Received: 31 January 2020 | Revised: 18 June 2020 | Accepted: 22 July 2020

DOI: 10.1002/wics.1529

ADVANCED REVIEW

WIREs COMPUTATIONAL STATISTICS WILEY

Competing risks analysis for discrete time-to-event data

Matthias Schmid | **Moritz Berger**

Institute of Medical Biometry, Informatics and Epidemiology, Faculty of Medicine, University of Bonn, Bonn, Germany

Correspondence
Matthias Schmid, Institute of Medical Biometry, Informatics and Epidemiology, Faculty of Medicine, University of Bonn, 53127 Bonn, Germany.
Email: matthias.c.schmid@uni-bonn.de

Funding information
German Research Foundation, Grant/Award Number: SCHM 2966/2-1

Abstract

This article presents an overview of statistical methods for the analysis of discrete failure times with competing events. We describe the most commonly used modeling approaches for this type of data, including discrete versions of the cause-specific hazards model and the subdistribution hazard model. In addition to discussing the characteristics of these methods, we present approaches to nonparametric estimation and model validation. Our literature review suggests that discrete competing-risks analysis has gained substantial interest in the research community and is used regularly in econometrics, biostatistics, and educational research.

This article is categorized under:
Statistical Models > Survival Models
Statistical Models > Semiparametric Models
Statistical Models > Generalized Linear Models

KEYWORDS

cause-specific hazards model, competing events, cumulative incidence function, discrete time-to-event analysis, subdistribution hazard model

1 | INTRODUCTION

Longitudinal empirical studies often involve the statistical analysis of a set of observation times that are measured on a discrete time scale $t = 1, 2, ..., q$. Typical examples are given by clinical and epidemiological studies with prespecified follow-up times where the values of t represent a set of fixed follow-up intervals $[a_{t-1}, a_t)$ with continuous boundaries $a_0 = 0 < a_1 < ... < a_q = \infty$ (e.g., Steinberg et al., 2016). Other examples of discretized continuous event times are given by panel studies in econometrics where durations are often measured in weeks or months (e.g., unemployment spells; Fahrmeir & Wagenpfeil, 1996; Han & Hausman, 1990) and patient data collected in intensive care units, which often record measurements on a daily basis (e.g., time to death or infection; Barnett et al., 2009; Heyard et al., 2019). Discrete observation times are also encountered in studies with an *intrinsically* discrete time scale, for example, in clinical trials dealing with time to pregnancy (where the observation time is often defined by the number of menstrual cycles, Scheike & Keiding, 2006; Fehring, Schneider, Raviele, Rodriguez, & Pruszynski, 2013) and in the analysis of university outcomes in educational research (where time is, for example, measured in semesters, Scott & Kennedy, 2005; Meggiolaro, Giraldo, & Clerici, 2017; Vallejos & Steel, 2017). The methods presented here apply to both discretized continuous observation times and intrinsically discrete observation times, so that we will not distinguish between the two scenarios in the following.

If the interest of an empirical study is in one or more *target events* that may (or may not) have happened at the end of the observation times, statistical investigations usually involve the lengths of the observation times and their determinants. In this case, the observation times may be subject to *right-censoring*, as they either correspond to the minimum of the true event time (if the first target event was observed during the study), or to a censoring time (if an individual dropped out or was still event-free at the end of the study). As a consequence, *time-to-event analysis* (also termed "survival analysis", "failure time analysis" or "event history analysis") is concerned, among other issues, with the development of statistical models and hypothesis tests for censored longitudinal data.

Consider, for example, a data set that was collected for the Clinical Randomization of an Antifibrinolyticin Significant Hemorrhage 2 ("CRASH-2"; CRASH-2 Trial Collaborators, 2010) trial and that will be used for illustrative purposes in the next sections. The CRASH-2 trial was conducted in 274 hospitals in 40 countries between 2005 and 2010, collecting information on hospital death in adult trauma patients with or at risk of significant hemorrhage. Death was recorded during hospitalization of the patients for up to 28 days after randomization, and observation times were measured on a discrete scale (days) with $t = 1, 2, ...,28,29$. After randomization, patients had either died, been discharged alive, transferred to another hospital, or were still alive in hospital, with the latter three scenarios resulting in right-censored observation times. Note that the value $t = q = 29$ refers to patients that survived more than 28 days after randomization. For deceased patients, several causes of death (= target events) were recorded, of which three (bleeding, head injury, other) will be considered in our analysis. Furthermore, the study database contains several covariates (e.g., Glasgow Coma Score, age, blood pressure, and number of hours since injury) that were measured at the time of randomization. For this type of data, the aims of a time-to-event analysis are usually (a) to descriptively summarize information on event times and observed events, (b) to build a statistical model incorporating the covariate information, and (c) to conduct hypothesis tests on the effects of the covariates on event occurrence. A descriptive summary of the CRASH-2 data, which is based on the publicly available version of the study database, is presented in Table 1.

Traditional methods for time-to-event analysis have often dealt with building models for one target event only. For example, in case of the CRASH-2 data, one could focus on only one cause of death (e.g., bleeding) and treat all patients that experienced another target event (e.g., death due to head injury or other causes, in the following termed "competing events") as censored. This strategy has the advantage that it is simple to implement (see, e.g., the works by Kalbfleisch & Prentice, 2002; Singer & Willett, 2003; Box-Steffensmeier & Jones, 2004; Allison, 2014 and Tutz & Schmid, 2016, which provide comprehensive treatments of the available methodology for discrete failure time analysis with a single target event). On the other hand, it is often prone to biased analysis results. This is because the "censoring" times implicitly defined by this strategy are usually not independent of the event times, so that the assumptions of many existing methods for time-to-event analysis with a single target event are violated (e.g., the random censoring assumption or, more generally, the independent censoring assumption made by Kalbfleisch & Prentice, 2002). As a consequence, special methods that account for the presence of competing risks are needed.

The purpose of this article is to provide an overview of the currently available statistical methodology for the analysis of discrete time-to-event data with competing target events. Section 2 starts with basic notations and definitions. In Sections 3 and 4, two modeling approaches for right-censored discrete competing-risks data are presented, namely the cause-specific discrete hazards model and the discrete subdistribution hazard model. Usually, these models employ *parametric* (e.g., linear) terms to quantify the effects of the covariates on the respective hazard(s). Section 5 presents a set of alternative modeling approaches that involve *nonparametric* recursive partitioning techniques for competing-risks analysis. The latter techniques may provide a convenient alternative to parametric modeling in situations where linear (or additive) effects do not capture the true data structure very well (e.g., when interactions between covariates are present or when covariate effects are highly nonlinear). Section 6 contains methods for measuring the goodness-of-fit and the calibration of a fitted cause-specific or subdistribution hazard model. Section 7 provides an overview of extensions and alternative analysis approaches. Section 8 contains a summary of the article.

2 | NOTATION AND DEFINITIONS

Assume that the interest is in the analysis of a data set containing information on n independent individuals with covariate vectors $\mathbf{x}_i = (x_{i1}, ..., x_{ip})^\top$, $i = 1, ..., n$. Let T_i be the event time and C_i the censoring time of individual i. Both T_i and C_i are assumed to be random variables taking discrete values in $\{1, 2, ..., q\}$. It is further assumed that T_i and C_i are independent (random censoring) and that the censoring mechanism is noninformative in the sense that it does not depend on any parameters used to model the event times (Kalbfleisch & Prentice, 2002, Chapter 6). For right-censored

TABLE 1 Description and summary statistics of the observation times, the target events, and the covariates used for the analysis of the CRASH-2 data (Q1 = first quartile, Q3 = third quartile)

Observation times and target events	**Summary statistics**						
	Min	Q1	Median	Q3	Max	Mean	SD
Time (days)	1	3	8	16	29	10.7	9.0
Cause of death	Bleeding:	902	Head injury:	1,155 (6.1%)	Other:	746 (3.9%)	
Covariates	**Summary statistics**						
	Min	Q1	Median	Q3	Max	Mean	SD
Age (years)	14	24	31	43	99	34.6	14.3
Time since injury (Itime; hours)	0.1	1.0	2.0	4.0	72.0	2.9	2.3
Systolic blood pressure (Sbp; mmHg)	4	80	96	110	250	99.1	25.4
Heart rate (Hr; per minute)	3	90	105	120	220	104.2	20.9
Respiratory rate (Rr; per minute)	1	20	22	26	96	223.0	6.7
Central capillary refill time (Ccrt; seconds)	1	2	3	4	30	3.2	1.6
Glasgow coma score (Gcs)	3	11	15	15	15	12.5	3.6
Gender	Male (m):	15,985 (84%)	Female (f):	3,053 (16%)			
Type of injury (Itype)	Blunt (b):	10,558 (55.5%)	Penetrating (p):	6,169 (32.4%)	Blunt and penetrating (bp):	2,311 (12.1%)	

The numbers were computed from the publicly available version of the study database (http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets). After exclusion of patients with missing values in any of the presented variables, the analysis data set contained information on 19,038 patients. Note that the following analyses will be based on a random sample containing two-thirds of the data (12,692 patients). The remaining third will be used for the validation checks presented in Section 6. Also, note that several authors have suggested to consider alive discharge as an additional competing event in applications involving hospital data. They argued that a censoring event defined by alive discharge is impossible to happen after hospital death, and that time to death and time to alive discharge cannot be considered independent (see the assumptions on the censoring process in Section 2). For the sake of illustration, we will not investigate this issue here and refer to Beyersmann and Schrade (2017) for an in-depth discussion.

data, the observation time of individual i is denoted by $\tilde{T}_i = \min(T_i, C_i)$, that is, $\tilde{T}_i$ corresponds to the true event time if $T_i \leq C_i$ and to the censoring time otherwise. The status indicator $\Delta_i := I(T_i \leq C_i)$ denotes whether $\tilde{T}_i$ is right-censored ($\Delta_i = 0$) or not ($\Delta_i = 1$). The number of competing target events is denoted by J, and the event type of the i-th individual at T_i is represented by the random variable $\epsilon_i \in \{1, , J\}$.

A key quantity for competing risks analysis is the discrete *cumulative incidence function* (CIF), which for each of the events $j \in \{1, ..., J\}$ is defined by $F_j(t|\ \mathbf{x}_i) := P(T_i \leq t, \epsilon_i = j|\ \mathbf{x}_i)$. By definition, F_j is bounded between 0 and $F_j(q|\ \mathbf{x}_i) = P(\epsilon_i = j|\ \mathbf{x}_i) \leq 1$. A related quantity, on which the models presented in Section 3 are based, is the *cause-specific hazard function* $\lambda_j(t|\ \mathbf{x}_i) := P(T_i = t, \epsilon_i = j|\ T_i \geq t, \mathbf{x}_i)$, which quantifies the probability of experiencing a type-j event at time point t given survival up to t. Note that, in contrast to survival analysis with continuous event times, the cause-specific hazard function is defined in terms of a probability and not by a (possibly unbounded) function in $\mathbb{R}_0^+$. It follows that the *overall hazard function*, that is, the probability of experiencing any of the J events at t (given survival up to t), becomes $\lambda(t|\mathbf{x}_i) := P(T_i = t|T_i \geq t, \mathbf{x}_i) = \sum_{j=1}^{J} \lambda_j(t|\mathbf{x}_i)$. From this it can further be derived that the *survival function* (denoting the probability of experiencing the first event after t) can be expressed as $S(t|\mathbf{x}_i) := P(T_i > t|\mathbf{x}_i) = \prod_{t=1}^{q}(1 - \lambda(t|\mathbf{x}_i))$, cf. Tutz and Schmid (2016). Combining these terms, the CIF for the j-th event can be written as

$$F_j(t|\mathbf{x}_i) = \sum_{s=1}^{t} \lambda_j(s|\mathbf{x}_i) \cdot S(s-1|\mathbf{x}_i), \tag{1}$$

where $S(0|\ \mathbf{x}_i) := 1$. Importantly, $F_j(t|\ \mathbf{x}_i)$ depends on the overall survival function $S(t|\ \mathbf{x}_i)$ and thus on *all* cause-specific hazard functions $\lambda_j(t|\ \mathbf{x}_i)$. In particular, there is usually no one-to-one relationship between $\lambda_j(t|\ \mathbf{x}_i)$ and $F_j(t|\ \mathbf{x}_i)$. This is analogous to survival analysis with continuous event times, cf. Beyersmann, Allignol, and Schumacher (2011).

3 | DISCRETE CAUSE-SPECIFIC HAZARDS MODELS

A popular approach to modeling discrete-time competing-risks data is the *discrete cause-specific hazards model* (Tutz, 1995), which relates the cumulative incidence functions $F_j(t|\ \mathbf{x}_i)$ to a set of covariate values $\mathbf{x}_i$ by specifying a multivariate regression model for the cause-specific hazards $\lambda_j(t|\ \mathbf{x}_i)$. We first consider the parametric version of the discrete cause-specific hazards model, which is defined by

$$\lambda_j(t|\mathbf{x}_i) = \frac{\exp\left(\gamma_{0tj} + \mathbf{x}_i^\top \gamma_j\right)}{1 + \sum_{j=1}^{J} \exp\left(\gamma_{0tj} + \mathbf{x}_i^\top \gamma_j\right)} \tag{2}$$

for $j = 1, ..., J$, $t = 1, ..., q-1$. The parameters $\gamma_{01j}, ..., \gamma_{0(q-1)j}$ represent a set of cause-specific baseline coefficients, and $\gamma_j := (\gamma_{j1}, ..., \gamma_{jp})^\top$ is a cause-specific vector of regression coefficients. Technically, the baseline coefficients can be treated as additional factor variables in Model (2). Defining the probability $\lambda_0(t|\ \mathbf{x}_i) := P(T_i > t|\ T_i \geq t, \mathbf{x}_i)$ as the "hazard" for conditional survival beyond t, it can be shown that $\lambda_j(t|\mathbf{x}_i)/\lambda_0(t|\mathbf{x}_i) = \exp\left(\gamma_{0tj}\right) \cdot \exp\left(\gamma_{j1}\right)^{x_1} \cdot ... \cdot \exp\left(\gamma_{jp}\right)^{x_p}$, cf. Tutz and Schmid (2016), Chapter 8.1. As a consequence, the increase of covariate x_k, $k \in \{1, ..., p\}$, by one unit increases the cause-specific odds by the factor $\exp(\gamma_{jk})$. This is equivalent to the specification of a multinomial logistic regression model with $J+1$ outcome categories and reference category "survival beyond t". Defining binary variables by

$$\mathbf{y}_{it}^\top = \left(y_{it0}, y_{it1}, ..., y_{i\tilde{T}_i\epsilon_i}, ..., y_{itJ}\right) = \begin{cases} (1,0,...,0,...,0), \text{if } t < \tilde{T}_i, \\ (0,0,...,1,...,0), \text{if } t = \tilde{T}_i,\ \Delta_i = 1, \\ (1,0,...,0,...,0), \text{if } t = \tilde{T}_i,\ \Delta_i = 0 \end{cases} \tag{3}$$

with $y_{i\tilde{T}_i\epsilon_i} = 1$ in the second row of (3), the log-likelihood of Model (2) can be written as

$$l = \sum_{i=1}^{n} \sum_{t=1}^{\tilde{T}_i} \left\{ \sum_{j=1}^{J} y_{itj} \log\left(\lambda_j(t|\mathbf{x}_i)\right) + y_{it0} \log(1 - \lambda(t|\mathbf{x}_i)) \right\}, \tag{4}$$

cf. Tutz and Schmid (2016), Chapter 8.2. Specifically, it can be shown that (4) follows from the random and non-informative censoring assumptions stated in Section 2 and from the definition of $\Delta_i := I(T_i \leq C_i)$, which implies "censoring at the end of the interval" (Tutz & Schmid, 2016, p. 52) and which is equivalent to Kalbfleisch and Prentices's assumption that "censorings follow failures in $[t, t + dt)$" (Kalbfleisch & Prentice, 2002, p. 194). For a detailed derivation of (4), we refer to Chapter 8 of Kalbfleisch and Prentice (2002). Note that the function in (4) cannot be decomposed into separate terms for each target event, cf. Lee, Feuer, and Fine (2018). However, a closer inspection of (4) reveals that it is equivalent to the log-likelihood of $\sum_i \tilde{T}_i$ independent observations $\mathbf{y}_{it}$, $i = 1, ..., n$, $t = 1, ..., \tilde{T}_i$ in a multinomial logistic regression model with $J+1$ outcome categories. Thus, although the summands in (4) are not necessarily independent, the optimization of (4) (using software for multinomial logistic regression such as the `vglm` function in the R package VGAM or the `multinom` function in the R package nnet) yields maximum likelihood estimators of the parameters γ_{0tj} and γ_{jk}. In addition, the maximum likelihood estimators are consistent and asymptotically normal under the above assumptions (see section 5.8 of Kalbfleisch & Prentice, 2002). The latter properties give rise to Wald confidence intervals and likelihood ratio tests for the covariate effects.

The results of an illustrative analysis of the CRASH-2 data are presented in Table 2. For example, an increase in the Glasgow Coma Score by one unit resulted in an estimated 10% decrease in the odds for death due to bleeding and in an estimated 34% decrease in the odds for death due to head injury. Also, note that an increase in systolic blood pressure had a positive effect on the cause-specific hazard of death due to head injury while it decreased the cause-specific hazard of death due to bleeding.

During the past decades, the discrete cause-specific hazards model has been the predominant approach for modeling discrete-time competing-risks data. An early version of the model appeared, for example, in Yeh, Le, and McHugh (1984), where it was derived from a life table analysis with categorical covariates in the piecewise exponential modeling framework. Since the 1980s, numerous extensions of the discrete cause-specific hazards model have been proposed. For example, since the derivation of the likelihood function in (4) is not restricted to the logistic response function, Model (2) can easily be extended to other strictly monotone increasing response functions taking values in [0, 1], like the inverse complementary log–log function (e.g., McCall, 1996; Narendranathan & Stewart, 1993).

It should further be noted that an increased number of coefficients in Model (2) may result in numerical instabilities and may also affect inference (c.f. Jóźwiak & Moerbeek, 2012). This issue is particularly relevant in applications where the number of time intervals is large relative to the number of uncensored event times, so that some of the time intervals may contain very small event counts. To overcome this problem and to increase model stability, several authors have suggested to impose restrictions on the baseline coefficients, for example by assuming them to follow a polynomial function (McCall, 1996) or a spline function (Luo, Kong, & Nie, 2016; Berger & Schmid, 2018, implemented, for example, in SAS's `proc logistic` and in the `vgam` function of the R package VGAM). Narendranathan and Stewart (1993) specified a parametric function for the baseline coefficients that was motivated by the discretization of a Weibull distribution in continuous time. Möst, Pößnecker, and Tutz (2016) extended these approaches by developing a penalized estimation technique for smoothing the baseline coefficients that also incorporates variable selection (implemented in the R package MRSP). The penalty, which is subtracted from (4) for penalized maximum likelihood estimation, consists of a sum of the squared differences between neighboring baseline coefficients (or, alternatively, coefficients of neighboring spline basis functions) plus an L_1 penalty that enforces simultaneous inclusion or exclusion of a covariate from all cause-specific predictors. To improve variable selection, the authors used adaptive weights analogous to Zou (2006). An alternative approach was taken by Heyard et al. (2019), who carried out variable selection in a Bayesian framework

TABLE 2 Analysis of the CRASH-2 data

	Bleeding		Head injury		Other	
	$\exp(\gamma_1)$	95% CI	$\exp(\gamma_2)$	95% CI	$\exp(\gamma_3)$	95% CI
Gender: F	0.941	[0.745; 1.189]	1.388	[1.146; 1.682]	0.782	[0.603; 1.015]
Age	1.013	[1.007; 1.018]	1.017	[1.012; 1.022]	1.020	[1.015; 1.026]
Itime	0.890	[0.850; 0.932]	0.904	[0.873; 0.937]	0.971	[0.930; 1.014]
Itype:p	1.195	[0.978; 1.460]	0.571	[0.434; 0.751]	0.990	[0.789; 1.243]
Itype:bp	0.960	[0.746; 1.234]	1.009	[0.824; 1.236]	0.894	[0.684; 1.169]
Sbp	0.967	[0.963; 0.971]	1.005	[1.002; 1.007]	0.987	[0.983; 0.991]
Hr	1.004	[1.001; 1.008]	1.002	[0.999; 1.005]	1.006	[1.002; 1.010]
Rr	1.034	[1.024; 1.044]	1.016	[1.007; 1.025]	1.021	[1.010; 1.033]
CCrt	1.083	[1.035; 1.133]	1.034	[0.992; 1.079]	1.101	[1.050; 1.155]
Gcs	0.900	[0.882; 0.919]	0.660	[0.646; 0.675]	0.884	[0.864; 0.903]

The table presents the estimates of the regression coefficients (with 95% confidence intervals) obtained from fitting a discrete cause-specific hazards model with logistic link function to two-thirds of the analysis data. The target events were "death due to bleeding", "death due to head injury", and "death from other causes". The baseline coefficients γ_{0tj} were replaced by cubic functions of the form $\gamma_{0j}(t) = \gamma_{01j} + \gamma_{02j} \cdot t + \gamma_{03j} \cdot t^2 + \gamma_{04j} \cdot t^3, j = 1, ..., J$, to reduce the number of model parameters. The 95% confidence intervals refer to Wald intervals that were obtained from Fisher scoring (vglm function in R package VGAM). Their use is justified by the equivalence between the second derivative of the log-likelihood in (4) and the second derivative of the log-likelihood of a multinomial regression model.

using posterior inclusion frequencies (implemented in the R package TBFmultinomial). As demonstrated by the authors, this approach can be refined to give cause-specific sets of selected covariates.

In addition to restricting the shape of the baseline coefficients, further simplifications can be achieved by fitting an ordered logistic regression model. This type of model may be useful in situations where the target events have a natural order (for example when considering the events "still jobless", "return to part-time employment," and "return to full-time employment" in the modeling of unemployment spells, cf. Tutz, 1995).

To model unobserved heterogeneity and clustered data, several authors have proposed to extend Model (2) by random-effects terms. These include Gaussian random effects (Steele, Goldstein, & Browne, 2004; Meggiolaro et al., 2017; implemented for example, in SAS's `proc nlmixed`) and real-valued random effects that follow an unspecified discrete distribution with finite support (for instance to represent subgroups of the individuals, implemented in Stata's `hsmlogit` command, Troncoso-Ponce, 2018). Details on the interpretation of random-effects terms in time-to-event models, which has to be undertaken with some care, have been given in Austin (2017). Time-dependent covariates, which can be incorporated into Model (2) with relative ease due to the binary structure in (3), have been considered, among others, in Luo et al. (2016), Möst et al. (2016), and Heyard et al. (2019). Models with time-varying covariate effects have been considered in McCall (1996). Fahrmeir and Wagenpfeil (1996) proposed a state-space approach to model time-varying covariate effects, for example via first-order random walks. As demonstrated by the authors, this approach can also be used for the smoothing of baseline coefficients.

Instead of maximizing the full multinomial log-likelihood in (4), one could alternatively consider each event separately and fit separate binary models for each cause-specific hazard function. This approach, which would be based on "collapsed" versions of the log-likelihood using a set of J "one-vs-all" models, has been considered, for example, in Narendranathan and Stewart (1993), McCall (1996), and Lee, Feuer, and Fine (2018). While the resulting estimates of $\lambda_j(t|\,\mathbf{x}_i)$ remain consistent (if one assumes that the J models are all specified correctly), it must be noted that the binary components of $\mathbf{y}_{it}$ in (3) are not independent. This dependence, which is not an issue in continuous-time analysis (where the likelihood factors into a separate component for each cause-specific hazard, Kalbfleisch & Prentice, 2002, Chapter 8), has to be accounted for when constructing confidence intervals and hypothesis tests from the combined set of models. To address this issue, Lee, Feuer, and Fine (2018) proposed a GEE approach that consistently estimates the variances of the model parameters. The authors argued that their approach may result in an efficiency loss but is less sensitive to model misspecification when compared to the full multinomial model in (2).

4 | DISCRETE SUBDISTRIBUTION HAZARD MODELS

A disadvantage of the discrete cause-specific hazards model defined in (2) is the lack of a one-to-one relationship between $\lambda_j(t|\,\mathbf{x}_i)$ and $F_j(t|\,\mathbf{x}_i)$. In fact, Equation (1) implies that all J cause-specific hazard functions are needed to model the cumulative incidence function $F_j(t|\,\mathbf{x}_i)$ of the j-th event. This is a drawback especially when the interest is only in one of the J events and when the number of competing events (and hence the number of coefficients to be estimated) is large. In particular, it becomes difficult to infer the effects of the covariates on the CIF, as a change in the survival probability (with regard to $\mathbf{x}_i$) of the event of interest may be a result of a changed cause-specific hazard of a competing event (while the cause-specific hazard of the event of interest may remain unchanged).

In these situations, an alternative to the discrete cause-specific hazards model is the *discrete subdistribution hazard model* (Berger, Schmid, Welchowski, Schmitz-Valckenberg, & Beyersmann, 2020), which is a direct modeling approach for the cumulative incidence function of a specific event of interest. It is an extension of the respective modeling approach for continuous event times by Fine and Gray (1999) and has the advantage that only one model needs to be considered for interpretation of the covariate effects on event occurrence. In the following, it will be assumed w.l.o.g. that the event of interest is defined by $j = 1$. The discrete subdistribution hazard model links the cumulative incidence function $F_1(t|\,\mathbf{x}_i)$ to the *subdistribution time*

$$\vartheta_i := \begin{cases} T_i, \text{ if} \epsilon_i = 1, \\ \infty, \text{ if} \epsilon_i \neq 1. \end{cases} \tag{5}$$

By definition, ϑ_i measures the time to the occurrence of the type-1 event. Specifically, it assumes that the type-1 event will never be the first event to be observed once a competing event has occurred (Fine & Gray, 1999), implying

that there is no finite event time for the occurrence of a type-1 event if $\epsilon_i \neq 1$. Accordingly, the *discrete subdistribution hazard*, which is defined by

$$\xi_1(t|\mathbf{x}_i) := P(T_i = t, \epsilon_i = 1 | (T_i \geq t) \cup (T_i \leq t-1, \epsilon_i \neq 1), \mathbf{x}_i) = P(\vartheta_i = t | \vartheta_i \geq t, \mathbf{x}_i), \tag{6}$$

$t = 1, ..., q$, represents the discrete hazard function of the subdistribution time ϑ_i. Comparing $\xi_1(t)$ to the cause-specific hazard $\lambda_1(t)$, it follows with a little algebra that $\lambda_1(t) = \xi_1(t) \cdot (1 + P(T_i \leq t-1, \epsilon_i \neq 1)/P(T_i \geq t))$. This relationship, which is analogous to a relationship for continuous event times derived by Beyersmann and Schumacher (2007), implies that $\xi_1(t)$ is usually smaller than $\lambda_1(t)$. Conceptually, the subdistribution hazard may be considered as an approach to model the *occurrence* probability of the target event whereas the cause-specific hazards may be understood as the driving forces of competing risks data, providing, for example, insight in disease dynamics.

With a little algebra it can be shown that the subdistribution hazard $\xi_1(t|\,\mathbf{x}_i)$ is linked to $F_1(t|\,\mathbf{x}_i)$ by $F_1(t|\mathbf{x}_i) = 1 - \prod_{s=1}^{t}(1 - \xi_1(s|\mathbf{x}_i)) = 1 - S_1(t|\mathbf{x}_i)$, where $S_1(t|\,\mathbf{x}_i) = P(\vartheta_i > t|\,\mathbf{x}_i)$. Consequently, there is a one-to-one relationship between $\xi_1(t|\,\mathbf{x}_i)$ and $F_1(t|\,\mathbf{x}_i)$, and the effects of the covariates on $\xi_1(t|\,\mathbf{x}_i)$ have a direct interpretation in terms of increases or decreases in the CIF of a type-1 event. A parametric model for $\xi_1(t|\,\mathbf{x}_i)$ is given by

$$\xi_1(t|\mathbf{x}_i) = \frac{\exp\left(\beta_{0t1} + \mathbf{x}_i^\top \beta_1\right)}{1 + \exp(\beta_{0t1} + \mathbf{x}_i^\top \beta_1)}, \tag{7}$$

$t = 1, ..., q-1$, where the parameters $\beta_{011}, ..., \beta_{0(q-1)1}$ represent the baseline coefficients and $\beta_1 := (\beta_{11}, ..., \beta_{p1})^\top$ is a vector of regression coefficients. Analogous to the cause-specific hazards model (2), the increase of covariate x_k, $k \in \{1, ..., p\}$, by one unit increases the odds $\xi_1(t|\,\mathbf{x}_i)/(1 - \xi_1(t|\,\mathbf{x}_i))$ by the factor $\exp(\beta_{k1})$. The baseline and regression coefficients in (7) can be obtained by a weighted maximum likelihood estimation scheme that is based on an estimate of the censoring survival function $\hat{G}(t) = \hat{P}(C_i > t)$ and on a vector of weights

$$w_{it} := \frac{\hat{G}(t-1)}{\hat{G}\left(\min\left(\tilde{T}_i, t\right) - 1\right)} \cdot \left(I\left(t \leq \tilde{T}_i\right) + I\left(\tilde{T}_i \leq t-1, \Delta_i \epsilon_i > 1\right)\right), \tag{8}$$

$t = 1, ..., q-1$, that are defined analogously to the respective weights for subdistribution hazard models in continuous time (Fine & Gray, 1999). Setting

$$\left(y_{i1}, ..., y_{i,\tilde{T}_i}, ..., y_{i,q-1}\right) = \begin{cases} (0, ..., 0, 1, 0, ..., 0), & \text{if } \Delta_i \epsilon_i = 1, \\ (0, ..., 0, 0, 0, ..., 0), & \text{if } \Delta_i \epsilon_i \neq 1, \end{cases} \tag{9}$$

it can be shown (Berger et al., 2020) that the maximization of the weighted log-likelihood

$$l_w = \sum_{i=1}^{n} \sum_{t=1}^{q-1} w_{it} \{ y_{it} \log(\xi_1(t|\mathbf{x}_i)) + (1 - y_{it}) \log(1 - \xi_1(t|\mathbf{x}_i)) \} \tag{10}$$

results in a consistent and asymptotically normal estimator of the parameters β_{0t1} and β_1. Analogous to the cause-specific hazards model, estimation can be carried out by using standard software for weighted binary regression. The weights w_{it} are included in (10) because they take into consideration that the censoring times C_i are unobserved for individuals with a competing event first ($\vartheta_i = \infty$). In fact, if $\vartheta_i = \infty$, these individuals continue to be at risk beyond $\tilde{T}_i$ until they eventually experience the censoring event. As a consequence, w_{it} corresponds to the estimated probability of individual i being still at risk at t, given that it is known to be at risk at least up to $\tilde{T}_i$. The estimates $\hat{G}(t)$, $t = 1, ..., q-1$, can be obtained by fitting a (possibly covariate-dependent) discrete hazard model for the censoring event. This model is obtained from the cause-specific hazards model (2) by replacing Δ_i with $1 - \Delta_i$ and setting $J = 1$, cf. Schmid, Tutz, and Welchowski (2018). The covariate-free estimation procedure for w_{it} is implemented in the `dataLongSubDist` function of the R package discSurv.

The results of an illustrative analysis of the CRASH-2 data are presented in Table 3. For example, an increase in the Glasgow Coma Score by one unit resulted in an estimated 7.8% decrease in the odds $\xi_1(t|\,\mathbf{x}_i)/(1-\xi_1(t|\,\mathbf{x}_i))$ for death due to bleeding. This effect is smaller than the respective effect on the cause-specific hazard for death due to bleeding in Table 2. Generally, the estimates in Table 3 are directly interpretable in terms of increases or decreases in the cumulative incidence function, see Figure 1b. For comparison purposes, Figure 1a also includes the estimated cumulative incidence function for death due to bleeding that was obtained from the cause-specific hazards Model (2) using Equation (1).

TABLE 3 Analysis of the CRASH-2 data

	Bleeding	
	exp(β_1)	**95% CI**
Gender: F	0.934	[0.739; 1.179]
Age	1.011	[1.005; 1.017]
Itime	0.898	[0.858; 0.941]
Itype:p	1.222	[0.999; 1.493]
Itype:bp	0.952	[0.740; 1.223]
Spb	0.968	[0.964; 0.972]
Hr	1.003	[0.999; 1.007]
Rr	1.032	[1.022; 1.042]
CCrt	1.082	[1.035; 1.132]
Gcs	0.922	[0.904; 0.941]

The table presents the estimates of the regression coefficients (with 95% confidence intervals) obtained from fitting a discrete subdistribution hazard model with logistic link function to two-thirds of the analysis data. The target event was "death due to bleeding". The baseline coefficients β_{0t1} were replaced by a cubic function of the form $\beta_{01}(t) = \beta_{011} + \beta_{021} \cdot t + \beta_{031} \cdot t^2 + \beta_{041} \cdot t^3$ to reduce the number of model parameters. The 95% confidence intervals refer to Wald intervals that were computed using variance estimates obtained from Fisher scoring (glm function in R). As demonstrated in Berger et al. (2020), these estimates provide a good approximation of the true variance of the weighted maximum likelihood estimator.

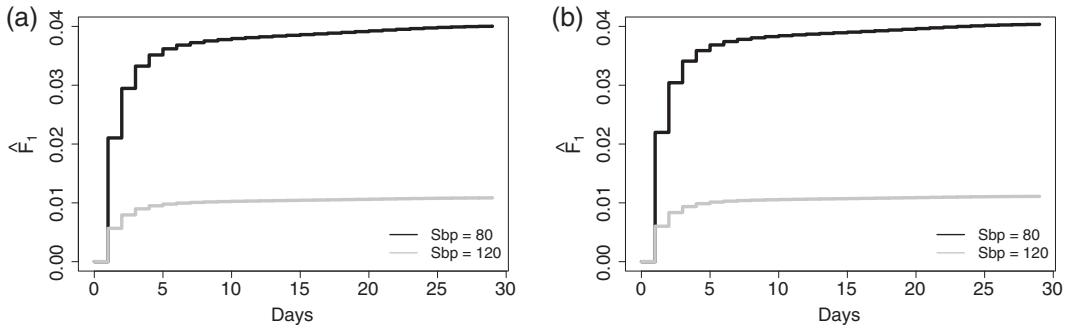

FIGURE 1 Analysis of the CRASH-2 data. The left panel presents the estimated CIF for death due to bleeding of a randomly selected female patient aged 41 years (Itime = 0.5, Itype = bp, Spb = 120, Hr = 70, Rr = 15, CCrt = 3, Gcs = 15), as obtained from the discrete cause-specific hazards model in Table 2. The right panel presents the corresponding CIF obtained from the discrete subdistribution hazard model in Table 3. Gray lines refer to the true sample data (Sbp = 120) whereas the black lines were obtained by switching the Sbp value from 120 to 80 mmHg. The estimates presented in the left panel were computed with the help of Formula 1, that is, by using all cause-specific hazard estimates shown in Table 2. The estimates presented in the right panel were computed with the help of the formula $F_1(t|\mathbf{x}_i) = 1 - \prod_{s=1}^{t}(1-\xi_1(s|\mathbf{x}_i))$, see also p. 8. Hence the calculations for the left panel involved three times as many coefficients as the calculations for the right panel, which were based on the values in Table 3

Analogous to the cause-specific hazards model, Model (7) can be extended in various ways. For example, it is possible to impose smoothness constraints on the baseline coefficients, to specify penalties for variable selection, and to include random effects terms and/or time-varying coefficients in the model equation. In addition, one could use other response functions, for example, the inverse complementary log–log function resulting in the *Gompertz model* or *complementary log–log model*. A convenient property of this model is its equivalence to the Cox proportional hazards model in continuous time. As shown, for example, in Tutz and Schmid (2016), the Gompertz model holds when continuous time-to-event data satisfying the proportional hazards assumption have been grouped.

It should finally be noted that the inclusion of time-dependent covariates in Model (7) is usually difficult. This is because the weighted log-likelihood in (10) requires the time-dependent values of $\mathbf{x}_i$ to be known up to time point $q-1$ (at least for those individuals with $w_{it} > 0$) and thus possibly beyond the observed event times $\tilde{T}_i$. When the covariates are not external, the latter requirement is often unrealistic or even impossible to meet, for example, in the analysis of causes of death. In particular, the CIF cannot be written as a function of the hazards when the model includes random (internal) time-dependent covariates (Cortese & Andersen, 2010). Finding an adequate strategy for the analysis of such covariates in the subdistribution hazard modeling framework remains a challenging problem (Poguntke, Schumacher, Beyersmann, & Wolkewitz, 2018).

5 | NONPARAMETRIC ESTIMATION WITH TREES

The modeling approaches described in Sections 3 and 4 are based on the assumption that the predictors $\mathbf{x}_i^\top \gamma_j$ and $\mathbf{x}_i^\top \beta_1$ are linear functions of the covariates. Although this assumption facilitates interpretation of the models, it may happen that linear predictors do not capture the true data structure very well. This is, for example, the case when interactions between covariates are present or when the relationship(s) between the hazard(s) and the predictors are highly nonlinear. In the latter case, it is often convenient to extend the linear predictors by smooth nonlinear functions in the covariates, for example, by polynomials or splines. Due to the structure of the likelihood functions in (4) and (10), this approach results in the fitting of a generalized additive model for multicategorical or binary outcome data.

When covariates affect the predictors in (2) and (7) in more than just an additive way (e.g., via unknown interactions that are difficult to pre-specify), an alternative strategy is to apply recursive partitioning methods, also termed "tree" methods. A popular tree method is the Classification and Regression Trees (CART) approach, which sequentially subdivides the covariate space into a set of disjoint rectangles. In each rectangle, a simple model (e.g., a constant) is fitted. These simple models are then combined to give nonparametric estimates of the cause-specific hazards or the subdistribution hazard. To find the best partition of the covariate space, the CART algorithm starts with a "root node" representing the whole covariate space. From the root node, the tree is grown in a hierarchical way by generating "children nodes", that is, by recursively splitting the covariate space into two smaller subsets. In each split, a single explanatory variable x_k (or the categorical time variable t, see below) is selected to define the split rule. Generally, the split rules depend on the measurement scales of x_k: For continuous x_k, the covariate subspace $\mathcal{M}$ represented by a node is split into two children nodes of the form $\mathcal{M}_{\text{left}} := \mathcal{M} \cap \{x_k \le c\}$ and $\mathcal{M}_{\text{right}} := \mathcal{M} \cap \{x_k > c\}$, where c is a suitably defined split point. For a multicategorical variable without ordering, that is, $x_k \in \{1, ..., R\}$, $R \in \mathbb{N}$, the subspaces represented by the children nodes have the form $\mathcal{M}_{\text{left}} = \mathcal{M} \cap \{x_k \in \mathcal{C}_1\}$ and $\mathcal{M}_{\text{right}} = \mathcal{M} \cap \{x_k \in \mathcal{C}_2\}$, where $\mathcal{C}_1$ and $\mathcal{C}_2$ are disjoint non-empty subsets defined by $\mathcal{C}_1 \subset \{1, ..., R\}$ and $\mathcal{C}_2 = \{1, ..., R\} \backslash \mathcal{C}_1$, respectively.

Since the cause-specific hazards and subdistribution hazard models in (2) and (7) can be represented by multinomial and (weighted) binary regression models with independent variables y_{itj} and y_{it}, respectively, it is convenient to base tree algorithms for discrete competing risks data on CART approaches for multicategorical or binary outcomes. Furthermore, to model the dependency of the hazard functions on time, it is convenient to consider time as an additional ordered candidate variable for splitting. In each node, the best split rule and split point are then found by minimizing a suitably defined impurity criterion. In the following we will describe the cause-specific hazards modeling approach by Berger, Welchowski, Schmitz-Valckenberg, and Schmid (2019), who proposed to use impurity criteria for classification trees with a multicategorical outcome variable for tree building. The data structure for these trees ("person-period format", Scott & Kennedy, 2005) is represented by $\sum_i \tilde{T}_i$ data lines, in the following termed "observations". In a node representing the subspace $\mathcal{M}$ one may consider, for instance, the *Gini impurity* measure defined by

$$GI(\mathcal{M}) = \sum_{j=0}^{J} \sum_{j \neq l} p_j(\mathcal{M}) \cdot p_l(\mathcal{M}), \tag{11}$$

where $p_j(\mathcal{M})$ denotes the proportion of observations with outcome category j falling in $\mathcal{M}$. (Analogous to Model (2), $j = 0$ defines the reference category, that is, the first entries in the vectors $\mathbf{y}_{it}$ in (3).) Then, in each step of the tree-building algorithm, the idea is to choose the split rule (among all covariates and subspaces of $\mathcal{M}$) that minimizes the pooled Gini impurity $GI_{\text{pooled}}(\mathcal{M}_{\text{left}}, \mathcal{M}_{\text{right}}) := |\mathcal{M}_{\text{left}}| \cdot GI(\mathcal{M}_{\text{left}}) + |\mathcal{M}_{\text{right}}| \cdot GI(\mathcal{M}_{\text{right}})$, where $|\cdot|$ denotes the number of observations falling in a node. When tree building is stopped (to give a set of *terminal nodes*), the relative frequencies of the outcome categories (= target events) in the terminal nodes yield estimates of the cause-specific hazards. In particular, since the time t is a candidate variable for splitting, each terminal node represents a specific interval on the discrete time scale, allowing the tree algorithm to model interactions between the covariates and time. Finally, a prediction of the cause-specific hazard functions of a new individual with covariate values $\mathbf{x}_{\text{new}}$ is obtained by (a) combining each time point $t = 1, ..., q$ with $\mathbf{x}_{\text{new}}$, (b) dropping the vectors $\left(t, \mathbf{x}_{\text{new}}^{\top}\right)$ down the tree, and (c) using the relative frequencies of target events in the terminal nodes attained by $\left(t, \mathbf{x}_{\text{new}}^{\top}\right)$ as predictions of the discrete cause-specific hazard functions.

An important tuning parameter of CART is the tree size, which is usually optimized using pruning techniques. In case of trees for discrete cause-specific hazard functions, controlling the tree size ensures a sufficiently large number of observations in each terminal node. This prevents the resulting cause-specific hazard estimates from having a too large variance (which is inversely related to the terminal node size, see Schmid, Küchenhoff, Hoerauf, & Tutz, 2016). Berger et al. (2019) therefore considered the minimum number of observations in the terminal nodes ("minimum node size") as the main parameter for pruning. The minimum node size can be determined by either cross-validation of the log-likelihood (4) or by information criteria such as AIC and BIC, which are defined in terms of weighted sums of the log-likelihood and the number of terminal nodes. An implementation of this procedure is available on GitHub, see Berger et al. (2019).

It must be emphasized that the tree-building framework described above is generic in the sense that it is not restricted to a particular impurity measure or a particular algorithm for the selection of split rules and/or split points. For example, one could use an alternative criterion for node splitting (such as Hellinger's distance, which is an appropriate criterion for unbalanced sets of outcome values, such as the ones defined in (3)). Instead of CART, one could also employ the conditional inference framework by Hothorn, Hornik, and Zeileis (2006) or a likelihood-based rule (Bou-Hamad et al., 2009) for selecting the split variables. Furthermore, the tree-based estimates of $\lambda_j(t|\ \mathbf{x}_i)$ can be smoothed using a Laplace correction, which prevents the estimates from becoming too close to either zero or one. More specifically, the *Laplace-corrected* cause-specific hazard estimate in a terminal node $\mathcal{M}$ is given by $\hat{\lambda}_j(\mathcal{M}) = \left(p_j(\mathcal{M}) \cdot |\mathcal{M}| + 1\right) / (|\mathcal{M}| + J + 1)$. A tree-building framework for the nonparametric estimation of subdistribution hazards can be obtained analogously, for instance by running the CART algorithm with a binary outcome variable as defined in (9) and the observation weights w_{it} defined in (8).

As an example, Figure 2 visualizes the CART-based procedure by Berger et al. (2019) when applied to the CRASH-2 data. In the course of tree building, all nine available covariates were selected to estimate the cause-specific hazards for "death due to bleeding", "death due to head injury", and "death due to other causes". After pruning with the BIC criterion, the final tree had 45 terminal nodes. The resulting cumulative incidence function for the target event "death due to bleeding" is illustrated in Figure 2 (c).

6 | GOODNESS-OF-FIT AND CALIBRATION

The appropriateness of a modeling approach for discrete competing-risks data can be investigated by assessing the *goodness-of-fit* and the *calibration* of a fitted cause-specific hazards or subdistribution hazard model. Usually, the term "goodness-of-fit" refers to the performance of a model on the same data that were used for fitting the model, whereas "calibration" refers to the generalization performance of the model on a set of independent validation data. The graphical methods presented in this section can be used for assessing both goodness-of-fit and calibration. In the following it will be assumed that the focus is on calibration, that is, that model performance is assessed using an independent i.i.d. validation sample containing data $\left(\tilde{T}_m, \Delta_m, \epsilon_m, \mathbf{x}_m^{\top}\right)$, $m = 1, ..., N$, collected from N individuals. Individual-specific predictions of the discrete cause-specific or subdistribution hazards are obtained by applying the fitted model (either parametric or nonparametric) to the validation data.

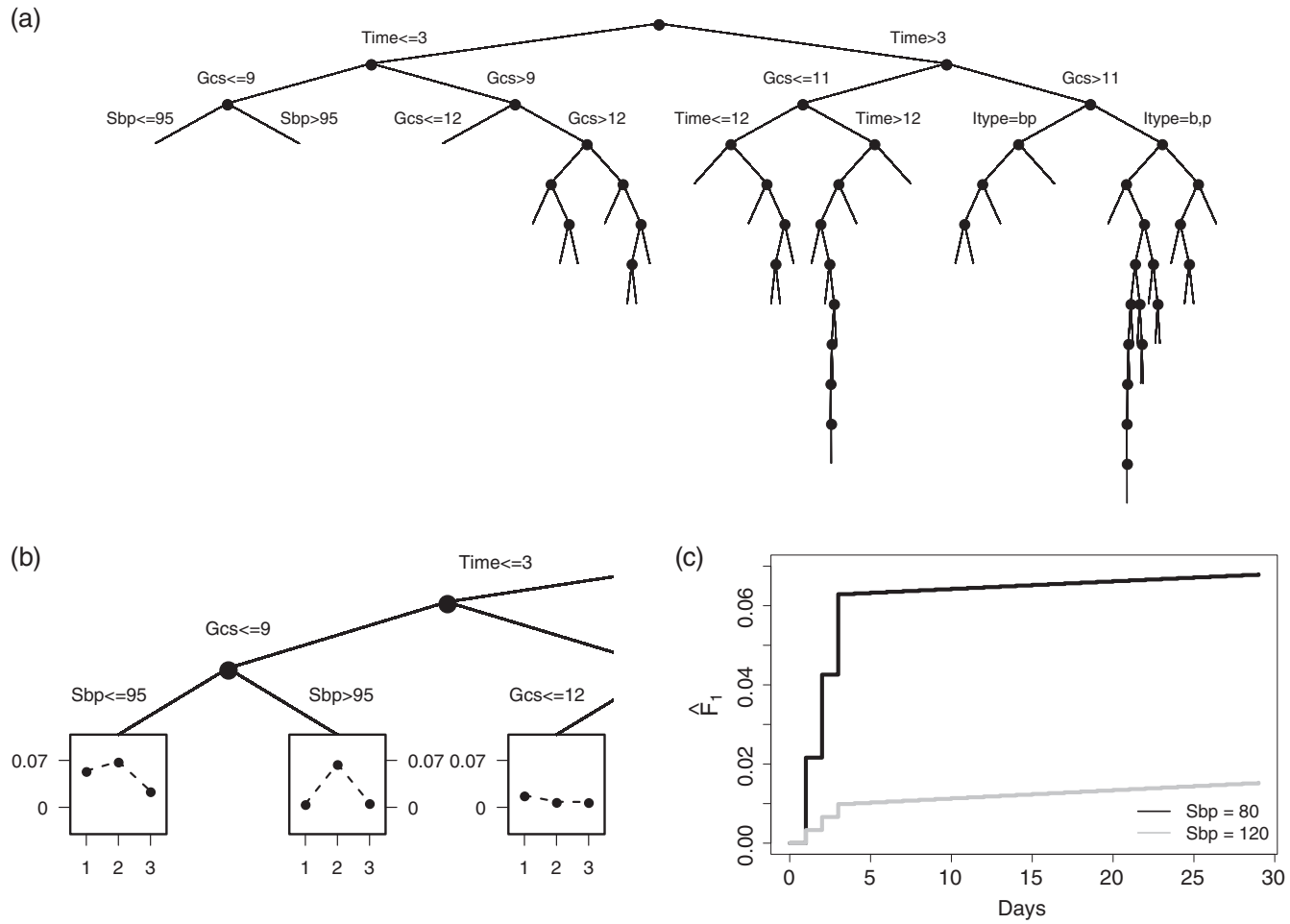

FIGURE 2 Analysis of the CRASH-2 data. The figure visualizes the results of the CART-based procedure by Berger et al. (2019) when applied to two-thirds of the analysis data. Hellinger's distance was used for node splitting. Panel (a) depicts the tree structure that resulted from pruning with the BIC. Part of the left branch of the tree (corresponding to time points $t \leq 3$) is illustrated in Panel (b). Estimates of the cause-specific hazards are presented in each of the terminal nodes (1 = death due to bleeding, 2 = death due to head injury, 3 = death due to other causes). The tree-based CIF estimate (referring to death due to bleeding) for the example patient from Figure 1 is visualized in Panel (c)

First consider the discrete cause-specific hazards model. The calibration of this model can be investigated graphically by checking how well the predicted discrete cause-specific hazard values agree with their corresponding observed proportions in the validation data. For each target event (keeping $j \in \{1, ..., J\}$ fixed), the idea underlying this method is to split the validation data into G subsets D_{gj}, $g = 1, ..., G$, defined by the empirical quantiles of the *predicted* cause-specific hazards $\hat{\lambda}_j(t|\mathbf{x}_m)$, $m = 1, ..., N$, $t = 1, ..., \tilde{T}_m$. Common choices are $G = 10$ or $G = 20$. In each of the G subsets, the predicted cause-specific hazards $\hat{\lambda}_j(t|\mathbf{x}_m)$ are averaged and plotted versus the *empirical* cause-specific hazards of the j-th target event. The latter are defined by $p_{gj} := \sum_{m,t:\hat{\lambda}_j(t|\mathbf{x}_m) \in D_{gj}} y_{mtj} / |D_{gj}|$, $g = 1, ..., G$, where $|D_{gj}|$ corresponds to the number of observations in subset D_{gj}. Put differently, the empirical cause-specific hazards are given by the relative frequencies of the j-th target event in the G subsets. For each j, a well calibrated cause-specific hazards model is indicated by a set of points that are close to the 45° line. Note that the methodology underlying the calibration plot is generic in the sense that it does not depend on the modeling approach that was used to predict the cause-specific hazards $\hat{\lambda}_j(t|\mathbf{x}_m)$.

As an example, Figure 3 contains a calibration plot that was obtained from the parametric discrete cause-specific hazards model reported in Table 2. The dots are close to the 45° line, indicating good calibration.

In addition to the graphical checks described above, calibration can be further investigated by fitting a multinomial logistic recalibration model. This model, which jointly considers all cause-specific hazards, is defined by

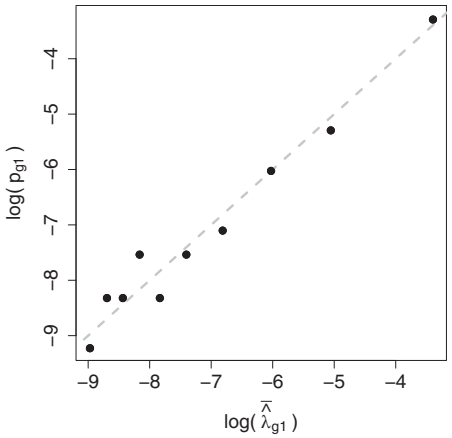

FIGURE 3 Analysis of the CRASH-2 data. The figure depicts a calibration plot for the predicted cause-specific hazard of the target event "death due to bleeding" (log scale). Predicted cause-specific hazard values were obtained by applying the model in Table 2 to the 6,346 patients that were not used for model fitting (one-third of the analysis data set). The plot is based on $G = 10$ subsets of the patients. A constant value of 0.0001 was added to the predicted and observed cause-specific hazard values before application of the log transformation. The p-value of the likelihood ratio test (joint null hypothesis "$H_0 : a_j = 0$, $b_{jj} = 1\ \forall\, j$", Heyard et al., 2020) was 0.247, indicating no significant deviations from a well-calibrated model

$$\log\left(\frac{\lambda_j(t|\mathbf{x}_m)}{\lambda_0(t|\mathbf{x}_m)}\right) = a_j + \sum_{r=1}^{J} b_{rj}\eta_{mtr} \tag{12}$$

with $\eta_{mtr} := \log\left(\hat{\lambda}_r(t|\mathbf{x}_m)/\left(1-\hat{\lambda}_r(t|\mathbf{x}_m)\right)\right)$ and $b_{rj} = 0$ if $r \neq j$. The intercepts a_j measure whether the predicted cause-specific hazards are systematically too high ($a_j < 0$) or too low ($a_j > 0$). Analogously, the slopes b_{jj} indicate whether the predicted cause-specific hazards do not show enough variation ($b_{jj} > 1$), vary too much ($0 < b_{jj} < 1$), or show the wrong general direction ($b_{jj} < 0$). A perfectly calibrated model should thus have $a_j = 0$ and $b_{jj} = 1$ for all j. Note that Model (12) can be fitted in the same way as the multinomial regression model in Section 3, but with the additional restriction that $b_{rj} = 0$ for all $r \neq j$. The latter constraint is needed to ensure that the cause-specific hazard of the j-th target event does not affect the calibration of the r-th cause-specific hazard model unless $r = j$. In the framework of this model, an overall check for calibration is given by a likelihood ratio test of the joint null hypothesis "$H_0 : a_j = 0,\ b_{jj} = 1\,\forall j$", cf. Heyard et al. (2020). Unlike calibration plots, which can be used to either assess goodness-of-fit or calibration, multinomial logistic recalibration models are only meaningful when fitted to a set of validation data that are independent of the data used to fit the discrete cause-specific hazards model.

Calibration checks for the discrete subdistribution hazard model have been developed by Berger and Schmid (2020). To generate a calibration plot, the authors proposed to compute predictions of the subdistribution hazard $\hat{\xi}_1(t|\mathbf{x}_m)$ for all $m \in \{1, ..., N\}$ and all $t \in \{1, ..., q-1\}$. These predictions can be split into G subsets D_g, $g = 1, ..., G$, in the same way as described above. Based on the subsets, the predicted and empirical subdistribution hazards can be defined by

$$\bar{\hat{\xi}}_{g1} = \frac{1}{\sum_{m,t:\hat{\xi}_1(t|\mathbf{x}_m)\in D_g} w_{mt}} \sum_{m,t:\hat{\xi}_1(t|\mathbf{x}_m)\in D_g} \hat{\xi}_1(t|\mathbf{x}_m) w_{mt} \text{ and } p_g = \frac{1}{\sum_{m,t:\hat{\xi}_1(t|\mathbf{x}_m)\in D_g} w_{mt}} \sum_{m,t:\hat{\xi}_1(t|\mathbf{x}_m)\in D_g} y_{mt} w_{mt}, \tag{13}$$

respectively, where w_{mt} and y_{mt} denote the weights and the binary values specified in (8) and (9), respectively. Analogous to the multinomial model in (12), a recalibration model for the predicted subdistribution hazards can be defined by a logistic regression model with observation weights w_{mt}.

7 | OTHER TOPICS

In addition to the modeling and estimation approaches described in Sections 3 to 6, numerous other methods for discrete-time competing-risks modeling have been proposed. This section presents a brief overview of these methods.

7.1 | Alternative modeling approaches

Analogous to the life table estimator for single target events, it is possible to obtain covariate-free estimates of the discrete cause-specific hazard and cumulative incidence functions by applying nonparametric maximum likelihood estimation techniques (Lee, 2017). As an alternative to the discrete subdistribution hazard model, Lee (2017) further proposed to model the set of cumulative incidence functions jointly by means of a transformation model. For this, the author expressed the log-likelihood function in terms of products of differences of the cumulative incidence functions. A nonparametric test for the (covariate-free) comparison of cumulative incidence functions has been developed by Sreedevi, Sankaran, and Dhanavanthan (2014). Han and Hausman (1990) proposed an alternative modeling approach for discretized versions of the continuous time variables U_{ij}, $j = 1, ..., J$, $i = 1, ..., n$, where U_{ij} is defined as the (latent) duration to the j-th target event and $T_i = \min(U_{i1}, ..., U_{iJ})$. Within this framework, the authors developed a likelihood-based modeling approach that yields estimates of the hazards of the (possibly correlated) duration times U_{ij}. Similar to (2), these estimates can be decomposed into a set of baseline coefficients and a predictor of the form $\mathbf{x}_i^\top \gamma_j$. In contrast to the model in Section 3, however, the model by Han and Hausman (1990) is defined in terms of the hazards of the latent variables U_{ij} and not in terms of the cause-specific hazards $\lambda_j(t|\,\mathbf{x}_i)$ that condition on overall survival up to t. Generally, the latent variable approach requires additional assumptions about the data-generating process to address identifiability and interpretability issues (Beyersmann et al., 2011). Lee, Feuer, and Fine (2018) considered a discrete cause-specific hazards model under left truncation. For this they specified an adjusted log-likelihood function that is essentially obtained by removing the terms corresponding to time points earlier than the truncation time from (4). The authors also incorporated different time scales in their model, which, for example, may be convenient in cancer research where the target events "death due to cancer" and "death from other causes" are often measured on the scales "time from diagnosis" and "time since birth", respectively. Heyard et al. (2019) proposed a landmarking approach for the discrete cause-specific hazards model that enables dynamic predictions (i.e., the prediction of events conditional on being at risk until a fixed point in time). Their approach has been implemented in the R package TBFmultinomial. Barnett et al. (2009) and Vallejos and Steel (2017) proposed Bayesian discrete cause-specific hazards models with suitably specified priors for the baseline and regression coefficients.

7.2 | Alternative methods for model validation

To evaluate the calibration of CIF estimates, Lee (2017) generated a plot that compared the model-based predictions of the CIF at various time points to the respective nonparametric CIF estimates. In addition to assessing cause-specific calibration (Section 6), Heyard et al. (2020) proposed measures for the evaluation of discrimination accuracy and prediction error of a discrete cause-specific hazards model. In particular, the authors developed estimators for cause-specific versions of the discrete time-dependent area under the curve and for the discrete-time concordance index. The authors further showed how to linearly combine the cause-specific concordance indices to give a global concordance measure. To quantify cause-specific prediction error, Heyard et al. (2020) estimated an average of the individual-specific squared differences between the predicted and observed cumulative incidence functions, where the latter were defined in terms of binary variables indicating whether the individual experienced a type-j event at or before t or not. An alternative measure of prediction error is the predictive deviance, which is obtained by evaluating twice the negative of the multinomial log-likelihood in an independent test sample (Janitza & Tutz, 2015).

7.3 | Machine learning techniques for discrete time-to-event data with competing target events

Sparapani, Logan, McCulloch, and Laud (2016) proposed the Bayesian Additive Regression Trees (BART) approach for obtaining nonparametric predictions of the discrete cumulative incidence functions. Generally, the BART method employs an ensemble of Bayesian regression trees with suitably specified priors for the algorithmic components that define the tree structure. Sparapani et al. (2016) adapted this method to the discrete-time competing risks case by optimizing separate models for parts of the multinomial log-likelihood that they defined in terms of binary outcome variables. After training the ensemble, a combination of the fitted models was used to construct predictions of the cumulative incidence functions. Janitza and Tutz (2015) proposed a random forest approach for the prediction of the

discrete cause-specific hazard functions. The approach is based on the multinomial representation of the model with binary outcome values in (3). Unlike Berger et al. (2019), the authors used the conditional inference framework (Hothorn et al., 2006) for tree building. Deep neural network methods for competing-risks modeling of discretized continuous time-to-event data have been considered in Lee, Zame, Yoon, and van der Schaar (2018) and Lee, Yoon, and van der Schaar (2020).

7.4 | Causal inference

Young, Stensrud, Tchetgen Tchetgen, and Hernán (2020) developed a framework for modeling causal treatment effects in the presence of competing events. Within this framework, they defined various estimands that are based directly on the discrete hazard and cumulative incidence functions presented in Sections 3 and 4. One of the main results derived by Young et al. (2020) is that differences in cause-specific and subdistribution hazards cannot be interpreted as causal effects, as hazards may differ simply because of differences in individuals who survive until a specific time point due to treatment effects before this time point (Young et al., 2020, p. 1204). On the other hand, causal treatment effects may be defined by differences in the respective cumulative incidence functions.

8 | DISCUSSION

The literature cited in this article demonstrates that discrete-time competing-risks analysis is still a very active field of research. Although the methods presented here are less well known than their "continuous" counterparts (Kalbfleisch & Prentice, 2002; Klein, 2010; Kleinbaum & Klein, 2012; Putter, Fiocco, & Geskus, 2007), numerous authors have emphasized their advantages. In particular, discrete-time methods often result in less bias and more accurate estimates when the number of tied event times is large (Hess & Persson, 2012). Also, they come with a high modeling flexibility, for example, in regards to nonproportional hazards and time-dependent covariates. Another advantage of discrete-time methods is that even the more complex modeling approaches are often implemented easily using standard software for multinomial or binary regression. Conceptually, discrete hazard functions have an intuitive probability interpretation that is often more accessible to applied researchers than the respective hazard functions in continuous time.

A general issue, which is not restricted to competing-risks analysis but also concerns time-to-event analysis with a single target event, is the distinction between discrete and continuous event times. While the methods presented in this article naturally apply to studies with an intrinsically discrete time scale, the choice of an appropriate (continuous or discrete) statistical model is considerably more difficult in situations where continuous times have been grouped or rounded. In the literature, the use of discrete-time methods has been generally recommended in situations where the number of ties is large and/or where the number of unique time measurements is small (e.g., Singer & Willett, 2003, Chapter 9). On the other hand, continuous-time methods may serve as a numerically convenient approximation of discrete-time methods in situations with a large number of unique time measurements. Currently, there is little guidance in the literature as to when such an approximation should be considered "appropriate" and when the number of ties should be considered "large" enough to recommend the use of discrete-time methods. We also note that even in the absence of grouping effects there may be reasons for researchers to divide a continuous time scale into a pre-specified number of discrete intervals and to use discrete-time methods for statistical analysis. This strategy, may, for instance, be driven by practical considerations (Austin, 2017) or may be required by the analysis technique itself (e.g., by neural network methods, see Gorgi Zadeh & Schmid, 2020 or Kvamme & Borgan, 2019, who provided guidance for the choice of an appropriate discretization scheme).

A remaining problem of discrete-time competing-risks methods is due to the size of the person-period data sets (also termed "augmented" data sets, Tutz & Schmid, 2016) that usually need to be generated before model fitting. If the number of discrete time points is large, the size of these data sets (consisting of up to $n \cdot q$ observations) may reduce the computational performance of the aforementioned modeling approaches. Furthermore, discrete-time competing-risks modeling may be affected by the heavy imbalance of the binary outcome values in (3) and (9). It should also be noted that the literature currently provides no unified theoretical treatment of the discrete-time competing-risks methods presented in this article. While Kalbfleisch and Prentice (2002) present a theory for cause-specific hazards modeling that is based on the counting process formulation by Andersen, Borgan, Gill, and Keiding (1993) (covering both

continuous and discrete event times, also for less restrictive censoring schemes than random censoring, cf. Chapters 5 and 8 of Kalbfleisch & Prentice, 2002), the asymptotics of the subdistribution hazard modeling approach presented in Section 4 have been established within a different framework involving unbiased estimation equations. On the other hand, a theoretical investigation of the properties of the tree-based methods presented in Section 5 (in particular, with respect to the dependence structure of the binary variables y_{itj}) does not exist to date. More research is warranted to embed the modeling approaches presented in this article in a common theoretical framework.

We finally note that the methodology presented in Sections 3 to 5 can be generalized in a number of ways. For example, it is natural to extend the cause-specific hazards model from Section 3 to situations involving multiple states and recurrent events. This approach has been taken, for example, by Enberg, Gottschalk, and Wolf (1990) and Steele et al. (2004). If the discrete-time points represent intervals $[a_{t-1}, a_t)$ on a continuous time scale, another extension is to allow for more flexible (possibly random) interval boundaries. This approach refers to the more general methodology for competing-risks analysis with interval-censored data (see for example, the references in Chen, Sun, & Peace, 2012 or Li, 2016).

ACKNOWLEDGMENTS

We thank two anonymous referees for very helpful comments and suggestions. Support by the German Research Foundation (DFG), grant SCHM 2966/2-1, is gratefully acknowledged. Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS

Matthias Schmid: Conceptualization; formal analysis; methodology; writing-original draft. **Moritz Berger:** Conceptualization; formal analysis; methodology; writing-original draft.

ORCID

Matthias Schmid https://orcid.org/0000-0002-0788-0317

RELATED WIREs ARTICLE

Competing risks

REFERENCES

Allison, P. D. (2014). *Event history and survival analysis* (2nd ed.). Los Angeles, CA: SAGE Publications.

Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (Eds.). (1993). *Statistical models based on counting processes*. New York, NY: Springer.

Austin, P. C. (2017). A tutorial on multilevel survival analysis: Methods, models and applications. *International Statistical Review*, *85*, 185–203.

Barnett, A. G., Batra, R., Graves, N., Edgeworth, J., Robotham, J., & Cooper, B. (2009). Using a longitudinal model to estimate the effect of methicillin-resistant *Staphylococcus aureus* infection on length of stay in an intensive care unit. *American Journal of Epidemiology*, *170*, 1186–1194.

Berger, M., & Schmid, M. (2018). Semiparametric regression for discrete time-to-event data. *Statistical Modelling*, *18*, 322–345.

Berger, M. & Schmid, M. (2020). *Assessing the calibration of subdistribution hazard models in discrete time*. Technical Report arXiv:2001.11240 [stat.ME].

Berger, M., Schmid, M., Welchowski, T., Schmitz-Valckenberg, S., & Beyersmann, J. (2020). Subdistribution hazard models for competing risks in discrete time. *Biostatistics*, *21*, 449–466.

Berger, M., Welchowski, T., Schmitz-Valckenberg, S., & Schmid, M. (2019). A classification tree approach for the modeling of competing risks in discrete time. *Advances in Data Analysis and Classification*, *13*, 965–990.

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

Beyersmann, J., & Schrade, C. (2017). Florence nightingale, William Farr and competing risks. *Journal of the Royal Statistical Society, Series A. Statistics in Society*, *180*, 285–293.

Beyersmann, J., & Schumacher, M. (2007). Misspecified regression model for the subdistribution hazard of a competing risk. *Statistics in Medicine*, *26*, 1649–1652.

Bou-Hamad, I., Larocque, D., Ben-Ameur, H., Masse, L., Vitaro, F., & Tremblay, R. (2009). Discrete-time survival trees. *The Canadian Journal of Statistics*, *37*, 17–32.

Box-Steffensmeier, J. M., & Jones, B. S. (2004). *Event history modeling*. Cambridge, MA: Cambridge University Press.

Chen, D.-G., Sun, J., & Peace, K. E. (Eds.). (2012). *Interval-censored time-to-event data*. Boca Raton, FL: Chapman & Hall/CRC.

Cortese, G., & Andersen, P. K. (2010). Competing risks and time-dependent covariates. *Biometrical Journal*, *52*, 138–158.

CRASH-2 Trial Collaborators. (2010). Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): A randomised, placebo-controlled trial. *The Lancet*, *376*, 23–32.

Enberg, J., Gottschalk, P., & Wolf, D. (1990). A random-effects logit model of work-welfare transitions. *Journal of Econometrics*, *43*, 63–75.

Fahrmeir, L., & Wagenpfeil, S. (1996). Smoothing hazard functions and time-varying effects in discrete duration and competing risks models. *Journal of the American Statistical Association*, *91*, 1584–1594.

Fehring, R., Schneider, M., Raviele, K., Rodriguez, D., & Pruszynski, J. (2013). Randomized comparison of two internet-supported fertility-awareness-based methods of family planning. *Contraception*, *88*, 24–30.

Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, *94*, 496–509.

Gorgi Zadeh, S., & Schmid, M. (2020). *Bias in cross-entropy-based training of deep survival networks*. Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: https://doi.org/10.1109/TPAMI.2020.2979450.

Han, A., & Hausman, J. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, *5*, 1–28.

Hess, W., & Persson, M. (2012). The duration of trade revisited—Continuous-time versus discrete-time hazards. *Empirical Economics*, *43*, 1083–1107.

Heyard, R., Timsit, J.-F., Essaied, W., Held, L., & COMBACTE-MAGNET Consortium. (2019). Dynamic clinical prediction models for discrete time-to-event data with competing risks—A case study on the OUTCOMEREA database. *Biometrical Journal*, *61*, 514–534.

Heyard, R., Timsit, J.-F., Held, L., & COMBACTE-MAGNET Consortium. (2020). Validation of discrete time-to-event prediction models in the presence of competing risks. *Biometrical Journal*, *62*, 643–657.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*, 651–674.

Janitza, S., & Tutz, G. (2015). *Prediction models for time discrete competing risks*. Technical Report 177. Department of Statistics, Ludwig-Maximilians-Universität München.

Jóźwiak, K., & Moerbeek, M. (2012). Power analysis for trials with discrete-time survival endpoints. *Journal of Educational and Behavioral Statistics*, *37*, 630–654.

Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data* (2nd ed.). Hoboken, NJ: Wiley.

Klein, J. P. (2010). Competing risks. *WIREs Computational Statistics*, *2*, 333–339.

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis* (3rd ed.). New York, NY: Springer.

Kvamme, H.,& Borgan, O. (2019). *Continuous and discrete-time survival prediction with neural networks*. Technical Report arXiv:1910.06724 [stat.ML].

Lee, C., Yoon, J., & van der Schaar, M. (2020). Dynamic-DeepHit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Transactions on Biomedical Engineering*, *67*, 122–133.

Lee, C., Zame, W. R., Yoon, J., & van der Schaar, M. (2018). DeepHit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence* (pp. 2314–2321). Palo Alto: AAAI Press.

Lee, M. (2017). Inference for cumulative incidence on discrete failure times with competing risks. *Journal of Statistical Computation and Simulation*, *87*, 1989–2001.

Lee, M., Feuer, E., & Fine, J. (2018). On the analysis of discrete time competing risks data. *Biometrics*, *74*, 1468–1481.

Li, C. (2016). Cause-specific hazard regression for competing risks data under interval censoring and left truncation. *Computational Statistics & Data Analysis*, *104*, 197–208.

Luo, S., Kong, X., & Nie, T. (2016). Spline based survival model for credit risk modeling. *European Journal of Operational Research*, *253*, 869–879.

McCall, B. (1996). Unemployment insurance rules, joblessness, and part-time work. *Econometrica*, *64*, 647–682.

Meggiolaro, S., Giraldo, A., & Clerici, R. (2017). A multilevel competing risks model for analysis of university students' careers in Italy. *Studies in Higher Education*, *42*, 1259–1274.

Möst, S., Pößnecker, W., & Tutz, G. (2016). Variable selection for discrete competing risks models. *Quality & Quantity*, *50*, 1589–1610.

Narendranathan, W., & Stewart, M. (1993). Modelling the probability of leaving unemployment: Competing risks models with flexible baseline hazards. *Journal of the Royal Statistical Society, Series C*, *42*, 63–83.

Poguntke, I., Schumacher, M., Beyersmann, J., & Wolkewitz, M. (2018). Simulation shows undesirable results for competing risks analysis with time-dependent covariates for clinical outcomes. *BMC Medical Research Methodology*, *18*, 79.

Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, *26*, 2389–2430.

Scheike, T. H., & Keiding, N. (2006). Design and analysis of time-to-pregnancy. *Statistical Methods in Medical Research*, *15*, 127–140.

Schmid, M., Küchenhoff, H., Hoerauf, A., & Tutz, G. (2016). A survival tree method for the analysis of discrete event times in clinical and epidemiological studies. *Statistics in Medicine*, *35*, 734–751.

Schmid, M., Tutz, G., & Welchowski, T. (2018). Discrimination measures for discrete time-to-event predictions. *Econometrics and Statistics*, *7*, 153–164.

Scott, M. A., & Kennedy, B. B. (2005). Pitfalls in pathways: Some perspectives on competing risks event history analysis in education research. *Journal of Educational and Behavioral Statistics*, *30*, 413–442.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. Oxford: Oxford University Press.

Sparapani, R., Logan, B., McCulloch, R., & Laud, P. (2016). Nonparametric survival analysis using Bayesian additive regression trees (BART). *Statistics in Medicine*, *35*, 2741–2753.

Sreedevi, E., Sankaran, P., & Dhanavanthan, P. (2014). A nonparametric test for comparing cumulative incidence functions of current status competing risks data. *Journal of Statistical Theory and Practice*, *8*, 743–759.

Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics. *Statistical Modelling*, *4*, 145–159.

Steinberg, J. S., Göbel, A. P., Thiele, S., Fleckenstein, M., Holz, F. G., & Schmitz-Valckenberg, S. (2016). Development of intraretinal cystoid lesions in eyes with intermediate age-related macular degeneration. *Retina*, *36*, 1548–1556.

Troncoso-Ponce, D. (2018). Estimation of competing risks duration models with unobserved heterogeneity using hsmlogit. Available at SSRN: https://doi.org/10.2139/ssrn.3114159.

Tutz, G. (1995). Competing risks models in discrete time with nominal or ordinal categories of response. *Quality and Quantity*, *29*, 405–420.

Tutz, G., & Schmid, M. (2016). *Modeling discrete time-to-event data*. New York: Springer.

Vallejos, C. A., & Steel, M. F. J. (2017). Bayesian survival modelling of university outcomes. *Journal of the Royal Statistical Society, Series A*, *180*, 613–631.

Yeh, C., Le, C., & McHugh, R. (1984). Competing risk analysis for life table data with known observation times. *Biometrical Journal*, *26*, 111–118.

Young, J. G., Stensrud, M. J., Tchetgen Tchetgen, E. J., & Hernán, M. A. (2020). A causal framework for classical statistical estimands in failure-time settings with competing events. *Statistics in Medicine*, *39*, 1199–1236.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*, 1418–1429.

How to cite this article: Schmid M, Berger M. Competing risks analysis for discrete time-to-event data. *WIREs Comput Stat*. 2020;e1529. https://doi.org/10.1002/wics.1529