{x}_i) ds\Big)\right) \\
&= \log\left(\prod_{i=1}^{n} \prod_{j=1}^{J_i} \exp\left(d_{i,j}\eta_{i,j} - \exp(o_{i,j} + \eta_{i,j})\right)\right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{J_i} \left(d_{i,j}\eta_{i,j} - \exp(o_{i,j} + \eta_{i,j})\right),
\end{aligned}$$

using $h(y_i|\mathbf{x}_i)^{d_i} = \prod_{j=1}^{J_i} \exp(d_{i,j}\eta_{i,j})$ and $\int_0^{y_i} h(s|\mathbf{x}_i)ds = \sum_{j=1}^{J_i} y_{i,j} \exp(\eta_{i,j}) = \sum_{j=1}^{J_i} \exp(o_{i,j} + \eta_{i,j})$ (Friedman, 1982; Bender, 2018). The reduction into a standard Poisson regression task then follows by assuming $d_{i,j} \overset{\text{i.i.d.}}{\sim} Po(\mu_{i,j})$ with $\mu_{i,j} = h_{i,j}y_{i,j}$ and density $f(d_{i,j}) = \mu_{i,j}^{d_{i,j}} \cdot \frac{1}{d_{i,j}!} \cdot \exp(-\mu_{i,j})$ and observing

$$\begin{aligned} l_{Po}(\boldsymbol{\beta}, \boldsymbol{\beta}_0) &= \log\left(\prod_{i=1}^{n}\prod_{j=1}^{J_i} f(d_{ij})\right) = \sum_{i=1}^{n}\sum_{j=1}^{J_i}(d_{ij}\log(\mu_{ij}) - \mu_{ij}) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{J_i}(d_{ij}\log(h_{ij}) + d_{ij}\log(y_{ij}) - h_{ij}y_{ij}) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{J_i}(d_{ij}\eta_{ij} - \exp(o_{ij} + \eta_{ij}) + d_{ij}o_{ij}) \propto l(\boldsymbol{\beta}, \boldsymbol{\beta}_0), \end{aligned}$$

since $d_{ij}o_{ij}$ does not depend on the parameters of interest. As a consequence, the parameters $\boldsymbol{\beta}$ of the linear predictor η_i – and, more generally, the parametrization of any predictor $g(\tau, \mathbf{x}_i)$ – can be obtained via optimization of a standard Poisson likelihood (Bender, 2018; Piller et al., 2025).

The discrete-time reduction employs the exact same follow-up partitioning and data transformation as the PEM reduction, but subsequently uses discrete hazards (cf. Equation (3.4)) to predict the binary responses $d_{i,j}$. The likelihood of a discrete-time hazard model is then equivalent to the likelihood of the binary responses d_{ij} from a binary response model $g(P(d_{i,j} = 1 \mid \mathbf{x}_i)) = f(\mathbf{x}_i)$, where $g()$ is a suitable link function, $f()$ is some prediction function, and $d_{i,j} \overset{\text{i.i.d.}}{\sim} Ber(h_d(j \mid \mathbf{x}_i))$ (Tutz et al., 2016). Therefore, analogously to the PEM reduction, the discrete-time reduction enables any (statistical or machine learning) binary classification algorithm to be used for survival analysis (Piller et al., 2025).

While both PEM and discrete-time reductions rely on partitioning the follow-up time and learn the entire event time distribution, other reductions are designed to predict point estimates of a chosen quantity of interest. For instance, the pseudo-value reduction routinely uses the KM estimator to construct continuous pseudo-observations from the underlying survival data, thus enabling their direct estimation via regression (Andersen et al., 2003). This reduction technique is especially useful for estimating (covariate effects on) quantities of interest that are not straightforward to compute using hazard-based approaches – in particular, the increasingly popular restricted mean survival time (RMST; Irwin, 1949; Royston and Parmar, 2011, 2013; Zhao et al., 2016), defined as $\mu_\tau = \mathbb{E}(min(T, \tau)) = \int_0^\tau S(u)du$.

The contributing article Piller et al. (2025) (Section 8) provides a unified framework for multiple such reduction techniques, which share the core idea of a dedicated pre-processing step to reshape the data and redefine the outcome variable to be estimated. The article further categorizes these reduction techniques and details their respective strengths and weaknesses in terms of various idiosyncrasies of survival tasks (e.g., supported types of censoring and truncation or applicability beyond single-risk, single-event settings). In doing so, this unified framework of reduction techniques offers a practical solution to the abovementioned problem of adopting machine learning algorithms to survival analysis.

4. Machine Learning And Deep Learning

This section focuses on machine learning and deep learning in the context of *supervised* learning, which refers to learning an unknown functional relationship $f : \mathcal{X} \to \mathbb{R}^g$ between the feature space $\mathcal{X} \subseteq \mathbb{R}^p$ and the target space $\mathcal{Y} \subseteq \mathbb{R}^g$ (Hastie et al., 2009). Supervised learning algorithms aim to learn this function f from observed data $\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_i, y_i), \ldots, (\mathbf{x}_n, y_n)\}$. An observation $(\mathbf{x}_i, y_i)$, $i \in \{1, \ldots, n\}$, consists of a p-dimensional feature (or covariate) vector $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^\top \in \mathcal{X}$ and the target (or response) $y_i \in \mathcal{Y}$. Each observation is assumed to be drawn i.i.d. from an unknown probability distribution (i.e., data generating process) $\mathcal{P}_{xy}$ on the joint space $\mathcal{X} \times \mathcal{Y}$; that is, $(\mathbf{x}_i, y_i) \overset{\text{i.i.d.}}{\sim} \mathcal{P}_{xy}$. A learning algorithm thus learns an approximation (i.e., prediction model) $\hat{f}$ of f, usually by minimizing the empirical risk $\mathcal{R}_{emp}(f)$, that is,

$$\hat{f} = \arg\min_{f \in \mathcal{H}} \mathcal{R}_{\text{emp}}(f) = \arg\min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(\mathbf{x}_i)). \tag{4.1}$$

The hypothesis space $\mathcal{H}$ is the the set of all possible candidate functions f available to the learning algorithm, while the loss function $L : \mathcal{Y} \times \mathbb{R}^g \to \mathbb{R}$ measures the discrepancy between predictions $f(\mathbf{x}_i)$ and true target values y_i (see below). Learning thus means finding a prediction model $\hat{f}$ that minimizes the average loss function across all observations in $\mathcal{D}$, that is, solving the above optimization problem (Casalicchio, 2019). This supervised learning process is called empirical risk minimization (ERM). The ERM framework provides a unifying principle for machine and deep learning as well as classical statistical learning algorithms. Indeed, maximum likelihood estimation, as routinely done in classical statistics (cf. Section 2), is identical to ERM with the negative log-likelihood as loss function L (Hastie et al., 2009; Murphy, 2012).

The specific learning task is determined by the target space $\mathcal{Y}$, with regression tasks (where the target space is continuous; e.g., $\mathcal{Y} = \mathbb{R}$, implying $g = 1$) and classification tasks (where the target space consists of discrete categories; e.g., $\mathcal{Y} = \{0, 1\}$ for binary classification, implying $g = 1$) being the two most fundamental types of learning tasks (Casalicchio, 2019). As a consequence, most machine learning algorithms are designed to handle such tasks (Bishop and Nasrabadi, 2006; Hastie et al., 2009). Finally, survival tasks are learning problems in which the target variable is the time-to-event T, subject to potential censoring and/or truncation (see Section 3).

For regression, a popular loss function is the L_2-loss, defined as $L_2(y, f(\mathbf{x})) = (y - f(\mathbf{x}))^2$; minimizing the L_2-loss is equivalent to minimizing the negative Gaussian log-likelihood. Other regression losses are the L_1-loss $L_1(y, f(\mathbf{x})) = |y - f(\mathbf{x})|$ and the Poisson loss $L_{Po}(y, f(\mathbf{x})) = f(\mathbf{x}) - y \cdot \log(f(\mathbf{x}))$, which are proportional to the negative Laplace and Poisson log-likelihood, respectively. For binary classification using probabilistic classifiers (e.g., logistic regression), the most common loss function is the Bernoulli loss $L_{Ber}(y, f(\mathbf{x})) = -y \cdot \log f(\mathbf{x}) - (1-y) \cdot \log(1 - f(\mathbf{x}))$, which is equivalent to the negative Bernoulli log-likelihood as well as to the binary cross-entropy loss (Goodfellow et al., 2016). For survival tasks, owing to the popularity of the Cox PH model,

the negative Cox partial likelihood – referred to as Cox loss – is a frequently used loss function among deep learning-based methods (Wiegrebe et al., 2024b). It is defined as

$$L_{Cox} = -\frac{1}{\sum_{i=1}^{n} d_i} \sum_{i:d_i=1}^{n} \Big(f(\mathbf{x}_i) - \log \sum_{j \in \mathcal{R}_i} \exp(f(\mathbf{x}_j)) \Big), \tag{4.2}$$

with y_i here being the observed time (see Section 3) and $\mathcal{R}_i := \{j : y_j \geq y_i\}$ the risk set of subject i. $f(\mathbf{x})$ is simply the linear predictor $\mathbf{x}^\top \boldsymbol{\beta}$ in case of the Cox PH model (cf. Equation (3.10)), but can also be parametrized, for example, by a neural network, as done in many deep learning-based survival models (e.g., *DeepSurv*; Katzman et al., 2018). For discrete-time survival analysis, the negative log-likelihood loss of a survival model can be written as

$$L_{nll} = -\sum_{i=1}^{n} \Big(d_i \cdot \log(h_d(J_i \mid \mathbf{x}_i)) + d_i \cdot \log(S_d(J_i - 1 \mid \mathbf{x}_i)) + (1 - d_i) \cdot \log(S_d(J_i \mid \mathbf{x}_i)) \Big), \tag{4.3}$$

where h_d and S_d are as defined in Equations (3.4) and (3.5), respectively (Zadeh and Schmid, 2020). By contrast, the binary cross-entropy loss in discrete-time survival analysis, as defined by Ren et al. (2019), is

$$L_{ce} = -\sum_{i=1}^{n} \Big(d_i \cdot \log(1 - S_d(J_i \mid \mathbf{x}_i)) + (1 - d_i) \cdot \log(S_d(J_i \mid \mathbf{x}_i)) \Big). \tag{4.4}$$

Subsections 4.1 and 4.2 discuss machine learning and deep learning algorithms, as well as their application to survival analysis, in more detail. We note that, strictly speaking, deep learning is a subset of machine learning which uses multi-layered (i.e., deep) neural networks. In the following, we refer to machine learning as the set of classical, "non-deep" learning techniques, and to deep learning as the set of neural network-based techniques.

4.1. Machine learning

Classical machine learning refers to supervised learning algorithms that learn prediction models $\hat{f}$ from data $\mathcal{D}$ via shallow, non-deep model structures; in fact, standard statistical approaches such as linear and logistic regression models are usually also considered machine learning algorithms (Hastie et al., 2009). Beyond that, kernel methods (in particular, support vector machines (SVMs); Cortes and Vapnik, 1995), gradient boosting approaches such as component-wise boosting (Friedman, 2001; Bühlmann and Yu, 2003; Bühlmann and Hothorn, 2007), tree-based ensemble methods including decision trees and random forests (Breiman et al., 2017; Breiman, 2001), as well as lazy learning methods such as k-nearest neighbors (Cover and Hart, 1967) are among the most widely used types of machine learning algorithms. Many of these algorithms can be embedded into the ERM framework; for example, SVMs for binary classification can be expressed as L_2-regularized minimization of the Hinge loss. This is consistent with the fact that machine learning algorithms are usually designed to tackle regression or classification tasks.

In order to apply them to survival tasks, machine learning algorithms need survival-specific modifications. The Random Survival Forest (Ishwaran et al., 2008), for instance, uses the log-rank statistic as splitting criterion, as opposed to the L_2 loss or the Bernoulli loss commonly used for

regression or binary classification tasks, respectively; and Barnwal et al. (2022) introduce explicit loss functions for AFT survival regression in order to apply XGBoost to survival tasks. Wang et al. (2019) present an overview of machine learning approaches for survival analysis.

However, as illustrated in Figure 1 of Piller et al. (2025), the development of such custom modifications to make machine learning algorithms applicable to survival tasks usually takes years. This again underlines the benefit of being able to employ standard loss functions and off-the-shelf implementations of machine learning algorithms when handling survival tasks – by previously applying reduction techniques as demonstrated in the contributing article Piller et al. (2025) (Section 8).

4.2. Deep learning

Deep learning uses multi-layered neural networks to learn a prediction model $\hat{f}$ from data $\mathcal{D}$ by composing multiple (non-)linear transformations (neurons) of the input features $\mathbf{x}$. Neural networks are trained by minimizing a loss function; because of their composite structure consisting of simple building blocks, neural networks are generally compatible with any differentiable loss. Since overfitting is a common issue in deep learning, caused by the high dimensionality of the parameter space, deep learning models often also incorporate regularization (or penalization) terms into their loss functions, such as L_1 or L_2 weight decay (Bishop, 1995; Ng, 2004; Goodfellow et al., 2016), dropout (Srivastava et al., 2014), or batch normalization (Ioffe and Szegedy, 2015). Model optimization is typically performed via stochastic gradient descent or one of its adaptive variants (e.g., Adam; Kingma and Ba, 2017), where weight updates are efficiently computed using the backpropagation algorithm (Rumelhart et al., 1986).

The earliest and most fundamental type of neural network architecture is the feed-forward neural network (FFNN; Rosenblatt, 1958; Ivakhnenko, 1968), where information only flows forward, from the input layer through a sequence of fully-connected layers towards the output layer. The hypothesis space of an FFNN can be written as

$$\mathcal{H} = \left\{ f(\mathbf{x}) : f(\mathbf{x}) = \tau \circ \phi \circ \sigma^{(h)} \circ \phi^{(h)} \circ \sigma^{(h-1)} \circ \phi^{(h-1)} \circ \cdots \circ \sigma^{(1)} \circ \phi^{(1)}(\mathbf{x}) \right\}, \tag{4.5}$$

where h is the number of hidden layers, $\sigma^{(i)}$ and $\phi^{(i)}$ are the activation function and the weighted sum of hidden layer i, respectively, with τ and ϕ being the corresponding components of the output layer. This network architecture is suitable for tabular data. With no hidden layers and a single output neuron, an FFNN reduces to a standard linear regression model when the activation function τ (cf. Equation (4.5)) is set to the identity function and the L_2-loss is selected. It corresponds to a logistic regression when τ is the logistic sigmoid function and the Bernoulli loss is applied, and to a Cox PH regression model when τ is the identity function and the Cox loss (Equation (4.2)) is used. This illustrates how FFNNs can be considered extensions of standard statistical learning algorithms (Goodfellow et al., 2016). While architecturally simple, FFNNs are, in theory, capable of approximating any continuous function, as stated by the universal approximation theorem (Hornik et al., 1989). Other popular neural network architectures are convolutional neural networks (CNNs), invented in the late 1980s (LeCun et al., 1989), which are particularly successful in computer vision applications and often leverage transfer learning approaches based on large pre-trained networks (see, e.g., ResNet18; He et al., 2016); recurrent

neural networks (RNNs; Rumelhart et al., 1985; Elman, 1990; Jordan, 1997), which are capable of incorporating memory mechanisms that allow them to retain and process information from previous inputs, thus making them applicable to sequential data; and transformers (Vaswani et al., 2017), which employ an attention mechanism to capture contextual representations within sequential data that are subsequently used to generate context-aware output sequences.

Due to the modular structure of deep learning models and their compatibility with any loss function, their application to survival tasks is relatively straightforward: it merely requires parametrizing the desired quantity of interest (often the hazard) by a neural network and choosing an adequate survival loss. A popular choice for continuous-time deep learning-based survival methods is the Cox loss. However, from Equation (4.2) it follows that minimizing the Cox loss requires recomputation of the risk set $\mathcal{R}_i$ for each uncensored subject, which is computationally expensive, as noted in Section 3.2. This is why some deep learning-based survival methods instead use modifications of the Cox loss, such as the restriction of the risk set $\mathcal{R}_i$ to a (sufficiently large) risk subset $\tilde{\mathcal{R}}_i$ (Kvamme et al., 2019). Alternatively, many deep learning-based survival models discretize time (cf. Section 3.2) and, accordingly, use discrete-time survival losses (Wiegrebe et al., 2024b), in particular the negative log-likelihood loss L_{nll} (Equation (4.3)) or the binary cross-entropy loss L_{ce} (Equation (4.4)); see, e.g., Gensheimer and Narasimhan (2019) or Ren et al. (2019). As opposed to binary classification, however, L_{nll} is now distinct from L_{ce}, except for the extreme case where $d_i = 0, \forall i$. In fact, as shown by Zadeh and Schmid (2020), use of L_{ce} causes large prediction error along with biased predictions and poor calibration, because the information from uncensored individuals is not fully exploited. Nevertheless, the binary cross-entropy loss remains popular among deep learning-based discrete-time survival methods (cf. Wiegrebe et al., 2024b). Deep learning-based survival methods sometimes also use combinations of multiple loss functions, for example by adding a ranking loss to the original survival loss (see, e.g., Lee et al., 2018; Jing et al., 2019). This is likely inspired by the fact that the C-index, a popular evaluation metric for survival tasks, is based on pairwise rankings of subjects (Harrell et al., 1982; Harrell Jr et al., 1996). Other methods (e.g., Huang et al., 2018) directly construct a loss function derived from survival evaluation metrics.

The contributing article Kopper et al. (2022) (Section 9) uses the PEM reduction technique (cf. Section 3.2) to develop a deep learning-based survival model with a penalized Poisson loss. Due to the additive structure of the predictor, combining input from structured (tabular) data and unstructured data (e.g., images or text), the model preserves its interpretability. The contributing article Wiegrebe et al. (2024b) (Section 7) provides a comprehensive, structured overview of the various deep learning-based survival methods that have been developed in recent years as a consequence of the adoption of deep learning techniques to time-to-event analysis. The article characterizes all methods according to both deep learning-related attributes (e.g., model class or network architecture) and survival-related aspects (e.g., supported types of censoring and truncation or handling of competing risks). This enables practitioners to quickly identify which methods are adequate for their particular use case, while also helping researchers assess potential areas for future research.

Part II.

Modeling Trait Change And Its Genetics In Longitudinal Data Via Regression Techniques

5. Analyzing Longitudinal Trait Trajectories Using GWAS Identifies Genetic Variants For Kidney Function Decline

Contributing article:

Wiegrebe, S., Gorski, M., Herold, J. M., Stark, K. J., Thorand, B., Gieger, C., Böger, C. A., Schödel, J., Hartig, F., Chen, H., Winkler, T. W., Küchenhoff, H., and Heid, I. M. (2024a). Analyzing longitudinal trait trajectories using GWAS identifies genetic variants for kidney function decline. *Nature communications*, 15(1):10061.

Supplementary material available at:

`https://www.nature.com/articles/s41467-024-54483-9`

Author contributions:

S.W. conceived the experiments, was responsible for development and implementation of the statistical methods, conducted all main analyses and wrote the first draft of the manuscript. M.G. contributed to data preparation and GWAS analyses. J.M.H. conducted PGS analyses. K.J.S. contributed to biological follow-up. B.T. and C.G. provided data for the KORA study. J.S. and C.A.B. helped interpreting the biological results. F.H. conceived part of the experiments and supervised simulation analyses. H.C. contributed to GWAS analyses. T.W.W. contributed to GWAS analyses, biological follow-up, and helped writing the first draft. H.K. conceived the experiments and co-supervised the project. I.M.H. conceived the experiments, supervised the project, and wrote the first draft of the manuscript. All authors contributed to the writing, critically read and commented the manuscript.

6. Genetic-By-Age Interaction Analyses On Complex Traits In UK Biobank And Their Potential To Identify Effects On Longitudinal Trait Change

Contributing article:

Winkler, T. W., Wiegrebe, S., Herold, J. M., Stark, K. J., Küchenhoff, H., and Heid, I. M. (2024). Genetic-by-age interaction analyses on complex traits in UK Biobank and their potential to identify effects on longitudinal trait change. *Genome Biology*, 25(1):300.

Supplementary material available at:

`https://link.springer.com/article/10.1186/s13059-024-03439-9`

Author contributions:

TWW conceived the experiments, conducted all main analyses, and contributed to the writing group. IMH supervised the project and contributed to the writing group. SW provided the derivation of equivalence of genetic-by-age interaction and longitudinal change effects and contributed to the writing group. JH conducted power computations and created the heatmap of genetic effect sizes. KJS contributed to the interpretation of biological follow-up. All authors reviewed the manuscript.

Part III.

Deep Learning And Reduction Techniques For Survival Analysis

7. Deep Learning For Survival Analysis: A Review

Contributing article:

Wiegrebe, S., Kopper, P., Sonabend, R., Bischl, B., and Bender, A. (2024b). Deep learning for survival analysis: a review. *Artificial Intelligence Review*, 57(3):65.

Supplementary material available at:

`https://link.springer.com/article/10.1007/s10462-023-10681-3`
`https://survival-org.github.io/DL4Survival`

Author contributions:

AB proposed the research idea. SW and AB developed the analysis plan. SW performed the initial literature search and methods inclusion and performed most of the screening, supported by PK and AB. SW wrote the initial draft, supported by AB. AB and PK contributed to and reviewed the manuscript. SW and PK created the figures and tables. SW created the open-source interactive table. RS and BB reviewed and edited the manuscript and provided valuable feedback.

8. Reduction Techniques For Survival Analysis

Contributing article:

Piller, J., Orsini, L., Wiegrebe, S., Zobolas, J., Burk, L., Langbein, S. H., Studener, P., Goeswein, M., and Bender, A. (2025). Reduction Techniques for Survival Analysis. *arXiv preprint arXiv:2508.05715*.

Supplementary material available at:

`https://arxiv.org/abs/2508.05715`

Author contributions:

JP was responsible for the piecewise-exponential reduction technique, including its application and the coordination of the partitioned based approaches. Selected data examples and set up the data analysis pipeline for the application. Created and edited application figures (supported by SW). Created figure 1. Wrote sections 2, 3.2, 3.4, supported writing the introduction, reviewed the manuscript (supported by SW). LO was responsible for the pseudo-values reduction technique, including its application. Wrote section 4.3 and helped review the manuscript. SW was responsible for the discrete-time reduction technique, including its application. Helped create application figures, created Tables 1, 2 and 4. Helped write Introduction. Wrote Sections 3.1 (supported by JP) and 3.3 as well as Discussion. Helped review the manuscript. SHL supported JP for the theoretical part of the piecewise-exponential reduction technique and helped review the manuscript. LB was responsible for planning and conducting the benchmark comparison and the resulting tables and figures, with support from AB and JZ for the implementation of the reduction technique learners. JZ supervised PS and MG for the implementation of the reduction techniques in mlr3proba, revised Figures 1 and 2 of the manuscript, wrote the sections on the IPCW and CRM reductions and the Software section. Helped review the manuscript. PS was responsible for implementing DT, IPCW and PEM reduction techniques in mlr3proba. MG was responsible for implementing PEM reduction techniques in mlr3proba. AB proposed the idea to conceptually combine different methods to reduction techniques, provided the initial outline, contributed to the first draft, supervised the team, reviewed and edited the manuscript.

9. DeepPAMM: Deep Piecewise Exponential Additive Mixed Models For Complex Hazard Structures In Survival Analysis

Contributing article:

Kopper, P., Wiegrebe, S., Bischl, B., Bender, A., and Rügamer, D. (2022). DeepPAMM: Deep Piecewise Exponential Additive Mixed Models for Complex Hazard Structures in Survival Analysis. In Gama, J., Li, T., Yu, Y., Chen, E., Zheng, Y., and Teng, F., editors, *Advances in Knowledge Discovery and Data Mining*, pages 249–261, Cham. Springer International Publishing.

Supplementary material available at:

`https://link.springer.com/chapter/10.1007/978-3-031-05936-0_20`

Copyright information:

Author contributions:

he first draft, performed parts of the experiments, and orally presented the work. SW helped write the related literature section and helped perform the benchmark analyses. BB actively supervised the work. AB developed the initial idea to combine semi-structured deep distributional regression with piecewise-exponential additive modeling for flexible analysis of time-to-event data, helped write the first draft, and actively supervised the work. DR co-developed the methodological idea, helped write the first draft, and actively supervised the work. All authors reviewed the manuscript.

Part IV.

Modeling Disease Histories Using Multi-State Models

10. Multi-State Models For Modeling Disease Histories Based On Longitudinal Data

Contributing article:

Wiegrebe, S., Piller, J., Gorski, M., Behr, M., Küchenhoff, H., Heid, I. M., and Bender, A. (2025). Multi-state Models For Modeling Disease Histories Based On Longitudinal Data. *arXiv preprint arXiv:2509.19956*.

Supplementary material available at:

`https://arxiv.org/abs/2509.19956`

Copyright information:

Author contributions:

SW and AB proposed the research idea. SW reviewed the literature, perceived and conducted the simulation studies (supported by AB), and conducted the UK Biobank analyses (supported by MB, HK, IMH, and AB). SW wrote the initial draft of the manuscript and created all figures and tables, supported by AB. JP developed the multi-state framework for PAMs, provided the software, wrote parts of the methods, and helped discuss the findings of the simulation studies and of the UK Biobank analyses. MG provided the UK Biobank data. MB, HK and IMH helped discuss the findings of the simulation studies and supported the UK Biobank analyses. AB helped design the simulation studies and discuss their findings and supported the UK Biobank analyses. AB supported the writing of the initial draft of the manuscript. All authors contributed to the writing, critically read and commented on the manuscript.

Contributing Publications

Wiegrebe, S., Gorski, M., Herold, J. M., Stark, K. J., Thorand, B., Gieger, C., Böger, C. A., Schödel, J., Hartig, F., Chen, H., et al. (2024a). Analyzing longitudinal trait trajectories using GWAS identifies genetic variants for kidney function decline. *Nature communications*, 15(1):10061.

Winkler, T. W., Wiegrebe, S., Herold, J. M., Stark, K. J., Küchenhoff, H., and Heid, I. M. (2024). Genetic-by-age interaction analyses on complex traits in UK Biobank and their potential to identify effects on longitudinal trait change. *Genome Biology*, 25(1):300.

Wiegrebe, S., Kopper, P., Sonabend, R., Bischl, B., and Bender, A. (2024b). Deep learning for survival analysis: a review. *Artificial Intelligence Review*, 57(3):65.

Piller, J., Orsini, L., Wiegrebe, S., Zobolas, J., Burk, L., Langbein, S. H., Studener, P., Goeswein, M., and Bender, A. (2025). Reduction Techniques for Survival Analysis. *arXiv preprint arXiv:2508.05715.*

Kopper, P., Wiegrebe, S., Bischl, B., Bender, A., and Rügamer, D. (2022). DeepPAMM: Deep Piecewise Exponential Additive Mixed Models for Complex Hazard Structures in Survival Analysis. In J. Gama, T. Li, Y. Yu, E. Chen, Y. Zheng, and F. Teng (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 249–261). Cham: Springer International Publishing. ISBN 978-3-031-05936-0.

Wiegrebe, S., Piller, J., Gorski, M., Behr, M., Küchenhoff, H., Heid, I. M., and Bender, A. (2025). Multi-state Models For Modeling Disease Histories Based On Longitudinal Data. *arXiv preprint arXiv:2509.19956.*

References

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

Aalen, O. O. and Johansen, S. (1978). An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations. *Scandinavian Journal of Statistics*, 5(3):141–150.

Abdellaoui, A., Yengo, L., Verweij, K. J., and Visscher, P. M. (2023). 15 years of GWAS discovery: realizing the promise. *The American Journal of Human Genetics*, 110(2):179–194.

All of Us Research Program Investigators (2019). The "All of Us" research program. *New England Journal of Medicine*, 381(7):668–676.

Andersen, P. K., Klein, J. P., and Rosthøj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, 90(1):15–27.

Armoni, M. (2009). Reduction in CS: A (mostly) quantitative analysis of reductive solutions to algorithmic problems. *Journal on Educational Resources in Computing (JERIC)*, 8(4):1–30.

Artigas, M. S., Loth, D. W., Wain, L. V., Gharib, S. A., Obeidat, M., Tang, W., Zhai, G., Zhao, J. H., Smith, A. V., Huffman, J. E., et al. (2011). Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nature genetics*, 43(11):1082–1090.

Aschard, H. (2016). A perspective on interaction effects in genetic association studies. *Genetic epidemiology*, 40(8):678–688.

Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nature reviews genetics*, 7(10):781–791.

Barnwal, A., Cho, H., and Hocking, T. (2022). Survival regression with accelerated failure time model in XGBoost. *Journal of Computational and Graphical Statistics*, 31(4):1292–1302.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278.

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of statistical software*, 67:1–48.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., and Bolker, M. B. (2015c). Package 'lme4'. *convergence*, 12(1):2.

Bender, A. (2018). *Flexible modeling of time-to-event data and exposure-lag-response associations*. PhD thesis, Ludwig-Maximilians-Universität München, Germany.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Bentley, A. R., Sung, Y. J., Brown, M. R., Winkler, T. W., Kraja, A. T., Ntalla, I., Schwander, K., Chasman, D. I., Lim, E., Deng, X., et al. (2019). Multi-ancestry genome-wide gene–smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. *Nature genetics*, 51(4):636–648.

Beyersmann, J., Allignol, A., and Schumacher, M. (2011). *Competing risks and multistate models with R*. Springer Science & Business Media.

Bishop, C. M. (1995). Training with noise is equivalent to Tikhonov regularization. *Neural computation*, 7(1):108–116.

Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L., Friedman, J., Olshen, R. A., and Stone, C. J. (2017). *Classification and regression trees*. Chapman and Hall/CRC.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, pages 89–99.

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research*, 47(D1):D1005–D1012.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209.

Bühlmann, P. and Hothorn, T. (2007). Boosting Algorithms: Regularization, Prediction and Model Fitting. *Statistical Science*, 22(4):477–505.

Bühlmann, P. and Yu, B. (2003). Boosting with the L 2 loss: regression and classification. *Journal of the American Statistical Association*, 98(462):324–339.

Casalicchio, G. (2019). *On benchmark experiments and visualization methods for the evaluation and interpretation of machine learning models*. PhD thesis, Ludwig-Maximilians-Universität München, Germany.

Casci, T. (2010). SNPs that come in threes. *Nature Reviews Genetics*, 11(1):8–8.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1):s13742–015.

Chen, H., Conomos, M., and Pham, D. T. (2023). *GMMAT: Generalized Linear Mixed Model Association Tests*. R package version 1.4.2, `https://CRAN.R-project.org/package=GMMAT`.

Chen, H., Wang, C., Conomos, M. P., Stilp, A. M., Li, Z., Sofer, T., Szpiro, A. A., Chen, W., Brehm, J. M., Celedón, J. C., et al. (2016). Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *The American Journal of Human Genetics*, 98(4):653–666.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Chen, Z., Boehnke, M., Wen, X., and Mukherjee, B. (2021). Revisiting the genome-wide significance threshold for common variant GWAS. *G3*, 11(2):jkaa056.

Cheng, J., Edwards, L. J., Maldonado-Molina, M. M., Komro, K. A., and Muller, K. E. (2010). Real longitudinal data analysis for real people: building a good enough mixed model. *Statistics in medicine*, 29(4):504–520.

Cheung, L. C., Pan, Q., Hyun, N., Schiffman, M., Fetterman, B., Castle, P. E., Lorey, T., and Katki, H. A. (2017). Mixture models for undiagnosed prevalent disease and interval-censored incident disease: applications to a cohort assembled from electronic health records. *Statistics in medicine*, 36(22):3583–3595.

CKDGen Consortium, KidneyGen Consortium, EchoGen Consortium, CHARGE-HF Consortium, Aspelund, T., Garcia, M., Chang, Y.-P. C., O'Connell, J. R., Steinle, N. I., et al. (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*, 478(7367):103–109.

Coens, F., Knops, N., Tieken, I., Vogelaar, S., Bender, A., Kim, J. J., Krupka, K., Pape, L., Raes, A., Tönshoff, B., et al. (2024). Time-varying determinants of graft failure in pediatric kidney transplantation in Europe. *Clinical Journal of the American Society of Nephrology*, 19(3):345–354.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Cottin, A., Pecuchet, N., Zulian, M., Guilloux, A., and Katsahian, S. (2022). IDNetwork: A deep illness-death network based on multi-state event history process for disease prognostication. *Statistics in Medicine*, 41(9):1573–1598.

Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

Diggle, P. (2002). *Analysis of longitudinal data*. Oxford university press.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS genetics*, 9(3):e1003348.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American statistical association*, 56(293):52–64.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American statistical Association*, 72(359):557–565.

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

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, 14(2):179–211.

Evangelou, E. and Ioannidis, J. P. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*, 14(6):379–389.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. D. (2022). Regression models. In *Regression: Models, methods and applications*, pages 23–84. Springer.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, 94(446):496–509.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012). *Applied longitudinal analysis*. John Wiley & Sons.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

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

Gallagher, M. D. and Chen-Plotkin, A. S. (2018). The post-GWAS era: from association to function. *The American Journal of Human Genetics*, 102(5):717–730.

Garrett, R., Patel, M., Feldman, B., and Pullenayegum, E. (2024). Why recommended visit intervals should be extracted when conducting longitudinal analyses using electronic health record data: examining visit mechanism and sensitivity to assessment not at random. *arXiv preprint arXiv:2406.04077*.

Gelman, A. and Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gensheimer, M. F. and Narasimhan, B. (2019). A scalable discrete-time survival model for neural networks. *PeerJ*, 7:e6257.

Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. *Nature genetics*, 45(11):1274–1283.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.

Gorski, M., Rasheed, H., Teumer, A., Thomas, L. F., Graham, S. E., Sveinbjornsson, G., Winkler, T. W., Günther, F., Stark, K. J., Chai, J.-F., et al. (2022). Genetic loci and prioritization of genes for kidney function decline derived from a meta-analysis of 62 longitudinal genome-wide association studies. *Kidney International*, 102(3):624–639.

Gorski, M., Wiegrebe, S., Burkhardt, R., Behr, M., Küchenhoff, H., Stark, K. J., Böger, C. A., and Heid, I. M. (2025). Bias-corrected serum creatinine from UK Biobank electronic medical records generates an important data resource for kidney function trajectories. *Scientific Reports*, 15(1):3540.

Greiser, K., Bohn, B., Hilger-Kolb, J., Panreck, L., Breunig, E., Lieb, W., Pischon, T., Schikowski, T., and Völzke, H. (2023). The German National Cohort (NAKO): Design, current state, and further follow-up data collection. *European Journal of Public Health*, 33(Supplement_2):ckad160–499.

Groha, S., Schmon, S. M., and Gusev, A. (2020). A general framework for survival analysis and multi-state modelling. *arXiv preprint arXiv:2006.04893*.

Günther, F. (2021). *Statistical models for the analysis of response variables with limited data quality due to misclassification and missing information*. PhD thesis, Ludwig-Maximilians-Universität München, Germany.

Hagar, Y., Albers, D., Pivovarov, R., Chase, H., Dukic, V., and Elhadad, N. (2014). Survival analysis with electronic health record data: Experiments with chronic kidney disease. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 7(5):385–403.

Hardin, J. W. and Hilbe, J. M. (2002). *Generalized estimating equations*. chapman and hall/CRC.

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., and Rosati, R. A. (1982). Evaluating the yield of medical tests. *Jama*, 247(18):2543–2546.

Harrell Jr, 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.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American statistical association*, 72(358):320–338.

Hastie, T., Tibshirani, R., Friedman, J., et al. (2009). The elements of statistical learning.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *elife*, 7:e34408.

Henderson, C. R., Kempthorne, O., Searle, S. R., and Von Krosigk, C. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 15(2):192–218.

Hofmeister, R. J., Rubinacci, S., Ribeiro, D. M., Buil, A., Kutalik, Z., and Delaneau, O. (2022). Parent-of-Origin inference for biobanks. *Nature Communications*, 13(1):6668.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366.

Huang, C., Zhang, A., and Xiao, G. (2018). Deep integrative analysis for survival prediction. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2018: Proceedings of the Pacific Symposium*, pages 343–352. World Scientific.

Hyun, N., Cheung, L. C., Pan, Q., Schiffman, M., and Katki, H. A. (2017). Flexible risk prediction models for left or interval-censored data from electronic health records. *The annals of applied statistics*, 11(2):1063.

Iacobelli, S. and Carstensen, B. (2013). Multiple time scales in multi-state models. *Statistics in medicine*, 32(30):5315–5327.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr.

Irwin, J. (1949). The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. *Epidemiology & Infection*, 47(2):188–189.

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., and Lauer, M. S. (2008). Random Survival Forests. *The Annals of Applied Statistics*, 2(3):841–860.

Ivakhnenko, A. G. (1968). The group method of data handling; a rival of the method of stochastic approximation. *Soviet Automatic Control*, 13(3):43–55.

Jing, B., Zhang, T., Wang, Z., Jin, Y., Liu, K., Qiu, W., Ke, L., Sun, Y., He, C., Hou, D., et al. (2019). A deep survival analysis method based on ranking. *Artificial intelligence in medicine*, 98:1–9.

Jordan, M. I. (1997). Serial order: A parallel distributed processing approach. In *Advances in psychology*, volume 121, pages 471–495. Elsevier.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data*. John Wiley & Sons.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809):434–443.

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., and Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC medical research methodology*, 18(1):24.

Kingma, D. P. and Ba, J. (2017). Adam: A Method for Stochastic Optimization.

Ko, S., German, C. A., Jensen, A., Shen, J., Wang, A., Mehrotra, D. V., Sun, Y. V., Sinsheimer, J. S., Zhou, H., and Zhou, J. J. (2022). GWAS of longitudinal trajectories at biobank scale. *The American Journal of Human Genetics*, 109(3):433–445.

Kopper, P., Wiegrebe, S., Bischl, B., Bender, A., and Rügamer, D. (2022). DeepPAMM: Deep Piecewise Exponential Additive Mixed Models for Complex Hazard Structures in Survival Analysis. In Gama, J., Li, T., Yu, Y., Chen, E., Zheng, Y., and Teng, F., editors, *Advances in Knowledge Discovery and Data Mining*, pages 249–261, Cham. Springer International Publishing.

Köttgen, A., Glazer, N. L., Dehghan, A., Hwang, S.-J., Katz, R., Li, M., Yang, Q., Gudnason, V., Launer, L. J., Harris, T. B., et al. (2009). Multiple loci associated with indices of renal function and chronic kidney disease. *Nature genetics*, 41(6):712–717.

Kvamme, H., Borgan, Ø., and Scheel, I. (2019). Time-to-event prediction with neural networks and Cox regression. *Journal of machine learning research*, 20(129):1–30.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.

Lammi, V., Nakanishi, T., Jones, S. E., Andrews, S. J., Karjalainen, J., Cortés, B., O'Brien, H. E., Ochoa-Guzman, A., Fulton-Howard, B. E., Broberg, M., et al. (2025). Genome-wide association study of long COVID. *Nature Genetics*, pages 1–16.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.

Lee, C., Zame, W., Yoon, J., and Van Der Schaar, M. (2018). Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Levin, A., Stevens, P. E., Bilous, R. W., Coresh, J., De Francisco, A. L., De Jong, P. E., Griffith, K. E., Hemmelgarn, B. R., Iseki, K., Lamb, E. J., Levey, A. S., Riella, M. C., Shlipak, M. G., Wang, H., White, C. T., Winearls, C. G., and Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group (2013). KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney International Supplements*, 3(1):1–150.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Liu, C.-T., Estrada, K., Yerges-Armstrong, L. M., Amin, N., Evangelou, E., Li, G., Minster, R. L., Carless, M. A., Kammerer, C. M., Oei, L., et al. (2012). Assessment of gene-by-sex interaction effect on bone mineral density. *Journal of Bone and Mineral Research*, 27(10):2051–2064.

Liu, H. Y., Alyass, A., Abadi, A., Peralta-Romero, J., Suarez, F., Gomez-Zamudio, J., Audirac, A., Parra, E. J., Cruz, M., and Meyre, D. (2019). Fine-mapping of 98 obesity loci in Mexican children. *International Journal of Obesity*, 43(1):23–32.

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518(7538):197–206.

Lokku, A., Lim, L. S., Birken, C. S., Pullenayegum, E. M., and Collaboration, T. K. (2020). Summarizing the extent of visit irregularity in longitudinal data. *BMC Medical Research Methodology*, 20:1–9.

Manning, A. K., Hivert, M.-F., Scott, R. A., Grimsby, J. L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.-T., Bielak, L. F., Prokopenko, I., et al. (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature genetics*, 44(6):659–669.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of memory and language*, 94:305–315.

Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J. A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., et al. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. *Nature genetics*, 53(7):1097–1103.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., and Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature reviews genetics*, 9(5):356–369.

McClelland, G. H. and Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological bulletin*, 114(2):376.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

NCI-NHGRI Working Group on Replication in Association Studies (2007). Replicating genotype–phenotype associations. *Nature*, 447(7145):655–660.

Nelson, M. R., Tipney, H., Painter, J. L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P. C., Li, M. J., Wang, J., et al. (2015). The support of human genetic evidence for approved drug indications. *Nature genetics*, 47(8):856–860.

Ng, A. Y. (2004). Feature selection, L 1 vs. L 2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*, page 78.

Nießl, A., Allignol, A., Beyersmann, J., and Mueller, C. (2023). Statistical inference for state occupation and transition probabilities in non-Markov multi-state models subject to both random left-truncation and right-censoring. *Econometrics and Statistics*, 25:110–124.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *Bmj*, 316(7139):1236–1238.

Peterhoff, D., Wiegrebe, S., Einhauser, S., Patt, A. J., Beileke, S., Günther, F., Steininger, P., Niller, H. H., Burkhardt, R., Küchenhoff, H., et al. (2023). Population-based study of the durability of humoral immunity after SARS-CoV-2 infection. *Frontiers in Immunology*, 14:1242536.

Piller, J., Küchenhoff, H., and Bender, A. (2024). Flexible additive models for multi-event survival analysis. In *38th International Workshop on Statistical Modelling*, page 241.

Piller, J., Orsini, L., Wiegrebe, S., Zobolas, J., Burk, L., Langbein, S. H., Studener, P., Goeswein, M., and Bender, A. (2025). Reduction Techniques for Survival Analysis. *arXiv preprint arXiv:2508.05715*.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, 38(8):904–909.

Rau, E., Tischendorf, T., and Mitzscherlich, B. (2024). Implementation of the electronic health record in the German healthcare system: an assessment of the current status and future development perspectives considering the potentials of health data utilisation by representatives of different stakeholder groups. *Frontiers in Health Services*, 4:1370759.

Ren, K., Qin, J., Zheng, L., Yang, Z., Zhang, W., Qiu, L., and Yu, Y. (2019). Deep recurrent survival analysis. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 4798–4805.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 54(3):507–554.

Rizvi, A. A., Karaesmen, E., Morgan, M., Preus, L., Wang, J., Sovic, M., Hahn, T., and Sucheston-Campbell, L. E. (2019). gwasurvivr: an R package for genome-wide survival analysis. *Bioinformatics*, 35(11):1968–1970.

Robinson-Cohen, C., Triozzi, J. L., Rowan, B., He, J., Chen, H. C., Zheng, N. S., Wei, W.-Q., Wilson, O. D., Hellwege, J. N., Tsao, P. S., et al. (2023). Genome-wide association study of CKD progression. *Journal of the American Society of Nephrology*, 34(9):1547–1559.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.

Royston, P. and Parmar, M. K. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in medicine*, 30(19):2409–2421.

Royston, P. and Parmar, M. K. (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC medical research methodology*, 13(1):152.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1985). Learning Internal Representations by Error Propagation. Technical Report ADA164453, University of California, San Diego, Institute for Cognitive Science. Technical report, March–September 1985.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533–536.

Sham, P. C. and Purcell, S. M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics*, 15(5):335–346.

Slawomirski, L. et al. (2023). Progress on implementing and using electronic health record systems: Developments in OECD countries as of 2021. OECD Health Working Papers, No. 160, OECD Publishing, Paris.

Smith, G. D. and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology*, 32(1):1–22.

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., Allen, H. L., Lindgren, C. M., Luan, J., Mägi, R., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics*, 42(11):937–948.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.

Stanzick, K. J., Li, Y., Schlosser, P., Gorski, M., Wuttke, M., Thomas, L. F., Rasheed, H., Rowan, B. X., Graham, S. E., Vanderweff, B. R., et al. (2021). Discovery and prioritization of variants and genes for kidney function in¿ 1.2 million individuals. *Nature communications*, 12(1):4350.

Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3):e1001779.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, 20(8):467–484.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466(7307):707–713.

Thomas, D. (2010). Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. *Annual review of public health*, 31(1):21–36.

Torkamani, A., Wineinger, N. E., and Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*, 19(9):581–590.

Tutz, G., Schmid, M., et al. (2016). *Modeling discrete time-to-event data*, volume 3. Springer.

Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., and Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1):59.

UK Biobank (2023). Primary Care Linked Data: Data Coding and Quality Documentation. Technical report, UK Biobank. Accessed October 22, 2025.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., and Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics*, 101(1):5–22.

Wain, L. V., Shrine, N., Artigas, M. S., Erzurumluoglu, A. M., Noyvert, B., Bossini-Castillo, L., Obeidat, M., Henry, A. P., Portelli, M. A., Hall, R. J., et al. (2017). Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nature genetics*, 49(3):416–425.

Wang, P., Li, Y., and Reddy, C. K. (2019). Machine learning for survival analysis: A survey. *ACM Computing Surveys (CSUR)*, 51(6):1–36.

Wang, X., Chen, H., and Pham, D. T. (2025). *MAGEE: Mixed Model Association Test for GEne-Environment Interaction*. R package version 1.4.3, `https://cran.r-project.org/package=MAGEE`.

Wang, X., Lim, E., Liu, C.-T., Sung, Y. J., Rao, D. C., Morrison, A. C., Boerwinkle, E., Manning, A. K., and Chen, H. (2020). Efficient gene–environment interaction tests for large biobank-scale sequencing studies. *Genetic epidemiology*, 44(8):908–923.

Warren, H. R., Evangelou, E., Cabrera, C. P., Gao, H., Ren, M., Mifsud, B., Ntalla, I., Surendran, P., Liu, C., Cook, J. P., et al. (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nature genetics*, 49(3):403–415.

Wiegrebe, S., Gorski, M., Herold, J. M., Stark, K. J., Thorand, B., Gieger, C., Böger, C. A., Schödel, J., Hartig, F., Chen, H., Winkler, T. W., Küchenhoff, H., and Heid, I. M. (2024a). Analyzing longitudinal trait trajectories using GWAS identifies genetic variants for kidney function decline. *Nature communications*, 15(1):10061.

Wiegrebe, S., Kopper, P., Sonabend, R., Bischl, B., and Bender, A. (2024b). Deep learning for survival analysis: a review. *Artificial Intelligence Review*, 57(3):65.

Wiegrebe, S., Piller, J., Gorski, M., Behr, M., Küchenhoff, H., Heid, I. M., and Bender, A. (2025). Multi-state Models For Modeling Disease Histories Based On Longitudinal Data. *arXiv preprint arXiv:2509.19956*.

Willer, C. J., Li, Y., and Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17):2190–2191.

Winkler, T. W., Wiegrebe, S., Herold, J. M., Stark, K. J., Küchenhoff, H., and Heid, I. M. (2024). Genetic-by-age interaction analyses on complex traits in UK Biobank and their potential to identify effects on longitudinal trait change. *Genome Biology*, 25(1):300.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(1):95–114.

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

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. chapman and hall/CRC.

Zadeh, S. G. and Schmid, M. (2020). Bias in cross-entropy-based training of deep survival networks. *IEEE transactions on pattern analysis and machine intelligence*, 43(9):3126–3137.

Zhao, L., Claggett, B., Tian, L., Uno, H., Pfeffer, M. A., Solomon, S. D., Trippa, L., and Wei, L.-J. (2016). On the restricted mean survival time curve in survival analysis. *Biometrics*, 72(1):215–221.

Eidesstattliche Versicherung (Affidavit)

(Siehe Promotionsordnung vom 12. Juli 2011, § 8 Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eides statt, dass diese Dissertation von mir selbstständig und ohne unerlaubte Beihilfe angefertigt wurde. Während des Überarbeitungsprozesses habe ich DeepL (DeepL SE), ChatGPT (OpenAI, GPT-5) und Google AI Studio (Google, Gemini 2.5 Pro) verwendet, um Rechtschreibfehler und grammatikalische Fehler zu korrigieren und bereits verfasste Inhalte sprachlich zu verbessern. Außerdem habe ich ChatGPT und Google AI Studio zur Unterstützung bei der Literaturrecherche verwendet.

München, den 26.01.2026 Simon Wiegrebe