

Combined multiple testing of multivariate survival times by censored empirical likelihood

Judith H. Parkinson 

Department of Mathematics, University of Salzburg

Correspondence

Judith Parkinson, Department of Mathematics, University of Salzburg, Hellbrunner Str. 34, 5020 Salzburg, Austria.
Email: judith.parkinson@sbg.ac.at

Funding information

Austrian Science Fund, FWF I 2697

Abstract

In each study testing the survival experience of one or more populations, one must not only choose an appropriate class of tests, but further an appropriate weight function. As the optimal choice depends on the true shape of the hazard ratio, one is often not capable of getting the best results with respect to a specific dataset. For the univariate case several methods were proposed to conquer this problem. However, most of the interesting datasets contain multivariate observations nowadays. In this work we propose a multivariate version of a method based on multiple constrained censored empirical likelihood where the constraints are formulated as linear functionals of the cumulative hazard functions. By considering the conditional hazards, we take the correlation between the components into account with the goal of obtaining a test that exhibits a high power irrespective of the shape of the hazard ratio under the alternative hypothesis.

KEYWORDS

conditional hazards, multiple constraints, multivariate survival analysis, weighted log-rank test

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Scandinavian Journal of Statistics* published by John Wiley & Sons Ltd on behalf of The Board of the Foundation of the Scandinavian Journal of Statistics.

1 | INTRODUCTION

In each study testing the survival experience of one or more populations, one must choose an appropriate test. Many of those tests rely on the choice of a weight function. One can select an optimal weight function if the true shape of the hazard ratio under the alternative hypothesis is known, which is not the case in most applications.

As a result one is not capable of getting the best results with regard to a specific dataset. The choice of the weight function influences the error rates. In certain situations a wrong choice of the weight function may lead to a considerable power reduction as pointed out by Kosorok and Lin (1999) and Klein and Moeschberger (1997). In the univariate case different methods were proposed to solve the problem, for example, by Brendel, Janssen, Mayer, and Pauly (2014).

Another problem arises from the complexity of multivariate survival times. Even though in current times of massive data ascertainment, data typically contains multivariate observations, one might lack an appropriate method to model the true structure of the survival times.

For bivariate samples, many methods to analyze survival times have been proposed. The methods by Lin and Ying (1993) and Wang and Wells (1997) deal with nonparametric estimators of the bivariate survival function under univariate censoring. An earlier approach by Dabrowska (1988) can even deal with general bivariate censoring. Theoretical and computational performance comparisons of different estimators were conducted by Gill, van der Laan, and Wellner (1995) and Wang and Zafra (2009), among others. A recent paper by Huang and Zhao (2018) also deals with the bivariate survival function and uses empirical likelihood to obtain asymptotic confidence intervals.

While we mainly focus on the bivariate time-to-event times, our method can easily be extended to a higher dimensional case. For multivariate data with more than two response variables, methods have been proposed as well. For the multivariate case there exist less methods than in the bivariate case due to the higher complexity and technical challenges. The most common model for multivariate time-to-event data is the Cox model. Further, the accelerated failure time model is also often applied. A comparison of the Cox model and the accelerated failure time model, among others, was performed, for example, by Clark, Bradburn, Love, and Altman (2003b). But most multivariate methods have limitations. For example many models make parametric assumptions which may not be met, as does the Cox model which assumes proportional hazards. Other methods might lack the ability to deal correctly with dependent or correlated samples to their full extent, as their design might not pick up the correlation structure. The assumption of independence between the components is sometimes implicitly assumed, even though it is likely to be violated when dealing with different times-to-event of a single subject. Tests assuming independence will fail, for example, when the marginal distributions are similar for two samples, but the correlation differs. Most methods, which are primary, based on univariate test statistics can solve this problem by integrating the covariance matrices into the multivariate test statistics, as performed by Wei and Lachin (1984). An overview of multivariate survival analysis and what to consider when choosing a method for a specific dataset was given, for instance, in Clark, Bradburn, Love, and Altman (2003a, 2003c, 2003d) and Clark et al. (2003b).

The focus in this paper is motivated by the approach proposed by Bathke, Kim, and Zhou (2009). Their method is based on censored, multiply-constrained empirical likelihood, where the constraints are formulated as linear functionals of the cumulative hazard functions. Their simulations suggested that the test performed well along with an easy computation compared to, for example, the function-indexed weighted log-rank tests which use extensive Monte-Carlo simulations. The proposed method unfortunately could only deal with univariate observations.

In this paper, following the approach of Bathke et al. (2009), we propose a multivariate test statistic. In case of independent components, the approach introduced here reduces to that by Bathke et al. (2009). Test procedures designed in an univariate manner are naturally not capable of picking up correlations or dependencies. By considering the conditional hazards, we take the correlation between the components into account with the goal of obtaining a test with a high power, irrespective of the shape of the hazard ratio under the alternative hypothesis. When considering real time-to-event data it is unlikely that the observations within one subject are uncorrelated. It is critical to have test methods that are capable of detecting these data structures. This is why an extension of the method proposed by Bathke et al. (2009) to the multivariate case is important. Combining the conditional hazards and the constraints offers a new and well-functioning test procedure that could be the foundation for further tests on multivariate survival times.

The paper is structured as follows. Section 2 deals with the theoretical aspects of the proposed method for the one-sample case. Section 3 describes the results for the two-sample case. In Section 4 we give examples on how to apply the proposed method. The simulations and an application to real data can be found in Section 5. The regularity conditions and the proofs of the main theorems are given in the Appendix.

2 | ONE-SAMPLE CENSORED EMPIRICAL LOG-LIKELIHOOD WITH MULTIPLE CONSTRAINTS

We start by giving the results for the one-sample case. We propose two methods: one assumes independence between the components, while the other one takes possible dependencies into account.

Let X_1, \dots, X_n be independent, identically distributed d -dimensional observations with $X_i := (X_{1i}, \dots, X_{di})^T$, distribution function $F_{X_0}(t)$, marginal distribution functions $F_{X_{01}}(t), \dots, F_{X_{0d}}(t)$, cumulative hazard function $\Lambda_{X_0}(t)$, and marginal cumulative hazard functions $\Lambda_{X_{01}}(t), \dots, \Lambda_{X_{0d}}(t)$. Due to right censoring we only observe

$$T_{ji} = \min\{X_{ji}, C_{ji}^X\} \text{ and } \delta_{ji} = \mathbb{1}\{X_{ji} \leq C_{ji}^X\}, \quad i = 1, \dots, n, j = 1, \dots, d, \tag{1}$$

with $C_i^X = (C_{1i}^X, \dots, C_{di}^X)^T$ the censoring times, which are assumed to be independent, identically distributed, and independent of the X_i 's.

2.1 | One-sample EL with independent components

We assume in this section that the components of one observation vector are independent from each other. Based on the censored observations, the empirical log-likelihood (EL) for component j pertaining to its distribution F_{X_j} is given by

$$\log \text{EL}(\hat{F}_{X_j}) = \sum_i [\delta_{ji} \log(\Delta \hat{F}_{X_j}(T_{ji})) + (1 - \delta_{ji}) \log(1 - \hat{F}_{X_j}(T_{ji}))], \tag{2}$$

with \hat{F}_{X_j} the empirical marginal distribution function of component $j = 1, \dots, d$. The analogue of Equation (2) based on the cumulative hazard function is given by

$$\log \text{EL}(\hat{\Lambda}_{X_j}) = \sum_i [d_{ji} \log v_{ji} + (R_{ji} - d_{ji}) \log(1 - v_{ji})], \tag{3}$$

where

$$d_{ji} = \sum_{l=1}^n \mathbb{1}\{T_{jl} = t_{ji}\} \delta_{jl} \text{ and } R_{ji} = \sum_{l=1}^n \mathbb{1}\{T_{jl} \geq t_{ji}\},$$

with t_{ji} denoting the ordered, distinct values of T_{ji} . Here, $0 < v_{ji} \leq 1$ are the discrete hazards of component j at time t_{ji} . It is well known that the maximum of Equation (3) with respect to v_{ji} is attained at the jumps of the Nelson–Aalen estimator for each component, that is, $v_{ji} = d_{ji}/R_{ji}$.

Let $\theta := (\theta_{11}, \dots, \theta_{1k}, \dots, \theta_{d1}, \dots, \theta_{dk})^T$ be a $(k \cdot d)$ -dimensional parameter defined via the components' cumulative hazard functions $\Lambda_{X_{01}}, \dots, \Lambda_{X_{0d}}$,

$$\theta_{jr} = \int g_{jr}(t) \log(1 - d\Lambda_{X_{0j}}(t)), \quad r = 1, \dots, k, \quad j = 1, \dots, d,$$

and a hypothesis testing problem

$$H_0 : \theta_{jr} = \mu_{jr} \quad \forall j = 1, \dots, d, r = 1, \dots, k \quad \text{versus} \quad H_A : \theta_{jr} \neq \mu_{jr} \text{ for some } j \text{ and } r, \quad (4)$$

where g_{jr} , $j = 1, 2, r = 1, \dots, k$ are nonnegative functions, and $\theta_j = (\theta_{j1}, \dots, \theta_{jk})^T$ and $\mu_j = (\mu_{j1}, \dots, \mu_{jk})^T$ are two vectors of constants.

The proposed constraints on the hazard v_{ji} for given functions g_{j1}, \dots, g_{jk} and constants $\mu_{j1}, \dots, \mu_{jk}$ are given by

$$\sum_i g_1(t_{ji}) \log(1 - v_{ji}) = \mu_{j1}, \dots, \sum_i g_k(t_{ji}) \log(1 - v_{ji}) = \mu_{jk}, \quad j = 1, 2, \quad (5)$$

where the sum is taken over all i for which $v_{ji} \neq 1$.

Denote the maximum empirical likelihood estimators of $\Delta \hat{\Lambda}_{X_j}(t_{ji})$ under constraints as v_{ji}^* , where $\hat{\Lambda}_{X_j}$ is the empirical cumulative hazard function. Application of the Lagrange multiplier method leads to

$$v_{ji}^* = v_{ji}(\lambda_j) = \frac{d_{ji}}{R_{ji} + n\lambda_j^T \mathbf{G}_j(t_{ji})}, \quad j = 1, \dots, d,$$

where $\mathbf{G}_j(t_{ji}) = \{g_{j1}(t_{ji}), \dots, g_{jk}(t_{ji})\}^T$, and $\lambda_j \in \mathbb{R}^k$ is a vector of Lagrange multipliers obtained by solving the constraints for component j . Then the component-wise test statistic in terms of hazards is given by

$$W_j = -2 \left[\max_{v_{ji}} \text{EL}(\Lambda_{X_j}) \text{ (when (5) holds)} - \max_{v_{ji}} \text{EL}(\Lambda_{X_j}) \right]. \quad (6)$$

Theorem 1. Define a test statistic for the complete data set by $W := \sum_{j=1}^d W_j$ with W_j defined as in Equation (6). Then, under regularity conditions specified in the Appendix and under H_0 as given in Equation (4), W is asymptotically chi-squared distributed with $k \cdot d$ degrees of freedom for $n \rightarrow \infty$.

Proof. Bathke et al. (2009) showed that the component-wise test statistics W_j are chi-squared distributed with k degrees of freedom. As we assumed independence between the components, and thus the test statistics, we obtain the desired result. ■

Remark 1. Although the notation suggests that the number of constraints per component is equal for all components, they may differ for the individual components, leading to different degrees of freedom. More precisely let k_j be the number of constraints for component j and define

$K = k_1 + \dots + k_d$. Then, the test statistic W is asymptotically chi-squared distributed with K degrees of freedom under the null hypothesis.

Even though this yields a rather easy and intuitively interpretable test statistic, we encounter serious limitations. We assumed independence between the components, which may be questionable for many data sets. We tackle this problem using a more general approach in the next section.

2.2 | General bivariate one-sample EL

Following the approach of Dabrowska (1988) and Vonta, Nikulin, Limnios, and Huber-Carol (2008) for bivariate survival models using conditional hazards, we derive the asymptotic distributions of the nonparametric maximum-likelihood estimator (NPMLE) for the hazards of the bivariate survival function without constraints for the two-dimensional case and the asymptotic distribution of the actual test statistics based on the empirical log-likelihood function. We point out that the following sections are based on the assumption that the true underlying cumulative hazard function is continuous, even if we do not explicitly state so. Let

$$\mathcal{L} = \{ \Lambda : \mathbb{R}^+ \rightarrow \mathbb{R}^+, \text{continuous, nondecreasing, } \Lambda(0) = 0, \Lambda(t) \rightarrow \infty \text{ for } t \rightarrow \infty \}$$

be the class of continuous univariate cumulative hazard functions on \mathbb{R}^+ . We define a joint bivariate survival function S on $\mathbb{R}^+ \times \mathbb{R}^+$ as proposed in Dabrowska (1988) by

$$\begin{aligned} dS(x, y) &= \exp \left\{ -\Lambda_{11}^{10}(x) - \Lambda_{11}^{01}(x) - \Lambda_{11}^{11}(x) \right\} d\Lambda_{11}^{10}(x) \\ &\quad \exp \left\{ -(\Lambda_{01}^{01}(y) - \Lambda_{01}^{01}(x)) \right\} d\Lambda_{01}^{01}(y) \text{ for } x < y, \\ dS(x, y) &= \exp \left\{ -\Lambda_{11}^{10}(y) - \Lambda_{11}^{01}(y) - \Lambda_{11}^{11}(y) \right\} d\Lambda_{11}^{01}(y) \\ &\quad \exp \left\{ -(\Lambda_{10}^{10}(x) - \Lambda_{10}^{10}(y)) \right\} d\Lambda_{10}^{10}(x) \text{ for } x > y, \\ dS(x, y) &= \exp \left\{ -\Lambda_{11}^{10}(x) - \Lambda_{11}^{01}(x) - \Lambda_{11}^{11}(x) \right\} d\Lambda_{11}^{11}(x) \text{ for } x = y, \end{aligned}$$

where $\Lambda_{ab}^{cd} \in \mathcal{L}$ with a and b indicating whether component 1 or 2, respectively, are event free up to x and c and d indicating whether an event occurred in components 1 and 2 at time point x . The notation is analogue to the one of the hazard rates, which are given below. This survival function $S(x, y) = P(X \geq x, Y \geq y)$ is indeed a bivariate survival function. A necessary condition for the proposed method is that $\Lambda_{11}^{11} \equiv 0$ holds true, that is, the survival times of the components are not identical

The observed bivariate time is given by $T = (T_{1i}, T_{2i})$ and the censoring indicator is given by $\delta_i = (\delta_{1i}, \delta_{2i})$. Denote the hazards by

$$\lambda_{11}^{10}(t)dt = P(t \leq X_1 \leq t + dt | X_1 \geq t, X_2 > t),$$

$$\lambda_{11}^{01}(t)dt = P(t \leq X_2 \leq t + dt | X_1 > t, X_2 \geq t),$$

$$\lambda_{10}^{10}(t)dt = P(t \leq X_1 \leq t + dt | X_1 \geq t, X_2 < t),$$

$$\lambda_{01}^{01}(t)dt = P(t \leq X_2 \leq t + dt | X_1 < t, X_2 \geq t).$$

The empirical likelihood for n observations is the product $V = \prod_{i=1}^n V_i$ with

$$V_i = \prod_t \left[(1 - \lambda_{11}^{10}(t)dt)^{\mathbb{1}\{T_{1i} > t\} \mathbb{1}\{T_{2i} > t\}} \cdot (\lambda_{11}^{10}(t)dt)^{\mathbb{1}\{T_{1i} = t\} \mathbb{1}\{T_{2i} > t\} \delta_{1i}} \right. \\ \cdot (1 - \lambda_{11}^{01}(t)dt)^{\mathbb{1}\{T_{1i} > t\} \mathbb{1}\{T_{2i} > t\}} \cdot (\lambda_{11}^{01}(t)dt)^{\mathbb{1}\{T_{1i} > t\} \mathbb{1}\{T_{2i} = t\} \delta_{2i}} \\ \cdot (1 - \lambda_{10}^{10}(t)dt)^{\mathbb{1}\{T_{1i} > t\} \mathbb{1}\{T_{2i} < t\} \delta_{2i}} \cdot (\lambda_{10}^{10}(t)dt)^{\mathbb{1}\{T_{1i} = t\} \mathbb{1}\{T_{2i} < t\} \delta_{2i} \delta_{1i}} \\ \left. \cdot (1 - \lambda_{01}^{01}(t)dt)^{\mathbb{1}\{T_{1i} < t\} \mathbb{1}\{T_{2i} > t\} \delta_{1i}} \cdot (\lambda_{01}^{01}(t)dt)^{\mathbb{1}\{T_{1i} < t\} \mathbb{1}\{T_{2i} = t\} \delta_{2i} \delta_{1i}} \right].$$

For simplicity, denote the hazards with $a_l := \lambda_{11}^{10}(t_{1l})dt$, $b_l := \lambda_{11}^{01}(t_{2l})dt$, $c_l := \lambda_{10}^{10}(t_{1l})dt$, and $d_l := \lambda_{01}^{01}(t_{2l})dt$, where t_{jl} are the ordered distinct time points of jumps in component j . It follows that the empirical log-likelihood EL is given by

$$\sum_{t_{1l}} \left[\log(1 - a_l) \sum_i (\mathbb{1}\{T_{1i} > t_{1l}\} \mathbb{1}\{T_{2i} > t_{1l}\}) + \log(a_l) \sum_i (\mathbb{1}\{T_{1i} = t_{1l}\} \mathbb{1}\{T_{2i} > t_{1l}\} \delta_{1i}) \right] \\ + \sum_{t_{2l}} \left[\log(1 - b_l) \sum_i (\mathbb{1}\{T_{1i} > t_{2l}\} \mathbb{1}\{T_{2i} > t_{2l}\}) + \log(b_l) \sum_i (\mathbb{1}\{T_{1i} > t_{2l}\} \mathbb{1}\{T_{2i} = t_{2l}\} \delta_{2i}) \right] \quad (7) \\ + \sum_{t_{1l}} \left[\log(1 - c_l) \sum_i (\mathbb{1}\{T_{1i} > t_{1l}\} \mathbb{1}\{T_{2i} < t_{1l}\} \delta_{2i}) + \log(c_l) \sum_i (\mathbb{1}\{T_{1i} = t_{1l}\} \mathbb{1}\{T_{2i} < t_{1l}\} \delta_{1i} \delta_{2i}) \right] \\ + \sum_{t_{2l}} \left[\log(1 - d_l) \sum_i (\mathbb{1}\{T_{1i} < t_{2l}\} \mathbb{1}\{T_{2i} > t_{2l}\} \delta_{1i}) + \log(d_l) \sum_i (\mathbb{1}\{T_{1i} < t_{2l}\} \mathbb{1}\{T_{2i} = t_{2l}\} \delta_{1i} \delta_{2i}) \right],$$

where the sums over t_{jl} are taken over those time points where a_l , respectively b_l , c_l , and d_l are strictly greater than zero, and excluding the last time point, as the hazards would be 1. Denote the sums over "i" by u_{1l}, \dots, u_{8l} in the order they are given above, where l indicates that the sum was taken at the l th time point of the first component, respectively, the second component.

Lemma 1. *Maximizing the empirical log-likelihood function with respect to a_l , b_l , c_l , and d_l , one obtains the NPMLE*

$$\hat{a}_l = \frac{u_{2l}}{u_{1l} + u_{2l}}, \quad \hat{b}_l = \frac{u_{4l}}{u_{3l} + u_{4l}}, \quad \hat{c}_l = \frac{u_{6l}}{u_{5l} + u_{6l}}, \quad \hat{d}_l = \frac{u_{8l}}{u_{7l} + u_{8l}}.$$

Proof. Taking the derivatives of EL with respect to a_l , we obtain

$$-(1 - a_l)^{-1} u_{1l} + a_l^{-1} u_{2l}.$$

The derivative is zero for $\hat{a}_l = u_{2l}/(u_{1l} + u_{2l})$. For the other hazards similar calculations will lead to the given NPMLE. ■

When fixing the constraints we do not consider the four hazards individually, but combine them such that only one hazard per component remains for the constraints. This will lead to a reduction of the degrees of freedom, as for each constraint one degree of freedom is added to the asymptotic sampling distribution of the test statistic. Define

$$v_{1l} = P(T_2 > t_{1l} | T_1 \geq t_{1l}) \cdot a_l + P(T_2 < t_{1l} | T_1 \geq t_{1l}) \cdot c_l.$$

We estimate v_{1l} by $z_{1l}a_l + z_{3l}c_l$, where

$$z_{1l} = (u_{1l} + u_{2l}) / (u_{1l} + u_{2l} + u_{5l} + u_{6l}) = 1 - z_{3l}.$$

The term v_{2l} is defined similarly, namely,

$$v_{2l} = P(T_1 > t_{2l} | T_2 \geq t_{2l}) \cdot b_l + P(T_1 < t_{2l} | T_2 \geq t_{2l}) \cdot d_l$$

and estimated by $z_{2l} \cdot b_l + z_{4l} \cdot d_l$ with

$$z_{2l} = (u_{3l} + u_{4l}) / (u_{3l} + u_{4l} + u_{7l} + u_{8l}) = 1 - z_{4l}.$$

The constraints are then imposed directly on v_{1l} and v_{2l} and are given in the form of

$$\sum_i g_{jr}(t_{ji}) \log(1 - v_{ji}) = \mu_{jr}, \quad j = 1, 2, \quad r = 1, \dots, k.$$

The function to maximize is given by

$$G_{\text{constraint}} = \text{EL}(\lambda) + \sum_{j=1}^2 \sum_{r=1}^k n \lambda_{jr} \left[\mu_{jr} - \sum_i g_{jr}(t_{ji}) \log(1 - v_{ji}) \right],$$

where $\text{EL}(\lambda)$ denotes the log-likelihood function, as given in Equation (7), without the constraints but with the modified hazards $a_l(\lambda_1), \dots, d_l(\lambda_2)$ instead of the Nelson–Aalen estimators plugged in and $\lambda = (\lambda_1, \lambda_2)^T$, $\lambda_j = (\lambda_{j1}, \dots, \lambda_{jk})^T$, $j = 1, 2$. Again, the last sum ranges over the distinct time points excluding the last.

Remark 2. For survival times with absolutely continuous distribution, there is one event per time point such that either a_l or c_l will be zero. This also implies that $v_{1l} \leq a_l$ if c_l is zero or $v_{1l} \leq c_l$ if a_l is zero. Taking additionally into account that we exclude the last time points such that $u_{2l} / (u_{1l} + u_{2l}) \neq 1$ and $u_{6l} / (u_{5l} + u_{6l}) \neq 1$, we can even go as far as stating $a_l, c_l \leq 0.5$ and thus $v_{1l} \leq 0.5$.

Remark 3. As the true NPMLE of $G_{\text{constraint}}$ is rather complex to calculate, we will approximate it. Using Remark 2, we approximate $\log(1 - v_{1l})$ by $z_{1l} \log(1 - a_l) + z_{3l} \log(1 - c_l)$ with an error of order v_{1l}^2 , where one of the two terms is always equal to zero. In total, we approximate $G_{\text{constraint}}$ by

$$G_{\text{constraint}}^* = \text{EL}(\lambda) + \sum_{r=1}^k n \lambda_{1r} \left[\mu_{1r} - \sum_i g_{1r}(t_{1i}) (z_{1i} \log(1 - a_i(\lambda_1)) + z_{3i} \log(1 - c_i(\lambda_1))) \right] + \sum_{r=1}^k n \lambda_{2r} \left[\mu_{2r} - \sum_i g_{2r}(t_{2i}) (z_{2i} \log(1 - b_i(\lambda_2)) + z_{4i} \log(1 - d_i(\lambda_2))) \right], \quad (8)$$

where the sum is taken over those time points t_{1l} for which $a_i(\lambda_1) \neq 1$ and $c_i(\lambda_1) \neq 1$ and over t_{2i} for which $b_i(\lambda_1) \neq 1$ and $d_i(\lambda_1) \neq 1$.

Remark 4. For noncontinuous distributions v_{jl} can be close to zero just like in the continuous case. It may, however, occur in such cases that we obtain high hazard rates at some of the jump points. Thus the approximation as described in Remark 3 would not hold in general. Therefore, we will exclude noncontinuous distributions in the following.

Remark 5. The approximation is a true equation if the components are independent, that is, the conditional hazards can be replaced by the regular hazards for the individual components. In this case, the conditional hazard model simplifies to the sum of the univariate models and thus we obtain the same results as in Section 2.1.

Remark 6. Instead of approximating $G_{\text{constraint}}$ one can also simply define the parameter θ under the null hypothesis to be the weighted sum of the integrals with respect to the conditional component-wise cumulative hazard function instead of the component-wise cumulative hazard function. That is, define

$$\theta_{1r} = \int P(T_2 > t | T_1 > t) g_r(t) \log(1 - d\Lambda_{X_{01|2>}}) \\ + \int P(T_2 < t | T_1 > t) g_r(t) \log(1 - d\Lambda_{X_{01|2<}}),$$

for the constraints on the first component, with $\Lambda_{X_{01|2>}}$ and $\Lambda_{X_{01|2<}}$ the conditional component-wise cumulative hazard functions. Note that this approach might be more precise. However, in order to test the hypothesis one must make assumptions on the conditional probabilities, which are usually unknown and difficult to estimate in advance in the absence of a pilot study.

Theorem 2. *The NPMLE of $G_{\text{constraints}}$, as given in Equation (8), are given by*

$$\hat{a}_l^* = a_l(\lambda_1) = \frac{u_{2l}}{u_{1l} + u_{2l} + z_{1l}n\lambda_1^T \mathbf{G}_1(t_{1l})}, \quad \hat{b}_l^* = b_l(\lambda_2) = \frac{u_{4l}}{u_{3l} + u_{4l} + z_{2l}n\lambda_2^T \mathbf{G}_2(t_{2l})}, \\ \hat{c}_l^* = c_l(\lambda_1) = \frac{u_{6l}}{u_{5l} + u_{6l} + z_{3l}n\lambda_1^T \mathbf{G}_1(t_{1l})}, \quad \hat{d}_l^* = d_l(\lambda_2) = \frac{u_{8l}}{u_{7l} + u_{8l} + z_{4l}n\lambda_2^T \mathbf{G}_2(t_{2l})},$$

where λ_1 and λ_2 are obtained as the solution of the following $2k$ equations:

$$\sum_i g_{11}(t_{1i}) \log(1 - v_{1i}(\lambda_1)) = \mu_{11}, \dots, \sum_i g_{2k}(t_{2i}) \log(1 - v_{2i}(\lambda_2)) = \mu_{2k}, \quad (9)$$

with $v_{1i}(\lambda_1) := z_{1l}a_l(\lambda_1) + z_{3l}c_l(\lambda_1)$ and $v_{2i}(\lambda_2) := z_{2l}b_l(\lambda_2) + z_{4l}d_l(\lambda_2)$.

Further $\mathbf{G}_1(t_{1i}) := (g_{11}(t_{1i}), \dots, g_{1k}(t_{1i}))^T$ and $\mathbf{G}_2(t_{2i}) := (g_{21}(t_{2i}), \dots, g_{2k}(t_{2i}))^T$.

Proof. Deriving the approximation of $G_{\text{constraint}}$ with respect to the λ s simply results in the constraints. For instance, the derivatives of $G_{\text{constraint}}$ with respect to (w.r.t.) a_l leads to

$$\frac{u_{2l}}{a_l} - \frac{u_{1l}}{1 - a_l} - n\lambda_1^T \mathbf{G}_1(t_{1l}) \cdot \frac{z_{1l}}{1 - a_l}.$$

Equalizing the previous expression to zero and solving for a_l leads to the NPMLE

$$\hat{a}_l^* = \frac{u_{2l}}{u_{1l} + u_{2l} + z_{1l}n\lambda_1^T \mathbf{G}_1(t_{1l})}.$$

The remaining derivations follow analogously. ■

Let the test statistic in terms of hazards be given by

$$W = -2\{\max G_{\text{constraint}} - \max \text{EL}(\Lambda_X)\},$$

where $\text{EL}(\Lambda_X)$ is empirical log-likelihood as given in Equation (7).

Theorem 3. *Suppose that the null hypothesis H_0 , as defined in Equation (4), holds for nonnegative, random functions $g_{jr}(t)$ that are predictable w.r.t. the filtration $\mathcal{F}_t = \sigma\{T_{1i}\mathbf{1}\{T_{1i} \leq t\}; \delta_{1i}\mathbf{1}\{T_{1i} \leq t\}; T_{2i}\mathbf{1}\{T_{2i} \leq t\}; \delta_{2i}\mathbf{1}\{T_{2i} \leq t\}; i = 1, \dots, n\}$. Then, under regularity conditions specified in the Appendix, the test statistic W has asymptotically a chi-squared distribution with $2k$ degrees of freedom, where k is the number of constraints per dimension.*

As in the case of independent components we can vary the number of constraints per component, such that for every component a fitting number of constraints can be chosen.

2.3 | General d -dimensional one-sample EL

As mentioned earlier, the presented approach can be extended to any arbitrary number of components. However, the notation in a case beyond the bivariate model gets rather complex. In general, one considers all conditional hazards that could possibly occur. For the three-dimensional case this would be

$$\begin{aligned} \lambda_{111}^{100}(t)dt &= P(t \leq X_1 \leq t + dt | X_1 \geq t, X_2 > t, X_3 > t) =: a_1(t), \\ \lambda_{110}^{100}(t)dt &= P(t \leq X_1 \leq t + dt | X_1 \geq t, X_2 > t, X_3 < t) =: a_2(t), \\ \lambda_{101}^{100}(t)dt &= P(t \leq X_1 \leq t + dt | X_1 \geq t, X_2 < t, X_3 > t) =: a_3(t), \\ \lambda_{100}^{100}(t)dt &= P(t \leq X_1 \leq t + dt | X_1 \geq t, X_2 < t, X_3 < t) =: a_4(t), \\ \lambda_{111}^{010}(t)dt &= P(t \leq X_2 \leq t + dt | X_1 > t, X_2 \geq t, X_3 > t) =: b_1(t), \\ \lambda_{111}^{001}(t)dt &= P(t \leq X_3 \leq t + dt | X_1 > t, X_2 > t, X_3 \geq t) =: c_1(t), \end{aligned}$$

and so forth. In general, the number of conditional hazards in the d -dimensional case is given by $d \cdot 2^{(d-1)}$. The remaining derivations follow analogously to the bivariate case. Note that the complete hazard is then summed up over all conditions, which are four for the three-dimensional case. Explicitly this means for the hazard of the first component that

$$v_{1l} = P(T_2 > t_{1l}, T_3 > t_{1l} | T_1 \geq t_{1l}) \cdot a_{1l} + \dots + P(T_2 < t_{1l}, T_3 < t_{1l} | T_1 \geq t_{1l}) \cdot a_{4l}.$$

Denote the log-likelihood function, similar to Equation (7), by EL and consider constraints of the form

$$\sum_i g_{jr}(t_{ji}) \log(1 - v_{ji}) = \mu_{jr}, \quad j = 1, \dots, d, \quad r = 1, \dots, k_j.$$

Then the function to maximize is given by

$$G_{\text{constraint}}(\boldsymbol{\mu}) = \text{EL}(\lambda) + \sum_{j=1}^d \sum_{r=1}^{k_j} n \lambda_{jr} \left[\mu_{jr} - \sum_i g_{jr}(t_{ji}) \log(1 - v_{ji}) \right].$$

Theorem 4. Suppose that the null hypothesis as defined in Equation (4) holds for nonnegative, random functions $g_{jr}(t)$. Under the regularity conditions specified in the Appendix, the test statistic defined by

$$W = W(\boldsymbol{\mu}) := -2\{\max G_{\text{constraint}}(\boldsymbol{\mu}) - \max \text{EL}(\Lambda_X)\}$$

has asymptotically a chi-squared distribution with $\sum_{j=1}^d k_j$ degrees of freedom with k_j the number of constraints in dimension j .

Remark 7. Although the multiple testing procedure is only applied to hypothesis testing here, it can be used for other forms of inference as well. For example it could be used to obtain confidence regions for a collection of the population quantiles. A straightforward way to construct confidence regions for the true $\boldsymbol{\theta}_0 = (\theta_{011}, \dots, \theta_{01k_1}, \dots, \theta_{0d1}, \dots, \theta_{0dk_d})^T$ works as follows. Define the confidence region to contain all $\boldsymbol{\mu} = (\mu_{11}, \dots, \mu_{dk_d})^T$ for which the test statistic $W(\boldsymbol{\mu})$ as given in Theorem 4 is smaller than the $1 - \alpha$ -quantile of the chi-squared distribution with $\sum_{j=1}^d k_j$ degrees of freedom.

3 | TWO-SAMPLE CENSORED EL WITH MULTIPLE CONSTRAINTS

As it is often of interest to not only test certain aspects of one sample but further to compare it with a second sample, we expand the model to a two-sample censored EL. Now, additionally to our partly censored sample of observations X_1, \dots, X_n , we have a second d -dimensional i.i.d. sample Y_1, \dots, Y_m with distribution function $F_{Y_0}(t)$, marginal distribution functions $F_{Y_{01}}(t), \dots, F_{Y_{0d}}(t)$, cumulative hazard function $\Lambda_{Y_0}(t)$, and marginal cumulative hazard functions $\Lambda_{Y_{01}}(t), \dots, \Lambda_{Y_{0d}}(t)$. Assume that the Y_i s, $i = 1, \dots, m$, are independent of the X_j s, $j = 1, \dots, n$. Again, the sample may be censored, such that we only observe $S_{ji} = \min(Y_{ji}, C_{ji}^Y)$ and $\tau_{ji} = \mathbb{1}\{Y_{ji} \leq S_{ji}\}$, where the $C_i^Y = (C_{1i}^Y, \dots, C_{di}^Y)$ denote the censoring times for the second sample, which shall be independent from the Y_i . Denote the ordered, distinct values of the S_{ji} as s_{jl} .

Note that we do not require independence between the individual components for this method.

3.1 | Bivariate two-sample EL

The EL function based on the two samples is given by

$$\log \text{EL}(\Lambda_X + \Lambda_Y) = \text{EL}_X + \text{EL}_Y, \quad (10)$$

where EL_X corresponds to the EL function of the observations X , as given in Equation (7), and EL_Y corresponds to the EL function of the observations Y , analogously to EL_X . More explicitly, EL_Y is given by

$$\sum_{s_{1l}} \left[\log(1 - \alpha_l) \sum_i (\mathbb{1}\{S_{1i} > s_{1l}\} \mathbb{1}\{S_{2i} > s_{1l}\}) + \log(\alpha_l) \sum_i (\mathbb{1}\{S_{1i} = s_{1l}\} \mathbb{1}\{S_{2i} > s_{1l}\} \tau_{1i}) \right] \\ \sum_{s_{2l}} \left[\log(1 - \beta_l) \sum_i (\mathbb{1}\{S_{1i} > s_{2l}\} \mathbb{1}\{S_{2i} > s_{2l}\}) + \log(\beta_l) \sum_i (\mathbb{1}\{S_{1i} > s_{2l}\} \mathbb{1}\{S_{2i} = s_{2l}\} \tau_{2i}) \right]$$

$$\sum_{s_{1l}} \left[\log(1 - \gamma_l) \sum_i (\mathbb{1}\{S_{1i} > s_{1l}\} \mathbb{1}\{S_{2i} < s_{1l}\} \tau_{2i}) + \log(\gamma_l) \sum_i (\mathbb{1}\{S_{1i} = s_{1l}\} \mathbb{1}\{S_{2i} < s_{1l}\} \tau_{1i} \tau_{2i}) \right]$$

$$\sum_{s_{2l}} \left[\log(1 - \varepsilon_l) \sum_i (\mathbb{1}\{S_{1i} < s_{2l}\} \mathbb{1}\{S_{2i} > s_{2l}\} \tau_{1i}) + \log(\varepsilon_l) \sum_i (\mathbb{1}\{S_{1i} < s_{2l}\} \mathbb{1}\{S_{2i} = s_{2l}\} \tau_{1i} \tau_{2i}) \right],$$

where the sums over s_{jl} are only taken over those time points where α_l , respectively β_l , γ_l , and ε_l are truly greater than zero, and excluding the last time points, as there the hazards would be 1. Again, we will denote the eight sums over i with $\nu_{1l}, \dots, \nu_{8l}$, in the same order as they appear above.

Lemma 2. Maximizing the log-likelihood function w.r.t. $a_l, \dots, \varepsilon_l$, one obtains the NPMLE

$$\hat{a}_l = \frac{u_{2l}}{u_{1l} + u_{2l}}, \quad \hat{b}_l = \frac{u_{4l}}{u_{3l} + u_{4l}}, \quad \hat{c}_l = \frac{u_{6l}}{u_{5l} + u_{6l}}, \quad \hat{d}_l = \frac{u_{8l}}{u_{7l} + u_{8l}},$$

$$\hat{\alpha}_l = \frac{\nu_{2l}}{\nu_{1l} + \nu_{2l}}, \quad \hat{\beta}_l = \frac{\nu_{4l}}{\nu_{3l} + \nu_{4l}}, \quad \hat{\gamma}_l = \frac{\nu_{6l}}{\nu_{5l} + \nu_{6l}}, \quad \hat{\varepsilon}_l = \frac{\nu_{8l}}{\nu_{7l} + \nu_{8l}},$$

where l indicates the time point of the jump.

Let us now consider a hypothesis testing problem for a $2k$ -dimensional parameter $\theta = (\theta_{11}, \dots, \theta_{1k}, \theta_{21}, \dots, \theta_{2k})^T$ w.r.t. the cumulative hazard functions Λ_X and Λ_Y such that

$$H_0 : \theta = \mu \text{ vs. } H_A : \theta \neq \mu, \tag{11}$$

where

$$\theta_{jr} = \int g_{jr}(t) \log(1 - d\Lambda_{X0j}(t)) - \int h_{jr}(t) \log(1 - d\Lambda_{Y0j}(t)), \quad r = 1, \dots, k, \quad j = 1, 2,$$

for some given functions $g_{jr}(t)$ and $h_{jr}(t)$. Then, the constraints imposed on v_{ji} and w_{ji} shall be given by

$$\mu_{jr} = \sum_i g_{jr}(t_{ji}) \log(1 - v_{ji}) - \sum_i h_{jr}(s_{ji}) \log(1 - w_{ji}), \tag{12}$$

with $j = 1, 2, r = 1, \dots, k$, and where the sum is taken over all distinct time points in each sample, excluding the last value as the hazard would be 1 at that time point. The w_{ji} s are defined analogously to v_{ji} s,

$$w_{1i}(\lambda_1) := \zeta_{1l} \alpha_l(\lambda_1) + \zeta_{3l} \gamma_l(\lambda_1) \text{ and } w_{2i}(\lambda_2) := \zeta_{2l} \beta_l(\lambda_2) + \zeta_{4l} \varepsilon_l(\lambda_2).$$

Application of the Lagrange multiplier method on the likelihood with constraints as given in Equation (12) shows that the NPMLE are given through

$$\hat{a}_l^* = \frac{u_{2l}}{u_{1l} + u_{2l} + z_{1l} n^* \lambda_1^T \mathbf{G}_1(t_{1l})}, \quad \hat{b}_l^* = \frac{u_{4l}}{u_{3l} + u_{4l} + z_{2l} n^* \lambda_2^T \mathbf{G}_2(t_{2l})},$$

$$\hat{c}_l^* = \frac{u_{6l}}{u_{5l} + u_{6l} + z_{3l} n^* \lambda_1^T \mathbf{G}_1(t_{1l})}, \quad \hat{d}_l^* = \frac{u_{8l}}{u_{7l} + u_{8l} + z_{4l} n^* \lambda_2^T \mathbf{G}_2(t_{2l})},$$

$$\hat{\alpha}_l^* = \frac{\nu_{2l}}{\nu_{1l} + \nu_{2l} - \zeta_{1l} n^* \lambda_1^T \mathbf{H}_1(s_{1l})}, \quad \hat{\beta}_l^* = \frac{\nu_{4l}}{\nu_{3l} + \nu_{4l} - \zeta_{2l} n^* \lambda_2^T \mathbf{H}_2(s_{2l})},$$

$$\hat{\gamma}_l^* = \frac{\nu_{6l}}{\nu_{5l} + \nu_{6l} - \zeta_{3l} n^* \lambda_1^T \mathbf{H}_1(s_{1l})}, \quad \hat{\varepsilon}_l^* = \frac{\nu_{8l}}{\nu_{7l} + \nu_{8l} - \zeta_{4l} n^* \lambda_2^T \mathbf{H}_2(s_{2l})}, \tag{13}$$

where $\mathbf{G}_1(t_{1l})$, $\mathbf{G}_2(t_{2l})$, $\mathbf{H}_1(s_{1l})$, and $\mathbf{H}_2(s_{2l})$ denote the vectors of constraints imposed on the hazards at the time point t_{jl} , respectively s_{jl} , and n^* is defined as $\min(n, m)$.

Let $G_{\text{constraint}}$ be given by

$$G_{\text{constraint}} = \text{EL}_X(\lambda) + \text{EL}_Y(\lambda) + \sum_{j=1}^2 \sum_{r=1}^k n^* \lambda_{jr} \left[\mu_{jr} - \left(\sum_i g_{jr}(t_{ji}) \log(1 - v_{ji}) - \sum_i h_{jr}(s_{ji}) \log(1 - w_{ji}) \right) \right],$$

where $\text{EL}_X(\lambda)$ and $\text{EL}_Y(\lambda)$ denote the log-likelihood function without constraints based on the modified hazards $a_l(\lambda), \dots, \varepsilon_l(\lambda)$ instead of the Nelson–Aalen estimators.

The two-sample test statistic is given by

$$W^* := -2(\max G_{\text{constraint}} - \max(\text{EL}_X + \text{EL}_Y)). \quad (14)$$

Theorem 5. *Suppose that the null hypothesis $H_0 : \theta_{jr} = \mu_{jr}$ holds for all $j = 1, 2$ and $r = 1, \dots, k$, that is,*

$$\mu_{jr} = \int g_{jr}(t) \log(1 - d\Lambda_{X0j}(t)) - \int h_{jr}(t) \log(1 - d\Lambda_{Y0j}(t)),$$

for nonnegative random functions $g_{jr}(t)$ and $h_{jr}(t)$ that are predictable w.r.t. the filtration \mathcal{F}_t specified in Appendix. Then, under further conditions specified in the regularity specification for the two-sample case, as $n^* \rightarrow \infty$ and $n/m \rightarrow c \in (0, \infty)$, W^* , as given in Equation (14), has asymptotically a chi-squared distribution with $2k$ degrees of freedom.

Again the number of constraints can be fitted individually to match each component. Needless to mention, the number of constraints for the two samples must match.

3.2 | General d -dimensional two-sample EL

As in the one-sample case, the proposed method works for any arbitrary number of components. Due to complexity of notation we only sketch the idea of the general d -dimensional two-sample test.

Let $\text{EL}(\lambda)$ be the sum of the individual EL functions in each component with again the modified hazards instead of the Nelson–Aalen estimators. Further let EL be the sum of the individual EL functions in each component with the Nelson–Aalen estimators for the hazards. The function to maximize under the constraints, $G_{\text{constraint}}$, is given by

$$G_{\text{constraint}} = \text{EL}(\lambda) + \sum_{j=1}^d \sum_{r=1}^{k_j} n^* \lambda_{jr} \left[\mu_{jr} - \left(\sum_i g_{jr}(t_{ji}) \log(1 - v_{ji}) - \sum_i h_{jr}(s_{ji}) \log(1 - w_{ji}) \right) \right].$$

Theorem 6. *Suppose that the null hypothesis $H_0 : \theta_{jr} = \mu_{jr}$ holds for all $j = 1, \dots, d$ and $r = 1, \dots, k_j$, that is,*

$$\mu_{jr} = \int g_{jr}(t) \log(1 - d\Lambda_{X0j}(t)) - \int h_{jr}(t) \log(1 - d\Lambda_{Y0j}(t)),$$

for nonnegative random functions $g_{jr}(t)$ and $h_{jr}(t)$ that are predictable w.r.t. the filtration \mathcal{F}_t specified in Appendix. Then under further conditions specified in the regularity specification for the two-sample case, as $n^* \rightarrow \infty$ and $n/m \rightarrow c \in (0, \infty)$,

$$W^* = -2 \{ \max G_{\text{constraint}} - \max EL \},$$

has asymptotically a chi-squared distribution with $\sum_{j=1}^d k_j$ degrees of freedom.

4 | APPLICATIONS

In this section we discuss different applications of Theorem 5. The constraints proposed in the following are based on random yet predictable constraint functions g_{jr} and h_{jr} .

One application of Theorem 5 is based on constraints obtained by combining the log-rank test with the Gehan test. We consider a two-sample problem with $H_0 : \Lambda_X(t) = \Lambda_Y(t)$ for all $t \geq 0$ versus $H_1 : \Lambda_X(t) \neq \Lambda_Y(t)$ for some t . Specifically the test statistics of the individual tests can be given in the form of

$$\sum_i h_j^*(t_{ji}, \rho, \gamma) \log(1 - v_{ji}) - \sum_l h_j^*(s_{jl}, \rho, \gamma) \log(1 - w_{jl}) = \mu_j, \quad j = 1, 2,$$

where

$$h_j^*(u, \rho, \gamma) = \left(\frac{n+m}{nm} \right)^{1/2} W_j(u)^\rho (1 - W_j(u))^\gamma \frac{R_{j1}(u)R_{j2}(u)}{R_{j1}(u) + R_{j2}(u)}, \text{ for } \rho, \gamma \geq 0,$$

as well as $R_{j1}(u) = \sum \mathbb{1}\{T_{ji} \geq u\}$ and $R_{j2}(u) = \sum \mathbb{1}\{S_{ji} \geq u\}$.

If $W_j(u)$ equals $(R_{j1}(u) + R_{j2}(u))/(n + m)$ then $(\rho, \gamma) = (0, 0)$ corresponds to the log-rank statistic and $(\rho, \gamma) = (1, 0)$ corresponds to the Gehan–Wilcoxon statistic. Note that $h_j^*(u, \rho, \gamma), j = 1, 2$, are indeed predictable functions.

For the proposed method, we used

$$\sum_i h_j^*(t_{ji}, \rho, \gamma) \log(1 - v_{ji}) - \sum_l h_j^*(s_{jl}, \rho, \gamma) \log(1 - w_{jl}) = 0,$$

as constraints for the two combinations of (ρ, γ) mentioned above. More specifically, we applied the constraint functions

$$g_{j1}(u) = h_{j1}(u) = \left(\frac{n+m}{nm} \right) \frac{R_{j1}(u)R_{j2}(u)}{R_{j1}(u) + R_{j2}(u)}, \quad j = 1, 2, \tag{15}$$

namely the log-rank statistic, and

$$g_{j2}(u) = h_{j2}(u) = \left(\frac{n+m}{nm} \right) \frac{R_{j1}(u) + R_{j2}(u)}{n+m}, \quad j = 1, 2, \tag{16}$$

the Gehan–Wilcoxon statistic, to each component leading to four degrees of freedom.

A comparison of the combined test based on these constraints with the two individual tests was performed in a simulation study. It corresponds to Simulation 4 in Section 5.

Further, it might be of interest to test whether two distributions which have identical marginals are identically distributed also in the multivariate sense. In practice, it is possible that

two datasets differ only in the correlation structure, which can easily go undetected. To this end, one may apply functions of the form

$$f(t) = \mathbb{1}\{\exists j \text{ s.t. } T_{2j} = t, \delta_{2j} = 1 \text{ and } T_{1j} > t\}g(t), \quad (17)$$

with $g(t)$ a nonnegative predictable function. Applying, for instance, five constraints we can use four to focus on testing the two datasets for equal marginals. Two may be used for the first component, where one of them puts weight on early events, and the other one on late events, and the same for the second component. The last constraint can be used to control whether the correlation within each of the two datasets is similar, and it can be applied to either component. If it is of the form Equation (17) it should be applied to the second component. As one can see in Section 5 this leads to a high rejection level if the correlation structure is not the same, even if the samples are drawn from identical marginal distributions.

Instead of using random, predictable functions, one can also apply various deterministic functions as constraints. Those can be chosen such that the test focuses on the part of the data where differences are to be expected or are considered as critical time frames. For instance, if one is only interested in detecting differences at the beginning of the study, constraints of the form $g_1(t) = \exp\{-t\}$ or $g_2(t) = \mathbb{1}\{t \leq t^*\}$, $t^* \in \mathbb{R}^+$ can be applied. While $g_1(t)$ assigns greater weight to earlier time points and lesser weight to later ones, $g_2(t)$ weights all time points up to a pre-specified time point t^* equally and does not consider later time points at all. On the other hand, if one is interested in differences at later time points, one can apply constraints of the form $g_1(t) = \mathbb{1}\{t \geq t^*\}$, $t^* \in \mathbb{R}^+$ or $g_2(t) = \log(t + 1)$, among others. Obviously, it is possible to combine two constraints so that one detects differences at early time points and one at later time points. Constraints of this form were used in the first simulation as given in Section 5.

5 | SIMULATIONS AND DATA EXAMPLE

We provide simulation results that confirm the chi-squared limiting distributions of W and W^* and illustrate the small sample performance. Further, we ran some simulations to show how dependency and correlation structures influence the results. We then compared the performance of our test to various others. A real data example is provided to illustrate the proposed method.

All computations were performed using R (R version 3.3.2, R Core Team, 2017).

Simulation 1 (one-sample case). This simulation illustrates the empirical distribution of W when the constraints are given by the functions

$$g_1(t) = \exp\{-t\}, \quad g_2(t) = 0.5t\mathbb{1}\{t \leq 1\}, \quad g_3(t) = \mathbb{1}\{0.5 < t < 1.5\}.$$

All constraints were applied to both components. We used the following distributions to generate the bivariate random variables:

$$X_1 \sim \exp(0.9), \quad X_2 \sim \exp(0.6), \quad C_1 \sim \exp(0.5), \quad C_2 \sim \exp(0.3).$$

The actual observed data was then created analogously to Equation (1). Censoring was roughly one-third in each of the components. The observations were uncensored in both components in 42% of the cases.

The Q-Q plot (Figure 1) is based on 10,000 runs and a sample size of 200. The resulting empirical distribution of W agrees well with the theoretically derived \mathcal{X}_6^2 -distribution.

Simulation 2 (two-sample case). In this simulation we illustrate the empirical distribution of W^* for the two-sample case. The constraints are given by the functions

$$g_1(t) = \exp\{-t\}, \quad g_2(t) = 0.5t\mathbb{1}\{t \leq 1\},$$

$$g_3(t) = \mathbb{1}\{t \leq 0.9\}, \quad g_4(t) = 2\exp\{-t\}\mathbb{1}\{t > 0.5\}.$$

The functions g_1 and g_2 were applied to the first component and the remaining two to the second component. The two samples were drawn from the following marginal distributions:

$$X_1, Y_1 \sim \exp(0.9), \quad X_2, Y_2 \sim \exp(0.6), \quad C_1^X, C_1^Y \sim \exp(0.5), \quad C_2^X, C_2^Y \sim \exp(0.3).$$

The simulated actual observations after censoring were again created according to Equation (1). The amount of censoring was 36% for the first component and 73% for the second component.

The four Q-Q plots in Figure 2 are based on 10,000 runs and show different combinations of sample sizes. The resulting empirical distribution of W generally agreed well with the theoretically derived \mathcal{X}_4^2 -distribution. However, in the case of $m = n = 50$ it only agreed well with the theoretical distribution up to the 90% quantile.

Table 1 shows the Type I errors at various significance levels and different sample sizes. The proposed combined test attained the Type I error quite well at the nominal levels for sample sizes of $n, m \geq 70$.

Simulation 3 (Dependency/correlation structure). The previous simulations were performed with random variables which had independent components. The following simulations show the influence of high dependency on the test statistic and the consequences of identically distributed components but with different correlation structures in the two datasets. First, we considered bivariate random variables which were identically distributed in both samples. For the first component we drew samples according to

$$X_1, Y_1 \sim \exp(0.9), \quad C_1^X, C_1^Y \sim \exp(0.3).$$

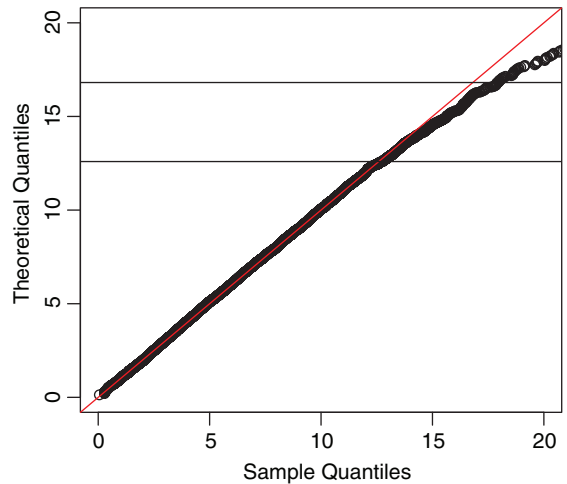


FIGURE 1 Q-Q plot of the empirical quantiles versus \mathcal{X}_6^2 percentiles for sample size 200 in the one sample case, corresponding to the theoretical result in Theorem 3 [Colour figure can be viewed at wileyonlinelibrary.com]

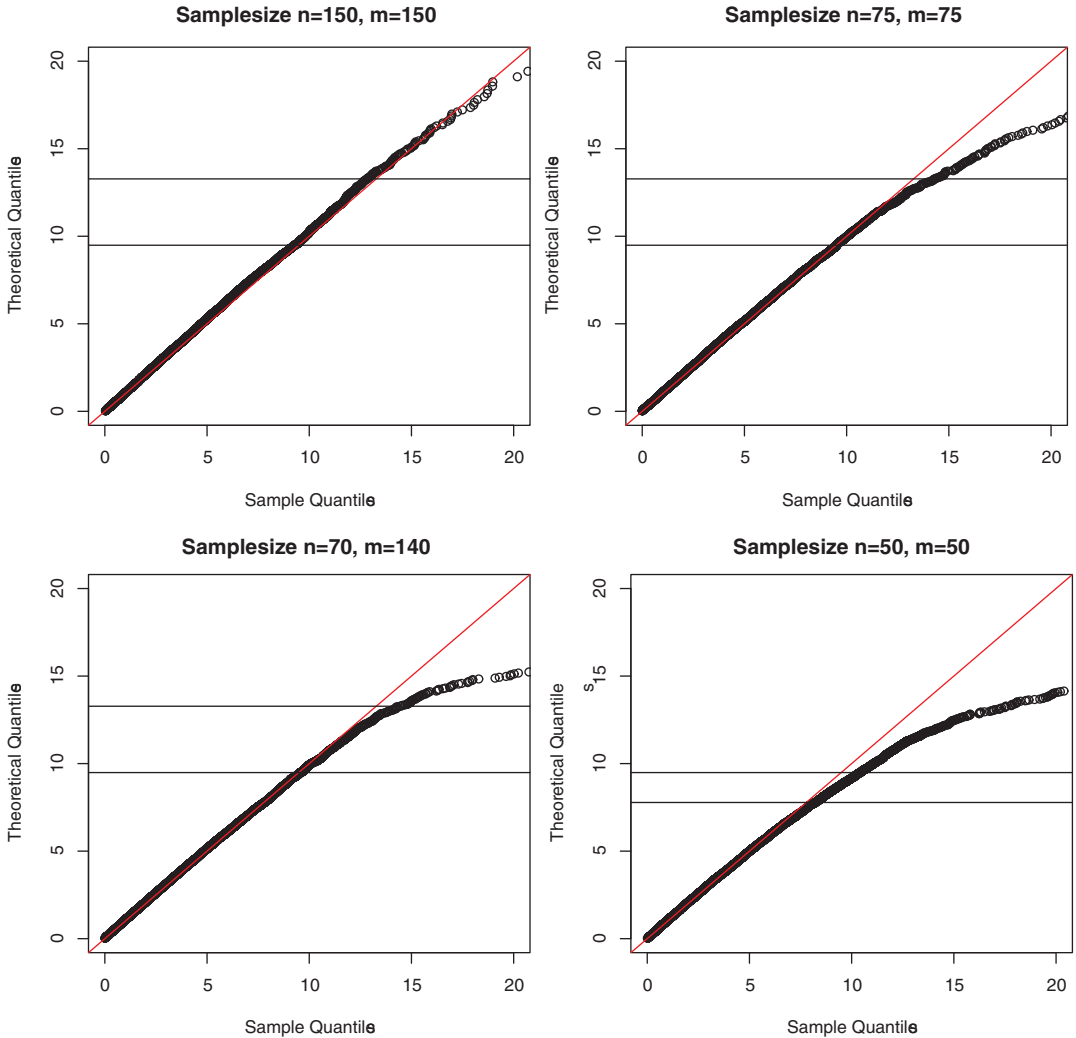


FIGURE 2 Q-Q plot of the empirical quantiles versus \mathcal{X}_4^2 percentiles for different sample sizes in the two-sample case, corresponding to the result in Theorem 5 [Colour figure can be viewed at wileyonlinelibrary.com]

| | | α | | | | |
|-----|-----|----------|--------|--------|--------|--------|
| n | m | 0.01 | 0.05 | 0.1 | 0.15 | 0.2 |
| 50 | 50 | 0.0210 | 0.0667 | 0.1140 | 0.1593 | 0.2060 |
| 75 | 75 | 0.0128 | 0.0505 | 0.0945 | 0.1416 | 0.1886 |
| 150 | 150 | 0.0082 | 0.0468 | 0.0884 | 0.1302 | 0.1769 |
| 70 | 140 | 0.0132 | 0.0517 | 0.1007 | 0.1463 | 0.1925 |

TABLE 1 Estimated Type I errors at various significance levels α and different sample sizes for two samples, corresponding to the theoretical results in Theorem 5

The second components were obtained by $X_2 = 0.65X_1 + 0.35X$ and $Y_2 = 0.65Y_1 + 0.35Y$, with $X, Y \sim \exp(0.6)$. Censoring was still considered to be i.i.d., such that $C_2^X, C_2^Y \sim \exp(0.3)$. The censoring accumulated to 25% in the first component and 30% in the second. Overall 55% of the observations could be observed completely. For each dataset we simulated 200 observations. The constraints were given by the functions

$$\begin{aligned} g_1(t) &= \exp\{-t\}, \quad g_2(t) = 0.5t\mathbb{1}\{t \leq 1\}, \\ g_3(t) &= \mathbb{1}\{t \leq 0.9\}, \quad g_4(t) = 2\exp\{-t\}\mathbb{1}\{t > 0.5\}. \end{aligned}$$

Here, g_1 and g_2 were applied to the first component, the remaining two to the second component.

The Q-Q plot (Figure 3) is based on 10,000 runs. The resulting empirical distribution of W^* agreed with the theoretically derived \mathcal{X}_4^2 -distribution up to the 90% quantile. For even stronger dependencies, higher sample sizes are needed to obtain similar empirical results. Yet we can conclude that with a sufficient number of observations, the proposed method can deal with dependencies.

Correlation between the components may also be of interest in the case of multivariate data. Extreme cases of testing two samples for equality could be given by having identical marginals but different correlation structure. A test should reject the null hypothesis of identical joint survival distributions in such cases. Both samples were simulated by

$$X_1, Y_1 \sim \exp(0.9), \quad X_2, Y_2 \sim \exp(0.9), \quad C_1^X, C_1^Y \sim \exp(0.2), \quad C_2^X, C_2^Y \sim \exp(0.2).$$

To obtain a different correlation structure, the samples of X_1 and of X_2 were sorted. The correlation of the true observations of X was given by .99 and for Y by 0. The actual observed data of the X -sample, T_X , was only correlated by .68. The correlation for the actual observed data for the Y -sample remained the same. The constraints were given by

$$g_1(t) = \exp\{-t\}, \quad g_2(t) = \mathbb{1}\{t \leq 1.5\}, \quad g_3(t) = 2\mathbb{1}\{0.5 < t < 1.5\},$$

and g_4 as defined in Equation (17) with $g(t) = 1$. Again, the first two constraints were imposed on component 1 and the other two on component 2. The simulated power was 81.79% indicating that the proposed method indeed is capable of detecting differences in the correlation structure. As one can see in Simulation 4, other methods have limited capability to detect such differences.

Simulation 4 (Comparison with other methods). We compare the small and moderate sample size behaviors of the proposed combined tests, as described in the first application in Section 4, to two tests from the $G^{p,\gamma}$ family of Harrington and Fleming (1982). As the constraints correspond to the weight functions of the log-rank statistic and the Gehan statistic, we compared the combined constraints to the individual tests. For all three tests, we estimated the Type I errors for one null hypothesis and the power for five different alternatives. The size of the simulation was 10,000 for the null model and 5,000 for each of the alternative models A–E, as defined in Table 2. The sample size for the small sample size setting was given by 50 X -observations and 50 Y -observations. For the moderate sample size setting we had 150 X -observations and 150 Y -observations. The five different survival configurations for the power analyses A–E are described as follows. A corresponds to identical marginals but different correlation structure, B to one component identically distributed and one different, C to crossing hazards, D to ordered

hazards with early differences, and E to ordered hazards with late differences. Samples were generated both without censoring and with exponentially distributed censoring with parameter $\lambda_1 = 0.5$ for the first component and $\lambda_2 = 0.3$ for the second component such that censoring amounted to 66% and 64% in the two components in case of the null model. Further information on the modeling can be found in Table 2.

For alternative A, an additional constraint was added to the second component, which was given by the constraint function in Equation (17) with $g(t) = 1$. The other constraints correspond to Equations (15) and (16), which are random predictable functions w.r.t. the filtration we specified in the Appendix B1. Both constraints were applied to each component.

From Tables 3 and 4, the proposed method had comparable performance to the two individual test statistics. In configurations A and C, it performed significantly better than the two individual tests, even though the sample size of 50 was fairly small. In scenarios where only the marginals differed in one component and the correlation structure was similar, it performed poorly compared to the log-rank test and the Gehan–Wilcoxon test. In the scenarios for early differences in the hazard functions, it obtained almost the performance level of Gehan–Wilcoxon, while the log-rank test lost a lot of power. For late differences in the hazards, it was the opposite. As an additional comparison between the amount of constraints, we performed additional tests on alternative E, where we added a fifth constraint corresponding to the constraint function in Equation (17) with $g(t) \equiv 1$. The resulting Type II error dropped in that case from 73.52 to 60.72 in the censored case and from 27.14 to 21.24 in the uncensored case. In the later case it slightly outperformed the log-rank test. Adding an additional constraint led to a lower error rate in this scenario. All tests performed better when no censoring occurred, yet one could observe similar patterns as in the censored case for the power comparisons.

Tables 5 and 6 show that similar results were obtained for moderate sample sizes. All test obtained lower error rates than in the small sample case, but the same properties as in the small sample simulations can be noticed. The strengths of the newly proposed test can be seen here even more clearly in scenarios A and C, which correspond to correlation and crossing hazards, respectively.

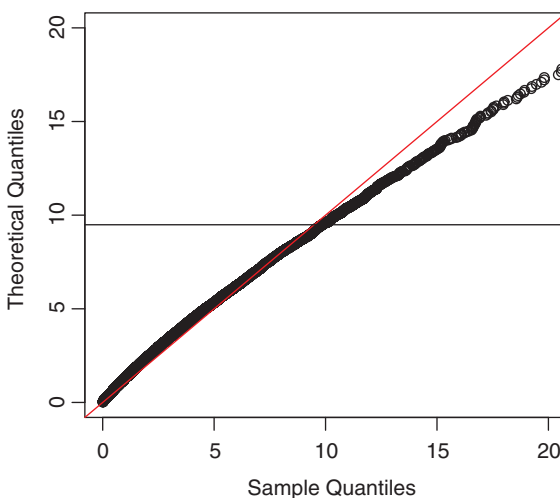


FIGURE 3 Q-Q plot of the empirical quantiles versus χ^2_4 percentiles for sample size 200 in the two-sample case, corresponding to the theoretical result in Theorem 5 [Colour figure can be viewed at wileyonlinelibrary.com]

5.1 | Data example

In order to illustrate our results on a real dataset, we chose the dataset *retinopathy* from the R-package *survival*. This package contains the core survival analysis routines. The dataset includes 394 observations out of a total of 197 patients from the Diabetic Retinopathy Study, which was conducted by the National Eye Institute. The observed data was part of a trial on laser coagulation as a treatment to delay diabetic retinopathy. Each patient received laser treatment on one

TABLE 2 Hazard functions for the configurations used in the power studies (Simulation 4)

| Configuration | Hazard functions | df |
|---------------|--|----|
| A | $\lambda_{X1}(t) = \lambda_{X2}(t) = \lambda_{Y1}(t) = \lambda_{Y2}(t) = 0.9, t \geq 0$ $\text{cor}(X1, X2) = 0.99, \text{cor}(Y1, Y2) = 0$ | 5 |
| B | $\lambda_{X1}(t) = \lambda_{X2}(t) = \lambda_{Y1}(t) = 0.9, t \geq 0,$ $\lambda_{Y2}(t) = 0.5, t \geq 0$ | 4 |
| C | $\lambda_{X1}(t) = \begin{cases} 0.9, & t \leq 0.5 \\ 0.6, & t > 0.5 \end{cases}, \lambda_{X2}(t) = \begin{cases} 0.7, & t \leq 0.5 \\ 0.4, & t > 0.5 \end{cases}$ $\lambda_{Y1}(t) = \begin{cases} 0.6, & t \leq 0.5 \\ 0.9, & t > 0.5 \end{cases}, \lambda_{Y2}(t) = \begin{cases} 0.4, & t \leq 0.5 \\ 0.7, & t > 0.5 \end{cases}$ | 4 |
| D | $\lambda_{X1}(t) = \begin{cases} 0.9, & t \leq 0.5 \\ 0.7, & t > 0.5 \end{cases}, \lambda_{X2}(t) = \begin{cases} 0.4, & t \leq 0.5 \\ 0.6, & t > 0.5 \end{cases}$ $\lambda_{Y1}(t) = \begin{cases} 0.4, & t \leq 0.5 \\ 0.7, & t > 0.5 \end{cases}, \lambda_{Y2}(t) = \begin{cases} 0.9, & t \leq 0.5 \\ 0.6, & t > 0.5 \end{cases}$ | 4 |
| E | $\lambda_{X1}(t) = \begin{cases} 0.7, & t \leq 0.4 \\ 0.9, & t > 0.4 \end{cases}, \lambda_{X2}(t) = \begin{cases} 0.6, & t \leq 0.4 \\ 0.5, & t > 0.4 \end{cases}$ $\lambda_{Y1}(t) = \begin{cases} 0.7, & t \leq 0.4 \\ 0.5, & t > 0.4 \end{cases}, \lambda_{Y2}(t) = \begin{cases} 0.6, & t \leq 0.4 \\ 0.9, & t > 0.4 \end{cases}$ | 4 |

TABLE 3 Small sample simulation results of Type I error and Type II error in percent with $n = m = 50$ for a null (N) and for five different alternative survival distribution configurations (A-E) with censored data for three test statistics

| | N | A | B | C | D | E |
|----------------------|------|-------|-------|-------|-------|-------|
| Combined Constraints | 4.83 | 38.42 | 66.14 | 70.94 | 45.90 | 73.52 |
| Log-rank | 5.62 | 93.44 | 46.60 | 91.48 | 56.16 | 57.68 |
| Gehan | 5.29 | 94.24 | 59.12 | 91.28 | 35.04 | 82.40 |

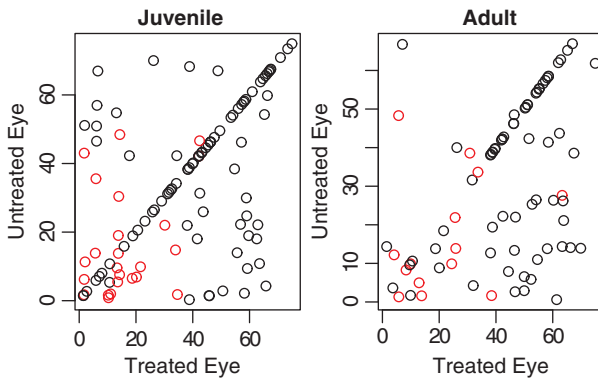


FIGURE 4 Scatter plot of the two groups. Observations which were completely observed were highlighted in red [Colour figure can be viewed at wileyonlinelibrary.com]

of their eyes, while the other remained untreated to obtain a baseline. The eye for treatment was randomly chosen. The two observations per patient contain some basic information which is the same in both and the time to vision loss in the treated and untreated eye, as well as the risk factor for each eye. The time to vision loss was measured from initiation of treatment to the time when visual acuity dropped below 5/200 for two visits in a row. As no event could occur within the first 6.5 months of the study, all survival times were reduced by this duration. The survival times were subject to univariate censoring. All patients were considered to have "high-risk" for diabetic retinopathy as defined by the Diabetic Retinopathy Study.

The patients can be split into two groups, *juvenile* and *adult* diabetes. As the types of diabetes are assumed to have differing progress, it is of interest to analyze the joint survival function of both eyes. The null hypothesis which was analyzed in the following was that both juvenile as well as adult diabetes had the same joint survival function of both eyes. Denote the times to blindness in the treated eye by X_1 for the juvenile group and by Y_1 for the adult group. Further let X_2 be the survival time of the untreated eyes in the juvenile group and let Y_2 be the time in the adult group. With 114 *juvenile* patients and 83 *adult* patients, we have enough observations to narrow the error rate. Censoring rates were 68% for X_1 , 55% for X_2 , 78% for Y_1 , and 40% for Y_2 . In the juvenile group 21% of the patients could be completely observed, in the adult group 17% of the patients.

For the newly proposed method, the constraints (15) and (16) were chosen. In a second calculation we added a fifth constraint, corresponding to Equation (17) with $g(t) = 1$ applied to the second component.

Both tests did not reject the null hypothesis at level 5%. The p values of the new test method were given by $p_1 = .2297$ and $p_2 = .1733$. Former analyses by Huang and Zhao (2018) and Huster, Brookmeyer, and Self (1989) concluded that the null hypothesis should be rejected. However, both did not actually conduct tests on the similarity of the survival functions. Huster et al. (1989) conducted tests on the similarities of the covariates and rejected the null hypothesis of equal influence of the covariates, while Huang and Zhao (2018) only conducted some visual analysis

TABLE 4 Small sample simulation results of Type I error and Type II error in percent with $n = m = 50$ for a null (N) and for five different alternative survival distribution configurations (A–E) with uncensored data for three test statistics

| | N | A | B | C | D | E |
|----------------------|------|-------|-------|-------|-------|-------|
| Combined Constraints | 5.00 | 29.32 | 40.50 | 43.16 | 40.10 | 27.14 |
| Log-rank | 5.60 | 93.60 | 26.28 | 75.74 | 64.36 | 21.74 |
| Gehan | 4.99 | 94.68 | 41.14 | 93.72 | 38.84 | 56.72 |

TABLE 5 Moderate sample simulation results of Type I error and Type II error in percent with $n = m = 150$ for a null (N) and for five different alternative survival distribution configurations (A–E) with censored data for three test statistics

| | N | A | B | C | D | E |
|----------------------|------|-------|-------|-------|-------|-------|
| Combined Constraints | 4.99 | 29.76 | 15.00 | 22.00 | 2.82 | 18.46 |
| Log-rank | 5.38 | 94.16 | 3.90 | 85.54 | 10.68 | 8.12 |
| Gehan | 5.22 | 94.10 | 10.82 | 83.92 | 0.86 | 53.62 |

TABLE 6 Moderate sample simulation results of Type I error and Type II error in percent with $n = m = 150$ for a null (N) and for five different alternative survival distribution configurations (A–E) with uncensored data for three test statistics

| | N | A | B | C | D | E |
|----------------------|------|-------|------|-------|-------|-------|
| Combined Constraints | 5.40 | 25.58 | 0.98 | 1.38 | 1.82 | 0.06 |
| Log-rank | 5.22 | 94.00 | 0.40 | 34.38 | 20.60 | 0.12 |
| Gehan | 5.23 | 93.92 | 2.28 | 91.08 | 1.44 | 10.00 |

of the individual estimators of the survival function and their confidence intervals. By examining Figure 4, not rejecting the null hypothesis does not seem entirely unreasonable.

The proposed test tends to be conservative for very small sample sizes and small sample sizes with a high censoring rate. This could already be observed in the univariate case.

6 | CONCLUSION

On the basis of a previous univariate approach based on empirical likelihood, we developed a multivariate testing procedure for hazards in the survival context with censored observations. Using constraints and conditional hazards, our method is capable of testing various aspects of a dataset containing time-to-event observations. It was shown that our proposed method is asymptotically consistent. The finite sample simulation studies further showed that the power of the proposed method depends on the underlying power associated with the individual constraints. For higher sample sizes, it has approximately the power of the better underlying individual test. In scenarios in which the greatest difference of two datasets lies in the correlation structure, the proposed test outperformed the individual tests from the family of test statistics introduced by Harrington and Fleming (1982).

ACKNOWLEDGEMENTS

This work was supported by a grant from the Austrian Science Fund, FWF I 2697. I thank Arne Bathke for his valuable comments and advice.

ORCID

Judith H. Parkinson  <https://orcid.org/0000-0003-2942-0204>

REFERENCES

Bathke, A., Kim, M.-O., & Zhou, M. (2009). Combined multiple testing by censored empirical likelihood. *Journal of Statistical Planning and Inference*, 139, 814–827.

- Brendel, M., Janssen, A., Mayer, C.-D., & Pauly, M. (2014). Weighted logrank permutation tests for randomly right censored life science data. *Scandinavian Journal of Statistics*, *41*, 742–761.
- Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003a). Survival analysis part I: Basic concepts and first analyses. *British Journal of Cancer*, *89*(2), 232–238.
- Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003b). Survival analysis part II: Multivariate data analysis – An introduction to concepts and methods. *British Journal of Cancer*, *89*(3), 431–436.
- Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003c). Survival analysis part III: Multivariate data analysis—Choosing a model and assessing its adequacy and fit. *British Journal of Cancer*, *89*(4), 605–611.
- Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003d). Survival analysis part IV: Further concepts and methods in survival analysis. *British Journal of Cancer*, *89*(5), 781–786.
- Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *Ann. Statist.*, *16*(4), 1475–1489.
- Gill, R., van der Laan, M., & Wellner, J. (1995). Inefficient estimators of the bivariate survival function for three models. *Annales de l'I.H.P. Probabilités et Statistique*, *31*(3), 545–597.
- Gill, R. (1983). Large Sample Behaviour of the Product-Limit Estimator on the Whole Line. *The Annals of Statistics*, *11*(1), 49–58. <https://doi.org/10.1214/aos/1176346055>
- Harrington, D., & Fleming, T. (1982). A class of rank test procedures for censored survival data. *Biometrika*, *69*, 553–556.
- Huang, H., & Zhao, Y. (2018). Empirical likelihood for the bivariate survival function under univariate censoring. *Journal of Statistical Planning and Inference*, *194*, 32–46.
- Huster, W. J., Brookmeyer, R., & Self, S. G. (1989). Modelling paired survival data with covariates. *Biometrics*, *45*, 145–156.
- Klein, J. P., & Moeschberger, M. (1997). *Survival analysis - Techniques for censored and truncated data*. New York, NY: Springer.
- Kosorok, M. R., & Lin, C.-Y. (1999). The versatility of function-indexed weighted log-rank statistics. *Journal of the American Statistical Association*, *94*(445), 330–332.
- Lin, D., & Ying, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, *80*(3), 573–581.
- Pan, X. R., & Zhou, M. (2002). Empirical likelihood ratio in terms of cumulative hazard function for censored data. *Journal of Multivariate Analysis*, *80*(1), 166–188.
- Spreij, P. (1990). *Counting process systems: Identification and stochastic realization*. Amsterdam, Netherlands: Centrum voor Wiskunde en Informatica.
- Vonta, F., Nikulin, M., Limnios, N., & Huber-Carol, C. (2008). *Statistical models and methods for biomedical and technical systems* (2008th ed.). Basel: Birkhäuser.
- Wang, W., & Wells, M. T. (1997). Nonparametric estimators of the bivariate survival function under simplified censoring conditions. *Biometrika*, *84*(4), 863–880.
- Wang, W., & Zafra, P. (2009). Estimating bivariate survival function by volterra estimator using dynamic programming techniques. *Journal of Data Science*, *7*(3), 365–380.
- Wei, L., & Lachin, J. (1984). Two-sample asymptotically distribution-free tests for incomplete multivariate observations. *Journal of the American Statistical Association*, *79*(387), 653–661.

How to cite this article: Parkinson JH. Combined multiple testing of multivariate survival times by censored empirical likelihood. *Scand J Statist.* 2020;47:757–786. <https://doi.org/10.1111/sjost.12423>

APPENDIX A. PROOF OF THE ONE-SAMPLE CASE

A1. Regularity specification for the one-sample case

Let X_1, \dots, X_n be i.i.d. d -dimensional random variables with cumulative distribution function $F_{X_0}(t)$ and marginals $F_{X_{01}}(t), \dots, F_{X_{0d}}(t)$. The cumulative hazard function is denoted as $\Lambda_{X_0}(t)$ and the marginals as $\Lambda_{X_{01}}(t), \dots, \Lambda_{X_{0d}}(t)$. We observe the d -dimensional vectors $T_i = \min(X_i, C_i^X)$ and

$\delta_i = \mathbb{1}\{X_i \leq C_i^X\}$, where C_i^X are i.i.d. censoring times, independent of X_i , and the minimum and the indicator function are considered component-wise. The cumulative distribution function of the C_i^X is $F_{CX}(t)$. The distribution functions $F_{X0}(t)$ and $F_{CX}(t)$ do not have common discontinuities.

Let $g_{j1}(t), \dots, g_{jk}(t), j = 1, \dots, d$, be nonnegative left continuous functions such that none can be expressed by a linear combination of the remaining functions. Further it shall hold for all functions

$$0 < \int \frac{|g_{jr}(t)|^w (1 - \Delta\Lambda_{X0j}(t))}{(1 - F_{X0j}(t)) \prod_j (1 - F_{CXj}(t))} d\Lambda_{X0j}(t) < \infty, \quad w = 1, 2, \quad j = 1, \dots, d, \quad r = 1, \dots, k.$$

This condition guarantees the asymptotic normality of the Nelson–Aalen estimators of the individual components (cf. Gill, 1983, theorem 2.1). Note that the functions $g_{jr}(t)$ may be random, yet they need to be predictable w.r.t. a filtration. Considering $d = 2$, the filtration is given by $\mathcal{F}_t = \sigma\{T_{1l}\mathbb{1}\{T_{1l} \leq t\}; \delta_{1l}\mathbb{1}\{T_{1l} \leq t\}; T_{2l}\mathbb{1}\{T_{2l} \leq t\}; \delta_{2l}\mathbb{1}\{T_{2l} \leq t\}; l = 1, \dots, n\}$. Through this we are capable of applying the martingale central limit theorem, as $\hat{\Lambda}_{NA,j}^{(X)}(t) - \Lambda_{X0j}(t)$ is a martingale, where $\hat{\Lambda}_{NA,j}^{(X)}$ denotes the Nelson–Aalen estimator of component j . Furthermore, for all random functions $g_{jr}(t)$ there must exist a nonrandom left continuous function $g_{jr0}(t)$ such that $\sup_{t \leq \max T_{ji}} |g_{jr}(t)/g_{jr0}(t)| = o_p(1)$ for $j = 1, \dots, d$ and $r = 1, \dots, k$ as $n \rightarrow \infty$.

A2. Preparation of the proof of Theorem 3

Let $d = 2$ and $f(\lambda)$ be given by

$$\sum_l \log(1 - a_l(\lambda_1))u_{1l} + \log(a_l(\lambda_1))u_{2l} + \log(1 - b_l(\lambda_2))u_{3l} + \log(b_l(\lambda_2))u_{4l} + \log(1 - c_l(\lambda_1))u_{5l} + \log(c_l(\lambda_1))u_{6l} + \log(1 - d_l(\lambda_2))u_{7l} + \log(d_l(\lambda_2))u_{8l}, \tag{A1}$$

with $\lambda = (\lambda_1^T, \lambda_2^T)^T$, $\lambda_j = (\lambda_{j1}, \dots, \lambda_{jk})^T, j = 1, 2, k$ the number of constraints per component, and l ranging over all distinct time points an event occurred excluding again the last one as it would be 1. For cases where a hazard is zero we will define the corresponding summand to be zero. Note that we will drop the indices of λ if it is clear from the context whether λ_1 or λ_2 is considered.

In order to show that the NPMLs of L are indeed the maximums we compute the derivatives of f w.r.t. all λ s and evaluate it at zero.

Without loss of generality we derive f only for $\lambda_{1r}, r = 1, \dots, k$, then

$$\begin{aligned} \frac{\partial}{\partial \lambda_{1r}} f(\lambda) &= \sum_l \frac{u_{1l}}{1 - a_l(\lambda)} \frac{\partial(1 - a_l(\lambda))}{\partial \lambda_{1r}} + \frac{u_{2l}}{a_l(\lambda)} \frac{\partial a_l(\lambda)}{\partial \lambda_{1r}} + \frac{u_{5l}}{1 - c_l(\lambda)} \frac{\partial(1 - c_l(\lambda))}{\partial \lambda_{1r}} + \frac{u_{6l}}{c_l(\lambda)} \frac{\partial c_l(\lambda)}{\partial \lambda_{1r}} \\ &= \sum_l \left(\frac{-u_{1l}}{1 - a_l(\lambda)} + \frac{u_{2l}}{a_l(\lambda)} \right) \frac{\partial a_l(\lambda)}{\partial \lambda_{1r}} + \left(\frac{-u_{5l}}{1 - c_l(\lambda)} + \frac{u_{6l}}{c_l(\lambda)} \right) \frac{\partial c_l(\lambda)}{\partial \lambda_{1r}}. \end{aligned}$$

Evaluating this at $\lambda = \mathbf{0}$ we obtain

$$\begin{aligned} \frac{\partial}{\partial \lambda_{1r}} f(\lambda)|_{\lambda=0} &= \sum_l \left(-\frac{u_{1l}}{u_{1l} + u_{2l}} + \frac{u_{2l}}{u_{1l} + u_{2l}} \right) \frac{\partial a_l(\lambda)}{\partial \lambda_{1r}} + \left(-\frac{u_{5l}}{u_{5l} + u_{6l}} + \frac{u_{6l}}{u_{5l} + u_{6l}} \right) \frac{\partial c_l(\lambda)}{\partial \lambda_{1r}} \\ &= \sum_l 0 \cdot \frac{\partial a_l(\lambda)}{\partial \lambda_{1r}} + 0 \cdot \frac{\partial c_l(\lambda)}{\partial \lambda_{1r}} = 0. \end{aligned}$$

The derivation w.r.t. the second component follows analogously. Computing the second derivative $f''(\mathbf{0}) =: D$ we can already see from the first derivative that if we take partial derivatives of $f(\lambda)$ w.r.t. λ_{1r} and then λ_{2s} , $r, s = 1, \dots, k$ we will obtain zero for all values of λ . Thus we will only calculate it for λ_{1r} and λ_{1s} , $r, s = 1, \dots, k$. The $1r \times 1s$ -th element of the $2k \times 2k$ matrix D is given by

$$D_{1r \times 1s} = \frac{\partial^2}{\partial \lambda_{1r} \partial \lambda_{1s}} f(\lambda)|_{\lambda=\mathbf{0}}.$$

Denote $u_{1l} + u_{2l} + z_{1l}n\lambda_1^T \mathbf{G}_1(t_{1l})$ by γ_{1l} and $u_{5l} + u_{6l} + z_{3l}n\lambda_1^T \mathbf{G}_1(t_{1l})$ by γ_{3l} . Now the second derivative is given by

$$\begin{aligned} & \frac{\partial}{\partial \lambda_{1s}} \left[\sum_l \left(\frac{-u_{1l}}{1-a_l(\lambda)} + \frac{u_{2l}}{a_l(\lambda)} \right) (-u_{2l}z_{1l}ng_{1r}\gamma_{1l}^{-2}) + \left(\frac{-u_{5l}}{1-c_l(\lambda)} + \frac{u_{6l}}{c_l(\lambda)} \right) (-u_{6l}z_{3l}ng_{1r}\gamma_{3l}^{-2}) \right] \\ &= \sum_l u_{2l}z_{1l}ng_{1r} \frac{\partial}{\partial \lambda_{1s}} \left[\left(\frac{u_{1l}}{u_{1l} + z_{1l}n\lambda_1^T \mathbf{G}_1(t_{1l})} - \frac{u_{2l}}{u_{2l}} \right) \gamma_{1l}^{-1} \right] \\ &+ \sum_l u_{6l}z_{3l}ng_{1r} \frac{\partial}{\partial \lambda_{1s}} \left[\left(\frac{u_{5l}}{u_{5l} + z_{3l}n\lambda_1^T \mathbf{G}_1(t_{1l})} - \frac{u_{6l}}{u_{6l}} \right) \gamma_{3l}^{-1} \right] \\ &= \sum_l u_{2l}z_{1l}^2 n^2 g_{1r} g_{1s} \gamma_{1l}^{-1} (-u_{1l}^{-1}) + \sum_l u_{6l}z_{3l}^2 n^2 g_{1r} g_{1s} \gamma_{3l}^{-1} (-u_{5l}^{-1}). \end{aligned}$$

For $\lambda = \mathbf{0}$ we then obtain

$$D_{1r \times 1s} = - \sum_l \left(\frac{u_{2l} z_{1l}^2 n^2 g_{1r} g_{1s}}{u_{1l} u_{1l} + u_{2l}} + \frac{u_{6l} z_{3l}^2 n^2 g_{1r} g_{1s}}{u_{5l} u_{5l} + u_{6l}} \right).$$

Using the fact that $z_{1l}/(u_{1l} + u_{2l}) = z_{3l}/(u_{5l} + u_{6l})$ we can rearrange $D_{1r \times 1s}$ to

$$- \sum_l n^2 g_{1r} g_{1s} \frac{z_{1l}}{u_{1l} + u_{2l}} \left(\frac{u_{2l}}{u_{1l}} z_{1l} + \frac{u_{6l}}{u_{5l}} z_{3l} \right).$$

The values of $D_{2r \times 2s}$, $r, s = 1, \dots, k$ can be calculated analogously. Then D is the block diagonal matrix

$$D = \begin{pmatrix} D_{1 \times 1} & \mathbf{0} \\ \mathbf{0} & D_{2 \times 2} \end{pmatrix},$$

where $D_{1 \times 1}$ is the matrix with the entries $D_{1r \times 1s}$, $r, s = 1, \dots, k$ and $D_{2 \times 2}$ analogously. By a now standard counting process martingale assumption we see that the elements of $-D/n$ converge in probability to the elements of D^* .

Define

$$Y_{1l} := \frac{\delta_{1l} n z_{1l}(T_{1l}) \tilde{\lambda}_1^T \mathbf{G}_1(T_{1l})}{u_{1l}(T_{1l}) + u_{2l}(T_{1l})} \quad \text{and} \quad Y_{3l} := \frac{\delta_{1l} n z_{3l}(T_{1l}) \tilde{\lambda}_1^T \mathbf{G}_1(T_{1l})}{u_{5l}(T_{1l}) + u_{6l}(T_{1l})}, \quad l = 1, \dots, n,$$

where $z_1(T_{1l})$ is simply z_{1l}^* with $t_{1l}^* = T_{1l}$, analogously for $z_3(T_{1l})$. As T_{1l} are defined to be i.i.d. we can conclude that Y_{1l} and Y_{3l} are i.i.d. as well. Due to the regularity specification it holds that $E(Y_{1l}^2) < \infty$ and $E(Y_{3l}^2) < \infty$.

Lemma 3. Let Y_{1i} and Y_{3i} be as above. Under the regularity specifications it holds that

$$1/n \sum_{i=1}^n (Y_{1i}^2 + Y_{3i}^2) \tag{A2}$$

converges to

$$\int \frac{(\tilde{\lambda}_1^T \mathbf{G}_1)^2 d\Lambda_{X01}}{(1 - F_{X01}) \prod_j (1 - F_{CXj})}$$

The convergence is of order $O_p(1)$.

Proof. The sum in Equation (A2) can be rearranged to

$$\sum_{t_{1l}} (\tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}))^2 (u_{2l} + u_{6l}) \frac{n}{u_{1l} + u_{2l} + u_{5l} + u_{6l}} \delta_{1l}$$

Define $Z(t) = \sum_{i=1}^n \mathbb{1}\{T_{1i} \geq t\}$. Then the term $u_{1l} + u_{2l} + u_{5l} + u_{6l}$ can be replaced by $\hat{Z}(t_{1l})$. Further $\delta_{1l} = 1$ if $u_{2l} + u_{6l} \neq 0$ and zero otherwise. We obtain

$$\sum_{t_{1l}} \frac{(\tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}))^2}{\hat{Z}(t_{1l})/n} (u_{2l} + u_{6l})$$

Now we can rewrite this in terms of the Nelson–Aalen estimator,

$$\int_0^\infty \frac{(\tilde{\lambda}_1^T \mathbf{G}_1(t))^2}{\hat{Z}(t)/n} d\hat{\Lambda}_{NA,1}$$

where $\hat{\Lambda}_{NA,1}$ denotes the Nelson–Aalen estimator of the first component. Analogue to Lemma 4.1 of Dabrowska (1988) the supremum norm of the component-wise cumulative hazard function converges to zero, which follows from the Glivenko–Cantelli Theorem and simple algebra under the regularity specifications.

Applying Lengart’s inequality and using lemma 3.1.3.9 of Spreij (1990) we can switch the integral w.r.t. the true component-wise cumulative hazard function.

Further, the empirical distribution given through $\hat{Z}(t)/n$ converges uniformly to $(1 - F_{X01}) (1 - F_{CX})$. Thus the convergence of the sum holds true. ■

Lemma 4. Define $Z_n := \max_{1 \leq i \leq n} Y_{1i}$, with Y_{1i} i.i.d. random variables, and $E(Y_{1i}^2) < \infty$. It follows that $Z_n = o_p(n^{1/2})$ a.s. as $n \rightarrow \infty$.

Proof. This follows from lemma A1 of Pan and Zhou (2002). ■

Lemma 5. Assume the data are such that the NPMLs of EL are asymptotically normal and the variance–covariance matrix Σ defined in the proof is invertible. Then for the solution $\lambda_X = (\lambda_1^T, \lambda_2^T)^T$ of the constrained problem (9), corresponding to the null hypothesis H_0 , as given in Equation (4), it holds that $n^{1/2} \lambda_X$ converges in distribution to $N(0, \Sigma)$.

Proof. Define $\boldsymbol{\phi}(\mathbf{s}) = (\phi_{11}(s_1), \dots, \phi_{1k}(s_1), \phi_{21}(s_2), \dots, \phi_{2k}(s_2))$ with $\phi_{jr}(t_{ji}) = \sum_i g_{jr}(t_{ji}) \log(1 - v_{ji}(s_j)) - \mu_{jr}$ for $j = 1, 2, r = 1, \dots, k$ and $\mathbf{s} = (s_1^T, s_2^T)^T \in \mathbb{R}^{2k}$. Then λ_X is the unique solution of $\boldsymbol{\phi}(\mathbf{s}) = 0$ under the regularity specification. Thus we can expand the term $\boldsymbol{\phi}(\lambda)$ to $\boldsymbol{\phi}(\mathbf{0}) + \boldsymbol{\phi}'(\mathbf{0})\lambda + o_p(n^{-1/2})$, where $\boldsymbol{\phi}'(\mathbf{0})$ is a $2k \times 2k$ matrix. The next part proves that this indeed holds true. Rewrite $\lambda_j = p_j \tilde{\lambda}_j$, where $\|\tilde{\lambda}_j\| = 1, j = 1, 2$. Then,

$$0 = \tilde{\lambda}^T \boldsymbol{\phi}(\lambda) = \sum_l \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}) \log(1 - v_{1l}(\lambda_1)) - \tilde{\lambda}_1^T \boldsymbol{\mu}_1 \\ + \sum_l \tilde{\lambda}_2^T \mathbf{G}_2(t_{2l}) \log(1 - v_{2l}(\lambda_2)) - \tilde{\lambda}_2^T \boldsymbol{\mu}_2.$$

We can split these terms up in a part without the Lagrange multiplier and a part with,

$$\sum_l \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}) \log(1 - v_{1l}(\mathbf{0})) - \tilde{\lambda}_1^T \boldsymbol{\mu}_1 \quad (i) \\ + \sum_l (\tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}) [\log(1 - v_{1l}(\lambda_1)) - \log(1 - v_{1l}(\mathbf{0}))]) \quad (ii) \\ + \sum_l \tilde{\lambda}_2^T \mathbf{G}_2(t_{2l}) \log(1 - v_{2l}(\mathbf{0})) - \tilde{\lambda}_2^T \boldsymbol{\mu}_2 \quad (iii) \\ + \sum_l (\tilde{\lambda}_2^T \mathbf{G}_2(t_{2l}) [\log(1 - v_{2l}(\lambda_2)) - \log(1 - v_{2l}(\mathbf{0}))]) \quad (iv).$$

Again we will only consider the terms (i) and (ii) in the following as the other terms follow analogously.

It holds that term (i) is of order $o_p(n^{-1/2})$ under the null hypothesis. This follows directly from the properties of the NPMLEs.

The term (ii) can be split into the two terms

$$B_1 = \sum_l \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}) z_{1l} [\log(1 - a_l(\lambda_1)) - \log(1 - a_l(\mathbf{0}))],$$

and

$$B_2 = \sum_l \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}) z_{3l} [\log(1 - c_l(\lambda_1)) - \log(1 - c_l(\mathbf{0}))],$$

with an error of order $o_p(\max v_{1l})$. We will rearrange and estimate B_1 . It holds, when applying $|\varepsilon_1 - \varepsilon_2| \leq |\log(\varepsilon_1) - \log(\varepsilon_2)|$ for $\varepsilon_1, \varepsilon_2 \in (0, 1]$, that

$$|B_1| \geq \left| \sum_l \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}) z_{1l} [1 - a_l(\lambda_1) - 1 + a_l(\mathbf{0})] \right| \\ = \left| \sum_l z_{1l} \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}) \frac{u_{2l} z_{1l} n \lambda_1^T \mathbf{G}_1(t_{1l})}{(u_{1l} + u_{2l})(u_{1l} + u_{2l} + z_{1l} n \lambda_1^T \mathbf{G}_1(t_{1l}))} \right| \\ = \left| \sum_l \frac{p_1 n u_{2l} (z_{1l} \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}))^2}{(u_{1l} + u_{2l})^2 + (u_{1l} + u_{2l})^2 \frac{z_{1l} p_1 n \lambda_1^T \mathbf{G}_1(t_{1l})}{u_{1l} + u_{2l}}} \right|.$$

Analogously for B_1 and B_2 we obtain in total

$$\begin{aligned}
 |B_1 + B_2| &\geq \left| \sum_l \frac{p_1 n u_{2l} (z_{1l} \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}))^2}{(u_{1l} + u_{2l})^2 + (u_{1l} + u_{2l})^2 \frac{z_{1l} p_1 n \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l})}{u_{1l} + u_{2l}}} \right. \\
 &\quad \left. + \sum_l \frac{p_1 n u_{6l} (z_{3l} \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}))^2}{(u_{5l} + u_{6l})^2 + (u_{5l} + u_{6l})^2 \frac{z_{3l} p_1 n \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l})}{u_{5l} + u_{6l}}} \right| \\
 &= \left| \sum_l \left[\frac{z_{1l}^2}{(u_{1l} + u_{2l})^2} \frac{p_1 n u_{2l} (\tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}))^2}{1 + \frac{z_{1l}}{u_{1l} + u_{2l}} p_1 n \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l})} \right. \right. \\
 &\quad \left. \left. + \frac{z_{3l}^2}{(u_{5l} + u_{6l})^2} \frac{p_1 n u_{6l} (\tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}))^2}{1 + \frac{z_{3l}}{u_{5l} + u_{6l}} p_1 n \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l})} \right] \right|.
 \end{aligned}$$

As it holds that $z_{1l}/(u_{1l} + u_{2l}) = z_{3l}/(u_{5l} + u_{6l})$, we can merge the two sums and obtain

$$\begin{aligned}
 &\left| \sum_l \frac{p_1}{1 + p_1 n \frac{z_{1l}}{u_{1l} + u_{2l}} \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l})} \frac{z_{1l}^2 n (u_{2l} + u_{6l}) (\tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}))^2}{(u_{1l} + u_{2l})^2} \right| \\
 &\geq \frac{|p_1|}{1 + |p_1| n \max_l \frac{z_{1l} \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l})}{u_{1l} + u_{2l}}} \sum_l \frac{(z_{1l} \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}))^2 n (u_{2l} + u_{6l})}{(u_{1l} + u_{2l})^2}.
 \end{aligned}$$

From Lemma 3 it follows that the sum in the last expression is of order $O_p(1)$, under the regularity specification the requirements for Lemma 4 are fulfilled and the maximum in the denominator is of order $o_p(n^{1/2})$. Hence $|p_1|$ is of order $O_p(n^{-1/2})$. The proof for the remaining terms follows analogously and the expansion above is valid. ■

Therefore,

$$n^{1/2} \lambda = \boldsymbol{\phi}'(\mathbf{0})^{-1} (-n^{1/2} \boldsymbol{\phi}(\mathbf{0})) + o_p(1).$$

The elements ϕ'_{rs} of $\boldsymbol{\phi}'(\mathbf{0})$ can easily be computed, and are equivalent to $-D_{rl}/n$ where $r, l = 1, \dots, 2k$. By the counting process martingale central limit theorem, we can show that $n^{1/2} \lambda$ converges in distribution to $N(\mathbf{0}, \Sigma_\phi)$ with $\Sigma_\phi = \lim \boldsymbol{\phi}'(\mathbf{0})$. All together we obtain that $n^{1/2} \lambda$ converges in distribution to $N(\mathbf{0}, \Sigma)$ with $\Sigma := \lim (\boldsymbol{\phi}'(\mathbf{0}))^{-1}$. Note that $\Sigma^{-1} = D^*$.

A3. Proof of Theorem 3

Let $f(\lambda)$ be defined as in Equation (A1), then $W = -2(f(\lambda_x) - f(\mathbf{0}))$, where λ_x denotes the Lagrange multiplier of the optimum. By Taylor expansion, which is valid as λ_x is close to zero, we obtain

$$W = 2 \left\{ f(\mathbf{0}) - f(\mathbf{0}) - f'(\mathbf{0}) \lambda_x - \frac{1}{2} \lambda_x^T D \lambda_x + o_p(1) \right\},$$

where D is again the matrix of the second derivative of $f(\lambda)$ w.r.t. λ . From the preparation part we have $f'(\mathbf{0}) = 0$ and we can simplify W to $-\lambda_x^T \Sigma \lambda_x + o_p(1)$. Note that $-\Sigma$ is symmetric and positive definite for large enough n . This follows from the convergence of $-D/n$ to a positive definite matrix $D^* = \Sigma^{-1}$, see proof of Lemma 5. Now $W = \lambda_x^T (-D)^{1/2} (-D)^{1/2} \lambda_x + o_p(1)$ and as $n^{1/2} \lambda_x^T (D^{1/2} n^{-1/2})$ converges in distribution to $N(\mathbf{0}, I)$, we obtain that W converges in distribution to \mathcal{X}_{2k}^2 .

APPENDIX B. PROOF OF THE TWO-SAMPLE CASE

B1. Regularity specification for the two-sample case

Let X_1, \dots, X_n be i.i.d. d -dimensional random variables as already defined in Section A.1. Let Y_1, \dots, Y_n be i.i.d. d -dimensional random variables with cumulative distribution function $F_{Y_0}(t)$ and marginals $F_{Y_{01}}(t), \dots, F_{Y_{0d}}(t)$. The cumulative hazard function is denoted by $\Lambda_{Y_0}(t)$ and the marginals by $\Lambda_{Y_{01}}(t), \dots, \Lambda_{Y_{0d}}(t)$. We observe the d -dimensional vectors $T_i = \min(X_i, C_i^X)$, $\delta_i = \mathbb{1}\{X_i \leq C_i^X\}$, $S_i = \min(Y_i, C_i^Y)$ and $\tau_i = \mathbb{1}\{Y_i \leq C_i^Y\}$ where C_i^X and C_i^Y are i.i.d. censoring times, independent of X_i and Y_i , and the minimum and the indicator function are considered component-wise. The cumulative distribution function of the C_i^X is given by $F_{CX}(t)$ and for C_i^Y it is $F_{CY}(t)$. The distribution functions $F_{X_0}(t)$ and $F_{CX}(t)$ do not have common discontinuities, respectively, the same holds for the distribution functions $F_{Y_0}(t)$ and $F_{CY}(t)$.

Let $g_{j1}(t), \dots, g_{jk}(t), h_{j1}(t), \dots, h_{jk}(t), j = 1, \dots, d$, be nonnegative left continuous functions such that none can be expressed by a linear combination of the rest of the functions. Further it shall hold for all functions

$$0 < \int \frac{|g_{jr}(t)|^w (1 - \Delta\Lambda_{X_{0j}}(t))}{(1 - F_{X_{0j}}(t)) \prod_j (1 - F_{CX_j}(t))} d\Lambda_{X_{0j}}(t) < \infty,$$

and

$$0 < \int \frac{|h_{jr}(t)|^w (1 - \Delta\Lambda_{Y_{0j}}(t))}{(1 - F_{Y_{0j}}(t)) \prod_j (1 - F_{CY_j}(t))} d\Lambda_{Y_{0j}}(t) < \infty,$$

with $w = 1, 2, j = 1, \dots, d, r = 1, \dots, k$. Note that the functions $g_{jr}(t)$ and $h_{jr}(t)$ may be random, yet they need to be predictable w.r.t. a filtration. For $d = 2$ the filtration is given by

$$\begin{aligned} \mathcal{F}_t = \sigma \{ & T_{1l} \mathbb{1}\{T_{1l} \leq t\}; \delta_{1l} \mathbb{1}\{T_{1l} \leq t\}; T_{2l} \mathbb{1}\{T_{2l} \leq t\}; \delta_{2l} \mathbb{1}\{T_{2l} \leq t\}; \\ & S_{1l} \mathbb{1}\{S_{1l} \leq t\}; \tau_{1l} \mathbb{1}\{S_{1l} \leq t\}; S_{2l} \mathbb{1}\{S_{2l} \leq t\}; \tau_{2l} \mathbb{1}\{S_{2l} \leq t\}; l = 1, \dots, n \}. \end{aligned} \quad (\text{B1})$$

Through this we are capable of applying the martingale central limit theorem, as $\hat{\Lambda}_{NA_j}^{(X)}(t) - \Lambda_{X_{0j}}(t)$ is then a martingale, where $\hat{\Lambda}_{NA_j}^{(X)}$ denotes the Nelson–Aalen estimator of component j for samples of X , respectively, for $\hat{\Lambda}_{NA_j}^{(Y)}(t) - \Lambda_{Y_{0j}}(t)$.

Furthermore, for all random functions $g_{jr}(t)$ and $h_{jr}(t)$ there must exist a nonrandom left continuous function $g_{jr0}(t)$ and $h_{jr0}(t)$, respectively, such that

$$\sup_{t \leq \max T_{ji}} |g_{jr}(t)/g_{jr0}(t)| = o_p(1) \text{ and } \sup_{t \leq \max S_{ji}} |h_{jr}(t)/h_{jr0}(t)| = o_p(1)$$

for $j = 1, \dots, d$ and $r = 1, \dots, k$ as $n \rightarrow \infty$.

B2. Mathematical derivations and proof of Theorem 5

The constraints are defined as in Equation (12). Define $\mathbf{G}_j(t_i) = (g_{j1}(t_i), \dots, g_{jk}(t_i))^T$ and $\mathbf{H}_j(s_i) = (h_{j1}(s_i), \dots, h_{jk}(s_i))^T$, $j = 1, 2$, the vectors of constraint-function values. By application of the Lagrange multiplier method we know that the maximum-likelihood estimators are given by the NPMLE as defined in Equation (13).

The function $f(\lambda)$, similar to Equation (A1), is given by

$$\begin{aligned} & \sum_l \log(1 - a_l(\lambda_1))u_{1l} + \log(a_l(\lambda_1))u_{2l} + \log(1 - b_l(\lambda_2))u_{3l} + \log(b_l(\lambda_2))u_{4l} \\ & + \log(1 - c_l(\lambda_1))u_{5l} + \log(c_l(\lambda_1))u_{6l} + \log(1 - d_l(\lambda_2))u_{7l} + \log(d_l(\lambda_2))u_{8l} \\ & + \sum_l \log(1 - \alpha_l(\lambda_1))v_{1l} + \log(\alpha_l(\lambda_1))v_{2l} + \log(1 - \beta_l(\lambda_2))v_{3l} + \log(\beta_l(\lambda_2))v_{4l} \\ & + \log(1 - \gamma_l(\lambda_1))v_{5l} + \log(\gamma_l(\lambda_1))v_{6l} + \log(1 - \varepsilon_l(\lambda_2))v_{7l} + \log(\varepsilon_l(\lambda_2))v_{8l}. \end{aligned}$$

Taking the derivatives of this function we obtain that $f'(\mathbf{0}) = \mathbf{0}$ and $f''(\mathbf{0}) = D$. The elements of D are zero where one first takes the derivative w.r.t λ_{1r} and then w.r.t. λ_{2s} or vice versa, $r, s = 1, \dots, k$, that is, D is again a block diagonal matrix. The element $D_{1r \times 1s}$, obtained by taking the derivatives of $f(\lambda)$ w.r.t. λ_{1r} and λ_{1s} , $r, s = 1, \dots, k$, is given by

$$- \left[\sum_l \left(\frac{u_{2l} z_{1l}^2 n^2 g_{1r} g_{1s}}{u_{1l} u_{1l} + u_{2l}} + \frac{u_{6l} z_{3l}^2 n^2 g_{1r} g_{1s}}{u_{5l} u_{5l} + u_{6l}} \right) + \sum_l \left(\frac{v_{2l} \zeta_{1l}^2 n^2 h_{1r} h_{1s}}{v_{1l} v_{1l} + v_{2l}} + \frac{v_{6l} \zeta_{3l}^2 n^2 h_{1r} h_{1s}}{v_{5l} v_{5l} + v_{6l}} \right) \right].$$

As we assumed that $n/m \rightarrow c \in (0, \infty)$ as $\min(m, n) \rightarrow \infty$ we obtain that the elements of $-D/n$ converge in probability to the elements of D^{**} . Again we need to show the asymptotic normality of $n^{1/2} \lambda_{xy}$, with λ_{xy} the solution of the constrained problem (12). Define $\phi(\mathbf{s}) = (\phi_{11}(s_1), \dots, \phi_{1k}(s_1), \phi_{21}(s_2), \dots, \phi_{2k}(s_2))$ with

$$\phi_{jr}(s_j) = \sum_l g_{jr}(t_{jl}) \log(1 - v_{jl}(s_j)) - \sum_l h_{jr}(s_{jl}) \log(1 - w_{jl}(s_j)) - \mu_{jr}$$

for $j = 1, 2, r = 1, \dots, k$ and $\mathbf{s} = (s_1, s_2)^T \in \mathbb{R}^2$ where the two sums are taken over all time points t_{jl} , respectively, s_{jl} , excluding the last.

Set $\lambda_j = p_j \tilde{\lambda}_j$ with $\|\tilde{\lambda}_j\| = 1, j = 1, 2$. Now

$$\begin{aligned} 0 &= \tilde{\lambda}^T \phi(\lambda) = \sum_l \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}) \log(1 - v_{1l}(\lambda_1)) - \sum_l \tilde{\lambda}_1^T \mathbf{H}_1(s_{1l}) \log(1 - w_{1l}(\lambda_1)) - \tilde{\lambda}_1^T \boldsymbol{\mu}_1 \\ &+ \sum_l \tilde{\lambda}_2^T \mathbf{G}_2(t_{2l}) \log(1 - v_{2l}(\lambda_2)) - \sum_l \tilde{\lambda}_2^T \mathbf{H}_2(s_{2l}) \log(1 - w_{2l}(\lambda_1)) - \tilde{\lambda}_2^T \boldsymbol{\mu}_2, \end{aligned}$$

with $\boldsymbol{\mu}_1 = (\mu_{11}, \dots, \mu_{1k})^T$ and $\boldsymbol{\mu}_2 = (\mu_{21}, \dots, \mu_{2k})^T$. Splitting this into the four terms A_1, A_2, B_1 , and B_2 , which are given by

$$\begin{aligned} A_1 &= \sum_l \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}) \log(1 - v_{1l}(\mathbf{0})) - \sum_l \tilde{\lambda}_1^T \mathbf{H}_1(s_{1l}) \log(1 - w_{1l}(\mathbf{0})) - \tilde{\lambda}_1^T \boldsymbol{\mu}_1 = O_p(n^{-1/2}), \\ A_2 &= \sum_l \tilde{\lambda}_2^T \mathbf{G}_2(t_{2l}) \log(1 - v_{2l}(\mathbf{0})) - \sum_l \tilde{\lambda}_2^T \mathbf{H}_2(s_{2l}) \log(1 - w_{2l}(\mathbf{0})) - \tilde{\lambda}_2^T \boldsymbol{\mu}_2 = O_p(n^{-1/2}), \end{aligned}$$

where the last equation in A_1 and A_2 holds under the null hypothesis,

$$\begin{aligned} B_1 &= \sum_l \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}) [\log(1 - v_{1l}(\lambda_1)) - \log(1 - v_{1l}(\mathbf{0}))] \\ &- \sum_l \tilde{\lambda}_1^T \mathbf{H}_1(s_{1l}) [\log(1 - w_{1l}(\lambda_1)) - \log(1 - w_{1l}(\mathbf{0}))] \end{aligned}$$

and

$$B_2 = \sum_l \tilde{\lambda}_2^T \mathbf{G}_2(t_{2l}) [\log(1 - v_{2l}(\lambda_2)) - \log(1 - v_{2l}(\mathbf{0}))] \\ - \sum_l \tilde{\lambda}_2^T \mathbf{H}_2(s_{2l}) [\log(1 - w_{2l}(\lambda_1)) - \log(1 - w_{2l}(\mathbf{0}))].$$

For B_1 we obtain, with similar calculations as in the one-sample case,

$$|B_1| \geq \frac{|p_1|}{1 + |p_1| \min(m, n) \max \left\{ \max_l \frac{z_{1l} \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l})}{u_{1l} + u_{2l}}; \max_l \frac{\zeta_{1l} \tilde{\lambda}_1^T \mathbf{H}_1(s_{1l})}{v_{1l} + v_{2l}} \right\}} \\ \times \left(\sum_l \frac{(z_{1l} \tilde{\lambda}_1^T \mathbf{G}_1(t_{1l}))^2 \min(m, n) (u_{2l} + u_{6l})}{(u_{1l} + u_{2l})^2} + \sum_l \frac{(\zeta_{1l} \tilde{\lambda}_1^T \mathbf{H}_1(s_{1l}))^2 \min(m, n) (v_{2l} + v_{6l})}{(v_{1l} + v_{2l})^2} \right).$$

Again the sum in the last expression is of order $O_p(1)$, and $|p_1|$ is of order $O_p(n^{-1/2})$. Thus

$$0 = \phi(\lambda) = \phi(\mathbf{0}) + \phi'(\mathbf{0})\lambda + o_p(n^{-1/2})$$

and by the counting process martingale central limit theorem we obtain that $n^{1/2}\lambda$ converges in distribution to $N(\mathbf{0}, \Sigma)$ with $\Sigma = \lim \{ \phi'(\mathbf{0}) \}^{-1}$. Note that $\phi'(\mathbf{0}) = -D/n$.

By Taylor expansion of $W^* = -2(f(\lambda_{xy}) - f(\mathbf{0}))$, as λ_{xy} is close to zero, it holds

$$W^* = -2(f'(\mathbf{0})\lambda_{xy} + \frac{1}{2}\lambda_{xy}^T D \lambda_{xy}) + o_p(1) = \lambda_{xy}^T (-D)^{1/2} (-D)^{1/2} \lambda_{xy} + o_p(1).$$

Now as $\lambda_{xy}^T (-D)^{1/2}$ converges in distribution to $N(0, 1)$ we obtain the result of Theorem 5.

B3. Notes on Proof of Theorem 4 and Theorem 6

Due to notation we will not explicitly give the proof of those two theorems. For a fixed dimension the proof follows analogue to the two-dimensional case. First defines the function f and calculates the derivatives. Second defines the function ϕ and calculates, by skillfully splitting it up in several sums, the rates of convergence. Last, one applies a Taylor expansion to obtain the results.