UNIVERSITY OF COPENHAGEN

Confidence bands for multiplicative hazards models

Flexible resampling approaches

Dobler, Dennis; Pauly, Markus; Scheike, Thomas H.

WILEY *Biometrics*
A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY

Confidence bands for multiplicative hazards models: Flexible resampling approaches

Dennis Dobler[1] | Markus Pauly[2] | Thomas H. Scheike[3]

[1]Department of Mathematics, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[2]Faculty of Statistics, Technical University of Dortmund, Dortmund, Germany

[3]Section of Biostatistics, University of Copenhagen, Copenhagen, Denmark

Correspondence
Dennis Dobler, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, Netherlands.
Email: d.dobler@vu.nl

Abstract

We propose new resampling-based approaches to construct asymptotically valid time-simultaneous confidence bands for cumulative hazard functions in multistate Cox models. In particular, we exemplify the methodology in detail for the simple Cox model with time-dependent covariates, where the data may be subject to independent right-censoring or left-truncation. We use simulations to investigate their finite sample behavior. Finally, the methods are utilized to analyze two empirical examples with survival and competing risks data.

KEYWORDS

competing risks, counting processes, Cox regression, hazards, Martingale theory, multistate models, survival analysis

1 | INTRODUCTION

Multiplier resampling (also known as the wild bootstrap) has evolved as a state-of-the-art choice for inference on cumulative incidences or hazards in nonparametric and semiparametric survival and multistate models in event history analysis; see Lin et al. (1993) and Lin et al. (1994) for Cox models and Lin (1997) for competing risks set-ups. The basic idea is to mimic martingale representations of the estimators by replacing nonobservable martingale residuals with randomly weighted counting processes. This approach has been extended in various directions, allowing for a greater flexibility (Beyersmann et al., 2013; Dobler & Pauly, 2014; Dobler et al., 2017) and multiple, possibly recurrent, states (Bluhmki et al., 2018, 2019). We will here focus on semiparametric regression models. The most used regression model in survival analysis is Cox's proportional hazard model (Cox, 1972). It is highly useful to predict survival probabilities based

on covariates. Such survival functions are typically provided with point-wise confidence intervals, and this is implemented in all major software packages. However, when interest is in the survival function as a whole, it is preferable to report it together with uniform confidence intervals. These so-called confidence bands describe the uncertainty of the whole survival function estimation. But this is infrequently done in practice because there are few programs that construct such uniform bands. In addition, apart from only few exceptions such as Lin (1997), systematic evaluations of finite sample results with demonstrations of the bands' performance are rarely available in the literature. In the present article we provide such results in Cox models and we investigate various new resampling bands that exhibit a reliable performance even for small sample sizes.

Our main achievements are the introduction of resampling strategies that jointly mimic the unknown distribution of baseline and parameter estimators in Cox

models and corresponding multistate versions. We provide theoretical justification for the resampling approaches based on martingale arguments, thereby allowing for different mechanisms of incomplete observations and simplifying many arguments.

There exist many resampling approaches for the Cox model with independent right-censoring. One approach is based on the martingale representations of the Breslow estimator for the cumulative hazard and the parameter estimator, where all unobservable martingale residuals $dM_i(u)$ are replaced with reweighted counting processes $G_i dN_i(u)$ (e.g., Lin et al., 1994). Another possibilty that works more generally, for example for rate estimation, is to replace the martingale instead with $G_i d\widehat{M}_i(u)$ (e.g., Spiekerman & Lin, 1998). This exploits the model's semiparametric structure more actively.

We will also use a more natural approach, which starts one step earlier by rewriting the score equations for the baseline function and the Euclidean parameter: we apply the multiplier techniques from above to martingale representations of the score equations. This leads to new equations, which are solved by quantities depending on the multipliers. Hence, paralleling the same steps as for the original estimators, their resampling counterparts are obtained in an elementary way. Throughout the article, we follow the flexible approach of Beyersmann et al. (2013) and allow for general multiplier distributions.

For ease of presentation, we exemplify the new methodology mainly for the Cox model in a survival setting but we also discuss extensions to more general multistate and other regression models. We are able to base our theoretical derivations for the resampling approaches on martingale arguments, which are novel in the semiparametric context. Thus, intricate derivations for verifying conditional tightness, that is, tightness of the resampling process conditioned on the observed data, are no longer required and are handled conveniently by the martingale central limit theorems. The martingale arguments for the multiplier bootstrap are motivated by those for resampling nonparametric Aalen-Johansen estimators (Bluhmki et al., 2018). In particular, we prove similar martingale properties of the resampling counterparts as for the original estimators. Moreover, mirroring the martingale structure in the bootstrap world allows for a simple interpretation and easy incorporation of missingness mechanisms, for example, independent right-censoring or left-truncation. To conclude, our findings allow for a wide range of applications some of which will be discussed in more detail in future papers.

The paper is organized as follows: Section 2 sketches how estimators in Cox models are found and it lists the technical conditions required for all of their large sample properties. Section 3 contains a derivation of all considered multiplier bootstrap procedures and also theoretical statements about their validity. In addition, we discuss several important extensions to more general multistate or other regression models. In Section 4, we compare the resampling procedures in an extensive simulation study. Section 5 has a brief demonstration of the methodology in a survival setting, where interest is on constructing confidence bands for the survival function for patients with acute myocardial infarction, and another demonstration of confidence bands for cumulative incidence functions for patients after a bone marrow transplantation. Finally, we discuss the findings in Section 6.

The online supporting information contains all proofs whose novelty lies in the considerable simplification of the technical arguments by transferring martingale methods to the resampling procedures in the present semiparametric framework. It also contains a second small simulation study, algorithmic descriptions of the multiplier bootstraps, and links to GitHub pages with example R code for user-friendly applicability of our procedures.

2 | JOINT LARGE SAMPLE PROPERTIES IN THE COX MODEL

We consider multiplicative *Cox* (or *proportional hazards*) models given by the intensity

$$\lambda_i(t, \boldsymbol{\beta}_0) = Y_i(t)\lambda_0(t)\exp\{\boldsymbol{X}'_i(t)\boldsymbol{\beta}_0\}. \qquad (1)$$

of the counting process $N_i(t)$ of subject $i = 1, ..., n$, possibly subject to independent left-truncation and/or right-censoring, given its possibly time-dependent p-dimensional vector of predictable covariates $\boldsymbol{X}_i(t) = \{X_{i1}(t), ..., X_{ip}(t)\}'$. That is, $E\{dN_i(t)|\boldsymbol{X}_i(t)\} = \lambda_i(t, \boldsymbol{\beta}_0)dt$. Here, $N_i(t) \in \{0, 1\}$ counts the occurence of a particular event for subject i not later than treatment time t, $Y_i(t)$ is the corresponding at-risk indicator at that time, λ_0 is the baseline hazard function, and $\boldsymbol{\beta}_0 = (\beta_1, ..., \beta_p)' \in \mathbb{R}^p$ is an unknown regression parameter (Cox, 1972). Let $\tau > 0$ be a terminal evaluation time on the treatment time-scale. Throughout we assume all $\boldsymbol{X}_i(t)$ to be contained in a bounded set and that the cumulative baseline hazard $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ is finite at $t = \tau$. Standard arguments yield the *Breslow estimator* $\widehat{\Lambda}_0$ of Λ_0 and the maximum partial likelihood parameter estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$. To illustrate this, we simplify derivations of Scheike and Zhang (2002) for Cox-Aalen models to the Cox model Equation (1): the score equation for Λ_0 is

$$\sum_{i=1}^{n} [dN_i(t) - Y_i(t)\exp\{X'_i(t)\boldsymbol{\beta}\}d\Lambda_0(t)] = 0, \quad (2)$$

and solved by $\widehat{\Lambda}_0(t, \boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^t J(u)S_0^{-1}(u, \boldsymbol{\beta})dN_i(u)$. We will make use of the definitions

$$\boldsymbol{S}_k(t, \boldsymbol{\beta}) = \sum_{i=1}^{n} X_i^{\otimes k}(t)Y_i(t)\exp\{X'_i(t)\boldsymbol{\beta}\}, \quad k = 0, 1, 2,$$

where $\boldsymbol{y}^{\otimes 2} = \boldsymbol{y}\boldsymbol{y}' \in \mathbb{R}^{p \times p}$, $\boldsymbol{y}^{\otimes 1} = \boldsymbol{y} \in \mathbb{R}^p$ and $\boldsymbol{y}^{\otimes 0} = 1 \in \mathbb{R}$ for any vector $\boldsymbol{y} \in \mathbb{R}^p$, and $J(u)$ is the indicator that any individual is under risk shortly before u. For simplicity, the notion of $J(u)$ is usually suppressed. Replacing $\widehat{\Lambda}_0(t, \boldsymbol{\beta})$ for Λ_0 in the score equation for $\boldsymbol{\beta}$, that is,

$$\sum_{i=1}^{n} \int_0^\tau X_i(u)Y_i(u)[dN_i(u) - \exp\{X'_i(u)\boldsymbol{\beta}\}d\Lambda_0(u)]$$
$$= 0, \quad (3)$$

and defining $\boldsymbol{E}(t, \boldsymbol{\beta}) = \boldsymbol{S}_1(t, \boldsymbol{\beta})S_0^{-1}(t, \boldsymbol{\beta})$, we obtain a solvable score equation for $\boldsymbol{\beta}$:

$$\boldsymbol{U}_\tau(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^\tau \{X_i(u) - \boldsymbol{E}(u, \boldsymbol{\beta})\}dN_i(u) = 0.$$

Through its solution $\widehat{\boldsymbol{\beta}}$ we obtain the *Breslow estimator* $\widehat{\Lambda}_0(t, \widehat{\boldsymbol{\beta}})$ of $\Lambda_0(t)$. The central limit theorem for $\widehat{\boldsymbol{\beta}}$ stated below requires the negative Jacobian of $\boldsymbol{U}_\tau$ with respect to $\boldsymbol{\beta}$, that is, $\boldsymbol{I}_\tau(\boldsymbol{\beta}) = -D\boldsymbol{U}_\tau(\boldsymbol{\beta}) = \int_0^\tau \boldsymbol{V}(u, \boldsymbol{\beta})dN(u)$, where $\boldsymbol{V}(t, \boldsymbol{\beta}) = \boldsymbol{S}_2(t, \boldsymbol{\beta})S_0^{-1}(t, \boldsymbol{\beta}) - \boldsymbol{E}^{\otimes 2}(t, \boldsymbol{\beta})$ and $N = \sum_{i=1}^{n} N_i$. Due to the assumed boundedness of the covariates, it follows from Theorems VII.2.2 (p. 498) and VII.2.3 (p. 503f.) in Andersen et al. (1993) that $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and $\sqrt{n}\{\widehat{\Lambda}_0(\cdot, \widehat{\boldsymbol{\beta}}) - \Lambda_0(\cdot)\}$ are asymptotically Gaussian provided the following regularity conditions hold; see Condition VII.2.1 in Andersen et al. (1993, p. 497). Let $\xrightarrow{\text{p}}$ denote convergence in probability as $n \to \infty$ and ∇ the gradient operator with respect to $\boldsymbol{\beta}$.

Condition 1. *There exists a neighborhood $\mathscr{B}$ of $\boldsymbol{\beta}_0$ and functions $s_0: [0, \tau] \times \mathscr{B} \to \mathbb{R}$, $s_1: [0, \tau] \times \mathscr{B} \to \mathbb{R}^p$, and such that for each $k = 0, 1, 2$:*

(a) $\sup_{\boldsymbol{\beta} \in \mathscr{B}, t \in [0,\tau]} \|n^{-1}\boldsymbol{S}_k(t, \boldsymbol{\beta}) - \boldsymbol{s}_k(t, \boldsymbol{\beta})\| \xrightarrow{\text{p}} 0$, *where* $\|\cdot\|$ *denotes the Euclidean norm;*

(b) $\boldsymbol{s}_k$ *is a continuous function of $\boldsymbol{\beta} \in \mathscr{B}$ uniformly in $t \in [0, \tau]$ and bounded on $[0, \tau] \times \mathscr{B}$;*

(c) $s_0(\cdot, \boldsymbol{\beta}_0)$ *is bounded away from zero on $[0, \tau]$;*

(d) $\boldsymbol{s}_2(t, \boldsymbol{\beta}) = D\boldsymbol{s}_1(t, \boldsymbol{\beta})$ *and* $\boldsymbol{s}_1(t, \boldsymbol{\beta}) = \{\nabla s_0(t, \boldsymbol{\beta})\}'$ *for $\boldsymbol{\beta} \in \mathscr{B}$, $t \in [0, \tau]$;*

(e) $\boldsymbol{Sigma}_\tau = \int_0^\tau \boldsymbol{v}(t, \boldsymbol{\beta}_0)s_0(t, \boldsymbol{\beta}_0)d\Lambda_0(t)$ *is positive definite, where* $\boldsymbol{v}(t, \boldsymbol{\beta}) = \boldsymbol{s}_2(t, \boldsymbol{\beta})s_0^{-1}(t, \boldsymbol{\beta}) - \{\boldsymbol{s}_1(t, \boldsymbol{\beta})s_0^{-1}(t, \boldsymbol{\beta})\}^{\otimes 2}$.

Throughout, we assume Condition 1 to be satisfied. Note that (a) and (b) immediately imply

$$\sup_{t \in [0,\tau]} \|\boldsymbol{S}_k(t, \tilde{\boldsymbol{\beta}}) - \boldsymbol{s}_k(t, \boldsymbol{\beta}_0)\| \xrightarrow{\text{p}} 0 \quad (4)$$

for each $k = 0, 1, 2$ as long as $\tilde{\boldsymbol{\beta}} \xrightarrow{\text{p}} \boldsymbol{\beta}_0$. The proofs of the aforementioned theorems in Andersen et al. (1993) make use of the following asymptotic representations of the estimators:

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \left\{\frac{1}{n}\boldsymbol{I}_\tau(\boldsymbol{\beta}_0)\right\}^{-1} \frac{1}{\sqrt{n}}\boldsymbol{U}_\tau(\boldsymbol{\beta}_0) + o_p(1) \quad (5)$$

$$\sqrt{n}\{\widehat{\Lambda}_0(\cdot, \widehat{\boldsymbol{\beta}}) - \Lambda_0(\cdot)\}$$
$$= -\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \int_0^\cdot \boldsymbol{e}(u, \boldsymbol{\beta}_0)d\Lambda_0(u)$$
$$+ \sqrt{n} \int_0^\cdot \frac{dM(u)}{S_0(u, \boldsymbol{\beta}_0)} + o_p(1). \quad (6)$$

These representations motivate the first bootstrap approach in Section 3. Here, $\boldsymbol{e}$ is the probability limit of $\boldsymbol{E}$, and $M(t) = \sum_{i=1}^{n} M_i(t) = \sum_{i=1}^{n} \{N_i(t) - \int_0^t \lambda_i(u, \boldsymbol{\beta}_0)du\}$ is a square-integrable martingale in $t \in [0, \tau]$; cf. Andersen et al. (1993, Section VII.2.2; p. 496ff) .

3 | MULTIPLIER RESAMPLING APPROACHES AND MAIN THEOREMS

While inference about $\boldsymbol{\beta}_0$ can be based on the asymptotic normality of its estimator (e.g., Martinussen & Scheike, 2006; p. 184ff.), the complicated limit process of the normalized Breslow estimator hinders straightforward time-simultaneous inference about Λ_0 or functionals thereof, for example, the survival function. As a remedy we will investigate two resampling strategies, both of which make use of independent i.i.d. zero mean and unit variance random variables $G_1, ..., G_n$, which are henceforth called *multipliers*. Since $\widehat{\Lambda}_0$ and $\widehat{\boldsymbol{\beta}}$ are dependent, we have to ensure that the bootstrap procedures respect their dependence structure.

3.1 | The 'Classical' multiplier bootstrap

The *first method* is inspired by the resampling procedures of Lin et al. (1994) and Spiekerman and Lin (1998). They use standard normal multipliers, which is motivated from the asymptotic martingale representation of the normalized

estimators. The idea is to replace the martingale increments $dM_i(t) = dN_i(t) - Y_i(t)\exp\{X'_i(t)\beta_0\}d\Lambda_0(t)$ with $G_i dN_i(t)$ (Lin et al., 1994) or with $G_i \widehat{M}_i(t) = G_i[dN_i(u) - Y_i(u)\exp\{X'_i(u)\widehat{\beta}\}d\widehat{\Lambda}_0(u, \widehat{\beta})]$, using estimated martingale increments (Spiekerman & Lin, 1998). We extend these approaches in several ways: our approaches are not restricted to normal multipliers G_i; cf. Beyersmann et al. (2013) where centered Poisson multipliers exhibited a better finite sample behavior in a nonparametric setting. Also, to be in line with the usual guidelines for bootstrap-based inference (Hall & Wilson, 1991), we provide and analyze suitable bootstrap versions of the covariance estimators. Finally, in Section 3.4 an extension to more general models is outlined.

We exemplify the general resampling idea for $G_i dN_i(u)$ by introducing multiplier bootstrap versions of the score equation defining vector U_τ and its negative Jacobian I_τ.

$$U_\tau^*(\widehat{\beta}) = \sum_{i=1}^n G_i \int_0^\tau \{X_i(t) - E(t, \widehat{\beta})\}dN_i(t) \quad \text{and} \quad (7)$$

$$\frac{1}{n}I_\tau^*(\widehat{\beta}) = \frac{1}{n}\sum_{i=1}^n G_i^2 \int_0^\tau \{X_i(t) - E(t, \widehat{\beta})\}\{X_i(t) - E(t, \widehat{\beta})\}'dN_i(t). \quad (8)$$

Here and below, all objects with asterisks refer to bootstrap quantities. Following the asymptotic representation Equations (5) and (6), we find their following multiplier bootstrap counterparts:

$$\sqrt{n}(\widehat{\beta}^* - \widehat{\beta}) = \left\{\frac{1}{n}I_\tau^*(\widehat{\beta})\right\}^{-1}\frac{1}{\sqrt{n}}U_\tau^*(\widehat{\beta}), \quad (9)$$

$$\sqrt{n}\{\widehat{\Lambda}_0^*(\cdot, \widehat{\beta}^*) - \widehat{\Lambda}_0(\cdot, \widehat{\beta})\}$$
$$= -\sqrt{n}(\widehat{\beta}^* - \widehat{\beta})' \int_0^\cdot E(u, \widehat{\beta})\widehat{\Lambda}_0(du, \widehat{\beta})$$
$$+ \sqrt{n}\sum_{i=1}^n G_i \int_0^\cdot \frac{dN_i(u)}{S_0(u, \widehat{\beta})}. \quad (10)$$

Note that Equations (9) and (10) are essentially definitions of $\widehat{\beta}^*$ and $\widehat{\Lambda}_0^*(\cdot, \widehat{\beta}^*)$, respectively, and that the $o_p(1)$ terms in Equations (5) and (6) have been dropped. Alternatively to this multiplier bootstrap choice, the Spiekerman and Lin (1998)-type martingale increment estimates $G_i d\widehat{M}_i(u)$ may replace $G_i dN_i(u)$ in Equations (7) and (10). A bootstrap-type covariance estimate similar to Equation (8) has been suggested by Dobler and Pauly (2014) in a nonparametric competing risks context. Here, it is additionally motivated from martingale arguments: defining I_t^* and U_t^* as in Equations (7)

and (8) but t replacing τ, we found that $\{I_t^*(\widehat{\beta})\}_{t\in[0,\tau]}$ is the optional variation process of the square-integrable martingale $\{n^{-1/2}U_t^*(\widehat{\beta})\}_{t\in[0,\tau]}$; see the supporting information for details.

The next subsection offers an alternative resampling strategy, which does not ignore the $o_p(1)$ terms in Equations (5) and (6). Instead, they are naturally incorporated into the resampling step.

3.2 | Bootstrapping the score equations

The idea of the *second* approach is to replace martingale representations of *score equations* with their multiplier counterparts. To this end, we parallel the classic approach of jointly solving two score equations and first expand the score equation in Equation (2) to $\sum_{i=1}^n dM_i(t) + \sum_{i=1}^n Y_i(t)[\exp\{X'_i(t)\beta_0\} - \exp\{X'_i(t)\beta\}]d\Lambda_0(t) = 0$. A multiplier bootstrap counterpart thereof is now given by replacing $dM_i(t)$ with $G_i dN_i(t)$, β_0 with $\widehat{\beta}$, and $\Lambda_0(t)$ with $\widehat{\Lambda}_0(t, \widehat{\beta})$:

$$\sum_{i=1}^n G_i dN_i(t) + \sum_{i=1}^n Y_i(t)[\exp\{X'_i(t)\widehat{\beta}\} - \exp\{X'_i(t)\beta\}]d\widehat{\Lambda}_0(t, \widehat{\beta}) \quad (11)$$
$$= \sum_{i=1}^n (G_i + 1)dN_i(t) - S_0(t, \beta)d\widehat{\Lambda}_0(t, \widehat{\beta}) \overset{!}{=} 0.$$

Keeping β fixed, its "solution" for $\widehat{\Lambda}_0(t, \widehat{\beta})$ is

$$\widehat{\Lambda}_0^*(t, \beta) = \sum_{i=1}^n (G_i + 1) \int_0^t \frac{dN_i(t)}{S_0(t, \beta)}. \quad (12)$$

Next, consider a martingale representation of the score Equation (3) for β:

$$\sum_{i=1}^n \int_0^\tau X_i(t)Y_i(t)dM_i(t) + \sum_{i=1}^n \int_0^\tau X_i(t)Y_i(t)[\exp\{X'_i(t)\beta_0\} - \exp\{X'_i(t)\beta\}]d\Lambda_0(t) = 0.$$

This motivates a multiplier resampling version given by

$$\sum_{i=1}^n G_i \int_0^\tau X_i(t)Y_i(t)dN_i(t) + \sum_{i=1}^n \int_0^\tau X_i(t)Y_i(t)[\exp\{X'_i(t)\widehat{\beta}\} - \exp\{X'_i(t)\beta\}]d\widehat{\Lambda}_0(t, \widehat{\beta})$$
$$= \sum_{i=1}^n \int_0^\tau \{X_i(t)Y_i(t)G_i + E(t, \widehat{\beta})\}dN_i(t)$$
$$- \int_0^\tau S_1(t, \beta)d\widehat{\Lambda}_0(t, \widehat{\beta}) \overset{!}{=} 0.$$

Inserting $\widehat{\Lambda}_0^*(t, \beta)$ for $\widehat{\Lambda}_0(t, \widehat{\beta})$ eventually yields the final *multiplier bootstrap score equation*

$$U_\tau^*(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\tau \{X_i(t)Y_i(t)G_i + E(t, \widehat{\boldsymbol{\beta}})\} dN_i(t)$$

$$- \int_0^\tau E(t, \boldsymbol{\beta}) \sum_{i=1}^n (G_i + 1) dN_i(t)$$

$$= \sum_{i=1}^n (G_i + 1) \int_0^\tau \{X_i(t) - E(t, \boldsymbol{\beta})\} dN_i(t) \overset{!}{=} 0.$$

$$(13)$$

The last equality is due to $\sum_{i=1}^n \int_0^\tau \{X_i(t) - E(t, \widehat{\boldsymbol{\beta}})\} dN_i(t) = U_\tau(\widehat{\boldsymbol{\beta}}) = 0$. Hence, $U_\tau^*(\widehat{\boldsymbol{\beta}})$ coincides with formula Equation (7). Define $\widehat{\boldsymbol{\beta}}^*$ as the solution to Equation (13). Similarly as in the proof of Theorem VII.2.1 in Andersen et al. (1993, p. 497f.), it can be shown that the probability of the existence of tends to one and that $\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}$ converges to zero in probability as $n \to \infty$; see also the proof of Theorem 1 below for similar arguments.

Finally, we combine Equation (12) with $\widehat{\boldsymbol{\beta}}^*$ to find a resampling version of the Breslow estimator, denoted $\widehat{\Lambda}_0^*(\cdot, \widehat{\boldsymbol{\beta}}^*)$. After centering it at $\widehat{\Lambda}_0(t, \widehat{\boldsymbol{\beta}})$ and multiplying it with $\sqrt{n}$, it equals

$$\sqrt{n} \sum_{i=1}^n (G_i + 1) \int_0^t \{S_0^{-1}(u, \widehat{\boldsymbol{\beta}}^*) - S_0^{-1}(u, \widehat{\boldsymbol{\beta}})\} dN_i(u)$$

$$+ \sqrt{n} \sum_{i=1}^n G_i \int_0^t \frac{dN_i(u)}{S_0(u, \widehat{\boldsymbol{\beta}})}$$

$$= \sqrt{n} \sum_{i=1}^n \int_0^t \{S_0^{-1}(u, \widehat{\boldsymbol{\beta}}^*) - S_0^{-1}(u, \widehat{\boldsymbol{\beta}})\} dN_i(u)$$

$$+ \sqrt{n} \sum_{i=1}^n G_i \int_0^t \frac{dN_i(u)}{S_0(u, \widehat{\boldsymbol{\beta}}^*)}.$$

$$(14)$$

A Taylor expansion around $\widehat{\boldsymbol{\beta}}$ of the first term on the right-hand side reveals the striking similarity to decomposition Equation (10). However, the current multiplier bootstrap approach does not ignore the $o_p(1)$ term resulting from the Taylor expansion. Another nice property of this "estimating equation" approach is the similar treatment for bootstrap and original estimator, which is in line with general recommendations for constructing resampling algorithms (Efron & Tibshirani, 1994). We are going to use the mean value theorem (cf. Feng et al., 2013) to analyze the asymptotic behavior of $\widehat{\boldsymbol{\beta}}^*$. To this end, we introduce a modified Jacobian operator. For a continuously differentiable function $\boldsymbol{H} = (H_1, ..., H_p)' : \mathbb{R}^p \to \mathbb{R}^p$ and a matrix $\boldsymbol{B} = (\boldsymbol{\beta}^{(1)}|...|\boldsymbol{\beta}^{(p)}) \in \mathbb{R}^{p \times p}$ we define $\bar{D}\boldsymbol{H}(\boldsymbol{B}) = (\nabla H_1(\boldsymbol{\beta}^{(1)})', ..., \nabla H_p(\boldsymbol{\beta}^{(p)})')'$. The mean value theorem yields that $-U_\tau^*(\widehat{\boldsymbol{\beta}}) = U_\tau^*(\widehat{\boldsymbol{\beta}}^*) - U_\tau^*(\widehat{\boldsymbol{\beta}}) = \bar{D}U_\tau^*(\widetilde{\boldsymbol{B}})(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})$, where every column of $\widetilde{\boldsymbol{B}}$ is on the line segment between $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\boldsymbol{\beta}}$.

3.3 | Consistency and confidence bands for the cumulative hazard

The validity proof of both resampling strategies from Sections 3.1 and 3.2 is based on

Lemma 1. *Under Conditions 1(a)-(e) it holds that, as $n \to \infty$*

(I) $\|\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}\| \overset{p}{\to} 0$,

(II) $-n^{-1}\bar{D}U_\tau^*(\widetilde{\boldsymbol{B}}) \overset{p}{\to} \Sigma_\tau$ *if each column of $\widetilde{\boldsymbol{B}}$ converges in probability to $\boldsymbol{\beta}_0$,*

(III) *given all observations, $n^{-1/2}U_\tau^*(\widehat{\boldsymbol{\beta}}) \overset{d}{\to} \mathcal{N}(\boldsymbol{0}, \Sigma_\tau)$ in probability, i.e. it is asymptotically multivariately normally distributed,*

if the resampling is done via method Equation (9) or (13).

The next theorem shows that both bootstraps based on the $G_i dN_i$ approach (Lin et al., 1994) are asymptotically valid. Therein, d denotes a distance that metrizes weak convergence on $\mathbb{R}^p \times D[0, \tau]$, e.g. the Prohorov distance (Dudley, 2002, p. 309ff.), and $\mathcal{L}(T)$ and $\mathcal{L}(T|\text{data})$ are, respectively, the unconditional and conditional law of a random variable T.

Theorem 1. *Under Condition 1 it holds for both resampling strategies Equation (9) or (13) that the asymptotic distributions of $\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}, \widehat{\Lambda}_0^* - \widehat{\Lambda}_0)$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \widehat{\Lambda}_0 - \Lambda_0)$ coincide, that is,*

$$d[\mathcal{L}\{\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}, \widehat{\Lambda}_0^* - \widehat{\Lambda}_0)|\text{data}\}, \mathcal{L}\{\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \widehat{\Lambda}_0 - \Lambda_0)\}] \overset{p}{\to} 0 \quad (15)$$

as $n \to \infty$, where $\widehat{\Lambda}_0(t) = \widehat{\Lambda}_0(t, \widehat{\boldsymbol{\beta}})$ and $\widehat{\Lambda}_0^*(t) = \widehat{\Lambda}_0^*(t, \widehat{\boldsymbol{\beta}}^*)$.

The asymptotic variance function $t \mapsto \sigma^2(t)$ of $\sqrt{n}(\widehat{\Lambda}_0 - \Lambda_0)$ (and thus also of $\sqrt{n}(\widehat{\Lambda}_0^* - \widehat{\Lambda}_0)$) can be found in Andersen et al. (1993), Corollary VII.2.4, p. 505, where a consistent estimator $\widehat{\sigma}^2(t)$ is given. In our simulation study in Section 4, we chose the multiplier bootstrap counterpart of $\widehat{\sigma}^2(t)$ to be the empirical variance of the bootstrap realizations of $\sqrt{n}\{\widehat{\Lambda}_0^*(t) - \widehat{\Lambda}_0(t)\}$.

We use the theorem to construct time-simultaneous confidence bands for Λ_0 on fixed intervals $I = [t_1, t_2] \subset [0, \tau]$. In particular, we obtain results similar to those of Lin et al. (1994): denoting by ϕ a continuously differentiable and strictly increasing or

decreasing function, asymptotic level $1 - \alpha$ confidence bands for Λ_0 on I are given by

$$\phi^{-1}[\phi\{\widehat{\Lambda}_0(t, \widehat{\boldsymbol{\beta}})\} \mp c_\phi^*(\alpha)/g_n(t)],$$

where $g_n: I \to (0, \infty)$ is a possibly random weight function. Typical choices are

$$g_n^{(1)}(t) = \sqrt{n}/\widehat{\sigma}(t) \quad \text{and} \quad g_n^{(2)}(t) = \sqrt{n}/\{1 + \widehat{\sigma}^2(t)\}$$

in case of the transformation $\phi_1(x) = x$, and

$$\widetilde{g}_n^{(1)}(t) = \sqrt{n}\widehat{\Lambda}_0(t, \widehat{\boldsymbol{\beta}})/\widehat{\sigma}(t) \quad \text{and} \quad \widetilde{g}_n^{(2)}(t)$$
$$= \sqrt{n}\widehat{\Lambda}_0(t, \widehat{\boldsymbol{\beta}})/\{1 + \widehat{\sigma}^2(t)\}$$

for the transformation $\phi_2(x) = \log(x)$. The resulting confidence bands correspond to the so-called *equal precision* (for $g_n^{(1)}$ or $\widetilde{g}_n^{(1)}$) and *Hall-Wellner bands* (for $g_n^{(2)}$ or $\widetilde{g}_n^{(2)}$), respectively. Let $g^*(t)$ and $\widetilde{g}^*(t)$ be the multiplier bootstrap analogs of $g_n^{(j)}(t)$ and $\widetilde{g}_n^{(j)}(t)$, respectively, $j \in \{1, 2\}$, that is, we utilized $\widehat{\Lambda}_0^*$ instead of $\widehat{\Lambda}_0$ and the corresponding empirical variances instead of $\widehat{\sigma}$. We chose $c_\phi^*(\alpha) = c_{\phi_1}^*(\alpha)$ to be the $(1 - \alpha)$ quantile of the conditional distribution of $\sup_{t \in I} g^*(t) |\widehat{\Lambda}_0^*(t, \widehat{\boldsymbol{\beta}}) - \widehat{\Lambda}_0(t, \widehat{\boldsymbol{\beta}})|$, and the naïve choice for $c_{\phi_2}^*(\alpha)$ is the corresponding quantile of $\sup_{t \in I} \widetilde{g}^*(t) |\log\{\widehat{\Lambda}_0^*(t, \widehat{\boldsymbol{\beta}})\} - \log\{\widehat{\Lambda}_0(t, \widehat{\boldsymbol{\beta}})\}|$. However, the latter resulted in numerical instabilities, and we preferred the asymptotically equivalent choice $c_{\phi_2}^*(\alpha) = c_{\phi_1}^*(\alpha)$.

It follows from Theorem 1 that all confidence bands are valid for large sample sizes. To additionally assess their small sample properties, we compare them in Monte-Carlo simulations in Section 4. There, we also analyze the analogous behavior of the resampling approaches based on $d\widehat{M}_i$. It should be noted that the confidence bands obtained from the 'classical' multiplier bootstrap from Section 3.1 with standard normal multipliers are not equal to those proposed in Lin et al. (1994). The key difference is that we additionally use a resampling version of the weight functions (particularly of the variance estimator) to be in line with the usual guidelines for bootstrap-based inference (Hall & Wilson, 1991). We additionally refer to Allignol et al. (2018), where a similar recommendation was made.

3.4 | Extensions to more general models and more on inference

A related multiplier bootstrap approach carries over to more general models in multistate set-ups. In particular,

as long as the counting process martingale can be mimicked with the help of bootstrap multipliers, the asymptotics of the resampled estimators result in almost the same way as for the original estimators. Thus, the above methodology can straightforwardly be extended to multistate models with K states and multiplicative intensity processes

$$\lambda_{ih}(t, \boldsymbol{\theta}) = Y_{ih}(t)\lambda_{h0}(t, \gamma)\exp\{\boldsymbol{X}'_i(t)\boldsymbol{\beta}_0\} \quad (16)$$

for each transition $h = 1, ..., K(K - 1)$, where $\boldsymbol{\theta} = (\gamma, \boldsymbol{\beta}'_0)'$. Different to above, this model allows for an arbitrary number of transitions between different states. Following Bluhmki et al. (2018), the above multiplier bootstrap approach remains applicable: instead of the previously used multipliers G_i, more general white noise processes $\{G_{ih}(u)\}_u$ with zero mean and unit variance are required. Randomly weighting the counting process increments leads to $G_{ih}(u)dN_{ih}(u)$. The evaluation occurs at a finite number of times, so no existence problems arise for the white noise processes, and the implementation is straightforward.

The martingale arguments still hold, so that the multiplier bootstrap mimics the joint stochastic behavior of the parameter and multivariate hazard function estimators. Indeed, according to Bluhmki et al. (2018), independent white noise processes $\{G_{ih}(u)\}_u$ and $\{G_{i\tilde{h}}(u)\}_u$ for transitions $h \neq \tilde{h}$ give rise to orthogonal square-integrable martingales in t, $\int_0^t f_{ih}(u)G_{ih}(u)dN_{ih}(u)$ and $\int_0^t f_{i\tilde{h}}(u)G_{i\tilde{h}}(u)dN_{i\tilde{h}}(u)$, with respect to the filtration

$$\mathscr{F} = [\mathscr{F}_t := \sigma\{Y_{ih}(u), N_{ih}(u), Y_{i\tilde{h}}(u), N_{i\tilde{h}}(u)$$
$$: 0 \leq u \leq \tau; \ G_{ih}(v), G_{i\tilde{h}}(v): 0 \leq v \leq t\}]_{t \in [0,\tau]}$$

Here, f_{ih} and $f_{i\tilde{h}}$ are $\mathscr{F}$-predictable random functions. The predictable variation process of the first martingale is $\int_0^t f_{ih}^2(u)dN_{ih}(u)$, which nicely reflects the properties of the original estimators: the corresponding orthogonal counting process martingales $\int_0^t f_{ih}(u)dM_{ih}(u)$ and $\int_0^t f_{i\tilde{h}}(u)dM_{i\tilde{h}}(u)$ have predictable variation $\int_0^t f_{ih}^2(u)Y_{ih}(u)\lambda_{h0}(u, \gamma)\exp\{\boldsymbol{X}'_i(u)\boldsymbol{\beta}_0\}du$, and similarly for $\tilde{h}$. In this sense, the bootstrap-based predictable variation processes estimate the original predictable variation processes. Extending the proofs in the supporting information based on these findings, large-sample properties of estimators in multistate problems transfer to their multiplier bootstrap versions in an obvious way, as long as the original estimators admit martingale representations. Similar arguments extend to Cox-Aalen multiplicative-additive intensities (Scheike & Zhang, 2002) or other models with martingale structures.

The incorporation of certain filtered (e.g., right-censored) observations is again allowed and important inferential applications follow: confidence bands for cumulative transition hazards or incidence functions, and tests for null hypotheses formulated in terms of the parameters can be constructed. Here, bootstrap-based versions of score or Wald-type test statistics (Martinussen & Scheike, 2006, p. 185f.) might ensure a proper finite sample behavior. However, detailed evaluations of these applications require additional extensive simulations and further elaborations in future research. For simplicity, we continue with our focus on the Cox model Equation (1) and assess the impact of the proposed methods in simulations.

4 | SIMULATION STUDY

To compare the performances of the various resampling approaches described in Section 3.3, we conducted a simulation study in which we covered situations of small to large sample sizes: $n = 100, 200, 400$. The generated data follow the Cox survival model with baseline hazard rate $\lambda_0 \equiv 1$, normally distributed covariates $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 16)$ with standard deviation $\sigma = 4$, and regression parameter $\beta = 0.3$. Censoring times are standard exponentially distributed, truncated at $\tau = 3$. The time interval, along which 95% confidence bands for the cumulative baseline hazard function Λ_0 shall be constructed, was $[t_1, t_2] = [0.5, 3]$. Here we chose the start time of $t_1 = 0.5$ because "the approximations tend to be poor for t close to 0" (Lin et al., 1994, p. 77). Another simulation study that evaluates the multiplier bootstrap for confidence bands for baseline cumulative incidence functions is given in Suporting Information.

As bootstrap multipliers $G_1, ..., G_n$, we considered the common choice $G_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, centered unit Poisson variables $G_i \overset{\text{i.i.d.}}{\sim} Poi(1) - 1$ with unit skewness, and centered unit exponential variables $G_i \overset{\text{i.i.d.}}{\sim} Exp(1) - 1$ with a skewness of 2. We simulated all confidence bands for Λ_0 that were introduced in Section 3.3, i.e. log- and nontransformed Hall-Wellner and equal precision bands. We considered replacing the martingale increments dM_i with $G_i dN_i$ or $G_i d\widehat{M}_i$, denoted in Table 1 as "dN" and "dM", respectively, and also both kinds of resampling algorithms, the direct resampling method of Section 3.1 and the method of Section 3.2 in which the estimating equations were bootstrapped. We note that, apart from the specific variance estimation scheme, the bands resulting from the cox.aalen function in the R package *timereg* are essentially the same as the multiplier bootstrap from Section 3.1 with no transformation and

standard normal multipliers and the results for these are therefore not shown. For each set-up and type of band, we constructed 10,000 confidence bands, where each was based on 1,000 resampling iterations. The obtained empirical coverage probabilities given in Table 1 are the observed frequencies of constructed bands containing Λ_0 completely.

When the sample size is 400, all methods give a reasonable performance. For smaller samples there are notable differences, and it seems that the log-transform improves the bands' performances considerably. Whether the bootstrap is based on $G_i dN_i$ or $G_i d\widehat{M}_i$ does not seem important, and in terms of computations it is considerably easier and faster to use the multipliers based on $G_i dN_i$. Despite theoretical and practical advantages of Poisson variables over standard normal multipliers in nonparametric competing risks models (Dobler et al., 2017), this choice of bootstrap multipliers does not seem highly important here. However, centered exponential multipliers tend to perform the best. Also, the particular resampling method, be it the direct approach of Section 3.1 or the estimating equation approach of Section 3.2, does not seem to have a clearly positive or negative impact on the outcomes.

Apart from the coverage probabilities, we also computed the bands' median widths for $t \in [0.5, 3]$. Because of the marginal differences among multiple resampling approaches, Figure 1 only shows the median widths for the $G_i dN_i$ method with standard normal multipliers, based on resampling the estimating equation. Again, we let $n = 100, 200, 400$. In all panels the trend is the same: at early time points the equal precision bands are slightly wider than the Hall-Wellner bands. But for later time points it is the opposite; and the widths of the Hall-Wellner bands grow rapidly due to quickly increasing variances. Also, the log-transformed bands are slightly wider than their untransformed counterparts. All in all, we recommend to use log-transformed equal precision bands for Λ_0 because of their good coverage probabilities (log-transformation) and their reasonable slimness.

Finally, we remark that our simulation results are only partially comparable to those of Lin et al. (1994): we construct confidence bands for Λ_0, that is, for an individual with $X_i = 0$; their bands are for survival curves for multiple covariates using also different transformantions.

Overall, however, the bands here and in Lin et al. (1994) seem to work well.

5 | DATA EXAMPLES

In this section, we illustrate how the resampling approach can be used in two practical settings, based on data that are available in the *timereg* R-package. The

TABLE 1 Simulated coverage probabilities (in %, rounded) of various 95% confidence bands for the baseline cumulative hazard function and sample sizes $n = 100, 200, 400$

| | | | Hall-Wellner | | | | Equal precision | | | |
| | | | Estimating equation | | Direct resampling | | Estimating equation | | Direct resampling | |
Distribution of G_i	n	Resampling approach	id	log	id	log	id	log	id	log
$\mathcal{N}(0, 1)$	100	dN	86.8	93.4	87.2	93.4	85.6	92.7	85.5	92.5
		dM	86.9	93.2	87.2	93.6	85.7	92.6	85.2	92.4
	200	dN	90.8	94.2	90.9	94.2	89.8	93.9	89.8	93.8
		dM	90.6	94.0	90.7	94.2	89.4	93.3	89.3	93.2
	400	dN	93.2	95.1	93.3	95.0	92.4	94.1	92.4	94.1
		dM	93.2	94.9	93.2	95.1	92.3	94.1	92.4	94.0
$Exp(1) - 1$	100	dN	88.4	94.4	88.5	94.5	88.8	95.3	89.0	95.3
		dM	88.1	94.3	88.4	94.5	88.7	95.4	87.9	94.6
	200	dN	91.3	94.6	91.3	94.7	91.3	95.4	91.4	95.3
		dM	91.2	94.5	91.2	94.6	91.0	94.9	90.8	94.6
	400	dN	93.6	95.3	93.5	95.2	93.1	94.9	93.1	95.0
		dM	93.9	95.4	94.1	95.5	93.5	95.4	93.4	95.2
$Poi(1) - 1$	100	dN	87.0	93.3	87.1	93.6	86.1	93.3	86.2	93.0
		dM	87.0	93.6	87.4	93.7	86.0	93.2	86.0	92.8
	200	dN	90.6	94.1	90.7	94.2	89.6	93.6	89.7	93.7
		dM	91.0	94.2	91.2	94.3	90.0	94.1	90.0	94.0
	400	dN	93.3	95.0	93.3	95.1	92.4	94.2	92.6	94.2
		dM	93.3	95.1	93.1	95.1	92.4	94.3	92.3	94.2

Distributions of the multipliers G_i : standard normal (top panel), centered unit exponential (middle panel), centered unit Poisson (bottom panel)

R-code is given in the supporting information. First, we show how confidence bands can be used and constructed in a standard survival setting. The key point is that a simultaneous confidence band is the type of interval estimate most often of interest unless focus is on a particular survival probability at a specific time such as, for example, 5-year survival. In addition, we find confidence intervals for derived measures, in particular, for the restricted mean life. The second example illustrates in a competing risks setting how simple it is to combine the multiplier bootstrap based on Cox models for cause-specific hazards to construct confidence bands for the cumulative incidence.

5.1 | Survival estimation

We consider the TRACE study (Jensen et al., 1997), where interest is on survival after acute myocardial infarction for 1878 consecutive patients included in the study. The data-set is available in the *timereg* R-package. Here for sake of illustration we focus interest on the covariates diabetes (1/0), sex and age. We depict the

survival predictions with confidence bands for a male with average age (66.9 years) and with or without diabetes, as well as the standard 95% point-wise confidence intervals for comparison.

We note that the hazard ratio related to diabetes is 1.82 with 95% confidence interval (1.50, 2.18). Thus reflecting that diabetes is a factor that leads to increased mortality. More interestingly, in regard to absolute mortality rates, the higher risk of diabetics is reflected in our estimated survival curves for males with average age and with diabetes (lower set of curves, with confidence bands (region) and point-wise intervals (dotted lines) or without diabetes (upper set of curves). The regions are standard *timereg*-bands based on direct resampling and via $G_i d\widehat{M}_i$, and for comparison we also show the resampling-based equal precision bands based on a log-transform (broken lines). We note a slight difference between the different confidence bands. In addition we note that the bands are quite a bit wider than the point-wise intervals. As the latter do not provide simultaneous coverage, the bands should be used to provide uncertainty about the entire survival curve as shown in Figure 2.
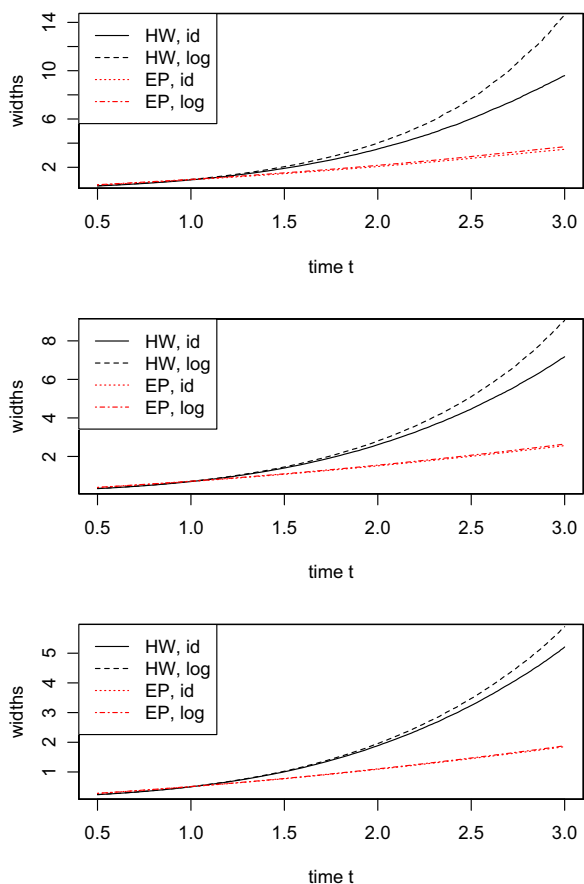
FIGURE 1 Median widths of untransformed and log-transformed confidence bands for sample sizes $n = 100$ (top), $n = 200$ (middle), $n = 400$ (bottom panel). Solid and dashed: Hall-Wellner, pointed and point-dashed: equal precision. This figure appears in color in the electronic version of this article

We finally illustrate how the joint asymptotic distribution of the baseline and parameter estimators can be used with other functionals. To this end, consider the restricted mean life of an individual with lifetime T and covariate vector X, that is,

$$\Psi(\Lambda_0, \beta_0, X) = \int_0^\tau \exp\{-\Lambda_0(s)\exp(X'\beta_0)\}ds$$
$$= E\{\min(T, \tau)|X\}$$

with estimator $\Psi(\widehat{\Lambda}_0, \widehat{\beta}, X)$. We refer to Karrison (1987) for the concept of adjusting the restricted mean life to covariates. To get a description of the uncertainty of $\Psi(\widehat{\Lambda}_0, \widehat{\beta}, X)$ based on multiplier bootstrap constructions we can simply apply the functional to the obtained bootstrap samples. It follows that $\sqrt{n}\{\Psi(\widehat{\Lambda}_0, \widehat{\beta}, X) - \Psi(\Lambda_0, \beta_0, X)\}$ has the same asymptotic distribution as $\sqrt{n}\{\Psi(\widehat{\Lambda}_0^*, \widehat{\beta}^*, X) - \Psi(\widehat{\Lambda}_0, \widehat{\beta}, X)\}$ due to Hadamard differentiability. Thus, we can easily construct symmetric 95% confidence intervals for the restricted mean and their differences based on the boot-

strap. The key point being that these are very easy to get at when the bootstrap estimates are at hand. In addition, still due to Hadamard differentiability, differences in restricted mean life $\Psi(\widehat{\Lambda}_0, \widehat{\beta}, X_1) - \Psi(\widehat{\Lambda}_0, \widehat{\beta}, X_2)$ and uncertainties thereof are also estimable using the obtained bootstrap estimates.

For example, the direct multiplier bootstrap approach based on $G_i dN_i$ and standard normal multipliers yields that males with diabetes have a restricted mean life within the first 5 years at 3.87 (3.74, 4.00) for males without diabetes and 3.15 (2.91, 3.41) with diabetes. Males with diabetes thus lose 0.71 (0.49, 0.93) years within the first 5 years. Confidence intervals for other functionals are similarly found by means of the functional Δ method.

5.2 | Cumulative incidence for competing risks

We here consider data on bone marrow transplantation from patients that are treated for myelodysplasia (Sierra et al., 2002). This data is also available in the the *timereg* R-package. Of 408 patients 161 died from treatment-related causes, and 87 had relapse before dying, and the remaining were censored. We consider these as two competing risks of death.

We wish to estimate the risk F_1 of dying from treatment-related causes by a certain point in time, and do this by modeling the two cause-specific hazards using Cox
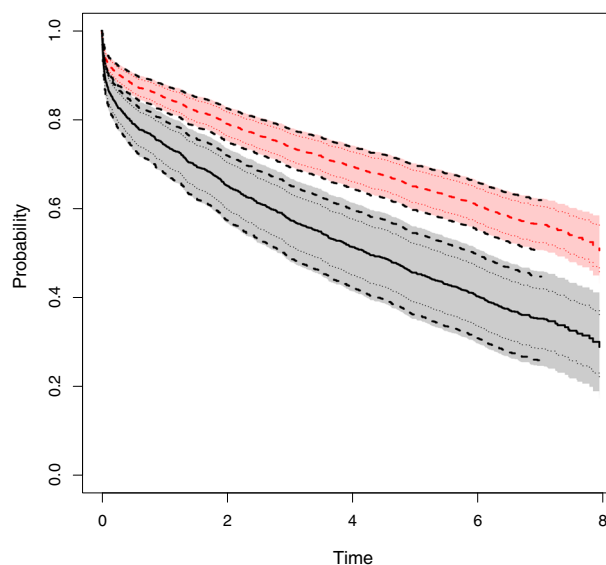

FIGURE 2 Survival function estimates for males with average age, with (fat broken line) or without diabetes (fat solid line), with 95 % confidence intervals (dotted lines), multiplier bootstrapped log-transform 95 % confidence bands (broken lines) and 95 % confidence bands of *timereg* (regions). This figure appears in color in the electronic version of this article

regression models based on the covariates platelet count (High 1/Low 0), tcell-depletion (yes 1/no 0), and standardized age ([age-35]/15). We then combine the estimates for a particular covariate combination (Low platelets, no tcell-depletion, and average age (35 years) to estimate the cumulative incidence $F_1(t|X) = \int_0^t S(u|X)\lambda_1(u|X)dt$, where $S(u|X)$ is the probability of surviving both causes, and $\lambda_1(u|X)$ is the intensity for a treatment-related death.

We denote the baseline cumulative hazards and the parameters of both Cox models by $\theta = (\Lambda_1, \beta_1, \Lambda_2, \beta_2)$. Note that the Cox models for each cause-specific hazard are resampled independently. Their combination yields inference on the cumulative incidence functional

$$F_1(\theta)(t|X) = \int_0^t \exp\{-\Lambda_1(t)\exp(X'\beta_1)$$
$$- \Lambda_2(t)\exp(X'\beta_2)\}\exp(X'\beta_1)d\Lambda_1(dt).$$

Again, $\sqrt{n}\{F_1(\widehat{\theta}) - F_1(\theta)\}$ has the same asymptotic distribution as $\sqrt{n}\{F_1(\widehat{\theta}^*) - F_1(\widehat{\theta})\}$ due to Hadamard differentiability of the functional. We thus simply need to apply the multiplier bootstrap to the two Cox models and then combine estimates. Subsequently, we can construct a confidence band for $F_1(\theta)$ based on the realizations at hand. Figure 3 shows the cumulative incidence curve for a subject with covariate vector $X = 0$ with 95% equal precision confidence bands with (dotted lines) and without log-transform (broken lines).
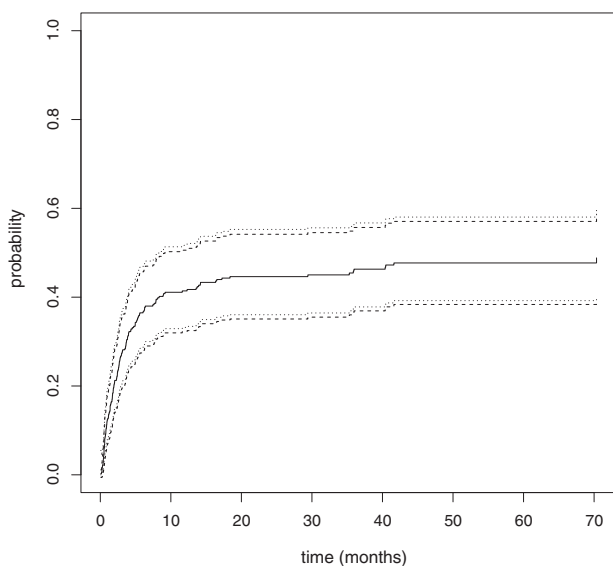

FIGURE 3 Cumulative incidence curve for a subject with covariate $X = 0$ with 95% equal precision confidence bands with (dotted lines) and without log-transform (broken lines)

6 | DISCUSSION AND FURTHER RESEARCH

Despite their importance and considerable interest in making survival predictions based on semiparametric regression models such as the Cox model, confidence bands are not used much in practice. This is probably because the most commonly-used Cox model routines do not implement simultaneous confidence bands. The aim of this work is to investigate some natural and simple multiplier bootstrap approaches for filling this gap. In particular, we show in the supporting information that the proposed bootstrap solutions do asymptotically have the desired properties. A key point in our proofs is that major characteristics of the procedure can be deduced from martingale arguments. This enormously facilitates the transfer of classical proofs for the estimators to their multiplier bootstrap counterparts. The approach evidently generalizes to more complex models as long as martingale structures for the relevant counting processes are preserved, that is, Cox models in multistate models or Fine-Gray regression models for subdistribution functions (Fine & Gray, 1999) with administrative censoring. Future work will focus on the adaptation to more complex designs.

In addition, we consider the finite sample performance of various confidence bands. It turns out that the choice of transformation and weight function considerably influence widths and coverage of the bands. Among the bootstrap weights, centered exponential multipliers exhibit the best coverage probability. In particular, their combination with the log-transformation leads to very accurate results, even for a small sample size ($n = 100$). Furthermore, the equal precision bands exhibit much smaller widths in comparison to the Hall-Wellner bands.

Another nice feature of our approach is its simple extensibility to find confidence intervals for functionals. We illustrated this with the restricted mean life based on a Cox model.

ORCID

Dennis Dobler http://orcid.org/0000-0002-9040-0854

Markus Pauly ⓘ http://orcid.org/0000-0002-0976-7190
Thomas H. Scheike ⓘ http://orcid.org/0000-0002-2148-4740

REFERENCES

Allignol, A., Königsbauer, F. and Pauly, M. (2018). The Effect of Left-Truncation on the Coverage of Confidence Intervals for Pregnancy Outcome Probabilities. Ulm, Germany: Ulm University.

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

Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993). Statistical Models Based on Counting Processes. New York: Springer.

Beyersmann, J., DiTermini, S. and Pauly, M. (2013) Weak convergence of the wild bootstrap for the Aalen-Johansen estimator of the cumulative incidence function of a competing risk. Scandinavian Journal of Statistics, 78, 387–402.

Bluhmki, T., Dobler, D., Beyersmann, J. and Pauly, M. (2019) The wild bootstrap for multivariate Nelson-Aalen estimators. Lifetime Data Analysis, 25, 97–127.

Bluhmki, T., Schmoor, C., Dobler, D., Pauly, M., Finke, J., Schumacher, M. and Beyersmann, J. (2018) A wild bootstrap approach for the Aalen-Johansen estimator. Biometrics, 74, 977–985.

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

Dobler, D. and Pauly, M. (2014) Bootstrapping Aalen-Johansen processes for competing risks: Handicaps, solutions, and limitations. Electronic Journal of Statistics, 8, 2779–2803.

Dobler, D., Beyersmann, J. and Pauly, M. (2017) Non-strange weird resampling for complex survival data. Biometrika, 104, 699–711.

Dudley, R.M. (2002). Real Analysis and Probability. Cambridge, UK: Cambridge Studies in Advanced Mathematics.

Efron, B. and Tibshirani, R.J. (1994). An Introduction to the Bootstrap. Boca Raton, FL: CRC Press.

Feng, C., Wang, H., Han, Y., Xia, Y. and Tu, X.M. (2013) The mean value theorem and Taylor's expansion in statistics. The American Statistician, 67, 245–248.

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, 496–509.

Hall, P. and Wilson, S.R. (1991) Two guidelines for bootstrap hypothesis testing. Biometrics, 47, 757–762.

Jensen, G.V., Torp-Pedersen, C., Hildebrandt, P., Kober, L., Nielsen, F.E., Melchior, T., ..., Andersen, P.K. (1997) Does in-hospital ventricular fibrillation affect prognosis after myocardial infarction? European Heart Journal, 18, 919–924.

Karrison, T. (1987) Restricted mean life with adjustment for covariates. Journal of the American Statistical Association, 82, 1169–1176.

Lin, D.Y. (1997) Non-parametric inference for cumulative incidence functions in competing risks studies. Statistics in Medicine, 16, 901–910.

Lin, D.Y., Wei, L.Y. and Ying, Z. (1993) Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika, 80, 557–572.

Lin, D.Y., Fleming, T.R. and Wei, L.J. (1994) Confidence bands for survival curves under the proportional hazards model. Biometrika, 81, 73–81.

Martinussen, T. and Scheike, T.H. (2006). Dynamic Regression Models for Survival Data, Statistics for Biology and Health. New York: Springer.

Scheike, T.H. and Zhang, M.J. (2002) An additive-multiplicative Cox-Aalen regression model. Scandinavian Journal of Statistics, 29, 75–88.

Sierra, J., Pérez, W.S., Rozman, C., Carreras, E., Klein, J.P., Rizzo, J.D., ..., Horowitz, M.M. (2002) Bone marrow transplantation from HLA-identical siblings as treatment fro myelodysplasia. Blood, 100, 1997–2004.

Spiekerman, C.F. and Lin, D.Y. (1998) Marginal regression models for multivariate failure time data. Journal of the American Statistical Association, 93, 1164–1175.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.