Inference for meaningful estimands in factorial survival designs and competing risks settings

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von M.Sc. Merle Munko
(akademischer Grad, Vorname, Name)

geb. am 26.02.2000 in Duisburg

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. Markus Pauly
(akademischer Grad, Vorname, Name)
Prof. Dr. Dennis Dobler
(akademischer Grad, Vorname, Name)
Prof. Dr. Eric Beutner
(akademischer Grad, Vorname, Name)

eingereicht am: 08.05.2025

Verteidigung am: 01.09.2025

Abstract

Estimands are highly used in survival analysis, e.g., to compare the effects of different treatments in a clinical study. Beyond the popular hazard ratio, which relies on the rather restrictive proportional hazards assumption, there are various estimands that do not rely on this assumption as, for example, the Mann-Whitney effect and the restricted mean survival time. Several inference procedures for estimands in simple one- and two-sample survival problems have already been developed in the literature. However, since the underlying designs of real data are often more complex, there is a lack of adequate testing procedures in more complex survival models. Moreover, ties easily occur in real data if time is measured in whole days, months, or years. While many existing methods require continuous survival distributions, we prove all methods without any continuity assumption on the survival and censoring times by empirical process theory. Thus, we explicitly allow for ties in the data. Furthermore, in many practical applications, not only one hypothesis is of interest, e.g., if not only the existence of an effect of any treatment is of interest but also which treatment groups have a different effect. In this case, multiple tests need to be performed to infer several hypotheses simultaneously.

To close all above-mentioned gaps, we construct tests for a version of the Mann-Whitney effect and restricted mean survival times in the paired survival setup, multiple tests for restricted mean survival times in general factorial designs, and multiple tests for restricted mean time losts of competing risks in general factorial designs. For the multiple tests, we incorporate the multivariate limit distribution of the test statistics to gain more power in contrast to a simple Bonferroni-correction. Additionally, we apply different resampling procedures, as permutation and bootstrap approaches, for all tests to improve the small sample performance of the tests. Moreover, for proving the validity of the resampling tests, we design a flexible conditional delta-method for resampling empirical processes.

Zusammenfassung

Estimands werden in der Überlebenszeitanalyse häufig verwendet, z.B. um die Auswirkungen verschiedener Behandlungen in einer klinischen Studie zu vergleichen. Neben dem weit verbreiteten Hazard Ratio, das auf der eher restriktiven Proportional-Hazard-Annahme beruht, gibt es verschiedene Estimands, die nicht auf dieser Annahme beruhen, wie z.B. der Mann-Whitney-Effekt und die 'restricted mean survival time'. In der Literatur wurden bereits mehrere Inferenzverfahren für Estimands bei einfachen Ein- oder Zwei-Stichproben-Problemen entwickelt. Da die den realen Daten zugrunde liegenden Designs jedoch häufig komplexer sind, fehlt es an geeigneten Testverfahren für komplexere Überlebenszeitmodelle. Außerdem treten bei realen Daten Bindungen auf, wenn die Zeit in ganzen Tagen, Monaten oder Jahren gemessen wird. Während viele bereits existierende Methoden stetige Überlebenszeitverteilungen voraussetzen, beweisen wir alle Methoden ohne Stetigkeitsannahme für die Überlebens- und Zensierungszeiten durch empirische Prozesstheorie. Somit erlauben wir Bindungen in den Daten explizit. Außerdem ist in vielen praktischen Anwendungen nicht nur eine Hypothese von Interesse, z.B. wenn nicht nur die Existenz eines Effekts einer Behandlung von Interesse ist, sondern auch, welche Behandlungsgruppen einen unterschiedlichen Effekt haben. In diesem Fall müssen multiple Tests durchgeführt werden, um mehrere Hypothesen simultan testen zu können.

Um alle oben erwähnten Lücken zu schließen, konstruieren wir Tests für eine Version des Mann-Whitney-Effekts und 'restricted mean survival times' für gepaarte Überlebenszeiten, multiple Tests für 'restricted mean survival times' in allgemeinen faktoriellen Designs und multiple Tests für 'restricted mean time losts' von konkurrierenden Risiken in allgemeinen faktoriellen Designs. Bei den multiplen Tests beziehen wir die multivariate Grenzverteilung der Teststatistiken ein, um im Gegensatz zu einer einfachen Bonferroni-Korrektur eine höhere Güte zu erzielen. Darüber hinaus wenden wir für alle Tests verschiedene Resampling-Verfahren an, wie Permutation und Bootstrap-Ansätze, um die Performance der Tests bei kleinen Stichprobenumfängen zu verbessern. Um die Gültigkeit der Resampling-Tests zu beweisen, entwickeln wir außerdem eine flexible bedingte Delta-Methode für Resampling bei empirischen Prozessen.

Contents

Preface: Structure, Personal Contributions and Authorship

This thesis includes six sections. The first section is an introduction in which the problems covered in this thesis are motivated. In the second section, methodological preliminaries are presented. Besides the general notation of this thesis, this includes a new conditional delta-method for resampling empirical processes in multiple sample problems in Section 2.2 and general methodology for simultaneous inference in Section 2.3. Then, three sections containing the main contributions of this thesis follow, that are

- Section 3, where two approaches are proposed to infer consecutive survival times properly and randomization tests therefore are constructed,
- Section 4, where simultaneous inference procedures for general factorial survival designs are constructed based on the restricted mean survival time,
- Section 5, where simultaneous inference procedures for competing risks data in general factorial designs are developed based on the restricted mean time lost.

In Section 6, the results of this thesis are discussed. Moreover, an extensive appendix is attached, which includes applications of the conditional delta-method in Section A, a correction of the limit distribution for Aalen-Johansen estimators that was stated in [26] in Section B, and additional simulation results of the simulations in Sections 4 and 5 in Sections D and C, respectively. Furthermore, in Section E, the two R Packages GFDrmst [20] and GFDrmtl [21] that were developed for applications of the methods in Sections 4 and 5, respectively, are presented.
The R code that was used for the simulations and data examples as well as more detailed tables for the simulation results are provided on GitHub (`https://github.com/MerleMunko/supplement_thesis`). The R Packages GFDrmst [20] and GFDrmtl [21] have been published on CRAN.
Sections 1, 2.1, 2.3, 3, and 6 have not been published yet. Sections 4 and D have been published in [58]. Moreover, Section B has been published in [28]. Section E contains some parts of the documentations in [20, 21]. For the remaining sections, arXiv preprints exist, that are [59] for Sections 2.2 and A and [60] for Sections 5 and C. The contents of the already published and preprinted sections coincides highly with the corresponding published and preprinted papers. Only some editorial changes were made as well as a correction on simultaneous non-inferiority and equivalence tests in Section 4.3.
The contents of the present thesis has been produced under the supervision of Marc Ditzhaus and Dennis Dobler. In detail, Marc Ditzhaus and Dennis Dobler suggested the idea for the topic of Section 3 while Marc Ditzhaus and Markus Pauly suggested the ideas for the topics of Sections 4 and 5. The need of the methodologies of Sections 2.5 and B occurred during the work on the previous named sections and the original ideas of these sections were jointly suggested by Dennis Dobler and Merle Munko. Dennis Dobler helped throughout the work on all topics and Marc Ditzhaus helped specifically during the work on Sections 4 and 5.
Sections 1, 2.1, 2.3, 6, D, C, and E have solely been written by Merle Munko. Sections 2.2, 3, 4, 5, A, and B were written by Merle Munko with only occasional editorial suggestions by Dennis Dobler for Sections 2.2, 3, A, and B and Marc Ditzhaus and Dennis Dobler for Sections 4 and 5. Furthermore, Section 4.5 was slightly revised by Jon Genuneit as well, who provided the data set for the data example.
The R code for the simulations and data analyses has been written by Merle Munko, where some of the used functions are based on functions of Marc Ditzhaus. Both R packages also contain some underlying basis functions of Marc Ditzhaus. All documentation files were solely written by Merle Munko. Marc Kindop and Merle Munko worked jointly on the functions exported by the R package GFDrmst [20]. In detail, Merle Munko mainly implemented the functions for the tests and Marc Kindop mainly implemented the shiny app and the plot and summary function advised by Merle Munko. Dennis Dobler and Marc Kindop tested the final version of the R package GFDrmst and suggested helpful improvements. For the R package GFDrmtl [21], Merle Munko implemented the functions for the tests. Jannes Walter helped to implement the shiny app advised by Merle Munko and tested the final version of the R package GFDrmtl.

Funding

Merle Munko gratefully acknowledges support by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) - GRK 2297 MathCoRe and DI 2906/1–2.

1 Introduction

1.1 Motivation

In survival analysis, the time of a specific event is the object of interest that should be analyzed. Classical applications can be found in medicine, where the time until death or progress of a disease is measured in clinical studies. However, survival analysis can also be applied in several other fields, where the time until an event of interest is measured as, e.g., for the analysis of material fatigue. Also if the times until the event of interest do not point to the duration of survival, the times are usually called survival times. When collecting survival data, it can happen that the event of interest can not be observed for all individuals, for example due to drop-outs of patients from a clinical study. As it still is an information that the event did not happen up to a so-called (right-)censoring time, censored data points should not be ignored but can be used in survival analysis to obtain asymptotically unbiased results.

However, real data is usually more complex and, thus, there is a need of more **complex survival models** to handle this data. Let us consider, for example, the GABRIELA study [37, 38], where the occurrence of asthma, hay fever and neurodermatitis was measured for 2234 children. First of all, we note that more than one event time is recorded per individual as more than one disease is considered. If two survival times are considered for one individual, we observe so-called **paired survival times**. Moreover, despite the time of the occurrence of the diseases, also other factors as the sex and whether the children grew up on a farm were recorded to analyze possible influence factors on the occurrence of the diseases. This is an example for a **factorial design**, where factors are observed for the individuals. Possible questions of interest could be whether the factors, that are sex and/or growing up on a farm in the example, have a significant effect on the survival time and whether there are interaction effects between the factors. A second example is about the survival times of 8966 leukemia patients with a bone and marrow transplantation [36]. Among others, the factors whether the gender of donor and recipient match and whether a T-cell depletion took place were observed. Furthermore, the cause of death, which contains relapse, graft-versus-host disease, and other causes, was recorded for all non-censored patients. If multiple event types as, e.g., different causes of death, are considered, we obtain so-called **competing risks** data.

1.2 Goals of the Thesis

Various methods from survival analysis for comparing different groups rely on the proportional hazards assumption, e.g., the famous Cox proportional hazards model [17] and the logrank test [57]. However, verifying this assumption can be challenging, and its fulfillment is not always guaranteed. Hence, alternatives that do not require the proportional hazards assumption are of great interest. Furthermore, easy-to-interpret effect **estimands**, which summarize treatment and interaction effects in factorial designs, are desired. While the often-used average hazard ratio [13, 46] relies on the proportional hazards assumption, there are also alternatives as the concordance and Mann-Whitney effect [31, 32, 48], the median survival time [12, 15, 22], the restricted mean survival time [68, 43, 24] and the restricted mean time lost [2, 55, 77, 78, 79] that do not rely on this assumption. In this thesis, we focus on a version of the Mann-Whitney effect for paired survival data in Section 3.1, the restricted mean survival time for paired survival data in Section 3.2, the restricted mean survival time in factorial survival designs in Section 4 and the restricted mean time lost for factorial competing risks data in Section 5.

The main goal of this thesis is the construction of hypothesis tests for estimands in complex survival models to provide adequate statistical tools for, e.g., comparing estimands across several groups.

It should be noted that many real data examples contain ties in the data as times are measured in whole days, months or years. Exemplarily, both above mentioned examples, i.e., the GABRIELA study as well as the example about the leukemia patients, contain tied data. Hence, we aim to develop methodology that explicitly allows for ties in the data. This implies that **no continuity assumptions** on the cumulative distribution functions of the survival and censoring times should be required for the validity of the methods. Technically, this can be realized by using empirical process theory [74] for the proofs.

Furthermore, it was shown in several works that the performance of an asymptotically valid test can be improved dramatically for small samples if resampling methods as, e.g., permutation and bootstrap procedures, are applied, see for example [24, 31, 32, 43, 63]. Thus, we aim to investigate and develop **resampling tests** to improve the small sample performance.

Additionally, often more than one hypothesis is of interest. For example, thinking of a factorial design with two factors A and B, hypotheses of interest could be whether (a) factor A has no effect, (b) factor B has no effect and (c) whether there is no interaction effect between A and B. If those three hypotheses are tested with a global null hypothesis, a rejection do not provide the information which of the three hypotheses (a)–(c) is rejected. To infer multiple hypotheses simultaneously, powerful **multiple testing** procedures are desired. General methodology for multiple testing procedures is developed in Section 2.3 and applied in Sections 4 and 5.

2 Methodological Preliminaries

In this section, we state the methodological preliminaries needed in the following. Firstly, the notation used in this thesis is introduced in Section 2.1. As one of the aims is to construct resampling-based tests, we develop a conditional delta-method in Section 2.2 that has a wide range of applications and will be needed to show consistency of the resampling test statistics in the following sections. Moreover, methodology for simultaneous inference in a general setup, which will be used for the construction of multiple tests in Sections 4 and 5, is developed in Section 2.3.

2.1 Notation

While most of the notation used in this thesis can be found in the list of symbols, we want to clarify the remaining notation in this section.
Throughout this thesis, we use the convention $0/0 := 0$. Furthermore, integrals of the form $\int_A f(t)\ \mathrm{d}t$ are understood as Lebesgue integrals. Additionally, integrals of the form $\int_A f\ \mathrm{d}F = \int_A f(t)\ \mathrm{d}F(t)$ are interpreted as Lebesgue-Stieltjes integral [72] for a Lebesgue-measurable set A whenever F is of bounded variation and right-continuous and f is Borel-measurable and bounded; or F is monotone and right-continuous and f is Borel-measurable and non-negative. However, if F is a càdlàg function of unbounded variation but f is of bounded variation and right-continuous, the integral $\int_{[a,b]} f\ \mathrm{d}F$ is defined via integration by parts as

$$f(b)F(b) - f_-(a)F_-(a) - \int_{[a,b]} F_-\ \mathrm{d}f,$$

where here and throughout F_- denotes the left-continuous version of a càdlàg function. For stochastic processes f, F, the integral is defined pathwisely. Throughout, we use the notation $\int_a^b f(t)\ \mathrm{d}t$ to denote integration over the interval $A = [a, b]$. Whenever it makes a difference, we explicitly indicate which endpoints are included by writing $\int_{[a,b]}, \int_{[a,b)}, \int_{(a,b]}$, or $\int_{(a,b)}$.
Moreover, we give a precise definition of conditional weak convergence in outer probability for potentially non-measurable maps. The intuition in the following definition is that $\mathbf{M}_n$ will stand for additional randomness, e.g., induced by random permutation or bootstrapping, whereas $\mathbf{X}_n$ will represent the original data.

Definition 2.1 (Conditional Weak Convergence in Outer Probability)**.** *Let $\mathbf{X}_n : \Omega_1 \to \chi_{1n}, \mathbf{M}_n : \Omega_2 \to \chi_{2n}$ be sequences of maps, where $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, Q_1 \otimes Q_2)$ denotes a product probability space and χ_{1n}, χ_{2n} denote arbitrary sets for $n \in \mathbb{N}$. Furthermore, assume that $\mathbf{y}_n : \chi_{1n} \times \chi_{2n} \to \mathbb{E}$ is a function taking values in a metric space $\mathbb{E}$ for all $n \in \mathbb{N}$ and $\mathbf{Y} : \Omega_1 \times \Omega_2 \to \mathbb{E}$ is a Borel measurable random variable. We say that $\mathbf{y}_n(\mathbf{X}_n, \mathbf{M}_n)$* converges weakly conditionally on $\mathbf{X}_n$ in outer probability to $\mathbf{Y}$, *write $\mathbf{y}_n(\mathbf{X}_n, \mathbf{M}_n) \rightsquigarrow \mathbf{Y}$ conditionally on $\mathbf{X}_n$ in outer probability as $n \to \infty$ or $\mathbf{y}_n(\mathbf{X}_n, \mathbf{M}_n) \xrightarrow{d*} \mathbf{Y}$ (conditionally on $\mathbf{X}_n$) as $n \to \infty$, if*

$$\sup_{h \in BL_1(\mathbb{E})} \left| \mathrm{E}_2 \left[h\left(\mathbf{y}_n(\mathbf{X}_n, \mathbf{M}_n)\right)^{2*} \right] - \mathrm{E}_2\left[h(\mathbf{Y})\right] \right| \xrightarrow{Q_1} 0 \quad \textit{and} \tag{2.1}$$

$$\mathrm{E}_2\left[h\left(\mathbf{y}_n(\mathbf{X}_n, \mathbf{M}_n)\right)^*\right] - \mathrm{E}_2\left[h\left(\mathbf{y}_n(\mathbf{X}_n, \mathbf{M}_n)\right)_*\right] \xrightarrow{Q_1} 0 \quad \textit{as } n \to \infty \textit{ for all } h \in BL_1(\mathbb{E}). \tag{2.2}$$

Here, E_2 denotes the conditional expectation with respect to Ω_2, $BL_1(\mathbb{E})$ denotes the set of all real functions on $\mathbb{E}$ with a Lipschitz norm bounded by 1 and the super- and subscript asterisks denote the minimal measurable majorants and maximal measurable minorants, respectively, with respect to $\Omega_1 \times \Omega_2$ jointly for * and with respect to Ω_2 for ^{2*}, see [74] for details.

2.2 Conditional Delta-Method for Resampling Empirical Processes in Multiple Sample Problems

The functional delta-method has a wide range of applications in statistics. Applications to functionals of empirical processes yield various limit results for classical statistics. To improve the finite sample properties of statistical inference procedures that are based on the limit results, resampling procedures such as random permutation and bootstrap methods are a popular solution. In order to analyze the behavior of the functionals of the resampling empirical processes, corresponding conditional functional delta-methods are desirable. While conditional functional delta-methods for some special cases already exist, there is a lack of generalizations for resampling procedures for empirical processes, such as the permutation and pooled bootstrap method. This gap is addressed in this section. Thereby, a general multiple sample problem is considered. The flexible application of the developed conditional delta-method is shown in various relevant examples.
Many applications of statistics involve comparisons of multiple samples. Section 3.8 of the monograph by [74] is devoted to a related empirical process treatment. In addition to an analysis of differences of two independent

empirical processes, they also explained how to analyze a random permutation and a pooled bootstrap version of the empirical processes. Most statistical applications involve a functional that is applied to these empirical processes. The statistical properties of a functional of one empirical process can be derived with the help of the functional delta-method; cf. Section 3.10 in [74]. However, the application of random permutation or the pooled bootstrap to a multiple sample problem requires a conditional variant of a delta-method with a varying reference point.

Several extensions of the functional delta-method in different directions have already been investigated in the literature. For example, [35, 71] studied the inference of functionals that are only directionally differentiable and, recently, [61] proposed a generalization of Hadamard differentiability for applications of the functional delta-method to the empirical copula processes. Under measurability assumptions, [5] developed a modified functional delta-method for quasi-Hadamard differentiable functionals. Additionally, there exist conditional delta-methods for the bootstrap (in one sample) in Section 3.10.3 of [74] and also extensions on (uniformly) quasi-Hadamard differentiable functionals for the bootstrap under measurability assumptions [6, 7]. However, as far as we know, a two- or multiple sample equivalent of such delta-methods for resampling empirical processes is not available in the literature. In detail, most of the existing methods require some of the following:

(1) measurability assumptions,

(2) that the resampling counterpart converges weakly to the same limit as the empirical process, and

(3) a fixed centering element of the empirical process, particularly independent of the sample sizes.

However, these requirements are usually not satisfied for resampling methods for empirical processes in multiple sample problems, such as random permutation and pooled bootstrapping.

Hence, we will develop a conditional delta-method in outer probability without assuming (1)–(3) for applications to the randomly permuted and pooled bootstrapped empirical processes in multiple independent sample problems. To this end, we require the uniform Hadamard differentiability of the functionals applied to the empirical processes. In several examples, we show its applicability and usefulness. This includes conditional central limit theorems for the permutation and pooled bootstrap counterparts of the Wilcoxon statistic, the Nelson-Aalen estimator and the Kaplan-Meier estimator.

The remainder of this section is organized as follows. In Section 2.2.1 the model of our multiple sample problem is presented and the notation of this section is introduced. Moreover, existing convergence results of the resampling empirical processes are restated and a limit theorem for the permutation empirical process of multiple samples as an extension of Theorem 3.8.1 in [74] is developed in Section 2.2.2. A functional delta-method for the empirical processes of the multiple sample problem is obtained in Section 2.2.3. Furthermore, uniform Hadamard differentiability is defined and some properties are investigated in Section 2.2.4. Section 2.2.5 contains the main results of this section that cover a flexible conditional delta-method. Particularly, this delta-method is applicable for deriving the limit of functionals of permutation and pooled bootstrap empirical processes. Exemplary functionals, applications, and limitations of our main result are given in Section A. This includes the Wilcoxon functional, the product integral, and the inverse map.

2.2.1 Model and Notation

Let

$$\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i} \sim P_i, \quad i \in \{1, ..., k\}$$

be $k \geqslant 2$ independent samples of independent identically distributed (i.i.d.) random elements on a measurable space $(\chi, \mathcal{A})$ with distributions $P_1, ..., P_k$ on $(\chi, \mathcal{A})$ and let $\mathbb{P}_{i,n_i} := \frac{1}{n_i}\sum_{j=1}^{n_i} \delta_{\boldsymbol{X}_{ij}}$ be the i-th empirical measure, $i \in \{1, \ldots, k\}$, where $\delta_{\boldsymbol{X}_{ij}}$ denotes the Dirac measure centered on $\boldsymbol{X}_{ij}$.

The introduction of the resampling techniques for the empirical process requires the pooled data. To this end, denote the pooled sample by

$$(\boldsymbol{Z}_{N1}, \ldots, \boldsymbol{Z}_{NN}) := (\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n_1}, \ldots, \boldsymbol{X}_{k1}, \ldots, \boldsymbol{X}_{kn_k}),$$

where $N := \sum_{i=1}^k n_i$ is the total sample size. Let $\boldsymbol{R} = (R_1, \ldots, R_N)$ be a vector that is uniformly distributed on the set of all permutations of $\{1, 2, \ldots, N\}$ and independent of the data $\boldsymbol{Z}_{N1}, \ldots, \boldsymbol{Z}_{NN}$. Also, let $N_i := \sum_{\ell=1}^{i} n_\ell$ be the total sample size of the first i samples, with $N_0 := 0$. Then, the multiple sample permutation empirical measures are defined as

$$\mathbb{P}^{\pi}_{i,n_i} := \frac{1}{n_i} \sum_{j=N_{i-1}+1}^{N_i} \delta_{\boldsymbol{Z}_{NR_j}}, \quad i \in \{1, \ldots, k\}.$$

Next, $\mathbb{H}_N := \frac{1}{N}\sum_{j=1}^{N}\delta_{\boldsymbol{Z}_{Nj}} = \sum_{i=1}^{k}\frac{n_i}{N}\mathbb{P}_{i,n_i}$ denotes the pooled empirical measure. The multiple sample bootstrap empirical measures are defined as

$$\hat{\mathbb{P}}_{i,n_i} := \frac{1}{n_i}\sum_{j=N_{i-1}+1}^{N_i}\delta_{\hat{\boldsymbol{Z}}_{Nj}}, \quad i \in \{1,\dots,k\},$$

where $\hat{\boldsymbol{Z}}_{N1},...,\hat{\boldsymbol{Z}}_{NN} \sim \mathbb{H}_N$ is given the data $\boldsymbol{Z}_{N1},\dots,\boldsymbol{Z}_{NN}$ an i.i.d. sample drawn from the pooled empirical measure.

Throughout, we assume that $\frac{n_i}{N} \to \kappa_i \in (0,1)$, as $\min_{i=1,...,k} n_i \to \infty$. Denote $\boldsymbol{\kappa} := (\kappa_1,\dots,\kappa_k)$ and $H := \sum_{i=1}^{k}\kappa_i P_i$. Furthermore, let $\mathcal{F}$ denote a class of measurable functions $f : \chi \to \mathbb{R}$ that is P_i-Donsker for all $i \in \{1,...,k\}$, i.e., $\sqrt{n_i}(\mathbb{P}_{i,n_i} - P_i) \rightsquigarrow \mathbb{G}_i$ in the space $\ell^\infty(\mathcal{F})$ of all bounded real-valued functions on $\mathcal{F}$ as $n_i \to \infty$, where here and throughout this section $\mathbb{G}_i$ is a tight P_i-Brownian bridge for all $i \in \{1,...,k\}$ and $\rightsquigarrow$ denotes weak convergence in the sense of Section 1.3 in [74]. In the following, let $\mathbb{G}_1,\dots,\mathbb{G}_k$ be independent. An immediate consequence of the above is

$$\sqrt{N}(\mathbb{H}_N - H_{\boldsymbol{n}}) = \sum_{i=1}^{k}\frac{\sqrt{n_i}}{\sqrt{N}}\sqrt{n_i}(\mathbb{P}_{i,n_i} - P_i) \rightsquigarrow \mathbb{G}_{\boldsymbol{\kappa}} \quad \text{in } \ell^\infty(\mathcal{F}) \tag{2.3}$$

as $\min_{i=1,...,k} n_i \to \infty$, where $H_{\boldsymbol{n}} := \sum_{i=1}^{k}\frac{n_i}{N}P_i$ and $\mathbb{G}_{\boldsymbol{\kappa}} := \sum_{i=1}^{k}\sqrt{\kappa_i}\mathbb{G}_i$. It should be noted that the centering element $H_{\boldsymbol{n}}$ in the previous display generally depends on the sample sizes.

2.2.2 Weak Convergence Results of Resampling Empirical Processes

Now, we turn to the asymptotic behavior of the resampling empirical processes. We will see that the limits of the permutation and pooled bootstrap empirical process generally do not coincide with the limit of the empirical processes or the pooled empirical process.

Theorem 3.8.6 in [74] provides that $\sqrt{n_1}(\hat{\mathbb{P}}_{1,n_1} - \mathbb{H}_N) \rightsquigarrow \mathbb{G}_H$ in $\ell^\infty(\mathcal{F})$ conditionally on $\boldsymbol{X}_{11},\boldsymbol{X}_{12},\dots,\boldsymbol{X}_{21},\boldsymbol{X}_{22},\dots$ in outer probability in the two-sample case $k = 2$ under $||P_i||_{\mathcal{F}} := \sup_{f\in\mathcal{F}}|P_i f| < \infty, i \in \{1,2\}$, with $P_i f := \int f \,\mathrm{d}P_i$. Here and throughout this section, convergence results are always meant as $\min_{i=1,...,k} n_i \to \infty$ if not stated otherwise. In the following theorem, the joint convergence of the pooled bootstrap empirical processes is studied.

Theorem 2.1. *Let $\mathcal{F}$ satisfy $||P_i||_{\mathcal{F}} < \infty$ for all $i \in \{1,...,k\}$. Then, we have*

$$\sqrt{N}(\hat{\mathbb{P}}_{1,n_1} - \mathbb{H}_N,\dots,\hat{\mathbb{P}}_{k,n_k} - \mathbb{H}_N) \rightsquigarrow (\kappa_1^{-1/2}\mathbb{G}_{H,1},\dots,\kappa_k^{-1/2}\mathbb{G}_{H,k}) \quad \text{in } (\ell^\infty(\mathcal{F}))^k$$

conditionally on the data

$$\boldsymbol{X}_{11},\boldsymbol{X}_{12},\dots,\boldsymbol{X}_{21},\boldsymbol{X}_{22},\dots,\dots,\boldsymbol{X}_{k1},\boldsymbol{X}_{k2},\dots \tag{2.4}$$

in outer probability, where $\mathbb{G}_{H,1},...,\mathbb{G}_{H,k}$ denote independent tight H-Brownian bridges on $\ell^\infty(\mathcal{F})$.

For the permutation empirical measure, Theorem 3.8.1 in [74] yields under $||P_i||_{\mathcal{F}} := \sup_{f\in\mathcal{F}}|P_i f| < \infty, i \in \{1,2\}$, that $\sqrt{n_1}(\mathbb{P}^\pi_{1,n_1} - \mathbb{H}_N) \rightsquigarrow \sqrt{1-\kappa_1}\mathbb{G}_H$ in $\ell^\infty(\mathcal{F})$ conditionally on (2.4) in outer probability, where $\mathbb{G}_H$ denotes a tight H-Brownian bridge on $\ell^\infty(\mathcal{F})$. In Lemma S.6 in the supplement of [23], the almost sure version of this theorem is generalized for multiple samples. Here, we state the corresponding extension in probability which is sufficient for most statistical applications.

Theorem 2.2. *Let $\mathcal{F}$ satisfy $||P_i||_{\mathcal{F}} < \infty$ for all $i \in \{1,...,k\}$. Then, we have*

$$\sqrt{N}(\mathbb{P}^\pi_{1,n_1} - \mathbb{H}_N,\dots,\mathbb{P}^\pi_{k,n_k} - \mathbb{H}_N) \rightsquigarrow \mathbb{G}^\pi_H \quad \text{in } \ell^\infty(\mathcal{F})$$

conditionally on (2.4) *in outer probability, where $\mathbb{G}^\pi_H$ denotes a tight zero-mean Gaussian process on $(\ell^\infty(\mathcal{F}))^k$ with covariance function $\boldsymbol{\Sigma}^\pi_H : \mathcal{F}^{k\times k} \to \mathbb{R}^{k\times k}$. The component functions of $\boldsymbol{\Sigma}^\pi_H$ at $(f,g) = ((f_1,\dots,f_k),(g_1,\dots,g_k))$ are given by*

$$(\boldsymbol{\Sigma}^\pi_H(f,g))_{ij} := \left(\kappa_i^{-1}\mathbb{1}\{i=j\} - 1\right) H\left((f_i - Hf_i)(g_j - Hg_j)\right)$$

for all $i,j \in \{1,\dots,k\}$.

2.2.3 Functional Delta-Method in the Multiple Sample Problem

In statistical applications, usually a functional is applied to the empirical processes. Delta-methods can be used to analyze the asymptotic behavior of the functionals of empirical processes. A delta-method for the empirical processes of the multiple sample problem can be easily proved by applying Theorem 3.10.4 of [74], but we wished to make it more explicit, tailored to the problem at hand.

Theorem 2.3. *Let $\mathbb{E}$ be a metrizable topological vector space and $\phi : (\ell^\infty(\mathcal{F}))^k \to \mathbb{E}$ such that*

$$\sqrt{N}(\phi(\mathbf{P} + N^{-1/2}\mathbf{h}_{\boldsymbol{n}}) - \phi(\mathbf{P})) \to \phi'_{\mathbf{P}}(\mathbf{h})$$

as $\min_{i=1,\ldots,k} n_i \to \infty$ holds for every converging sequence $\mathbf{h}_{\boldsymbol{n}} \to \mathbf{h} \in (\ell^\infty(\mathcal{F}))^k$ with $\boldsymbol{n} := (n_1, \ldots, n_k)$, $\mathbf{P} := (P_1, \ldots, P_k)$, $\mathbf{P} + N^{-1/2}\mathbf{h}_{\boldsymbol{n}} \in (\ell^\infty(\mathcal{F}))^k$ for all $\boldsymbol{n}$ and for an arbitrary map $\phi'_{\mathbf{P}} : (\ell^\infty(\mathcal{F}))^k \to \mathbb{E}$. Then, we have

$$\sqrt{N}(\phi(\mathbb{P}_{1,n_1}, \ldots, \mathbb{P}_{k,n_k}) - \phi(\mathbf{P})) \rightsquigarrow \phi'_{\mathbf{P}}(\kappa_1^{-1/2}\mathbb{G}_1, \ldots, \kappa_k^{-1/2}\mathbb{G}_k).$$

If $\phi'_{\mathbf{P}}$ is linear and continuous, the sequence

$$\sqrt{N}(\phi(\mathbb{P}_{1,n_1}, \ldots, \mathbb{P}_{k,n_k}) - \phi(\mathbf{P})) - \phi'_{\mathbf{P}}(\sqrt{N}(\mathbb{P}_{1,n_1} - P_1, \ldots, \mathbb{P}_{k,n_k} - P_k))$$

converges to zero in outer probability.

The condition on ϕ in the previous theorem is satisfied if ϕ is Hadamard differentiable at $\mathbf{P}$. To define Hadamard differentiability of a functional $\phi : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{E}$, let $\mathbb{D}$ and $\mathbb{E}$ be metrizable topological vector spaces.

Definition 2.2 (Hadamard differentiability)**.** *The functional ϕ is called* Hadamard differentiable at $\theta \in \mathbb{D}_\phi$ tangentially to a subspace $\mathbb{D}_0 \subset \mathbb{D}$ *if*

$$t_n^{-1}(\phi(\theta + t_n h_n) - \phi(\theta)) \to \phi'_\theta(h)$$

holds for all $t_n \to 0$ and every converging sequence $h_n \to h \in \mathbb{D}_0$ with $\theta + t_n h_n \in \mathbb{D}_\phi$ for all n and for a continuous, linear map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$.

In order to obtain a delta-method for the permutation and pooled bootstrap empirical processes, we need to introduce the uniform Hadamard differentiability in the following paragraph.

2.2.4 Uniform Hadamard differentiability

We aim to develop functional delta-methods that are suitable for applications to $(\mathbb{P}^\pi_{1,n_1}, \ldots, \mathbb{P}^\pi_{k,n_k})$ and to $(\hat{\mathbb{P}}_{1,n_1}, \ldots, \hat{\mathbb{P}}_{k,n_k})$, conditionally on (2.4). To this end, we will consider again a functional $\phi : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{E}$, where $\mathbb{D}$ and $\mathbb{E}$ are metrizable topological vector spaces.

Definition 2.3 (Uniform Hadamard differentiability)**.** *The functional ϕ is called* uniformly Hadamard differentiable at $\theta \in \mathbb{D}_\phi$ tangentially to a subspace $\mathbb{D}_0 \subset \mathbb{D}$ *if*

$$t_n^{-1}(\phi(\theta_n + t_n h_n) - \phi(\theta_n)) \to \phi'_\theta(h)$$

holds for all $t_n \to 0$ and every converging sequence $h_n \to h \in \mathbb{D}_0$ and $\theta_n \to \theta$ with $\theta_n, \theta_n + t_n h_n \in \mathbb{D}_\phi$ for all n and for a continuous, linear map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$.

If the subspace $\mathbb{D}_0$ is not specified, we assume $\mathbb{D}_0 = \mathbb{D}$ in the following. For example, $\mathbb{D}$ can be chosen as product space $(\ell^\infty(\mathcal{F}))^k$ equipped with the max-sup-norm for applications to the empirical measures.
In the following, we investigate different properties of uniform Hadamard differentiable functionals. The following remarks address the classical Hadamard derivative as a special case, the more restrictive Fréchet differentiability, and the aggregation of multiple functionals.

Remark 2.1. *Let $\phi : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{E}$ be Hadamard differentiable at $\theta \in \mathbb{D}_\phi$ tangentially to a subspace $\mathbb{D}_0 \subset \mathbb{D}$ with Hadamard derivative $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$. If ϕ is uniformly Hadamard differentiable at $\theta \in \mathbb{D}_\phi$ tangentially to $\mathbb{D}_0$, then the (uniform) Hadamard derivative is ϕ'_θ, which can easily be seen by setting $\theta_n = \theta$ in the definition.*

Remark 2.2 (Uniform Fréchet differentiability and other sufficient criteria for uniform Hadamard differentiability)**.** *Let $(\mathbb{D}, ||.||_{\mathbb{D}})$ and $(\mathbb{E}, ||.||_{\mathbb{E}})$ be normed spaces. We call a functional ϕ uniformly Fréchet differentiable at $\theta \in \mathbb{D}_\phi$ with continuous and linear derivative ϕ'_θ if*

$$\|\phi(\theta + h) - \phi(\theta + k) - \phi'_\theta(h - k)\|_{\mathbb{E}} = o(\|h - k\|_{\mathbb{D}}), \quad \text{as } \|h\|_{\mathbb{D}}, \|k\|_{\mathbb{D}} \to 0.$$

To see that the uniform Fréchet differentiability implies the uniform Hadamard differentiability of ϕ, insert $h = \theta_n + t_n h_n - \theta$, $k = \theta_n - \theta$. According to a variant of Problem 3.10.1 in [74], uniform Fréchet differentiability at θ is implied by the Fréchet differentiability on a neighborhood of θ and the uniform norm-continuity of $\vartheta \mapsto \phi'_\vartheta$ at θ if there exists a convex neighborhood of θ as a subset of $\mathbb{D}_\phi$. Other criteria for the uniform Hadamard differentiability of ϕ at θ are the convexity of $\mathbb{D}_\phi$, the Hadamard differentiability on a neighborhood of θ, $\lim_{\eta\to\theta} \phi'_\eta(h) = \phi'_\theta(h)$ for every $h \in \text{lin } \mathbb{D}_\phi$, and $\lim_{\eta\to\theta} \phi'_\eta(h) = \phi'_\theta(h)$ uniformly in $h \in K$ for every totally bounded subset $K \subset \mathbb{D}_\phi$, where $\eta \in \mathbb{D}_\phi$; cf. (3.10.6) in [74]. Here, lin $\mathbb{D}_\phi$ *denotes the linear span of $\mathbb{D}_\phi$.*

Remark 2.3. *If $\phi_1 : \mathbb{D}_\phi \to \mathbb{E}_1, \ldots, \phi_k : \mathbb{D}_\phi \to \mathbb{E}_k$ are uniformly Hadamard differentiable at $\theta \in \mathbb{D}_\phi \subset \mathbb{D}$ tangentially to $\mathbb{D}_0 \subset \mathbb{D}$ with Hadamard derivatives $\phi'_{1,\theta} : \mathbb{D}_0 \to \mathbb{E}_1, \ldots, \phi'_{k,\theta} : \mathbb{D}_0 \to \mathbb{E}_k$, it follows directly from the definition of uniform Hadamard differentiability that $\phi := (\phi_1, \ldots, \phi_k) : \mathbb{D}_\phi \to \mathbb{E}_1 \times \cdots \times \mathbb{E}_k$ is uniformly Hadamard differentiable at θ tangentially to $\mathbb{D}_0$ with Hadamard derivative $\phi'_\theta = (\phi'_{1,\theta}, \ldots, \phi'_{k,\theta}) : \mathbb{D}_0 \to \mathbb{E}_1 \times \cdots \times \mathbb{E}_k$. Here, the product space $\mathbb{E}_1 \times \cdots \times \mathbb{E}_k$ is equipped with the product topology.*

Finally, the following theorem provides a chain rule for uniformly Hadamard differentiable functionals. It should be noted that [7] proved a chain rule for uniformly quasi-Hadamard differentiable functionals; see Lemma A.1 therein. That chain rule implies the chain rule statement below. For the sake of completeness, however, we shall present a version of the chain rule which is relevant for the remainder of this section.

Theorem 2.4 (Chain rule)**.** *Let $\mathbb{L}, \mathbb{D}, \mathbb{E}$ be metrizable topological vector spaces. If $\psi : \mathbb{L}_\psi \subset \mathbb{L} \to \mathbb{D}_\phi \subset \mathbb{D}$ is uniformly Hadamard differentiable at $\theta \in \mathbb{L}_\psi$ tangentially to $\mathbb{L}_0 \subset \mathbb{L}$ with Hadamard derivative $\psi'_\theta : \mathbb{L}_0 \to \mathbb{D}$ and $\phi : \mathbb{D}_\phi \to \mathbb{E}$ is uniformly Hadamard differentiable at $\psi(\theta)$ tangentially to $\psi'(\mathbb{L}_0)$ with Hadamard derivative $\phi'_{\psi(\theta)} : \psi'(\mathbb{L}_0) \to \mathbb{E}$, then $\phi \circ \psi : \mathbb{L}_\psi \to \mathbb{E}$ is uniformly Hadamard differentiable at θ tangentially to $\mathbb{L}_0$ with derivative $\phi'_{\psi(\theta)} \circ \psi'_\theta$.*

2.2.5 Main Results

In this section, we aim to prove a conditional delta-method, e.g., for applications to the permutation and pooled bootstrap empirical processes. For proving the main theorem, we first need the following auxiliary lemma to obtain joint unconditional convergence of two maps. The result is similar to the results in Sections 2 and 3 in [14] but allows arbitrary maps in general metric spaces.

Lemma 2.1. *Let $\mathbb{D}, \mathbb{E}$ be metric spaces, $\mathbf{X}_n : \Omega_1 \to \chi_{1n}, \mathbf{M}_n : \Omega_2 \to \chi_{2n}$ be sequences of functions, where $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, Q_1 \otimes Q_2)$ denotes a product probability space, and $\mathbf{h}_n : \chi_{1n} \to \mathbb{D}$ be such that $\mathbf{h}_n(\mathbf{X}_n) \rightsquigarrow \mathbf{H}$ as $n \to \infty$ for some separable Borel measurable random element $\mathbf{H} : \Omega_1 \to \mathbb{D}$. Moreover, let $\mathbf{y}_n : \chi_{1n} \times \chi_{2n} \to \mathbb{E}$ with $\mathbf{y}_n(\mathbf{X}_n, \mathbf{M}_n) \rightsquigarrow \mathbf{Y}$ conditionally on $\mathbf{X}_n$ in outer probability as $n \to \infty$ for some separable Borel measurable random element $\mathbf{Y} : \Omega_2 \to \mathbb{E}$. Then, it follows that $(\mathbf{h}_n(\mathbf{X}_n), \mathbf{y}_n(\mathbf{X}_n, \mathbf{M}_n)) \rightsquigarrow (\mathbf{H}, \mathbf{Y})$ unconditionally as $n \to \infty$ for independent $\mathbf{H}, \mathbf{Y}$.*

The following theorem is an extension of Theorem 3.10.11 in [74], where $\mathbb{P}_n, \widehat{\mathbb{P}}_n$ may be arbitrary maps instead of random elements, different limits $\mathbb{G}$ and $\widehat{\mathbb{G}}$ are allowed for the empirical process and its resampling counterpart, and the centering element, say P_n, may depend on n. This theorem is in particular applicable for $\widehat{\mathbb{P}}_n = \widehat{\mathbb{P}}_n(\mathbf{X}_n, \mathbf{M}_n)$ being the permutation empirical measure, i.e., $\mathbb{P}^\pi_{\boldsymbol{n}} := (\mathbb{P}^\pi_{1,n_1}, \ldots, \mathbb{P}^\pi_{k,n_k})$, or the pooled bootstrap empirical measure, i.e., $\hat{\mathbb{P}}_{\boldsymbol{n}} := (\hat{\mathbb{P}}_{1,n_1}, \ldots, \hat{\mathbb{P}}_{k,n_k})$. Here, $\mathbf{X}_n$ denotes the data and $\mathbf{M}_n$ denotes the randomness of the resampling procedures. However, we do not restrict to the cases of permutation and pooled bootstrap empirical processes in the following theorem but allow more general processes $\widehat{\mathbb{P}}_n$.

Theorem 2.5 (Conditional Delta-Method)**.** *Let $(\mathbb{D}, ||.||_{\mathbb{D}}), (\mathbb{E}, ||.||_{\mathbb{E}})$ be normed spaces, $\mathbf{X}_n : \Omega_1 \to \chi_{1n}, \mathbf{M}_n : \Omega_2 \to \chi_{2n}$ be sequences of functions, where $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, Q_1 \otimes Q_2)$ denotes a product probability space. Furthermore, let $\phi : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{E}$ be uniformly Hadamard differentiable at $P \in \mathbb{D}_\phi$ tangentially to a subspace $\mathbb{D}_0 \subset \mathbb{D}$. Moreover, let r_n be a sequence of constants tending to infinity, P_n be a sequence in $\mathbb{D}_\phi$ with $P_n \to P$ and $\mathbb{P}_n = \mathbb{P}_n(\mathbf{X}_n) : \Omega_1 \to \mathbb{D}_\phi$ a map with $r_n(\mathbb{P}_n - P_n) \rightsquigarrow \mathbb{G}$ as $n \to \infty$ for some separable Borel measurable random element $\mathbb{G} : \Omega_1 \to \mathbb{D}_0$. Additionally, let $\widehat{\mathbb{P}}_n = \widehat{\mathbb{P}}_n(\mathbf{X}_n, \mathbf{M}_n) : \Omega_1 \times \Omega_2 \to \mathbb{D}_\phi$ be maps with*

$$r_n(\widehat{\mathbb{P}}_n - \mathbb{P}_n) \rightsquigarrow \widehat{\mathbb{G}} \tag{2.5}$$

conditionally on $\mathbf{X}_n$ in outer probability as $n \to \infty$ for some separable Borel measurable random element $\widehat{\mathbb{G}} : \Omega_2 \to \mathbb{D}_0$. Then, we have $r_n\left(\phi(\widehat{\mathbb{P}}_n) - \phi(\mathbb{P}_n)\right) \rightsquigarrow \phi'_P(\widehat{\mathbb{G}})$ conditionally on $\mathbf{X}_n$ in outer probability as $n \to \infty$.

The following assertions will be stated in terms of the permutation and pooled bootstrap empirical measures.

Corollary 2.1 (Conditional Delta-Method for the Permutation Empirical Process)**.** *Let $\mathcal{F}$ satisfy $||P_i||_{\mathcal{F}} < \infty$ for all $i \in \{1, ..., k\}$ and $(\mathbb{E}, ||.||_{\mathbb{E}})$ be a normed space, $\phi : (\ell^\infty(\mathcal{F}))^k \to \mathbb{E}$ be uniformly Hadamard differentiable at*

$(H, \ldots, H) \in (\ell^\infty(\mathcal{F}))^k$ *tangentially to a subspace* $\mathbb{D}_0 \subset (\ell^\infty(\mathcal{F}))^m$ *with* $\mathbb{G}_H^\pi$ *taking values in* $\mathbb{D}_0$ *almost surely, where again* $H = \sum_{i=1}^k \kappa_i P_i$. *Then, we have*

$$\sqrt{N}\left(\phi(\mathbb{P}_{1,n_1}^\pi, \ldots, \mathbb{P}_{k,n_k}^\pi) - \phi(\mathbb{H}_N, \ldots, \mathbb{H}_N)\right) \rightsquigarrow \phi'_{(H,\ldots,H)}(\mathbb{G}_H^\pi)$$

conditionally on (2.4) *in outer probability.*

This is an application of Theorem 2.5 together with (2.3) and the conditional convergence of the permutation empirical process, see Theorem 2.2. The very same holds for the pooled bootstrap.

Corollary 2.2 (Conditional Delta-Method for the Pooled Bootstrap Empirical Process). *Let* $\mathcal{F}$ *satisfy* $||P_i||_{\mathcal{F}} < \infty$ *for all* $i \in \{1, ..., k\}$ *and* $(\mathbb{E}, ||.||_{\mathbb{E}})$ *be a normed space,* $\phi : (\ell^\infty(\mathcal{F}))^k \to \mathbb{E}$ *be uniformly Hadamard differentiable at* $(H, \ldots, H) \in (\ell^\infty(\mathcal{F}))^k$ *tangentially to a subspace* $\mathbb{D}_0 \subset (\ell^\infty(\mathcal{F}))^m$ *with* $(\lambda_1^{-1/2}\mathbb{G}_{H,1}, \ldots, \lambda_m^{-1/2}\mathbb{G}_{H,m})$ *taking values in* $\mathbb{D}_0$ *almost surely. Then, we have*

$$\sqrt{N}\left(\phi(\hat{\mathbb{P}}_{1,n_1}, \ldots, \hat{\mathbb{P}}_{k,n_k}) - \phi(\mathbb{H}_N, \ldots, \mathbb{H}_N)\right) \rightsquigarrow \phi'_{(H,\ldots,H)}(\kappa_1^{-1/2}\mathbb{G}_{H,1}, \ldots, \kappa_k^{-1/2}\mathbb{G}_{H,k})$$

conditionally on (2.4) *in outer probability.*

Note that the functional of the permutation and pooled bootstrap empirical processes in Corollaries 2.1 and 2.2 cannot mimic the same limit distribution as the functional of the original empirical process in Theorem 2.3; this is similar as for the randomization empirical process in [27]. Hence, the corollaries are not directly applicable for inference methodologies on $\phi(\mathbf{P})$ due to altered (co-)variance structures. However, a studentization can yield the consistency of the permutation and pooled bootstrap techniques with asymptotically pivotal distributions in many cases; cf. [27] for a similar approach.

2.2.6 Proofs of Section 2.2

Proof of Theorem 2.1 As in the proof of Theorem 2.9.4 in [74], it suffices to show conditional weak convergence almost surely of the marginals and

$$\limsup_{n\to\infty} \mathrm{E}^*[||\sqrt{N}(\hat{\mathbb{P}}_{i,n_i} - \mathbb{H}_N)||_{\mathcal{F}_\delta}] \xrightarrow{\delta \searrow 0} 0 \tag{2.6}$$

for all $i \in \{1, ..., k\}$, where $\mathcal{F}_\delta := \{f - g \mid f, g \in \mathcal{F}, H(f - g)^2 < \delta^2\}$.
As for Theorem 3.8.6 of [74], the Lindeberg-Feller theorem yields the conditional weak convergence almost surely of the marginals of $\sqrt{N}(\hat{\mathbb{P}}_{i,n_i} - \mathbb{H}_N)$ for all $i \in \{1, ..., k\}$. Then, the conditional independence provides the conditional weak convergence almost surely of the marginals of $\sqrt{N}(\hat{\mathbb{P}}_{1,n_1} - \mathbb{H}_N, \ldots, \hat{\mathbb{P}}_{k,n_k} - \mathbb{H}_N)$.
For (2.6), we can proceed as in the proof of Theorem 3.8.1 in [74]. In contrast to the equicontinuity condition considered there, we look at the outer expectation in terms of the joint probability space in (2.6). Therefore, note that Lemmas 2.3.1, 2.3.11, 3.6.5 and, thus, also Lemma 3.7.6 in [74] all hold for outer expectations in terms of the joint probability space. ☐

Proof of Theorem 2.2 Again, it suffices to show conditional weak convergence almost surely of the marginals and

$$\limsup_{n\to\infty} \mathrm{E}^*[||\sqrt{N}(\mathbb{P}_{i,n_i}^\pi - \mathbb{H}_N)||_{\mathcal{F}_\delta}] \xrightarrow{\delta \searrow 0} 0 \tag{2.7}$$

for all $i \in \{1, ..., k\}$.
For the conditional weak convergence almost surely of the marginals, we proceed similar to the proof of (S.18) in the supplement of [23] with a Cramér-Wold argument, where $P = H$ and $\mathbb{P} = \mathbb{H}_N$. Let $g_i = c_i f_i$ with $c_i \in [-1, 1]$ and $f_i \in \mathcal{F}$ for all $i \in \{1, \ldots, k\}$. Then,

$$\frac{1}{n_i} \max\{g_r(\boldsymbol{X}_{ij})^2 : j \in \{1, \ldots, n_i\}\} \to 0$$

holds almost surely for all $r, i \in \{1, \ldots, k\}$, which can be shown with the three steps (i)–(iii) in the beginning of the proof of Lemma S.6 in the supplement of [23] by using g_r instead of the envelope function $\tilde{F}$. In detail, the three steps are the following:

(i) dividing g_r into $g_{r,1,M} := g_r 1\{|g_r| \leqslant M\}$ and $g_{r,2,M} := g_r 1\{|g_r| > M\}$ for $M \in \mathbb{N}$,

(ii) using the inequalities $(a + b)^2 \leqslant 2a^2 + 2b^2$ and

$$\max_{j=1,\ldots,n_i} g_{r,2,M}(\boldsymbol{X}_{ij})^2 \leqslant \sum_{j=1}^{n_i} g_{r,2,M}(\boldsymbol{X}_{ij})^2,$$

(iii) letting first $n \to \infty$ and finally $M \to \infty$.

In the last step (iii) we need $P_i g_r^2 < \infty$ for an application of the dominated convergence theorem as $M \to \infty$, which holds due to the assumption that $\mathcal{F}$ is P_i-Donsker. Moreover, we have $\mathbb{H}_N g_r \to H g_r$ almost surely for all $r \in \{1, \ldots, k\}$ by the strong law of large numbers since $\mathcal{F}$ is P_i-Donsker for all $i \in \{1, \ldots, k\}$, and, thus, H-Donsker. Condition (S.19) in the supplement of [23] follows almost surely by the same arguments as given there. For proving condition (S.20) almost surely, it remains to show that $\mathbb{H}_N(g_i g_r) \to H(g_i g_r)$ almost surely for all $i, r \in \{1, \ldots, k\}$ by the last display in the proof of Lemma S.6 in the supplement of [23]. Since $\mathcal{F}$ is P_i-Donsker for every $i \in \{1, \ldots, k\}$, $H(g_i g_r)$ exists and, thus, the almost sure convergence follows by the strong law of large numbers. Hence, the conditional weak convergence almost surely follows from (S.18) in the supplement of [23] given (2.4) almost surely.
Now, we turn to condition (2.7). First, we need a version of Hoeffding's inequality for outer expectations; cf. Proposition A.1.10 in [74] for a similar statement for expectations. For $i \in \{1, ..., k\}$, let $(M_1, ..., M_N)$ denote a multinomially distributed random variable with n_i trials and probabilities $(N^{-1}, ..., N^{-1})$ independent of the data (2.4) and of the random variable $\boldsymbol{R}$. Moreover, define

$$\pi(j) := \min\{\operatorname{argmax}_{j' \in \{1,\ldots,N\} \setminus \{\pi(1),\ldots,\pi(j-1)\}} M_{j'}\}$$

for all $j \in \{1, ..., N\}$ and π' a random permutation of $\{1, ..., n_i\}$ independent of $(M_1, ..., M_N)$, (2.4) and $\boldsymbol{R}$. Note that π is a permutation of $\{1, ..., N\}$ with $M_{\pi(j)} = 0$ for all $j > n_i$. Thus, one can show

$$\begin{aligned}\sum_{j'=N_{i-1}+1}^{N_i} \delta_{\hat{\boldsymbol{Z}}_{Nj'}} &\stackrel{d}{=} \sum_{j=1}^{N} M_j \delta_{\boldsymbol{Z}_{NR_j}} = \sum_{j=1}^{N} M_{\pi(j)} \delta_{\boldsymbol{Z}_{NR_{\pi(j)}}} = \sum_{j=1}^{n_i} M_{\pi(j)} \delta_{\boldsymbol{Z}_{NR_{\pi(j)}}} = \sum_{j=1}^{n_i} M_{\pi(\pi'(j))} \delta_{\boldsymbol{Z}_{NR_{\pi(\pi'(j))}}} \\ &\stackrel{d}{=} \sum_{j=1}^{n_i} M_{\pi(\pi'(j))} \delta_{\boldsymbol{Z}_{NR_{N_{i-1}+j}}},\end{aligned}$$

where $\hat{\boldsymbol{Z}}_{Nj}, j \in \{N_{i-1}+1, ..., N_i\}$, has the same distribution as i.i.d. observations drawn from the pooled empirical measure. Furthermore, note that

$$\mathrm{E}_M \mathrm{E}_{\pi'} \left[M_{\pi(\pi'(j))}\right] = \mathrm{E}_M \left[\frac{1}{n_i} \sum_{j'=1}^{n_i} M_{\pi(j')}\right] = \frac{1}{n_i} \mathrm{E}_M \left[\sum_{j'=1}^{N} M_{j'}\right] = 1$$

for all $j \in \{1, ..., n_i\}$ with $\mathrm{E}_M, \mathrm{E}_{\pi'}$ denoting the expectation regarding $(M_1, ..., M_N), \pi'$, respectively. Hence, we obtain

$$\begin{aligned}\mathrm{E}^*[||\sqrt{N}(\mathbb{P}^{\pi}_{i,n_i} - \mathbb{H}_N)||_{\mathcal{F}_\delta}] &= \mathrm{E}\left[\left\|\sqrt{N}\left(\frac{1}{n_i}\sum_{j=N_{i-1}+1}^{N_i} \delta_{\boldsymbol{Z}_{NR_j}} - \mathbb{H}_N\right)\right\|^*_{\mathcal{F}_\delta}\right] \\ &= \mathrm{E}\left[\left\|\sqrt{N}\left(\frac{1}{n_i}\sum_{j=1}^{n_i} \mathrm{E}_M \mathrm{E}_{\pi'}\left[M_{\pi(\pi'(j))}\right] \delta_{\boldsymbol{Z}_{NR_{N_{i-1}+j}}} - \mathbb{H}_N\right)\right\|^*_{\mathcal{F}_\delta}\right] \\ &\leqslant \mathrm{E}\mathrm{E}_M \mathrm{E}_{\pi'}\left[\left\|\sqrt{N}\left(\frac{1}{n_i}\sum_{j=1}^{n_i} M_{\pi(\pi'(j))} \delta_{\boldsymbol{Z}_{NR_{N_{i-1}+j}}} - \mathbb{H}_N\right)\right\|^*_{\mathcal{F}_\delta}\right] \\ &= \mathrm{E}\left[\left\|\sqrt{N}\left(\frac{1}{n_i}\sum_{j'=N_{i-1}+1}^{N_i} \delta_{\hat{\boldsymbol{Z}}_{Nj'}} - \mathbb{H}_N\right)\right\|^*_{\mathcal{F}_\delta}\right] \\ &= \mathrm{E}^*[||\sqrt{N}(\hat{\mathbb{P}}_{i,n_i} - \mathbb{H}_N)||_{\mathcal{F}_\delta}]\end{aligned}$$

with E denoting the joint expectation. Now, (2.6) implies (2.7). □

Proof of Theorem 2.4 Let $t \to 0, h_t \to h \in \mathbb{L}_0, \theta_t \to \theta$ with $\theta_t, \theta_t + t h_t \in \mathbb{L}_\psi$. Write

$$(\phi \circ \psi)(\theta_t + t h_t) - (\phi \circ \psi)(\theta_t) = \phi(\psi(\theta_t) + t k_t) - \phi(\psi(\theta_t)),$$

where $k_t = (\psi(\theta_t + t h_t) - \psi(\theta_t))/t$. Now, the uniform Hadamard differentiability of ψ yields that $k_t \to \psi'_\theta(h)$. Next, the uniform Hadamard differentiabiliy implies the theorem. □

Proof of Lemma 2.1 We aim to apply Corollary 1.4.5 in [74]. Hence, it remains to show

$$\mathrm{E}^* f(\mathbf{h}_n(\mathbf{X}_n))g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n)) \to \mathrm{E}f(\mathbf{H})\mathrm{E}g(\mathbf{Y})$$

for all bounded, nonnegative Lipschitz functions $f:\mathbb{D}\to\mathbb{R}, g:\mathbb{E}\to\mathbb{R}$. We have

$$\begin{aligned}
&|\mathrm{E}^* f(\mathbf{h}_n(\mathbf{X}_n))g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n)) - \mathrm{E}f(\mathbf{H})\mathrm{E}g(\mathbf{Y})| \\
&\leqslant \left|\mathrm{E}_1\mathrm{E}_2\left(f(\mathbf{h}_n(\mathbf{X}_n))g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))\right)^* - \mathrm{E}_1\left(f(\mathbf{h}_n(\mathbf{X}_n))^*\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^*\right)\right| \\
&\quad + \left|\mathrm{E}_1\left(f(\mathbf{h}_n(\mathbf{X}_n))^*\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^*\right) - \mathrm{E}_1 f(\mathbf{H})\mathrm{E}_2 g(\mathbf{Y})\right|,
\end{aligned}$$

where $\mathrm{E}_1,\mathrm{E}_2$ denote the expectations regarding $(\Omega_1,\mathcal{A}_1,Q_1)$ and $(\Omega_2,\mathcal{A}_2,Q_2)$, respectively. By Lemma 1.2.2 (v) in [74] and the nonnegativity of f and g, it holds that

$$\left(f(\mathbf{h}_n(\mathbf{X}_n))g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))\right)^* \leqslant f(\mathbf{h}_n(\mathbf{X}_n))^* g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^*$$

almost surely and, hence, it follows that

$$\mathrm{E}_2\left(f(\mathbf{h}_n(\mathbf{X}_n))g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))\right)^* \leqslant f(\mathbf{h}_n(\mathbf{X}_n))^*\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^*$$

almost surely. Thus, we get

$$\begin{aligned}
&|\mathrm{E}^* f(\mathbf{h}_n(\mathbf{X}_n))g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n)) - \mathrm{E}f(\mathbf{H})\mathrm{E}g(\mathbf{Y})| \\
&\leqslant \mathrm{E}_1\left((f(\mathbf{h}_n(\mathbf{X}_n))^*\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^*) - \mathrm{E}_2\left(f(\mathbf{h}_n(\mathbf{X}_n))g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))\right)^*\right) \\
&\quad + \left|\mathrm{E}_1\left(f(\mathbf{h}_n(\mathbf{X}_n))^*\left(\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^* - \mathrm{E}_2 g(\mathbf{Y})\right)\right)\right| \\
&\quad + \left|\mathrm{E}_1\left((f(\mathbf{h}_n(\mathbf{X}_n))^* - f(\mathbf{H}))\,\mathrm{E}_2 g(\mathbf{Y})\right)\right| \\
&\leqslant \mathrm{E}_1\left((f(\mathbf{h}_n(\mathbf{X}_n))^*\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^*) - f(\mathbf{h}_n(\mathbf{X}_n))_*\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))_*\right) && (2.8) \\
&\quad + ||f||_\infty \mathrm{E}_1\left(\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^* - \mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^{2*} + \left|\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^{2*} - \mathrm{E}_2 g(\mathbf{Y})\right|\right) && (2.9) \\
&\quad + ||g||_\infty \left|\mathrm{E}_1\left(f(\mathbf{h}_n(\mathbf{X}_n))^* - f(\mathbf{H})\right)\right|. && (2.10)
\end{aligned}$$

By $g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))_* \leqslant g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^{2*} \leqslant g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^*$, (2.1) and (2.2) imply

$$\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^* - \mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^{2*} + \left|\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^{2*} - \mathrm{E}_2 g(\mathbf{Y})\right| \to 0$$

in outer probability. Hence, the dominated convergence theorem provides that (2.9) converges to zero. Due to $\mathbf{h}_n(\mathbf{X}_n) \rightsquigarrow \mathbf{H}$, (2.10) converges to zero. Hence, (2.8) remains to consider. First note that (2.8) can be written as

$$\begin{aligned}
&\mathrm{E}_1\left((f(\mathbf{h}_n(\mathbf{X}_n))^*\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^*) - f(\mathbf{h}_n(\mathbf{X}_n))_*\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))_*\right) \\
&= \mathrm{E}_1\left(f(\mathbf{h}_n(\mathbf{X}_n))^*\left(\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^* - \mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))_*\right)\right) \\
&\quad + \mathrm{E}_1\left((f(\mathbf{h}_n(\mathbf{X}_n))^* - f(\mathbf{h}_n(\mathbf{X}_n))_*)\,\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))_*\right) \\
&\leqslant ||f||_\infty \mathrm{E}_1\left(\mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))^* - \mathrm{E}_2 g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n))_*\right) \\
&\quad + ||g||_\infty \mathrm{E}_1\left(f(\mathbf{h}_n(\mathbf{X}_n))^* - f(\mathbf{h}_n(\mathbf{X}_n))_*\right).
\end{aligned}$$

By (2.2) and the dominated convergence theorem, the first summand converges to zero. The second summand converges to zero since $\mathbf{h}_n(\mathbf{X}_n)$ is asymptotically measurable. Consequently, $\mathrm{E}^* f(\mathbf{h}_n(\mathbf{X}_n))g(\mathbf{y}_n(\mathbf{X}_n,\mathbf{M}_n)) \to \mathrm{E}f(\mathbf{H})\mathrm{E}g(\mathbf{Y})$ follows. □

Proof of Theorem 2.5 We proceed analogously as in the proof of Theorem 3.10.11 in [74] by applying their Theorem 3.10.5 (rather than Theorem 3.10.4). First note that we may assume without loss of generality that the derivative $\phi'_P:\mathbb{D}\to\mathbb{E}$ is not only defined and continuous but also linear on the whole space $\mathbb{D}$ by their Problem 3.10.18. For all $h\in BL_1(\mathbb{E})$, we have $h\circ\phi'_P \in BL_{\max\{1,||\phi'_P||\}}(\mathbb{D})$ where $||\phi'_P||$ is the operator norm of ϕ'_P. By (2.5), it follows that

$$\sup_{h\in BL_1(\mathbb{E})}\left|\mathrm{E}_2 h\left(\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right)^{2*} - \mathrm{E}h(\phi'_P(\widehat{\mathbb{G}}))\right| \to 0$$

in outer probability. Let $\varepsilon>0$ be arbitrary. Since $|h(A)^{2*}-h(B)^{2*}| \leqslant |h(A)-h(B)|^* \leqslant ||A-B||_{\mathbb{E}}^*$ holds for all $h\in BL_1(\mathbb{E}), A,B\in\mathbb{E}$, it follows that

$$\begin{aligned}
&\sup_{h\in BL_1(\mathbb{E})}\left|\mathrm{E}_2 h\left(r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(\mathbb{P}_n)\right)\right)^{2*} - \mathrm{E}_2 h\left(\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right)^{2*}\right| \\
&\leqslant \varepsilon + 2Q_2\left(\left\|r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(\mathbb{P}_n)\right) - \phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right\|_{\mathbb{E}}^* > \varepsilon\right).
\end{aligned}$$

Theorem 3.10.5 in [74] implies that

$$\left\|r_n\left(\phi(\mathbb{P}_n)-\phi(P_n)\right)-\phi'_P\left(r_n(\mathbb{P}_n-P_n)\right)\right\|_{\mathbb{E}}^{*}=o_{Q_1}(1).$$

By Lemma 2.1 and the continuous mapping theorem, we have

$$r_n(\widehat{\mathbb{P}}_n-P_n)=r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)+r_n(\mathbb{P}_n-P_n)\rightsquigarrow\widehat{\mathbb{G}}+\mathbb{G}$$

unconditionally for independent $\widehat{\mathbb{G}},\mathbb{G}$. Hence, Theorem 3.10.5 in [74] implies that

$$\left\|r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(P_n)\right)-\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-P_n)\right)\right\|_{\mathbb{E}}^{*}=o_{Q_1\otimes Q_2}(1)$$

and, by the triangle inequality, that

$$\left\|r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(\mathbb{P}_n)\right)-\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right\|_{\mathbb{E}}^{*}=o_{Q_1\otimes Q_2}(1).$$

Markov's inequality yields that

$$\begin{aligned}
&Q_1\left(Q_2\left(\left\|r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(\mathbb{P}_n)\right)-\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right\|_{\mathbb{E}}^{*}>\varepsilon\right)>\delta\right)\\
&\leqslant \mathrm{E}_1Q_2\left(\left\|r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(\mathbb{P}_n)\right)-\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right\|_{\mathbb{E}}^{*}>\varepsilon\right)/\delta\\
&=(Q_1\otimes Q_2)\left(\left\|r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(\mathbb{P}_n)\right)-\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right\|_{\mathbb{E}}^{*}>\varepsilon\right)/\delta\\
&\to 0
\end{aligned}$$

for all $\delta>0$. Thus, it follows

$$\sup_{h\in BL_1(\mathbb{E})}\left|\mathrm{E}_2h\left(r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(\mathbb{P}_n)\right)\right)^{2*}-\mathrm{E}_2h\left(\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right)^{2*}\right|\to 0$$

in outer probability. Analogously, we can conclude

$$\sup_{h\in BL_1(\mathbb{E})}\left|\mathrm{E}_2h\left(r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(\mathbb{P}_n)\right)\right)^{*}-\mathrm{E}_2h\left(\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right)^{*}\right|\to 0 \qquad (2.11)$$

in outer probability.
For the asymptotic measurability in outer probability, write

$$\begin{aligned}
&\mathrm{E}_2h\left(r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(\mathbb{P}_n)\right)\right)^{*}-\mathrm{E}_2h\left(r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(\mathbb{P}_n)\right)\right)_{*}\\
&\leqslant\left|\mathrm{E}_2h\left(r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(\mathbb{P}_n)\right)\right)^{*}-\mathrm{E}_2h\left(\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right)^{*}\right|\\
&\quad+\mathrm{E}_2h\left(\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right)^{*}-\mathrm{E}_2h\left(\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right)_{*}\\
&\quad+\left|\mathrm{E}_2\left(-h\left(r_n\left(\phi(\widehat{\mathbb{P}}_n)-\phi(\mathbb{P}_n)\right)\right)\right)^{*}-\mathrm{E}_2\left(-h\left(\phi'_P\left(r_n(\widehat{\mathbb{P}}_n-\mathbb{P}_n)\right)\right)\right)^{*}\right|.
\end{aligned}$$

Then, the asymptotic measurability (2.2) in outer probability follows from (2.5) and (2.11). □

2.3 General Methodology for Simultaneous Inference

Statistical hypothesis tests offer a technique to make decisions about a null hypothesis. Tests are required to control the type-I error rate, that is the probability that the null hypothesis is rejected although the null hypothesis is true. However, often more than one hypothesis is of interest in practice and, hence, more than one test is needed to infer the hypotheses simultaneously. By just controlling the type-I error rate for each test, the probability that at least one true null hypothesis is rejected increases generally. Hence, we aim to control the family-wise error rate (FWER), which is the probability to reject at least one true null hypothesis, in the strong sense for multiple testing problems in the following thesis. Here, *in the strong sense* means that the FWER is controlled for any set of true and false null hypotheses.
Therefore, we derive the general methodology that is used in Sections 4 and 5 to infer multiple hypotheses simultaneously in this section. Firstly, we present the multiple testing setup and state a general theorem in Section 2.3.1 which provides a multiple testing procedure that controls the FWER in the strong sense asymptotically. Moreover, we show that the critical values for the tests can be determined by consistent resampling schemes in Section 2.3.2.

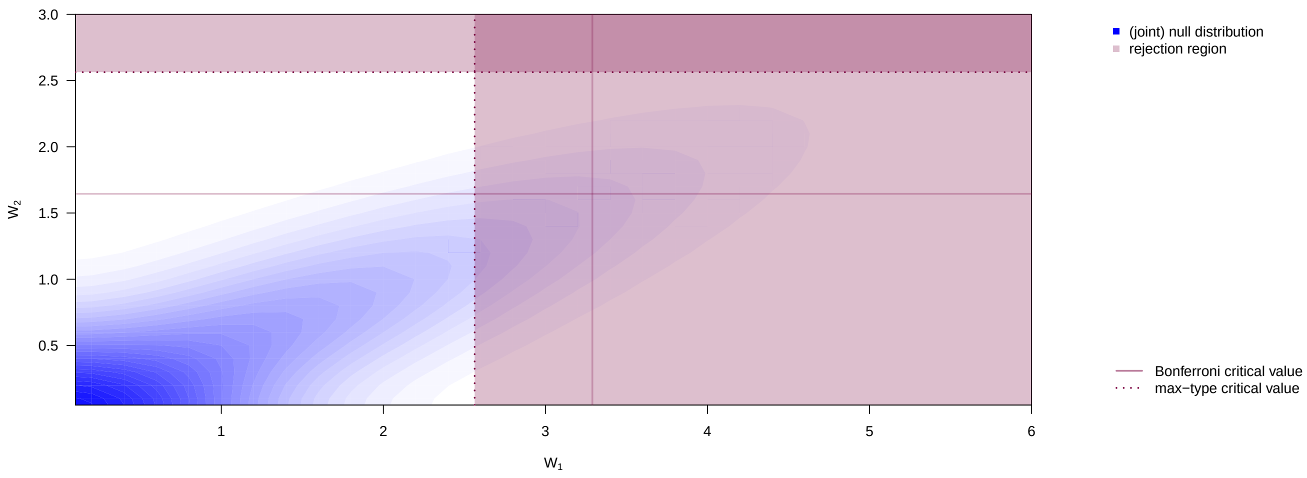


Figure 1: Exemplary illustration of the max-type testing procedure (dotted lines) for two test statistics with different distributions compared to the Bonferroni-correction (solid lines).

2.3.1 General Multiple Testing Setup

In this section, we are considering a general multiple testing problem with $L \in \mathbb{N}$ local null hypotheses $\mathcal{H}_{0,\ell}$ and global null hypothesis $\mathcal{H}_0 := \bigcap_{\ell=1}^{L} \mathcal{H}_{0,\ell}$. Suppose that sequences of maps $W_{\ell,n} : \Omega \to \mathbb{R}, \ell \in \{1, \ldots, L\}$, on a probability space $(\Omega, \mathcal{A}, P)$ are present, which we will call test statistics in the following. Despite the common definition of test statistics, we do not assume measurability of $W_{\ell,n}, \ell \in \{1, \ldots, L\}$, in the following. In the applications of this thesis, the measurability of the test statistics will usually be given. However, nonmeasurable test statistics might be of interest, for example, a Kolmogorov-Smirnov statistic of empirical processes over an uncountable function class, see Section 3.8 in [74]. We assume that large values of the ℓth test statistic $W_{\ell,n}$ indicate a rejection of the ℓth null hypothesis $\mathcal{H}_{0,\ell}$ for every $\ell \in \{1, \ldots, L\}$. In the following, we will discuss how *large* can be quantified, i.e., how the critical values for the local test decisions can be constructed.
A naive approach for compatible local and global test decisions would be to calculate a (max-type) critical value for the maximum statistic $\max_{\ell \in \{1,\ldots,L\}} W_{\ell,n}$. The ℓth null hypothesis $\mathcal{H}_{0,\ell}$ is rejected whenever the corresponding local test statistic $W_{\ell,n}$ exceeds the critical value. Suppose that

$$(W_{\ell,n})_{\ell \in \mathcal{T}} \xrightarrow{d} (W_\ell)_{\ell \in \mathcal{T}} \quad \text{as } n \to \infty \text{ under } \bigcap_{\ell \in \mathcal{T}} \mathcal{H}_{0,\ell} \text{ for all index sets } \mathcal{T} \subset \{1, \ldots, L\}, \tag{2.12}$$

where $W_\ell, \ell \in \{1, ..., L\}$, denote random variables. In the special case that the distributions of $W_1, \ldots, W_L$ are equal, every local hypothesis has asymptotically the same probability to be wrongly rejected. However, if the distributions are not equal, the local hypotheses may have different asymptotic probabilities to be wrongly detected by considering the maximum statistic and, thus, are not treated in the same way, cf. Figure 1. For a fair comparison, we therefore adopt the idea of balanced simultaneous confidence sets as in [4].
In detail, we aim to find individual critical values $q_{1,n}, ..., q_{L,n}$ for the local hypotheses such that

$$\limsup_{n\to\infty} P^* \left(\exists \ell \in \mathcal{T} :\ W_{\ell,n} > q_{\ell,n}\right) \leqslant \alpha \quad \text{under } \bigcap_{\ell \in \mathcal{T}} \mathcal{H}_{0,\ell} \text{ for all index sets } \mathcal{T} \subset \{1, \ldots, L\} \tag{2.13}$$

for a global level $\alpha \in (0, 1)$ and

$$\limsup_{n\to\infty} P^* \left(W_{\ell,n} > q_{\ell,n}\right) =: \beta \quad \text{under } \mathcal{H}_{0,\ell} \text{ for all } \ell \in \{1, \ldots, L\}, \tag{2.14}$$

where $\beta \in (0, 1)$ does not depend on ℓ. Here and throughout, P^* denotes the outer probability to avoid measurability issues. Hence, we do not restrict to the case that $W_{\ell,n}$ and $q_{\ell,n}$ are measurable. The first condition (2.13) ensures the asymptotic FWER control in the strong sense. The second condition (2.14) guarantees that all local hypotheses are treated in the same way and is referred to as *asymptotically balanced* multiple tests, cf. [4]. If the joint distribution of $(W_\ell)_{\ell \in \{1,\ldots,L\}}$ is known with joint cumulative distribution function $\mathcal{F} : \mathbb{R}^L \to [0, 1]$ and continuous marginal cumulative distribution functions $\mathcal{F}_\ell : \mathbb{R} \to [0, 1], \ell \in \{1, ..., L\}$, the critical values can be determined as $q_{\ell,n} = \mathcal{F}_{\ell-}^{-1}(1 - \beta)$ and $\beta \in (0, 1)$ is chosen such that $1 - \mathcal{F}(\mathcal{F}_{1-}^{-1}(1 - \beta), ..., \mathcal{F}_{L-}^{-1}(1 - \beta)) \leqslant \alpha$ holds. Here and throughout, $F_-^{-1}(p) := \sup\{x \in \mathbb{R} \mid F_-(x) \leqslant p\}$ denotes the largest p-quantile and F_- denotes the left-continuous version of a monotone function $F : \mathbb{R} \to [0, 1]$ for $p \in \mathbb{R}$. Note that even for left-continuous

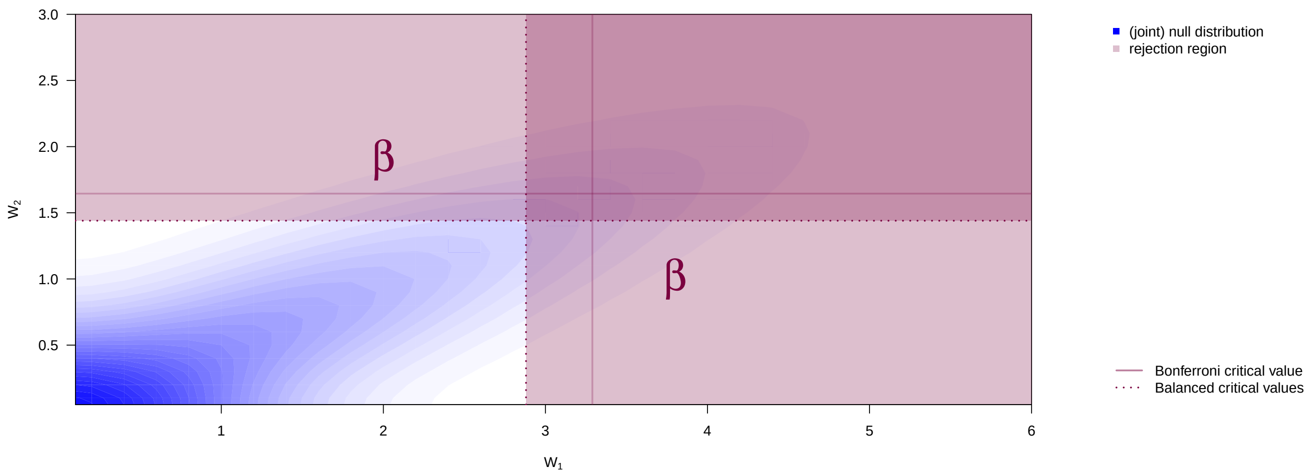


Figure 2: Exemplary illustration of the balanced multiple testing procedure (dotted lines) for two test statistics with different distributions compared to the Bonferroni-correction (solid lines). Integrating over the rejection regions for the first (right part) and second (upper part) hypothesis yields the same value β.

monotone functions $F : \mathbb{R} \to [0,1]$, we have $F_-^{-1}(p) \neq F^{-1}(p)$ in general for $p \in \mathbb{R}$, which is why we explicitly write $F_-^{-1}(p)$ even for left-continuous functions. Furthermore, it should be noted that β can always be chosen as the Bonferroni-corrected local level α/L, cf. [33]. However, larger values for β may increase the power of the multiple testing procedure. An exemplary illustration of the balanced multiple testing procedure compared to the Bonferroni-correction can be found in Figure 2.

In many applications, the joint distribution of $(W_\ell)_{\ell \in \{1,\dots,L\}}$ depends on unknown parameters or is approximated by a Monte-Carlo method due to the complexity of the distribution and quantile functions, see for example Sections 4 and 5. Hence, the joint cumulative distribution function $\mathcal{F}$ is usually approximated by (random) sequences $(F_n)_{n\in\mathbb{N}}$ of cumulative distribution functions in practice. Moreover, resampling procedures are often used to approximate the limit distribution to improve the small sample performance. In this case, even the marginal cumulative distribution functions $\mathcal{F}_\ell, \ell \in \{1,\dots,L\}$, are approximated by (random) sequences $(F_{\ell,n})_{n\in\mathbb{N}}, \ell \in \{1,\dots,L\}$, of cumulative distribution functions. The following lemma ensures that the critical values $q_{1,n},\dots,q_{L,n}$ can be approximated through random critical values based on F_n and $F_{\ell,n}, \ell \in \{1,\dots,L\}$, as long as the sequences converge in outer probability to the true distribution functions.

Theorem 2.6. *Let $L \in \mathbb{N}$ and $\alpha \in (0,1)$. Moreover, let F_n denote a map on $(\Omega,\mathcal{A},P)$ taking values in the space of all cumulative distribution functions on $\mathbb{R}^L$ and $F_{\ell,n}, \ell \in \{1,\dots,L\}$, denote maps on $(\Omega,\mathcal{A},P)$ taking values in the space of all cumulative distribution functions on $\mathbb{R}$ for all $n \in \mathbb{N}$. Additionally, let $\mathcal{F} : \mathbb{R}^L \to [0,1]$ denote a cumulative distribution function of a random vector $(W_1,\dots,W_L)$ with continuous marginal distribution functions $\mathcal{F}_1,\dots,\mathcal{F}_L : \mathbb{R} \to [0,1]$. Furthermore, set $\mathrm{FWER}(\zeta) := 1 - \mathcal{F}\left(\mathcal{F}_{1-}^{-1}(1-\zeta),\dots,\mathcal{F}_{L-}^{-1}(1-\zeta)\right)$ for all $\zeta \in \mathbb{R}$ and assume that FWER is strictly increasing on $[a,b]$ with $0 \leqslant a < \alpha/L \leqslant \alpha < b \leqslant 1$. Let*

$$F_n(\mathbf{t}) \xrightarrow{P} \mathcal{F}(\mathbf{t}) \text{ as } n \to \infty \text{ for all } \mathbf{t} \in \mathbb{R}^L \tag{2.15}$$

and

$$F_{\ell,n}(t) \xrightarrow{P} \mathcal{F}_\ell(t) \text{ as } n \to \infty \text{ for all } t \in \mathbb{R}, \ell \in \{1,\dots,L\} \tag{2.16}$$

hold. Furthermore, suppose we have a sequence of maps β_n on $(\Omega,\mathcal{A},P)$ taking values in $[0,1]$ and satisfying $\beta_n \in \left[\mathrm{FWER}_{n+}^{-1}(\alpha) - \varepsilon_n, \mathrm{FWER}_{n+}^{-1}(\alpha + \varepsilon_n)\right]$ for all $n \in \mathbb{N}$ with

$$\mathrm{FWER}_{n+}(\zeta) := 1 - F_n\left(F_{1,n}^{-1}(1-\zeta),\dots,F_{L,n}^{-1}(1-\zeta)\right) \qquad \text{for all } \zeta \in \mathbb{R}$$

and some null sequence $(\varepsilon_n)_{n\in\mathbb{N}} \subset [0,\infty)$. If additionally (2.12) holds, it follows

$$\lim_{n\to\infty} P^*\left(\exists \ell \in \mathcal{T} : W_{\ell,n} > q_{\ell,n}\right) \leqslant \alpha \qquad \text{under } \bigcap_{\ell \in \mathcal{T}} \mathcal{H}_{0,\ell} \text{ for all index sets } \mathcal{T} \subset \{1,\dots,L\}, \tag{2.17}$$

$$\lim_{n\to\infty} P^*\left(\exists \ell \in \{1,\dots,L\} : W_{\ell,n} > q_{\ell,n}\right) = \alpha \qquad \text{under the global null hypothesis } \mathcal{H}_0 \text{ and} \tag{2.18}$$

$$\lim_{n\to\infty} P^*\left(W_{\ell,n} > q_{\ell,n}\right) = \mathrm{FWER}^{-1}(\alpha) \qquad \text{under } \mathcal{H}_{0,\ell} \text{ for all } \ell \in \{1,\dots,L\}, \tag{2.19}$$

where $q_{\ell,n} := F_{\ell,n-}^{-1}(1-\beta_n)$.

Note that (2.13) and (2.14) are direct consequences from (2.17) and (2.19).
The decision rules for the global and local hypotheses are then constructed as follows:

- For each $\ell \in \{1, ..., L\}$, we reject $\mathcal{H}_{0,\ell}$ if and only if $W_{\ell,n} > q_{\ell,n}$ or, equivalently, $W_{\ell,n}/q_{\ell,n} > 1$. Here, we set $0/0 := 0$.
- We reject the global null hypothesis $\mathcal{H}_0$ whenever at least one of the hypotheses $\mathcal{H}_{0,1}, ..., \mathcal{H}_{0,L}$ is rejected. Hence, we reject the global null hypothesis $\mathcal{H}_0$ if and only if $\max_{\ell \in \{1,...,L\}} W_{\ell,n}/q_{\ell,n} > 1$.

Here, each test statistic $W_{\ell,n}, \ell \in \{1, ..., L\}$, is treated in the same way and has asymptotically the same impact since we use the same local level of significance β_n for each local hypothesis. The resulting tests can be formulated accordingly as $\varphi_\ell := \mathbb{1}\{W_{\ell,n} > q_{\ell,n}\}$ for $\mathcal{H}_{0,\ell}$ for all $\ell \in \{1, ..., L\}$ and as $\varphi := \mathbb{1}\{\max_{\ell \in \{1,...,L\}} W_{\ell,n}/q_{\ell,n} > 1\}$ for $\mathcal{H}_0$. Then, (2.17)–(2.19) can be formulated equivalently as

$$\lim_{n\to\infty} \mathrm{E}^*\left(\max_{\ell\in\mathcal{T}} \varphi_\ell\right) \leqslant \alpha \qquad \text{under } \bigcap_{\ell\in\mathcal{T}} \mathcal{H}_{0,\ell} \text{ for all index sets } \mathcal{T} \subset \{1,\ldots,L\}, \tag{2.20}$$

$$\lim_{n\to\infty} \mathrm{E}^*\left(\max_{\ell\in\{1,\ldots,L\}} \varphi_\ell\right) = \alpha \qquad \text{under the global null hypothesis } \mathcal{H}_0 \text{ and} \tag{2.21}$$

$$\lim_{n\to\infty} \mathrm{E}^*\left(\varphi_\ell\right) = \beta \qquad \text{under } \mathcal{H}_{0,\ell} \text{ for all } \ell \in \{1,\ldots,L\}. \tag{2.22}$$

In the following remark, we give a condition under which (2.16) easily follows from (2.15).

Remark 2.4. *If $F_{1,n}, \ldots, F_{L,n}$ in Theorem 2.6 are the marginal cumulative distribution functions of F_n,* (2.16) *is a direct consequence of* (2.15).

The following lemma ensures that the function FWER is strictly increasing in our applications in Sections 4 and 5.

Lemma 2.2. *Let $k \in \mathbb{N}$ and $W_\ell := \mathbf{Z}^\top \mathbf{A}_\ell \mathbf{Z}, \ell \in \{1, ..., L\}$, for a random vector $\mathbf{Z}$ taking values in $\mathbb{R}^k$ with a positive Lebesgue density on all of $\mathbb{R}^k$ and $\mathbf{A}_1, ..., \mathbf{A}_L \in \mathbb{R}^{k\times k}$ being symmetric positive semi-definite matrices with* $\mathrm{rank}(\mathbf{A}_\ell) > 0, \ell \in \{1, ..., L\}$. *Moreover, let $\mathcal{F} : \mathbb{R}^L \to [0, 1]$ denote the cumulative distribution function of $(W_1, ..., W_L)$ and $\mathcal{F}_\ell : \mathbb{R} \to [0, 1], \ell \in \{1, ..., L\}$, denote the continuous marginal distribution functions. Then,*

$$[0,1] \ni \zeta \mapsto \mathrm{FWER}(\zeta) := 1 - \mathcal{F}\left(\mathcal{F}_{1-}^{-1}(1-\zeta), \ldots, \mathcal{F}_{L-}^{-1}(1-\zeta)\right)$$

is strictly increasing.

It is well known that the closed testing procedure may improve the power of multiple tests. Hence, we propose a stepwise extension of the multiple testing procedure in the following remark.

Remark 2.5 (Stepwise Extension)**.** *Our methodologies can be combined with the closed testing procedure as in [10] to gain more power. Therefore, for each $\ell \in \{1, ..., L\}$, the hypothesis $\mathcal{H}_{0,\ell}$ is rejected at level α if and only if for each $\mathcal{J} \subset \{1, ..., L\}$ with $\mathcal{J} \ni \ell$ the intersection hypothesis $\mathcal{H}_{0,\mathcal{J}} := \bigcap_{j\in\mathcal{J}} \mathcal{H}_{0,j}$ is rejected at level α. For testing an intersection hypothesis $\mathcal{H}_{0,\mathcal{J}}$, we can use the procedure as described above. To be specific, $\mathcal{H}_{0,\mathcal{J}}$ is rejected at level α whenever $\max_{j\in\mathcal{J}} W_{j,n}/F_{j,n-}^{-1}\left(1-\beta_n^{\mathcal{J}}\right) > 1$ holds, where $\beta_n^{\mathcal{J}} \in \left[\mathrm{FWER}_{n+}^{\mathcal{J},-1}(\alpha) - \varepsilon_n, \mathrm{FWER}_{n+}^{\mathcal{J},-1}(\alpha + \varepsilon_n)\right]$ for all $n \in \mathbb{N}$ with some null sequence $(\varepsilon_n)_{n\in\mathbb{N}} \subset [0, \infty)$,*

$$\mathrm{FWER}_{n+}^{\mathcal{J}}(\zeta) := 1 - F_n^{\mathcal{J}}\left((F_{j,n}^{-1}(1-\zeta))_{j\in\mathcal{J}}\right) \qquad \textit{for all } \zeta \in \mathbb{R}$$

and $F_n^{\mathcal{J}}$ denoting the marginal cumulative distribution function of F_n with respect to the components with indices $\mathcal{J}$.

Moreover, multiple tests for estimands yield simultaneous confidence regions. This is explained in more detail in the following remark.

Remark 2.6 (Simultaneous Confidence Regions)**.** *Let us consider the local hypotheses $\mathcal{H}_{0,\ell} : \mathbf{h}_\ell(P) = \mathbf{c}_\ell, \ell \in \{1, ..., L\}$, about estimands $\mathbf{h}_\ell(P) \in \Xi_\ell$ and $\mathbf{c}_\ell \in \Xi_\ell$ for all $\ell \in \{1, ..., L\}$, where $\Xi_1, ..., \Xi_L$ denote arbitrary sets (e.g., $\mathbb{R}^{r_1}, ..., \mathbb{R}^{r_L}$). Moreover, we write $W_{\ell,n}(\mathbf{c}_\ell)$ for the ℓth local test statistic to express the dependence of $W_{\ell,n}$ on $\mathbf{c}_\ell$. Then, we can use the constructed multiple testing procedure to define simultaneous confidence regions for $\mathbf{h}_\ell(P)$ with asymptotic global confidence level $1 - \alpha$. Under the notation and conditions of Theorem 2.6, we define the ℓth confidence region as*

$$CR_{\ell,n} := \{\boldsymbol{\xi} \in \Xi_\ell \mid W_{\ell,n}(\boldsymbol{\xi}) \leqslant q_{\ell,n}\}$$

for all $\ell \in \{1, ..., L\}$. Then, it can be easily checked that $\lim_{n\to\infty} P^*(\exists \ell \in \{1, \ldots, L\} : \mathbf{h}_\ell(P) \notin CR_{\ell,n}) = \alpha$ holds.

2.3.2 Consistent Resampling Schemes

When considering a consistent resampling scheme, the cumulative distribution function $\mathcal{F}$ can be approximated through the empirical distribution function of the resampled test statistics by a Monte Carlo method. In this section, we show that the consistency implies (2.15) and construct adjusted p-values that lead to the same test decisions as in Section 2.3.1.
A resampling scheme is called consistent if the resampled test statistics converge weakly conditionally on the data in outer probability to $(W_1, \ldots, W_L)$, i.e., to the same limit distribution as the original test statistics under the global null hypothesis. We aim to formulate a lemma that (2.15) in Theorem 2.6 follows for consistent resampling schemes, where F_n denotes the empirical distribution function of B_n independently resampled test statistics. Here, $\mathbf{X}_n$ represents the randomness of the data while $\mathbf{M}_n$ can be interpreted as the randomness of the resampling method.

Lemma 2.3. *Let $\mathbf{X}_n : \Omega_0 \to \chi_{1n}, \mathbf{M}_n : \Omega \to \chi_{2n}$ denote sequences of maps, where $(\Omega_0 \times \Omega^{\mathbb{N}}, \mathcal{A}_i \otimes \mathcal{A}^{\otimes\mathbb{N}}, P_0 \otimes P^{\otimes\mathbb{N}})$ denotes a product probability space and χ_{1n}, χ_{2n} are arbitrary sets for $n \in \mathbb{N}$. Furthermore, assume that $\mathbf{W}_n : \chi_{1n} \times \chi_{2n} \to \mathbb{R}^L$ is a function for $L \in \mathbb{N}$ and all $n \in \mathbb{N}$. Suppose that F_n is the empirical distribution function of*

$$\mathbf{W}_n^{(b)} : \Omega_0 \times \Omega^{\mathbb{N}} \to \mathbb{R}^L, \quad \mathbf{W}_n^{(b)}(\omega_0, \omega_1, ...) := \mathbf{W}_n(\mathbf{X}_n(\omega_0), \mathbf{M}_n(\omega_b)), \qquad b \in \{1, ..., B_n\},$$

for $B_n, n \in \mathbb{N}$ satisfying $\mathbf{W}_n^{(1)} \xrightarrow{d*} \mathbf{W} \sim \mathcal{F}$ conditionally on $\mathbf{X}_n$ in outer probability as $n \to \infty$, where $\mathcal{F} : \mathbb{R}^L \to [0,1]$ denotes a cumulative distribution function with continuous marginal cumulative distribution functions. Then, (2.15) *is satisfied if $B_n \to \infty$ as $n \to \infty$.*

The lemma ensures that the empirical distribution function of independently resampled test statistics $\mathbf{W}_n^{(b)} = (W_{1,n}^{(b)}, \ldots, W_{L,n}^{(b)})$, $b \in \{1, ..., B_n\}$, that is

$$F_n : \mathbb{R}^L \to [0,1], \quad F_n(\mathbf{w}) = \frac{1}{B_n} \sum_{b=1}^{B_n} \mathbb{1}\left\{\mathbf{W}_n^{(b)} \leqslant \mathbf{w}\right\},$$

fulfills the condition to approximate the function $\mathcal{F}$ in Theorem 2.6 for a consistent resampling scheme, where $\mathbf{W}_n^{(b)} \leqslant \mathbf{w}$ means that all components of $\mathbf{W}_n^{(b)}$ are less than or equal to the corresponding values in the vector $\mathbf{w}$. For these choices, we have

$$\mathrm{FWER}_{n+}(\zeta) = \frac{1}{B_n} \sum_{b=1}^{B_n} \mathbb{1}\left\{\exists \ell \in \{1, ..., L\} : W_{\ell,n}^{(b)} > F_{\ell,n}^{-1}(1-\zeta)\right\}$$

for all $\zeta \in \mathbb{R}$. Regarding Remark 2.4, the marginal distribution functions can be approximated by the marginals of the empirical distribution function, that are

$$F_{\ell,n} : \mathbb{R} \to [0,1], \quad F_{\ell,n}(w) = \frac{1}{B_n} \sum_{b=1}^{B_n} \mathbb{1}\left\{W_{\ell,n}^{(b)} \leqslant w\right\}, \quad \ell \in \{1, \ldots, L\}.$$

Then, we can define the local level β_n as the largest value such that the estimated family-wise type I error rate is bounded by the level of significance α, i.e.,

$$\begin{aligned} \beta_n &:= \max\left\{\zeta \in \left\{\frac{1}{B_n}, \frac{2}{B_n}, ..., 1\right\} \mid \mathrm{FWER}_{n+}(\zeta - B_n^{-1}) \leqslant \alpha\right\} \\ &= \begin{cases} \mathrm{FWER}_{n+}^{-1}(\alpha + 1/B_n) - 1/B_n & \text{if } \alpha B_n \in \mathbb{N} \\ \mathrm{FWER}_{n+}^{-1}(\alpha) - 1/B_n & \text{if } \alpha B_n \notin \mathbb{N}. \end{cases} \end{aligned}$$

Note that we only have to consider $\zeta \in \left\{\frac{1}{B_n}, \frac{2}{B_n}, ..., 1\right\}$ since the quantiles can only take B_n different values, respectively. Additionally, we only have to search for β_n within the interval $\left[\frac{1}{B_n}\left\lfloor\frac{B_n\alpha}{L}\right\rfloor + \frac{1}{B_n}, 1\right]$. The lower bound can be interpreted as Bonferroni bound and results from the following inequalities:

$$\mathrm{FWER}_{n+}\left(\frac{1}{B_n}\left\lfloor\frac{B_n\alpha}{L}\right\rfloor\right) \leqslant \sum_{\ell=1}^{L} \frac{1}{B_n} \sum_{b=1}^{B_n} \mathbb{1}\left\{W_{\ell,n}^{(b)} > F_{\ell,n}^{-1}\left(1 - \frac{1}{B_n}\left\lfloor\frac{B_n\alpha}{L}\right\rfloor\right)\right\} \leqslant L\frac{1}{B_n}\left\lfloor\frac{B_n\alpha}{L}\right\rfloor \leqslant \alpha.$$

With the above choices for $F_n, F_{\ell,n}$ and β_n, Theorem 2.6 implies that $q_{\ell,n} := F_{\ell,n-}^{-1}(1-\beta_n), \ell \in \{1, ..., L\}$, fulfills (2.17)–(2.19) under the conditions of Theorem 2.6 and Lemma 2.3.

Adjusted p-values The method described in Theorem 2.6 for constructing multiple tests based on a consistent resampling scheme is accompanied by an adjustment of p-values. To see this, we determine the local p-values by

$$\beta_{\ell,n} := 1 - F_{\ell,n-}(W_{\ell,n})$$

for all $\ell \in \{1, ..., L\}$. Comparing the local p-values to β_n yields multiple test decisions that are consistent to the method in Theorem 2.6. Translating this comparison to a comparison with the level of significance α is intuitive due to the definition of β_n. Hence, by plugging the local p-value in FWER_{n+}, the adjusted p-value for the ℓth hypothesis can be defined by

$$p_\ell := \mathrm{FWER}_{n+}(\beta_{\ell,n})$$

for all $\ell \in \{1, ..., L\}$ and the global p-value by $p := \min\{p_1, ..., p_L\}$. The following proposition ensures that the test decisions based on these p-values are unchanged.

Proposition 2.1. *With the notation of Section 2.3.2,*

(1) for each $\ell \in \{1, ..., L\}$, it holds $p_\ell \leqslant \alpha$ whenever $W_{\ell,n} > q_{\ell,n}$,

(2) it holds $p \leqslant \alpha$ whenever $\max_{\ell \in \{1,...,L\}} W_{\ell,n}/q_{\ell,n} > 1$.

2.3.3 Proofs of Section 2.3

We start by proving Lemma 2.3, as we will use one of the techniques in the proof of Theorem 2.6.

Proof of Lemma 2.3 Let $\mathbf{t} \in \mathbb{R}^L$ be arbitrary. Approximate $f : \mathbb{R}^L \to \{0, 1\}, \mathbf{w} \mapsto \mathbb{1}\{\mathbf{w} \leqslant \mathbf{t}\}$ through sequences of Lipschitz functions $(g_m)_{m\in\mathbb{N}}, (h_m)_{m\in\mathbb{N}}$ with $1 \geqslant g_m \geqslant f \geqslant h_m \geqslant 0$ and $\mathrm{E}[g_m(\mathbf{W}) - h_m(\mathbf{W})] \leqslant m^{-1}$ for all $m \in \mathbb{N}$, where $\mathbf{x} \leqslant \mathbf{t}$ means that all components of $\mathbf{x}$ are less than or equal to the corresponding values in the vector $\mathbf{t}$. Let E_2^* denote the outer expectation with respect to the product space $\Omega^{\mathbb{N}}$, cf. Section 1.2 in [74]. Then, it holds

$$\begin{aligned}
&\left|\mathrm{E}_2\left[f(\mathbf{W}_n^{(1)})^*\right] - \mathrm{E}\left[f(\mathbf{W})\right]\right| \\
&\leqslant \mathrm{E}_2\left[f(\mathbf{W}_n^{(1)})^*\right] - \mathrm{E}_2^*\left[f(\mathbf{W}_n^{(1)})\right] + \left|\mathrm{E}_2^*\left[f(\mathbf{W}_n^{(1)})\right] - \mathrm{E}\left[f(\mathbf{W})\right]\right| \\
&\leqslant \mathrm{E}_2\left[g_m(\mathbf{W}_n^{(1)})^*\right] - \mathrm{E}_2^*\left[h_m(\mathbf{W}_n^{(1)})\right] \\
&\quad + \max\left\{\mathrm{E}_2^*\left[g_m(\mathbf{W}_n^{(1)})\right] - \mathrm{E}\left[g_m(\mathbf{W})\right], \mathrm{E}\left[h_m(\mathbf{W})\right] - \mathrm{E}_2^*\left[h_m(\mathbf{W}_n^{(1)})\right]\right\} \\
&\quad + \mathrm{E}\left[g_m(\mathbf{W})\right] - \mathrm{E}\left[h_m(\mathbf{W})\right] \\
&\leqslant \mathrm{E}_2\left[g_m(\mathbf{W}_n^{(1)})^* - g_m(\mathbf{W}_n^{(1)})_*\right] + \mathrm{E}_2^*\left[g_m(\mathbf{W}_n^{(1)})\right] - \mathrm{E}_2^*\left[h_m(\mathbf{W}_n^{(1)})\right] \\
&\quad + \max\left\{\mathrm{E}_2^*\left[g_m(\mathbf{W}_n^{(1)})\right] - \mathrm{E}\left[g_m(\mathbf{W})\right], \mathrm{E}\left[h_m(\mathbf{W})\right] - \mathrm{E}_2^*\left[h_m(\mathbf{W}_n^{(1)})\right]\right\} + m^{-1} \\
&\xrightarrow{P_0} 0 + \mathrm{E}\left[g_m(\mathbf{W})\right] - \mathrm{E}\left[h_m(\mathbf{W})\right] + 0 + m^{-1} \leqslant 2m^{-1},
\end{aligned}$$

where here and throughout A^* and A_* denote the minimal measurable majorant and maximal measurable minorant, respectively, for a map A with respect to all probability spaces jointly. By choosing m sufficiently large, it follows

$$\mathrm{E}_2\left[f(\mathbf{W}_n^{(1)})^*\right] \xrightarrow{P_0} \mathrm{E}\left[f(\mathbf{W})\right] = \mathcal{F}(\mathbf{t})$$

as $n \to \infty$. Analogously, one can show $\mathrm{E}_2\left[f(\mathbf{W}_n^{(1)})_*\right] \xrightarrow{P_0} \mathcal{F}(\mathbf{t})$ as $n \to \infty$. Since $B_n \to \infty$ as $n \to \infty$, we have

$$
\begin{aligned}
&\mathrm{E}_2\left[\left(|F_n(\mathbf{t}) - \mathcal{F}(\mathbf{t})|^2\right)^*\right] \\
&\leqslant \mathrm{E}_2\left[\left(F_n^2(\mathbf{t})\right)^*\right] + 2\mathrm{E}_2\left[(-F_n(\mathbf{t}))^*\right]\mathcal{F}(\mathbf{t}) + \mathcal{F}^2(\mathbf{t}) \\
&\leqslant \frac{1}{B_n^2}\sum_{b_1,b_2=1}^{B_n} \mathrm{E}_2\left[\left(f(\mathbf{W}_n^{(b_1)})f(\mathbf{W}_n^{(b_2)})\right)^*\right] + \frac{2}{B_n}\sum_{b=1}^{B_n} \mathrm{E}_2\left[\left(-f(\mathbf{W}_n^{(b)})\right)^*\right]\mathcal{F}(\mathbf{t}) + \mathcal{F}^2(\mathbf{t}) \\
&\leqslant \frac{1}{B_n} + \frac{1}{B_n^2}\sum_{b_1,b_2=1,b_1\neq b_2}^{B_n} \mathrm{E}_2\left[f(\mathbf{W}_n^{(b_1)})^*\right]\mathrm{E}_2\left[f(\mathbf{W}_n^{(b_2)})^*\right] - 2\mathrm{E}_2\left[f(\mathbf{W}_n^{(1)})_*\right]\mathcal{F}(\mathbf{t}) + \mathcal{F}^2(\mathbf{t}) \\
&\leqslant \frac{1}{B_n} + \frac{B_n - 1}{B_n}\mathrm{E}_2\left[f(\mathbf{W}_n^{(1)})^*\right]^2 - 2\mathrm{E}_2\left[f(\mathbf{W}_n^{(1)})_*\right]\mathcal{F}(\mathbf{t}) + \mathcal{F}^2(\mathbf{t}) \\
&\xrightarrow{P_0} \mathcal{F}^2(\mathbf{t}) - 2\mathcal{F}(\mathbf{t})\mathcal{F}(\mathbf{t}) + \mathcal{F}^2(\mathbf{t}) = 0
\end{aligned}
$$

as $n \to \infty$. Thus, $\mathrm{E}^*\left[|F_n(\mathbf{t}) - \mathcal{F}(\mathbf{t})|^2\right] \to 0$ as $n \to \infty$ for all $\mathbf{t} \in \mathbb{R}^L$ by the dominated convergence theorem (dominated by 1), where E^* denotes the outer expectation with respect to all probability measures jointly. Hence, (2.15) in Theorem 2.6 follows. □

Proof of Theorem 2.6 For proving Theorem 2.6, we firstly need two lemmas. The first one is a multivariate version of Polya's theorem.

Lemma 2.4. *Let $L \in \mathbb{N}$, $A := \times_{\ell=1}^{L}[a_\ell, b_\ell] \subset (\mathbb{R} \cup \{-\infty, \infty\})^L$, $\mathcal{F} : (\mathbb{R} \cup \{-\infty, \infty\})^L \to [0,1]$ be a cumulative distribution function that is continuous on A and $(F_n)_{n\in\mathbb{N}}$ denote a sequence of maps on $(\Omega, \mathcal{A}, P)$ taking values in the space of all cumulative distribution functions satisfying*

$$
F_n(\mathbf{t}) \xrightarrow{P} \mathcal{F}(\mathbf{t}) \text{ as } n \to \infty \text{ for all } \mathbf{t} \in A. \tag{2.23}
$$

Then, we have $\sup_{\mathbf{t}\in A} |F_n(\mathbf{t}) - \mathcal{F}(\mathbf{t})| \xrightarrow{P} 0$ *as* $n \to \infty$.

Proof of Lemma 2.4. We aim to apply Proposition 2.1 in [9]. Therefore, let $f_{\mathbf{t}} : \mathbb{R}^L \to \mathbb{R}, f_{\mathbf{t}}(x_1, ..., x_L) = \mathbb{1}\{x_1 \leqslant t_1, ..., x_L \leqslant t_L\}$ for all $\mathbf{t} = (t_1, ..., t_L)' \in A$, $\mathbf{F} := \{f_{\mathbf{t}} \mid \mathbf{t} \in A\}$ and $\varepsilon > 0$ be arbitrary. Furthermore, let $\mathcal{F}_1, ..., \mathcal{F}_L : \mathbb{R} \to [0,1]$ denote the marginal cumulative distribution functions of $\mathcal{F}$, $m \in \mathbb{N}$ with $\varepsilon/L \geqslant 1/m$ and define

$$
a_\ell =: t_{\ell,0} < t_{\ell,1} < ... < t_{\ell,m} := b_\ell
$$

such that $\mathcal{F}_\ell(t_{\ell,i}) - \mathcal{F}_\ell(t_{\ell,i-1}) \leqslant \varepsilon/L$ for all $i \in \{1, ..., m\}, \ell \in \{1, ..., L\}$. Set $\mathbf{t}_{i_1,...,i_L} := (t_{1,i_1}, ..., t_{L,i_L})'$ for all $i_1, ..., i_L \in \{0, ..., m\}$. Then, it holds that

$$
\int_{\mathbb{R}^L} f_{\mathbf{t}_{i_1,...,i_L}} - f_{\mathbf{t}_{i_1-1,...,i_L-1}} \, \mathrm{d}\mathcal{F} = \mathcal{F}(\mathbf{t}_{i_1,...,i_L}) - \mathcal{F}(\mathbf{t}_{i_1-1,...,i_L-1}) \leqslant \sum_{\ell=1}^{L} \left(\mathcal{F}_\ell(t_{\ell,i_\ell}) - \mathcal{F}_\ell(t_{\ell,i_\ell-1})\right) \leqslant \varepsilon
$$

for all $i_1, ..., i_L \in \{1, ..., m\}$. Thus, the bracketing number is bounded by $(m+1)^L < \infty$. As in the proof of Proposition 2.1 in [9], it holds

$$
\sup_{\mathbf{t}\in A} |F_n(\mathbf{t}) - \mathcal{F}(\mathbf{t})| \leqslant 3\max\{|F_n(\mathbf{t}) - \mathcal{F}(\mathbf{t})|\} + 2\varepsilon,
$$

where the maximum is taken over $2(m+1)^L$ different values $\mathbf{t} \in A$. Hence, we get

$$
P^*\left(\sup_{\mathbf{t}\in A} |F_n(\mathbf{t}) - \mathcal{F}(\mathbf{t})| > 3\varepsilon\right) \leqslant P^*\left(\max\{|F_n(\mathbf{t}) - \mathcal{F}(\mathbf{t})|\} > \varepsilon\right),
$$

where P^* denotes the outer probability. The latter tends to 0 by (2.23) since the maximum is taken over a finite number of values $\mathbf{t} \in A$. □

The following lemma ensures that the quantiles of a converging sequence of cumulative distribution functions converge.

Lemma 2.5. *Let $\mathcal{F} : \mathbb{R} \to [0,1]$ be a distribution function that is continuous and strictly increasing on $[a,b] \subset \mathbb{R}$ and $(F_n)_{n\in\mathbb{N}}$ denote a sequence of maps on $(\Omega, \mathcal{A}, P)$ taking values in the space of all cumulative distribution functions satisfying*

$$F_n(t) \xrightarrow{P} \mathcal{F}(t) \quad \text{for all } t \in [a,b] \tag{2.24}$$

as $n \to \infty$. Furthermore, let $\mathcal{F}(a) < p \leqslant q < \mathcal{F}(b)$. Then, we have

$$\sup_{r\in[p,q]} |F_n^{-1}(r) - \mathcal{F}^{-1}(r)| \xrightarrow{P} 0$$

as $n \to \infty$.

Proof of Lemma 2.5. First of all, Lemma 2.4 implies $\sup_{t\in[a,b]} |F_n(t) - \mathcal{F}(t)| \xrightarrow{P} 0$ as $n \to \infty$. Then, by Theorem 1.9.2 (ii) in [74], every subsequence has a further subsequence such that

$$\sup_{t\in[a,b]} |F_n(t) - \mathcal{F}(t)| \to 0 \quad \text{as } n \to \infty \text{ outer almost surely} \tag{2.25}$$

holds along the latter subsequence. In the following, we are considering the sequence only along this subsequence and proceed similarly as in [74]. Let $(\delta_n)_{n\in\mathbb{N}}$ be a positive sequence with $\delta_n \to 0$ as $n \to \infty$. By (2.25), there exists an $N \in \mathbb{N}$ such that

$$\mathcal{F}(b) - F_n(b) \leqslant \mathcal{F}(b) - q \quad \text{and} \quad F_n(a+\delta_n) - \mathcal{F}(a+\delta_n) < (p - \mathcal{F}(a))/2$$

holds for all $n \geqslant N$ outer almost surely. Due to the continuity of $\mathcal{F}$, we can choose N sufficiently large such that $\mathcal{F}(a+\delta_n) \leqslant \mathcal{F}(a) + (p - \mathcal{F}(a))/2$ holds for all $n \geqslant N$. Hence, it follows that $F_n(b) \geqslant q$ and $F_n(a+\delta_n) < p$ for all $n \geqslant N$ outer almost surely. Since

$$F_n^{-1}(r) \leqslant x \quad \Leftrightarrow \quad r \leqslant F_n(x)$$

holds for all $r \in [p,q], x \in \mathbb{R}$ due to the definition of the inverse map, we have $F_n^{-1}(r) \leqslant b$ and $F_n^{-1}(r) > F_n^{-1}(r) - \delta_n > a$ for all $r \in [p,q]$ and $n \geqslant N$ outer almost surely.
Moreover, it holds

$$F_n(F_n^{-1}(r) - \delta_n) \leqslant r \leqslant F_n(F_n^{-1}(r))$$

for all $r \in [p,q], n \in \mathbb{N}$ by the definition of the inverse map. Hence, it follows

$$\mathcal{F}(F_n^{-1}(r)) - F_n(F_n^{-1}(r)) \leqslant \mathcal{F}(F_n^{-1}(r)) - r \leqslant \mathcal{F}(F_n^{-1}(r)) - F_n(F_n^{-1}(r) - \delta_n) \tag{2.26}$$

for all $r \in [p,q], n \in \mathbb{N}$. The left side of (2.26) is converging to 0 uniformly in r as $n \to \infty$ outer almost surely by (2.25). Since $\mathcal{F}$ is continuous on the compact set $[a,b]$, it is also uniformly continuous. The right side of (2.26) can be rewritten as

$$\begin{aligned}\mathcal{F}(F_n^{-1}(r)) - F_n(F_n^{-1}(r) - \delta_n) =& \mathcal{F}(F_n^{-1}(r)) - \mathcal{F}(F_n^{-1}(r) - \delta_n)\\ &+ \mathcal{F}(F_n^{-1}(r) - \delta_n) - F_n(F_n^{-1}(r) - \delta_n),\end{aligned}$$

where the first part vanishes asymptotically uniformly in r due to the uniform continuity of $\mathcal{F}$ and the second part outer almost surely due to (2.25). Thus, (2.26) implies

$$\sup_{r\in[p,q]} \left|\mathcal{F}(F_n^{-1}(r)) - r\right| \to 0 \quad \text{as } n \to \infty \text{ outer almost surely.}$$

By the strict monotony of $\mathcal{F}$ on $[a,b]$, $\mathcal{F}^{-1}$ is continuous on $[(\mathcal{F}(a)+p)/2, \mathcal{F}(b)]$ and, thus, uniformly continuous on $[(\mathcal{F}(a)+p)/2, \mathcal{F}(b)]$. Let $\varepsilon > 0$ be arbitrary and $\delta \in (0, (p - \mathcal{F}(a))/2]$ such that

$$|\mathcal{F}^{-1}(x) - \mathcal{F}^{-1}(y)| < \varepsilon$$

holds for all $x, y \in [(\mathcal{F}(a)+p)/2, \mathcal{F}(b)]$ with $|x - y| < \delta$. There exists an $M \in \mathbb{N}$ such that

$$\sup_{r\in[p,q]} |\mathcal{F}(F_n^{-1}(r)) - r| < \delta$$

holds for all $n \geqslant M$ outer almost surely. This further implies that $\mathcal{F}(F_n^{-1}(r)) > r - \delta \geqslant (\mathcal{F}(a)+p)/2$ for all $r \in [p,q], n \geqslant M$ outer almost surely. Since ${F_n}^{-1}(r) \leqslant b$ for all $r \in [p,q], n \geqslant N$, we also have $\mathcal{F}(F_n^{-1}(r)) \leqslant \mathcal{F}(b)$ for all $r \in [p,q], n \geqslant N$. Hence, it follows that

$$\sup_{r\in[p,q]} |F_n^{-1}(r) - \mathcal{F}^{-1}(r)| = \sup_{r\in[p,q]} |\mathcal{F}^{-1}(\mathcal{F}(F_n^{-1}(r))) - \mathcal{F}^{-1}(r)| < \varepsilon$$

for all $n \geqslant \max\{N, M\}$ outer almost surely. Applying Theorem 1.9.2 (ii) in [74] again completes the proof. ☐

Now, we aim to show that β_n converges in outer probability. Since $\varepsilon_n \to \infty$ as $n \to \infty$, it remains to show

$$\sup_{r\in[\alpha,c]} \left|\mathrm{FWER}_{n+}^{-1}(r) - \mathrm{FWER}^{-1}(r)\right| \xrightarrow{P} 0 \quad \text{as } n \to \infty \tag{2.27}$$

for some $c > \alpha$. Therefore, we apply Lemma 2.5. Note that FWER_{n+} and FWER can be seen as distribution functions and FWER is continuous and strictly increasing on $[a, b]$ by assumption. By Lemma 2.4, we have

$$\sup_{t_1,...,t_R\in\mathbb{R}} |F_n(t_1, ..., t_R) - \mathcal{F}(t_1, ..., t_R)| \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

Moreover, Lemma 2.4 implies the uniform convergence of the marginal distribution functions, that is

$$\sup_{t\in\mathbb{R}} |F_{\ell,n}(t) - \mathcal{F}_\ell(t)| \xrightarrow{P} 0 \tag{2.28}$$

as $n \to \infty$ for all $\ell \in \{1, ..., R\}$. For $\mathrm{FWER}_{n-} := (\mathrm{FWER}_{n+})_-$ being the left-continuous version of FWER_{n+}, we have

$$\begin{aligned}
|\mathrm{FWER}_{n-}(\zeta) - \mathrm{FWER}(\zeta)| &= \left|F_n\left(F_{1,n-}^{-1}(1-\zeta), \dots, F_{L,n-}^{-1}(1-\zeta)\right) - \mathcal{F}\left(\mathcal{F}_{1-}^{-1}(1-\zeta), \dots, \mathcal{F}_{L-}^{-1}(1-\zeta)\right)\right| \\
&\leqslant \sup_{t_1,...,t_R\in\mathbb{R}} |F_n(t_1, ..., t_R) - \mathcal{F}(t_1, ..., t_R)| \\
&\quad + \left|\mathcal{F}\left(F_{1,n-}^{-1}(1-\zeta), \dots, F_{L,n-}^{-1}(1-\zeta)\right) - \mathcal{F}\left(\mathcal{F}_{1-}^{-1}(1-\zeta), \dots, \mathcal{F}_{L-}^{-1}(1-\zeta)\right)\right|
\end{aligned}$$

for all $\zeta \in [0, 1]$. Here, the first summand converges to zero in outer probability. For the second summand, note that

$$\mathcal{F}_\ell\left(\mathcal{F}_{\ell-}^{-1}(1-\zeta)\right) = 1 - \zeta, \tag{2.29}$$

$$F_-(w) \leqslant 1 - \zeta \Leftrightarrow w \leqslant F_-^{-1}(1-\zeta), \text{ and} \tag{2.30}$$

$$|\mathcal{F}(x_1, ..., x_L) - \mathcal{F}(y_1, ..., y_L)| \leqslant \sum_{\ell=1}^{L} |\mathcal{F}_\ell(x_\ell) - \mathcal{F}_\ell(y_\ell)| \tag{2.31}$$

hold for all $\zeta \in [0, 1], w, x_1, ..., x_L, y_1, ..., y_L \in \mathbb{R}$ and all cumulative distribution functions $F : \mathbb{R} \to [0, 1]$. Let $\varepsilon > 0$ and $\ell \in \{1, ..., L\}$ be arbitrary. Then, it holds

$$\mathcal{F}_\ell\left(F_-^{-1}(1-\zeta)\right) = P\left(W_\ell \leqslant F_-^{-1}(1-\zeta)\right) = P\left(F_-(W_\ell) \leqslant 1-\zeta\right) \begin{cases} \leqslant P\left(\mathcal{F}_\ell(W_\ell) \leqslant 1-\zeta+\varepsilon\right) = 1-\zeta+\varepsilon \\ \geqslant P\left(\mathcal{F}_\ell(W_\ell) \leqslant 1-\zeta-\varepsilon\right) = 1-\zeta-\varepsilon \end{cases}$$

by (2.29) and (2.30) for all cumulative distribution functions $F : \mathbb{R} \to \mathbb{R}$ with $\sup_{t\in\mathbb{R}} |F(t) - \mathcal{F}_\ell(t)| \leqslant \varepsilon$ and all $\zeta \in [0, 1]$. Thus, it follows $\left|\mathcal{F}_\ell\left(F_-^{-1}(1-\zeta)\right) - \mathcal{F}_\ell\left(\mathcal{F}_{\ell-}^{-1}(1-\zeta)\right)\right| \leqslant \varepsilon$ by (2.29) for all cumulative distribution functions $F : \mathbb{R} \to \mathbb{R}$ with $\sup_{t\in\mathbb{R}} |F(t) - \mathcal{F}_\ell(t)|$ and all $\zeta \in [0, 1]$. Hence, (2.28) implies

$$P^*\left(\left|\mathcal{F}_\ell\left(F_{\ell,n-}^{-1}(1-\zeta)\right) - \mathcal{F}_\ell\left(\mathcal{F}_{\ell-}^{-1}(1-\zeta)\right)\right| > \varepsilon\right) \leqslant P^*\left(\sup_{t\in\mathbb{R}} |F_{\ell,n-}(t) - \mathcal{F}_\ell(t)| > \varepsilon\right) \to 0$$

as $n \to \infty$ for all $\zeta \in [0, 1]$. Thus, by (2.31), the second summand converges to zero in outer probability as well. Hence, it follows

$$|\mathrm{FWER}_{n-}(\zeta) - \mathrm{FWER}(\zeta)| \xrightarrow{P} 0 \quad \text{as } n \to \infty \text{ for all } \zeta \in [0, 1].$$

Lemma 2.4 implies $\sup_{\zeta\in[a,b]} |\mathrm{FWER}_{n-}(\zeta) - \mathrm{FWER}(\zeta)| \xrightarrow{P} 0$ as $n \to \infty$ and, thus,

$$\begin{aligned}
|\mathrm{FWER}_{n+}(\zeta) - \mathrm{FWER}(\zeta)| \leqslant &\max\{\mathrm{FWER}_{n-}(\zeta + n^{-1}) - \mathrm{FWER}(\zeta + n^{-1}) + \mathrm{FWER}(\zeta + n^{-1}) - \mathrm{FWER}(\zeta), \\
&\mathrm{FWER}(\zeta) - \mathrm{FWER}_{n-}(\zeta)\} \\
\xrightarrow{P} 0&
\end{aligned}$$

as $n \to \infty$ for all $\zeta \in [a, b)$. Applying Lemma 2.5 yields (2.27) and, thus, $\beta_n \xrightarrow{P} \mathrm{FWER}^{-1}(\alpha)$ as $n \to \infty$.
In order to proof the statements (2.17)–(2.19), let $\mathcal{T} \subset \{1, ..., L\}$ be arbitrary and $\bigcap_{\ell\in\mathcal{T}} \mathcal{H}_{0,\ell}$ be true. Note that (2.30) implies

$$W_{\ell,n} > q_{\ell,n} \quad \Leftrightarrow \quad F_{\ell,n-}(W_{\ell,n}) > 1 - \beta_n \qquad \text{for all } \ell \in \{1, ..., L\}. \tag{2.32}$$

By $\beta_n \xrightarrow{P} \mathrm{FWER}^{-1}(\alpha)$ as $n \to \infty$, it follows

$$\left((W_{\ell,n})_{\ell\in\mathcal{T}}, \beta_n\right) \xrightarrow{d} \left((W_\ell)_{\ell\in\mathcal{T}}, \mathrm{FWER}^{-1}(\alpha)\right)$$

as $n \to \infty$ by (2.12) and Slutsky's lemma. Moreover, the continuous mapping theorem implies

$$\left((\mathcal{F}_\ell(W_{\ell,n}))_{\ell\in\mathcal{T}}, \beta_n\right) \xrightarrow{d} \left((\mathcal{F}_\ell(W_\ell))_{\ell\in\mathcal{T}}, \mathrm{FWER}^{-1}(\alpha)\right)$$

as $n \to \infty$ due to the continuity of $\mathcal{F}_1, ..., \mathcal{F}_L$. By (2.28), we have $|F_{\ell,n-}(W_{\ell,n}) - \mathcal{F}_\ell(W_{\ell,n})| \leqslant \sup_{t\in\mathbb{R}} |F_{\ell,n}(t) - \mathcal{F}_\ell(t)| \xrightarrow{P} 0$ for all $\ell \in \mathcal{T}$ and, thus, Slutsky's lemma implies

$$\left((F_{\ell,n-}(W_{\ell,n}))_{\ell\in\mathcal{T}}, \beta_n\right) \xrightarrow{d} \left((\mathcal{F}_\ell(W_\ell))_{\ell\in\mathcal{T}}, \mathrm{FWER}^{-1}(\alpha)\right)$$

as $n \to \infty$.
By approximating the function

$$f : \mathbb{R}^{|\mathcal{T}|} \times \mathbb{R} \ni ((t_\ell)_{\ell\in\mathcal{T}}, b) \mapsto \max_{\ell\in\mathcal{T}} \mathbb{1}\{t_\ell > 1 - b\} \in \{0, 1\}$$

by sequences of Lipschitz functions $(g_m)_{m\in\mathbb{N}}, (h_m)_{m\in\mathbb{N}}$ with $1 \geqslant g_m \geqslant f \geqslant h_m \geqslant 0$ and

$$\mathrm{E}\left[g_m\left((\mathcal{F}_\ell(W_\ell))_{\ell\in\mathcal{T}}, \mathrm{FWER}^{-1}(\alpha)\right) - h_m\left(\mathcal{F}_\ell(W_\ell))_{\ell\in\mathcal{T}}, \mathrm{FWER}^{-1}(\alpha)\right)\right] \leqslant m^{-1}$$

for all $m \in \mathbb{N}$, one can follow similarly as in the proof of Lemma 2.3 that

$$\begin{aligned}
P^*\left(\exists \ell \in \mathcal{T} :\ F_{\ell,n-}(W_{\ell,n}) > 1 - \beta_n\right) &= \mathrm{E}^*\left[f\left((F_{\ell,n-}(W_{\ell,n}))_{\ell\in\mathcal{T}}, \beta_n\right)\right] \\
&\to \mathrm{E}\left[f\left((\mathcal{F}_\ell(W_\ell))_{\ell\in\mathcal{T}}, \mathrm{FWER}^{-1}(\alpha)\right)\right] \\
&= P(\exists \ell \in \mathcal{T} :\ \mathcal{F}_\ell(W_\ell) > 1 - \mathrm{FWER}^{-1}(\alpha)) \\
&\leqslant 1 - P(\forall \ell \in \{1, \ldots, L\} :\ W_\ell \leqslant \mathcal{F}_{\ell-}^{-1}(1 - \mathrm{FWER}^{-1}(\alpha))) \\
&= \mathrm{FWER}(\mathrm{FWER}^{-1}(\alpha)) = \alpha
\end{aligned}$$

holds as $n \to \infty$ under $\bigcap\limits_{\ell\in\mathcal{T}} \mathcal{H}_{0,\ell}$, yielding (2.17) by (2.32). The inequality is an equality if $\mathcal{T} = \{1, ..., L\}$, yielding (2.18) by (2.32). Moreover, we observe that

$$P^*\left(F_{\ell,n-}(W_{\ell,n}) > 1 - \beta_n\right) \to P(W_\ell > \mathcal{F}_{\ell-}^{-1}(1 - \mathrm{FWER}^{-1}(\alpha))) = 1 - \mathcal{F}_\ell(\mathcal{F}_{\ell-}^{-1}(1 - \mathrm{FWER}^{-1}(\alpha))) = \mathrm{FWER}^{-1}(\alpha)$$

follows as $n \to \infty$ for $\mathcal{T} = \{\ell\}$ for some $\ell \in \{1, ..., L\}$, yielding (2.19) by (2.32). □

Remark 2.7. *It results from the proof that $\beta_n \xrightarrow{P} \mathrm{FWER}^{-1}(\alpha)$ as $n \to \infty$.*
Furthermore, by applying Lemma 2.5, one can show $q_{\ell,n} \xrightarrow{P} \mathcal{F}_\ell^{-1}(1-\mathrm{FWER}^{-1}(\alpha))$ as $n \to \infty$ for all $\ell \in \{1, ..., L\}$ if $\mathcal{F}_\ell$ is strictly increasing on $\left[\widetilde{a}, \widetilde{b}\right]$ with $\mathcal{F}_\ell(\widetilde{a}) < 1 - \alpha \leqslant 1 - \alpha/L < \mathcal{F}_\ell(\widetilde{b})$ for all $\ell \in \{1, ..., L\}$.

Proof of Lemma 2.2 Let $0 \leqslant x < y \leqslant 1$. We aim to show $\mathrm{FWER}(y) - \mathrm{FWER}(x) > 0$. We have

$$\begin{aligned}
&\mathrm{FWER}(y) - \mathrm{FWER}(x) \\
&= \mathcal{F}\left(\mathcal{F}_{1-}^{-1}(1-x), ..., \mathcal{F}_{L-}^{-1}(1-x)\right) - \mathcal{F}\left(\mathcal{F}_{1-}^{-1}(1-y), ..., \mathcal{F}_{L-}^{-1}(1-y)\right) \\
&= P\left(\exists \ell \in \{1, ..., L\} : W_\ell \in (\mathcal{F}_{\ell-}^{-1}(1-y), \mathcal{F}_{\ell-}^{-1}(1-x)], \forall \ell \in \{1, ..., L\} : W_\ell \leqslant \mathcal{F}_{\ell-}^{-1}(1-x)\right).
\end{aligned}$$

Let $b := \min_{\ell\in\{1,...,L\}} \frac{\mathcal{F}_{\ell-}^{-1}(1-x)}{||\mathbf{A}_\ell||}$, where here and throughout $||\mathbf{A}_\ell|| > 0$ denotes the spectral norm of $\mathbf{A}_\ell$, $\ell^* \in \arg\min_{\ell\in\{1,...,L\}} \frac{\mathcal{F}_{\ell-}^{-1}(1-x)}{||\mathbf{A}_\ell||}$, and $a := \frac{\mathcal{F}_{\ell^*\_}^{-1}(1-y)}{||\mathbf{A}_{\ell^*}||}$. Moreover, let $\mathbf{A}_{\ell^*} = \mathbf{U}\mathrm{diag}(\lambda_1, ..., \lambda_k)\mathbf{U}^\top$ denote the eigen-decomposition of $\mathbf{A}_{\ell^*}$ with an orthogonal matrix $\mathbf{U} = [\mathbf{u}_1, ..., \mathbf{u}_k]$, where $\lambda_1 \geqslant ... \geqslant \lambda_k \geqslant 0$ and $\lambda_1 = ||\mathbf{A}_{\ell^*}||$. Then, we have for all $\mathbf{z} \in \mathbb{R}^k$ that

$$||\mathbf{z}||^2 \leqslant b \text{ and } (\mathbf{u}_1^\top\mathbf{z})^2 > a \ \Rightarrow\ \mathbf{z}^\top\mathbf{A}_{\ell^*}\mathbf{z} \in (\mathcal{F}_{\ell^*\_}^{-1}(1-y), \mathcal{F}_{\ell^*\_}^{-1}(1-x)], \mathbf{z}^\top\mathbf{A}_\ell\mathbf{z} \leqslant \mathcal{F}_{\ell-}^{-1}(1-x), \ell \in \{1, ..., L\}.$$

In order to show this, note that $\mathbf{z}^\top\mathbf{A}_\ell\mathbf{z} \leqslant ||\mathbf{A}_\ell|| \cdot ||\mathbf{z}||^2 \leqslant b||\mathbf{A}_\ell|| \leqslant \mathcal{F}_{\ell-}^{-1}(1-x)$ holds for $||\mathbf{z}||^2 \leqslant b$ for all $\ell = 1, ..., L$. Moreover, for $(\mathbf{u}_1^\top\mathbf{z})^2 > a$, it holds

$$\mathbf{z}^\top\mathbf{A}_{\ell^*}\mathbf{z} = \sum_{i=1}^k \lambda_i(\mathbf{u}_i^\top\mathbf{z})^2 \geqslant ||\mathbf{A}_{\ell^*}||(\mathbf{u}_1^\top\mathbf{z})^2 > a||\mathbf{A}_{\ell^*}|| = \mathcal{F}_{\ell^*\_}^{-1}(1-y).$$

Thus, we get

$$\mathrm{FWER}(y) - \mathrm{FWER}(x) \geqslant P\left(||\mathbf{Z}||^2 \leqslant b, (\mathbf{u}_1^\top \mathbf{Z})^2 > a\right) = P\left(||\mathbf{U}^\top \mathbf{Z}||^2 \leqslant b, (\mathbf{u}_1^\top \mathbf{Z})^2 > a\right).$$

Now, it remains to show that $P\left(||\mathbf{U}^\top \mathbf{Z}||^2 \leqslant b, (\mathbf{u}_1^\top \mathbf{Z})^2 > a\right) > 0$. Integration by substitution yields

$$\begin{aligned} P\left(||\mathbf{U}^\top \mathbf{Z}||^2 \leqslant b, (\mathbf{u}_1^\top \mathbf{Z})^2 > a\right) &= \int_{\{\mathbf{z}\in\mathbb{R}^k \mid ||\mathbf{U}^\top \mathbf{z}||^2 \leqslant b, (\mathbf{u}_1^\top \mathbf{z})^2 > a\}} f(\mathbf{z})\, \mathrm{d}\mathbf{z} \\ &= \int_{\{\mathbf{y}=(y_1,\dots,y_k)^\top \in\mathbb{R}^k \mid ||\mathbf{y}||^2 \leqslant b, y_1^2 > a\}} f(\mathbf{U}\mathbf{y})\, \mathrm{d}\mathbf{y}, \end{aligned}$$

where $f : \mathbb{R}^k \to (0, \infty)$ denotes the Lebesgue density of $\mathbf{Z}$. Since the density is strictly positive, we only need to show that the Lebesgue measure of $\{\mathbf{y} = (y_1, \dots, y_k)^\top \in \mathbb{R}^k \mid ||\mathbf{y}||^2 \leqslant b, y_1^2 > a\}$ is strictly positive. Due to the continuity of $F_{\ell *}$, we have $a = \frac{\mathcal{F}^{-1}_{\ell * -}(1-y)}{||\mathbf{A}_{\ell *}||} < \frac{\mathcal{F}^{-1}_{\ell * -}(1-x)}{||\mathbf{A}_{\ell *}||} = b$. Let $\tilde{b}$ fulfill $a < \tilde{b} < b$. Then, we have $||\mathbf{y}||^2 \leqslant b, y_1^2 > a$ for all $\mathbf{y} = (y_1, \dots, y_k)^\top \in \mathbb{R}^k$ with $\sqrt{a} < |y_1| \leqslant \sqrt{\tilde{b}}, |y_2|, \dots, |y_k| \leqslant \sqrt{\frac{b-\tilde{b}}{k-1}}$. Hence, the Lebesgue measure of $\{\mathbf{y} = (y_1, \dots, y_k)^\top \in \mathbb{R}^k \mid ||\mathbf{y}||^2 \leqslant b, y_1^2 > a\}$ is at least the Lebesgue measure of

$$\left\{\mathbf{y} = (y_1, \dots, y_k)^\top \in \mathbb{R}^k \mid \sqrt{a} < |y_1| \leqslant \sqrt{\tilde{b}}, |y_2|, \dots, |y_k| \leqslant \sqrt{\frac{b-\tilde{b}}{k-1}}\right\},$$

which equals $2^k\left(\sqrt{\tilde{b}} - \sqrt{a}\right)\left(\sqrt{\frac{b-\tilde{b}}{k-1}}\right)^{k-1} > 0$. □

Proof of Proposition 2.1 For (1), let $\ell \in \{1, \dots, L\}$ be fixed. Firstly, we aim to show

$$p_\ell \leqslant \alpha \quad \Rightarrow \quad W_{\ell,n} > q_{\ell,n}.$$

Therefore, assume that $p_\ell \leqslant \alpha$ holds. Since $p_\ell \leqslant \alpha$ implies that $\beta_{\ell,n}$ satisfies $\mathrm{FWER}_{n+}(\beta_{\ell,n}) \leqslant \alpha$, it follows

$$1 - F_{\ell,n-}(W_{\ell,n}) = \beta_{\ell,n} \leqslant \beta_n - B_n^{-1} < \beta_n$$

by the definition of β_n. Thus, we have $W_{\ell,n} > F^{-1}_{\ell,n-}(1 - \beta_n) = q_{\ell,n}$ by (2.30).
Secondly, we aim to prove

$$W_{\ell,n} > q_{\ell,n} \quad \Rightarrow \quad p_\ell \leqslant \alpha.$$

The inequality $W_{\ell,n} > q_{\ell,n} = F^{-1}_{\ell,n-}(1 - \beta_n)$ implies

$$\beta_{\ell,n} = 1 - F_{\ell,n-}(W_{\ell,n}) \leqslant 1 - F_{\ell,n}(F^{-1}_{\ell,n-}(1 - \beta_n)) < \beta_n.$$

Due to $\beta_{\ell,n}, \beta_n \in \{0, \frac{1}{B_n}, \frac{2}{B_n}, \dots, 1\}$, it follows $\beta_{\ell,n} \leqslant \beta_n - B_n^{-1}$. Thus, the definition of β_n yields that $\beta_{\ell,n}$ fulfills

$$p_\ell = \mathrm{FWER}_{n+}(\beta_{\ell,n}) \leqslant \alpha.$$

For (2), we note that $p = \min\{p_1, \dots, p_L\} \leqslant \alpha$ if and only if there exists an $\ell \in \{1, \dots, L\}$ such that $p_\ell \leqslant \alpha$. Due to (1), this holds whenever there exists an $\ell \in \{1, \dots, L\}$ such that $W_{\ell,n} > q_{\ell,n}$ or, equivalently, $\max_{\ell \in \{1,\dots,L\}} W_{\ell,n}/q_{\ell,n} > 1$. □

3 Inference for Paired Survival Times

Throughout this section, we consider i.i.d. pairs of survival times $(T_{1j}, T_{2j}), j \in \{1, \dots, n\}$, i.e., T_{1j}, T_{2j} are non-negative random variables, with survival functions $S_i : [0, \infty) \ni t \mapsto P(T_{i1} > t) \in [0, 1], i \in \{1, 2\}$. Let $(T_{1j}, T_{2j}), j \in \{1, \dots, n\}$ model the times to the progression of a disease: first, after the initiation of one treatment, T_{1j} is measured; then, a second treatment phase begins, and T_{2j} is recorded. Here, we do not assume that the survival functions are continuous and, thus, we allow for ties in the data. In general, T_{1j} and T_{2j} are allowed to be correlated. Furthermore, we assume that the event times are subject to right-censoring, i.e., we only observe $(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j}), j \in \{1, \dots, n\}$, where $X_{1j} := \min\{T_{1j}, C_{1j}\}, X_{2j} := \min\{T_{2j}, C_{2j}\}$ are the censored event times and $\delta_{1j} := \mathbb{1}\{X_{1j} \leqslant C_{1j}\}, \delta_{2j} := \mathbb{1}\{X_{2j} \leqslant C_{2j}\}$ are the corresponding non-censoring indicators. Moreover, we assume that right-censoring is independent, i.e., (T_{1j}, T_{2j}) and (C_{1j}, C_{2j}) are stochastically independent, and the censoring times $(C_{1j}, C_{2j}), j \in \{1, ..., n\}$, are i.i.d. At times, we omit the index j when it is not necessary to distinguish the pairs.

Our aim is to propose methods that are based on relative and absolute estimands for quantifying the efficacy of an experimental treatment compared to a standard treatment: a variant of the probability as in von Hoff's method [75, 76] and functions of restricted mean survival times, respectively.

3.1 Variant of von Hoff's Method

Von Hoff's method [75, 76] is based on the probability that the ratio of T_2 and T_1 exceeds a preliminarily chosen threshold δ. The most common choice is $\delta = 1.3$ [76]. Instead of $P(T_2/T_1 > \delta)$ or $P(T_2/T_1 \geqslant \delta)$, we propose to focus on the estimand

$$P(T_2/T_1 > \delta) + \frac{1}{2} P(T_2/T_1 = \delta).$$

The second term, $P(T_2/T_1 = \delta)$, which has the weight 1/2, is important to take into account that the distribution of T_2/T_1 is allowed to have an atom at δ. Even in the case of continuous survival functions, this is possible, as can be seen from the perfectly correlated case $T_2 = \delta T_1$.

Due to the limited time horizon of studies, one typically cannot identify this probability. Instead, we consider the estimand

$$\begin{aligned}\theta :=& P(\min\{T_2, \tau_2\}/\min\{T_1, \tau_1\} > \delta) + \frac{1}{2} P(\min\{T_2, \tau_2\}/\min\{T_1, \tau_1\} = \delta) \\ =& P(\min\{T_2, \tau_2\} > \delta \cdot \min\{T_1, \tau_1\}) + \frac{1}{2} P(\min\{T_2, \tau_2\} = \delta \cdot \min\{T_1, \tau_1\})\end{aligned}$$

which is closely related to the estimand in [30]. Here, τ_1 and τ_2 denote the maximum follow-up times. The experimental treatment is then considered effective if that probability exceeds a certain probability $\theta_0 \in (0, 1)$, the choice of which might depend on the particular medical application. Thus, we aim to test the hypothesis

$$H_0^\theta : \theta \leqslant \theta_0 \quad \text{vs.} \quad H_1^\theta : \theta > \theta_0. \tag{3.1}$$

Since von Hoff's method [75, 76] ignores the censoring indicator, it yields a negatively biased estimator [49]. Thus, we propose an approach that takes the proper handling of right-censoring into account. This will lead to an approximately unbiased estimator of θ and, as a consequence, it is expected to improve the reliability and the power of the method.

Based on the competing risks-based approach in [30], θ can be estimated with the help of the Aalen-Johansen estimator [1]. To see the connection to the method developed in [30], define the pair of survival times $(\tilde{T}_1, \tilde{T}_2) = (\delta \cdot \min\{T_1, \tau_1\}, \min\{T_2, \tau_2\})$. The underlying competing risks data set can be written as

$$(Z_j, \varepsilon_j) = (\min\{\delta T_{1j}, T_{2j}, \tau, \delta C_{1j}, C_{2j}\}, \check{\varepsilon}_j \mathbb{1}\{\min\{\delta T_{1j}, T_{2j}, \tau\} \leqslant \min\{\delta C_{1j}, C_{2j}\}\}), \quad j \in \{1, ..., n\},$$

where $\tau := \min\{\delta \tau_1, \tau_2\}$ and $\check{\varepsilon}_j \in \{1, 2, 3\}$ denotes the event indicator; see Section 3.3 for details. Now, θ can be represented with the help of the cumulative incidence functions for type 2 and type 3 events, i.e., F_2 and F_3, as $\theta = F_2(\tau) + \frac{1}{2} F_3(\tau)$. Here, an event of type 1 is present if $\tilde{T}_{1j} > \tilde{T}_{2j}$ has been *observed*, an event of type 2 is present if $\tilde{T}_{1j} < \tilde{T}_{2j}$ has been *observed*, and a type-3 event is present if $\tilde{T}_{1j} = \tilde{T}_{2j}$ is *observed*. In other cases, the data point is censored from a competing risks point of view.

We define the number of individuals at risk just before time $t \geqslant 0$ by $Y(t) := \sum_{j=1}^n \mathbb{1}\{Z_j \geqslant t\}$ and the number of individuals with an event of type m before or at time $t \geqslant 0$ by $N_m(t) := \sum_{j=1}^n \mathbb{1}\{Z_j \leqslant t, \varepsilon_j = m\}$ for all $m \in \{1, 2, 3\}$. Moreover, we set

$$\widehat{A}_m(t) := \int_{[0,t]} \frac{1}{Y} \,\mathrm{d}N_m, \quad \widehat{A} := \sum_{m=1}^{3} \widehat{A}_m \quad \text{and} \quad \widehat{S}(t) := \prod_{x \in [0,t]} \left\{1 - \mathrm{d}\widehat{A}(x)\right\}$$

for all $t \geqslant 0, m \in \{1,2,3\}$, where here and throughout $\prod$ denotes the product integral as in [39]. These estimators are the cause-specific and all-cause Nelson–Aalen estimators and the Kaplan–Meier estimator, respectively. Thus, we obtain the Aalen–Johansen estimator [1] at t for $F_m(t)$ as $\widehat{F}_m(t) := \int_{[0,t]} \widehat{S}_- \, \mathrm{d}\widehat{A}_m, m \in \{1,2,3\}$ for all $t \geqslant 0$ and, hence,

$$\widehat{\theta} = \widehat{F}_2(\tau) + \frac{1}{2}\widehat{F}_3(\tau)$$

for θ. Note that one could allow either $\tau_1 = \infty$ or $\tau_2 = \infty$ as long as the respective other terminal time is finite. An adaptation of Theorems 1 and 2 in [30] justifies the asymptotic normality of the estimation approach under the following assumption.

Assumption 3.1. *We assume $P(\delta T_1 \geqslant \tau, T_2 \geqslant \tau) > 0$ and $P(\delta C_1 \geqslant \tau, C_2 \geqslant \tau) > 0$.*

Theorem 3.1. *Under Assumption 3.1, we have $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma_\theta^2)$ as $n \to \infty$, where σ_θ^2 is defined in Section 3.3.*

For technical reasons, we need $\sigma_\theta^2 > 0$. Therefore, we suppose the following.

Assumption 3.2. *We assume $\sigma_\theta^2 > 0$. Under Assumption 3.1, this is, e.g., the case if at least one of the following holds, which is shown in Lemma 3.2:*

(1) $P(T_2 < \min\{\delta T_1, \tau\}) > 0$ and $P(\tau \leqslant \min\{\delta T_1, \delta\tau_1\} \leqslant \min\{T_2, \tau_2\}) > 0$,

(2) $P(\delta T_1 < \min\{T_2, \tau\}) > 0$ and $P(\min\{\delta T_1, \delta\tau_1\} \geqslant \min\{T_2, \tau_2\} \geqslant \tau) > 0$,

(3) $P(\delta T_1 = T_2 < \tau) > 0$ and $P(\min\{\delta T_1, \delta\tau_1\} > \min\{T_2, \tau_2\} > u) \neq P(u < \min\{\delta T_1, \delta\tau_1\} < \min\{T_2, \tau_2\})$ for all $u \in [0, \tau)$.

This preliminary work and Slutsky's theorem imply the following result.

Theorem 3.2. *Under Assumptions 3.1 and 3.2, we have $\sqrt{n}(\widehat{\theta} - \theta)/\widehat{\sigma}_\theta \xrightarrow{d} \mathcal{N}(0,1)$ as $n \to \infty$. The definition of $\widehat{\sigma}_\theta^2$ is given in (3.4) below.*

With this theorem, we can construct an asymptotic level-α test for (3.1), that is,

$$\varphi^\theta := \mathbb{1}\left\{\sqrt{n}(\widehat{\theta} - \theta_0)/\widehat{\sigma}_\theta > z_{1-\alpha}\right\},$$

where here and throughout $z_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution.
Instead of the standard normal quantile, it is typically beneficial to use a resampling-based quantile. In particular, we propose a randomization approach, i.e., the observable event indicator ε_j is randomly re-labeled as 1 or 2 with probability 1/2, respectively, whenever an event of type 1 or 2 occurred; cf. [30] for a similar approach. This is equivalent to randomly permuting the paired (censored) event times (X_{1j}, δ_{1j}) and (X_{2j}, δ_{2j}) within each pair $j \in \{1, ..., n\}$. This results in the randomized data set $(Z_j, \widetilde{\varepsilon}_j), j \in \{1, ..., n\}$, and corresponding randomized estimators $\widetilde{\widehat{\theta}}, \widetilde{\widehat{\sigma}}_\theta^2$ based on our randomized sample $(Z_j, \widetilde{\varepsilon}_j), j \in \{1, ..., n\}$.
Analogously to the proof of Theorem 2 in the supplement of [30], we obtain that $\sqrt{n}(\widetilde{\widehat{\theta}} - 1/2) \xrightarrow{d^*} \mathcal{N}(0, \widetilde{\sigma}_\theta^2)$ conditionally on the data $(Z_j, \varepsilon_j), j \in \{1, ..., n\}$, in outer probability as $n \to \infty$, where $\widetilde{\sigma}_\theta^2$ is given in Section 3.3. Again, we need to assume a positive variance of the limit.

Assumption 3.3. *We assume $\widetilde{\sigma}_\theta^2 > 0$. Under Assumption 3.1, this is, e.g., the case if $P(T_2 < \min\{\delta T_1, \tau\}) > 0$ or $P(\delta T_1 < \min\{T_2, \tau\}) > 0$ holds, which is shown in Lemma 3.3.*

Then, we obtain the following result.

Theorem 3.3. *Under Assumptions 3.1 and 3.3, we have $\sqrt{n}(\widetilde{\widehat{\theta}} - 1/2)/\widetilde{\widehat{\sigma}}_\theta \xrightarrow{d^*} \mathcal{N}(0,1)$ conditionally on the data $(Z_j, \varepsilon_j), j \in \{1, ..., n\}$, in outer probability as $n \to \infty$.*

Theorem 3.3 provides that the randomization test

$$\widetilde{\varphi}^\theta := \mathbb{1}\left\{\sqrt{n}(\widehat{\theta} - \theta_0)/\widehat{\sigma}_\theta > \widetilde{z}_{1-\alpha}\right\}$$

is an asymptotic level α test, where $\widetilde{z}_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the conditional distribution of $\sqrt{n}(\widetilde{\widehat{\theta}} - 1/2)/\widetilde{\widehat{\sigma}}_\theta$ given the data $(Z_j, \varepsilon_j), j \in \{1, ..., n\}$. In practice, the quantile $\widetilde{z}_{1-\alpha}$ can be approximated by a Monte Carlo method, cf. Section 2.3.2.

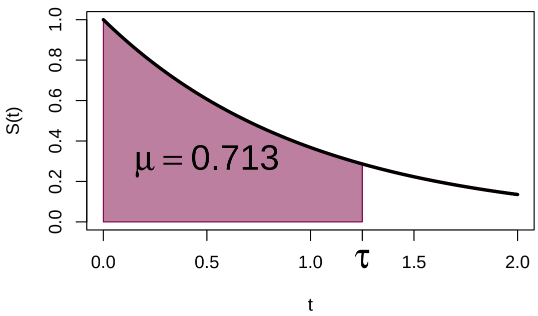


Figure 3: An exemplary illustration of the RMST μ.

3.2 Restricted Mean Survival Times

An alternative approach for comparing paired survival times is the comparison of the restricted mean survival times (RMSTs) of the two event times. The RMST is defined as the area under the survival curve up to a prespecified time point $\tau > 0$ as illustrated in Figure 3 and it has an intuitive interpretation as the expected minimum of the survival time and the specified time point. Thus, the RMST reduces the whole survival curve to a meaningful estimand. In detail, the RMSTs are

$$\mu_i := \int_0^\tau S_i(t) \,\mathrm{d}t \in [0,\tau], \quad i \in \{1,2\}.$$

For comparing two RMSTs, we can consider the hypotheses

$$H_0^{\text{diff}} : \mu_1 - \mu_2 \geqslant \xi \quad \text{vs.} \quad H_1^{\text{diff}} : \mu_1 - \mu_2 < \xi$$

for the difference of the RMSTs with $\xi \in [-\tau, \tau]$, or

$$H_0^{\text{rat}} : \frac{\mu_1}{\mu_2} \geqslant 1 + \zeta \quad \text{vs.} \quad H_1^{\text{rat}} : \frac{\mu_1}{\mu_2} < 1 + \zeta$$

for the ratio of RMSTs with $\zeta \in (-1, \infty)$. A natural estimator for μ_i is given by

$$\widehat{\mu}_i := \int_0^\tau \widehat{S}_i(t) \,\mathrm{d}t$$

for $i \in \{1,2\}$, where $\widehat{S}_i$ denotes the Kaplan-Meier estimator of S_i. Hence, we get the estimator $\widehat{\mu}_1 - \widehat{\mu}_2$ for $\mu_1 - \mu_2$ and $\widehat{\mu}_1/\widehat{\mu}_2$ for μ_1/μ_2.
For technical reasons, we need the following assumptions.

Assumption 3.4. *Throughout this section, we assume*

(1) $P(C_1 \geqslant \tau), P(C_2 \geqslant \tau) > 0$, *and*

(2) $P(T_1 \geqslant \tau), P(T_2 \geqslant \tau) > 0$.

Under the stated assumptions, the estimators can be shown to be asymptotically normal, that is

$$\sqrt{n}((\widehat{\mu}_1 - \widehat{\mu}_2) - (\mu_1 - \mu_2)) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\text{diff}})$$
$$\text{and} \quad \sqrt{n}(\log(\widehat{\mu}_1/\widehat{\mu}_2) - \log(\mu_1/\mu_2)) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\text{rat}})$$

as $n \to \infty$ for some $\sigma^2_{\text{diff}}, \sigma^2_{\text{rat}} \geqslant 0$; see Section 3.3 for details.

Theorem 3.4. *Under Assumptions 3.4 and* $\sigma^2_{\text{diff}}, \sigma^2_{\text{rat}} > 0$, *we have*

$$\sqrt{n}((\widehat{\mu}_1 - \widehat{\mu}_2) - (\mu_1 - \mu_2))/\widehat{\sigma}_{\text{diff}} \xrightarrow{d} \mathcal{N}(0,1)$$

and

$$\sqrt{n}(\log(\widehat{\mu}_1/\widehat{\mu}_2) - \log(\mu_1/\mu_2))/\widehat{\sigma}_{\text{rat}} \xrightarrow{d} \mathcal{N}(0,1)$$

as $n \to \infty$. *The definitions of the variance estimators* $\widehat{\sigma}^2_{\text{diff}}$ *and* $\widehat{\sigma}^2_{\text{rat}}$ *are given in Section 3.3.*

This theorem yields that the tests

$$\varphi^{\text{diff}} := \mathbb{1}\left\{\sqrt{n}((\widehat{\mu}_1 - \widehat{\mu}_2) - \xi)/\widehat{\sigma}_{\text{diff}} < z_\alpha\right\}$$
$$\text{and} \quad \varphi^{\text{rat}} := \mathbb{1}\left\{\sqrt{n}(\log(\widehat{\mu}_1/\widehat{\mu}_2) - \log(1+\zeta))/\widehat{\sigma}_{\text{rat}} < z_\alpha\right\}$$

are asymptotic level α tests for H_0^{diff} and H_0^{rat}, respectively.
The randomization approach described in Section 3.1 can be adopted to construct a randomization test. To this end, let $(X_{1j}^\pi, \delta_{1j}^\pi)$ and $(X_{2j}^\pi, \delta_{2j}^\pi)$ denote the permuted (censored) event times of the paired (censored) event times (X_{1j}, δ_{1j}) and (X_{2j}, δ_{2j}) within each pair $j \in \{1, ..., n\}$. Furthermore, we denote all estimators based on the permuted (censored) event times $(X_{1j}^\pi, \delta_{1j}^\pi), (X_{2j}^\pi, \delta_{2j}^\pi), j \in \{1, ..., n\}$, with a π in the superscript in the following. E.g., $\widehat{\mu}_i^\pi$ denotes the RMST estimator based on the permuted (censored) event times $(X_{ij}^\pi, \delta_{ij}^\pi), j \in \{1, ..., n\}$, for $i \in \{1, 2\}$. The following theorem yields the consistency of this randomization approach.

Theorem 3.5. *Under Assumption 3.4 and $\sigma_{\text{diff}}^\pi, \sigma_{\text{rat}}^\pi > 0$, we have, as $n \to \infty$,*

$$\sqrt{n}(\widehat{\mu}_1^\pi - \widehat{\mu}_2^\pi)/\widehat{\sigma}_{\text{diff}}^\pi \xrightarrow{d^*} \mathcal{N}(0,1)$$

and

$$\sqrt{n}\log(\widehat{\mu}_1^\pi/\widehat{\mu}_2^\pi)/\widehat{\sigma}_{\text{rat}}^\pi \xrightarrow{d^*} \mathcal{N}(0,1)$$

conditionally on the data $(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j}), j \in \{1, ..., n\}$ in outer probability, where $\sigma_{\text{diff}}^\pi, \sigma_{\text{rat}}^\pi$ are defined in Section 3.3.

Hence, the validity of the randomization tests

$$\varphi^{\pi,\text{diff}} := \mathbb{1}\left\{\sqrt{n}((\widehat{\mu}_1 - \widehat{\mu}_2) - \xi)/\widehat{\sigma}_{\text{diff}} < z_\alpha^{\pi,\text{diff}}\right\}$$
$$\text{and} \quad \varphi^{\pi,\text{rat}} := \mathbb{1}\left\{\sqrt{n}(\log(\widehat{\mu}_1/\widehat{\mu}_2) - \log(1+\zeta))/\widehat{\sigma}_{\text{rat}} < z_\alpha^{\pi,\text{rat}}\right\}$$

is provided, where $z_\alpha^{\pi,\text{diff}}$ and $z_\alpha^{\pi,\text{rat}}$ denote the α-quantiles of the conditional distributions of $\sqrt{n}(\widehat{\mu}_1^\pi - \widehat{\mu}_2^\pi)/\widehat{\sigma}_{\text{diff}}^\pi$ and $\sqrt{n}\log(\widehat{\mu}_1^\pi/\widehat{\mu}_2^\pi)/\widehat{\sigma}_{\text{rat}}^\pi$, respectively, given the data $(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j}), j \in \{1, ..., n\}$. By Section 2.3.2, the quantiles can also be approximated by a Monte Carlo method.

3.3 Proofs of Section 3

Remark 3.1. *Instead of assuming that (C_1, C_2) is stochastically independent of (T_1, T_2), we also may assume that the first survival time T_1 is always uncensored and C_2 is independent of (T_1, T_2). This case is a special case of the independent censoring case since we can set $C_1 = \tau_1 + 1$.*

Proof of Theorem 3.1 As in [30], let

$$\check{\varepsilon} := \begin{cases} 1 & \text{if } \min\{\delta T_1, \delta\tau_1\} > \min\{T_2, \tau_2\} \\ 2 & \text{if } \min\{\delta T_1, \delta\tau_1\} < \min\{T_2, \tau_2\} \\ 3 & \text{if } \min\{\delta T_1, \delta\tau_1\} = \min\{T_2, \tau_2\} \end{cases}$$

denote the (uncensored) event indicator, $\check{T} := \min\{\delta T_1, T_2, \tau\}$ and $\check{C} := \min\{\delta C_1, C_2\}$. Thus, we can write the censored competing risks data set as

$$(Z_j, \varepsilon_j) = (\min\{\check{T}_j, \check{C}_j\}, \check{\varepsilon}_j \mathbb{1}\{\check{T}_j \leqslant \check{C}_j\}), \quad j \in \{1, ..., n\}.$$

Furthermore, let $F_m(t) := P(\check{T} \leqslant t, \check{\varepsilon} = m), S(t) := P(\check{T} > t)$, $A_m(t) := \int_{[0,t]} \frac{1}{S_-(u)} \, \mathrm{d}F_m(u)$, $A(t) := A_1(t) + A_2(t) + A_3(t)$ and $G(t) := P(\check{C} > t)$ for all $t \geqslant 0, m \in \{1, 2, 3\}$. Note that

$$F_2(\tau) = P(\delta \min\{T_1, \tau_1\} < \min\{T_2, \tau_2\})$$
$$\text{and} \quad F_3(\tau) = P(\delta \min\{T_1, \tau_1\} = \min\{T_2, \tau_2\}).$$

Firstly, we emphasize that the Aalen–Johansen estimator $\widehat{A}_m$ consistently estimates the cause-specific cumulative hazard function A_m, i.e., no relevant information is lost by the above-described competing risks data. Regarding Theorem 4.2 in [26], we need to show the following statement.

Lemma 3.1. *It holds $A_m(t) = \int_{[0,t]} \frac{1}{P(Z_1 \geqslant .)} \, \mathrm{d}P(Z_1 \leqslant ., \varepsilon_1 = m)$ for all $m \in \{1, 2, 3\}, t \geqslant 0$ with $G_-(t) > 0$.*

Proof of Lemma 3.1. Let $m \in \{1,2,3\}, t \geqslant 0$ be arbitrary with $G_-(t) > 0$. Due to the definition of Z_1, we have

$$P(Z_1 \geqslant u) = P(\check{T}_1 \geqslant u, \check{C}_1 \geqslant u) = P(\check{T}_1 \geqslant u)P(\check{C}_1 \geqslant u) = S_-(t)G_-(t)$$

for all $u \in [0,t]$. Moreover,

$$P(Z_1 \leqslant u, \varepsilon_1 = m) = P(\check{T}_1 \leqslant u, \check{\varepsilon}_1 = m, \check{C}_1 \geqslant \check{T}_1) = \int_{[0,u]} G_- \,\mathrm{d}F_m$$

for all $u \in [0,t]$. Hence, it follows

$$\int_{[0,t]} \frac{1}{P(Z_1 \geqslant .)} \,\mathrm{d}P(Z_1 \leqslant ., \varepsilon_1 = m) = \int_{[0,t]} \frac{G_-}{S_- G_-} \,\mathrm{d}F_m = \int_{[0,t]} \frac{1}{S_-} \,\mathrm{d}F_m = A_m(t).$$

□

To analyze the asymptotic behavior of $\widehat{\theta}$, we do not use the results of [30] for two reasons: First, we suppose weaker assumptions on the survival and censoring distributions and, second, the variance formulas given in [30] and in the earlier version of the linked GitHub repository (`https://github.com/dennis-dobler/relative_treatment_effect_paired_survival/tree/30fa79a`) are both wrongly stated, as the following examples show.
In the GitHub repository, there is just a bracket missing, such that the second and third stated summand of the final result should be multiplied with 1/4 as well. Then, the formula is correct under $\Delta F_1(\tau) = \Delta F_2(\tau) = 0$, which follows from the assumptions in [30]. Here and throughout, $\Delta F(u) := F(u) - F_-(u)$ denotes the increment of F at u. However, since we aim to allow mass in τ for all event types, we state a different variance formula later. Our formula coincides to the formula in the GitHub repository with added bracket under $\Delta F_1(\tau) = \Delta F_2(\tau) = 0$.

Example 3.1. *Let*

$$F_1(t) := 0, \quad F_2(t) := \begin{cases} 0 & \text{if } t < 1 \\ 1/2 & \text{if } t \geqslant 1 \end{cases}, \quad \text{and } F_3(t) := \begin{cases} 0 & \text{if } t < 2 \\ 1/2 & \text{if } t \geqslant 2 \end{cases}$$

for all $t \geqslant 0$ with $\tau = 2$ and $C_1, C_2 \geqslant 2$ almost surely. Then, one can show $\widehat{\theta} = \frac{1}{n}\sum_{i=1}^n \left(\mathbb{1}\{\check{\varepsilon}_i = 2\} + \frac{1}{2}\mathbb{1}\{\check{\varepsilon}_i = 3\}\right)$ and, thus, by the central limit theorem, $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, 1/16)$. However, the formula in the supplement of [30] yields

$$\begin{aligned} \sigma_\theta^2 &= S_-(1)S_-(1)\sigma_2^2(1)\Delta A_2(1)\Delta A_2(1) + S_-(1)S_-(2)\sigma_2^2(1)\Delta A_2(1)\Delta A_3(2) \\ &\quad + S_-(2)S_-(2)\left(\frac{\Delta\sigma_2^2(1)}{1-\Delta A_2(1)}\left(\frac{1}{2}\Delta A_3(2)\right)^2 - 2\frac{\Delta\sigma_2^2(1)}{1-\Delta A_2(1)}\left(\frac{1}{2}\Delta A_3(2)\right)\Delta A_3(2) + \frac{1}{4}\sigma_3^2(2)\Delta A_3(2)\Delta A_3(2)\right) \\ &= \frac{1}{4}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{4}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2}\cdot\left(\frac{1/4}{1-1/2}\cdot\left(\frac{1}{2}\right)^2 - 2\frac{1/4}{1-1/2}\cdot\frac{1}{2} + 0\right) = 0 \end{aligned}$$

if the integrals $\int_0^\tau$ are meant as $\int_{[0,\tau]}$. The formula of the GitHub repository yields

$$\sigma_\theta^2 = \frac{(F_2(1) - F_{2,-}(2) - F_1(1) + F_{1,-}(2) + S(1))^2}{(1-\Delta A_2(1))^2}\Delta\sigma_2^2(1) = \frac{(1/2 - 1/2 + 1/2)^2}{(1-1/2)^2}\cdot\frac{1}{4} = \frac{1}{4}.$$

Note that this is not correct due to the missing factor 1/4.

Also if the integrals $\int_0^\tau$ are meant as $\int_{[0,\tau)}$, the formula in the supplement of [30] is not correct, as the following example shows.

Example 3.2. *Let*

$$F_1(t) := \begin{cases} 0 & \text{if } t < 1 \\ 1/2 & \text{if } t \geqslant 1 \end{cases}, \quad F_2(t) := 0, \quad \text{and } F_3(t) := \begin{cases} 0 & \text{if } t < 2 \\ 1/2 & \text{if } t \geqslant 2 \end{cases}$$

for all $t \geqslant 0$ with $\tau = 2$ and $C_1, C_2 \geqslant 2$ almost surely. Then, one can show $\widehat{\theta} = \frac{1}{2}\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{\check{\varepsilon}_i = 3\}$ and, thus, by the central limit theorem, $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, 1/16)$. However, the formula in the supplement of [30] yields $\sigma_\theta^2 = 0$ if the integrals $\int_0^\tau$ are meant as $\int_{[0,\tau)}$.

Now, to prove Theorem 3.1, note that θ and $\widehat{\theta}$ can be written as

$$\tilde{\psi}\left(\tilde{\phi}\left(-(A_1+A_2+A_3)\right)_-, A_2+\frac{1}{2}A_3\right)(\tau)=\int_{[0,\tau]} S_-\,\mathrm{d}\left(A_2+\frac{1}{2}A_3\right)=F_2(\tau)+\frac{1}{2}F_3(\tau)=\theta$$

and

$$\tilde{\psi}\left(\tilde{\phi}\left(-(\widehat{A}_1+\widehat{A}_2+\widehat{A}_3)\right)_-, \widehat{A}_2+\frac{1}{2}\widehat{A}_3\right)(\tau)=\int_{[0,\tau]} \widehat{S}_-\,\mathrm{d}\left(\widehat{A}_2+\frac{1}{2}\widehat{A}_3\right)=\widehat{F}_2(\tau)+\frac{1}{2}\widehat{F}_3(\tau)=\widehat{\theta}$$

with $\tilde{\psi}:\tilde{D}[0,\tau]\times BV_M[0,\tau]\to D[0,\tau], \tilde{\phi}:BV_{3M}[0,\tau]\to D[0,\tau)$ as in Section A for some $M<\infty$. Hence, we define

$$\Psi:(BV_M[0,\tau])^3\to\mathbb{R},\quad \Psi(\Lambda_1,\Lambda_2,\Lambda_3):=\tilde{\psi}\left(\tilde{\phi}\left(-(\Lambda_1+\Lambda_2+\Lambda_3)\right)_-, \Lambda_2+\frac{1}{2}\Lambda_3\right)(\tau).$$

To apply the delta-method, we show the Hadamard differentiability of Ψ at (A_1,A_2,A_3) by the chain rule. Note that Assumption 3.1 is equivalent to $S_-(\tau)>0$ and $P(Z_1\geqslant\tau)>0$. Analogously to Lemma 3.10.18 and Lemma 3.10.32 in [74], we obtain the Hadamard-derivatives

$$\tilde{\psi}'_{(\tilde{\phi}(-(A_1+A_2+A_3))_-,A_2+\frac{1}{2}A_3)}(\alpha,\beta)=\int_{[0,.]}\tilde{\phi}(-(A_1+A_2+A_3))_-\,\mathrm{d}\beta+\int_{[0,.]}\alpha\,\mathrm{d}\left(A_2+\frac{1}{2}A_3\right)$$

$$\text{and}\qquad \tilde{\phi}'_{-(A_1+A_2+A_3)}(\beta)=\tilde{\phi}(-(A_1+A_2+A_3))(.)\int_{[0,.]}\frac{1}{1-\Delta(A_1+A_2+A_3)}\,\mathrm{d}\beta$$

for all $\alpha\in\tilde{D}[0,\tau],\beta\in D[0,\tau]$ under Assumption 3.1. Here, we consider $\tilde{\phi}:BV_M[0,\tau)\to D[0,\tau)$ as function mapping to $D[0,\tau)$ instead of $D[0,\tau]$ to guarantee that the weaker assumption $S_-(\tau)>0$ instead of $S(\tau)>0$ suffices, cf. Section A.2. Moreover, $(BV_M[0,\tau])^3\ni(\Lambda_1,\Lambda_2,\Lambda_3)\mapsto\Lambda_1+\Lambda_2+\Lambda_3$, $(BV_M[0,\tau])^3\ni(\Lambda_1,\Lambda_2,\Lambda_3)\mapsto\Lambda_2+\frac{1}{2}\Lambda_3$, $D[0,\tau)\ni\Lambda\mapsto\Lambda_-\in\tilde{D}[0,\tau]$ and $D[0,\tau]\ni\Lambda\mapsto\Lambda(\tau)\in\mathbb{R}$ are linear and, thus, their Hadamard-derivatives equals the functionals, respectively. Hence, the chain rule implies that Ψ is Hadamard differentiable at (A_1,A_2,A_3) with Hadamard-derivative

$$\begin{aligned}\Psi'_{(A_1,A_2,A_3)}(\alpha_1,\alpha_2,\alpha_3)=&\tilde{\psi}'_{(\tilde{\phi}(-(A_1+A_2+A_3))_-,A_2+\frac{1}{2}A_3)}\left(\tilde{\phi}'_{-(A_1+A_2+A_3)}\left(-(\alpha_1+\alpha_2+\alpha_3)\right)_-,\alpha_2+\frac{1}{2}\alpha_3\right)(\tau)\\
=&\int_{[0,\tau]}\tilde{\phi}\left(-(A_1,A_2,A_3)\right)_-\,\mathrm{d}\left(\alpha_2+\frac{1}{2}\alpha_3\right)\\
&-\int_{[0,\tau)}\frac{\int_{(u,\tau]}\tilde{\phi}\left(-(A_1,A_2,A_3)\right)_-\,\mathrm{d}\left(A_2+\frac{1}{2}A_3\right)}{1-\Delta(A_1+A_2+A_3)(u)}\,\mathrm{d}(\alpha_1+\alpha_2+\alpha_3)(u)\\
=&\int_{[0,\tau]}S_-\,\mathrm{d}\left(\alpha_2+\frac{1}{2}\alpha_3\right)-\int_{[0,\tau]}\frac{\int_{(u,\tau]}S_-\,\mathrm{d}\left(A_2+\frac{1}{2}A_3\right)}{1-\Delta A(u)}\,\mathrm{d}(\alpha_1+\alpha_2+\alpha_3)(u)\end{aligned}$$

for all $\alpha_1,\alpha_2,\alpha_3\in D[0,\tau]$ by the chain rule, where we set $0/0:=0$.
Furthermore, Theorem 4.1 in [26] provides that

$$\sqrt{n}\left(\widehat{A}_1-A_1,\widehat{A}_2-A_2,\widehat{A}_3-A_3\right)\xrightarrow{d}(U_1,U_2,U_3)\tag{3.2}$$

holds as $n\to\infty$ on $D^3[0,\tau]$, where U_1,U_2,U_3 are zero-mean Gaussian-martingales with

$$\begin{aligned}\mathbb{C}ov(U_m(t),U_m(s))&=\int_{[0,\min\{t,s\}]}\frac{1-\Delta A_m}{y}\,\mathrm{d}A_m=:\sigma_m^2(\min\{t,s\}),\\
\mathbb{C}ov(U_m(t),U_\ell(s))&=-\int_{[0,\min\{t,s\}]}\frac{\Delta A_\ell}{y}\,\mathrm{d}A_m=:\sigma_{m\ell}(\min\{t,s\})\end{aligned}$$

with $y(t):=S_-(t)G_-(t)$ for all $t,s\in[0,\tau],m,\ell\in\{1,2,3\},m\neq\ell$. By Section B, the limit variable is separable. Thus, the delta-method (Theorem 3.10.4 in [74]) implies $\sqrt{n}(\widehat{\theta}-\theta)\xrightarrow{d}\Psi'_{(A_1,A_2,A_3)}(U_1,U_2,U_3)$ as $n\to\infty$, where $\Psi'_{(A_1,A_2,A_3)}(U_1,U_2,U_3)$ follows a centered normal distribution. The variance of $\Psi'_{(A_1,A_2,A_3)}(U_1,U_2,U_3)$ can be calculated as

$$
\begin{aligned}
\sigma_\theta^2 &:= \mathbb{V}ar\left(\int_{[0,\tau]} S_- \,\mathrm{d}\left(U_2 + \frac{1}{2}U_3\right) - \int_{[0,\tau]} \frac{\int_{(u,\tau]} S_- \,\mathrm{d}\left(A_2 + \frac{1}{2}A_3\right)}{1-\Delta A(u)} \,\mathrm{d}(U_1+U_2+U_3)(u)\right) \\
&= \int_{[0,\tau]} S_-^2 \,\mathrm{d}\left(\sigma_2^2 + \sigma_{23} + \frac{1}{4}\sigma_3^2\right) \\
&\quad - 2\int_{[0,\tau]} \frac{S_-(u)}{1-\Delta A(u)} \int_{(u,\tau]} S_- \,\mathrm{d}\left(A_2 + \frac{1}{2}A_3\right) \,\mathrm{d}\left(\sigma_{12} + \sigma_2^2 + \frac{3}{2}\sigma_{23} + \frac{1}{2}\sigma_{13} + \frac{1}{2}\sigma_3^2\right)(u) \\
&\quad + \int_{[0,\tau]} \frac{\left(\int_{(u,\tau]} S_- \,\mathrm{d}\left(A_2 + \frac{1}{2}A_3\right)\right)^2}{(1-\Delta A(u))^2} \,\mathrm{d}\sigma_\bullet^2(u) \\
&= \int_{[0,\tau]} S_-^2 \,\mathrm{d}\left(\sigma_2^2 + \sigma_{23} + \frac{1}{4}\sigma_3^2\right) \\
&\quad - 2\int_{[0,\tau]}\int_{[0,v)} \frac{S_-(u)S_-(v)}{1-\Delta A(u)} \,\mathrm{d}\left(\sigma_{12} + \sigma_2^2 + \frac{3}{2}\sigma_{23} + \frac{1}{2}\sigma_{13} + \frac{1}{2}\sigma_3^2\right)(u)\mathrm{d}\left(A_2 + \frac{1}{2}A_3\right)(v) \\
&\quad + \int_{[0,\tau]}\int_{[0,\tau]} S_-(u)S_-(v) \int_{[0,\min\{u,v\})} \frac{1}{(1-\Delta A(w))^2} \,\mathrm{d}\sigma_\bullet^2(w)\mathrm{d}\left(A_2 + \frac{1}{2}A_3\right)(u)\mathrm{d}\left(A_2 + \frac{1}{2}A_3\right)(v)
\end{aligned}
$$

with $\sigma_\bullet^2 := \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2\sigma_{12} + 2\sigma_{13} + 2\sigma_{23}$. □

Proof of Theorem 3.2

Lemma 3.2. *Under at least one of (1)–(3) in Assumption 3.2, we have* $\sigma_\theta^2 > 0$.

Proof of Lemma 3.2. By the proof of Theorem 3.1, it holds that

$$\sigma_\theta^2 = \mathbb{V}ar\left(\sum_{m=1}^{3} \int\limits_{[0,\tau]} h_m \,\mathrm{d}U_m\right)$$

with

$$
\begin{aligned}
h_1(u) &:= \frac{\int_{(u,\tau]} S_- \,\mathrm{d}\left(A_2 + \frac{1}{2}A_3\right)}{1-\Delta A(u)}, \\
h_2(u) &:= S_-(u) - \frac{\int_{(u,\tau]} S_- \,\mathrm{d}\left(A_2 + \frac{1}{2}A_3\right)}{1-\Delta A(u)} \\
\text{and} \quad h_3(u) &:= \frac{S_-(u)}{2} - \frac{\int_{(u,\tau]} S_- \,\mathrm{d}\left(A_2 + \frac{1}{2}A_3\right)}{1-\Delta A(u)}
\end{aligned}
$$

for all $u \in [0,\tau]$, where $0/0 := 0$. We can calculate this variance further as

$$
\begin{aligned}
\sigma_\theta^2 &= \sum_{m=1}^{3} \mathrm{E}\left(\left(\int\limits_{[0,\tau]} h_m \,\mathrm{d}U_m\right)^2\right) + \sum_{m=1}^{3}\sum_{\tilde{m}\neq m} \mathrm{E}\left(\int\limits_{[0,\tau]} h_m \,\mathrm{d}U_m \int\limits_{[0,\tau]} h_{\tilde{m}} \,\mathrm{d}U_{\tilde{m}}\right) \\
&= \sum_{m=1}^{3} \int\limits_{[0,\tau]} h_m^2 \frac{1-\Delta A_m}{y} \,\mathrm{d}A_m - \sum_{m=1}^{3}\sum_{\tilde{m}\neq m} \int\limits_{[0,\tau]} h_m h_{\tilde{m}} \frac{\Delta A_m}{y} \,\mathrm{d}A_{\tilde{m}} \\
&= \sum_{m=1}^{3} \int\limits_{[0,\tau]} \frac{h_m^2}{y} \,\mathrm{d}A_m - \sum_{m=1}^{3}\sum_{\tilde{m}=1}^{3} \int\limits_{[0,\tau]} h_m h_{\tilde{m}} \frac{\Delta A_m}{y} \,\mathrm{d}A_{\tilde{m}} \\
&= \sum_{m=1}^{3} \int\limits_{[0,\tau]} \frac{h_m^2}{y} \,\mathrm{d}A_m^c + \sum_{x\in\mathcal{D}} \frac{\sum_{m=1}^3 h_m^2(x)\Delta A_m(x) - \left(\sum_{m=1}^3 h_m(x)\Delta A_m(x)\right)^2}{y(x)}
\end{aligned}
\tag{3.3}
$$

where $\mathcal{D} = \{x \in [0,\tau] : \Delta A(x) > 0\}$ is the set of discontinuity time points and

$$A_m^c(x) := A_m(x) - \sum_{y\leqslant x, y\in\mathcal{D}} \Delta A_m(y), m \in \{1,2,3\},$$

denotes the continuous part of A_m at $x \in [0, \tau]$. The Cauchy-Schwarz inequality yields

$$\left(\sum_{m=1}^{3} h_m(x)\Delta A_m(x)\right)^2 \leqslant \left(\sum_{m=1}^{3} h_m^2(x)\Delta A_m(x)\right)\left(\sum_{m=1}^{3} \Delta A_m(x)\right)$$

and, thus,

$$\sum_{m=1}^{3} h_m^2(x)\Delta A_m(x) - \left(\sum_{m=1}^{3} h_m(x)\Delta A_m(x)\right)^2 \geqslant \sum_{m=1}^{3} h_m^2(x)\Delta A_m(x)\left(1 - \Delta A(x)\right) \geqslant 0$$

for all $x \in \mathcal{D}$, where $1 - \Delta A(u) \geqslant S_-(\tau) = P(\delta T_1 \geqslant \tau, T_2 \geqslant \tau) > 0$ due to Assumption 3.2.
Under (1), we have $F_{1,-}(\tau) = P(T_2 < \min\{\delta T_1, \tau\}) > 0$ and

$$h_1(u) \geqslant \frac{S_-(\tau)\Delta(A_2 + \frac{1}{2}A_3)(\tau)}{1 - \Delta A(u)} \geqslant \frac{\frac{1}{2}\Delta(F_2 + F_3)(\tau)}{1 - \Delta A(u)} = \frac{\frac{1}{2}P(\min\{\delta T_1, \delta\tau_1\} \leqslant \min\{T_2, \tau_2\}, \delta T_1 \geqslant \tau, T_2 \geqslant \tau)}{1 - \Delta A(u)} > 0$$

for all $u \in [0, \tau)$. Hence, at least one of the summands in (3.3) with $m = 1$ is strictly positive.
Similarly, under (2), we have $F_{2,-}(\tau) = P(\delta T_1 < \min\{T_2, \tau\}) > 0$ and

$$\begin{aligned}
h_2(u) &= \frac{S(u) - F_2(\tau) + F_2(u) - \frac{1}{2}(F_3(\tau) - F_3(u))}{1 - \Delta A(u)} \\
&= \frac{F_1(\tau) - F_1(u) + \frac{1}{2}(F_3(\tau) - F_3(u))}{1 - \Delta A(u)} \\
&\geqslant \frac{\frac{1}{2}\Delta(F_1 + F_3)(\tau)}{1 - \Delta A(u)} \\
&= \frac{\frac{1}{2}P(\min\{\delta T_1, \delta\tau_1\} \geqslant \min\{T_2, \tau_2\}, \delta T_1 \geqslant \tau, T_2 \geqslant \tau)}{1 - \Delta A(u)} > 0
\end{aligned}$$

for all $u \in [0, \tau)$. Then, at least one of the summands in (3.3) with $m = 2$ is strictly positive.
For (3), we have $F_{3,-}(\tau) = P(\delta T_1 = T_2 < \tau) > 0$ and

$$\begin{aligned}
|h_3(u)| &= \frac{\left|\frac{1}{2}S(u) - F_2(\tau) + F_2(u) - \frac{1}{2}(F_3(\tau) - F_3(u))\right|}{1 - \Delta A(u)} \\
&= \frac{\frac{1}{2}\left|F_1(\tau) - F_1(u) - F_2(\tau) + F_2(u)\right|}{1 - \Delta A(u)} \\
&= \frac{|P(\min\{\delta T_1, \delta\tau_1\} > \min\{T_2, \tau_2\}, \min\{\delta T_1, T_2\} > u) - P(\min\{\delta T_1, \delta\tau_1\} < \min\{T_2, \tau_2\}, \min\{\delta T_1, T_2\} > u)|}{2 - 2\Delta A(u)} \\
&> 0
\end{aligned}$$

for all $u \in [0, \tau)$. Thus, at least one of the summands in (3.3) with $m = 3$ is strictly positive. □

Analogously to [30], we can use the Greenwood-type variance estimators $\widehat{\sigma}_m^2, \widehat{\sigma}_{m\ell}, \widehat{\sigma}_\bullet^2$ for $\sigma_m^2, \sigma_{m\ell}, \sigma_\bullet^2$, cf. [3], (4.4.17) and (4.4.18), and $\widehat{A} = \widehat{A}_1 + \widehat{A}_2 + \widehat{A}_3$ for A to obtain the variance estimator

$$\begin{aligned}
\widehat{\sigma}_\theta^2 = &\int_{[0,\tau]} \widehat{S}_-^2 \,\mathrm{d}\left(\widehat{\sigma}_2^2 + \widehat{\sigma}_{23} + \frac{1}{4}\widehat{\sigma}_3^2\right) \\
&- 2\int_{[0,\tau]}\int_{[0,v)} \frac{\widehat{S}_-(u)\widehat{S}_-(v)}{1 - \Delta\widehat{A}(u)} \,\mathrm{d}\left(\widehat{\sigma}_{12} + \widehat{\sigma}_2^2 + \frac{3}{2}\widehat{\sigma}_{23} + \frac{1}{2}\widehat{\sigma}_{13} + \frac{1}{2}\widehat{\sigma}_3^2\right)(u)\mathrm{d}\left(\widehat{A}_2 + \frac{1}{2}\widehat{A}_3\right)(v) \\
&+ \int_{[0,\tau]}\int_{[0,\tau]} \widehat{S}_-(u)\widehat{S}_-(v)\int_{[0,\min\{u,v\})} \frac{1}{(1 - \Delta\widehat{A}(w))^2} \,\mathrm{d}\widehat{\sigma}_\bullet^2(w)\mathrm{d}\left(\widehat{A}_2 + \frac{1}{2}\widehat{A}_3\right)(u)\mathrm{d}\left(\widehat{A}_2 + \frac{1}{2}\widehat{A}_3\right)(v).
\end{aligned} \tag{3.4}$$

The variance estimator $\widehat{\sigma}_\theta^2$ is a continuous functional of $(\widehat{A}_1, \widehat{A}_2, \widehat{A}_3)$ and $\widehat{\sigma}_m^2, \widehat{\sigma}_{m\ell}, \widehat{\sigma}_\bullet^2, m, \ell \in \{1, 2, 3\}, m \neq \ell$. By the uniform consistency of $(\widehat{A}_1, \widehat{A}_2, \widehat{A}_3)$, cf. (3.2), and $\widehat{\sigma}_m^2, \widehat{\sigma}_{m\ell}, \widehat{\sigma}_\bullet^2, m, \ell \in \{1, 2, 3\}, m \neq \ell$, the continuous mapping theorem yields the consistency of the variance estimator. The theorem follows then by applying Slutsky's lemma. □

Proof of Theorem 3.3 Analogously to the proof of Theorem 2 in the supplement of [30], we obtain by Theorem 2 in [27] that $\sqrt{n}(\widetilde{\widetilde{\theta}} - 1/2)$ converges weakly conditionally on the data $(Z_j, \varepsilon_j), j \in \{1, \ldots, n\}$, in outer probability as $n \to \infty$ to a centered normal variable with variance

$$\begin{aligned}
\widetilde{\sigma}_\theta^2 :=& \int_{[0,\tau]} S_-^2 \,\mathrm{d}\left(\widetilde{\sigma}_2^2 + \widetilde{\sigma}_{23} + \frac{1}{4}\widetilde{\sigma}_3^2\right) \\
&- \int_{[0,\tau]} \int_{[0,v)} \frac{S_-(u)S_-(v)}{1-\Delta A(u)} \,\mathrm{d}\left(\widetilde{\sigma}_{12} + \widetilde{\sigma}_2^2 + \frac{3}{2}\widetilde{\sigma}_{23} + \frac{1}{2}\widetilde{\sigma}_{13} + \frac{1}{2}\widetilde{\sigma}_3^2\right)(u)\mathrm{d}A(v) \\
&+ \frac{1}{4} \int_{[0,\tau]} \int_{[0,\tau]} S_-(u)S_-(v) \int_{[0,\min\{u,v\})} \frac{1}{(1-\Delta A(w))^2} \,\mathrm{d}\widetilde{\sigma}_\bullet^2(w)\mathrm{d}A(u)\mathrm{d}A(v),
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{\sigma}_1^2(t) := \widetilde{\sigma}_2^2(t) :=& \frac{1}{2}\int_{[0,t]} \frac{1-\Delta\frac{A_1+A_2}{2}}{y} \,\mathrm{d}(A_1+A_2), \quad \widetilde{\sigma}_3^2(t) := \sigma_3^2(t), \\
\widetilde{\sigma}_{12}(t) :=& -\frac{1}{4}\int_{[0,t]} \frac{\Delta(A_1+A_2)}{y} \,\mathrm{d}(A_1+A_2), \quad \widetilde{\sigma}_{13}(t) := \widetilde{\sigma}_{23}(t) := -\frac{1}{2}\int_{[0,t]} \frac{\Delta A_3}{y} \,\mathrm{d}(A_1+A_2)
\end{aligned}$$

for all $t \in [0, \tau]$.

Lemma 3.3. *If Assumption 3.1 and* $\max\{P(T_2 < \min\{\delta T_1, \tau\}), P(\delta T_1 < \min\{T_2, \tau\})\} > 0$ *hold, we have* $\widetilde{\sigma}_\theta^2 > 0$.

Proof of Lemma 3.3. By proceeding similarly as in the proof of Lemma 3.2, we obtain

$$\widetilde{\sigma}_\theta^2 = \frac{1}{2}\left(\sum_{m=1}^{2} \int_{[0,\tau]} \frac{\widetilde{h}_m^2}{y} \,\mathrm{d}(A_1^c + A_2^c) + \sum_{x \in \mathcal{D}} \frac{\sum_{m=1}^2 \widetilde{h}_m^2(x)\Delta(A_1+A_2)(x) - \left(\sum_{m=1}^2 \widetilde{h}_m(x)\Delta(A_1+A_2)(x)\right)^2}{y(x)}\right),$$

where

$$\widetilde{h}_1(u) := \frac{\int_{(u,\tau]} S_- \,\mathrm{d}A}{2 - 2\Delta A(u)} \quad \text{and} \quad \widetilde{h}_2(u) := S_-(u) - \widetilde{h}_1(u)$$

for all $u \in [0, \tau]$ with $0/0 := 0$. As in the proof of Lemma 3.2, one can show

$$\widetilde{h}_1(u) \geqslant \frac{S_-(\tau)\Delta A(\tau)}{2-2\Delta A(u)} = \frac{\Delta(F_1+F_2+F_3)(\tau)}{2-2\Delta A(u)} = \frac{P(\delta T_1 \geqslant \tau, T_2 \geqslant \tau)}{2-2\Delta A(u)} > 0$$

for all $u \in [0, \tau)$ by Assumption 3.1. Furthermore, it holds

$$\begin{aligned}
F_{1,-}(\tau) + F_{2,-}(\tau) &= P(T_2 < \min\{\delta T_1, \tau\} \vee \delta T_1 < \min\{T_2, \tau\}) \\
&\geqslant \max\{P(T_2 < \min\{\delta T_1, \tau\}), P(\delta T_1 < \min\{T_2, \tau\})\} > 0.
\end{aligned}$$

Hence, at least one of the summands with $m = 1$ is strictly positive. □

The consistency of the variance estimator $\widetilde{\widetilde{\sigma}}_\theta^2$ for $\widetilde{\sigma}_\theta^2$ follows analogously as in the proof of Theorem 3.2. Therefore note that S, G and A_3 as well as their estimators remain the same for the randomized data $(Z_j, \tilde{\varepsilon}_j), j \in \{1, \ldots, n\}$. Moreover, $\frac{A_1+A_2}{2}, \frac{A_1+A_2}{2}, A_3$ can be calculated as the cause-specific cumulative hazard functions of the randomized data $(Z_j, \tilde{\varepsilon}_j), j \in \{1, \ldots, n\}$. Consequently, the cause-specific Nelson-Aalen estimators based on the randomized data converge uniformly in outer probability to $\frac{A_1+A_2}{2}, \frac{A_1+A_2}{2}, A_3$, respectively, on $[0, \tau]$ by Theorem 4.1 in [26] and the separability of the limit by Section B. This implies the consistency of the variance estimator. Applying Slutsky's lemma completes the proof. □

Proof of Theorem 3.4 Let $G_i : [0, \infty) \ni t \mapsto P(C_i > t) \in [0, 1]$ denote the survival function of the censoring times C_i at time $t \geqslant 0$ in the following for $i \in \{1, 2\}$.
By Appendix D.1 of [27], the influence function of the Kaplan-Meier estimator $\widehat{S}_i$ at the Dirac measure in $(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j})$ is given by

$$t \mapsto S_i(t)\left[\frac{\delta_{ij}\mathbb{1}\{X_{ij} \leqslant t\}}{G_{i-}(X_{ij})S_i(X_{ij})} - \int_{[0,\min\{t,X_{ij}\}]} \frac{1}{G_{i-}(u)S_i(u)} \,\mathrm{d}A_i(u)\right]$$

for $i \in \{1,2\}, j \in \{1,...,n\}$. Hence, one can calculate the influence function of the estimator $\widehat{\mu}_i$ for the RMST μ_i at the Dirac measure in $(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j})$ as

$$\mathrm{IF}_i(X_{ij}, \delta_{ij}) := \int_0^\tau S_i(t) \left[\frac{\delta_{ij} \mathbb{1}\{X_{ij} \leqslant t\}}{G_{i-}(X_{ij}) S_i(X_{ij})} - \int_{[0,\min\{t,X_{ij}\}]} \frac{1}{G_{i-}(u) S_i(u)} \, \mathrm{d}A_i(u) \right] \mathrm{d}t$$

for $i \in \{1,2\}, j \in \{1,...,n\}$. Furthermore, we set

$$\mathrm{IF}_i(x, \delta) := \int_0^\tau S_i(t) \left[\frac{\delta \mathbb{1}\{x \leqslant t\}}{G_{i-}(x) S_i(x)} - \int_{[0,\min\{t,x\}]} \frac{1}{G_{i-}(u) S_i(u)} \, \mathrm{d}A_i(u) \right] \mathrm{d}t$$

for all $i \in \{1,2\}, x \in [0,\tau], \delta \in \{0,1\}$. Note that $\sup_{x \in [0,\tau], \delta \in \{0,1\}} |\mathrm{IF}_1(x,\delta)|$ and $\sup_{x \in [0,\tau], \delta \in \{0,1\}} |\mathrm{IF}_2(x,\delta)|$ are bounded due to Assumption 3.4. The influence function of $\widehat{\mu}_1 - \widehat{\mu}_2$ at the Dirac measure in $(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j})$ is then given by

$$\mathrm{IF}_j^{\mathrm{diff}} := \mathrm{IF}_1(X_{1j}, \delta_{1j}) - \mathrm{IF}_2(X_{2j}, \delta_{2j})$$

for $j \in \{1,...,n\}$. Furthermore, the chain rule (Theorem 3.10.3 in [74]) implies that $\Phi'_{\theta_0}(\cdot)/\Phi(\theta_0)$ is the Hadamard derivative of $\theta \mapsto \log(\Phi(\theta))$ at θ_0 for a Hadamard-differentiable map $\Phi : \mathbb{D}_\Phi \to \mathbb{R}$ with $\theta_0 \in \mathbb{D}_\Phi$. Hence, the influence function of $\log(\widehat{\mu}_1) - \log(\widehat{\mu}_2)$ at the Dirac measure in $(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j})$ is given by

$$\mathrm{IF}_j^{\mathrm{rat}} := \frac{\mathrm{IF}_1(X_{1j}, \delta_{1j})}{\mu_1} - \frac{\mathrm{IF}_2(X_{2j}, \delta_{2j})}{\mu_2}$$

for $j \in \{1,...,n\}$. By [27], it holds

$$\sqrt{n}((\widehat{\mu}_1 - \widehat{\mu}_2) - (\mu_1 - \mu_2)) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \mathrm{IF}_j^{\mathrm{diff}} + o_p(1)$$

and

$$\sqrt{n}((\log(\widehat{\mu}_1) - \log(\widehat{\mu}_2)) - (\log(\mu_1) - \log(\mu_2))) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \mathrm{IF}_j^{\mathrm{rat}} + o_p(1)$$

as $n \to \infty$. The advantage of this representation is that the summands are independent and identically distributed. Therefore, the asymptotic normality follows by an application of the central limit theorem.
Now, it remains to show that we have consistent estimators $\widehat{\sigma}^2_{\mathrm{diff}}, \widehat{\sigma}^2_{\mathrm{rat}}$ for the limit variances $\sigma^2_{\mathrm{diff}} := \mathbb{V}ar(\mathrm{IF}_1^{\mathrm{diff}})$ and $\sigma^2_{\mathrm{rat}} := \mathbb{V}ar(\mathrm{IF}_1^{\mathrm{rat}})$, respectively. Since $\sup_{x \in [0,\tau], \delta \in \{0,1\}} |\mathrm{IF}_1(x,\delta)|$ and $\sup_{x \in [0,\tau], \delta \in \{0,1\}} |\mathrm{IF}_2(x,\delta)|$ are bounded, the variances exist. We define

$$\widehat{\mathrm{IF}}_i(x, \delta) := \int_0^\tau \widehat{S}_i(t) \left[\frac{\delta \cdot \mathbb{1}\{x \leqslant t\}}{\widehat{G}_{i-}(x) \widehat{S}_i(x)} - \int_{[0,\min\{t,x\}]} \frac{1}{\widehat{G}_{i-}(u) \widehat{S}_i(u)} \, \mathrm{d}\widehat{A}_i(u) \right] \mathrm{d}t$$

for $i \in \{1,2\}, x \in [0,\tau], \delta \in \{0,1\}$, where $\widehat{S}_i, \widehat{G}_i$ denote the Kaplan-Meier estimators of S_i, G_i, respectively, and $\widehat{A}_i$ denotes the Nelson-Aalen estimator of A_i. Then, we set $\widehat{\mathrm{IF}}_j^{\mathrm{diff}} := \widehat{\mathrm{IF}}_1(X_{1j}, \delta_{1j}) - \widehat{\mathrm{IF}}_2(X_{2j}, \delta_{2j})$ and $\widehat{\mathrm{IF}}_j^{\mathrm{rat}} := \widehat{\mathrm{IF}}_1(X_{1j}, \delta_{1j})/\widehat{\mu}_1 - \widehat{\mathrm{IF}}_2(X_{2j}, \delta_{2j})/\widehat{\mu}_2$ for all $j \in \{1,...,n\}$. Consequently, $\widehat{\sigma}^2_{\mathrm{diff}}, \widehat{\sigma}^2_{\mathrm{rat}}$ can be estimated as the empirical variances

$$\widehat{\sigma}^2_{\mathrm{diff}} := \frac{1}{n} \sum_{j=1}^n \left(\widehat{\mathrm{IF}}_j^{\mathrm{diff}} - \frac{1}{n} \sum_{\ell=1}^n \widehat{\mathrm{IF}}_\ell^{\mathrm{diff}} \right)^2 \quad \text{and} \quad \widehat{\sigma}^2_{\mathrm{rat}} := \frac{1}{n} \sum_{j=1}^n \left(\widehat{\mathrm{IF}}_j^{\mathrm{rat}} - \frac{1}{n} \sum_{\ell=1}^n \widehat{\mathrm{IF}}_\ell^{\mathrm{rat}} \right)^2.$$

It is well known that $\widehat{S}_i, \widehat{G}_i, \widehat{A}_i$ are uniformly consistent for S_i, G_i, A_i on $[0,\tau]$, respectively, for $i \in \{1,2\}$, see, e.g., the supplement of [24] for details. Due to the continuity of the functionals, the continuous mapping theorem implies

$$\sup_{x \in [0,\tau], \delta \in \{0,1\}} \left| \widehat{\mathrm{IF}}_i(x,\delta) - \mathrm{IF}_i(x,\delta) \right| \xrightarrow{P} 0$$

as $n \to \infty$ for $i \in \{1,2\}$. Thus, easy calculations and an application of Slutsky's lemma yield

$$\left| \widehat{\sigma}^2_{\mathrm{diff}} - \frac{1}{n} \sum_{j=1}^n \left(\mathrm{IF}_j^{\mathrm{diff}} - \frac{1}{n} \sum_{\ell=1}^n \mathrm{IF}_\ell^{\mathrm{diff}} \right)^2 \right| \xrightarrow{P} 0 \quad \text{and} \quad \left| \widehat{\sigma}^2_{\mathrm{rat}} - \frac{1}{n} \sum_{j=1}^n \left(\mathrm{IF}_j^{\mathrm{rat}} - \frac{1}{n} \sum_{\ell=1}^n \mathrm{IF}_\ell^{\mathrm{rat}} \right)^2 \right| \xrightarrow{P} 0$$

as $n \to \infty$. Since $\mathrm{IF}_j^{\mathrm{diff}}, j \in \{1,...,n\}$, are i.i.d., it follows $\widehat{\sigma}^2_{\mathrm{diff}} \xrightarrow{P} \sigma^2_{\mathrm{diff}}$ and, analogously, $\widehat{\sigma}^2_{\mathrm{rat}} \xrightarrow{P} \sigma^2_{\mathrm{rat}}$ as $n \to \infty$.
□

Proof of Theorem 3.5 We aim to apply Theorem 2 in [27] and, hence, verifying the conditions in the following. Therefore, let

$$\mathcal{F} := \{(y_1, y_2, d_1, d_2) \mapsto d_i \cdot \mathbb{1}\{y_i \leqslant t\}, (y_1, y_2, d_1, d_2) \mapsto \mathbb{1}\{y_i > t\} \mid t \in [0, \tau], i \in \{1, 2\}\}$$

$$\text{and} \quad \widetilde{\mathcal{F}} := \left\{(y_1, y_2, d_1, d_2) \mapsto \frac{1}{2}\left(f(y_1, y_2, d_1, d_2) + f(y_2, y_1, d_2, d_1)\right) \mid f \in \mathcal{F}\right\}.$$

Analogously to the proof of Theorem 2 in [30], $\mathcal{F}$ and $\widetilde{\mathcal{F}}$ are VC-classes. Consequently, the sets are $\mathbb{P}$- and $\widetilde{\mathbb{P}}$-Donsker and Glivenko-Cantelli classes with $\mathbb{P} := P^{(X_{11}, X_{21}, \delta_{11}, \delta_{21})}$ and $\widetilde{\mathbb{P}} := P^{(X_{11}^\pi, X_{21}^\pi, \delta_{11}^\pi, \delta_{21}^\pi)}$. Furthermore, $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ have bounded supremum norms with respect to both sets $\mathcal{F}$ and $\widetilde{\mathcal{F}}$. Moreover, the Kaplan-Meier estimator is a Hadamard-differentiable functional as shown in Example 3.10.33 in [74]. Hence, the estimators for the RMST difference and ratio are also Hadamard-differentiable functionals, respectively, by the chain rule in Lemma 3.10.3 in [74]. Thus, Theorem 2 in [30] provides that $\sqrt{n}(\widehat{\mu}_1^\pi - \widehat{\mu}_2^\pi)$ and $\sqrt{n}\log(\widehat{\mu}_1^\pi / \widehat{\mu}_2^\pi)$ are converging weakly to centered normal distributions in outer probability conditionally on the data $(X_{1j}, X_{2j}, \delta_{1j}, \delta_{2j}), j \in \{1, ..., n\}$, as $n \to \infty$. For deriving the variances of the limit distributions, let $S^\pi := \frac{1}{2}(S_1 + S_2), G^\pi := \frac{1}{2}(G_1 + G_2), F^\pi := 1 - S^\pi$ and $A^\pi(.) := \int_{[0,.]} \frac{1}{S_-^\pi(t)} \, \mathrm{d}F^\pi(t)$ be the pooled survival, distribution and cumulative hazard functions. Note that permuting the data randomly leads to those survival, distribution and cumulative hazard functions for the permuted data. Furthermore, define

$$\mathrm{Q.IF}(x, \delta) := \int_0^\tau S^\pi(t) \left[\frac{\delta \cdot \mathbb{1}\{x \leqslant t\}}{G_-^\pi(x) S^\pi(x)} - \int_{[0, \min\{t, x\}]} \frac{1}{G_-^\pi(u) S^\pi(u)} \, \mathrm{d}A^\pi(u)\right] \mathrm{d}t$$

for all $x \in [0, \tau], \delta \in \{0, 1\}$. Note that $\sup_{x \in [0,\tau], \delta \in \{0,1\}} |\mathrm{Q.IF}(x, \delta)|$ is bounded under Assumption 3.4. One can show that the variances of the normal distributions are $\sigma_{\text{diff}}^\pi := \mathbb{V}ar\left(\mathrm{Q.IF}_1^{\text{diff}}\right)$ and $\sigma_{\text{rat}}^\pi := \mathbb{V}ar\left(\mathrm{Q.IF}_1^{\text{rat}}\right)$, where $\mathrm{Q.IF}_1^{\text{diff}} := \mathrm{Q.IF}(X_{11}, \delta_{11}) - \mathrm{Q.IF}(X_{21}, \delta_{21})$ and

$$\mathrm{Q.IF}_1^{\text{rat}} := \mathrm{Q.IF}(X_{11}, \delta_{11})/\overline{\mu} - \mathrm{Q.IF}(X_{21}, \delta_{21})/\overline{\mu} = (\mathrm{Q.IF}(X_{11}, \delta_{11}) - \mathrm{Q.IF}(X_{21}, \delta_{21}))/\overline{\mu}$$

similar as in the proof of Theorem 3.4 by the chain rule (Theorem 3.10.3 in [74]) with $\overline{\mu} := \int_0^\tau S^\pi(t) \, \mathrm{d}t = \frac{\mu_1 + \mu_2}{2}$. Hence, it remains to show that the permutation counterparts of the variance estimators are consistent. Since $\widehat{S}_i^\pi, \widehat{G}_i^\pi, \widehat{A}_i^\pi, i \in \{1, 2\}$, are continuous functionals of the empirical process of $(X_{1j}^\pi, X_{2j}^\pi, \delta_{1j}^\pi, \delta_{2j}^\pi)$ and $\mathcal{F}$ is a $\widetilde{\mathbb{P}}$-Glivenko-Cantelli class, it follows that $\widehat{S}_i^\pi, \widehat{G}_i^\pi, \widehat{A}_i^\pi$ are unconditionally uniformly consistent for S^π, G^π and A^π on $[0, \tau]$, respectively, for $i \in \{1, 2\}$. Thus, we get

$$\sup_{x \in [0,\tau], \delta \in \{0,1\}} |\mathrm{IF}_i^\pi(x, \delta) - \mathrm{Q.IF}(x, \delta)| \xrightarrow{P} 0$$

as $n \to \infty$ for $i \in \{1, 2\}$. Then, the consistency of the variance estimators follow analogously as in the proof of Theorem 3.4. $\square$

4 RMST-Based Inference in General Factorial Survival Designs

As we already developed tests based on restricted mean survival time comparisons in Section 3.2 for paired survival data, we now turn to complex factorial survival designs.
The asymptotic behavior and statistical inference of RMSTs in the two-sample case have already been examined in the literature. Due to an inflation of the type I error of the asymptotic RMST-based test for small samples as shown in [43] for the two-sample case, an unstudentized permutation approach was proposed by [43] under exchangeability. In [24], this approach was extended by developing a studentized permutation test to allow for different censoring distributions in the two groups. A similar approach has been further analyzed in the context of cure models, in both non- and semi-parametric models [29].
Such studentized permutation tests could be of interest for more complex factorial designs or more general linear hypotheses in practice, e.g., when more than two different treatments are to be compared in a clinical study. Thus, we aim to extend the studentized permutation test in [24] for general factorial designs and general linear hypotheses by employing a Wald-type test statistic. Furthermore, other resampling methods as the groupwise and the wild bootstrap are considered for this general setup.
On the other hand, when a global test detects a significant result by comparing the RMSTs of more than two groups, it is of interest which particular RMSTs differ significantly. Unfortunately, global tests do not yield this information. Therefore, multiple linear hypothesis testing procedures are desired. They offer the information which of the local hypotheses are rejected in addition to the global one. Moreover, their power is not necessarily lower than the power of a global testing procedure [47]. For gaining more power, we aim to take the exact asymptotic dependency structure between the different test statistics into account. In order to improve the small sample performance, we propose a groupwise and wild bootstrap procedure for approximating the limiting null distribution and we show their validity.
The remainder of this section is organized as follows. In Section 4.1, the factorial survival setup is presented. The global contrast testing problem is introduced in Section 4.2. Furthermore, a suitable test statistic is defined and studied in Section 4.2.1. The studentized permutation approach [24] is extended for more general factorial designs in Section 4.2.2. Furthermore, a groupwise and wild bootstrap procedure is investigated in Section 4.2.3 and 4.2.4, respectively. In Section 4.3, multiple contrast tests for the RMST are constructed and the consistency of the groupwise and wild bootstrap in this setup is shown. The small sample performance of the proposed RMST-based tests is analyzed in extensive simulation studies in Section 4.4. In Section 4.5, we illustrate the proposed methodologies by analyzing a real data example about the occurrence of hay fever.

4.1 Factorial Survival Setup

We consider the following factorial design as in [22], i.e., as k-sample setup, $k \in \mathbb{N}, k \geqslant 2$. We suppose that the survival and censoring times

$$T_{ij} \sim S_i, \quad C_{ij} \sim G_i, \quad j \in \{1, ..., n_i\}, i \in \{1, ..., k\},$$

respectively, are mutually independent. Here, $S_i : [0, \infty) \ni t \mapsto P(T_{i1} > t) \in [0, 1]$ and $G_i : [0, \infty) \ni t \mapsto P(C_{i1} > t) \in [0, 1]$ denote the survival functions of the survival and censoring times, respectively, and $n_i \in \mathbb{N}$ represent the numbers of individuals in group i for all $i \in \{1, ..., k\}$. Of note, we do not assume the continuity of the survival functions. Consequently, ties in the data are explicitly allowed. However, we assume that the S_i do not have jumps of size 1, i.e., the survival times are not deterministic. Moreover, we define the right-censored observable event times $X_{ij} := \min\{T_{ij}, C_{ij}\}$ and the censoring status $\delta_{ij} := \mathbb{1}\{X_{ij} = T_{ij}\}$ for all $j \in \{1, ..., n_i\}, i \in \{1, ..., k\}$.
The restricted mean survival time (RMST) of group i is defined as

$$\mu_i := \int_0^\tau S_i(t) \,\mathrm{d}t = \mathrm{E}[\min\{T_{i1}, \tau\}]$$

for all $i \in \{1, ..., k\}$. Here, $\tau > 0$ should be a pre-specified constant. By replacing S_i through the Kaplan-Meier estimator $\widehat{S}_i$, a natural estimator for the RMST of group i is

$$\widehat{\mu}_i := \int_0^\tau \widehat{S}_i(t) \,\mathrm{d}t$$

for all $i \in \{1, ..., k\}$. Let $\boldsymbol{\mu} := (\mu_1, ..., \mu_k)'$ be the vector of the RMSTs and $\widehat{\boldsymbol{\mu}} := (\widehat{\mu}_1, ..., \widehat{\mu}_k)'$ be the vector of their estimators.
Furthermore, we assume the following.

Assumption 4.1. *We assume that the group sizes do not vanish asymptotically, i.e.,*

$$\frac{n_i}{n} \to \kappa_i \in (0, 1) \tag{4.1}$$

as $n \to \infty$ for all $i \in \{1, ..., k\}$, where $n := \sum_{i=1}^k n_i$ represents the total sample size. Additionally, we assume that $\tau > 0$ fulfills $P(X_{i1} \geqslant \tau) = P(T_{i1} \geqslant \tau)P(C_{i1} \geqslant \tau) > 0$ and $P(T_{i1} < \tau) > 0$ for all $i \in \{1, ..., k\}$.

4.2 Global Tests

Let $r \in \mathbb{N}$, $\mathbf{c} \in \mathbb{R}^r$ be a fixed vector and $\mathbf{H} \in \mathbb{R}^{r\times k}$ be a contrast matrix, i.e., $\mathbf{H}\mathbf{1}_k = \mathbf{0}_r$. Moreover, we assume that $\operatorname{rank}(\mathbf{H}) > 0$. Then, we consider the null and alternative hypothesis

$$\mathcal{H}_0 : \mathbf{H}\boldsymbol{\mu} = \mathbf{c} \quad \text{vs.} \quad \mathcal{H}_1 : \mathbf{H}\boldsymbol{\mu} \neq \mathbf{c}. \tag{4.2}$$

The formulation of this testing framework is very general. In particular, it includes the null hypothesis of equal RMSTs in all groups by choosing, for example, $\mathbf{c} = \mathbf{0}_k$ and the *Grand-mean-type* contrast matrix [25] $\mathbf{H} := \mathbf{P}_k := \mathbf{I}_k - \mathbf{J}_k/k$. Here, $\mathbf{J}_k := \mathbf{1}_k\mathbf{1}_k' \in \mathbb{R}^{k\times k}$ represents the matrix of ones. Moreover, by splitting up indices, different kinds of factorial structures can be covered. For example, in a two-way design with factors A (a levels) and B (b levels), we set $k := ab$ and split up the group index i in two subindices $(i_1, i_2) \in \{1, ..., a\} \times \{1, ..., b\}$. Then, hypotheses about no main or interaction effect can be formulated by choosing $\mathbf{c}$ as the zero vector and one of the following contrast matrices:

- $\mathbf{H}_A := \mathbf{P}_a \otimes (\mathbf{1}_b'/b)$ (no main effect of factor A),
- $\mathbf{H}_B := (\mathbf{1}_a'/a) \otimes \mathbf{P}_b$ (no main effect of factor B),
- $\mathbf{H}_{AB} := \mathbf{P}_a \otimes \mathbf{P}_b$ (no interaction effect).

Here, $\otimes$ represents the Kronecker product. Higher-way designs or hierarchically nested layouts can be incorporated similarly as in [63].

4.2.1 The Wald-Type Test Statistic and its Asymptotic Behavior

In this section, a suitable test statistic for the testing problem (4.2) is constructed and its asymptotic behaviour is studied. First of all, let us introduce some notation. In the following, $Y_i(x) := \sum_{j=1}^{n_i} \mathbb{1}\{X_{ij} \geqslant x\}$ represents the number of individuals at risk just before time $x \geqslant 0$ and $N_i(x) := \sum_{j=1}^{n_i} \delta_{ij}\mathbb{1}\{X_{ij} \leqslant x\}$ denotes the number of observed individuals with an event before or at time $x \geqslant 0$ in group i with $i \in \{1, ..., k\}$. Furthermore, $\widehat{A}_i(x) := \int_{[0,x]} \frac{1}{Y_i}\, \mathrm{d}N_i$ denotes the Nelson-Aalen estimator of the cumulative hazard function $A_i(x) := \int_{[0,x]} \frac{1}{S_{i-}}\, \mathrm{d}F_i = \int_{[0,x]} \frac{1}{y_i}\, \mathrm{d}\nu_i$ at time x with $\nu_i(x) := \int_{[0,x]} G_{i-}\, \mathrm{d}F_i$, $y_i(x) := S_{i-}(x)G_{i-}(x)$ and $F_i(x) := 1 - S_i(x)$ for all $x \geqslant 0, i \in \{1, ..., k\}$.
Then, we define the Wald-type test statistic for the testing problem (4.2) as

$$W_n(\mathbf{H}, \mathbf{c}) := n(\mathbf{H}\widehat{\boldsymbol{\mu}} - \mathbf{c})'(\mathbf{H}\widehat{\boldsymbol{\Sigma}}\mathbf{H}')^+(\mathbf{H}\widehat{\boldsymbol{\mu}} - \mathbf{c}),$$

where $\widehat{\boldsymbol{\Sigma}} := \operatorname{diag}(\widehat{\sigma}_1^2, ..., \widehat{\sigma}_k^2)$ with

$$\widehat{\sigma}_i^2 := n \int_{[0,\tau)} \left(\int_x^\tau \widehat{S}_i(t)\, \mathrm{d}t \right)^2 \frac{1}{(1 - \Delta\widehat{A}_i(x))Y_i(x)}\, \mathrm{d}\widehat{A}_i(x) \tag{4.3}$$

being an estimator regarding the asymptotic variance of $\sqrt{n}(\widehat{\mu}_i - \mu_i)$ for all $i \in \{1, ..., k\}$ [24]. The following theorem provides the asymptotic distribution of the Wald-type test statistic.

Theorem 4.1. *Under Assumption 4.1 and the null hypothesis in (4.2), we have, as $n \to \infty$,*

$$W_n(\mathbf{H}, \mathbf{c}) \xrightarrow{d} \chi^2_{\operatorname{rank}(\mathbf{H})}.$$

Thus, we obtain an asymptotically valid level-α-test

$$\varphi_n := \mathbb{1}\{W_n(\mathbf{H}, \mathbf{c}) > \chi^2_{\operatorname{rank}(\mathbf{H}),1-\alpha}\}, \tag{4.4}$$

where $\chi^2_{\operatorname{rank}(\mathbf{H}),1-\alpha}$ denotes the $(1-\alpha)$-quantile of the $\chi^2_{\operatorname{rank}(\mathbf{H})}$ distribution for $\alpha \in (0, 1)$.

4.2.2 Studentized Permutation Test

For two-sample comparisons, it was pointed out in [43] that RMST-based tests derived from asymptotic methods have an increased type I error. Hence, we aim to improve the type I error control by extending the studentized permutation approach of [24] to the present general factorial design setting. When considering the already treated two-sample case, the approach has the advantage that it also works asymptotically without the assumption of exchangeable data. In this section, we will transfer these good properties to general factorial designs to construct a resampling-based test that serves as an alternative for (4.4).

For this purpose, let $(\mathbf{X}, \boldsymbol{\delta}) := (X_{ij}, \delta_{ij})_{j\in\{1,...,n_i\}, i\in\{1,...,k\}}$ denote the observed data and

$$(\mathbf{X}^\pi, \boldsymbol{\delta}^\pi) := (X_{ij}^\pi, \delta_{ij}^\pi)_{j\in\{1,...,n_i\}, i\in\{1,...,k\}}$$

be the permuted version. That is, the groups of the original data are randomly shuffled in the sense that the data pairs (X_{ij}, δ_{ij}) are permuted. In the following, we denote the permutation counterparts of the statistics $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ defined in the previous sections with a superscript π: $\widehat{\boldsymbol{\mu}}^\pi$ and $\widehat{\boldsymbol{\Sigma}}^\pi$. Then, we define the permutation counterpart of the Wald-type test statistic as

$$W_n^\pi(\mathbf{H}) := n(\mathbf{H}\widehat{\boldsymbol{\mu}}^\pi)'(\mathbf{H}\widehat{\boldsymbol{\Sigma}}^\pi\mathbf{H}')^+\mathbf{H}\widehat{\boldsymbol{\mu}}^\pi.$$

Note that the permutation counterpart of the Wald-type test statistic does not depend on $\mathbf{c}$.

Theorem 4.2. *Under Assumption 4.1,we have*

$$W_n^\pi(\mathbf{H}) \xrightarrow{d^*} \chi^2_{\text{rank}(\mathbf{H})} \tag{4.5}$$

as $n \to \infty$.

From this result, we can construct a permutation test

$$\varphi_n^\pi := \mathbb{1}\{W_n(\mathbf{H}, \mathbf{c}) > q_{1-\alpha}^\pi\},$$

where $q_{1-\alpha}^\pi$ denotes the $(1-\alpha)$-quantile of the conditional distribution of $W_n^\pi(\mathbf{H})$ given $(\mathbf{X}, \boldsymbol{\delta})$. Lemma 1 in [45] ensures that φ_n^π is asymptotically valid. Furthermore, Section 2.3.2 provides that the quantile may also be approximated by a Monte Carlo method.

4.2.3 Groupwise Bootstrap Test

Another possible solution for approximating the limiting distribution is the groupwise bootstrap. An advantage over the studentized permutation approach is that the groupwise bootstrap can mimic the different variance structures in the groups. This ensures that the groupwise bootstrap is also applicable for the multiple testing problem, see Section 4.3.
For the groupwise bootstrap, the bootstrap observations are drawn randomly with replacement from the observations of the corresponding group, i.e., $(X_{ij}^*, \delta_{ij}^*), j \in \{1, ..., n_i\}$, are drawn randomly from the ith sample $(X_{ij}, \delta_{ij}), j \in \{1, ..., n_i\}$, for all $i \in \{1, ..., k\}$. Then, we denote the groupwise bootstrap counterparts of the statistics $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ defined in Section 4.2.1 with a superscript $*$: $\widehat{\boldsymbol{\mu}}^*$ and $\widehat{\boldsymbol{\Sigma}}^*$. The groupwise bootstrap test statistic is defined by

$$W_n^*(\mathbf{H}) := n\left(\mathbf{H}(\widehat{\boldsymbol{\mu}}^* - \widehat{\boldsymbol{\mu}})\right)'(\mathbf{H}\widehat{\boldsymbol{\Sigma}}^*\mathbf{H}')^+\left(\mathbf{H}(\widehat{\boldsymbol{\mu}}^* - \widehat{\boldsymbol{\mu}})\right).$$

The following theorem provides the consistency of the groupwise bootstrap.

Theorem 4.3. *Under Assumption 4.1, we have*

$$W_n^*(\mathbf{H}) \xrightarrow{d^*} \chi^2_{\text{rank}(\mathbf{H})}$$

as $n \to \infty$.

Hence, we obtain a groupwise bootstrap test

$$\varphi_n^* := \mathbb{1}\{W_n(\mathbf{H}, \mathbf{c}) > q_{1-\alpha}^*\},$$

where $q_{1-\alpha}^*$ denotes the $(1-\alpha)$-quantile of the conditional distribution of $W_n^*(\mathbf{H})$ given $(\mathbf{X}, \boldsymbol{\delta})$. By Lemma 1 in [45], φ_n^* is an asymptotically valid level-α test. By Section 2.3.2, the quantile may also be approximated by a Monte Carlo method.
Note that we do not need the property that $\mathbf{H}$ is a contrast matrix in the proofs of Theorems 4.1 and 4.3. Hence, the groupwise bootstrap test is also valid for general matrices $\mathbf{H} \in \mathbb{R}^{r\times k}$ with $\text{rank}(\mathbf{H}) > 0$.

4.2.4 Wild Bootstrap Test

In this section, we use the wild bootstrap approach similar as described in [80] for approximating the distribution of the Wald-type test statistic. For developing the approach in [80], the ideas of [52] and [62] were adopted.

In [62], it was proposed to replace $\sqrt{n}(\widehat{S}_i(t) - S_i(t))$ by

$$\sqrt{n} \sum_{j=1}^{n_i} G_{ij} \widehat{S}_i(t) \int_{[0,t]} \frac{1}{Y_i(x)} \, \mathrm{d}N_{ij}(x)$$

for $t \geqslant 0$, where $G_{ij}, j \in \{1, ..., n_i\}, i \in \{1, ..., k\}$, are independent standard Gaussian random variables and $N_{ij}(x) := \delta_{ij} \mathbb{1}\{X_{ij} \leqslant x\}$ for all $x \geqslant 0$. We modify this procedure analogously to Greenwood's formula [40] and replace $\sqrt{n}(\widehat{S}_i(t) - S_i(t))$ by

$$\sqrt{n} \sum_{j=1}^{n_i} G_{ij} \widehat{S}_i(t) \int_{[0,t]} \frac{1}{\sqrt{(Y_i(x) - \Delta N_i(x)) Y_i(x)}} \, \mathrm{d}N_{ij}(x)$$

for all $t > 0$; also see [26]. For continuous survival functions S_i, this is asymptotically equivalent to the proposal of [62]. However, the modification becomes important for the extension to discontinuous distribution functions S_i. Moreover, we aim to weaken the assumption that $G_{ij}, j \in \{1, ..., n_i\}, i \in \{1, ..., k\}$, are standard Gaussian distributed. In fact, the multipliers only have to fulfill the following conditions:

(i) $G_{ij}, j \in \{1, ..., n_i\}, i \in \{1, ..., k\}$, are i.i.d. and independent of the data $(\mathbf{X}, \boldsymbol{\delta})$,

(ii) $\mathrm{E}\left[G_{ij}\right] = 0$,

(iii) $\mathrm{E}\left[G_{ij}^2\right] = 1$,

(iv) $\mathrm{E}\left[G_{ij}^4\right] \leqslant C$ for some constant $C < \infty$

for all $j \in \{1, ..., n_i\}, i \in \{1, ..., k\}$. Hence, we replace $\sqrt{n}(\widehat{\mu}_i - \mu_i)$ by

$$\sqrt{n} \widehat{\mu}_i^G := \sqrt{n} \sum_{j=1}^{n_i} G_{ij} \int_0^\tau \widehat{S}_i(t) \int_{[0,t]} \frac{1}{\sqrt{(Y_i(x) - \Delta N_i(x)) Y_i(x)}} \, \mathrm{d}N_{ij}(x) \, \mathrm{d}t$$

for all $i \in \{1, ..., k\}$. Furthermore, let

$$W_n^G(\mathbf{H}) := n(\mathbf{H}\widehat{\boldsymbol{\mu}}^G)'(\mathbf{H}\widehat{\boldsymbol{\Sigma}}^G\mathbf{H}')^+\mathbf{H}\widehat{\boldsymbol{\mu}}^G$$

be the wild bootstrap counterpart of the Wald-type test statistic, where $\widehat{\boldsymbol{\mu}}^G := (\widehat{\mu}_1^G, ..., \widehat{\mu}_k^G)'$ and $\widehat{\boldsymbol{\Sigma}}^G := \mathrm{diag}(\widehat{\sigma}_1^{G2}, ..., \widehat{\sigma}_k^{G2})$ with

$$\widehat{\sigma}_i^{G2} := n \sum_{j=1}^{n_i} G_{ij}^2 \int_{[0,\tau)} \left(\int_x^\tau \widehat{S}_i(t) \, \mathrm{d}t \right)^2 \frac{1}{(Y_i(x) - \Delta N_i(x)) Y_i(x)} \, \mathrm{d}N_{ij}(x).$$

Then, the following theorem ensures the wild bootstrap consistency.

Theorem 4.4. *Under Assumption 4.1, we have*

$$W_n^G(\mathbf{H}) \xrightarrow{d} \chi^2_{\mathrm{rank}(\mathbf{H})}$$

almost surely as $n \to \infty$ given the data $(\mathbf{X}, \boldsymbol{\delta})$. Mathematically, this means

$$\sup_{z \in \mathbb{R}} \left| P\left(W_n^G(\mathbf{H}) \leqslant z \mid (\mathbf{X}, \boldsymbol{\delta})\right) - P\left(Z \leqslant z\right) \right| \xrightarrow{a.s.} 0$$

as $n \to \infty$, where $Z \sim \chi^2_{\mathrm{rank}(\mathbf{H})}$.

We define a wild bootstrap test by

$$\varphi^G := \mathbb{1}\{W_n(\mathbf{H}, \mathbf{c}) > q_{1-\alpha}^G\},$$

where $q_{1-\alpha}^G$ denotes the $(1-\alpha)$-quantile of the conditional distribution of $W_n^G(\mathbf{H})$ given $(\mathbf{X}, \boldsymbol{\delta})$. Again, the asymptotic validity of this test is provided by Lemma 1 in [45] and the quantile may be approximated by a Monte Carlo method by Section 2.3.2.

4.3 Multiple Tests

Let us now interpret the contrast matrix $\mathbf{H}$ as a partitionized matrix $\mathbf{H} = [\mathbf{H}_1', ..., \mathbf{H}_L']'$ with $\mathbf{H}_\ell \in \mathbb{R}^{r_\ell \times k}$ for all $\ell \in \{1, ..., L\}$ such that $\sum_{\ell=1}^{L} r_\ell = r$ and, analogously, $\mathbf{c} = (\mathbf{c}_1', ..., \mathbf{c}_L')'$ with $\mathbf{c}_\ell \in \mathbb{R}^{r_\ell}$ for all $\ell \in \{1, ..., L\}$. Moreover, we assume $\operatorname{rank}(\mathbf{H}_\ell) > 0$ for all $\ell \in \{1, ..., L\}$. In this section, we aim to construct a testing procedure for the multiple testing problem with null and alternative hypotheses

$$\mathcal{H}_{0,\ell} : \mathbf{H}_\ell \boldsymbol{\mu} = \mathbf{c}_\ell \quad \text{vs.} \quad \mathcal{H}_{1,\ell} : \mathbf{H}_\ell \boldsymbol{\mu} \neq \mathbf{c}_\ell, \qquad \text{for } \ell \in \{1, ..., L\}. \tag{4.6}$$

In doing so, we aim to incorporate the asymptotically exact dependence structure between the test statistics of the L local tests to gain more power than, for example, by using a Bonferroni-correction.

Example 4.1. *A global null hypothesis which is of interest in many applications is the equality of the RMSTs, i.e.,* $\mathcal{H}_0 : \mu_1 = ... = \mu_k$ *versus the alternative* $\mathcal{H}_1 : \mu_{i_1} \neq \mu_{i_2}$ *for some* $i_1, i_2 \in \{1, ..., k\}$*. However, there are different possible choices of the contrast matrix* $\mathbf{H}$ *which lead to this global null hypothesis [47]. A popular choice is the Grand-mean-type contrast matrix as introduced in Section 4.1, where the RMSTs of the different groups are compared with the overall mean of the RMSTs* $\overline{\mu} := \frac{1}{k}\sum_{i=1}^{k} \mu_i$ *for the different contrasts, respectively. Many-to-one comparisons can be considered by choosing the Dunnett-type contrast matrix [34]*

$$\mathbf{H} = [-\mathbf{1}_{k-1}, \mathbf{I}_{k-1}] = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ -1 & 0 & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{(k-1)\times k} \tag{4.7}$$

and $\mathbf{c} = \mathbf{0}_{k-1}$*, where the RMSTs* $\mu_2, ..., \mu_k$ *are compared to the RMST* μ_1 *of the first group regarding the different contrasts. In order to compare all pairs of RMSTs* $\mu_{i_1}, \mu_{i_2}, i_1, i_2 \in \{1, ..., k\}$ *with* $i_1 \neq i_2$*, the Tukey-type contrast matrix [73]*

$$\mathbf{H} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ -1 & 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & 0 & 0 & 0 & \cdots & \cdots & 1 \\ 0 & -1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \in \mathbb{R}^{k(k-1)/2\times k} \tag{4.8}$$

and $\mathbf{c} = \mathbf{0}_{k(k-1)/2}$ *can be used. An overview of different contrast tests can be found in [11].*
Furthermore, the choice of the considered partition of the matrix $\mathbf{H} = [\mathbf{H}_1', ..., \mathbf{H}_R']'$ *and, therefore, the resulting local hypotheses depend on the question of interest. This general formulation of the multiple testing problem covers the post-hoc testing problem and includes, for example, the local null hypotheses* $\mathcal{H}_{0,\ell} : \mu_\ell = \overline{\mu}$*, for* $\ell \in \{1, \ldots, k\}$*, by choosing* $\mathbf{H}_\ell = \mathbf{e}_\ell' - \frac{1}{k}\mathbf{1}_k'$ *for all* $\ell \in \{1, ..., k\}$*, where* $\mathbf{e}_\ell \in \mathbb{R}^k$ *denotes the* ℓ*th unit vector. Analogously, we can perform many-to-one comparisons and all-pair comparisons of the mean functions simultaneously by considering the* r *rows of the Dunnett-type and Tukey-type contrast matrix, respectively, as blocks* $\mathbf{H}_1, ..., \mathbf{H}_r$*. Furthermore, the formulation of this testing problem allows to perform multiple tests with more than one contrast matrix simultanously. In a two-way design, we may choose* $\mathbf{H}_1 = \mathbf{H}_A, \mathbf{H}_2 = \mathbf{H}_B$ *and* $\mathbf{H}_3 = \mathbf{H}_{AB}$ *as introduced in Section 4.1, for example. This allows for simultaneous testing of the factors A and B and their interaction.*

For all local hypotheses in (4.6), we can calculate the Wald-type test statistics $W_n(\mathbf{H}_\ell, \mathbf{c}_\ell), \ell \in \{1, ..., L\}$. Since we aim to use the asymptotically exact dependence structure of the test statistics, we have to investigate the joint asymptotic behavior.
Therefore, let $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}_k, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} := \operatorname{diag}(\sigma_1^2, ..., \sigma_k^2)$ in the following, where

$$\sigma_i^2 := \frac{1}{\kappa_i}\int_{[0,\tau)} \left(\int_x^\tau S_i(t)\,\mathrm{d}t\right)^2 \frac{1}{(1-\Delta A_i(x))y_i(x)}\,\mathrm{d}A_i(x), \quad i \in \{1, ..., k\}.$$

In Section S.5 of the supplement of [24], it is shown that σ_i^2 is the almost sure limit of (4.3) for all $i \in \{1, ..., k\}$.

Theorem 4.5. *Let* $\mathcal{T}$ *denote the indices of true null hypotheses in* (5.2)*. Under Assumption 4.1, we have, as* $n \to \infty$*,*

$$\begin{aligned} (W_n(\mathbf{H}_\ell, \mathbf{c}_\ell))_{\ell\in\mathcal{T}} &= \Big(n\,(\mathbf{H}_\ell(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}))'\,(\mathbf{H}_\ell \widehat{\boldsymbol{\Sigma}}\mathbf{H}_\ell')^+ \mathbf{H}_\ell(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})\Big)_{\ell\in\mathcal{T}} \\ &\xrightarrow{d} \big((\mathbf{H}_\ell\mathbf{Z})'(\mathbf{H}_\ell\boldsymbol{\Sigma}\mathbf{H}_\ell')^+\mathbf{H}_\ell\mathbf{Z}\big)_{\ell\in\mathcal{T}}. \end{aligned} \tag{4.9}$$

4.3.1 Asymptotic Multiple Tests

Note that $\mathbf{\Sigma}$ is generally unknown such that we do not know the exact asymptotic joint limiting distribution of $(W_n(\mathbf{H}_1, \mathbf{c}_1), ..., W_n(\mathbf{H}_L, \mathbf{c}_L))$. However, we can estimate the joint limit distribution of the test statistics by estimating $\mathbf{\Sigma}$ through $\widehat{\mathbf{\Sigma}}$. This results in the local asymptotic tests

$$\varphi_\ell = \mathbb{1}\left\{W_n(\mathbf{H}_\ell, \mathbf{c}_\ell) > \chi^2_{\mathrm{rank}(\mathbf{H}_\ell),1-\beta_n}\right\} \quad , \ell \in \{1, \ldots, L\}, \tag{4.10}$$

where β_n denotes the local level for each test and can be derived from the conditional multivariate distribution

$$\left((\mathbf{H}_\ell \widehat{\mathbf{\Sigma}}^{1/2}\mathbf{Y})'(\mathbf{H}_\ell \widehat{\mathbf{\Sigma}}\mathbf{H}_\ell')^+ \mathbf{H}_\ell \widehat{\mathbf{\Sigma}}^{1/2}\mathbf{Y}\right)_{\ell\in\{1,\ldots,L\}}$$

given $\widehat{\mathbf{\Sigma}}$ as explained in Section 2.3 for $\mathbf{Y} \sim \mathcal{N}_k(\mathbf{0}_k, \mathbf{I}_k)$ independent of $\widehat{\mathbf{\Sigma}}$. In detail, the local level can be approximated by a Monte Carlo method as $\beta_n = \max\{\beta \in \{0, 1/B_n, \ldots, (B_n - 1)/B_n\} \mid \mathrm{FWER}_n(\beta) \leqslant \alpha\}$ with approximated family-wise error rate

$$\mathrm{FWER}_n(\beta) = \frac{1}{B_n}\sum_{b=1}^{B_n} \max_{\ell\in\{1,\ldots,L\}} \mathbb{1}\left\{(\mathbf{H}_\ell \widehat{\mathbf{\Sigma}}^{1/2}\mathbf{Y}^{(b)})'(\mathbf{H}_\ell \widehat{\mathbf{\Sigma}}\mathbf{H}_\ell')^+ (\mathbf{H}_\ell \widehat{\mathbf{\Sigma}}^{1/2}\mathbf{Y}^{(b)}) > \chi^2_{\mathrm{rank}(\mathbf{H}_\ell \widehat{\mathbf{\Sigma}}\mathbf{H}_\ell'),1-\beta}\right\}$$

for $\beta \in [0, 1)$ and $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(B_n)} \sim \mathcal{N}_{kM}(\mathbf{0}_k, \mathbf{I}_k)$ i.i.d. and independent of $\widehat{\mathbf{\Sigma}}$. Here, $(B_n)_{n\in\mathbb{N}}$ is a sequence of natural numbers with $B_n \to \infty$ as $n \to \infty$. Theorem 2.6 and Lemma 2.3 ensure the family-wise error control of the multiple tests.

Theorem 4.6. *Under Assumption 4.1, the asymptotic multiple tests* (4.10) *fulfill* (2.20)–(2.22)*, i.e., they control the family-wise error rate asymptotically in the strong sense and are asymptotically balanced.*

4.3.2 Multiple Resampling Tests

In order to improve the small sample performance of the tests, we aim to use consistent resampling procedures as in Section 4.2. Unfortunately, we cannot use the studentized permutation approach for approximating the joint limiting distribution. That is because

$$(W_n^\pi(\mathbf{H}_\ell))_{\ell\in\{1,\ldots,L\}} \xrightarrow{d^*} \left((\mathbf{H}_\ell \mathbf{Z}^\pi)'(\mathbf{H}_\ell \mathbf{\Sigma}^\pi \mathbf{H}_\ell')^+ \mathbf{H}_\ell \mathbf{Z}^\pi\right)_{\ell\in\{1,\ldots,L\}} \tag{4.11}$$

as $n \to \infty$ holds similarly as in the proof of Theorem 4.2, where $\mathbf{Z}^\pi \sim \mathcal{N}_k(\mathbf{0}_k, \mathbf{\Sigma}^\pi)$. Since the limiting distributions in (4.11) and (4.9) are generally not equal in distribution, the studentized permutation approach is not consistent for the multiple testing problem.
However, we can approximate the critical values via the groupwise bootstrap as introduced above. The difference here is that the covariance structures of the groups are not altered since the bootstrap observations are drawn within each group. The asymptotic validity is guaranteed by the following theorem.

Theorem 4.7. *Under Assumption 4.1, we have, as $n \to \infty$,*

$$(W_n^*(\mathbf{H}_\ell))_{\ell\in\{1,\ldots,L\}} \xrightarrow{d^*} \left((\mathbf{H}_\ell \mathbf{Z})'(\mathbf{H}_\ell \mathbf{\Sigma}\mathbf{H}_\ell')^+ \mathbf{H}_\ell \mathbf{Z}\right)_{\ell\in\{1,\ldots,L\}}.$$

Thus, we define the multiple groupwise bootstrap tests as

$$\varphi_\ell^* := \mathbb{1}\left\{W_n(\mathbf{H}_\ell, \mathbf{c}_\ell) > q^*_{\ell,1-\beta_n^*}\right\}, \quad \ell \in \{1, ..., L\}, \tag{4.12}$$

where $q^*_{\ell,1-\beta_n^*}, \beta_n^*$ denote the critical values and local level as in Section 2.3.2 for B_n Monte Carlo replicates of $(W_n^*(\mathbf{H}_\ell))_{\ell\in\{1,\ldots,L\}}$.
An analogous result can be found regarding the wild bootstrap.

Theorem 4.8. *Under Assumption 4.1, we have, as $n \to \infty$,*

$$(W_n^G(\mathbf{H}_\ell))_{\ell\in\{1,\ldots,L\}} \xrightarrow{d} \left((\mathbf{H}_\ell \mathbf{Z})'(\mathbf{H}_\ell \mathbf{\Sigma}\mathbf{H}_\ell')^+ \mathbf{H}_\ell \mathbf{Z}\right)_{\ell\in\{1,\ldots,L\}}$$

almost surely given the data $(\mathbf{X}, \boldsymbol{\delta})$.

The multiple wild bootstrap tests are given by

$$\varphi_\ell^G := \mathbb{1}\left\{W_n(\mathbf{H}_\ell, \mathbf{c}_\ell) > q^G_{\ell,1-\beta_n^G}\right\}, \quad \ell \in \{1, ..., L\}, \tag{4.13}$$

where $q^G_{\ell,1-\beta_n^G}, \beta_n^G$ denote the critical values and local level as in Section 2.3.2 for B_n Monte Carlo replicates of $(W_n^G(\mathbf{H}_\ell))_{\ell\in\{1,\ldots,L\}}$.
Hence, Theorem 2.6 and Lemma 2.3 provide that we obtain multiple tests for the bootstrap methods that control the family-wise error rate in the strong sense.

Theorem 4.9. *Under Assumption 4.1, the multiple groupwise bootstrap tests* (4.12) *as well as the mutiple wild bootstrap tests* (4.13) *fulfill* (2.20)–(2.22)*, respectively, i.e., they control the family-wise error rate asymptotically in the strong sense and are asymptotically balanced.*

Furthermore, by the methodologies in Section 2.3, we can construct more powerful multiple tests by using the closed testing procedure, cf. Remark 2.5, simultaneous confidence regions for $\mathbf{H}_\ell\boldsymbol{\mu}, \ell \in \{1, ..., L\}$, cf. Remark 2.6, that are

$$CR_{\ell,n} := \{\boldsymbol{\xi} \in \mathbb{R}^{r_\ell} \mid W_n(\mathbf{H}_\ell, \boldsymbol{\xi}) \leqslant q_{\ell,n}\}, \ell \in \{1, ..., L\},$$

with $q_{\ell,n}$ being one of $\chi^2_{\mathrm{rank}(\mathbf{H}_\ell),1-\beta_n}, q^*_{\ell,1-\beta^*_n}, q^G_{\ell,1-\beta^G_n}$, and adjusted p-values. In the case that $\mathbf{H}_\ell \in \mathbb{R}^{1\times k}$, i.e., $r_\ell = 1$, we can simplify the confidence regions to confidence intervals $CR_{n,\ell} := [L_{n,\ell}(\alpha/2), U_{n,\ell}(\alpha/2)]$ by solving the equation $W_n(\mathbf{H}_\ell, \xi) \leqslant q_{\ell,n}$ for $\xi \in \mathbb{R}$. This yields

$$L_{n,\ell}(\alpha/2) := \mathbf{H}_\ell\widehat{\boldsymbol{\mu}} - \frac{\sqrt{\mathbf{H}_\ell\widehat{\boldsymbol{\Sigma}}\mathbf{H}_\ell'}}{\sqrt{n}}\sqrt{q_{\ell,n}} \quad \text{and} \quad U_{n,\ell}(\alpha/2) := \mathbf{H}_\ell\widehat{\boldsymbol{\mu}} + \frac{\sqrt{\mathbf{H}_\ell\widehat{\boldsymbol{\Sigma}}\mathbf{H}_\ell'}}{\sqrt{n}}\sqrt{q_{\ell,n}}.$$

4.3.3 Counterexamples for Simultaneous Non-Inferiority and Equivalence Tests

In [58], we constructed simultaneous non-inferiority and equivalence tests. However, these procedures are only valid for $L = 1$ hypothesis and not for multiple hypotheses, as we will outline in the following two examples. Therefore, let us consider the case $r_\ell = 1$ for all $\ell \in \{1, ..., L\}$. In this special case, we write c_ℓ instead of $\mathbf{c}_\ell$ in non-bold type for all $\ell \in \{1, ..., L\}$. We defined simultaneous non-inferiority and equivalence tests by using the *two one-sided test procedure* [70]: let $\epsilon_1, ..., \epsilon_L > 0$ be prespecified equivalence bounds; the hypotheses of interest are

$$\mathcal{H}^i_{0,\ell} : \mathbf{H}_\ell\boldsymbol{\mu} - c_\ell \geqslant \epsilon_\ell \quad \text{vs.} \quad \mathcal{H}^i_{1,\ell} : \mathbf{H}_\ell\boldsymbol{\mu} - c_\ell < \epsilon_\ell, \qquad \text{for } \ell \in \{1, ..., L\} \tag{4.14}$$

for the non-inferiority testing problem and

$$\mathcal{H}^e_{0,\ell} : |\mathbf{H}_\ell\boldsymbol{\mu} - c_\ell| \geqslant \epsilon_\ell \quad \text{vs.} \quad \mathcal{H}^e_{1,\ell} : |\mathbf{H}_\ell\boldsymbol{\mu} - c_\ell| < \epsilon_\ell, \qquad \text{for } \ell \in \{1, ..., L\} \tag{4.15}$$

for the equivalence testing problem.
Let $q_{\ell,n}(2\alpha), \ell \in \{1, ..., L\}$, denote the used critical values at global level 2α in order to obtain critical values at level α for the one-sided testing problem. For each $\ell \in \{1, ..., L\}$, we reject $\mathcal{H}^i_{0,\ell}$ in (4.14) if and only if $U_{n,\ell}(\alpha) - c_\ell < \epsilon_\ell$. Furthermore, for each $\ell \in \{1, ..., L\}$, we reject $\mathcal{H}^e_{0,\ell}$ in (4.15) if and only if

$$U_{n,\ell}(\alpha) - c_\ell < \epsilon_\ell \quad \text{and} \quad L_{n,\ell}(\alpha) - c_\ell > -\epsilon_\ell.$$

However, these methods do not guarantee a family-wise error rate control of α for $L > 1$, not even in the weak sense, as the following two examples show.

Example 4.2 (Counterexample for the multiple non-inferiority tests)**.** *Let $L = 2$, $k = 1$, $\mathbf{H}_1 = 1, \mathbf{H}_2 = -1, \epsilon_1 = \epsilon_2 = 0$, $c_1 = \mu_1$, $c_2 = -\mu_1$, i.e., $\mathcal{H}^i_{0,1}$ and $\mathcal{H}^i_{0,2}$ are true. Furthermore, we denote $\widehat{\sigma}^2 := \widehat{\boldsymbol{\Sigma}} > 0$. Note that $\sqrt{n}(\widehat{\mu}_1 - \mu_1)/\widehat{\sigma} \xrightarrow{d} Z \sim \mathcal{N}(0,1)$ as $n \to \infty$ by Lemmas 4.1 and 4.2. Moreover, $q_{\ell,n}(2\alpha) \xrightarrow{P} z^2_{1-\alpha}$ by Remark 2.7 with $\beta = 2\alpha$, where $z^2_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of a $\mathcal{N}(0,1)$-distribution. Then, the family-wise error rate is given by*

$$\begin{aligned}
&P\left(\{U_{n,1}(\alpha) - c_1 < 0\} \cup \{U_{n,2}(\alpha) - c_2 < 0\}\right)\\
&= P\left(\left\{\widehat{\mu}_1 + \frac{\widehat{\sigma}}{\sqrt{n}}\sqrt{q_{\ell,n}(2\alpha)} - \mu_1 < 0\right\} \cup \left\{-\widehat{\mu}_1 + \frac{\widehat{\sigma}}{\sqrt{n}}\sqrt{q_{\ell,n}(2\alpha)} + \mu_1 < 0\right\}\right)\\
&= P\left(\left\{\sqrt{n}(\widehat{\mu}_1 - \mu_1)/\widehat{\sigma} < -\sqrt{q_{\ell,n}(2\alpha)}\right\} \cup \left\{\sqrt{n}(\widehat{\mu}_1 - \mu_1)/\widehat{\sigma} > \sqrt{q_{\ell,n}(2\alpha)}\right\}\right)\\
&\to P\left(\{Z < -z_{1-\alpha}\} \cup \{Z > z_{1-\alpha}\}\right) = 2\alpha > \alpha
\end{aligned}$$

as $n \to \infty$.

Example 4.3 (Counterexample for the multiple equivalence tests)**.** *Let $L = 2$, $k = 1$, $\mathbf{H}_1 = \mathbf{H}_2 = 1, \epsilon_1 = \epsilon_2 = 1$, $c_1 = \mu_1 - 1$, $c_2 = \mu_1 + 1$, i.e., $\mathcal{H}^e_{0,1}$ and $\mathcal{H}^e_{0,2}$ are true. Furthermore, we denote again $\widehat{\sigma}^2 := \widehat{\boldsymbol{\Sigma}} > 0$. Then, the* *family-wise error rate is given by*

$$
\begin{aligned}
&P\left(\{(U_{n,1}(\alpha)-c_1<\epsilon_1 \wedge L_{n,1}(\alpha)-c_1>-\epsilon_1)\} \cup \{(U_{n,2}(\alpha)-c_2<\epsilon_2 \wedge L_{n,2}(\alpha)-c_2>-\epsilon_2)\}\right)\\
&= P\left(\left\{\left(\widehat{\mu}_1+\frac{\widehat{\sigma}}{\sqrt{n}}\sqrt{q_{\ell,n}(2\alpha)}-\mu_1+1<1 \wedge \widehat{\mu}_1-\frac{\widehat{\sigma}}{\sqrt{n}}\sqrt{q_{\ell,n}(2\alpha)}-\mu_1+1>-1\right)\right\} \cup\right.\\
&\qquad \left.\left\{\left(\widehat{\mu}_1+\frac{\widehat{\sigma}}{\sqrt{n}}\sqrt{q_{\ell,n}(2\alpha)}-\mu_1-1<1 \wedge \widehat{\mu}_1-\frac{\widehat{\sigma}}{\sqrt{n}}\sqrt{q_{\ell,n}(2\alpha)}-\mu_1-1>-1\right)\right\}\right)\\
&= P\left(\left\{-2\frac{\sqrt{n}}{\widehat{\sigma}}+\sqrt{q_{\ell,n}(2\alpha)}<\sqrt{n}(\widehat{\mu}_1-\mu_1)/\widehat{\sigma}<-\sqrt{q_{\ell,n}(2\alpha)}\right\} \cup\right.\\
&\qquad \left.\left\{\sqrt{q_{\ell,n}(2\alpha)}<\sqrt{n}(\widehat{\mu}_1-\mu_1)/\widehat{\sigma}<2\frac{\sqrt{n}}{\widehat{\sigma}}-\sqrt{q_{\ell,n}(2\alpha)}\right\}\right)\\
&\to P\left(\{Z<-z_{1-\alpha}\} \cup \{Z>z_{1-\alpha}\}\right)=2\alpha>\alpha
\end{aligned}
$$

as $n \to \infty$ for $Z \sim \mathcal{N}(0,1)$.

4.4 Simulation Study

In order to analyze the small sample performance of our proposed methods, we conducted an extensive simulation study by using the computing environment R, version 4.2.1 [66].

4.4.1 Simulation Setup

The simulation setup is based on [24]. We simulated a factorial design with $k = 4$ groups and utilized the three different contrast matrices introduced in Example 4.1: the Dunnett-type, Tukey-type and Grand-mean-type contrast matrix. Here, the local hypotheses were constructed by the rows of the contrast matrix, i.e., the blocks $\mathbf{H}_1, ..., \mathbf{H}_R$ correspond to the rows of $\mathbf{H}$.
The survival times of the first three groups were always drawn from the same distribution. However, the survival distribution of the fourth group may differ. As in [24], the data were generated from the following survival distributions:

- Exponential distributions and early departures (*exp early*): $T_{11}, T_{21}, T_{31} \sim Exp(0.2)$ and T_{41} with piece-wise constant hazard function $t \mapsto \lambda_{\delta,1} \cdot \mathbb{1}\{t \leqslant 2\} + 0.2 \cdot \mathbb{1}\{t > 2\}$,
- exponential distributions and late departures (*exp late*): $T_{11}, T_{21}, T_{31} \sim Exp(0.2)$ and T_{41} with piece-wise constant hazard function $t \mapsto 0.2 \cdot \mathbb{1}\{t \leqslant 2\} + \lambda_{\delta,2} \cdot \mathbb{1}\{t > 2\}$,
- exponential distributions and proportional hazard alternative (*exp prop*): $T_{11}, T_{21}, T_{31} \sim Exp(0.2)$ and $T_{41} \sim Exp(\lambda_{\delta,3})$,
- lognormal distributions with scale alternatives (*logn*): $T_{11}, T_{21}, T_{31} \sim logN(2, 0.25)$ and $T_{41} \sim logN(\lambda_{\delta,4}, 0.25)$,
- exponential distributions and piece-wise exponential distributions (*pwExp*): $T_{11}, T_{21}, T_{31} \sim Exp(0.2)$ and T_{41} with piece-wise constant hazard function $t \mapsto 0.5 \cdot \mathbb{1}\{t \leqslant \lambda_{\delta,5}\} + 0.05 \cdot \mathbb{1}\{t > \lambda_{\delta,5}\}$,
- Weibull distributions and late departures (*Weib late*): $T_{11}, T_{21}, T_{31} \sim Weib(3, 8)$ and $T_{41} \sim Weib(3 \cdot \lambda_{\delta,6}, 8/\lambda_{\delta,6})$,
- Weibull distributions and proportional hazard alternative (*Weib prop*): $T_{11}, T_{21}, T_{31} \sim Weib(3, 8)$ and $T_{41} \sim Weib(3, \lambda_{\delta,7})$,
- Weibull distributions with crossing curves and scale alternatives (*Weib scale*): $T_{11}, T_{21}, T_{31} \sim Weib(3, 8)$ and $T_{41} \sim Weib(1.5, \lambda_{\delta,8})$,
- Weibull distributions with crossing curves and shape alternatives (*Weib shape*): $T_{11}, T_{21}, T_{31} \sim Weib(3, 8)$ and $T_{41} \sim Weib(\lambda_{\delta,9}, 14)$.

Here, the parameters $\lambda_{\delta,1}, ..., \lambda_{\delta,9}$ were determined such that the RMST difference equals $\delta = \mu_1 - \mu_4$. This difference was set to $\delta = 0$ for simulating under the null and to $\delta = 1.5$ for simulating under the alternative hypothesis.
Note that, under the null hypothesis, the scenarios *exp early*, *exp late* and *exp prop* as well as *Weib late* and *Weib prop* are respectively equal. Consequently, we only included the results for these scenarios once in the figures and tables, respectively. This is done by calculating the mean over the results.
For the censoring times, we chose the following three scenarios:

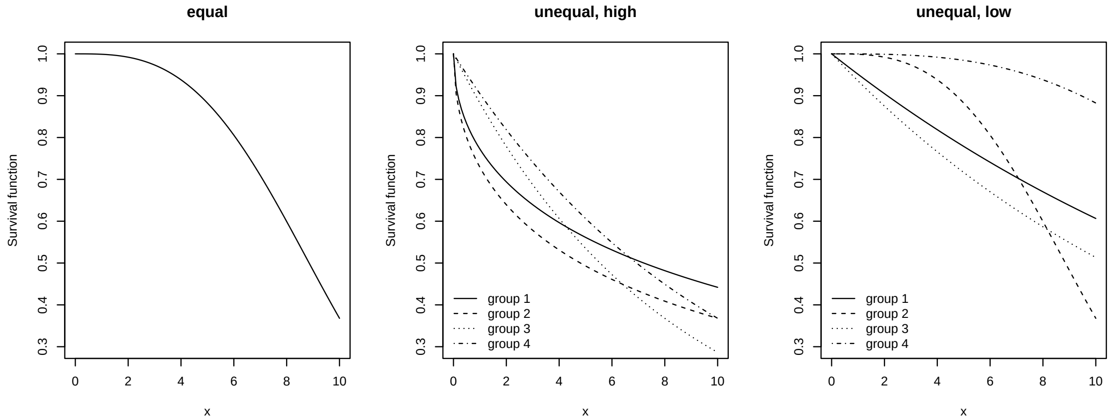


Figure 4: Survival functions of the censoring times.

- equally Weibull distributed censoring times (*equal*): $C_{11}, C_{21}, C_{31}, C_{41} \sim Weib(3, 10)$,
- unequally Weibull distributed censoring times with high censoring rates (*unequal, high*): $C_{11} \sim Weib(0.5, 15), C_{21} \sim Weib(0.5, 10), C_{31} \sim Weib(1, 8)$ and $C_{41} \sim Weib(1, 10)$,
- unequally Weibull distributed censoring times with low censoring rates (*unequal, low*): $C_{11} \sim Weib(1, 20), C_{21} \sim Weib(3, 10), C_{31} \sim Weib(1, 15)$ and $C_{41} \sim Weib(3, 20)$.

The survival functions of these censoring times are illustrated in Figure 4. The resulting censoring rates of the different groups are presented in Table 4 in the appendix. The censoring rates ranged from 20% up to 60% in groups 1-3 and from 1% up to 57% in group 4.
We considered *balanced* and *unbalanced* designs with sample sizes $\mathbf{n} = (n_1, n_2, n_3, n_4) = K \cdot (15, 15, 15, 15)$ and $\mathbf{n} = K \cdot (10, 20, 10, 20)$, where $K \in \{1, 2, 4\}$ for *small, medium* and *large* samples.
Furthermore, $N_{sim} = 5000$ simulation runs with $B = 1999$ resampling iterations were generated. The level of significance was set to $\alpha = 0.05$ and the upper integration bound to $\tau = 10$.
The following methods were compared:

- *asymptotic_global*: The global Wald-type test as in Section 4.2.1,
- *permutation*: The global studentized permutation test as in Section 4.2.2,
- *asymptotic*: Multiple asymptotic Wald-type tests as in Section 4.3.1,
- *wild, Rademacher*; *wild, Gaussian*: Multiple wild bootstrap tests as in Section 4.3.2 with Rademacher and Gaussian multipliers, respectively,
- *groupwise*: The multiple groupwise bootstrap test as in Section 4.3.2,
- *asymptotic_bonf*: Global Wald-type tests as in Section 4.2.1 adjusted with the Bonferroni-correction,
- *permutation_bonf*: Global studentized permutation tests as in Section 4.2.2 adjusted with the Bonferroni-correction.

Clearly, the first two methods (*asymptotic_global*, *permutation*) can only be compared to multiple testing procedures for the global testing problem. However, by using a Bonferroni-correction (*asymptotic_bonf*, *permutation_bonf*), we can also obtain test decisions for the local hypotheses.

4.4.2 Simulation Results under the Null Hypothesis

Figures 5 to 7 under $\mathcal{H}_0$ illustrate the global rejection rates, which coincide with the family-wise error rates for the multiple tests, over all settings for the different contrast matrices. Here, the dotted line represents the α-level of 0.05 and the dashed lines represent the borders of the binomial confidence interval $[0.044, 0.0562]$.
In all figures, it can be seen that only the permutation approach and the groupwise bootstrap seem to perform well over all simulation settings. Here, the permutation approach yields slightly better values than the groupwise bootstrap. Tables S1 to S36 on GitHub (`https://github.com/MerleMunko/supplement_thesis`) show the global rejection rates of the different settings. Under the null hypothesis, all values in the binomial confidence interval are printed in bold type. The permutation method is exact under exchangeability and, thus, most of the values of the permutation method with equal survival distributions across the groups under the null (*exp*

early, exp late, exp prop, logn, Weib late, Weib prop*) and *equal* censoring distributions fall within that interval. Furthermore, when exchangeability is violated, the permutation method still seems to perform quite accurately in terms of type I error control for all sample sizes. The groupwise bootstrap approach also results in very accurate family-wise error rates, especially for medium and large sample sizes. Moreover, we note that the three asymptotic approaches (*asymptotic_global, asymptotic, asymptotic_bonf*) and the wild bootstrap approaches are too liberal, as they exhibit too high rejection rates in nearly all settings. In Figures 15 to 17 in the appendix, it is observable that these methods exceed the desired level of significance particularly for settings with small sample sizes. By further analyzing the tables in the appendix, we observe that high censoring rates amplify the liberality of the tests. Note that the highest rejection rates occur for small sample size settings, where at least 49% of the data is censored.

It should be noted that the power of our multiple tests can be improved by using a stepwise procedure as described in Section 2.3. The power of the Bonferroni corrected methods can also be improved by a stepwise procedure, e.g., the Holm-correction [42]. However, stepwise procedures cannot be used for the construction of confidence regions and, hence, we did not focus on these in the simulation study.

We have proven that all approaches are asymptotically valid under the null hypothesis. Figures 15 to 17 in the appendix confirm this empirically: all methods seem to tend to the desired level of significance of 0.05 for increasing sample sizes. However, the convergence rates of the asymptotic and the wild bootstrap approaches appear to be very slow. This observation prompts an inquiry into analyzing how larger sample sizes might influence the type I error control for the naive methods, that are the three asymptotic approaches. Therefore, further simulations under the null hypothesis were conducted in Section C.2 in the appendix. Specifically, we increased the scaling factor for sample sizes, that is $K \in \{6, 8, 10\}$, resulting in sample sizes ranging from 60 to 200 per group.

4.4.3 Simulation Results under the Alternative Hypothesis

In the power assessment, we observed small differences between the different methods. The global asymptotic approach leads to the highest power in most settings, followed by the wild bootstrap with Gaussian and with Rademacher multipliers. However, in view of the bad type I error control of these methods, we cannot recommend their use.

Let us now review the multiple testing problem. Because of the bad type I error control of the wild bootstrap approaches and for the sake of clarity, we did not consider this method in the following. Moreover, the global approaches (*asymptotic global* and *permutation*) do not yield local decisions. Thus, we only compared the asymptotic, the groupwise bootstrap and the Bonferroni-corrected approaches for the multiple testing problem. Furthermore, only the settings under the alternative hypothesis are considered. Tables S37 to S54 on GitHub (`https://github.com/MerleMunko/supplement_thesis`) provide the rejection rates of the false local

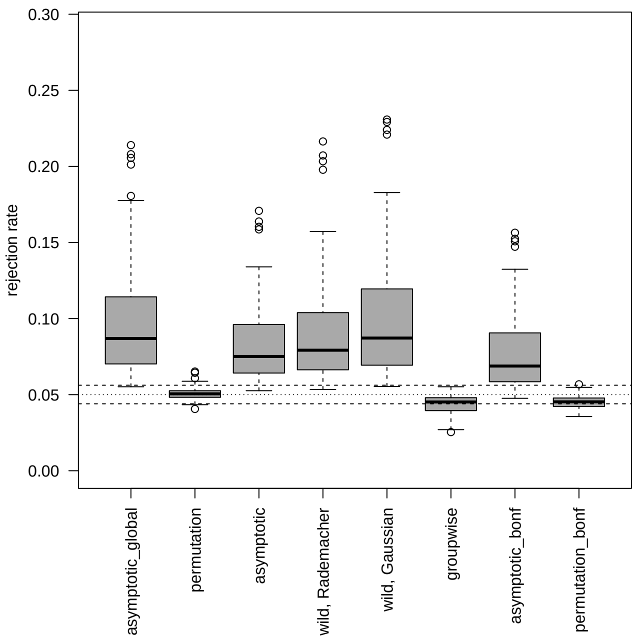


Figure 5: Rejection rates under $\mathcal{H}_0$ over all settings for the Dunnett-type contrast matrix. The dashed lines represent the borders of the binomial confidence interval $[0.044, 0.0562]$.

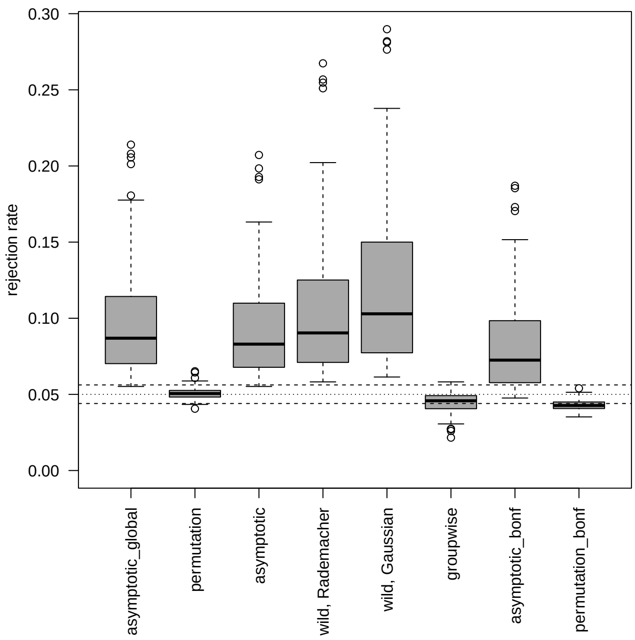


Figure 6: Rejection rates under $\mathcal{H}_0$ over all settings for the Tukey-type contrast matrix. The dashed lines represent the borders of the binomial confidence interval $[0.044, 0.0562]$.

hypotheses across all settings for the different sample sizes; they are further illustrated in Figures 18 to 20 in the appendix. Therein, it is apparent that the asymptotic approaches have a higher power for each false hypothesis than the groupwise bootstrap and the studentized permutation approach with the Bonferroni-correction. However, this difference is rather small, especially for large sample sizes. Additionally, by comparing the empirical power of the groupwise bootstrap test and of the studentized permutation test with Bonferroni-correction, the groupwise bootstrap test tends to be slightly more powerful for medium and large sample sizes. For small sample sizes, this trend reverses for the Dunnett-type and Tukey-type contrast matrix. However, it is important to note that the differences between the two methods regarding the empirical power are quite small and mainly not even visible in Figures 18 to 20.

Nevertheless, it is well-known that the Bonferroni-correction might lead to a loss of power [47]. In order to illustrate this, we conducted an additional simulation study under non-exchangeability; see Section C.3 in the appendix for details. Here, we saw that the groupwise bootstrap approach is able to outperform the permutation approach with Bonferroni-corrections in specific scenarios under non-exchangeability. This effect becomes particularly observable for the Tukey-type contrast matrix, where six hypotheses are tested simultaneously.

We conducted further investigations in order to assess the impact of censoring and sample sizes on the power. As expected, the power increases for larger sample sizes for each method. Additionally, settings with lower censoring rates tend to be more powerful. When comparing the power between the three false hypotheses $\mathcal{H}_{0,3}, \mathcal{H}_{0,5}$ and $\mathcal{H}_{0,6}$ of the Tukey-type contrast matrix, it becomes apparent that the fifth hypothesis $\mathcal{H}_{0,5}$ can be rejected more often, see, e.g., Figure 19. The reason behind this can be attributed to the unequal sample sizes in the unbalanced design: Groups 1 and 3 contain only $K \cdot 10$ observations, respectively, while groups 2 and 4 contain $K \cdot 20$ observations each, for $K \in \{1, 2, 4\}$. Consequently, when comparing the RMSTs of groups 2 and 4, we have a larger dataset compared to other pairwise comparisons leading to more power. This exemplifies how an unbalanced design can boost the power of specific local hypotheses. However, depending on the contrast matrix, this is often done at the cost of a reduced power for testing other local hypotheses.

It should be noted that the empirical power is very low in some scenarios. This is particularly the case when considering the groupwise bootstrap and the studentized permutation approach with Bonferroni-correction and small sample sizes. Moreover, an increasing number of hypotheses decreases the power for the local hypotheses in general. Consequently, multiple tests based on the Tukey-type contrast matrix have even less power than multiple tests based on the Dunnett-type contrast matrix. Furthermore, small differences to the null hypothesis are difficult to detect. This can be observed for the Grand-mean-type contrast matrix, see Figure 20 in the appendix, where the three null hypotheses $\mathcal{H}_{0,1} : \mu_1 = \overline{\mu}, \mathcal{H}_{0,2} : \mu_2 = \overline{\mu}$, and $\mathcal{H}_{0,3} : \mu_3 = \overline{\mu}$ have very low rejection rates under the alternative hypothesis due to a small difference of $\mu_i - \overline{\mu} = \delta/4 = 3/8$ for $i \in \{1, 2, 3\}$.

In conclusion, we recommend to use the studentized permutation method for the global testing problem. For the multiple testing problem, the groupwise bootstrap test and the studentized permutation method with Bonferroni-correction perform similarly and quite well in terms of the type I error control and the empirical

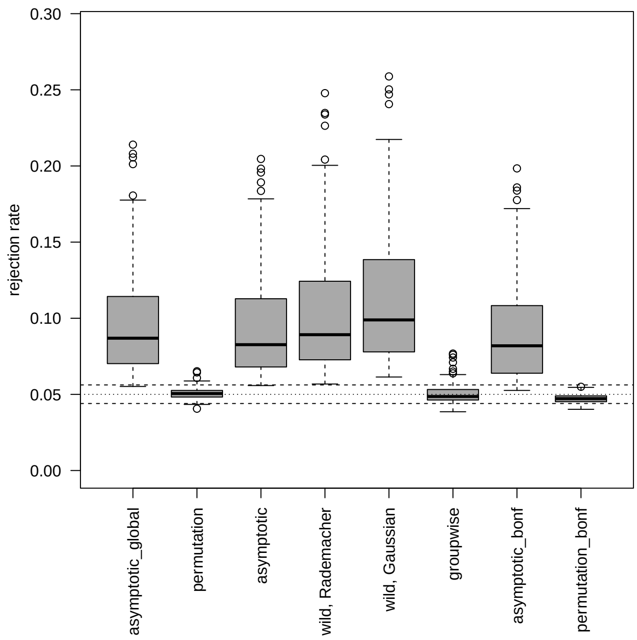


Figure 7: Rejection rates under $\mathcal{H}_0$ over all settings for the Grand-mean-type contrast matrix. The dashed lines represent the borders of the binomial confidence interval $[0.044, 0.0562]$.

power across all simulation scenarios. However, we recommend to use the groupwise bootstrap test for testing a large number of hypotheses since the Bonferroni-correction is known to have a lower power in this case [47].

4.5 Data Example about the Occurrence of Hay Fever

In order to illustrate our novel methods on real data, we consider a data set with data about the occurrence of hay fever of boys and girls with and without contact to farming environments [37, 38]. These data derive from an observational study and may be structured in a factorial 2-by-2 design: factor A represents whether the child was growing up on a farm; factor B represents the sex. The event of interest is the age at which hay fever occurred. Ties are present in the data as each measured age was rounded (down) to full years.
The children were included in the survey via primary schools in 2006. Hence, their age has been mainly between six and ten years at the beginning of the study. The medical diagnoses of hay fever together with the age at initial diagnosis *before* study entry were recorded retrospectively. The age at which the diagnosis was made is easy to remember so that no significant recall bias or inaccuracies were assumed here. Follow-up surveys took place in 2010 with retrospective recording of initial diagnoses since the last survey and from then on annually until 2016. For simultaneous testing on a main effect of the two factors as well as on an interaction effect, we define $\mathbf{H} := [\mathbf{H}_A', \mathbf{H}_B', \mathbf{H}_{AB}']'$ by using the notation of Section 4.2. Furthermore, we set $\alpha = 0.05$ as the level of significance and chose $\tau = 15$ years.
The data set consists of 2234 participants. In detail, 654 boys and 649 girls not growing up on a farm and 450 boys and 481 girls growing up on farms were observed. Note that we did not adjust for any confounding variables in order to simplify this application of our method to real data. This comes with the limitation that the results may not fully reflect the causal effects of sex or growing up on a farm on the incidence of hay fever. The censoring rates in the different groups ranged from 74% up to 93%. The Kaplan-Meier and Nelson-Aalen curves of all groups are illustrated in Figure 8. Here, it can be seen that the estimated cumulative hazard functions are crossing each other and, thus, the proportional hazards assumption is not justified. If we would perform a Cox proportional hazards model nevertheless, the resulting (unadjusted) p-values of the existence of an impact on the occurrence of hay fever are $p_A < 10^{-8}$ for a main effect of factor A, $p_B = 0.112$ for a main effect of factor B and $p_{AB} = 0.235$ for an interaction effect. By using a Bonferroni- or Holm-correction of the p-values, we could only establish that factor A (growing up on a farm) has a main effect on the occurrence of hay fever at global level 0.05.
However, since the proportional hazards assumption seems violated, we aimed to compare the RMSTs in the different groups. The estimated RMSTs respectively are 14.22 and 14.66 for boys and girls growing up on farms and 13.59 and 13.79 for boys and girls not growing up on a farm. This indicates that boys tend to be more prone to hay fever than girls until the age of 15. Furthermore, growing up on a farm seems to reduce the risk of getting hay fever until the age of 15. Performing the global asymptotic Wald-type test and its

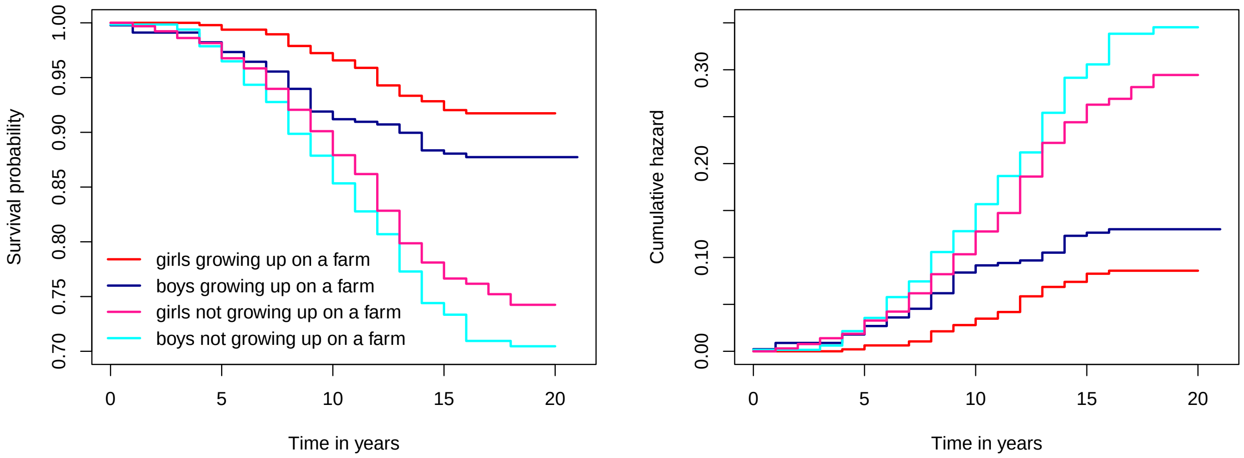


Figure 8: Kaplan-Meier and Nelson-Aalen curves of the different groups

Test	asymptotic	wild Rademacher	wild Gaussian	groupwise	asymptotic bonf	permutation bonf
Farm	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Sex	0.005	0.006	0.006	0.007	0.006	0.006
Interaction	0.605	0.599	0.603	0.597	0.811	0.800

Table 1: Adjusted p-values for the data example

global studentized permutation version with $B = 19999$ resampling iterations leads to p-values of $p < 0.003$ and, thus, the existence of at least one main or the interaction effect on the occurrence of hay fever is highly significant. However, these tests cannot provide the information whether there is a significant difference of hay fever occurrence between the groups regarding the sex and/or growing up on a farm and/or an interaction effect. Therefore, we applied multiple testing procedures. The resulting adjusted p-values of our proposed methods with $B = 19999$ resampling iterations are shown in Table 1. The p-values of the global asymptotic and permutation approach were adjusted by a Bonferroni-correction for enabling local test decisions. Here, we found that all methods rejected the local hypotheses of no main effect of the two factors simultaneously at the $\alpha = 0.05$ level. However, the interaction effect of the two factors was not significant.
The data from this example do not fit perfectly to the simulation design in Section 4.4 since, here, a 2-by-2 design with different hypothesis matrices and larger sample sizes and censoring rates is considered. Thus, additional simulation results inspired by this data example can be found in Section C.4.

4.6 Proofs of Section 4

Proof of Theorem 4.1 By Lemma S.1 in the supplement of [24], it holds

$$\sqrt{n}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \xrightarrow{d} \mathcal{N}_k(\mathbf{0}_k, \boldsymbol{\Sigma}) \tag{4.16}$$

as $n \to \infty$. Moreover, we have

$$\widehat{\sigma}_i^2 \xrightarrow{a.s.} \sigma_i^2 \tag{4.17}$$

as $n \to \infty$ for all $i \in \{1, ..., k\}$ by Section S.5 in the supplement of [24] under Assumption 4.1. Due to $P(T_{i1} < \tau) > 0$, it holds $\sigma_i^2 > 0$ for all $i \in \{1, ..., k\}$. Hence, it follows $\operatorname{rank}(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}') = \operatorname{rank}(\mathbf{H}\boldsymbol{\Sigma}^{1/2}) = \operatorname{rank}(\mathbf{H})$ and, analogously, $P(\operatorname{rank}(\mathbf{H}\widehat{\boldsymbol{\Sigma}}\mathbf{H}') \neq \operatorname{rank}(\mathbf{H})) \to 0$ for $n \to \infty$. Consequently, the Moore-Penrose inverse $(\mathbf{H}\widehat{\boldsymbol{\Sigma}}\mathbf{H}')^+$ converges in probability to $(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}')^+$. By Slutsky's lemma and Theorem 9.2.2 in [67], it follows

$$W_n(\mathbf{H}, \mathbf{c}) \xrightarrow{d} \chi^2_{\operatorname{rank}(\mathbf{H})}$$

as $n \to \infty$ under the null hypothesis in (4.2). $\square$

Proof of Theorem 4.2 First of all, we introduce some notation. Let $\nu(t) := \sum_{i=1}^k \kappa_i \nu_i(t)$, $y(t) := \sum_{i=1}^k \kappa_i y_i(t)$, $A(t) := \int_{[0,t]} 1/y \, \mathrm{d}\nu$ and $S(t) := \prod_{s\in[0,t]} (1 - \mathrm{d}A(s))$ for all $t \geqslant 0$. Moreover, let $\widehat{S}$ denote the Kaplan-Meier estimator of the pooled survival function S, see [24] for details, and $\widehat{\mu} := \int_0^\tau \widehat{S}(t) \, \mathrm{d}t$ denote the estimator regarding the RMST of the pooled sample.
As in the proof of Lemma S.2 in the supplement of [24], it holds

$$\sqrt{n}(\widehat{\boldsymbol{\mu}}^\pi - \widehat{\mu}\mathbf{1}_k) \xrightarrow{d^*} \mathcal{N}_k(\mathbf{0}_k, \mathbf{\Sigma}^\pi)$$

as $n \to \infty$, where

$$(\mathbf{\Sigma}^\pi)_{ii'} := \left(\frac{1}{\kappa_i}\mathbb{1}\{i = i'\} - 1\right)\sigma^{\pi 2}$$

for all $i, i' \in \{1, ..., k\}$ and

$$\sigma^{\pi 2} := \int_{[0,\tau)} \left(\int_x^\tau S(t) \, \mathrm{d}t\right)^2 \frac{1}{(1 - \Delta A(x))y(x)} \, \mathrm{d}A(x).$$

Moreover, in the proof of Lemma S.3 of the supplement of [24], it was shown that

$$\widehat{\sigma}_i^{\pi 2} \xrightarrow{P} {\kappa_i}^{-1}\sigma^{\pi 2}$$

as $n \to \infty$ for all $i \in \{1, ..., k\}$ under $P(X_{i1} \geqslant \tau) > 0$. Hence, it follows

$$\widehat{\mathbf{\Sigma}}^\pi \xrightarrow{P} \operatorname{diag}\left({\kappa_1}^{-1}\sigma^{\pi 2}, ..., {\kappa_k}^{-1}\sigma^{\pi 2}\right) = \mathbf{\Sigma}^\pi + \sigma^{\pi 2}\mathbf{J}_k$$

as $n \to \infty$. Since $\mathbf{H}\mathbf{J}_k = \mathbf{0}_{r\times k}$, it holds $\mathbf{H}\widehat{\mathbf{\Sigma}}^\pi \xrightarrow{P} \mathbf{H}\mathbf{\Sigma}^\pi$ as $n \to \infty$. Moreover, $P(T_{i1} < \tau) > 0$ implies $\sigma^{\pi 2} > 0$ and, thus, $P(\widehat{\sigma}_i^{\pi 2} > 0) \to 1$ as $n \to \infty$. Consequently, we have $(\mathbf{H}\widehat{\mathbf{\Sigma}}^\pi\mathbf{H}')^+ \xrightarrow{P} (\mathbf{H}\mathbf{\Sigma}^\pi\mathbf{H}')^+$ as $n \to \infty$. Hence, it follows by Slutsky's lemma and Theorem 9.2.2 in [67]

$$\begin{aligned} W_n^\pi(\mathbf{H}) &= n(\mathbf{H}\widehat{\boldsymbol{\mu}}^\pi)'(\mathbf{H}\widehat{\mathbf{\Sigma}}^\pi\mathbf{H}')^+\mathbf{H}\widehat{\boldsymbol{\mu}}^\pi \\ &= n[\mathbf{H}(\widehat{\boldsymbol{\mu}}^\pi - \widehat{\mu}\mathbf{1}_k)]'(\mathbf{H}\widehat{\mathbf{\Sigma}}^\pi\mathbf{H}')^+\mathbf{H}(\widehat{\boldsymbol{\mu}}^\pi - \widehat{\mu}\mathbf{1}_k) \xrightarrow{d^*} \chi^2_{\operatorname{rank}(\mathbf{H})} \end{aligned}$$

as $n \to \infty$. □

Proof of Theorem 4.3 For proving Theorem 4.3, let $\widehat{S}_i^*, \widehat{A}_i^*, \widehat{\sigma}_i^{*2}, Y_i^*$ and N_i^* denote the groupwise bootstrap counterparts of $\widehat{S}_i, \widehat{A}_i, \widehat{\sigma}_i^2, Y_i$ and N_i, respectively, for all $i \in \{1, ..., k\}$.

Lemma 4.1. *Under Assumption 4.1, we have*

$$\sqrt{n}(\widehat{\boldsymbol{\mu}}^* - \widehat{\boldsymbol{\mu}}) \xrightarrow{d^*} \mathbf{Z} = (Z_1, ..., Z_k)' \sim \mathcal{N}_k(\mathbf{0}_k, \mathbf{\Sigma}) \tag{4.18}$$

as $n \to \infty$.

Proof of Lemma 4.1. By Section A.2, it holds

$$\sqrt{n_i}(\widehat{S}_i^* - \widehat{S}_i) \xrightarrow{d^*} \mathbb{U}_i \sim GP(0, \Gamma_i)$$

on $D[0, \tau)$ as $n \to \infty$, where

$$\Gamma_i : [0, \tau)^2 \ni (t, s) \mapsto S_i(t)S_i(s) \int_{[0,\min\{t,s\}]} \frac{1}{(1 - \Delta A_i(x))y_i(x)} \, \mathrm{d}A_i(x) \in \mathbb{R}$$

for all $i \in \{1, ..., k\}$ and $GP(0, \Gamma_i)$ denotes a centered Gaussian process with covariance function Γ_i. Since the samples are independent, it follows

$$\sqrt{n}(\widehat{S}_i^* - \widehat{S}_i)_{i\in\{1,...,k\}} \xrightarrow{d^*} \mathbb{G}^* = \left(\frac{1}{\sqrt{\kappa_i}}\mathbb{U}_i\right)_{i\in\{1,...,k\}} \sim GP_k\left(\mathbf{0}_k, \operatorname{diag}\left(\frac{1}{\kappa_1}\Gamma_1, ..., \frac{1}{\kappa_k}\Gamma_k\right)\right)$$

on $D[0, \tau)^k$ as $n \to \infty$ by (4.1), where $GP_k(\mathbf{0}_k, \mathbf{D})$ denotes a k-dimensional centered Gaussian process with covariance function $\mathbf{D} : [0, \tau)^2 \to \mathbb{R}^{k\times k}$. Hence, the continuous mapping theorem provides

$$\sqrt{n}(\widehat{\boldsymbol{\mu}}^* - \widehat{\boldsymbol{\mu}}) \xrightarrow{d^*} \int_0^\tau \mathbb{G}^*(t) \, \mathrm{d}t = \mathbf{Z}$$

as $n \to \infty$. The limiting variable $\mathbf{Z}$ is normally distributed as linear transformation of a Gaussian process and its moments can be calculated by using Fubini's theorem. Thus, we get $\mathrm{E}[\mathbf{Z}] = \mathbf{0}_k$ and $\mathbb{C}ov(\mathbf{Z}) = \mathbf{\Sigma}$. □

Lemma 4.2. *Under Assumption 4.1, we have, as $n \to \infty$,*

$$\widehat{\mathbf{\Sigma}}^* \xrightarrow{P} \mathbf{\Sigma}. \tag{4.19}$$

Proof of Lemma 4.2. Let $i \in \{1, ..., k\}$ be arbitrary. Similarly as in the supplement of [31], we consider the $P^{(X_{i1},\delta_{i1})}$-Donsker classes

$$\mathcal{F}_1 := \{(x,d) \mapsto \mathbb{1}\{x \leqslant t, d = 1\} \mid t \in [0,\tau]\} \text{ and } \mathcal{F}_2 := \{(x,d) \mapsto \mathbb{1}\{x \geqslant t\} \mid t \in [0,\tau]\}$$

with finite envelope function $F \equiv 1$. By Theorem 3.7.1 in [74] and Slutsky's lemma, we obtain

$$\sup_{t\in[0,\tau]} \left| \frac{1}{n_i} Y_i^*(t) - \frac{1}{n_i} Y_i(t) \right| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{t\in[0,\tau]} \left| \frac{1}{n_i} N_i^*(t) - \frac{1}{n_i} N_i(t) \right| \xrightarrow{P} 0$$

as $n \to \infty$. Section S.6 in the supplement of [24] provides

$$\sup_{t\in[0,\tau]} \left| \frac{1}{n_i} Y_i(t) - y_i(t) \right| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{t\in[0,\tau]} \left| \frac{1}{n_i} N_i(t) - \nu_i(t) \right| \xrightarrow{P} 0$$

as $n \to \infty$. It follows

$$\sup_{t\in[0,\tau]} \left| \widehat{S}_i^*(t) - S_i(t) \right| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{t\in[0,\tau]} \left| \widehat{A}_i^*(t) - A_i(t) \right| \xrightarrow{P} 0$$

as $n \to \infty$ under $P(X_{i1} \geqslant \tau) > 0$. Hence, we have $\widehat{\sigma}_i^{*2} \xrightarrow{P} \sigma_i^2$ as $n \to \infty$. Since $i \in \{1, ..., k\}$ was arbitrary, (4.19) follows. ☐

Now, the statement of Theorem 4.3 follows with similar arguments as in the proofs of Theorem 4.1 and 4.2 by Lemma 4.1 and 4.2. In doing so, we apply Slutsky's lemma and Theorem 9.2.2 in [67] again. ☐

Proof of Theorem 4.4

Lemma 4.3. *Under Assumption 4.1, we have*

$$\sqrt{n}\widehat{\boldsymbol{\mu}}^G \xrightarrow{d} \mathbf{Z} = (Z_1, ..., Z_k)' \sim \mathcal{N}_k(\mathbf{0}_k, \mathbf{\Sigma})$$

almost surely as $n \to \infty$ given the data $(\mathbf{X}, \boldsymbol{\delta})$.

Proof of Lemma 4.3. Let $i \in \{1, ..., k\}$ be arbitrary. We aim to apply the Lindeberg-Feller theorem with

$$Z_{j,n_i} := \sqrt{n_i} G_{ij} \int_{[0,\tau)} \int_x^\tau \widehat{S}_i(t) \,\mathrm{d}t \frac{1}{\sqrt{(Y_i(x) - \Delta N_i(x))Y_i(x)}} \,\mathrm{d}N_{ij}(x)$$

for all $j \in \{1, ..., n_i\}$. Then, $Z_{1,n_i}, ..., Z_{n_i,n_i}$ are independent conditionally on $(\mathbf{X}, \boldsymbol{\delta})$. Moreover, we have

$$\mathrm{E}\left[Z_{j,n_i} \mid \mathbf{X}, \boldsymbol{\delta}\right] = \sqrt{n_i} \mathrm{E}\left[G_{ij} \mid \mathbf{X}, \boldsymbol{\delta}\right] \int_{[0,\tau)} \int_x^\tau \widehat{S}_i(t) \,\mathrm{d}t \frac{1}{\sqrt{(Y_i(x) - \Delta N_i(x))Y_i(x)}} \,\mathrm{d}N_{ij}(x) = 0$$

almost surely. It should be noted that all following statements about conditional expectations hold just almost surely but we will not always add this throughout, for the sake of clarity. Furthermore, it holds

$$\begin{aligned}
s_n^2 := \sum_{j=1}^{n_i} \mathbb{V}ar\left(Z_{j,n_i} \mid \mathbf{X}, \boldsymbol{\delta}\right) &= n_i \sum_{j=1}^{n_i} \mathbb{V}ar\left(G_{ij} \mid \mathbf{X}, \boldsymbol{\delta}\right) \left(\int_{[0,\tau)} \int_x^\tau \widehat{S}_i(t) \,\mathrm{d}t \frac{1}{\sqrt{(Y_i(x) - \Delta N_i(x))Y_i(x)}} \,\mathrm{d}N_{ij}(x) \right)^2 \\
&= n_i \int_{[0,\tau)} \left(\int_x^\tau \widehat{S}_i(t) \,\mathrm{d}t \right)^2 \frac{1}{(Y_i(x) - \Delta N_i(x))Y_i(x)} \,\mathrm{d}N_i(x) \\
&= \frac{n_i}{n} \widehat{\sigma}_i^2 \xrightarrow{a.s.} \kappa_i \sigma_i^2
\end{aligned}$$

as $n \to \infty$ under $P(X_{i1} \geqslant \tau) > 0$ by Section S.5 in the supplement of [24]. For showing Lindeberg's condition, let $\varepsilon > 0$ be arbitrary. Then, we have

$$\begin{aligned}
&\frac{1}{s_n^2} \sum_{j=1}^{n_i} \mathrm{E}\left[Z_{j,n_i}^2 \mathbb{1}\left\{Z_{j,n_i}^2 > \varepsilon^2 s_n^2\right\} \mid \mathbf{X}, \boldsymbol{\delta}\right] \\
&= \frac{n}{n_i \widehat{\sigma}_i^2} \sum_{j=1}^{n_i} n_i \mathrm{E}\left[G_{ij}^2 \mathbb{1}\left\{n_i G_{ij}^2 \int_{[0,\tau)} \frac{\left(\int_x^\tau \widehat{S}_i(t)\, \mathrm{d}t\right)^2}{(Y_i(x) - \Delta N_i(x)) Y_i(x)} \, \mathrm{d}N_{ij}(x) > \varepsilon^2 \frac{n_i}{n} \widehat{\sigma}_i^2\right\} \mid \mathbf{X}, \boldsymbol{\delta}\right] \cdot \\
&\quad \int_{[0,\tau)} \frac{\left(\int_x^\tau \widehat{S}_i(t)\, \mathrm{d}t\right)^2}{(Y_i(x) - \Delta N_i(x)) Y_i(x)} \, \mathrm{d}N_{ij}(x) \\
&\leqslant \frac{n}{\widehat{\sigma}_i^2} \sum_{j=1}^{n_i} \mathrm{E}\left[G_{ij}^2 \mathbb{1}\left\{G_{ij}^2 \sup_{x \in [0,\tau)} \left\{\frac{\left(\int_x^\tau \widehat{S}_i(t)\, \mathrm{d}t\right)^2}{(Y_i(x) - \Delta N_i(x)) Y_i(x)}\right\} > \varepsilon^2 \frac{1}{n} \widehat{\sigma}_i^2\right\} \mid \mathbf{X}, \boldsymbol{\delta}\right] \cdot \\
&\quad \int_{[0,\tau)} \frac{\left(\int_x^\tau \widehat{S}_i(t)\, \mathrm{d}t\right)^2}{(Y_i(x) - \Delta N_i(x)) Y_i(x)} \, \mathrm{d}N_{ij}(x) \\
&= \mathrm{E}\left[G_{i1}^2 \mathbb{1}\left\{G_{i1}^2 \sup_{x \in [0,\tau)} \left\{\frac{\left(\int_x^\tau \widehat{S}_i(t)\, \mathrm{d}t\right)^2}{(n_i^{-1} Y_i(x) - n_i^{-1} \Delta N_i(x)) n_i^{-1} Y_i(x)}\right\} > \varepsilon^2 \frac{n_i^2}{n} \widehat{\sigma}_i^2\right\} \mid \mathbf{X}, \boldsymbol{\delta}\right] \\
&\xrightarrow{a.s.} 0
\end{aligned}$$

as $n \to \infty$ by the dominated convergence theorem with integrable majorant G_{i1}^2, where the last equality follows from the definition of $\widehat{\sigma}_i^2$. Here, we use that

$$\sup_{x \in [0,\tau)} \left|n_i^{-1} Y_i(x) - y_i(x)\right| \xrightarrow{a.s.} 0, \quad \sup_{x \in [0,\tau)} \left|n_i^{-1} N_i(x) - \nu_i(x)\right| \xrightarrow{a.s.} 0 \quad \text{and} \quad \widehat{\sigma}_i^2 \xrightarrow{a.s.} \sigma_i^2 \tag{4.20}$$

as $n \to \infty$ holds under $P(X_{i1} \geqslant \tau) > 0$ by Section S.5 and S.6 in the supplement of [24] such that

$$\begin{aligned}
&P\left(\mathbb{1}\left\{G_{i1}^2 \sup_{x \in [0,\tau)} \left\{\frac{\left(\int_x^\tau \widehat{S}_i(t)\, \mathrm{d}t\right)^2}{(n_i^{-1} Y_i(x) - n_i^{-1} \Delta N_i(x)) n_i^{-1} Y_i(x)}\right\} > \varepsilon^2 \frac{n_i^2}{n} \widehat{\sigma}_i^2\right\} > \varepsilon \mid (\mathbf{X}, \boldsymbol{\delta})\right) \\
&\leqslant P\left(G_{i1}^2 \sup_{x \in [0,\tau)} \left\{\frac{\tau^2}{(n_i^{-1} Y_i(x) - n_i^{-1} \Delta N_i(x)) n_i^{-1} Y_i(x)}\right\} > \varepsilon^2 \frac{n_i^2}{n} \widehat{\sigma}_i^2 \mid (\mathbf{X}, \boldsymbol{\delta})\right) \xrightarrow{a.s.} 0
\end{aligned}$$

as $n \to \infty$ for all $\varepsilon > 0$ follows. Thus, the Lindeberg-Feller theorem implies

$$\sqrt{n_i} \widehat{\mu}_i^G = \sum_{j=1}^{n_i} Z_{j,n_i} \xrightarrow{d} \mathcal{N}(0, \kappa_i \sigma_i^2)$$

almost surely as $n \to \infty$ given the data $(\mathbf{X}, \boldsymbol{\delta})$. Hence, the statement of the lemma follows by Slutsky's lemma. □

Lemma 4.4. *Under Assumption 4.1, we have*

$$P\left(\left|\widehat{\sigma}_i^{G2} - \sigma_i^2\right| > \varepsilon \mid (\mathbf{X}, \boldsymbol{\delta})\right) \xrightarrow{a.s.} 0$$

as $n \to \infty$ for all $i \in \{1, ..., k\}$.

Proof of Lemma 4.4. Let $i \in \{1, ..., k\}$ be arbitrary. Then, it holds

$$\begin{aligned}
\mathrm{E}\left[\widehat{\sigma}_i^{G2} \mid \mathbf{X}, \boldsymbol{\delta}\right] &= \sum_{j=1}^{n_i} n \mathrm{E}\left[G_{ij}^2 \mid \mathbf{X}, \boldsymbol{\delta}\right] \int_{[0,\tau)} \left(\int_x^\tau \widehat{S}_i(t)\, \mathrm{d}t\right)^2 \frac{1}{(Y_i(x) - \Delta N_i(x)) Y_i(x)} \, \mathrm{d}N_{ij}(x) \\
&= \widehat{\sigma}_i^2 \xrightarrow{a.s.} \sigma_i^2
\end{aligned}$$

as $n \to \infty$ and, analogously,

$$\begin{aligned}
\mathrm{E}\left[(\widehat{\sigma}_i^{G2})^2 \mid \mathbf{X}, \boldsymbol{\delta}\right] &\leqslant (\widehat{\sigma}_i^2)^2 + (C-1)\sum_{j=1}^{n_i} n^2 \left(\int_{[0,\tau)} \left(\int_x^\tau \widehat{S}_i(t)\,\mathrm{d}t \right)^2 \frac{1}{(Y_i(x) - \Delta N_i(x))Y_i(x)}\,\mathrm{d}N_{ij}(x) \right)^2 \\
&= \widehat{\sigma}_i^4 + (C-1)n^2 \int_{[0,\tau)} \left(\int_x^\tau \widehat{S}_i(t)\,\mathrm{d}t \right)^4 \frac{1}{(Y_i(x) - \Delta N_i(x))^2 Y_i^2(x)}\,\mathrm{d}N_i(x) \\
&\leqslant \widehat{\sigma}_i^4 + (C-1)\frac{n^2}{n_i^3} \int_{[0,\tau)} \frac{\tau^4}{(n_i^{-1}Y_i(x) - n_i^{-1}\Delta N_i(x))^2 n_i^{-1} Y_i(x)}\,\mathrm{d}\widehat{A}_i(x) \\
&\xrightarrow{a.s.} \sigma_i^4
\end{aligned}$$

as $n \to \infty$ by (4.20). Thus, it follows

$$P\left(\left|\widehat{\sigma}_i^{G2} - \sigma_i^2\right| > \varepsilon \mid (\mathbf{X}, \boldsymbol{\delta})\right) \xrightarrow{a.s.} 0$$

as $n \to \infty$ for all $\varepsilon > 0$ by Chebyshev's inequality. Hence, the statement of the lemma follows. □

Lemma 4.3 and 4.4 provide that there exists a measurable set $\Omega' \subset \Omega$ with $P(\Omega') = 1$ such that

$$\sup_{z_1,\dots,z_k \in \mathbb{R}} \left| P\left(\sqrt{n}\widehat{\mu}_1^G \leqslant z_1, \dots, \sqrt{n}\widehat{\mu}_k^G \leqslant z_k \mid (\mathbf{X}, \boldsymbol{\delta})\right)(\omega) - P\left(Z_1 \leqslant z_1, \dots, Z_k \leqslant z_k\right) \right| \to 0$$

and

$$P\left(\left|\widehat{\sigma}_i^{G2} - \sigma_i^2\right| > \varepsilon \mid (\mathbf{X}, \boldsymbol{\delta})\right)(\omega) \to 0$$

as $n \to \infty$ for all $i \in \{1, \dots, k\}$, $\omega \in \Omega'$. Then, by running through the same steps as in the proof of Theorem 4.1, we get

$$\sup_{z \in \mathbb{R}} \left| P\left(W_n^G(\mathbf{H}) \leqslant z \mid (\mathbf{X}, \boldsymbol{\delta})\right)(\omega) - P\left(Z \leqslant z\right) \right| \to 0$$

as $n \to \infty$ for all $\omega \in \Omega'$, where $Z \sim \chi^2_{\mathrm{rank}(\mathbf{H})}$. □

Proofs of Theorem 4.5, 4.7 and 4.8 The theorems about the joint convergences follow now easily from the previous results. Therefore, we apply Slutsky's lemma. For Theorem 4.5, we combine (4.16) and (4.17), for Theorem 4.7 Lemma 4.1 and 4.2 and for Theorem 4.8 Lemma 4.3 and 4.4. Then, we use the continuous mapping theorem with maps

$$\mathbb{R}^k \times \mathbb{R}^{k \times k} \ni (\mathbf{m}, \mathbf{S}) \mapsto \left((\mathbf{H}_\ell \mathbf{m})'(\mathbf{H}_\ell \mathbf{S} \mathbf{H}_\ell')^+ \mathbf{H}_\ell \mathbf{m}\right)_{\ell \in \mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$$

and

$$\mathbb{R}^k \times \mathbb{R}^{k \times k} \ni (\mathbf{m}, \mathbf{S}) \mapsto \left((\mathbf{H}_\ell \mathbf{m})'(\mathbf{H}_\ell \mathbf{S} \mathbf{H}_\ell')^+ \mathbf{H}_\ell \mathbf{m}\right)_{\ell \in \{1,\dots,L\}} \in \mathbb{R}^L.$$

The maps are continuous on $\mathbb{R}^k \times \{\boldsymbol{\Sigma}\}$ due to $\sigma_i^2 > 0$ for all $i \in \{1, \dots, k\}$. The three theorems follow, respectively. □

Proof of Theorem 4.6 In order to prove Theorem 4.6, we aim to apply Theorem 2.6 and Lemma 2.3. Therefore, let $\mathbf{X}_n := (\mathbf{X}, \boldsymbol{\delta})$ denote the data, $\mathbf{M}_n^{(b)} := \mathbf{Y}^{(b)}$ and

$$\mathbf{W}_n^{(b)} := \left((\mathbf{H}_\ell \widehat{\boldsymbol{\Sigma}}^{1/2} \mathbf{Y}^{(b)})'(\mathbf{H}_\ell \widehat{\boldsymbol{\Sigma}} \mathbf{H}_\ell')^+ (\mathbf{H}_\ell \widehat{\boldsymbol{\Sigma}}^{1/2} \mathbf{Y}^{(b)})\right)_{\ell \in \{1,\dots,L\}}$$

for all $b \in \{1, \dots, B_n\}$. Moreover, let F_n be as in Lemma 2.3, i.e., denoting the empirical distribution function of $\mathbf{W}_n^{(1)}, \dots, \mathbf{W}_n^{(B_n)}$. Then,

$$\mathbf{W}_n^{(1)} \xrightarrow{d^*} \left((\mathbf{H}_\ell \mathbf{Z})'(\mathbf{H}_\ell \boldsymbol{\Sigma} \mathbf{H}_\ell')^+ (\mathbf{H}_\ell \mathbf{Z})\right)_{\ell \in \{1,\dots,L\}}$$

holds as $n \to \infty$ for $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}_k, \boldsymbol{\Sigma})$ due to the consistency of $\widehat{\boldsymbol{\Sigma}}$, cf. the proof of Theorem 4.1. The marginal limit distributions are $\chi^2_{\mathrm{rank}(\mathbf{H}_\ell)}$, $\ell \in \{1, \dots, L\}$, which have continuous distribution functions $\mathcal{F}_\ell : \mathbb{R} \to [0,1]$ that are strictly increasing on $[0, \infty)$ due to $\mathrm{rank}(\mathbf{H}_\ell) > 0$. Hence, Lemma 2.3 implies (2.15). Furthermore, let $F_{\ell,n}$ denote the cumulative distribution function of $\chi^2_{\mathrm{rank}(\mathbf{H}_\ell \widehat{\boldsymbol{\Sigma}} \mathbf{H}_\ell')}$ for all $\ell \in \{1, \dots, L\}, n \in \mathbb{N}$, which fulfills (2.16) since

$$P\left(\mathrm{rank}\left(\mathbf{H}_\ell \widehat{\boldsymbol{\Sigma}} \mathbf{H}_\ell'\right) \neq \mathrm{rank}\left(\mathbf{H}_\ell\right)\right) \to 0$$

as $n \to \infty$ follows from the consistency of the covariance estimator. Then, Theorem 2.6 yields the statement of the theorem. □

Proof of Theorem 4.9 Again, we aim to apply Theorem 2.6 and Lemma 2.3. Therefore, let $\mathbf{X}_n := (\mathbf{X}, \boldsymbol{\delta})$ denote the data, $\mathbf{M}_n^{(b)}$ denote the randomness of the bootstrap procedures and $\mathbf{W}_n^{(b)}$ denote the bth Monte Carlo replicate of the bootstrap Wald-type test statistic for all $b \in \{1, ..., B_n\}$. Moreover, let F_n be as in Lemma 2.3, i.e., denoting the empirical distribution function of $\mathbf{W}_n^{(1)}, ..., \mathbf{W}_n^{(B_n)}$, and $F_{n,\ell}, \ell \in \{1, ..., L\}$, be their marginal cumulative distribution functions. Then,

$$\mathbf{W}_n^{(1)} \xrightarrow{d^*} \big((\mathbf{H}_\ell \mathbf{Z})'(\mathbf{H}_\ell \boldsymbol{\Sigma} \mathbf{H}_\ell')^+(\mathbf{H}_\ell \mathbf{Z})\big)_{\ell \in \{1,...,L\}}$$

holds as $n \to \infty$ for $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}_k, \boldsymbol{\Sigma})$ by Theorems 4.7 and 4.8. The marginal limit distributions are $\chi^2_{\mathrm{rank}(\mathbf{H}_\ell)}, \ell \in \{1, ..., L\}$, which have continuous distribution functions $\mathcal{F}_\ell : \mathbb{R} \to [0, 1]$ that are strictly increasing on $[0, \infty)$ due to $\mathrm{rank}(\mathbf{H}_\ell) > 0$. Hence, Lemma 2.3 implies (2.15). Moreover, Remark 2.4 implies (2.16). Thus, Theorem 2.6 yields the statements of the theorem. $\square$

5 RMTL-Based Inference in Competing Risks Setups

As we have seen in the previous section, the restricted mean survival time (RMST) is an alternative effect measure to the popular hazard ratio, especially in situations where the proportional hazards assumption is violated [68]. It is defined as the area under the survival curve up to a prespecified end point τ and, thus, it offers a straightforward interpretation as the expected duration of time alive before τ. By integrating across the distribution function rather than the survival curve, we derive the restricted mean time lost (RMTL), which can be interpreted as expected time lost before τ. Naturally, it equals τ minus the RMST. In the context of competing-risks frameworks, where multiple events like death from various causes occur, the RMTL for a specific event can be defined simply as the area under the corresponding sub-distribution function. Then, the restricted mean survival time equals τ minus the RMTLs of all possible events. However, analyzing RMTLs instead of the restricted mean survival time in competing-risks frameworks offers the possibility to differ between the different risks. An exemplary illustration of the relation between the RMTL and RMST can be found in Figure 9.

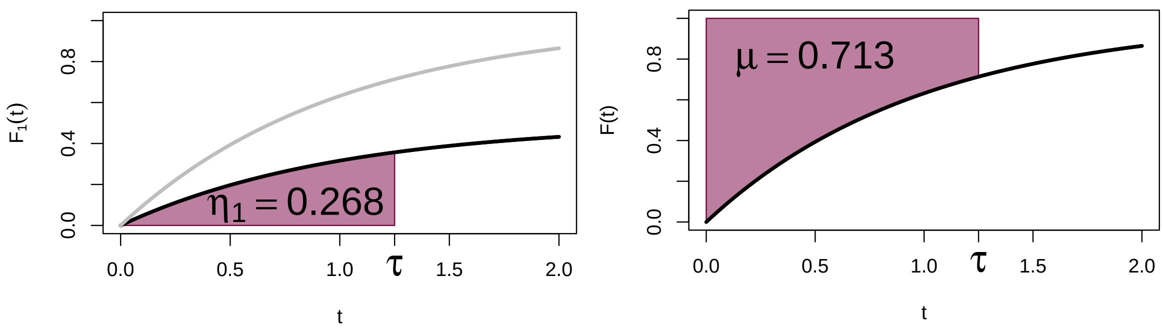


Figure 9: An exemplary illustration of the RMTL of the first event η_1 (left) and RMST μ (right).

In competing-risks settings, the RMTL has been studied in several papers [2, 55, 77, 78, 79]. However, the considered settings are limited to one- and two-sample cases and mostly allow for only two different event types such that there is a lack of suitable RMTL-based tests for more complex factorial designs, more general hypotheses and more event types to, e.g., compare the RMTLs of various event types across several groups. Additionally, all proposed methods seem to require existing sub-distribution hazards, i.e., in particular continuous sub-distribution functions. This assumption is often not justified in practice, e.g., when the event times are measured in whole days or weeks. Consequently, we aim to develop flexible Wald-type tests that are applicable (i) for general RMTL contrasts in factorial designs and (ii) without a continuity assumption on the event times. Moreover, for the RMST in the classical survival setup, resampling procedures have proven to be useful in ensuring an accurate type I error control for finite samples [24, 43, 58]. Hence, to improve the small sample performance of the constructed Wald-type test, a studentized permutation approach is applied and its asymptotic validity is shown in Proposition 5.4.

In many applications, the comparison of the RMTLs across several groups may be of interest. Here, the global null hypothesis might be that all RMTLs are equal across the groups, cf. Example 5.3. If a test rejects this global null hypothesis, it could also be of interest which specific RMTL differences cause the significant result. In order to answer such questions, multiple tests for pairwise RMTL comparisons need to be performed simultaneously. Recently, a maximum joint test for testing the equalities of two RMTLs in the two-sample case jointly was studied in [77]. However, the two-sample case and the two considered hypotheses, i.e., equal RMTLs of event type 1 and 2, respectively, are rather restrictive. Hence, there is still a lack of multiple testing procedures based on RMTLs for general multiple contrast hypotheses addressing (i) and (ii). Thus, we aim to develop powerful multiple tests for RMTL contrasts by taking the asymptotically exact dependence structure of the local Wald-type test statistics into account.

The remainder of this section is organized as follows. The general factorial competing risks setup is presented in Section 5.1 including the formal definition of the RMTL. The global testing problem is introduced in Section 5.2. The Wald-type test statistic is investigated in Section 5.2.1. In Section 5.2.2, the studentized permutation approach is introduced and its asymptotic validity is proven in Proposition 5.4. Multiple tests for several RMTL contrast hypotheses are developed in Section 5.3. In Section 5.4, the finite sample performance of our proposed methods is analyzed in extensive simulations. Additionally, we illustrate our methods by analyzing data of leukemia patients who underwent bone marrow transplantation in Section 5.5. All technical proofs of this section are given in Section 5.6. Moreover, an implementation of the proposed methods is freely available in the R package `GFDrmtl` [21], see Section E for a description.

5.1 Factorial Competing Risks Setup

In the following, we interpret a factorial competing risks design as a k-sample setup with M competing events; $k, M \in \mathbb{N}, k \geqslant 2$. We assume that there are independent event and right-censoring times $T_{ij} \sim S_i, C_{ij} \sim G_i, j \in \{1, \ldots, n_i\}, i \in \{1, \ldots, k\}$, respectively, and random variables $D_{ij}, j \in \{1, \ldots, n_i\}, i \in \{1, \ldots, k\}$ indicating the event types and taking values in $\{1, \ldots, M\}$. Here,

$$S_i : [0, \infty) \to [0, 1],\ S_i(t) := P(T_{i1} > t) \quad \text{and} \quad G_i : [0, \infty) \to [0, 1],\ G_i(t) := P(C_{i1} > t)$$

denote the survival functions of the event and censoring times, respectively, and $n_i \geqslant 2$ denotes the sample size of group i for all $i \in \{1, \ldots, k\}$. We do not suppose the continuity of the survival functions and, thus, we explicitly allow for ties in the data. Additionally, we assume that $(T_{ij}, C_{ij}, D_{ij}), j \in \{1, \ldots, n_i\}, i \in \{1, \ldots, k\}$ are mutually independent and that the censoring time C_{ij} is independent of the event time and event type (T_{ij}, D_{ij}) for all $j \in \{1, \ldots, n_i\}, i \in \{1, \ldots, k\}$. Due to right-censoring, we can only observe the right-censored event times $X_{ij} := \min\{T_{ij}, C_{ij}\}$ and the event indicator $\delta_{ij} := D_{ij} \mathbb{1}\{X_{ij} = T_{ij}\}, j \in \{1, \ldots, n_i\}, i \in \{1, \ldots, k\}$, where here and throughout $\mathbb{1}$ denotes the indicator function. Furthermore, let

$$\begin{aligned} &F_{im} : [0, \infty) \to [0, 1],\ F_{im}(t) := P(T_{i1} \leqslant t, D_{i1} = m) \\ \text{and} \quad &A_{im} : [0, \infty) \to [0, \infty],\ A_{im}(t) := \int_{[0,t]} \frac{1}{S_{i-}} \,\mathrm{d}F_{im} \end{aligned}$$

denote the cumulative incidence function and the cause-specific cumulative hazard functions, respectively, for all $i \in \{1, \ldots, k\}, m \in \{1, \ldots, M\}$. The sum of all cause-specific hazard functions of group i is denoted by $A_i := \sum_{m=1}^{M} A_{im}, i \in \{1, \ldots, k\}$, in the following.

In order to introduce suitable estimators for these quantities, we firstly define the number of individuals at risk just before time $t \geqslant 0$ by $Y_i(t) := \sum_{j=1}^{n_i} \mathbb{1}\{X_{ij} \geqslant t\}$ and the number of individuals with an event of type m before or at time $t \geqslant 0$ by $N_{im}(t) := \sum_{j=1}^{n_i} \mathbb{1}\{X_{ij} \leqslant t, \delta_{ij} = m\}$ for all $i \in \{1, \ldots, k\}, m \in \{1, \ldots, M\}$. Then, we set

$$\widehat{A}_{im}(t) := \int_{[0,t]} \frac{1}{Y_i} \,\mathrm{d}N_{im}, \quad \widehat{A}_i := \sum_{m=1}^{M} \widehat{A}_{im}, \quad \text{and} \quad \widehat{S}_i(t) := \prod_{x \in [0,t]} \left\{1 - \mathrm{d}\widehat{A}_i(x)\right\}$$

for all $t \geqslant 0, i \in \{1, \ldots, k\}, m \in \{1, \ldots, M\}$. These estimators are the cause-specific and all-cause Nelson–Aalen estimators and the Kaplan–Meier estimator, respectively. Thus, we obtain the Aalen–Johansen estimator at t for $F_{im}(t)$ as $\widehat{F}_{im}(t) := \int_{[0,t]} \widehat{S}_{i-} \,\mathrm{d}\widehat{A}_{im}, i \in \{1, \ldots, k\}, m \in \{1, \ldots, M\}$ for all $t \geqslant 0$.

The restricted mean time lost (RMTL) due to the event type m in group i is defined as the area under the corresponding cumulative incidence function up to a prespecified time point $\tau > 0$, that is,

$$\eta_{im} := \int_0^\tau F_{im}(t) \,\mathrm{d}t, \quad i \in \{1, \ldots, k\}, m \in \{1, \ldots, M\}.$$

Of note, in the case of only one event type, i.e., $M = 1$, the RMTL equals τ minus the more popular RMST. By replacing F_{im} with the corresponding Aalen–Johansen estimator, we obtain a natural estimator for the RMTL, that is,

$$\widehat{\eta}_{im} := \int_0^\tau \widehat{F}_{im}(t) \,\mathrm{d}t, \quad i \in \{1, \ldots, k\}, m \in \{1, \ldots, M\}.$$

5.2 Global Tests

Let

$$\boldsymbol{\eta} := (\eta_{11}, \ldots, \eta_{1M}, \eta_{21}, \ldots, \eta_{kM})'$$

denote the vector of the RMTLs and its estimator by

$$\widehat{\boldsymbol{\eta}} := (\widehat{\eta}_{11}, \ldots, \widehat{\eta}_{1M}, \widehat{\eta}_{21}, \ldots, \widehat{\eta}_{kM})'.$$

Moreover, let $r \in \mathbb{N}$, $\mathbf{c} \in \mathbb{R}^r$ and $\mathbf{H} \in \mathbb{R}^{r \times kM} \setminus \{\mathbf{0}_{r \times kM}\}$ satisfying $\mathbf{H}(\mathbf{1}_k \otimes \mathbf{e}_m) = \mathbf{0}_r, m \in \{1, \ldots, M\}$, where here and throughout $\mathbf{e}_m = (0, ..., 0, 1, 0, ..., 0)^\top \in \mathbb{R}^M$ denotes the mth standard unit vector and $\otimes$ denotes the Kronecker product. This ensures that $\mathbf{H}$ has the contrast property in terms of the different groups and not in terms of the different event types. Here, we consider the testing problem

$$\mathcal{H}_0 : \mathbf{H}\boldsymbol{\eta} = \mathbf{c} \quad \text{vs.} \quad \mathcal{H}_1 : \mathbf{H}\boldsymbol{\eta} \neq \mathbf{c}. \tag{5.1}$$

This testing problem is very general and covers various types of hypotheses and factorial designs as illustrated in the following examples.

Example 5.1 (Two-sample case)**.** *The simplest but perhaps most relevant case in practice is the two-sample case, i.e., $k = 2$. The null hypothesis of equal RMTLs of all event types, i.e., $\mathcal{H}_0 : \eta_{1m} = \eta_{2m}, m \in \{1, \ldots, M\}$, can be realized by choosing $\mathbf{c} := \mathbf{0}_r$ and $\mathbf{H} := [-1, 1] \otimes \mathbf{I}_M$. If not the RMTLs of all M event types but only the first $\widetilde{M} < M$ event types are of interest, we may choose $\mathbf{H} := [-1, 1] \otimes [\mathbf{I}_{\widetilde{M}}, \mathbf{0}_{\widetilde{M} \times (M - \widetilde{M})}]$ instead. This yields the null hypothesis $\mathcal{H}_0 : \eta_{1m} = \eta_{2m}, m \in \{1, \ldots, \widetilde{M}\}$.*

Example 5.2 (One-way design)**.** *In many applications, the hypothesis of equal RMTLs across the groups is of interest, that is,*

$$\mathcal{H}_0 : \eta_{1m} = \ldots = \eta_{km}, \quad m \in \{1, \ldots, M\}.$$

This hypothesis can be formulated with $\mathbf{c} := \mathbf{0}_r$ and various hypothesis matrices $\mathbf{H}$. For example, $\mathbf{H}$ may be chosen as the Kronecker product of the Dunnett-type [34] contrast matrix (4.7) *and the identity matrix $\mathbf{I}_M \in \mathbb{R}^{M \times M}$. Another possibility is the Kronecker product of the Tukey-type [73] contrast matrix* (4.8) *and the identity matrix $\mathbf{I}_M$.*

Example 5.3 (Factorial 2-by-2 design)**.** *In a factorial 2-by-2 design with factors A and B, $k = 4$ groups arise from the combinations* (A,B) $\in \{(1,1), (1,2), (2,1), (2,2)\}$. *The following null hypotheses and corresponding hypothesis matrices combined with $\mathbf{c} = \mathbf{0}_M$ are relevant:*

- *no main effect of factor A;* $\quad \mathbf{H}_A = [\mathbf{I}_M, \mathbf{I}_M, -\mathbf{I}_M, -\mathbf{I}_M]$;
- *no main effect of factor B;* $\quad \mathbf{H}_B = [\mathbf{I}_M, -\mathbf{I}_M, \mathbf{I}_M, -\mathbf{I}_M]$;
- *no interaction effect between A and B;* $\quad \mathbf{H}_{AB} = [\mathbf{I}_M, -\mathbf{I}_M, -\mathbf{I}_M, \mathbf{I}_M]$.

More general factorial designs can be incorporated easily by splitting up the indices similarly as in Example 5.3, see [63] for details.

5.2.1 The Wald-Type Test Statistic and its Asymptotic Behavior

In this section, we construct and study a suitable test statistic for the testing problem (5.1). For technical reasons, we need the following assumptions.

Assumption 5.1. *In the following, we assume $S_{i-}(\tau) > 0$, $G_{i-}(\tau) > 0$ and $n_i/n \to \kappa_i \in (0, 1)$ as $n \to \infty$ for all $i \in \{1, \ldots, k\}$, where here and throughout $n := \sum_{i=1}^k n_i$ denotes the total sample size.*

By applying empirical process theory and the delta method, we obtain the asymptotic normality of the vector of RMTL estimators $\widehat{\boldsymbol{\eta}}$:

Theorem 5.1. *Under Assumption 5.1, we have*

$$\sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{d} \mathbf{Z} \sim \mathcal{N}_{kM}(\mathbf{0}_{kM}, \boldsymbol{\Sigma})$$

as $n \to \infty$. The covariance matrix $\boldsymbol{\Sigma}$ is defined in (5.7) in Section 5.6.

Note that there is no notation clash regarding $\boldsymbol{\Sigma}$ in Section 4 since it coincides with $\boldsymbol{\Sigma}$ in the previous theorem in the special case of only one event type $M = 1$.
In the following, we also need that the limit distribution is not degenerated. In Lemma 5.2 in the appendix, we show that the following natural assumption together with Assumption 5.1 is sufficient to guarantee the positive definiteness of $\boldsymbol{\Sigma}$:

Assumption 5.2. *We assume that $F_{im-}(\tau) > 0$ for all $i \in \{1, \ldots, k\}, m \in \{1, \ldots, M\}$.*

As shown in (5.6), the entries of $\boldsymbol{\Sigma}$ depend on the unknown functions F_{im}, A_i and $\sigma_{im\widetilde{m}}$ for all $i \in \{1, \ldots, k\}, m, \widetilde{m} \in \{1, \ldots, M\}$, where $\sigma_{im\widetilde{m}}$ denotes the asymptotic covariance function of the cause-specific cumulative hazard functions $\widehat{A}_{im}, \widehat{A}_{i\widetilde{m}}$. Thus, the plug-in estimator

$$\widehat{\boldsymbol{\Sigma}} := \bigoplus_{i=1}^{k} \left(\frac{n}{n_i} \widehat{\boldsymbol{\Sigma}}_i \right)$$

for $\boldsymbol{\Sigma}$ can be obtained by replacing F_{im}, A_i and $\sigma_{im\widetilde{m}}$ in (5.6) by $\widehat{F}_{im}, \widehat{A}_i$ and $\widehat{\sigma}_{im\widetilde{m}}$, respectively, for all $i \in \{1, \ldots, k\}, m, \widetilde{m} \in \{1, \ldots, M\}$, with

$$\widehat{\sigma}_{imm}(t) := n_i \int_{[0,t]} \frac{1 - \Delta \widehat{A}_{im}}{Y_i} \, \mathrm{d}\widehat{A}_{im} \quad \text{and} \quad \widehat{\sigma}_{im\widetilde{m}}(t) := -n_i \int_{[0,t]} \frac{\Delta \widehat{A}_{im}}{Y_i} \, \mathrm{d}\widehat{A}_{i\widetilde{m}}$$

for all $t \geqslant 0, m \neq \widetilde{m}$. Here, $\bigoplus$ denotes the direct sum. Then, the Wald-type test statistic can be defined by

$$W_n(\mathbf{H}, \mathbf{c}) := n(\mathbf{H}\widehat{\boldsymbol{\eta}} - \mathbf{c})' \left(\mathbf{H}\widehat{\boldsymbol{\Sigma}}\mathbf{H}'\right)^+ (\mathbf{H}\widehat{\boldsymbol{\eta}} - \mathbf{c}).$$

Since the Wald-type test statistic is a quadratic form of the vector $(\mathbf{H}\widehat{\boldsymbol{\eta}} - \mathbf{c})$, we would reject the null hypothesis in (5.1) for large values of $W_n(\mathbf{H}, \mathbf{c})$. The following theorem provides the asymptotic distribution of the Wald-type test statistic.

Theorem 5.2. *Under Assumptions 5.1 and 5.2 and the null hypothesis in (5.1), we have, as $n \to \infty$,*

$$W_n(\mathbf{H}, \mathbf{c}) \xrightarrow{d} \chi^2_{\mathrm{rank}(\mathbf{H})}.$$

Thus, an asymptotic level-α test for (5.1) is $\varphi = \mathbb{1}\{W_n(\mathbf{H}, \mathbf{c}) > \chi^2_{\mathrm{rank}(\mathbf{H}),1-\alpha}\}$. Due to the direct connection between tests and confidence regions, we also obtain a confidence region with confidence level $1 - \alpha$ for $\mathbf{H}\boldsymbol{\eta}$ from Theorem 5.2, that is, $\{\boldsymbol{\xi} \in \mathbb{R}^r \mid W_n(\mathbf{H}, \boldsymbol{\xi}) \leqslant \chi^2_{\mathrm{rank}(\mathbf{H}),1-\alpha}\}$.

5.2.2 Studentized Permutation Test

We showed in the previous section that the proposed test for the RMTLs is asymptotically valid but this generally does not guarantee a good small sample performance of the test in terms of type I error control. As we will see in Section 5.4.2, the asymptotic test has in fact an increased type I error in simulations. For the RMST, permutation methods solved this problem [24, 43, 58]. Permutation tests are known to control the type I error exactly under exchangeable data [41, 51], which means $F_{im} = F_{jm}, G_i = G_j, i, j \in \{1, \ldots, k\}, m \in \{1, \ldots, M\}$, in our case. However, the null hypothesis in (5.1) may hold even if the data are not exchangeable. Thus, we develop a studentized permutation approach that not only preserves the finite exact control of the type I error under exchangeability but is also asymptotically valid under non-exchangeable data as the studentized permutation tests in [24, 58].
To this end, let

$$(\mathbf{X}, \boldsymbol{\delta}) = (X_j, \delta_j)_{j=1,\ldots,n} := (X_{ij}, \delta_{ij})_{j\in\{1,\ldots,n_i\}, i\in\{1,\ldots,k\}}$$

denote the pooled sample and $(X^\pi_{ij}, \delta^\pi_{ij})_{j\in\{1,\ldots,n_i\}, i\in\{1,\ldots,k\}}$ the permuted data. In detail, the data points are permuted as pairs (X_j, δ_j) by drawing the vector $(R_1, ..., R_n)$ uniformly on the set of all permutations of $(1, ..., n)$ independently of the data and defining

$$(X^\pi_{ij}, \delta^\pi_{ij})_{j\in\{1,\ldots,n_i\}, i\in\{1,\ldots,k\}} := (X^\pi_j, \delta^\pi_j)_{j\in\{1,\ldots,n\}} := (X_{R_j}, \delta_{R_j})_{j\in\{1,\ldots,n\}}.$$

This can also be interpreted as shuffling the groups of the original data randomly. Furthermore, we denote the statistics $\widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\Sigma}}$ re-calculated based on the permuted data with a π in the superscript, i.e., $\widehat{\boldsymbol{\eta}}^\pi, \widehat{\boldsymbol{\Sigma}}^\pi$. Finally, we define the permutation counterpart of the Wald-type test statistic by

$$W^\pi_n(\mathbf{H}) := n(\mathbf{H}\widehat{\boldsymbol{\eta}}^\pi)' \left(\mathbf{H}\widehat{\boldsymbol{\Sigma}}^\pi\mathbf{H}'\right)^+ (\mathbf{H}\widehat{\boldsymbol{\eta}}^\pi).$$

It asymptotically mimics the null distribution of $W_n(\mathbf{H}, \mathbf{c})$, as shown in the following theorem.

Theorem 5.3. *Under Assumptions 5.1 and 5.2, we have under both hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, as $n \to \infty$,*

$$W^\pi_n(\mathbf{H}) \xrightarrow{d^*} \chi^2_{\mathrm{rank}(\mathbf{H})}.$$

By using this result, we can construct a permutation test for (5.1). In practice, usually a Monte Carlo method is applied to approximate the resulting critical value, which is the $(1 - \alpha)$-quantile of the conditional distribution of $W^\pi_n(\mathbf{H})$ given the data $(\mathbf{X}, \boldsymbol{\delta})$. Therefore, the quantile is approximated by the empirical $(1 - \alpha)$-quantile $q^\pi_{1-\alpha}$ of B_n conditional independent random variables distributed as $W^\pi_n(\mathbf{H})$ given $(\mathbf{X}, \boldsymbol{\delta})$. Here and throughout, $(B_n)_{n\in\mathbb{N}}$ denotes a sequence of natural numbers with $B_n \to \infty$ as $n \to \infty$. Hence, we receive the permutation test $\varphi^\pi = \mathbb{1}\left\{W_n(\mathbf{H}, \mathbf{c}) > q^\pi_{1-\alpha}\right\}$. This permutation test is asymptotically valid:

Theorem 5.4. *Under Assumptions 5.1, 5.2 and the null hypothesis in* (5.1)*, we have*

$$\lim_{n\to\infty} \mathrm{E}\left(\varphi^\pi\right) = \lim_{n\to\infty} P\left(W_n(\mathbf{H}, \mathbf{c}) > q^\pi_{1-\alpha}\right) = \alpha.$$

The corresponding confidence region with level $1 - \alpha$ for $\mathbf{H}\boldsymbol{\eta}$ is $\left\{\boldsymbol{\xi} \in \mathbb{R}^r \mid W_n(\mathbf{H}, \boldsymbol{\xi}) \leqslant q^\pi_{1-\alpha}\right\}$, i.e.,

$$\lim_{n\to\infty} P\left(\mathbf{H}\boldsymbol{\eta} \in \left\{\boldsymbol{\xi} \in \mathbb{R}^r \mid W_n(\mathbf{H}, \boldsymbol{\xi}) \leqslant q^\pi_{1-\alpha}\right\}\right) = \lim_{n\to\infty} P\left(W_n(\mathbf{H}, \mathbf{H}\boldsymbol{\eta}) \leqslant q^\pi_{1-\alpha}\right) = 1 - \alpha.$$

Example 5.4. *Let* $\mathbf{H} = (\tilde{\mathbf{e}}_1 - \tilde{\mathbf{e}}_2)^\top \otimes \mathbf{e}_1^\top$*, where* $\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2 \in \mathbb{R}^k$ *denote the first and second unit vector in* $\mathbb{R}^k$*. Then a permutation-based asymptotic* $(1-\alpha)$*-confidence interval for* $\eta_{11} - \eta_{21} = \mathbf{H}\boldsymbol{\eta}$ *is*

$$\left[\widehat{\eta}_{11} - \widehat{\eta}_{21} - \left(\left(\widehat{\Sigma}_{111}/n_1 + \widehat{\Sigma}_{211}/n_2\right) q^{\pi}_{1-\alpha}\right)^{1/2}, \widehat{\eta}_{11} - \widehat{\eta}_{21} + \left(\left(\widehat{\Sigma}_{111}/n_1 + \widehat{\Sigma}_{211}/n_2\right) q^{\pi}_{1-\alpha}\right)^{1/2}\right],$$

where $\widehat{\Sigma}_{i11}$ *denotes the top-left entry of* $\widehat{\boldsymbol{\Sigma}}_i, i \in \{1, \dots, k\}$*. Analogously, an asymptotic* $(1-\alpha)$*-confidence interval for* $\eta_{11} - \eta_{21}$ *based on the asymptotic test in Section 5.2.1 is*

$$\left[\widehat{\eta}_{11} - \widehat{\eta}_{21} - \left(\left(\widehat{\Sigma}_{111}/n_1 + \widehat{\Sigma}_{211}/n_2\right) \chi^2_{1,1-\alpha}\right)^{1/2}, \widehat{\eta}_{11} - \widehat{\eta}_{21} + \left(\left(\widehat{\Sigma}_{111}/n_1 + \widehat{\Sigma}_{211}/n_2\right) \chi^2_{1,1-\alpha}\right)^{1/2}\right].$$

5.3 Multiple Tests

In many applications, not only the global test decisions for (5.1) are of interest but a more in-depth analysis of local hypotheses. By that, conclusions on which specific hypotheses cause a rejection of the global hypothesis can be drawn.
Formally, we split up the hypothesis matrix $\mathbf{H} = [\mathbf{H}_1', \dots, \mathbf{H}_L']'$ into L matrices with $\mathrm{rank}(\mathbf{H}_\ell) > 0, \ell \in \{1, \dots, L\}$, and the vector $\mathbf{c} = (\mathbf{c}_1', \dots, \mathbf{c}_L')'$ into L vectors of lengths corresponding to the number of rows of the matrices $\mathbf{H}_1, \dots, \mathbf{H}_L$, respectively. This covers but is not restricted to the case that $\mathbf{H}_1, \dots, \mathbf{H}_L$ can be chosen to be the r rows of $\mathbf{H}$. Then, the multiple testing problem is

$$\mathcal{H}_{0,\ell} : \mathbf{H}_\ell \boldsymbol{\eta} = \mathbf{c}_\ell \quad \text{vs.} \quad \mathcal{H}_{1,\ell} : \mathbf{H}_\ell \boldsymbol{\eta} \neq \mathbf{c}_\ell, \quad \ell \in \{1, \dots, L\}. \tag{5.2}$$

As we will see in the following examples, this formulation covers the most interesting cases for multiple hypotheses about the RMTLs in practice.

Example 5.5 (Two-sample case, continued)**.** *If it is also of interest which event type differences in Example 5.1 cause the significant result, multiple tests need to be performed for* $\widetilde{M} > 1$*. In our notation, this can be realized by choosing the matrices* $\mathbf{H}_m \in \mathbb{R}^{1 \times kM}, m \in \{1, \dots, \widetilde{M}\}$ *as the rows of the hypothesis matrix* $\mathbf{H}$ *given in Example 5.1, i.e.,* $\mathbf{H}_m := [-\mathbf{e}_m', \mathbf{e}_m'], m \in \{1, \dots, \widetilde{M}\}$*. Hence, we receive the multiple hypotheses* $\mathcal{H}_{0,m} : \eta_{1m} = \eta_{2m}, m \in \{1, \dots, \widetilde{M}\}$*. For* $\widetilde{M} = 2$*, we receive the hypotheses of [77] as special case.*

Example 5.6 (One-way design, continued)**.** *Now, the choice of the hypothesis matrix leading to the hypothesis of equal RMTLs across the groups in Example 5.2 becomes important and depends on the question of interest. E.g., if all RMTLs should be compared to the RMTLs of the first group (many-to-one), the Dunnett-type contrast matrix is the hypothesis matrix to go with. However, the Tukey-type contrast matrix should be used if the RMTLs of all pairs of groups should be compared.*
The second choice is how to split up the hypothesis matrix $\mathbf{H}$*. This, again, depends on the question of interest. If it is only of interest which groups have different RMTLs but it does not matter for which event types the RMTLs exhibit differences, it is enough to consider* $\mathbf{h}_\ell \otimes \mathbf{I}_M, \ell \in \{1, \dots, L\}$*, as the hypothesis matrices, where* $\mathbf{h}_\ell$ *denotes the* ℓ*th row of the Dunnett- and Tukey-type contrast matrix, respectively. However, if also the event types that cause a rejection of equal RMTLs should be detected, each row of the (global) hypothesis matrix* $\mathbf{H}$ *corresponds to a hypothesis matrix of the multiple tests, i.e.,* $\mathbf{H}_\ell \in \mathbb{R}^{1 \times kM}, \ell \in \{1, \dots, L\}$*.*

Example 5.7 (Factorial 2-by-2 design, continued)**.** *In Example 5.3, main effects A and B, and an interaction effect could be tested simultaneously with the help of* $\mathbf{H}_1 = \mathbf{H}_A, \mathbf{H}_2 = \mathbf{H}_B$ *and* $\mathbf{H}_3 = \mathbf{H}_{AB}$*, respectively. However, the resulting tests cannot determine which event type(s) caused a significant difference between the groups. For this, all* M *rows of each of the three hypotheses matrix must be considered as separate hypothesis matrices which results in* $3M$ *multiple hypotheses.*

The local test statistics $W_n(\mathbf{H}_\ell, \mathbf{c}_\ell), \ell \in \{1, \dots, L\}$, can be used to derive (local) test decisions for $\mathcal{H}_{0,\ell}, \ell \in \{1, \dots, L\}$, respectively. As we already developed global tests in Section 5.2, a simple application of the Bonferroni-correction can solve the multiple testing problem. However, the Bonferroni-correction is known to lead to conservative decisions and low power. Hence, we aim to incorporate the asymptotic exact dependence structure of the local test statistics as described in Section 2.3 for constructing powerful multiple tests. The multivariate limit distribution of the local test statistics is given by the following theorem.

Theorem 5.5. *Let* $\mathcal{T}$ *denote the indices of true null hypotheses in* (5.2) *and let* $\mathbf{Z}$ *be as in Theorem 5.1. Under Assumptions 5.1 and 5.2, we have, as* $n \to \infty$*,*

$$\left(W_n(\mathbf{H}_\ell, \mathbf{c}_\ell)\right)_{\ell \in \mathcal{T}} \xrightarrow{d} \left((\mathbf{H}_\ell \mathbf{Z})'(\mathbf{H}_\ell \boldsymbol{\Sigma} \mathbf{H}_\ell')^+ (\mathbf{H}_\ell \mathbf{Z})\right)_{\ell \in \mathcal{T}}.$$

5.3.1 Asymptotic Multiple Tests

Motivated by Theorem 5.5, asymptotic multiple tests are given by

$$\varphi_\ell = \mathbb{1}\left\{W_n(\mathbf{H}_\ell, \mathbf{c}_\ell) > \chi^2_{\mathrm{rank}(\mathbf{H}_\ell),1-\beta_n}\right\}, \quad \ell \in \{1, \ldots, L\}, \tag{5.3}$$

where β_n denotes the local level for each test and can be derived from the multivariate limit distribution in Theorem 5.5. In practice, this local level can be approximated by a Monte Carlo method as

$$\beta_n = \max\{\beta \in \{0, 1/B_n, \ldots, (B_n - 1)/B_n\} \mid \mathrm{FWER}_n(\beta) \leqslant \alpha\}$$

with approximated family-wise error rate

$$\mathrm{FWER}_n(\beta) = \frac{1}{B_n}\sum_{b=1}^{B_n} \max_{\ell\in\{1,\ldots,L\}} \mathbb{1}\left\{(\mathbf{H}_\ell\widehat{\mathbf{\Sigma}}^{1/2}\mathbf{Y}^{(b)})'(\mathbf{H}_\ell\widehat{\mathbf{\Sigma}}\mathbf{H}_\ell')^+(\mathbf{H}_\ell\widehat{\mathbf{\Sigma}}^{1/2}\mathbf{Y}^{(b)}) > \chi^2_{\mathrm{rank}(\mathbf{H}_\ell\widehat{\mathbf{\Sigma}}\mathbf{H}_\ell'),1-\beta}\right\}$$

for $\beta \in [0, 1)$ and $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(B_n)} \sim \mathcal{N}_{kM}(\mathbf{0}_{kM}, \mathbf{I}_{kM})$ i.i.d. and independent of $\widehat{\mathbf{\Sigma}}$. Here, $(B_n)_{n\in\mathbb{N}}$ is a sequence of natural numbers with $B_n \to \infty$ as $n \to \infty$.

Theorem 5.6. *Under Assumptions 5.1 and 5.2, the multiple asymptotic tests* (5.3) *fulfill* (2.20)–(2.22)*, i.e., they control the family-wise error rate asymptotically in the strong sense and are asymptotically balanced.*

By Remark 2.6 and Theorem 5.6, simultaneous confidence regions for $\mathbf{H}_\ell\boldsymbol{\eta}, \ell \in \{1, \ldots, L\}$, with asymptotic global confidence level $1 - \alpha$ of the following form are immediate:

$$\bigtimes_{\ell=1}^{L} \left\{\boldsymbol{\xi} \mid W_n(\mathbf{H}_\ell, \boldsymbol{\xi}) \leqslant \chi^2_{\mathrm{rank}(\mathbf{H}_\ell),1-\beta_n}\right\} \subset \mathbb{R}^r.$$

For row vectors $\mathbf{H}_\ell \in \mathbb{R}^{1\times k}$, the confidence region simplifies to a confidence interval, that is

$$\left[\mathbf{H}_\ell\widehat{\boldsymbol{\eta}} - \left(\frac{\mathbf{H}_\ell\widehat{\mathbf{\Sigma}}\mathbf{H}_\ell'}{n}\chi^2_{\mathrm{rank}(\mathbf{H}_\ell),1-\beta_n}\right)^{1/2}, \mathbf{H}_\ell\widehat{\boldsymbol{\eta}} + \left(\frac{\mathbf{H}_\ell\widehat{\mathbf{\Sigma}}\mathbf{H}_\ell'}{n}\chi^2_{\mathrm{rank}(\mathbf{H}_\ell),1-\beta_n}\right)^{1/2}\right].$$

Moreover, we can derive more powerful multiple tests by using the closed testing procedure, cf. Remark 2.5, and adjusted p-values by the methodologies in Section 2.3.

5.4 Simulation Study

5.4.1 Simulation Setup

For the simulation study, we used the computing environment R, version 4.2.1 [66]. The simulation setup is based on the simulation in Section 4.4.1 and adapted for competing risks data. We considered $k = 4$ groups with equal event time distributions for the first three groups while the distribution of the fourth group may differ, using the same survival and censoring distributions as in Section 4.4.1 with the same censoring rates stated there. An illustration of the survival curves of the event times can be found in [24] and of the censoring times in Section 4.4.1.

Beyond these continuous settings, we also added corresponding discrete settings. This is done since the proposed methods also work under the existence of ties as proven in Section 5.6. Therefore, we generated the event times as in the continuous case but round them up to obtain integer values. Of course, the rounding typically results in altered values of the RMSTs $\int_0^\tau S_i(t)\mathrm{d}t, i \in \{1, \ldots, k\}$ and RMTLs. However, it is still possible to obtain a specific RMST difference δ as in Section 4.4.1 by adjusting the parameters $\lambda_{\delta,1}, \ldots, \lambda_{\delta,9}$ adequately.

As in the data example in Section 5.5 below, we are considering $M = 3$ event types. The causes $D_{ij} \in \{1, 2, 3\}$ were drawn independently of the survival and censoring times with probabilities $p_1 = 33\%$, $p_2 = 25\%$, and $p_3 = 42\%$, respectively, across all $j \in \{1, \ldots, n_i\}, i \in \{1, \ldots, k\}$. This results in a direct connection between the survival function and the cumulative incidence functions, that is, $F_{im} = p_m(1 - S_i)$ for all $i \in \{1, \ldots, k\}, m \in \{1, \ldots, M\}$. Hence, the RMTLs $\eta_{1m}, \eta_{2m}, \eta_{3m}, \eta_{4m}$ of the event type m coincide whenever the RMSTs $\int_0^\tau S_i(t)\mathrm{d}t, i \in \{1, \ldots, k\}$ coincide. Furthermore, a RMST difference of $\delta = \int_0^\tau S_1(t)\mathrm{d}t - \int_0^\tau S_4(t)\mathrm{d}t$ results in an RMTL difference of $\eta_{4m} - \eta_{1m} = p_m\delta, m \in \{1, \ldots, M\}$.

Motivated by the data example in Section 5.5, the hypothesis matrix in Example 5.7 is considered for testing on the two main effects and an interaction effect simultaneously in a 2-by-2 design (2x2). This results in the local null hypotheses

$$\begin{aligned} &\mathcal{H}_{0,m}^{A} : \eta_{1m} + \eta_{2m} = \eta_{3m} + \eta_{4m}, \quad \mathcal{H}_{0,m}^{B} : \eta_{1m} + \eta_{3m} = \eta_{2m} + \eta_{4m}, \\ \text{and} \quad &\mathcal{H}_{0,m}^{AB} : \eta_{1m} + \eta_{4m} = \eta_{2m} + \eta_{3m}, \quad m \in \{1, \ldots, M\}. \end{aligned} \tag{5.4}$$

Moreover, Dunnett- and Tukey-type contrast matrices are used for many-to-one and all-pairs comparisons of the RMTLs, respectively, as in Example 5.6. The block matrices $\mathbf{H}_\ell, \ell \in \{1, \dots, L\}$, for the local hypotheses are always chosen to be the rows of the global hypothesis matrices. The global hypothesis matrices all lead to the same global null hypothesis, that is, all RMTLs are equal across the groups for each respective event type. However, the local hypotheses differ between the different matrices.
The RMST difference is chosen as $\delta = 0$ when simulating under the null and as $\delta = 1.5$ when simulating under the alternative hypothesis. Since the survival settings *exp early*, *exp late* and *exp prop* defined in the appendix result in the same survival functions under the null hypothesis, the results for these scenarios are only included once in the figures and tables, respectively. The same holds for the settings *Weib late* and *Weib prop*.
Balanced and unbalanced designs with sample sizes $\mathbf{n} = (n_1, n_2, n_3, n_4) = K \cdot (60, 60, 60, 60)$ and $\mathbf{n} = K \cdot (128, 44, 52, 16)$ are considered, where the factor $K \in \{1, 5, 25\}$ results in small, medium, and large samples, respectively.
In total, $N_{sim} = 2000$ simulation runs with $B = 1000$ resampling iterations were conducted. The level of significance is set to $\alpha = 0.05$ and the terminal time point to $\tau = 10$.
We included the following methods in our simulation study: multiple asymptotic Wald-type tests as in Section 5.3.1 (*asymptotic*), global asymptotic Wald-type tests as in Section 5.2.1 adjusted with the Bonferroni-correction (*asymptotic_bonf*), and global studentized permutation tests as in Section 5.2.2 adjusted with the Bonferroni-correction (*permutation_bonf*). In our first simulations, we also compared a pooled bootstrap, wild bootstrap, and groupwise bootstrap method similar to that in Section 4. Additionally, we considered a random p-value permutation approach similar as the prepivoting method in [16]. However, the results of all four methods were not as convincing in terms of type I error control and/or power. Moreover, the runtime of the random p-value permutation approach was quite high since, for each permutation sample, the calculations need to be done for several groupwise bootstrap samples. Consequently, we focused on the three above-mentioned methods.

5.4.2 Simulation Results

The boxplots in Figures 10, 11, and 12 summarize the rejection levels under all null hypotheses. It is observable that the multiple asymptotic Wald-type tests of Section 5.3.1 as well as the Bonferroni-corrected global asymptotic Wald-type tests of Section 5.2.1 can not control the type I error in unbalanced designs with smaller sample sizes as they perform too liberal in all scenarios. When considering Dunnett- and Tukey-type contrast matrices, the empirical family wise error rates are exceeding even 50% in some scenarios. The Bonferroni-corrected permutation tests also have a slight liberal behaviour in some of these scenarios but not nearly as dramatic. The highest empirical family wise error rates for the Bonferroni-corrected permutation tests are only up to 10%. Those are reached under the non-exchangeable survival distribution settings *Weib scale* and *Weib shape* for unbalanced small sample sizes. A possible reason for the liberal behaviour of the tests for unbalanced small sample sizes could be that it is more likely to observe no event of a specific type in at least one of the samples. In this case, the permutation approach may still use the information of the events of the same type that occur in other groups through the randomization across groups. However, the asymptotic approach can not benefit from observations in other samples and, thus, probably underestimates the variance systematically. Even for unbalanced designs and medium sample sizes of 80–640 observations, the liberality of the asymptotic approaches is still notable. In balanced designs, this issue is only slightly present, even for small sample sizes with 60 observations per group. For large sample sizes, all methods seem to perform quite well under the global null hypothesis in terms of family wise error rate control, which underlines the asymptotic validity of the proposed tests.
Figures 30, 31 and 32 in the appendix visualize the empirical rejection rates of the global null hypothesis under the alternative hypothesis, i.e., the empirical (global) powers. It is observable that all methods have a comparable power in all scenarios with balanced designs or large sample sizes. Moreover, the power naturally increases for larger sample sizes. In unbalanced designs with small to medium sample sizes, the multiple and Bonferroni-corrected asymptotic tests have usually a higher power than the Bonferroni-corrected permutation tests. Here, the multiple asymptotic tests that take the multivariate distribution of the test statistics into account (Section 5.3.1) are slightly more powerful than the asymptotic tests with Bonferroni-correction. However, both asymptotic testing procedures performed too liberal in unbalanced designs with small to medium sample sizes and, thus, we do not recommend their application in these scenarios. Furthermore, it is observable that only under a few scenarios, all methods can detect the alternative in unbalanced designs with small sample sizes as the most rejection rates are similar as under the null hypothesis. A possible explanation for this may be the small sample size of 16 in group four, which was sampled with different RMTLs under the alternative.

Recommendations and limitations In view of the present simulation results, we recommend the use of the Bonferroni-corrected permutation tests, especially if the sample sizes are small, due to the best family-wise error rate control. On the other hand, for large sample sizes, all methods yield rather similar results regarding

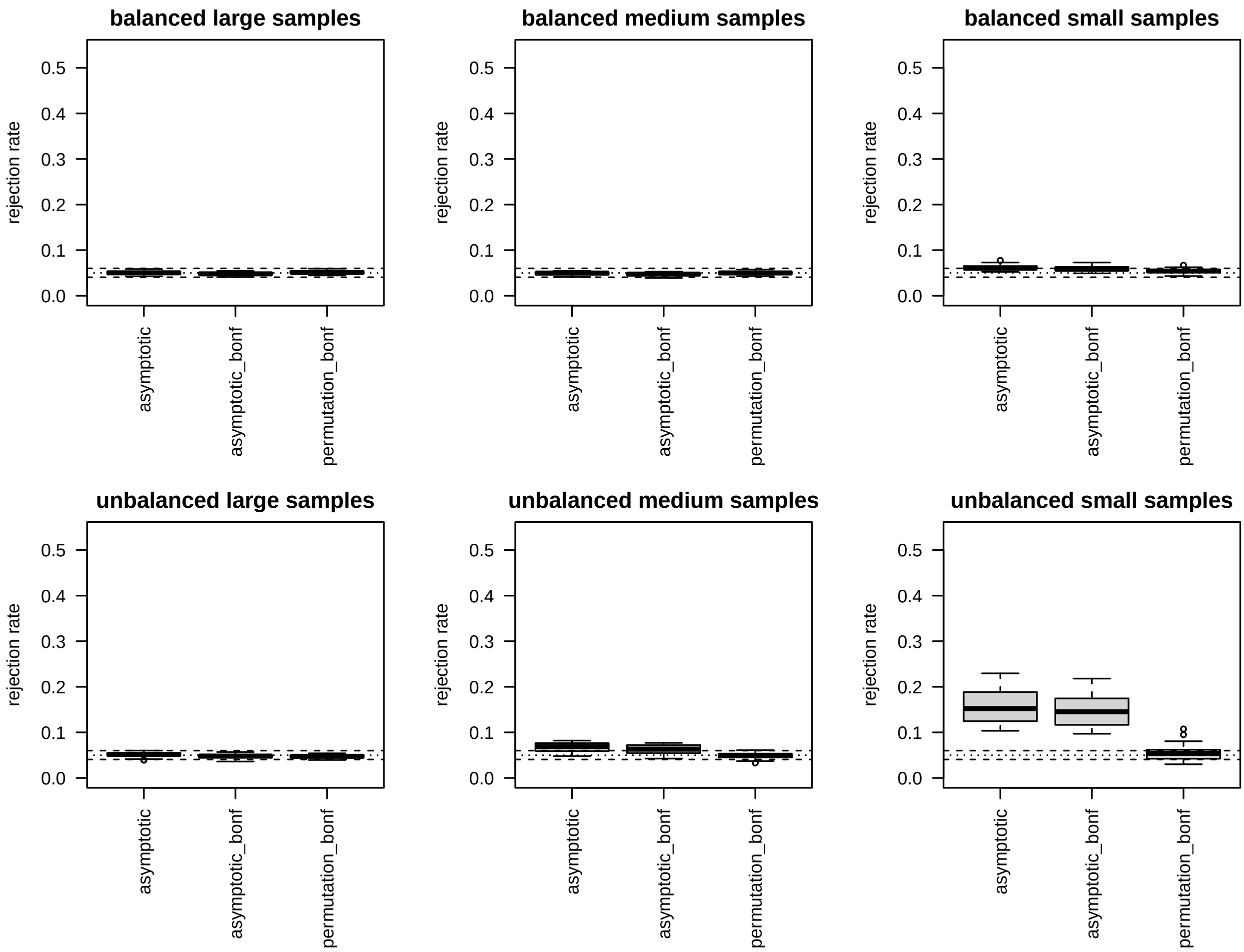


Figure 10: Empirical family wise error rates for the 2-by-2 design across all scenarios under the global null hypothesis. The dotted line represents the desired global level of 0.05 and the dashed lines represent the borders of the binomial interval [0.0405, 0.06].

family-wise error rate control and power, but asymptotic tests might be preferred due to lower computational demands. In this case, it should be noted that the multiple asymptotic tests are slightly more powerful than the Bonferroni-corrected asymptotic tests in general. Whether a sample size is considered small or large also depends on the specific study design, including the number of competing risks, groups, and hypotheses. For instance, in a balanced design with three competing risks, the simulation results indicate that a sample size of 60 individuals per group is sufficient for minor differences between methods. However, for unbalanced designs, notable differences persist for a sample size of 80 individuals in the smallest group but vanish when the smallest group contains 400 individuals. When faced with extremely unbalanced datasets with small sample sizes, additional challenges as, e.g., nearly no power under the alternative, may arise.

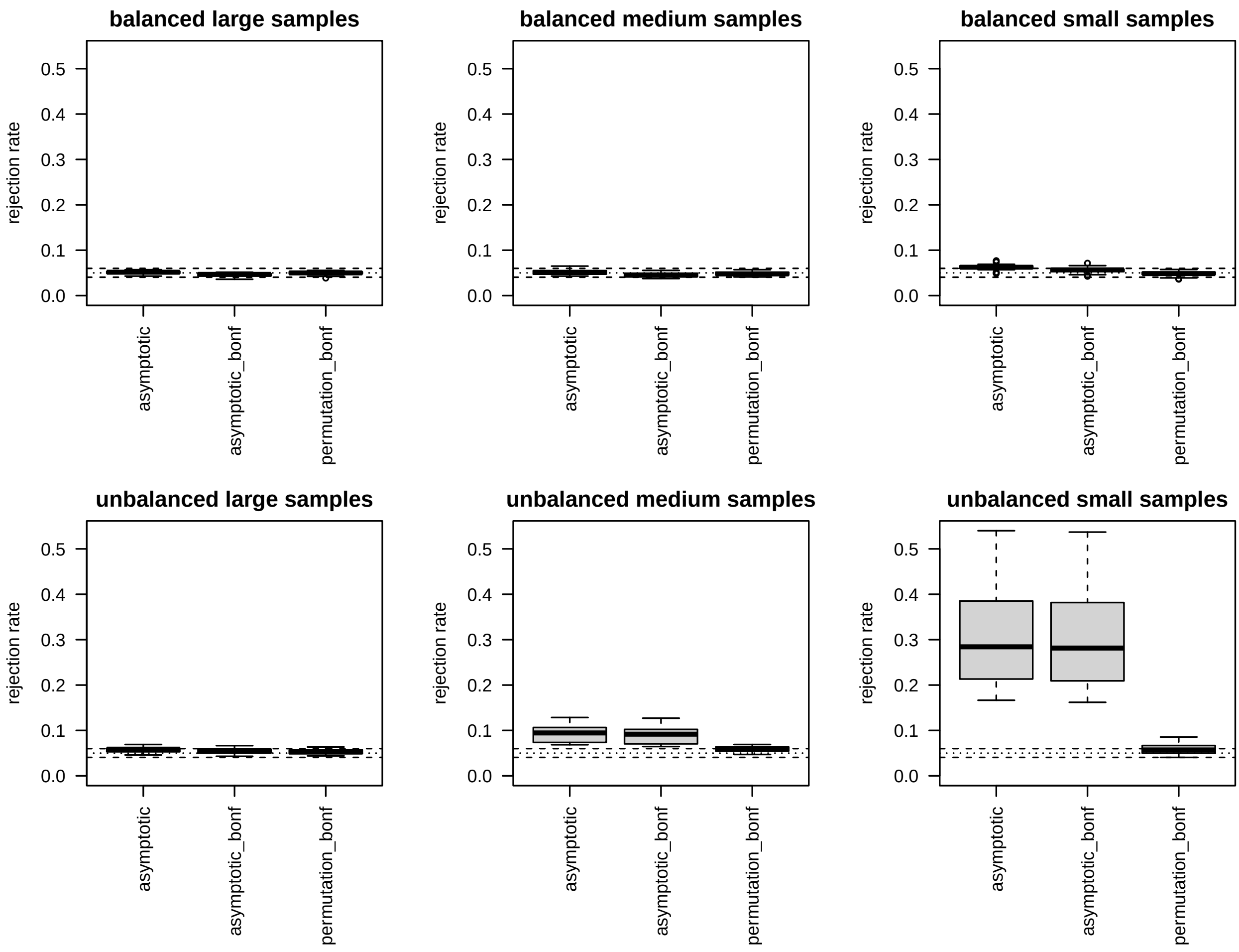


Figure 11: Empirical family wise error rates for the Dunnett-type contrast hypotheses across all scenarios under the global null hypothesis. The dotted line represents the desired global level of 0.05 and the dashed lines represent the borders of the binomial interval [0.0405, 0.06].

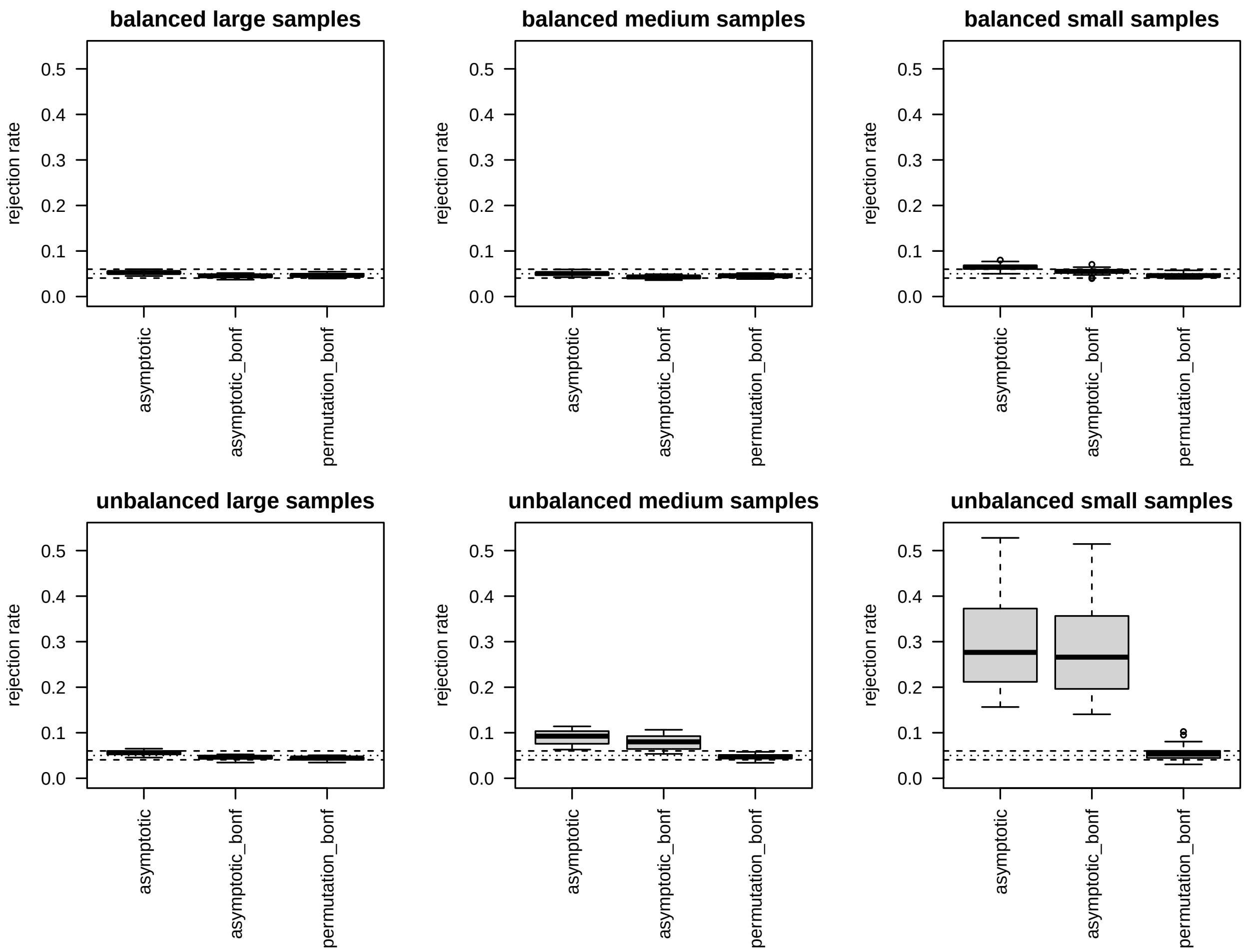


Figure 12: Empirical family wise error rates for the Tukey-type contrast hypotheses across all scenarios under the global null hypothesis. The dotted line represents the desired global level of 0.05 and the dashed lines represent the borders of the binomial interval [0.0405, 0.06].

5.5 Data Example about Blood and Marrow Transplantation

In order to illustrate the proposed methods, we analyze the data set `ebmt2` in the R package `mstate` [18, 19, 65] from the European Society for Blood and Marrow Transplantation. The data consists of 8966 leukemia patients who underwent a bone marrow transplantation. An initial statistical analysis [36] focused on reduced rank models for proportional cause-specific hazard models. First of all, the data set contains `time`, which is the time in months from transplantation to death or the last follow-up, and `status`, which indicates the survival status; for simplicity, we aggregate the status levels into the following $M = 3$ causes of the death, next to censoring: relapse (1), graft-versus-host disease (2), and all other causes (3). We included the following factor variables in our analysis within a factorial 2-by-2 design; see Example 5.3: (A) `match`: yes/no, according to whether the donor's and the recipient's genders matched; (B) `tcd`: yes/no, depending on whether a T-cell depletion took place. Because it was unknown for 2856 patients whether a T-cell depletion took place or not, we assumed the missingness to have been completely at random, and the incomplete records were removed from our further analysis. Hence, $n = 6110$ patients remained. Thereof, 1296 (3313) patients with donor-recipient gender match did (not) receive a T-cell depletion. For those without a match, the numbers were 424 and 1077, respectively. An illustration of the resulting Aalen-Johansen estimators of the cumulative incidence functions can be found in Figure 13.

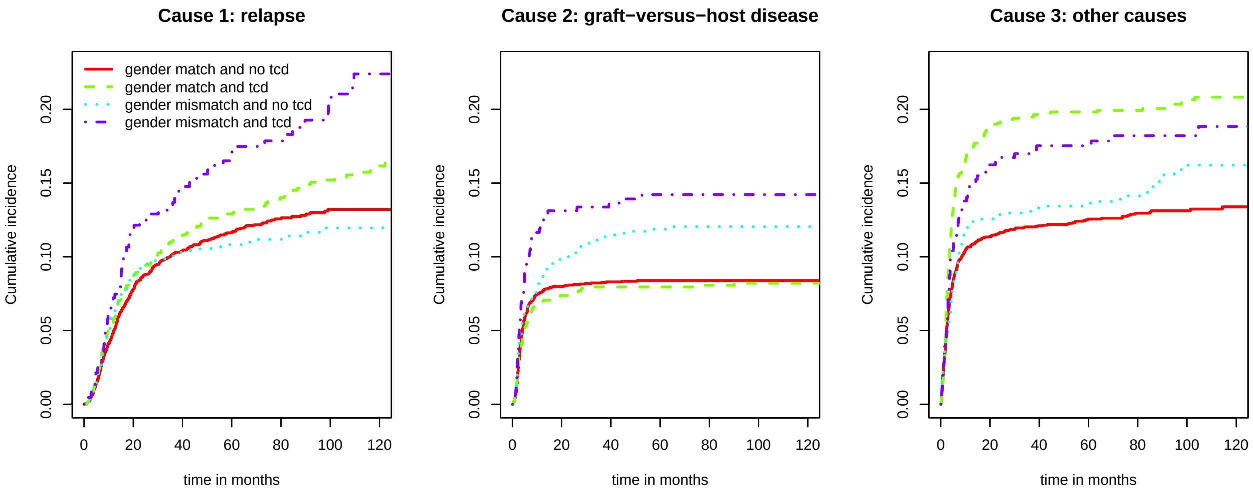


Figure 13: Aalen-Johansen estimators of the cumulative incidence functions regarding the data example for the different causes and groups

Here, the question of interest is whether the donor-recipient gender match and/or the T-cell depletion have a main or interaction effect on any event type-specific RMTL. If there is a significant effect, we are also interested in which specific main/interaction effects are present and which of the event types are affected. Hence, the nine hypothesis matrices for the multiple testing problem are, for the event type $m \in \{1, 2, 3\}$ and the effect ($\ell = 1$ for main A, $\ell = 2$ for main B, and $\ell = 3$ for interaction), $\mathbf{h}_\ell \otimes \mathbf{e}'_m, m \in \{1, 2, 3\}$, where $\mathbf{h}_1 = [1, 1, -1, -1]$, $\mathbf{h}_2 = [1, -1, 1, -1]$, and $\mathbf{h}_3 = [1, -1, -1, 1]$. The vector of no RMTL differences, i.e., $\mathbf{c} = \mathbf{0}_9$, is tested under the global null hypothesis. This results in the hypotheses (5.4) as in the simulation study. We used the terminal time point ten years, i.e., $\tau = 120$ months. The estimated RMTLs up to $\tau = 120$ for the different groups and causes can be found in Table 2.

The resulting method-specific and adjusted p-values based on $B = 19999$ resampling iterations are presented in Table 3. Comparing the adjusted p-values with the global level of significance allows for testing all nine local hypotheses simultaneously. Due to the large sample sizes, all methods should yield reliable results in terms of type I error control regarding the simulation results of Section 5.4. For $\alpha = 0.05$, all methods indicate that $\mathcal{H}^B_{0,1}, \mathcal{H}^B_{0,3}$, and $\mathcal{H}^A_{0,2}$ can be rejected simultaneously. Thus, there is a significant main effect of the T-cell

Group	Cause 1: relapse	Cause 2: graft-versus-host disease	Cause 3: other causes
gender match and no tcd	12.526	9.595	14.446
gender match and tcd	14.159	9.151	22.885
gender mismatch and no tcd	11.883	13.125	16.291
gender mismatch and tcd	18.518	15.982	20.296

Table 2: RMTL estimation for the data example

Method	$\mathcal{H}_{0,1}^{A}$	$\mathcal{H}_{0,2}^{A}$	$\mathcal{H}_{0,3}^{A}$	$\mathcal{H}_{0,1}^{B}$	$\mathcal{H}_{0,2}^{B}$	$\mathcal{H}_{0,3}^{B}$	$\mathcal{H}_{0,1}^{AB}$	$\mathcal{H}_{0,2}^{AB}$	$\mathcal{H}_{0,3}^{AB}$
asymptotic	0.663	< 0.001	1.000	0.008	0.950	< 0.001	0.295	0.773	0.614
asymptotic_bonf	1.000	< 0.001	1.000	0.008	1.000	< 0.001	0.396	1.000	1.000
permutation_bonf	1.000	< 0.001	1.000	0.009	1.000	< 0.001	0.407	1.000	1.000

Table 3: Adjusted p-values for the data example

depletion on the RMTL for relapse and other causes and a significant main effect of the donor-recipient gender match on the RMTL regarding the graft-versus-host disease.
As a word of caution, no clinical conclusions should be drawn from this analysis since the data were simplified. For example, aspects that would be relevant for causal interpretations were not taken into consideration. Instead, the present real data analysis was meant to illustrate the potential of our new statistical techniques. Moreover, since this data example is a classical application of competing risks analysis in a medical context, we want to emphasize that our methodology is also applicable and may also be relevant in non-medical domains as, e.g., reliability engineering [56].

5.6 Proofs of Section 5

Proof of Theorem 5.1 Firstly, let $i \in \{1, ..., k\}$ be arbitrary but fixed. By Theorem 4.1 in [26], we have

$$n_i^{1/2}\left(\widehat{A}_{i1} - A_{i1}, ..., \widehat{A}_{iM} - A_{iM}\right) \xrightarrow{d} (U_{i1}, ..., U_{iM}) \tag{5.5}$$

as $n \to \infty$ on $(D[0,\tau])^M$ equipped with the sup-norm, where $U_{i1}, ..., U_{iM}$ are centered Gaussian-martingales with

$$\begin{aligned}\mathbb{C}ov(U_{im_1}(t), U_{im_1}(s)) &= \int_{[0,\min\{t,s\}]} \frac{1-\Delta A_{im_1}}{y_i}\,\mathrm{d}A_{im_1} =: \sigma_{im_1m_1}(\min\{t,s\}),\\ \mathbb{C}ov(U_{im_1}(t), U_{im_2}(s)) &= -\int_{[0,\min\{t,s\}]} \frac{\Delta A_{im_1}}{y_i}\,\mathrm{d}A_{im_2} =: \sigma_{im_1m_2}(\min\{t,s\})\end{aligned}$$

and $y_i(t) := P(X_{i1} \geqslant t)$ for all $t, s \in [0,\tau]; m_1, m_2 \in \{1, \dots, M\}; m_1 \neq m_2$. By Section B, the limit $(U_{i1}, ..., U_{iM})$ is separable.
Then, we consider the functional

$$\begin{aligned}\Phi :\ (BV_K[0,\tau])^M \to \mathbb{R}^M, \quad \Phi(\Lambda_1, ..., \Lambda_M) &:= \left(\int_0^\tau \widetilde{\psi}\left(\widetilde{\phi}\left(-\sum_{\widetilde{m}=1}^M \Lambda_{\widetilde{m}}\right)_-, \Lambda_m\right)\mathrm{d}t\right)_{m\in\{1,...,M\}}\\ &= \left(\int_0^\tau \int_{[0,t]} \prod_{x\in[0,u)}\left\{1 - \mathrm{d}\left(\sum_{\widetilde{m}=1}^M \Lambda_{\widetilde{m}}\right)(x)\right\}\mathrm{d}\Lambda_m(u)\,\mathrm{d}t\right)_{m\in\{1,...,M\}}\end{aligned}$$

for some $K < \infty$ with $\widetilde{\psi} : \widetilde{D}[0,\tau] \times BV_K[0,\tau] \to D[0,\tau], \widetilde{\phi} : BV_{MK}[0,\tau] \to D[0,\tau)$ as in Section A. Here and throughout, we define the jump $\Delta\Lambda(0)$ of a function $\Lambda \in BV_K[0,\tau]$ at 0 as $\Lambda(0)$. Note that $\Phi(A_{i1}, ..., A_{iM}) = (\eta_{i1}, ..., \eta_{iM})$ and $\Phi(\widehat{A}_{i1}, ..., \widehat{A}_{iM}) = (\widehat{\eta}_{i1}, ..., \widehat{\eta}_{iM})$ holds. We aim to apply the delta-method. Therefore, we firstly show that Φ is Hadamard-differentiable at $(A_{i1}, ..., A_{iM})$ with Hadamard-derivative $\Phi'_{(A_{i1},...,A_{iM})}$ at $(\alpha_1, ..., \alpha_M) \in (D[0,\tau])^M$ given by

$$\left(\int_0^\tau \left(\int_{[0,t]} \prod_{x\in[0,u)} (1 - \mathrm{d}A_i(x))\,\mathrm{d}\alpha_m(u) - \int_{[0,t]} \prod_{x\in[0,u)} (1 - \mathrm{d}A_i(x)) \int_{[0,u)} \frac{\mathrm{d}\sum_{\widetilde{m}=1}^M \alpha_{\widetilde{m}}}{1-\Delta A_i}\,\mathrm{d}A_{im}(u)\right)\mathrm{d}t\right)_{m\in\{1,...,M\}}.$$

Here, the integrals with respect to α_m are defined via integration by parts because α_m need not have finite variation. The same holds for other integrals of this kind below. In order to prove the Hadamard-differentiability, we aim to apply the chain rule (Lemma 3.10.3 in [74]). This yields

$$\Phi'_{(A_{i1},...,A_{iM})}(\alpha_1, ..., \alpha_M) = \left(\int_0^\tau \widetilde{\psi}'_{(\widetilde{\phi}(-A_i)_-, A_{im})}\left(\widetilde{\phi}'_{-A_i}\left(-\sum_{\widetilde{m}=1}^M \alpha_{\widetilde{m}}\right)_-, \alpha_m\right)(t)\,\mathrm{d}t\right)_{m\in\{1,...,M\}}$$

for $\alpha_1, ..., \alpha_M \in D[0,\tau]$. Note here that $D[0,\tau) \ni \Lambda \mapsto \Lambda_- \in \widetilde{D}[0,\tau]$ and the final integral functional is linear, so it equals its Hadamard-derivative, respectively. As in Section 3.3, one can show that

$$\begin{aligned}
&\widetilde{\psi}'_{(\widetilde{\phi}(-A_i)_-, A_{im})}(\alpha,\beta) = \int_{[0,.]} S_{i-}\, \mathrm{d}\beta + \int_{[0,.]} \alpha\, \mathrm{d}A_{im}, \quad m \in \{1,\dots,M\},\\
\text{and}\qquad &\widetilde{\phi}'_{-A_i}(\beta) = S_i(.) \int_{[0,.]} \frac{1}{1-\Delta A_i}\, \mathrm{d}\beta
\end{aligned}$$

holds for all $\alpha \in \widetilde{D}[0,\tau], \beta \in D[0,\tau]$ analogously to Lemma 3.10.18 and Lemma 3.10.32 in [74]. Thus, it follows that

$$\begin{aligned}
&\Phi'_{(A_{i1},\dots,A_{iM})}(\alpha_1,...,\alpha_M)\\
&= \left(\int_0^\tau \left(\int_{[0,t]} S_{i-}\, \mathrm{d}\alpha_m - \int_{[0,t]} S_{i-}(s) \int_{[0,s)} \frac{\mathrm{d}\sum_{\widetilde{m}=1}^M \alpha_{\widetilde{m}}}{1-\Delta A_i}\, \mathrm{d}A_{im}(s) \right) \mathrm{d}t \right)_{m\in\{1,\dots,M\}}\\
&= \left(\int_0^\tau \left(\int_{[0,t]} S_{i-}\, \mathrm{d}\alpha_m - \int_{[0,t)} \frac{\int_{(u,t]} S_{i-}\, \mathrm{d}A_{im}}{1-\Delta A_i(u)}\, \mathrm{d} \sum_{\widetilde{m}=1}^M \alpha_{\widetilde{m}}(u) \right) \mathrm{d}t \right)_{m\in\{1,\dots,M\}}\\
&= \left(\int_0^\tau \left(\int_{[0,t]} \frac{S_i}{1-\Delta A_i}\, \mathrm{d}\alpha_m - \int_{[0,t]} \frac{F_{im}(t)-F_{im}(u)}{1-\Delta A_i(u)}\, \mathrm{d} \sum_{\widetilde{m}=1}^M \alpha_{\widetilde{m}}(u) \right) \mathrm{d}t \right)_{m\in\{1,\dots,M\}}\\
&= \left(\int_0^\tau \left(\int_{[0,t]} \frac{S_i(u)-F_{im}(t)+F_{im}(u)}{1-\Delta A_i(u)}\, \mathrm{d}\alpha_m(u) - \int_{[0,t]} \frac{F_{im}(t)-F_{im}(u)}{1-\Delta A_i(u)}\, \mathrm{d} \sum_{\widetilde{m}\neq m} \alpha_{\widetilde{m}}(u) \right) \mathrm{d}t \right)_{m\in\{1,\dots,M\}}\\
&= \left(\int_0^\tau \left(\int_{[0,t]} \frac{1-\sum_{\widetilde{m}\neq m} F_{i\widetilde{m}}(u) - F_{im}(t)}{1-\Delta A_i(u)}\, \mathrm{d}\alpha_m(u) - \int_{[0,t]} \frac{F_{im}(t)-F_{im}(u)}{1-\Delta A_i(u)}\, \mathrm{d} \sum_{\widetilde{m}\neq m} \alpha_{\widetilde{m}}(u) \right) \mathrm{d}t \right)_{m\in\{1,\dots,M\}}\\
&= \Bigg(\int_{[0,\tau)} \frac{(\tau-u)\left(1-\sum_{\widetilde{m}\neq m} F_{i\widetilde{m}}(u)\right) - \int_u^\tau F_{im}(t)\, \mathrm{d}t}{1-\Delta A_i(u)}\, \mathrm{d}\alpha_m(u)\\
&\qquad + \int_{[0,\tau)} \frac{(\tau-u)F_{im}(u) - \int_u^\tau F_{im}(t)\, \mathrm{d}t}{1-\Delta A_i(u)}\, \mathrm{d} \sum_{\widetilde{m}\neq m} \alpha_{\widetilde{m}}(u) \Bigg)_{m\in\{1,\dots,M\}}
\end{aligned}$$

for $\alpha_1, ..., \alpha_M \in D[0,\tau]$, where here and throughout, we write $\sum_{\widetilde{m}\neq m}$ instead of $\sum_{\widetilde{m}=1,\widetilde{m}\neq m}^M$ for the sake of brevity. Thus, an application of the delta-method (Theorem 3.10.4 in [74]) yields

$$\begin{aligned}
{n_i}^{1/2}(\widehat{\eta}_{im} - \eta_{im})_{m\in\{1,\dots,M\}} &\xrightarrow{d} \Phi'_{(A_{i1},\dots,A_{iM})}(U_{i1},...,U_{iM})\\
&= \Bigg(\int_{[0,\tau)} \frac{(\tau-u)\left(1-\sum_{\widetilde{m}\neq m} F_{i\widetilde{m}}(u)\right) - \int_u^\tau F_{im}(t)\, \mathrm{d}t}{1-\Delta A_i(u)}\, \mathrm{d}U_{im}(u)\\
&\qquad + \int_{[0,\tau)} \frac{(\tau-u)F_{im}(u) - \int_u^\tau F_{im}(t)\, \mathrm{d}t}{1-\Delta A_i(u)}\, \mathrm{d} \sum_{\widetilde{m}\neq m} U_{i\widetilde{m}}(u) \Bigg)_{m\in\{1,\dots,M\}}
\end{aligned}$$

as $n \to \infty$, where the limit is a centered M-dimensional random vector which follows a multivariate normal distribution. Its covariance matrix $\mathbf{\Sigma}_i := [\Sigma_{im_1m_2}]_{m_1,m_2\in\{1,\dots,M\}}$ has the following entries: $\Sigma_{im_1m_2}$ given by

$$\begin{aligned}
&\int_{[0,\tau)} \frac{\left\{(\tau-u)\left(1-\sum\limits_{\widetilde{m}\neq m_1} F_{i\widetilde{m}}(u)\right) - \int_u^\tau F_{im_1}(t)\, \mathrm{d}t\right\}\left\{(\tau-u)\left(1-\sum\limits_{\widetilde{m}\neq m_2} F_{i\widetilde{m}}(u)\right) - \int_u^\tau F_{im_2}(t)\, \mathrm{d}t\right\}}{(1-\Delta A_i(u))^2}\, \mathrm{d}\sigma_{im_1m_2}(u)\\
&+ \int_{[0,\tau)} \frac{\left\{(\tau-u)\left(1-\sum\limits_{\widetilde{m}\neq m_1} F_{i\widetilde{m}}(u)\right) - \int_u^\tau F_{im_1}(t)\, \mathrm{d}t\right\}\left\{(\tau-u)F_{im_2}(u) - \int_u^\tau F_{im_2}(t)\, \mathrm{d}t\right\}}{(1-\Delta A_i(u))^2}\, \mathrm{d}\sum_{\widetilde{m}\neq m_2} \sigma_{im_1\widetilde{m}}(u)\\
&+ \int_{[0,\tau)} \frac{\left\{(\tau-u)\left(1-\sum\limits_{\widetilde{m}\neq m_2} F_{i\widetilde{m}}(u)\right) - \int_u^\tau F_{im_2}(t)\, \mathrm{d}t\right\}\left\{(\tau-u)F_{im_1}(u) - \int_u^\tau F_{im_1}(t)\, \mathrm{d}t\right\}}{(1-\Delta A_i(u))^2}\, \mathrm{d}\sum_{\widetilde{m}\neq m_1} \sigma_{im_2\widetilde{m}}(u)\\
&+ \int_{[0,\tau)} \frac{\left\{(\tau-u)F_{im_1}(u) - \int_u^\tau F_{im_1}(t)\, \mathrm{d}t\right\}\left\{(\tau-u)F_{im_2}(u) - \int_u^\tau F_{im_2}(t)\, \mathrm{d}t\right\}}{(1-\Delta A_i(u))^2}\, \mathrm{d}\sum_{\widetilde{m}\neq m_1}\sum_{\check{m}\neq m_2} \sigma_{i\widetilde{m}\check{m}}(u)
\end{aligned} \tag{5.6}$$

for all $m_1, m_2 \in \{1, \ldots, M\}$. Since $i \in \{1, ..., k\}$ was arbitrary and the groups are independent, the statement of the theorem follows with

$$\mathbf{\Sigma} := \bigoplus_{i=1}^{k} \left(\kappa_i^{-1} \mathbf{\Sigma}_i\right) \tag{5.7}$$

by Assumption 5.1. □

Proof of Theorem 5.2 First of all, we investigate the covariance matrix estimator $\widehat{\mathbf{\Sigma}}$.

Lemma 5.1. *Under Assumption 5.1, we have $\widehat{\mathbf{\Sigma}} \xrightarrow{P} \mathbf{\Sigma}$ as $n \to \infty$.*

Proof of Lemma 5.1. By (5.5), Slutsky's lemma, and the continuous mapping theorem, we have

$$\sup_{t \in [0,\tau]} \left|\widehat{A}_{im}(t) - A_{im}(t)\right| \xrightarrow{P} 0, i \in \{1, \ldots, k\}, m \in \{1, \ldots, M\},$$

as $n \to \infty$. Moreover, it holds

$$\sup_{t \in [0,\tau]} |\widehat{\sigma}_{im_1m_2}(t) - \sigma_{im_1m_2}(t)| \xrightarrow{P} 0, i \in \{1, \ldots, k\}, m_1, m_2 \in \{1, \ldots, M\},$$

as $n \to \infty$ by [26]. Note that $\Delta A_i(u) \leqslant 1 - S_{i-}(\tau) < 1, i \in \{1, \ldots, k\}$, holds for all $u \in [0, \tau)$. Hence, the covariance estimator $\widehat{\mathbf{\Sigma}}_i$ is a continuous functional of $\widehat{A}_{im_1}$ and $\widehat{\sigma}_{im_1m_2}$, $m_1, m_2 \in \{1, \ldots, M\}$ for all $i \in \{1, \ldots, k\}$ and, thus, the consistency $\widehat{\mathbf{\Sigma}} \xrightarrow{P} \mathbf{\Sigma}$ as $n \to \infty$ follows by Assumption 5.1. □

Furthermore, we need that $\mathbf{\Sigma}$ is positive definite. The following lemma ensures this under Assumption 5.2.

Lemma 5.2. *Under Assumptions 5.1 and 5.2, $\mathbf{\Sigma}$ is positive definite.*

Proof of Lemma 5.2. Since $\mathbf{\Sigma}$ is positive definite whenever $\mathbf{\Sigma}_i$ is positive definite for all $i \in \{1, \ldots, k\}$, we fix $i \in \{1, ..., k\}$. Now, let $\mathbf{a} = (a_1, ..., a_M)' \in \mathbb{R}^M \backslash \{\mathbf{0}_M\}$ be arbitrary. We aim to show $\mathbf{a}'\mathbf{\Sigma}_i\mathbf{a} > 0$. By the proof of Theorem 5.1, it holds that

$$\mathbf{a}'\mathbf{\Sigma}_i\mathbf{a} = \mathbb{V}ar\left(\sum_{m=1}^{M} a_m \left(\int_{[0,\tau)} f_{im} \, \mathrm{d}U_{im} + \int_{[0,\tau)} g_{im} \, \mathrm{d}\sum_{\widetilde{m} \neq m} U_{i\widetilde{m}}\right)\right) = \mathbb{V}ar\left(\sum_{m=1}^{M} \int_{[0,\tau)} h_{im} \, \mathrm{d}U_{im}\right)$$

with

$$\begin{aligned} f_{im}(u) &:= \frac{(\tau - u)\left(1 - \sum_{\widetilde{m} \neq m} F_{i\widetilde{m}}(u)\right) - \int_u^\tau F_{im}(t) \, \mathrm{d}t}{1 - \Delta A_i(u)}, \\ g_{im}(u) &:= \frac{(\tau - u)F_{im}(u) - \int_u^\tau F_{im}(t) \, \mathrm{d}t}{1 - \Delta A_i(u)} \\ \text{and} \quad h_{im}(u) &:= a_m f_{im}(u) + \sum_{\widetilde{m} \neq m} a_{\widetilde{m}} g_{i\widetilde{m}}(u) \end{aligned}$$

for all $u \in [0, \tau)$. We can calculate this variance further as

$$\begin{aligned} \mathbf{a}'\mathbf{\Sigma}_i\mathbf{a} &= \sum_{m=1}^{M} \mathrm{E}\left(\left(\int_{[0,\tau)} h_{im} \, \mathrm{d}U_{im}\right)^2\right) + \sum_{m=1}^{M} \sum_{\widetilde{m} \neq m} \mathrm{E}\left(\int_{[0,\tau)} h_{im} \, \mathrm{d}U_{im} \int_{[0,\tau)} h_{i\widetilde{m}} \, \mathrm{d}U_{i\widetilde{m}}\right) \\ &= \sum_{m=1}^{M} \int_{[0,\tau)} h_{im}^2 \frac{1 - \Delta A_{im}}{y_i} \, \mathrm{d}A_{im} - \sum_{m=1}^{M} \sum_{\widetilde{m} \neq m} \int_{[0,\tau)} h_{im} h_{i\widetilde{m}} \frac{\Delta A_{im}}{y_i} \, \mathrm{d}A_{i\widetilde{m}} \\ &= \sum_{m=1}^{M} \int_{[0,\tau)} \frac{h_{im}^2}{y_i} \, \mathrm{d}A_{im} - \sum_{m=1}^{M} \sum_{\widetilde{m}=1}^{M} \int_{[0,\tau)} h_{im} h_{i\widetilde{m}} \frac{\Delta A_{im}}{y_i} \, \mathrm{d}A_{i\widetilde{m}} \\ &= \sum_{m=1}^{M} \int_{[0,\tau)} \frac{h_{im}^2}{y_i} \, \mathrm{d}A_{im}^c + \sum_{x \in \mathcal{D}_i} \frac{\sum_{m=1}^{M} h_{im}^2(x) \Delta A_{im}(x) - \left(\sum_{m=1}^{M} h_{im}(x) \Delta A_{im}(x)\right)^2}{y_i(x)} \end{aligned} \tag{5.8}$$

where $\mathcal{D}_i = \{x \in [0, \tau) : \Delta A_i(x) > 0\}$ is the set of discontinuity time points and

$$A_{im}^c(x) := A_{im}(x) - \sum_{y \leqslant x, y \in \mathcal{D}_i} \Delta A_{im}(y), m \in \{1, \ldots, M\},$$

denotes the continuous part of A_{im} at $x \in [0, \tau)$. The Cauchy-Schwarz inequality yields

$$\left(\sum_{m=1}^{M} h_{im}(x)\Delta A_{im}(x)\right)^2 \leqslant \left(\sum_{m=1}^{M} h_{im}^2(x)\Delta A_{im}(x)\right)\left(\sum_{m=1}^{M} \Delta A_{im}(x)\right)$$

and, thus,

$$\sum_{m=1}^{M} h_{im}^2(x)\Delta A_{im}(x) - \left(\sum_{m=1}^{M} h_{im}(x)\Delta A_{im}(x)\right)^2 \geqslant \sum_{m=1}^{M} h_{im}^2(x)\Delta A_{im}(x)\left(1 - \Delta A_i(x)\right) \geqslant 0$$

for all $x \in \mathcal{D}_i$. Let $m^* \in \{1, ..., M\}$ be the index with $|a_{m^*}| = \max\{|a_1|, ..., |a_M|\} > 0$. Then, it holds

$$\begin{aligned}
|h_{im^*}(u)| &= \left| a_{m^*} f_{im^*}(u) + \sum_{\widetilde{m}\neq m^*} a_{\widetilde{m}} g_{i\widetilde{m}}(u) \right| \geqslant |a_{m^*}| f_{im^*}(u) - \left| \sum_{\widetilde{m}\neq m^*} a_{\widetilde{m}} g_{i\widetilde{m}}(u) \right| \\
&\geqslant |a_{m^*}| \left(f_{im^*}(u) + \sum_{\widetilde{m}\neq m^*} g_{i\widetilde{m}}(u) \right) = |a_{m^*}| \frac{\int_u^\tau S_i(t)\,\mathrm{d}t}{1-\Delta A_i(u)} \geqslant |a_{m^*}|(\tau - u)\frac{S_{i-}(\tau)}{1-\Delta A_i(u)} > 0
\end{aligned}$$

for all $u \in [0, \tau)$ since $|f_{im^*}(u)| = f_{im^*}(u)$, $|g_{i\widetilde{m}}(u)| = -g_{i\widetilde{m}}(u)$ and $1 - \Delta A_i(u) \geqslant S_{i-}(\tau) > 0$ for all $u \in [0, \tau), \widetilde{m} \in \{1, \ldots, M\}$. Note that $A_{im^*-}(\tau) > 0$ due to Assumption 5.2. Hence, it follows that at least one of the summands in (5.8) is strictly positive and, thus, $\mathbf{a}'\mathbf{\Sigma}_i\mathbf{a} > 0$. □

For $\mathbf{\Sigma}$ positive definite, it holds that $\operatorname{rank}(\mathbf{H\Sigma H'}) = \operatorname{rank}(\mathbf{H})$. Furthermore, the consistency of the covariance matrix estimator provides

$$P\left(\operatorname{rank}\left(\mathbf{H}\widehat{\mathbf{\Sigma}}\mathbf{H}'\right) \neq \operatorname{rank}(\mathbf{H})\right) \to 0$$

as $n \to \infty$. Hence, it follows that $\left(\mathbf{H}\widehat{\mathbf{\Sigma}}\mathbf{H}'\right)^+ \xrightarrow{P} (\mathbf{H\Sigma H'})^+$ as $n \to \infty$. Theorem 5.1, Slutsky's lemma and Theorem 9.2.2 in [67] yield

$$\begin{aligned}
W_n(\mathbf{H}, \mathbf{c}) &= n(\mathbf{H}\widehat{\boldsymbol{\eta}} - \mathbf{c})'\left(\mathbf{H}\widehat{\mathbf{\Sigma}}\mathbf{H}'\right)^+(\mathbf{H}\widehat{\boldsymbol{\eta}} - \mathbf{c}) \\
&= \left(\mathbf{H}\left(\sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})\right)\right)'\left(\mathbf{H}\widehat{\mathbf{\Sigma}}\mathbf{H}'\right)^+\mathbf{H}\left(\sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})\right) \xrightarrow{d} \chi^2_{\operatorname{rank}(\mathbf{H})}
\end{aligned}$$

as $n \to \infty$ under the null hypothesis $\mathcal{H}_0$ in (5.1). □

Proof of Theorem 5.3 In the following, we denote all permutation counterparts of the counting processes and estimators with a π in the superscript, e.g., $Y_i^\pi, N_{im}^\pi, \widehat{A}_{im}^\pi, i \in \{1, \ldots, k\}, m \in \{1, \ldots, M\}$, and all counterparts for the pooled sample with a subscript $\bullet$ instead of i, e.g., $Y_\bullet := \sum_{i=1}^k Y_i, N_{\bullet m} := \sum_{i=1}^k N_{im}, \widehat{A}_{\bullet m}(t) := \int_{[0,t]} Y_\bullet^{-1}\,\mathrm{d}N_{\bullet m}, m \in \{1, \ldots, M\}$, as well as $\widehat{A}_\bullet := \sum_{m=1}^M \widehat{A}_{\bullet m}, \widehat{S}_\bullet(t) := \prod_{x \leqslant t}\left\{1 - \mathrm{d}\left(\sum_{m=1}^M \widehat{A}_{\bullet m}\right)(x)\right\}, \widehat{F}_{\bullet m}(t) := \int_{[0,t]} \widehat{S}_{\bullet -}(t)\ \mathrm{d}\widehat{A}_{\bullet m}, \widehat{\eta}_{\bullet m} := \int_0^\tau \widehat{F}_{\bullet m}(t)\ \mathrm{d}t$, and $\widehat{\boldsymbol{\eta}}_\bullet := \mathbf{1}_k \otimes (\widehat{\eta}_{\bullet 1}, ..., \widehat{\eta}_{\bullet M})'$ for all $t \geqslant 0$. Furthermore, set $y_\bullet := \sum_{i=1}^k \kappa_i y_i, \nu_{\bullet m}(t) := \sum_{i=1}^k \kappa_i P(X_{i1} \leqslant t, \delta_{i1} = m), F_{\bullet m} := \sum_{i=1}^k \kappa_i F_{im}, A_{\bullet m}(t) := \int_{[0,t]} y_\bullet^{-1}\,\mathrm{d}\nu_{\bullet m}, A_\bullet := \sum_{m=1}^M A_{\bullet m}$,

$$S_\bullet(t) := \prod_{x\in[0,t]}\{1 - \mathrm{d}A_\bullet(x)\}, \qquad \sigma_{\bullet m\widetilde{m}}(t) := \int_{[0,t]} (\mathbb{1}\{m = \widetilde{m}\} - \Delta A_{\bullet m}) y_\bullet^{-1}\,\mathrm{d}A_{\bullet\widetilde{m}}$$

for all $t \geqslant 0, m, \widetilde{m} \in \{1, \ldots, M\}$.

Lemma 5.3. *Under Assumption 5.1, we have*

$$n^{1/2}\left(\widehat{\boldsymbol{\eta}}^\pi - \widehat{\boldsymbol{\eta}}_\bullet\right) \xrightarrow{d^*} \mathbf{Z}^\pi \sim \mathcal{N}_{kM}(\mathbf{0}_{kM}, \mathbf{\Sigma}^\pi)$$

as $n \to \infty$. The definition of the covariance matrix $\mathbf{\Sigma}^\pi$ is given at the end of the proof.

Proof of Lemma 5.3. Let us consider the class

$$\mathcal{F} := \{(x, d) \mapsto \mathbb{1}\{x \geqslant t\}, (x, d) \mapsto \mathbb{1}\{x \leqslant t, d = m\} \mid t \in [0, \tau], m \in \{1, ..., M\}\}$$

with finite envelope function $F \equiv 1$. It can be shown that $\mathcal{F}$ is $P^{(X_{i1},D_{i1})}$-Donsker for all $i = 1, \dots, k$, e.g., by applying Theorem 2.6.8 in [74]. By Theorem 2.2, it follows that

$$n^{1/2}\left(n_i^{-1}(Y_i^\pi, N_{i1}^\pi, \dots, N_{iM}^\pi) - n^{-1}(Y_\bullet, N_{\bullet 1}, \dots, N_{\bullet M})\right)_{i\in\{1,\dots,k\}} \xrightarrow{d^*} (\overline{G}_i, G_{i1}, \dots, G_{iM})_{i\in\{1,\dots,k\}} \tag{5.9}$$

on $(D[0,\tau])^{k(M+1)}$ as $n \to \infty$, where $(\overline{G}_i, G_{i1}, \dots, G_{iM})_{i\in\{1,\dots,k\}}$ denotes a tight centered Gaussian process with covariance structure

$$\begin{aligned}
\mathrm{E}\left(\overline{G}_i(t)\overline{G}_j(s)\right) &= \left(\kappa_i^{-1}\mathbb{1}\{i=j\} - 1\right)\left(y_\bullet(\max\{t,s\}) - y_\bullet(t)y_\bullet(s)\right),\\
\mathrm{E}\left(\overline{G}_i(t)G_{jm_1}(s)\right) &= \left(\kappa_i^{-1}\mathbb{1}\{i=j\} - 1\right)\left((\nu_{\bullet m_1}(s) - \nu_{\bullet m_1 -}(t))\mathbb{1}\{s \geqslant t\} - y_\bullet(t)\nu_{\bullet m_1}(s)\right),\\
\mathrm{E}\left(G_{im_1}(t)G_{jm_2}(s)\right) &= \left(\kappa_i^{-1}\mathbb{1}\{i=j\} - 1\right)\left(\nu_{\bullet m_1}(\min\{t,s\})\mathbb{1}\{m_1 = m_2\} - \nu_{\bullet m_1}(t)\nu_{\bullet m_2}(s)\right)
\end{aligned}$$

at $(t,s) \in [0,\tau]^2$ for all $i,j \in \{1,\dots,k\}, m_1, m_2 \in \{1,\dots,M\}$. By the conditional delta-method (Theorem 2.5) and the uniform Hadamard-differentiability of the Wilcoxon functional (Example A.1), we get

$$\begin{aligned}
n^{1/2}\left(\widehat{A}_{im}^\pi - \widehat{A}_{\bullet m}\right)_{i\in\{1,\dots,k\}, m\in\{1,\dots,M\}} &\xrightarrow{d^*} \left(\int_{[0,.]} y_\bullet^{-1}\,\mathrm{d}G_{im} - \int_{[0,.]} \overline{G}_i y_\bullet^{-2}\,\mathrm{d}\nu_{\bullet m}\right)_{i\in\{1,\dots,k\}, m\in\{1,\dots,M\}}\\
&=: (U_{im}^\pi)_{i\in\{1,\dots,k\}, m\in\{1,\dots,M\}}
\end{aligned}$$

on $(D[0,\tau])^{kM}$ as $n \to \infty$. The limit variable $(U_{im}^\pi)_{i\in\{1,\dots,k\}, m\in\{1,\dots,M\}}$ is a separable centered Gaussian process and the covariance structure can be calculated similarly as in [26] as

$$\begin{aligned}
\mathrm{E}\left(U_{im_1}^\pi(t)U_{jm_2}^\pi(s)\right) &= \left(\kappa_i^{-1}\mathbb{1}\{i=j\} - 1\right)\int_{[0,\min\{t,s\}]} (\mathbb{1}\{m_1 = m_2\} - \Delta A_{\bullet m_1})y_\bullet^{-1}\,\mathrm{d}A_{\bullet m_2}\\
&=: \sigma_{ijm_1m_2}^\pi(\min\{t,s\})
\end{aligned}$$

for all $i,j \in \{1,\dots,k\}, m_1,m_2 \in \{1,\dots,M\}$ and $t,s \in [0,\tau]$. Furthermore, we aim to apply the conditional delta-method with the function

$$\mathbf{\Phi} : \mathbb{D}_\Phi^k \ni (\Lambda_{im})_{i\in\{1,\dots,k\}, m\in\{1,\dots,M\}} \mapsto (\Phi(\Lambda_{11}, \dots, \Lambda_{1M}), \dots, \Phi(\Lambda_{k1}, \dots, \Lambda_{kM})) \in \mathbb{R}^{kM},$$

where Φ is defined as in the proof of Theorem 5.2 and

$$\mathbb{D}_\Phi := \left\{(\Lambda_m)_{m\in\{1,\dots,M\}} \in (BV_K[0,\tau])^M \mid -\sum_{m=1}^M \Lambda_m \in BV_{MK}^{>-1}[0,\tau], \widetilde{\phi}\left(-\sum_{m=1}^M \Lambda_m\right) \in BV_K[0,\tau)\right\}.$$

Consequently, it remains to show the uniform Hadamard-differentiability of $\Phi : \mathbb{D}_\Phi \to \mathbb{R}^M$ at $(A_{\bullet 1}, \dots, A_{\bullet M})$. Therefore, we remind that Φ is a composition of linear functionals and $\widetilde{\psi}, \widetilde{\phi}$ in Section A. Hence, the chain rule (Theorem 2.4) together with the examples in Section A implies that it remains to show that $-\sum_{m=1}^M A_{\bullet m} : [0,\tau) \to \mathbb{R}$ is a càdlàg function of bounded variation with jumps contained in $(-1,\infty)$ and bounded away from -1, which follows by Assumption 5.1. Thus, we receive

$$n^{1/2}\left(\widehat{\boldsymbol{\eta}}^\pi - \widehat{\boldsymbol{\eta}}_\bullet\right) \xrightarrow{d^*} \left(\Phi'_{A_{\bullet 1},\dots,A_{\bullet M}}(U_{11}^\pi, \dots, U_{1M}^\pi), \dots, \Phi'_{A_{\bullet 1},\dots,A_{\bullet M}}(U_{k1}^\pi, \dots, U_{kM}^\pi)\right) =: \mathbf{Z}^\pi$$

as $n \to \infty$. The limit variable $\mathbf{Z}^\pi$ is a centered normal variable with covariance matrix $\mathbf{\Sigma}^\pi$, where the entries $\mathbf{\Sigma}_{im_1,jm_2}^\pi$ are given by

$$\begin{aligned}
&\int_{[0,\tau)} \frac{\left\{(\tau-u)\left(1 - \sum\limits_{\widetilde{m}\neq m_1} F_{\bullet\widetilde{m}}(u)\right) - \int_u^\tau F_{\bullet m_1}(t)\,\mathrm{d}t\right\}\left\{(\tau-u)\left(1 - \sum\limits_{\widetilde{m}\neq m_2} F_{\bullet\widetilde{m}}(u)\right) - \int_u^\tau F_{\bullet m_2}(t)\,\mathrm{d}t\right\}}{(1-\Delta A_\bullet(u))^2}\,\mathrm{d}\sigma_{ijm_1m_2}^\pi(u)\\
&+ \int_{[0,\tau)} \frac{\left\{(\tau-u)\left(1 - \sum\limits_{\widetilde{m}\neq m_1} F_{\bullet\widetilde{m}}(u)\right) - \int_u^\tau F_{\bullet m_1}(t)\,\mathrm{d}t\right\}\left\{(\tau-u)F_{\bullet m_2}(u) - \int_u^\tau F_{\bullet m_2}(t)\,\mathrm{d}t\right\}}{(1-\Delta A_\bullet(u))^2}\,\mathrm{d}\sum_{\widetilde{m}\neq m_2}\sigma_{ijm_1\widetilde{m}}^\pi(u)\\
&+ \int_{[0,\tau)} \frac{\left\{(\tau-u)\left(1 - \sum\limits_{\widetilde{m}\neq m_2} F_{\bullet\widetilde{m}}(u)\right) - \int_u^\tau F_{\bullet m_2}(t)\,\mathrm{d}t\right\}\left\{(\tau-u)F_{\bullet m_1}(u) - \int_u^\tau F_{\bullet m_1}(t)\,\mathrm{d}t\right\}}{(1-\Delta A_\bullet(u))^2}\,\mathrm{d}\sum_{\widetilde{m}\neq m_1}\sigma_{ijm_2\widetilde{m}}^\pi(u)\\
&+ \int_{[0,\tau)} \frac{\left\{(\tau-u)F_{\bullet m_1}(u) - \int_u^\tau F_{\bullet m_1}(t)\,\mathrm{d}t\right\}\left\{(\tau-u)F_{\bullet m_2}(u) - \int_u^\tau F_{\bullet m_2}(t)\,\mathrm{d}t\right\}}{(1-\Delta A_\bullet(u))^2}\,\mathrm{d}\sum_{\widetilde{m}\neq m_1}\sum_{\breve{m}\neq m_2}\sigma_{ij\widetilde{m}\breve{m}}^\pi(u)
\end{aligned}$$

for all $i,j \in \{1,\dots,k\}, m_1,m_2 \in \{1,\dots,M\}$. □

Now, we turn to the permutation counterpart of the covariance matrix estimator.

Lemma 5.4. *Under Assumption 5.1, we have*

$$\widehat{\mathbf{\Sigma}}^{\pi} \xrightarrow{P} \widetilde{\mathbf{\Sigma}}^{\pi} := \bigoplus_{i=1}^{k} \kappa_i^{-1} (\widetilde{\Sigma}^{\pi}_{m_1 m_2})_{m_1, m_2 \in \{1,\ldots,M\}}$$

as $n \to \infty$, where $\widetilde{\Sigma}^{\pi}_{m_1 m_2}$ is defined as in (5.10) *for all $m_1, m_2 \in \{1, \ldots, M\}$.*

Proof of Lemma 5.4. Due to the definition of $\widehat{\mathbf{\Sigma}}^{\pi}$, it remains to show $\widehat{\Sigma}^{\pi}_{im_1,im_2} \xrightarrow{P} \widetilde{\Sigma}^{\pi}_{m_1 m_2}$, $i \in \{1, \ldots, k\}, m_1, m_2 \in \{1, \ldots, M\}$. Therefore, let $i \in \{1, \ldots, k\}$ be arbitrary but fixed. Then, (5.9) implies

$$n^{1/2} \left(n_i^{-1} (Y_i^{\pi}, N_{i1}^{\pi}, \ldots, N_{iM}^{\pi}) - n^{-1} (Y_\bullet, N_{\bullet 1}, \ldots, N_{\bullet M}) \right) \xrightarrow{d} (\overline{G}_i, G_{i1}, \ldots, G_{iM})$$

on $(D[0,\tau])^{M+1}$ as $n \to \infty$ unconditionally by Lemma 2.1. Hence, Slutsky's lemma provides

$$\sup_{t \in [0,\tau]} \left| n_i^{-1} Y_i^{\pi}(t) - n^{-1} Y_\bullet(t) \right| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{t \in [0,\tau]} \left| n_i^{-1} N_{im}^{\pi}(t) - n^{-1} N_{\bullet m}(t) \right| \xrightarrow{P} 0, m \in \{1, \ldots, M\},$$

as $n \to \infty$. By the definitions of $Y_\bullet, N_{\bullet 1}, \ldots, N_{\bullet M}$, we further get

$$\begin{aligned} \sup_{t \in [0,\tau]} \left| n^{-1} Y_\bullet(t) - y_\bullet(t) \right| &= \sup_{t \in [0,\tau]} \left| \sum_{i=1}^{k} \left(\frac{n_i}{n} n_i^{-1} Y_i(t) - \kappa_i y_i(t) \right) \right| \xrightarrow{P} 0 \quad \text{and} \\ \sup_{t \in [0,\tau]} \left| n^{-1} N_{\bullet m}(t) - F_{\bullet m}(t) \right| &= \sup_{t \in [0,\tau]} \left| \sum_{i=1}^{k} \left(\frac{n_i}{n} n_i^{-1} N_{im}(t) - \kappa_i F_{im}(t) \right) \right| \xrightarrow{P} 0, m \in \{1, \ldots, M\}, \end{aligned}$$

as $n \to \infty$ since $\mathcal{F}$ in the proof of Lemma 5.3 is $P^{(X_{i1}, D_{i1})}$-Donsker. Thus, it follows

$$\sup_{t \in [0,\tau]} \left| n_i^{-1} Y_i^{\pi}(t) - y_\bullet(t) \right| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{t \in [0,\tau]} \left| n_i^{-1} N_{im}^{\pi}(t) - F_{\bullet m}(t) \right| \xrightarrow{P} 0, m \in \{1, \ldots, M\},$$

as $n \to \infty$. Since $\widehat{A}^{\pi}_{im}(\cdot) = \int_{[0,\cdot]} n_i (Y_i^{\pi})^{-1} \, \mathrm{d}(n_i^{-1} N_{im}^{\pi})$, $\widehat{F}^{\pi}_{im}$ and $\widehat{\sigma}^{\pi}_{im\widetilde{m}}$ are depending continuously on

$$n_i^{-1} (Y_i^{\pi}, N_{i1}^{\pi}, \ldots, N_{iM}^{\pi}),$$

we can conclude

$$\sup_{t \in [0,\tau]} \left| \widehat{A}^{\pi}_{im}(t) - A_{\bullet m}(t) \right| \xrightarrow{P} 0, \sup_{t \in [0,\tau]} \left| \widehat{F}^{\pi}_{im}(t) - F_{\bullet m}(t) \right| \xrightarrow{P} 0 \text{ and } \sup_{t \in [0,\tau]} \left| \widehat{\sigma}^{\pi}_{im\widetilde{m}}(t) - \sigma_{\bullet m \widetilde{m}}(t) \right| \xrightarrow{P} 0$$

as $n \to \infty$ for all $m, \widetilde{m} \in \{1, \ldots, M\}$. Hence, we get

$$\begin{aligned} &\widehat{\Sigma}^{\pi}_{im_1,im_2} \xrightarrow{P} \widetilde{\Sigma}^{\pi}_{m_1 m_2} := \\ &\int_{[0,\tau)} \frac{\left\{ (\tau - u) \left(1 - \sum\limits_{\widetilde{m} \neq m_1} F_{\bullet \widetilde{m}}(u) \right) - \int_u^\tau F_{\bullet m_1}(t) \, \mathrm{d}t \right\} \left\{ (\tau - u) \left(1 - \sum\limits_{\widetilde{m} \neq m_2} F_{\bullet \widetilde{m}}(u) \right) - \int_u^\tau F_{\bullet m_2}(t) \, \mathrm{d}t \right\}}{(1 - \Delta A_\bullet(u))^2} \, \mathrm{d}\sigma_{\bullet m_1 m_2}(u) \\ &+ \int_{[0,\tau)} \frac{\left\{ (\tau - u) \left(1 - \sum\limits_{\widetilde{m} \neq m_1} F_{\bullet \widetilde{m}}(u) \right) - \int_u^\tau F_{\bullet m_1}(t) \, \mathrm{d}t \right\} \left\{ (\tau - u) F_{\bullet m_2}(u) - \int_u^\tau F_{\bullet m_2}(t) \, \mathrm{d}t \right\}}{(1 - \Delta A_\bullet(u))^2} \, \mathrm{d} \sum_{\widetilde{m} \neq m_2} \sigma_{\bullet m_1 \widetilde{m}}(u) \\ &+ \int_{[0,\tau)} \frac{\left\{ (\tau - u) \left(1 - \sum\limits_{\widetilde{m} \neq m_2} F_{\bullet \widetilde{m}}(u) \right) - \int_u^\tau F_{\bullet m_2}(t) \, \mathrm{d}t \right\} \left\{ (\tau - u) F_{\bullet m_1}(u) - \int_u^\tau F_{\bullet m_1}(t) \, \mathrm{d}t \right\}}{(1 - \Delta A_\bullet(u))^2} \, \mathrm{d} \sum_{\widetilde{m} \neq m_1} \sigma_{\bullet m_2 \widetilde{m}}(u) \\ &+ \int_{[0,\tau)} \frac{\left\{ (\tau - u) F_{\bullet m_1}(u) - \int_u^\tau F_{\bullet m_1}(t) \, \mathrm{d}t \right\} \left\{ (\tau - u) F_{\bullet m_2}(u) - \int_u^\tau F_{\bullet m_2}(t) \, \mathrm{d}t \right\}}{(1 - \Delta A_\bullet(u))^2} \, \mathrm{d} \sum_{\widetilde{m} \neq m_1} \sum_{\check{m} \neq m_2} \sigma_{\bullet \widetilde{m} \check{m}}(u) \end{aligned} \tag{5.10}$$

as $n \to \infty$. ☐

As in the proof of Theorem 5.2, we need the positive definiteness of $\widetilde{\mathbf{\Sigma}}^{\pi}$, which is given by the following lemma.

Lemma 5.5. *Under Assumptions 5.1 and 5.2, $\widetilde{\mathbf{\Sigma}}^{\pi}$ is positive definite.*

Proof of Lemma 5.5. The proof is similar to the proof of Lemma 5.2. Firstly note that $\widetilde{\mathbf{\Sigma}}^\pi$ is positive definite if $(\widetilde{\Sigma}^\pi_{m_1m_2})_{m_1,m_2\in\{1,\ldots,M\}}$ is positive definite. Let $\mathbf{a} = (a_1,\ldots,a_M)' \in \mathbb{R}^M\backslash\{\mathbf{0}_M\}$ be arbitrary. With

$$\begin{aligned} f_m(u) &:= \frac{(\tau-u)(1-\sum_{\widetilde{m}\neq m} F_{\bullet\widetilde{m}}(u)) - \int_u^\tau F_{\bullet m}(t)\,\mathrm{d}t}{1-\Delta A_\bullet(u)}, \\ g_m(u) &:= \frac{(\tau-u)F_{\bullet m}(u) - \int_u^\tau F_{\bullet m}(t)\,\mathrm{d}t}{1-\Delta A_\bullet(u)} \\ \text{and} \quad h_m(u) &:= a_m f_m(u) + \sum_{\widetilde{m}\neq m} a_{\widetilde{m}} g_{\widetilde{m}}(u) \end{aligned}$$

for all $u \in [0,\tau)$, we get

$$\mathbf{a}'(\widetilde{\Sigma}^\pi_{m_1m_2})_{m_1,m_2\in\{1,\ldots,M\}}\mathbf{a} = \sum_{m=1}^M \int_{[0,\tau)} \frac{h_m^2}{y_\bullet}\,\mathrm{d}A^c_{\bullet m} \tag{5.11}$$

$$+ \sum_{x\in\mathcal{D}} \frac{\sum_{m=1}^M h_m^2(x)\Delta A_{\bullet m}(x) - \left(\sum_{m=1}^M h_m(x)\Delta A_{\bullet m}(x)\right)^2}{y_\bullet(x)} \tag{5.12}$$

analogously to the proof of Lemma 5.2, where $\mathcal{D} := \{x \in [0,\tau) : \Delta A_\bullet(x) > 0\}$ and

$$A^c_{\bullet m}(x) := A_{\bullet m}(x) - \sum_{y\leqslant x, y\in\mathcal{D}} \Delta A_{\bullet m}(y), m \in \{1,\ldots,M\},$$

for all $x \in [0,\tau)$. The Cauchy-Schwarz inequality implies

$$\sum_{m=1}^M h_m^2(x)\Delta A_{\bullet m}(x) - \left(\sum_{m=1}^M h_m(x)\Delta A_{\bullet m}(x)\right)^2 \geqslant \sum_{m=1}^M h_m^2(x)\Delta A_{\bullet m}(x)\,(1-\Delta A_\bullet(x)) \geqslant 0$$

for all $x \in \mathcal{D}$. For $m^* \in \{1,...,M\}$ with $|a_{m^*}| = \max\{|a_1|,\ldots,|a_M|\} > 0$, it holds

$$\begin{aligned} |h_{m^*}(u)| &= \left|a_{m^*}f_{m^*}(u) + \sum_{\widetilde{m}\neq m^*} a_{\widetilde{m}}g_{\widetilde{m}}(u)\right| \geqslant |a_{m^*}|f_{m^*}(u) - \left|\sum_{\widetilde{m}\neq m^*} a_{\widetilde{m}}g_{\widetilde{m}}(u)\right| \\ &\geqslant |a_{m^*}|\left(f_{im^*}(u) + \sum_{\widetilde{m}\neq m^*} g_{i\widetilde{m}}(u)\right) = |a_{m^*}|\frac{\sum_{i=1}^k \kappa_i \int_u^\tau S_i(t)\,\mathrm{d}t}{1-\Delta A_\bullet(u)} \\ &\geqslant |a_{m^*}|\frac{(\tau-u)\sum_{i=1}^k \kappa_i S_i(\tau)}{1-\Delta A_\bullet(u)} > 0 \end{aligned}$$

for all $u \in [0,\tau)$ since $|f_{m^*}(u)| = f_{m^*}(u)$, $|g_{\widetilde{m}}(u)| = -g_{\widetilde{m}}(u)$ and

$$\Delta A_\bullet(u) = \frac{\sum_{i=1}^k \kappa_i y_i(u)\Delta A_i(u)}{\sum_{i=1}^k \kappa_i y_i(u)} < 1$$

for all $u \in [0,\tau); \widetilde{m} \in \{1,\ldots,M\}$. Note that $A_{\bullet m^*-}(\tau) > 0$ due to Assumption 5.2. Thus, at least one of the summands (5.11) and (5.12) is strictly positive. □

Note that $\sigma^\pi_{ijm_1m_2} = (\kappa_i^{-1}\mathbb{1}\{i=j\}-1)\sigma_{\bullet m_1m_2}$ for all $i,j \in \{1,\ldots,k\}, m_1,m_2 \in \{1,\ldots,M\}$. Hence, it holds $\widetilde{\mathbf{\Sigma}}^\pi = \mathbf{\Sigma}^\pi + (\mathbf{1}_k\mathbf{1}_k') \otimes (\widetilde{\Sigma}^\pi_{m_1m_2})_{m_1,m_2\in\{1,\ldots,M\}}$ and, thus, $\mathbf{H}\widehat{\mathbf{\Sigma}}^\pi\mathbf{H}' \xrightarrow{P} \mathbf{H}\widetilde{\mathbf{\Sigma}}^\pi\mathbf{H}' = \mathbf{H}\mathbf{\Sigma}^\pi\mathbf{H}'$ as $n\to\infty$ due to the contrast property of the hypothesis matrix $\mathbf{H}$. The positive definiteness of $\widetilde{\mathbf{\Sigma}}^\pi$ ensures

$$\operatorname{rank}\left(\mathbf{H}\mathbf{\Sigma}^\pi\mathbf{H}'\right) = \operatorname{rank}\left(\mathbf{H}\widetilde{\mathbf{\Sigma}}^\pi\mathbf{H}'\right) = \operatorname{rank}(\mathbf{H}).$$

As in the proof of Theorem 5.2, we get $(\mathbf{H}\widehat{\mathbf{\Sigma}}^\pi\mathbf{H}')^+ \xrightarrow{P} (\mathbf{H}\mathbf{\Sigma}^\pi\mathbf{H}')^+$ as $n\to\infty$. Combining this with Lemma 5.3, Slutsky's lemma and Theorem 9.2.2 in [67] yields

$$\begin{aligned} W_n^\pi(\mathbf{H}) &= n(\mathbf{H}\widehat{\boldsymbol{\eta}}^\pi)'\left(\mathbf{H}\widehat{\mathbf{\Sigma}}^\pi\mathbf{H}'\right)^+(\mathbf{H}\widehat{\boldsymbol{\eta}}^\pi) \\ &= \left(\mathbf{H}(n^{1/2}(\widehat{\boldsymbol{\eta}}^\pi - \widehat{\boldsymbol{\eta}}_\bullet))\right)'\left(\mathbf{H}\widehat{\mathbf{\Sigma}}^\pi\mathbf{H}'\right)^+\left(\mathbf{H}(n^{1/2}(\widehat{\boldsymbol{\eta}}^\pi - \widehat{\boldsymbol{\eta}}_\bullet))\right) \\ &\xrightarrow{d^*} (\mathbf{H}\mathbf{Z}^\pi)'(\mathbf{H}\mathbf{\Sigma}^\pi\mathbf{H}')^+(\mathbf{H}\mathbf{Z}^\pi) \sim \chi^2_{\operatorname{rank}(\mathbf{H})} \end{aligned}$$

as $n\to\infty$. □

Proof of Theorem 5.4 We aim to apply Theorem 2.6 and Lemma 2.3. Therefore, let $\mathbf{X}_n := (\mathbf{X}, \boldsymbol{\delta})$ denote the data, $\mathbf{M}_n^{(b)}$ denote the randomness of the permutation method and $\mathbf{W}_n^{(b)}$ denote the bth Monte Carlo replicate of the permutation counterparts of the Wald-type test statistic for all $b \in \{1, ..., B_n\}$. Moreover, let $L = 1$ and $F_{n,1} = F_n$ be as in Lemma 2.3, i.e., denoting the empirical distribution function of $\mathbf{W}_n^{(1)}, ..., \mathbf{W}_n^{(B_n)}$. Then, $\mathbf{W}_n^{(1)} \xrightarrow{d*} \chi^2_{\mathrm{rank}(\mathbf{H})}$ holds as $n \to \infty$ by Theorem 5.3. The limit distribution $\chi^2_{\mathrm{rank}(\mathbf{H})}$ has a continuous distribution function $\mathcal{F}_1 : \mathbb{R} \to [0, 1]$ that is strictly increasing on $[0, \infty)$ due to $\mathrm{rank}(\mathbf{H}) > 0$ and the function FWER in Theorem 2.6 equals the identity on $[0, 1]$. Hence, Lemma 2.3 implies (2.15) and (2.16). Thus, Theorem 2.6 yields the statement of the theorem. $\square$

Proof of Theorem 5.5 By Slutsky's lemma, we combine Theorem 5.1 and Lemma 5.1 for Theorem 5.5. Since the map

$$\mathbb{R}^{kM} \times \mathbb{R}^{kM \times kM} \ni (\mathbf{m}, \mathbf{S}) \mapsto \left((\mathbf{H}_\ell \mathbf{m})'(\mathbf{H}_\ell \mathbf{S} \mathbf{H}_\ell')^+ \mathbf{H}_\ell \mathbf{m}\right)_{\ell \in \mathcal{T}} \in \mathbb{R}^{\mathcal{T}}$$

is continuous on $\mathbb{R}^{kM} \times \{\boldsymbol{\Sigma}\}$ due to Lemma 5.2, the statement follows by the continuous mapping theorem. $\square$

Proof of Theorem 5.6 In order to prove Theorem 5.6, we aim to apply Theorem 2.6 and Lemma 2.3. Therefore, let $\mathbf{X}_n := (\mathbf{X}, \boldsymbol{\delta})$ denote the data, $\mathbf{M}_n^{(b)} := \mathbf{Y}^{(b)}$ denote the randomness of the Monte Carlo method and $\mathbf{W}_n^{(b)} := (\mathbf{H}_\ell \widehat{\boldsymbol{\Sigma}}^{1/2} \mathbf{Y}^{(b)})'(\mathbf{H}_\ell \widehat{\boldsymbol{\Sigma}} \mathbf{H}_\ell')^+(\mathbf{H}_\ell \widehat{\boldsymbol{\Sigma}}^{1/2} \mathbf{Y}^{(b)})$ denote the bth Monte Carlo replicate for all $b \in \{1, ..., B_n\}$. Moreover, let F_n be as in Lemma 2.3, i.e., denoting the empirical distribution function of $\mathbf{W}_n^{(1)}, ..., \mathbf{W}_n^{(B_n)}$ and $F_{n,\ell}$ denote the cumulative distribution functions of the $\chi^2_{\mathrm{rank}(\mathbf{H}_\ell \widehat{\boldsymbol{\Sigma}} \mathbf{H}_\ell')}$-distribution for all $\ell \in \{1, ..., L\}$. Then, $\mathbf{W}_n^{(1)} \xrightarrow{d*} ((\mathbf{H}_\ell \mathbf{Z})'(\mathbf{H}_\ell \boldsymbol{\Sigma} \mathbf{H}_\ell')^+(\mathbf{H}_\ell \mathbf{Z}))_{\ell \in \{1,...,L\}}$ holds as $n \to \infty$ by Lemma 5.1. The marginal limit distributions $\chi^2_{\mathrm{rank}(\mathbf{H}_\ell)}$ have continuous distribution functions $\mathcal{F}_\ell : \mathbb{R} \to [0, 1]$ that are strictly increasing on $[0, \infty)$ due to $\mathrm{rank}(\mathbf{H}_\ell) > 0$ for all $\ell \in \{1, ..., L\}$. Hence, Lemma 2.3 implies (2.15). Moreover, (2.16) follows from the consistency of the covariance matrix estimator since

$$P\left(\mathrm{rank}\left(\mathbf{H}_\ell \widehat{\boldsymbol{\Sigma}} \mathbf{H}_\ell'\right) \neq \mathrm{rank}\left(\mathbf{H}_\ell\right)\right) \to 0$$

holds as $n \to \infty$. As a result, Theorem 2.6 yields the statement of the theorem. $\square$

6 Discussion and Outlook

In this thesis, we constructed tests for survival estimands in complex survival designs. In order to do so, we started with some methodological preliminaries in Section 2. There, we closed the gap of a suitable delta-method for resampling procedures in multiple sample problems as, e.g., the permutation and pooled bootstrap. In addition, we introduced a strategy to obtain multiple tests that are asymptotically balanced and control the family-wise error rate in the strong sense. In Section 3, we considered paired survival times and constructed suitable tests for a version of the Mann-Whitney effect and the restricted mean survival time (RMST). Multiple tests for RMSTs in general factorial survival designs are developed in Section 4. As a natural extension of the RMST in competing risks setups, we constructed multiple tests based on the restricted mean time lost (RMTL) in general factorial designs in Section 5.
All theoretical results were proven in detail. Furthermore, extensive simulation studies were conducted to analyze the finite sample behavior of the proposed methods of Section 4 and 5. Additionally, we successfully applied the methodology to different data examples to illustrate their usage.

Outlook When considering the methodology of Section 2.5, the weak convergence of uniform Hadamard differentiable functionals of resampling empirical processes can be derived in outer probability which is sufficient for most statistical applications. In view of the extensions of the classical delta-method for, e.g., quasi-Hadamard differentiable functionals [5, 6, 7] and directionally differentiable functionals [35], future research might include whether conditional delta-methods for more general functionals can be achieved that are applicable for resampling procedures in multiple sample problems. Moreover, the weak convergence could also be investigated in the outer almost sure case; cf. the supplement of [7] for an extension of the conditional delta-method outer almost surely under measurability assumptions.
Section 5 involved the construction of tests which cover RMTL comparisons for the same event types across different groups. However, one may also be interested in comparing two or more RMTLs within each group. A potentially suitable resampling procedure for this problem could be motivated from the randomization approach in [30]. As an adaption of this, the event indicators δ_{ij} are re-drawn as $\widetilde{\delta}_{ij}$ from $\{1, \ldots, M\}$ with equal probability $1/M$ if $\delta_{ij} \neq 0$, which leads to the randomized data $(X_{ij}, \widetilde{\delta}_{ij}), j \in \{1, \ldots, n_i\}, i \in \{1, \ldots, k\}$. By the theory of [27], the asymptotic validity of this randomization approach can be shown if the hypothesis matrix can be partitioned into a block matrices with one row block and k column blocks for k contrast matrices. Moreover, finitely exact tests could be achieved by this randomization approach under the event type exchangeability of the data, i.e., $F_{i1} \equiv \ldots \equiv F_{iM}, i \in \{1, ..., k\}$. Nonetheless, more analysis on this matter is a point of future research. Beyond the permutation approach described in the paper and the randomization approach of [30], further resampling approaches could be considered. For instance, an alternative permutation approach for factorial designs might be to only permute within the factor whose effect should be tested.
As a further outlook, tests regarding other effect estimands as, e.g., the median survival time [12, 15, 22], in complex survival designs could be developed. An estimand similar to the usual RMST studied in Sections 3.2 and 4 is the weighted version of the RMST. That is, $\mu_w := \int_0^\tau w(t)S(t)\,\mathrm{d}t$ with estimator $\widehat{\mu}_w := \int_0^\tau w(t)\widehat{S}(t)\,\mathrm{d}t$ for some weight function $w \in \mathcal{L}_1([0, \tau])$ and survival function S similar as in [80]. Additionally, the case of data dependent weight functions was already investigated in [80] for the two-sample case. For competing risks setups as in Section 5, for example cumulative incidence quantiles [8, 50, 64, 69], extensions of the probabilistic index (or relative treatment effect) [31, 32] in the presence of competing risks, and the area between curves statistic [53, 54] are interesting alternatives.
In addition, our survival models cover paired survival data and factorial designs so far, i.e., only factorial covariates can be incorporated. In future research, models for more complex covariates as, e.g., continuous, high-dimensional, and time-varying covariates, could be considered. While competing risks setups are an extension of classical survival models, they can be further extended as multi-state models [44]. Multi-state models allow for intermediate states or transitions back to the initial state once an event has occurred and, hence, are even more flexible than competing risks models. Here, (multiple) tests for estimands in multi-state models could be investigated as well. In this context, mean sojourn times could be considered as a natural extension of the RMST and RMTL.

References

[1] O. O. Aalen and S. Johansen. An empirical transition matrix for non-homogeneous markov chains based on censored observations. *Scandinavian Journal of Statistics*, 5(3):141–150, 1978.

[2] P. K. Andersen. Decomposition of number of life years lost according to causes of death. *Statistics in Medicine*, 32(30):5278–5285, 2013.

[3] P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes.* Springer, New York, 1993.

[4] R. Beran. Balanced simultaneous confidence sets. *Journal of the American Statistical Association*, 83(403):679–686, 1988.

[5] E. Beutner and H. Zähle. A modified functional delta method and its application to the estimation of risk functionals. *Journal of Multivariate Analysis*, 101(10):2452–2463, 2010.

[6] E. Beutner and H. Zähle. Functional delta-method for the bootstrap of quasi-Hadamard differentiable functionals. *Electronic Journal of Statistics*, 10(1):1181 – 1222, 2016.

[7] E. Beutner and H. Zähle. Bootstrapping average value at risk of single and collective risks. *Risks*, 6(3):96, 2018.

[8] J. Beyersmann and M. Schumacher. A note on nonparametric quantile inference for competing risks and more complex multistate models. *Biometrika*, 95(4):1006–1008, 2008.

[9] P. J. Bickel and P. W. Millar. Uniform convergence of probability measures on classes of functions. *Statistica Sinica*, 2(1):1–15, 1992.

[10] P. Blanche, J.-F. Dartigues, and J. Riou. A closed max-t test for multiple comparisons of areas under the roc curve. *Biometrics*, 78(1):352–363, 2022.

[11] F. Bretz, A. Genz, and L. A. Hothorn. On the numerical availability of multiple comparison procedures. *Biometrical Journal*, 43(5):645–656, 2001.

[12] R. Brookmeyer and J. J. Crowley. A confidence interval for the median survival time. *Biometrics*, 38(1):29–41, 1982.

[13] M. Brueckner and W. Brannath. Sequential tests for non-proportional hazards data. *Lifetime Data Analysis*, 23(3):339–352, 07 2017.

[14] A. Bücher and I. Kojadinovic. A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *Journal of Theoretical Probability*, 32(3):1145–1165, 2019.

[15] Z. Chen and G. Zhang. Comparing survival curves based on medians. *BMC Medical Research Methodology*, 16, 03 2016.

[16] E. Chung and J. P. Romano. Multivariate and multiple permutation tests. *Journal of Econometrics*, 193(1):76–91, 2016.

[17] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.

[18] L. C. de Wreede, M. Fiocco, and H. Putter. The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Computer Methods and Programs in Biomedicine*, 99:261–274, 2010.

[19] L. C. de Wreede, M. Fiocco, and H. Putter. mstate: An R package for the analysis of competing risks and multi-state models. *Journal of Statistical Software*, 38(7):1–30, 2011.

[20] M. Ditzhaus, D. Dobler, M. Kindop, and M. Munko. *GFDrmst: Multiple RMST-Based Tests in General Factorial Designs*, 2024. R package version 0.1.1.

[21] M. Ditzhaus, D. Dobler, M. Munko, and J. Walter. *GFDrmtl: Multiple RMTL-Based Tests for Competing Risks Data in General Factorial Designs*, 2024. R package version 0.1.0.

[22] M. Ditzhaus, D. Dobler, and M. Pauly. Inferring median survival differences in general factorial designs via permutation tests. *Statistical Methods in Medical Research*, 30(3):875–891, 2021. PMID: 33349152.

[23] M. Ditzhaus, R. Fried, and M. Pauly. QANOVA: quantile-based permutation methods for general factorial designs. *Test*, pages 1–20, 2021.

[24] M. Ditzhaus, M. Yu, and J. Xu. Studentized permutation method for comparing two restricted mean survival times with small sample from randomized trials. *Statistics in Medicine*, 42(13):2226–2240, 2023.

[25] G. D. Djira and L. A. Hothorn. Detecting relative changes in multiple comparisons with an overall mean. *Journal of Quality Technology*, 41(1):60–65, 2009.

[26] D. Dobler. A discontinuity adjustment for subdistribution function confidence bands applied to right-censored competing risks data. *Electronic Journal of Statistics*, 11(2):3673 – 3702, 2017.

[27] D. Dobler. Randomized empirical processes by algebraic groups, and tests for weak null hypotheses. *Bernoulli*, 29(2):1109 – 1136, 2023.

[28] D. Dobler and M. Munko. Erratum to "a discontinuity adjustment for subdistribution function confidence bands applied to right-censored competing risks data". *Electronic Journal of Statistics*, 18(2):4578–4581, 2024.

[29] D. Dobler and E. Musta. A two-sample comparison of mean survival times of uncured subpopulations. *Electronic Journal of Statistics*, 18(2):3107 – 3169, 2024.

[30] D. Dobler and K. Möllenhoff. A nonparametric relative treatment effect for direct comparisons of censored paired survival outcomes. *Statistics in Medicine*, 43(11):2216–2238, 2024.

[31] D. Dobler and M. Pauly. Bootstrap- and permutation-based inference for the Mann–Whitney effect for right-censored and tied data. *TEST*, 27(3):639–658, 2018.

[32] D. Dobler and M. Pauly. Factorial analyses of treatment effects under independent right-censoring. *Statistical Methods in Medical Research*, 29(2):325–343, 2020. PMID: 30834811.

[33] O. J. Dunn. Multiple Comparisons among Means. *Journal of the American Statistical Association*, 56(293):52–64, Mar. 1961.

[34] C. W. Dunnett. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50(272):1096–1121, 1955.

[35] Z. Fang and A. Santos. Inference on directionally differentiable functions. *The Review of Economic Studies*, 86(1):377–412, 2019.

[36] M. Fiocco, H. Putter, and J. Van Houwelingen. Reduced rank proportional hazards model for competing risks. *Biostatistics*, 6(3):465–478, 2005.

[37] J. Genuneit. Sex-specific development of asthma differs between farm and nonfarm children: a cohort study. *American Journal of Respiratory and Critical Care Medicine*, 190(5):588–590, 2014.

[38] J. Genuneit, G. Büchele, M. Waser, K. Kovacs, A. Debinska, A. Boznanski, C. Strunz-Lehner, E. Horak, P. Cullinan, D. Heederik, C. Braun-Fahrländer, and E. G. S. G. von Mutius. The gabriel advanced surveys: study design, participation and evaluation of bias. *Paediatric and Perinatal Epidemiology*, 25(5):436–447, 2011.

[39] R. D. Gill and S. Johansen. A survey of product-integration with a view toward application in survival analysis. *The Annals of Statistics*, 18(4):1501–1555, 1990.

[40] M. Greenwood. *A Report on the Natural Duration of Cancer.* Reports on public health and medical subjects. H.M. Stationery Office, 1926.

[41] J. Hemerik and J. Goeman. Exact testing with random permutations. *Test*, 27(4):811–825, 2018.

[42] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

[43] M. Horiguchi and H. Uno. On permutation tests for comparing restricted mean survival time with small sample from randomized trials. *Statistics in Medicine*, 39(20):2655–2670, 2020.

[44] P. Hougaard. Multi-state models: a review. *Lifetime data analysis*, 5:239–264, 1999.

[45] A. Janssen and T. Pauls. How do bootstrap and permutation tests work? *Annals of Statistics*, 31(3):768–806, 2003.

[46] J. D. Kalbfleisch and R. L. Prentice. Estimation of the average hazard ratio. *Biometrika*, 68(1):105–112, 04 1981.

[47] F. Konietschke, S. Bösiger, E. Brunner, and L. A. Hothorn. Are multiple contrast tests superior to the anova? *The International Journal of Biostatistics*, 9(1):63–73, 2013.

[48] J. Koziol and Z. Jia. The concordance index C and the mann-whitney parameter Pr(X>Y) with randomly censored data. *Biometrical journal*, 51(3):467–74, 07 2009.

[49] A.-L. Künster. Eine vergleichende Untersuchung von Methoden zur Bewertung des PFS Ratios in Phase-II-Studien. Bachelor's thesis, Hochschule Koblenz, RheinAhrCampus Remagen, conducted at the Deutsches Krebsforschungszentrum (DKFZ), not publicly available, 2023.

[50] M. Lee and J. P. Fine. Inference for cumulative incidence quantiles via parametric and nonparametric approaches. *Statistics in Medicine*, 30(27):3221–3235, 2011.

[51] E. L. Lehmann, J. P. Romano, and G. Casella. *Testing statistical hypotheses*, volume 3. Springer, 1986.

[52] D. Y. Lin, L. J. Wei, and Z. Ying. Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80(3):557–572, 1993.

[53] T. Liu, M. Ditzhaus, and J. Xu. A resampling-based test for two crossing survival curves. *Pharmaceutical Statistics*, 19(4):399–409, 2020.

[54] J. Lyu, J. Chen, Y. Hou, and Z. Chen. Comparison of two treatments in the presence of competing risks. *Pharmaceutical Statistics*, 19, 06 2020.

[55] J. Lyu, Y. Hou, and Z. Chen. The use of restricted mean time lost under competing risks data. *BMC Medical Research Methodology*, 20(1):1–11, 2020.

[56] Z. Ma and A. W. Krings. Competing risks analysis of reliability, survivability, and prognostics and health management (phm). In *2008 IEEE Aerospace Conference*, pages 1–21, 2008.

[57] N. Mantel. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports*, 50(3):163—170, March 1966.

[58] M. Munko, M. Ditzhaus, D. Dobler, and J. Genuneit. RMST-based multiple contrast tests in general factorial designs. *Statistics in Medicine*, 43(10):1849–1866, 2024.

[59] M. Munko and D. Dobler. Conditional delta-method for resampling empirical processes in multiple sample problems. *arXiv preprint arXiv:2408.10856*, 2024.

[60] M. Munko, D. Dobler, and M. Ditzhaus. Multiple tests for restricted mean time lost with competing risks data. *arXiv preprint arXiv:2409.07917*, 2024.

[61] N. Neumeyer and M. Omelka. Generalized hadamard differentiability of the copula mapping and its applications. *Bernoulli*, 31(2):1451–1474, 2025.

[62] M. I. Parzen, L. J. Wei, and Z. Ying. Simultaneous confidence intervals for the difference of two survival functions. *Scandinavian Journal of Statistics*, 24(3):309–314, 1997.

[63] M. Pauly, E. Brunner, and F. Konietschke. Asymptotic permutation tests in general factorial designs. *Journal of the Royal Statistical Society Series B*, 77(2):461–473, 2015.

[64] L. Peng and J. P. Fine. Nonparametric quantile inference with competing-risks data. *Biometrika*, 94(3):735–744, 2007.

[65] H. Putter, M. Fiocco, and R. B. Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11):2389–2430, 2007.

[66] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2022.

[67] C. Rao and S. K. Mitra. *Generalized Inverse of Matrices and Its Applications.* New York-London-Sydney: John Wiley & Sons, 1971.

[68] P. Royston and M. Parmar. Restricted mean survival time: An alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Medical Research Methodology*, 13:152, 12 2013.

[69] P. Sankaran, N. Unnikrishnan Nair, and E. Sreedevi. A quantile based test for comparing cumulative incidence functions of competing risks models. *Statistics & Probability Letters*, 80(9):886–891, 2010.

[70] D. J. Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, 15(6):657–680, 1987.

[71] A. Shapiro. On concepts of directional differentiability. *Journal of optimization theory and applications*, 66:477–487, 1990.

[72] G. Shilov, B. Gurevich, and R. Silverman. *Integral, Measure, and Derivative: A Unified Approach*. Dover books on advanced mathematics. Dover Publications, 1966.

[73] J. W. Tukey. The problem of multiple comparisons. 1953. Unpublished manuscript reprinted in: *The Collected Works of John W. Tukey, Volume 8 (1994)*, Braun, H.I. (ed.), Chapman and Hall, NewYork.

[74] A. v. d. Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, Cham, Switzerland, 2023.

[75] D. D. Von Hoff. There are no bad anticancer agents, only bad clinical trial designs–twenty-first Richard and Hinda Rosenthal Foundation Award Lecture. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 4(5):1079–1086, 1998.

[76] D. D. Von Hoff, J. J. Stephenson Jr, P. Rosen, D. M. Loesch, M. J. Borad, S. Anthony, G. Jameson, S. Brown, N. Cantafio, D. A. Richards, T. R. Fitch, E. Wasserman, C. Fernandez, S. Green, W. Sutherland, M. Bittner, A. Alarcon, D. Mallery, and R. Penny. Pilot study using molecular profiling of patients' tumors to find potential targets and select treatments for their refractory cancers. *Journal of clinical oncology*, 28(33):4877–4883, 2010.

[77] J. Wen, M.-C. Wang, and C. Hu. Simultaneous hypothesis testing for multiple competing risks in comparative clinical trials. *Statistics in Medicine*, 42(14):2394–2408, 2023.

[78] H. Wu, H. Yuan, Z. Yang, Y. Hou, and Z. Chen. Implementation of an alternative method for assessing competing risks: restricted mean time lost. *American Journal of Epidemiology*, 191(1):163–172, 2022.

[79] L. Zhao, L. Tian, B. Claggett, M. Pfeffer, D. H. Kim, S. Solomon, and L. Wei. Estimating treatment effect with clinical interpretation from a comparative clinical trial with an end point subject to competing risks. *JAMA Cardiology*, 3, 03 2018.

[80] L. Zhao, L. Tian, H. Uno, S. Solomon, M. Pfeffer, J. Schindler, and L. Wei. Utilizing the integrated difference of two survival functions to quantify the treatment contrast for designing, monitoring, and analyzing a comparative clinical study. *Clinical Trials*, 9(5):570–577, 2012. PMID: 22914867.

A Applications of the Conditional Delta-Method on Exemplary Functionals

In this section, we will verify the uniform Hadamard differentiability of exemplary functionals as the Wilcoxon functional and the product integral functional. The verifications of the uniform Hadamard differentiability of these functionals roughly follow the lines of Lemma 3.10.18 and Lemma 3.10.32 in [74].

A.1 Wilcoxon functional

Let $[a, b] \subset \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ and

$$\psi : BV_M[a, b] \times BV_M[a, b] \to D[a, b], \quad \psi(A, B) := \int_{(a,.]} A \, \mathrm{d}B$$

denote the Wilcoxon functional. We aim to show the uniform Hadamard differentiability at $(A, B) \in \mathbb{D}_\psi$ tangentially to $D[a, b] \times D[a, b]$ with Hadamard derivative

$$\psi'_{(A,B)} : D[a, b] \times D[a, b] \to D[a, b], \quad \psi'_{(A,B)}(\alpha, \beta) = \int_{(a,.]} A \, \mathrm{d}\beta + \int_{(a,.]} \alpha \, \mathrm{d}B;$$

cf. Lemma 3.10.18 in [74]. Here and below, the integral w.r.t. β is defined via integration by parts if β has unbounded variation. Let $t \to 0$, $A_t, A, B_t, B \in BV_M[a, b]$, $\alpha_t, \alpha, \beta_t, \beta \in D[a, b]$ such that $||A_t - A||_\infty \to 0, ||B_t - B||_\infty \to 0, ||\alpha_t - \alpha||_\infty \to 0, ||\beta_t - \beta||_\infty \to 0$ and $A_t + t\alpha_t, B_t + t\beta_t \in BV_M[a, b]$. As in the proof of Lemma 3.10.18 in [74], we consider

$$\frac{\psi(A_t + t\alpha_t, B_t + t\beta_t) - \psi(A_t, B_t)}{t} - \psi'_{(A,B)}(\alpha_t, \beta_t) = \int_{(a,.]} (A_t - A) \, \mathrm{d}\beta_t + \int_{(a,.]} \alpha_t \, \mathrm{d}(B_t + t\beta_t - B).$$

The second term converges to zero by proceeding as in the proof of Lemma 3.10.18 in [74]. For the first term, we apply integration by parts to obtain

$$\begin{aligned}\left|\int_{(a,.]} (A_t - A) \, \mathrm{d}\beta_t\right| &= \left|(A_t - A)(.)\beta_t(.) - (A_t - A)(a)\beta_t(a) - \int_{(a,.]} \beta_{t-}(u) \, \mathrm{d}(A_t - A)(u)\right| \\ &\leqslant 2||A_t - A||_\infty ||\beta_t||_\infty + \left|\int_{(a,.]} \beta_{t-}(u) \, \mathrm{d}(A_t - A)(u)\right|,\end{aligned}$$

where here and throughout $\beta_{t-}(u) := \lim_{s \nearrow u} \beta_t(s)$ denotes the left-continuous version of β_t at u. The first term converges to zero by $||A_t - A||_\infty \to 0$ and the second term converges to zero as in the proof of Lemma 3.10.18 in [74]. Hence, we showed that

$$\frac{\psi(A_t + t\alpha_t, B_t + t\beta_t) - \psi(A_t, B_t)}{t} - \psi'_{(A,B)}(\alpha_t, \beta_t) \to 0$$

and, by the continuity of $\psi'_{(A,B)}$, the uniform Hadamard differentiability of the Wilcoxon functional follows.

Example A.1 (Wilcoxon statistic)**.** *Let $a = -\infty, b = \infty$. We consider the case of two independent samples $X_1, \ldots, X_n \sim F$ and $Y_1, \ldots, Y_m \sim G$ taking values in $\mathbb{R}$ with empirical distribution functions $\mathbb{F}_n, \mathbb{G}_m$, respectively. The Wilcoxon statistic $\psi(\mathbb{F}_n, \mathbb{G}_k)(\infty) = \int_{-\infty}^{\infty} \mathbb{F}_n \, \mathrm{d}\mathbb{G}_m$ is an estimator of $\psi(F, G)(\infty) = P(X_1 \leqslant Y_1)$. In the following, we assume $n/(n + m) \to \kappa_1 > 0, m/(n + m) \to \kappa_2 > 0$. Furthermore, let us consider the P^{X_1}- and P^{Y_1}-Donsker class $\mathcal{F} := \{x \mapsto \mathbb{1}\{x \leqslant t\} \mid t \in \mathbb{R}\}$, cf. Example 2.1.3 in [74], with $||P^{X_1}||_\mathcal{F}, ||P^{Y_1}||_\mathcal{F} \leqslant 1$. As in Example 3.10.19, we get*

$$\sqrt{\frac{nm}{n + m}} \left(\int_\mathbb{R} \mathbb{F}_n \, \mathrm{d}\mathbb{G}_m - \int_\mathbb{R} F \, \mathrm{d}G\right) \rightsquigarrow \sqrt{\kappa_2} \int_\mathbb{R} F \, \mathrm{d}\mathbb{G}_G + \sqrt{\kappa_1} \int_\mathbb{R} \mathbb{G}_F \, \mathrm{d}G,$$

where $\mathbb{G}_F, \mathbb{G}_G$ denote independent tight F- and G-Brownian bridges.
Furthermore, we get

$$\sqrt{\frac{nm}{n + m}} (\mathbb{H}_{n+m} - H_{n+m}) \rightsquigarrow \sqrt{\kappa_2}\kappa_1 \mathbb{G}_F + \sqrt{\kappa_1}\kappa_2 \mathbb{G}_G \quad \text{in } D[-\infty, \infty]$$

and $H_{n+m} \to H := \kappa_1 F + \kappa_2 G$ for the pooled empirical distribution function $\mathbb{H}_{n+m} := \frac{n}{n+m}\mathbb{F}_n + \frac{m}{n+m}\mathbb{G}_m$ and $H_{n+m} := \frac{n}{n+m}F + \frac{m}{n+m}G$.

*For deriving the asymptotic behavior of the permutation and pooled bootstrap counterpart of the Wilcoxon statistic, we denote the empirical distribution functions of the permutation and pooled bootstrap samples as $\mathbb{F}_n^\pi, \mathbb{G}_m^\pi, \hat{\mathbb{F}}_n, \hat{\mathbb{G}}_m$, respectively. Then, Theorem 2.2 and (**??**) yield*

$$\sqrt{n+m}(\mathbb{F}_n^\pi - \mathbb{H}_{n+m}, \mathbb{G}_m^\pi - \mathbb{H}_{n+m}) \rightsquigarrow (\mathbb{G}_{H,1}^\pi, \mathbb{G}_{H,2}^\pi)$$
$$\sqrt{n+m}(\hat{\mathbb{F}}_n - \mathbb{H}_{n+m}, \hat{\mathbb{G}}_m - \mathbb{H}_{n+m}) \rightsquigarrow (\kappa_1^{-1/2}\mathbb{G}_{H,1}, \kappa_2^{-1/2}\mathbb{G}_{H,2})$$

in $(D[-\infty,\infty])^2$ conditionally in outer probability, where $\mathbb{G}_H^\pi$ denotes a tight zero-mean Gaussian process with

$$\mathrm{E}\left[\mathbb{G}_{H,i}^\pi(s)\mathbb{G}_{H,j}^\pi(t)\right] = (\kappa_i^{-1}\mathbb{1}\{i=j\} - 1)(H(\min\{s,t\}) - H(s)H(t))$$

and $\mathbb{G}_{H,1}, \mathbb{G}_{H,2}$ denote independent tight H-Brownian bridges. Since $H, H_{n+m}, \mathbb{H}_{n+m}, \mathbb{F}_n^\pi, \mathbb{G}_n^\pi, \hat{\mathbb{F}}_n, \hat{\mathbb{G}}_n$ are (empirical) distribution functions, the total variations are bounded by $M = 1$. By Theorem 2.5 and the uniform Hadamard differentiability of the Wilcoxon functional, we get

$$\sqrt{n+m}(\psi(\mathbb{F}_n^\pi, \mathbb{G}_m^\pi) - \psi(\mathbb{H}_{n+m}, \mathbb{H}_{n+m})) \rightsquigarrow \psi'_{(H,H)}(\mathbb{G}_{H,1}^\pi, \mathbb{G}_{H,2}^\pi),$$
$$\sqrt{n+m}(\psi(\hat{\mathbb{F}}_n, \hat{\mathbb{G}}_m) - \psi(\mathbb{H}_{n+m}, \mathbb{H}_{n+m})) \rightsquigarrow \psi'_{(H,H)}(\kappa_1^{-1/2}\mathbb{G}_{H,1}, \kappa_2^{-1/2}\mathbb{G}_{H,2})$$

in $(D[-\infty,\infty])^2$ conditionally in outer probability. Thus, it follows that

$$\sqrt{\frac{nm}{n+m}}\left(\int_{\mathbb{R}} \mathbb{F}_n^\pi \,\mathrm{d}\mathbb{G}_m^\pi - \int_{\mathbb{R}} \mathbb{H}_{n+m}\,\mathrm{d}\mathbb{H}_{n+m}\right) \rightsquigarrow \sqrt{\kappa_1\kappa_2}\left(\int_{\mathbb{R}} H\,\mathrm{d}\mathbb{G}_{H,2}^\pi + \int_{\mathbb{R}} \mathbb{G}_{H,1}^\pi\,\mathrm{d}H\right),$$
$$\sqrt{\frac{nm}{n+m}}\left(\int_{\mathbb{R}} \hat{\mathbb{F}}_n \,\mathrm{d}\hat{\mathbb{G}}_m - \int_{\mathbb{R}} \mathbb{H}_{n+m}\,\mathrm{d}\mathbb{H}_{n+m}\right) \rightsquigarrow \sqrt{\kappa_1}\int_{\mathbb{R}} H\,\mathrm{d}\mathbb{G}_{H,2} + \sqrt{\kappa_2}\int_{\mathbb{R}} \mathbb{G}_{H,1}\,\mathrm{d}H$$

conditionally in outer probability by Slutsky's lemma.

As we turn to survival analysis in Sections 3–5, we investigate the asymptotic behavior of the Nelson-Aalen estimator and its resampling versions here.

Example A.2 (Nelson-Aalen estimator)**.** *Let us consider a survival setup with multiple samples, i.e., each data point $\boldsymbol{X}_{ij} = (Z_{ij}, \Delta_{ij})$ consists of a (censored) failure time $Z_{ij} = \min\{X_{ij}, C_{ij}\}$ and a censoring status $\Delta_{ij} = \mathbb{1}\{X_{ij} \leqslant C_{ij}\}$, see Example 3.10.20 in [74] for details. Furthermore, let*

$$\overline{\mathbb{H}}_{i,n_i}(t) := \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbb{1}\{Z_{ij} \geqslant t\} \quad \textit{and} \quad \mathbb{H}_{i,n_i}^{uc}(t) := \frac{1}{n_i}\sum_{j=1}^{n_i}\Delta_{ij}\mathbb{1}\{Z_{ij} \leqslant t\}$$

denote the survival function of the observation times and the empirical subdistribution functions of the uncensored failure times, respectively, and $\overline{H}_i(t) := P(Z_{ij} \geqslant t), H_i^{uc}(t) := P(Z_{ij} \leqslant t, \Delta_{ij} = 1)$. The Nelson-Aalen estimator

$$\Lambda_{i,n_i}(.) := \int_{[0,.]} \frac{1}{\overline{\mathbb{H}}_{i,n_i}}\,\mathrm{d}\mathbb{H}_{i,n_i}^{uc}$$

estimates the cumulative hazard function

$$\Lambda_i(.) := \int_{[0,.]} \frac{1}{\overline{H}_i}\,\mathrm{d}H_i^{uc}.$$

To derive the asymptotic behavior of the Nelson-Aalen estimators, let us consider the $P_i := P^{(Z_{i1},\Delta_{i1})}$-Donsker class

$$\mathcal{F} := \left\{f_t^{(1)} : (x,d) \mapsto \mathbb{1}\{x \geqslant t\}, f_t^{(2)} : (x,d) \mapsto \mathbb{1}\{x \leqslant t, d = 1\} \mid t \in [0,\tau]\right\}$$

for some $\tau > 0$. Then, we have

$$\sqrt{N}(\overline{\mathbb{H}}_{i,n_i} - \overline{H}_i, \mathbb{H}_{i,n_i}^{uc} - H_i^{uc})_{i\in\{1,\ldots,k\}} \rightsquigarrow (\kappa_i^{-1/2}\overline{\mathbb{G}}_i, \kappa_i^{-1/2}\mathbb{G}_i^{uc})_{i\in\{1,\ldots,k\}} \quad \textit{in } (\tilde{D}[0,\tau] \times D[0,\tau])^k$$

as in Example 3.10.20 in [74]. Here, $(\overline{\mathbb{G}}_i, \mathbb{G}_i^{uc}), i \in \{1,\ldots,k\}$, denote independent tight, zero-mean Gaussian processes with covariance structure

$$\mathrm{E}\left[\overline{\mathbb{G}}_i(s)\overline{\mathbb{G}}_i(t)\right] = \overline{H}_i(\max\{s,t\}) - \overline{H}_i(s)\overline{H}_i(t),$$
$$\mathrm{E}\left[\mathbb{G}_i^{uc}(s)\overline{\mathbb{G}}_i(t)\right] = (H_i^{uc}(s) - H_{i-}^{uc}(t))\mathbb{1}\{t \leqslant s\} - H_i^{uc}(s)\overline{H}_i(t),$$
$$\mathrm{E}\left[\mathbb{G}_i^{uc}(s)\mathbb{G}_i^{uc}(t)\right] = H_i^{uc}(\min\{s,t\}) - H_i^{uc}(s)H_i^{uc}(t),$$

and throughout $\tilde{D}[0,\tau]$ denotes the subset of all functions $[0,\tau] \to \mathbb{R}$ that are everywhere left-continuous and have right limits everywhere, equipped with the sup-norm.
For the pooled empirical subdistribution functions $\overline{\mathbb{H}}_N := \sum_{i=1}^k \frac{n_i}{N} \overline{\mathbb{H}}_{i,n_i}, \mathbb{H}_N^{uc} := \sum_{i=1}^k \frac{n_i}{N} \mathbb{H}_{i,n_i}^{uc}$, it follows that

$$\sqrt{N}\left(\overline{\mathbb{H}}_N - \sum_{i=1}^k \frac{n_i}{N}\overline{H}_i, \mathbb{H}_N^{uc} - \sum_{i=1}^k \frac{n_i}{N} H_i^{uc}\right) \rightsquigarrow \left(\sum_{i=1}^k \kappa_i^{1/2}\overline{\mathbb{G}}_i, \sum_{i=1}^k \kappa_i^{1/2}\mathbb{G}_i^{uc}\right) \quad \textit{in } \tilde{D}[0,\tau] \times D[0,\tau].$$

Furthermore, we have that $\sum_{i=1}^k \frac{n_i}{N}\overline{H}_i \to \overline{H} := \sum_{i=1}^k \kappa_i \overline{H}_i$ and $\sum_{i=1}^k \frac{n_i}{N} H_i^{uc} \to H^{uc} := \sum_{i=1}^k \kappa_i H_i^{uc}$ in the sup-norm.
Let us assume $\overline{H}_i(\tau) > 0$ for all $i \in \{1,\ldots,k\}$ in the following. Then, the (classical) functional delta-method (Theorem 3.10.4 in [74]) implies

$$\sqrt{N}(\Lambda_{i,n_i} - \Lambda_i)_{i\in\{1,\ldots,k\}} \rightsquigarrow \left(\kappa_i^{-1/2}\mathbb{Z}_i(C_i)\right)_{i\in\{1,\ldots,k\}} \quad \textit{in } (D[0,\tau])^k,$$

where $\mathbb{Z}_i, i \in \{1,\ldots,k\}$, are independent standard Brownian motions and

$$C_i(.) = \int_{[0,.]} \frac{1-\Delta\Lambda_i}{\overline{H}_i}\, \mathrm{d}\Lambda_i.$$

as in Example 3.10.20 in [74].
*Now, we are considering the permutation and pooled bootstrap counterparts of the Nelson-Aalen estimators. Therefore, we denote all processes and statistics introduced above with a π in the superscript, if they are based on the permuted data $\boldsymbol{Z}_{NR_1}, \ldots, \boldsymbol{Z}_{NR_N}$ instead of the original data, and with a hat $\hat{}$, if they are based on the bootstrapped data $\hat{\boldsymbol{Z}}_{N1}, \ldots, \hat{\boldsymbol{Z}}_{NN}$. Theorem 2.2 and (**??**) imply the conditional weak convergence of the permutation and pooled bootstrap empirical processes*

$$\begin{aligned}
\sqrt{N}(\overline{\mathbb{H}}_{i,n_i}^{\pi} - \overline{\mathbb{H}}_N, \mathbb{H}_{i,n_i}^{uc,\pi} - \mathbb{H}_N^{uc})_{i\in\{1,\ldots,k\}} &\rightsquigarrow (\overline{\mathbb{G}}_i^{\pi}, \mathbb{G}_i^{uc,\pi})_{i\in\{1,\ldots,k\}} \quad \textit{in } (\tilde{D}[0,\tau]\times D[0,\tau])^k,\\
\sqrt{N}(\hat{\overline{\mathbb{H}}}_{i,n_i} - \overline{\mathbb{H}}_N, \hat{\mathbb{H}}_{i,n_i}^{uc} - \mathbb{H}_N^{uc})_{i\in\{1,\ldots,k\}} &\rightsquigarrow (\kappa_i^{-1/2}\hat{\overline{\mathbb{G}}}_i, \kappa_i^{-1/2}\hat{\mathbb{G}}_i^{uc})_{i\in\{1,\ldots,k\}} \quad \textit{in } (\tilde{D}[0,\tau]\times D[0,\tau])^k,
\end{aligned}$$

conditionally in outer probability due to $||P_i||_{\mathcal{F}} \leqslant 1$. Here, $(\overline{\mathbb{G}}_i^{\pi}, \mathbb{G}_i^{uc,\pi})_{i\in\{1,\ldots,k\}}$ is a tight, zero-mean Gaussian process with

$$\begin{aligned}
\mathrm{E}\left[\overline{\mathbb{G}}_i^{\pi}(s)\overline{\mathbb{G}}_j^{\pi}(t)\right] &= (\kappa_i^{-1}\mathbb{1}\{i=j\} - 1)(\overline{H}(\max\{s,t\}) - \overline{H}(s)\overline{H}(t)),\\
\mathrm{E}\left[\mathbb{G}_i^{uc,\pi}(s)\overline{\mathbb{G}}_j^{\pi}(t)\right] &= (\kappa_i^{-1}\mathbb{1}\{i=j\} - 1)\left((H^{uc}(s) - H_-^{uc}(t))\mathbb{1}\{t \leqslant s\} - H^{uc}(s)\overline{H}(t)\right),\\
\mathrm{E}\left[\mathbb{G}_i^{uc,\pi}(s)\mathbb{G}_j^{uc,\pi}(t)\right] &= (\kappa_i^{-1}\mathbb{1}\{i=j\} - 1)(H^{uc}(\min\{s,t\}) - H^{uc}(s)H^{uc}(t)),
\end{aligned}$$

and $(\hat{\overline{\mathbb{G}}}_i, \hat{\mathbb{G}}_i^{uc}), i \in \{1,\ldots,k\}$, are independent tight, zero-mean Gaussian processes with

$$\begin{aligned}
\mathrm{E}\left[\hat{\overline{\mathbb{G}}}_i(s)\hat{\overline{\mathbb{G}}}_i(t)\right] &= \overline{H}(\max\{s,t\}) - \overline{H}(s)\overline{H}(t),\\
\mathrm{E}\left[\hat{\mathbb{G}}_i^{uc}(s)\hat{\overline{\mathbb{G}}}_i(t)\right] &= (H^{uc}(s) - H_-^{uc}(t))\mathbb{1}\{t \leqslant s\} - H^{uc}(s)\overline{H}(t),\\
\mathrm{E}\left[\hat{\mathbb{G}}_i^{uc}(s)\hat{\mathbb{G}}_i^{uc}(t)\right] &= H^{uc}(\min\{s,t\}) - H^{uc}(s)H^{uc}(t).
\end{aligned}$$

The Nelson-Aalen functional is a composition of the functionals

$$\tilde{\psi} : \tilde{BV}_M[0,\tau] \times BV_M[0,\tau] \to D[0,\tau], \quad \tilde{\psi}(A,B) := \int_{[0,.]} A\, \mathrm{d}B$$

and $(A,B) \mapsto (1/A, B)$, where here and throughout $\tilde{BV}_M[0,\tau] \subset \tilde{D}[0,\tau]$ denotes the subset of functions with total variation bounded by $M < \infty$. Furthermore, we set $\Delta B(0) := B(0)$ for $B \in D[0,\tau]$ in the following to guarantee a well-defined jump in 0. Similarly to the above calculations for the Wilcoxon functional, it can be shown that $\tilde{\psi}$ is uniformly Hadamard differentiable at $(A,B) \in \mathbb{D}_{\tilde{\psi}}$ with Hadamard derivative

$$\tilde{\psi}'_{(A,B)} : \tilde{D}[0,\tau] \times D[0,\tau] \to D[0,\tau], \quad \tilde{\psi}'_{(A,B)}(\alpha,\beta) = \int_{[0,.]} A\, \mathrm{d}\beta + \int_{[0,.]} \alpha\, \mathrm{d}B.$$

Furthermore, it is easy to show that $(A,B) \mapsto (1/A,B)$ is uniformly Hadamard differentiable at $(A,B) \in \tilde{D}[0,\tau] \times D[0,\tau]$ such that $|A| \geqslant \varepsilon$ for some $\varepsilon > 0$ with Hadamard derivative $(\alpha,\beta) \mapsto (-\alpha/A^2, \beta)$. Hence, the

Nelson-Aalen functional $(A, B) \mapsto \tilde{\psi}(1/A, B)$ *is uniformly Hadamard differentiable at* (A, B) *with Hadamard derivative* $(\alpha, \beta) \mapsto \tilde{\psi}'_{(1/A,B)}(-\alpha/A^2, \beta)$ *by the chain rule (Theorem 2.4), where* $|A| \geqslant \varepsilon$ *and* $(1/A, B) \in \mathbb{D}_{\tilde{\psi}}$*. Since* $\overline{H}_i, H_i^{uc}, \overline{H}, H^{uc}$ *are positive monotone functions, we have*

$$\overline{H}_i, \overline{H} > \min\{\overline{H}_1(\tau), \ldots, \overline{H}_k(\tau)\} =: 2\varepsilon > 0$$

and the total variation of $1/\overline{H}_i, H_i^{uc}, 1/\overline{H}, H^{uc}$ *is bounded by* $M := \varepsilon^{-1}$*. Moreover, the considered empirical processes are contained in* $\{A \mid A \geqslant \varepsilon, 1/A \in BV_M[0,\tau]\} \times BV_M[0,\tau]$ *with probability tending to 1 by monotonicity and Glivenko-Cantelli arguments. Hence, the uniform Hadamard differentiability of the Nelson-Aalen functional and the conditional delta-method (Theorem 2.5) yield*

$$\sqrt{N}(\Lambda^\pi_{i,n_i} - \Lambda_N)_{i\in\{1,\ldots,k\}} \rightsquigarrow (\mathbb{Z}_i^\pi)_{i\in\{1,\ldots,k\}} \quad \text{in } (D[0,\tau])^k,$$
$$\sqrt{N}(\hat{\Lambda}_{i,n_i} - \Lambda_N)_{i\in\{1,\ldots,k\}} \rightsquigarrow \left(\kappa_i^{-1/2}\mathbb{Z}_i(C)\right)_{i\in\{1,\ldots,k\}} \quad \text{in } (D[0,\tau])^k$$

conditionally in outer probability similarly to the calculations in Example 3.10.20 in [74], where $\Lambda_N(.) := \int_{[0,.]} \frac{1}{\overline{\mathbb{H}}_N} \,\mathrm{d}\mathbb{H}_N^{uc}$ *denotes the pooled Nelson-Aalen estimator,*

$$C(.) := \int_{[0,.]} \frac{1 - \Delta\Lambda}{\overline{H}} \,\mathrm{d}\Lambda,$$

$\Lambda(.) := \int_{[0,.]} \frac{1}{\overline{H}} \,\mathrm{d}H^{uc}$ *and* $(\mathbb{Z}_i^\pi)_{i\in\{1,\ldots,k\}}$ *is a zero-mean Gaussian process with*

$$\mathrm{E}\left[\mathbb{Z}_i^\pi(s)\mathbb{Z}_j^\pi(t)\right] = (\kappa_i^{-1}\mathbb{1}\{i = j\} - 1)C(\min\{s,t\}).$$

A.2 Product integral

Consider ϕ, the product integral functional, i.e.,

$$\phi : BV_M^{>-1}[a,b] \to D[a,b], \quad A \mapsto \prod_{u\in(a,\cdot]} (1 + \mathrm{d}A(u)).$$

Here, $BV_M^{>-1}[a,b] \subset D[a,b]$ is the subset of functions $[a,b] \to \mathbb{R}$ with total variation bounded by M and whose jumps are contained in $(-1,\infty)$ and bounded away from -1 for each function. To analyze the uniform Hadamard differentiability of ϕ, let $t_n \to 0$, $A_n, A \in BV_M^{>-1}[a,b]$ such that $\|A_n - A\|_\infty \to 0$, and $\alpha_n, \alpha \in D[a,b]$ such that $\|\alpha_n - \alpha\|_\infty \to 0$ and $A_n + t_n\alpha_n \in BV_M^{>-1}[a,b]$. Let $\varepsilon > 0$ and $\tilde{\alpha} \in BV[a,b]$ such that $\|\alpha - \tilde{\alpha}\|_\infty < \varepsilon$, $\|\alpha_n - \alpha\|_\infty < \varepsilon$, and $\|A_n - A\|_\infty < \varepsilon$ for sufficiently large n. This function $\tilde{\alpha}$ can be defined piece-wise constant and with finitely many jumps because it approximates the càdlàg function α. Also, because the sequence $(\alpha_n)_n$ approximates α uniformly, it is clear that such a function $\tilde{\alpha}$ exists.
It is well-known that

$$\phi'_A : D[a,b] \to D[a,b], \quad \alpha \mapsto \int_{(a,\cdot]} \phi(A)_-(u)\frac{\phi(A)(\cdot)}{\phi(A)(u)} \,\mathrm{d}\alpha(u) = \int_{(a,\cdot]} \frac{1}{1 + \Delta A(u)} \,\mathrm{d}\alpha(u)\phi(A)(\cdot)$$

defines the Hadamard derivative of ϕ at A in the classical sense; cf. [39]. Note that the Hadamard derivative above may also be written as

$$\phi'_A(\alpha) = \phi(A)(\cdot)\left(\alpha(\cdot) - \alpha(a) - \sum_{u\in D_A\cap(a,\cdot]} \frac{\Delta A(u)\Delta\alpha(u)}{1 + \Delta A(u)}\right)$$

where $D_A \subset (a,b]$ is the set of discontinuities of A. Due to the finite variation of A and its boundedness of its jumps away from -1, this representation reveals that $\tilde{\phi}' := ((A,\alpha) \mapsto \phi'_A(\alpha))$ defines a continuous functional from $BV_M^{>-1}[a,b] \times D[a,b] \to D[a,b]$ with respect to the maximum-supremum norm. To see this, let us focus on the sum-term and notice that

$$\begin{aligned}
&\sum_{u\in D_{A_n}\cap(a,\cdot]} \frac{\Delta A_n(u)\Delta\alpha_n(u)}{1 + \Delta A_n(u)} - \sum_{u\in D_A\cap(a,\cdot]} \frac{\Delta A(u)\Delta\alpha(u)}{1 + \Delta A(u)} \\
&= \sum_{u\in(D_{A_n}\cup D_A)\cap(a,\cdot]} \frac{\Delta A_n(u)\Delta A(u)\Delta(\alpha_n(u) - \alpha(u)) + \Delta A_n(u)\Delta\alpha_n(u) - \Delta A(u)\Delta\alpha(u)}{(1 + \Delta A_n(u))(1 + \Delta A(u))}.
\end{aligned}$$

Choose $\delta > 0$ sufficiently small, i.e., $\delta < \min(\varepsilon/2, \inf_u(1 + \Delta A(u)))$ such that a finite, positive constant $K \geqslant \sup_u(1+\Delta A(u)-\delta)^{-1}$ exists. Now, choose n_0 sufficiently large such that $\sup_u |\Delta A_n(u)-\Delta A(u)| \leqslant 2\|A_n-A\|_\infty \leqslant$

$2\varepsilon < \delta$ for all $n \geqslant n_0$. Let us only consider such n henceforth. This implies that $K \geqslant \sup_u(1+\Delta A_n(u))^{-1}$. Since the jumps of A and A_n are bounded away from -1, the supremum norm of the previous display is bounded above by K^2 (due to the denominator) times the sum of

$2\|\alpha_n - \alpha\|_\infty \|A_n\|_{BV} \|A\|_{BV}$	for the term $\Delta A_n(u)\Delta A(u)\Delta(\alpha_n(u) - \alpha(u))$,
$2\|\alpha_n - \alpha\|_\infty \|A_n\|_{BV}$	for a term $\Delta A_n(u)\Delta(\alpha_n(u) - \alpha(u))$,
$2\|\alpha - \tilde{\alpha}\|_\infty(\|A_n\|_{BV} + \|A\|_{BV})$	for a term $\Delta(A_n(u) - A(u))\Delta(\alpha(u) - \tilde{\alpha}(u))$,
$2\|A_n - A\|_\infty \|\tilde{\alpha}\|_{BV}$	for a term $\Delta(A_n(u) - A(u))\Delta\tilde{\alpha}(u)$,

where $||.||_{BV}$ denotes the total variation. Hence, an upper bound is given by $K^2(8\max(M^2, 1) + 2\|\tilde{\alpha}\|_{BV})\varepsilon$ which is arbitrarily small because $\varepsilon > 0$ was arbitrary.
To verify the uniform Hadamard differentiability, we need to show that the following term converges to zero:

$$t_n^{-1}(\phi(A_n + t_n\alpha_n) - \phi(A_n)) - \phi'_A(\alpha) =: I + II$$

where

$$\begin{aligned} I &= t_n^{-1}(\phi(A_n + t_n\alpha_n) - \phi(A_n)) - \phi'_{A_n}(\alpha_n), \\ II &= \phi'_{A_n}(\alpha_n) - \phi'_A(\alpha) \end{aligned}$$

The second term can be written as $\tilde{\phi}'(A_n, \alpha_n) - \tilde{\phi}'(A, \alpha)$ which goes to zero as argued above. Hence, we focus on the first term which, by Duhamel's equation, equals:

$$\begin{aligned} & t_n^{-1}\int_{(a,\cdot]} \phi(A_n + t_n\alpha_n)_-(u)\frac{\phi(A_n)(\cdot)}{\phi(A_n)(u)}\,\mathrm{d}(A_n + t_n\alpha_n - A_n)(u) - \phi'_{A_n}(\alpha_n) \\ &= \int_{(a,\cdot]} (\phi(A_n + t_n\alpha_n)_-(u) - \phi(A_n)_-(u))\,\frac{\phi(A_n)(\cdot)}{\phi(A_n)(u)}\,\mathrm{d}(\alpha_n - \tilde{\alpha} + \tilde{\alpha})(u) \end{aligned}$$

The part with $\alpha_n - \tilde{\alpha}$ is arbitrarily small, which follows from integration-by-parts, combined with the fact that the involved product integrals have a finite variation; cf. the proof of Theorem 7 in [39] for similar arguments. The upper bound for the variation norm of the involved product integrals can be chosen independently of n. The remaining part with $\tilde{\alpha}$ converges to zero in supremum norm due to the uniform continuity of the product integral functional (Theorem 7 in 39) combined with $A_n + t_n\alpha_n \to A, A_n \to A$. Indeed, combine $\|\tilde{\alpha}\|_{BV} < \infty$ with $\|\phi(A_n + t_n\alpha_n)(\cdot) - \phi(A_n)(\cdot)\|_\infty \to 0$ and $\sup_{a\leqslant u\leqslant t\leqslant b} |\phi(A_n)(t)/\phi(A_n)(u)| < \tilde{K}$ for some finite constant $\tilde{K}$ independent of n; cf. the inequality in (20) of [39].

Example A.3 (Kaplan-Meier estimator)**.** *Let us consider the setup of Example A.2. The Kaplan-Meier estimator*

$$\widehat{S}_{i,n_i}(\cdot) := \prod_{u\in[0,\cdot]} (1 - \mathrm{d}\widehat{\Lambda}_{i,n_i}(u))$$

estimates the survival function

$$S_i(\cdot) := P(X_{ij} > \cdot) = \prod_{u\in[0,\cdot]} (1 - \mathrm{d}\Lambda_i(u)).$$

The (classical) functional delta-method (Theorem 3.10.4 in [74]) implies

$$\sqrt{N}(\widehat{S}_{i,n_i} - S_i)_{i\in\{1,\dots,k\}} \rightsquigarrow \left(\kappa_i^{-1/2}\mathbb{U}_i\right)_{i\in\{1,\dots,k\}} \quad in\ (D[0,\tau])^k,$$

where $\mathbb{U}_i, i \in \{1,\dots,k\}$, are independent zero-mean Gaussian processes with covariance structure

$$\mathrm{E}\left[\mathbb{U}_i(s)\mathbb{U}_i(t)\right] = S_i(s)S_i(t)\int_{[0,\min\{s,t\}]} \frac{1}{(1 - \Delta\Lambda_i)\overline{H}_i}\,\mathrm{d}\Lambda_i,$$

which can be shown as in Example 3.10.33 in [74].
The Kaplan-Meier functional is a composition of the functionals

$$\tilde{\phi} : BV_M^{>-1}[0,\tau] \to D[0,\tau], \quad A \mapsto \prod_{u\in[0,\cdot]} (1 + \mathrm{d}A(u)),$$

$A \mapsto -A$, and the Nelson-Aalen functional. Again, we set $\Delta A(0) := A(0)$ for $A \in D[0,\tau]$ to guarantee a well-defined jump in 0. Similarly as above, $\tilde{\phi}$ is uniformly Hadamard differentiable at each $A \in \mathbb{D}_{\tilde{\phi}}$ with Hadamard derivative

$$\tilde{\phi}'_A : D[0,\tau] \to D[0,\tau], \quad \alpha \mapsto \int_{[0,\cdot]} \tilde{\phi}(A)_-(u)\frac{\tilde{\phi}(A)(\cdot)}{\tilde{\phi}(A)(u)}\,\mathrm{d}\alpha(u).$$

Moreover, $A \mapsto -A$ is uniformly Hadamard differentiable at each $A \in D[0,\tau]$ with Hadamard derivative $\alpha \mapsto -\alpha$ due to its linearity. Thus, the Kaplan-Meier functional $A \mapsto \tilde{\phi}(-A)$ is uniformly Hadamard differentiable at each A such that $-A \in \mathbb{D}_{\tilde{\phi}}$ with Hadamard derivative $\alpha \mapsto -\tilde{\phi}'_A(\alpha)$ by Theorem 2.4. To apply the conditional delta-method in Theorem 2.5, we need to ensure that $-\Lambda_i, -\Lambda$ are elements in $\mathbb{D}_{\tilde{\phi}}$. This can be guaranteed by assuming $S_i(\tau) = P(X_{ij} > \tau) > 0$ in the following. Moreover, the Nelson-Aalen estimator and its permutation and pooled bootstrap counterpart are contained in $\{A \mid -A \in \mathbb{D}_{\tilde{\phi}}\}$ with probability tending to 1 by monotonicity and Glivenko-Cantelli arguments. By the uniform Hadamard differentiability of the Kaplan-Meier functional, Theorem 2.5 yields

$$\sqrt{N}(\widehat{S}^{\pi}_{i,n_i} - \widehat{S}_N)_{i\in\{1,\dots,k\}} \rightsquigarrow (\mathbb{U}^{\pi}_i)_{i\in\{1,\dots,k\}} \quad \text{in } (D[0,\tau])^k, \tag{A.1}$$

$$\sqrt{N}(\hat{\widehat{S}}_{i,n_i} - \widehat{S}_N)_{i\in\{1,\dots,k\}} \rightsquigarrow \left(\kappa_i^{-1/2}\hat{\mathbb{U}}_i\right)_{i\in\{1,\dots,k\}} \quad \text{in } (D[0,\tau])^k \tag{A.2}$$

conditionally in outer probability similarly to the calculations in Example 3.10.33 in [74], where $\widehat{S}_N := \tilde{\phi}(-\widehat{\Lambda}_N)$ denotes the pooled Kaplan-Meier estimator. Here, $(\mathbb{U}^{\pi}_i)_{i\in\{1,\dots,k\}}$ is a zero-mean Gaussian process with

$$\mathrm{E}\left[\mathbb{U}^{\pi}_i(s)\mathbb{U}^{\pi}_j(t)\right] = (\kappa_i^{-1}\mathbb{1}\{i=j\} - 1)S(s)S(t)\int_{[0,\min\{s,t\}]} \frac{1}{(1-\Delta\Lambda)\overline{H}}\,\mathrm{d}\Lambda$$

for $S := \tilde{\phi}(-\Lambda)$ and $\hat{\mathbb{U}}_i, i \in \{1,\dots,k\}$, are independent zero-mean Gaussian processes with

$$\mathrm{E}\left[\hat{\mathbb{U}}_i(s)\hat{\mathbb{U}}_i(t)\right] = S(s)S(t)\int_{[0,\min\{s,t\}]} \frac{1}{(1-\Delta\Lambda)\overline{H}}\,\mathrm{d}\Lambda.$$

From this example, we can deduce Theorem 4 and Theorem 5 in the supplement of [31] under $S_i(\tau) > 0, i \in \{1,2\}$. The uniform Hadamard differentiability of the Wilcoxon functional completes the proofs of the consistency for the permutation and pooled bootstrap counterpart of the Mann–Whitney statistic (Theorem 2 and Theorem 3 in [31]).

Furthermore, other works on resampling in survival analysis are based on Theorem 4 and Theorem 5 in the supplement of [31]. This includes the resampling tests for the restricted mean survival times (RMSTs) of [24] and [58]. For the RMSTs, the application of the continuous mapping theorem with continuous function

$$(D[0,\tau))^k \ni (A_1,\dots,A_k) \mapsto \left(\int_0^\tau A_1(t)\,\mathrm{d}t, \dots, \int_0^\tau A_k(t)\,\mathrm{d}t\right)$$

only requires (A.1) and (A.2) in $(D[0,\tau))^k$ instead of $(D[0,\tau])^k$. By replacing all intervals $[0,\tau]$ in Example A.3 by $[0,\tau)$, it is easy to see that the assumption $S_{i-}(\tau) > 0, i \in \{1,\dots,k\}$, (instead of $S_i(\tau) > 0, i \in \{1,\dots,k\}$) is sufficient to ensure (A.1) and (A.2) in $(D[0,\tau))^k$. Hence, the weaker assumption $S_{i-}(\tau) > 0, i \in \{1,\dots,k\}$, is enough for getting the consistency for the permutation and pooled bootstrap counterparts of the RMSTs as in [24] and [58].

A.3 Inverse map: counterexample and additional requirements

Let $p \in \mathbb{R}$ and $A \in D[a,b]$ nondecreasing with $A_-(y) \leqslant p \leqslant A(y)$ for some $y \in (a,b]$. Then, the inverse map Φ_p at A satisfies

$$A_-(\Phi_p(A)) \leqslant p \leqslant A(\Phi_p(A)),$$

where the exact value of $\Phi_p(A)$ is irrelevant if there is more than one solution. Let $\mathbb{D}_{\Phi_p}$ denote the set of all nondecreasing functions A with $A_-(y) \leqslant p \leqslant A(y)$ for some $y \in (a,b]$. As shown in Lemma 3.10.21 of [74], $\Phi_p : \mathbb{D}_{\Phi_p} \to (a,b]$ is Hadamard differentiable at a function $A \in \mathbb{D}_{\Phi_p}$ that is differentiable at $\Phi_p(A) =: \xi_p \in (a,b)$ such that $A(\xi_p) = p$ with positive derivative $A'(\xi_p) > 0$, tangentially to the set of functions $\alpha \in D[a,b]$ that are continuous at ξ_p, with Hadamard derivative $\Phi'_{p,A}(\alpha) = -\alpha(\Phi_p(A))/A'(\Phi_p(A))$ at α. However, the uniform Hadamard differentiability of the inverse map does not hold under these assumptions.

Example A.4. *Let $A : [0,2] \to \mathbb{R}, A(x) = x$,*

$$A_n : [0,2] \to \mathbb{R}, \quad A_n(x) := \begin{cases} x - 1/\sqrt{n} & \text{if } x \leqslant 1 - 1/\sqrt{n}, \\ 2x - 1 & \text{if } 1 + 1/\sqrt{n} > x > 1 - 1/\sqrt{n}, \\ x + 1/\sqrt{n} & \text{if } x \geqslant 1 + 1/\sqrt{n}, \end{cases}$$

$\alpha \equiv \alpha_n \equiv 1$ and $p = 1$. An exemplary illustration for the functions can be found in Figure 14. For $t_n = 1/\sqrt{n}$, we have $\Phi_p(A_n + t_n\alpha_n) = 1 - t_n/2$ since

$$(A_n + t_n\alpha_n)(1 - t_n/2) = A_n(1 - 1/(2\sqrt{n})) + 1/\sqrt{n} = 1 - 1/\sqrt{n} + 1/\sqrt{n} = 1$$

and $\Phi_p(A_n) = 1$. Hence, $(\Phi_p(A_n + t_n\alpha_n) - \Phi_p(A_n))/t_n = -1/2$. However, $\Phi'_{p,A}(\alpha) = -\alpha(\Phi_p(A))/A'(\Phi_p(A)) = -1 \neq -1/2$ and, thus, $\Phi_p : \mathbb{D}_{\Phi_p} \to (0,2]$ is not uniformly Hadamard differentiable at A tangentially to $\alpha \equiv 1$.

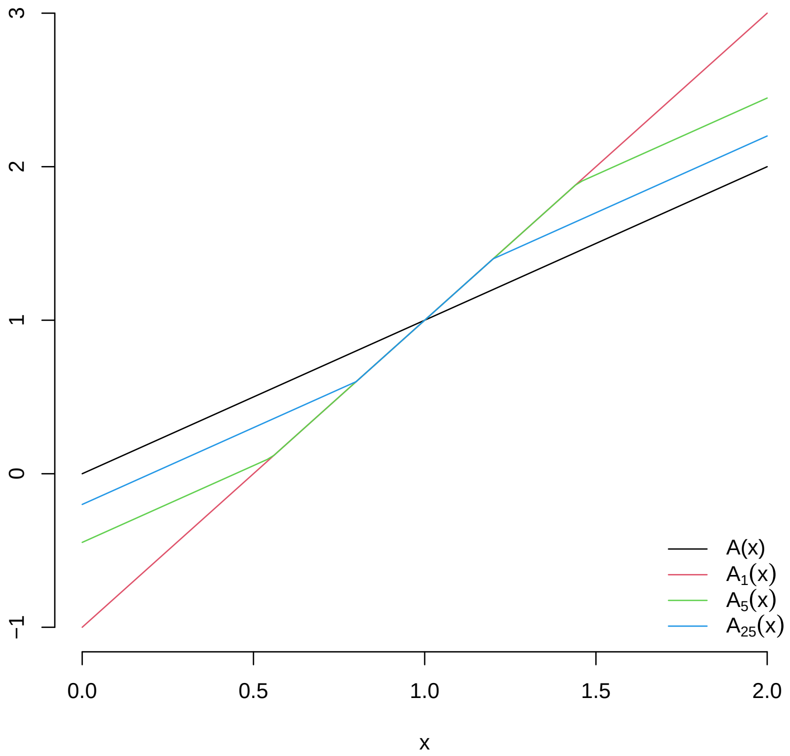


Figure 14: Illustration of the functions A, A_1, A_5, and A_{25}.

A restriction of the definition space $\mathbb{D}_{\Phi_p}$ can lead to uniform Hadamard differentiability. In view of Example A.4, such a restriction needs to exclude the rather simple sequence $(A_n)_n$. That is why such a restriction is not really of interest for applications.
However, in Lemma S.5 in [23], a version of uniform Hadamard differentiability of the inverse map is shown under stricter conditions, where the rate of convergence of the converging sequence $A_n \to A$ and its increments around ξ_p also needs to be controlled. Note that for Example A.4, the condition (A.3) does not hold.

Lemma A.1 (Lemma S.5 of [23])**.** *Let $A_n, A \in \mathbb{D}_{\Phi_p}$ such that A is continuously differentiable at $\xi_p \in \mathbb{R}$ with positive derivative $A'(\xi_p) > 0$. Suppose that $\sqrt{n}||A_n - A||_\infty \leqslant M$ for some $M > 0$ and*

$$\sqrt{n} \sup_{|x| \leqslant K/\sqrt{n}} |A_n(\xi_p + x) - A_n(\xi_p) - A(\xi_p + x) + A(\xi_p)| \to 0 \tag{A.3}$$

for every $K > 0$. Then,

$$\sqrt{n}\left(\Phi_p(A_n + n^{-1/2}\alpha_n) - \Phi_p(A_n)\right) \to \Phi'_{p,A}(\alpha),$$

where $A_n + n^{-1/2}\alpha_n \in \mathbb{D}_{\Phi_p}$ and $\alpha_n \to \alpha$ such that α is bounded and continuous at ξ_p.

As shown in Section S3.3.1 in the supplement of [23], the required conditions are fulfilled for applications on empirical distribution functions. Hence, a central limit theorem for permutation quantiles follows as shown in Lemma S.1 in the supplement of [23]. Analogously, a central limit theorem for pooled bootstrap quantiles could be followed, where $\hat{\gamma}(c,d) = \kappa_c^{-1/2}\mathbb{1}\{c = d\}$ replaces $\gamma^\pi(c,d) = \kappa_c^{-1}\mathbb{1}\{c = d\} - 1$ in the covariance formula given in Lemma S.1.

B Correction of the Limit Distribution for Aalen-Johansen Estimators

In this section, we correct Theorem 5.1 in [26], by mending the limit distribution of the Aalen-Johansen estimator under discontinuous survival distributions.
We consider the same competing risks setup as in [26], i.e., we assume that there are $k \in \mathbb{N}$ competing risks and $n \in \mathbb{N}$ i.i.d. random event times $T_1, ..., T_n$, which are independently right-censored and distributed as a random variable $T \sim S$. Here, S denotes the survival function, i.e., $S(t) = P(T > t)$ for all $t \geqslant 0$; S need not be continuous. Then, we denote the probability that an individual is under observation at time $t-$, that is, *just before* time t, by $\bar{H}(t) = P(\min(T, C) \geqslant t) = S_-(t)G_-(t)$ for all $t \geqslant 0$. Here, $C \sim G$ with survival function $G(t) = P(C > t)$ denotes a generic censoring time which is assumed to be independent of T. Furthermore, let $\widehat{A}_j$ denote the cause-specific Nelson-Aalen estimator for the cumulative hazard function A_j of type j events, $\widehat{S}$ the Kaplan-Meier estimator for the Survival function S, and $\widehat{F}_j(.) = \int_{[0,.]} \widehat{S}_-(u)\mathrm{d}\widehat{A}_j(u)$ the Aalen-Johansen estimator for the cumulative incidence function $F_j(.) = \int_{[0,.]} S_-(u)\mathrm{d}A_j(u)$ for all $j \in \{1, ..., k\}$, see [26] for details. In addition to the assumptions in [26], it is actually required that $\bar{H}(K) > 0$ for $K \geqslant 0$ to ensure finite variances $\sigma_j^2(K), j \in \{1, \ldots, k\}$, in Theorem 4.1 therein.
Theorem 5.1 in [26] states for $k = 2$ competing risks that

$$\sqrt{n}(\widehat{F}_1 - F_1) \xrightarrow{d} U_{F_1}$$

as $n \to \infty$ on the càdlàg space $D[0, K]$ equipped with the sup-norm, where U_{F_1} is a zero-mean Gaussian process with covariance function

$$\begin{aligned}
\sigma_{F_1}^2 : (s,t) \mapsto &\int_{[0,\min\{s,t\}]} \frac{(1 - F_{2-}(u) - F_1(s))(1 - F_{2-}(u) - F_1(t))}{\bar{H}(u)} \frac{\mathrm{d}A_1(u)}{1 - \Delta A(u)} \\
&+ \int_{[0,\min\{s,t\}]} \frac{(F_{1-}(u) - F_1(s))(F_{1-}(u) - F_1(t))}{\bar{H}(u)} \frac{\mathrm{d}A_2(u)}{1 - \Delta A(u)} \\
&+ \sum_{u \in D, u \leqslant s,t} \frac{S_-^2(u)}{\bar{H}(u)} \frac{\Delta A_1(u)\Delta A_2(u)}{(1 - \Delta A(u))^2},
\end{aligned}$$

where $A = \sum_{j=1}^k A_j$ and $D = \{t \in [0, K] : \Delta A(t) > 0\}$ is the set of discontinuity time points. However, we found that the right-continuous versions F_1, F_2, S must appear in the covariance function above in all occurrences of F_{1-}, F_{2-}, S_-.
In order to prove this, we go one step back: By Theorem 4.1 in [26],

$$\sqrt{n}\left(\widehat{A}_1 - A_1, ..., \widehat{A}_k - A_k\right) \xrightarrow{d} (U_1, ..., U_k)$$

holds as $n \to \infty$ on the product space $D^k[0, K]$ equipped with the max-sup norm, where $U_1, ..., U_k$ are zero-mean Gaussian-martingales with

$$\begin{aligned}
\mathbb{C}ov(U_j(t), U_j(s)) &= \int_{[0,\min\{t,s\}]} \frac{1 - \Delta A_j(u)}{\bar{H}(u)} \,\mathrm{d}A_j(u) =: \sigma_j^2(\min\{t, s\}), \\
\mathbb{C}ov(U_j(t), U_\ell(s)) &= -\int_{[0,\min\{t,s\}]} \frac{\Delta A_\ell(u)}{\bar{H}(u)} \,\mathrm{d}A_j(u) =: \sigma_{j\ell}(\min\{t, s\})
\end{aligned}$$

for all $t, s \in [0, K], j, \ell \in \{1, ..., k\}, j \neq \ell$. We further note that the limit $(U_1, ..., U_k)$ is separable since $G_1^{uc}, ..., G_k^{uc}$ and $\overline{G}$ in Appendix A of [26] are tight, which follows by the main empirical central limit theorems in [74], as in Example 3.10.20.
Now it holds that

$$\begin{aligned}
&\sqrt{n}(\widehat{F}_1(t) - F_1(t)) \\
&= \sqrt{n}\left(\int_{[0,t]} \widehat{S}_-(u)\mathrm{d}\widehat{A}_1(u) - \int_{[0,t]} S_-(u)\mathrm{d}A_1(u)\right) \\
&= \int_{[0,t]} \widehat{S}_-(u)\mathrm{d}\sqrt{n}(\widehat{A}_1 - A_1)(u) + \int_{[0,t]} \sqrt{n}(\widehat{S} - S)_-(u)\mathrm{d}A_1(u) \\
&= \sqrt{n}(\widehat{A}_1 - A_1)(t)\widehat{S}(t) - \int_{[0,t]} \sqrt{n}(\widehat{A}_1 - A_1)(u)\mathrm{d}\widehat{S}(u) + \int_{[0,t]} \sqrt{n}(\widehat{S} - S)_-(u)\mathrm{d}A_1(u)
\end{aligned}$$

for all $t \in [0, K]$ by integration by parts, that is

$$\int_{[0,t]} f_-(v)\, \mathrm{d}g(v) = (gf)(t) - (gf)_-(0) - \int_{[0,t]} g(v)\, \mathrm{d}f(v)$$

for $f \in BV_1[0, K], g \in D[0, K]$, where $BV_1[0, K]$ denote the set of all càdlàg functions $D[0, K]$ of total variation bounded by 1. As in Example 3.10.33 in [74], the functional delta method yields

$$\left(\sqrt{n}(\widehat{A}_1 - A_1), \sqrt{n}(\widehat{S} - S)\right) \xrightarrow{d} \left(U_1, -S(.) \int_{[0,.]} \frac{S_-(v)}{S(v)} \mathrm{d}(U_1 + U_2)(v)\right)$$

as $n \to \infty$ on $D^2[0, K]$, where the integral is defined by integration by parts since $U_1 + U_2$ is not of bounded variation. Hence, we get

$$\left(\sqrt{n}(\widehat{A}_1 - A_1), \sqrt{n}(\widehat{S} - S), \widehat{S}\right) \xrightarrow{d} \left(U_1, -S(.) \int_{[0,.]} \frac{S_-(v)}{S(v)} \mathrm{d}(U_1 + U_2)(v), S\right) \tag{B.1}$$

as $n \to \infty$ on $D^2[0, K] \times BV_1[0, K]$ by Slutsky's lemma. Note that the map

$$\begin{aligned} \psi : D^2[0, K] \times BV_1[0, K] &\to D[0, K], \\ (\tilde{A}, \tilde{B}, \tilde{C}) &\mapsto \tilde{A}(.)\tilde{C}(.) - \int_{[0,.]} \tilde{A}\mathrm{d}\tilde{C} - \int_{[0,.]} \tilde{B}_-(u)\mathrm{d}A_1(u) \end{aligned}$$

is continuous on $D^2[0, K] \times \{S\}$. Thus, an application of the continuous mapping theorem and changing the order of integration result in

$$\begin{aligned} &\sqrt{n}(\widehat{F}_1 - F_1) \\ &\xrightarrow{d} U_1(.)S(.) - \int_{[0,.]} U_1 \mathrm{d}S - \int_{[0,.]} S_-(u) \int_{[0,u)} \frac{S_-(v)}{S(v)} \mathrm{d}(U_1 + U_2)(v)\mathrm{d}A_1(u) \\ &= \int_{[0,.]} S_-(u)\mathrm{d}U_1(u) - \int_{[0,.]} \frac{S_-(v)}{S(v)} \int_{(v,.]} S_-(u)\mathrm{d}A_1(u)\mathrm{d}(U_1 + U_2)(v) \\ &= \int_{[0,.]} S_-(u)\mathrm{d}U_1(u) - \int_{[0,.]} \frac{S_-(v)}{S(v)} (F_1(.) - F_1(v))\mathrm{d}(U_1 + U_2)(v) \\ &= \int_{[0,.]} \frac{S_-(u)}{S(u)} \left(S(u) - F_1(.) + F_1(u)\right) \mathrm{d}U_1(u) + \int_{[0,.]} \frac{F_1(u) - F_1(.)}{1 - \Delta A(u)} \mathrm{d}U_2(u) \\ &= \int_{[0,.]} \frac{1 - F_2(u) - F_1(.)}{1 - \Delta A(u)} \mathrm{d}U_1(u) + \int_{[0,.]} \frac{F_1(u) - F_1(.)}{1 - \Delta A(u)} \mathrm{d}U_2(u) \end{aligned}$$

as $n \to \infty$ on $D[0, K]$.

Theorem B.1 (Corrected Theorem 5.1 in [26])**.** *As $n \to \infty$, we have on the càdlàg space $D[0, K]$*

$$\sqrt{n}(\widehat{F}_1 - F_1) \xrightarrow{d} U_{F_1} = \int_{[0,.]} \frac{1 - F_2(u) - F_1(.)}{1 - \Delta A(u)} \mathrm{d}U_1(u) + \int_{[0,.]} \frac{F_1(u) - F_1(.)}{1 - \Delta A(u)} \mathrm{d}U_2(u),$$

where U_{F_1} is a zero-mean Gaussian process with covariance function

$$\begin{aligned} \sigma^2_{F_1} : (s, t) \mapsto &\int_{[0,\min\{s,t\}]} \frac{(1 - F_2(u) - F_1(s))(1 - F_2(u) - F_1(t))}{\bar{H}(u)} \frac{\mathrm{d}A_1(u)}{1 - \Delta A(u)} \\ &+ \int_{[0,\min\{s,t\}]} \frac{(F_1(u) - F_1(s))(F_1(u) - F_1(t))}{\bar{H}(u)} \frac{\mathrm{d}A_2(u)}{1 - \Delta A(u)} \\ &+ \sum_{u \in D, u \leqslant s,t} \frac{S^2(u)}{\bar{H}(u)} \frac{\Delta A_1(u)\Delta A_2(u)}{(1 - \Delta A(u))^2}. \end{aligned}$$

The covariance function can be calculated analogously to Appendix E of [26]. Here, the last sum may be simplified to $\sum_{u \in D, u \leqslant s,t} \frac{S_-(u)}{G_-(u)} \Delta A_1(u)\Delta A_2(u)$.

C Details on the Simulation Results and Additional Simulations for Section 4

In this section, the results of additional simulation studies are provided. First, more detailed results of the simulation study in Section 4.4 can be found. Furthermore, the simulation setup from Section 4.4.1 is repeated for the asymptotic approaches by using larger sample sizes. Next, we show a setup where the groupwise bootstrap outperforms the permutation approach with Bonferroni-correction in terms of empirical power. Finally, a simulation study inspired by the data example in Section 4.5 is investigated.

C.1 Additional Tables and Figures

δ	distribution	censoring distribution	group 1	group 2	group 3	group 4
0.0	exp early, late, prop	equal	0.20	0.21	0.21	0.21
0.0	exp early, late, prop	unequal, high	0.38	0.44	0.38	0.33
0.0	exp early, late, prop	unequal, low	0.20	0.21	0.25	0.06
0.0	logn	equal	0.41	0.41	0.41	0.41
0.0	logn	unequal, high	0.51	0.58	0.60	0.53
0.0	logn	unequal, low	0.33	0.41	0.41	0.11
0.0	pwExp	equal	0.21	0.20	0.21	0.33
0.0	pwExp	unequal, high	0.38	0.44	0.39	0.37
0.0	pwExp	unequal, low	0.20	0.20	0.25	0.22
0.0	Weib late, Weib prop	equal	0.34	0.34	0.34	0.34
0.0	Weib late, Weib prop	unequal, high	0.49	0.56	0.57	0.49
0.0	Weib late, Weib prop	unequal, low	0.29	0.34	0.37	0.06
0.0	Weib scale	equal	0.34	0.34	0.34	0.45
0.0	Weib scale	unequal, high	0.49	0.56	0.57	0.52
0.0	Weib scale	unequal, low	0.29	0.34	0.37	0.15
0.0	Weib shape	equal	0.34	0.34	0.34	0.53
0.0	Weib shape	unequal, high	0.49	0.56	0.57	0.57
0.0	Weib shape	unequal, low	0.29	0.34	0.37	0.31
1.5	exp early	equal	0.21	0.21	0.21	0.12
1.5	exp early	unequal, high	0.38	0.44	0.38	0.22
1.5	exp early	unequal, low	0.20	0.21	0.25	0.03
1.5	exp late	equal	0.21	0.21	0.21	0.05
1.5	exp late	unequal, high	0.38	0.44	0.38	0.23
1.5	exp late	unequal, low	0.20	0.21	0.25	0.01
1.5	exp prop	equal	0.20	0.21	0.21	0.08
1.5	exp prop	unequal, high	0.38	0.44	0.39	0.23
1.5	exp prop	unequal, low	0.20	0.21	0.25	0.02
1.5	logn	equal	0.41	0.41	0.41	0.24
1.5	logn	unequal, high	0.51	0.58	0.60	0.43
1.5	logn	unequal, low	0.33	0.41	0.41	0.05
1.5	pwExp	equal	0.21	0.21	0.21	0.17
1.5	pwExp	unequal, high	0.38	0.44	0.38	0.25
1.5	pwExp	unequal, low	0.20	0.21	0.25	0.11
1.5	Weib late	equal	0.34	0.34	0.34	0.17
1.5	Weib late	unequal, high	0.49	0.56	0.57	0.41
1.5	Weib late	unequal, low	0.29	0.34	0.37	0.02
1.5	Weib prop	equal	0.34	0.34	0.34	0.19
1.5	Weib prop	unequal, high	0.49	0.56	0.57	0.41
1.5	Weib prop	unequal, low	0.29	0.34	0.37	0.03
1.5	Weib scale	equal	0.34	0.34	0.34	0.26
1.5	Weib scale	unequal, high	0.49	0.56	0.57	0.41
1.5	Weib scale	unequal, low	0.29	0.34	0.37	0.06
1.5	Weib shape	equal	0.34	0.34	0.34	0.44
1.5	Weib shape	unequal, high	0.49	0.56	0.57	0.48
1.5	Weib shape	unequal, low	0.29	0.34	0.37	0.34

Table 4: Censoring rates for the different settings.

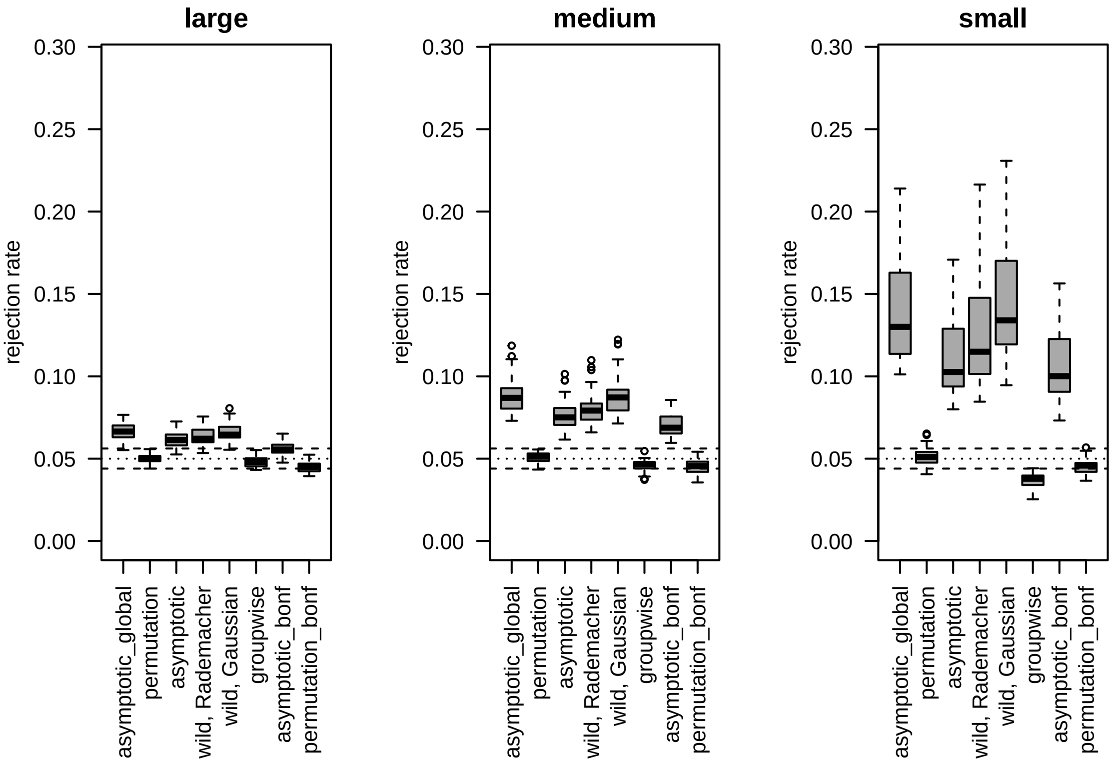


Figure 15: Rejection rates over all settings under the null hypothesis for the Dunnett-type contrast matrix. The dashed lines represent the borders of the binomial confidence interval $[0.044, 0.0562]$.

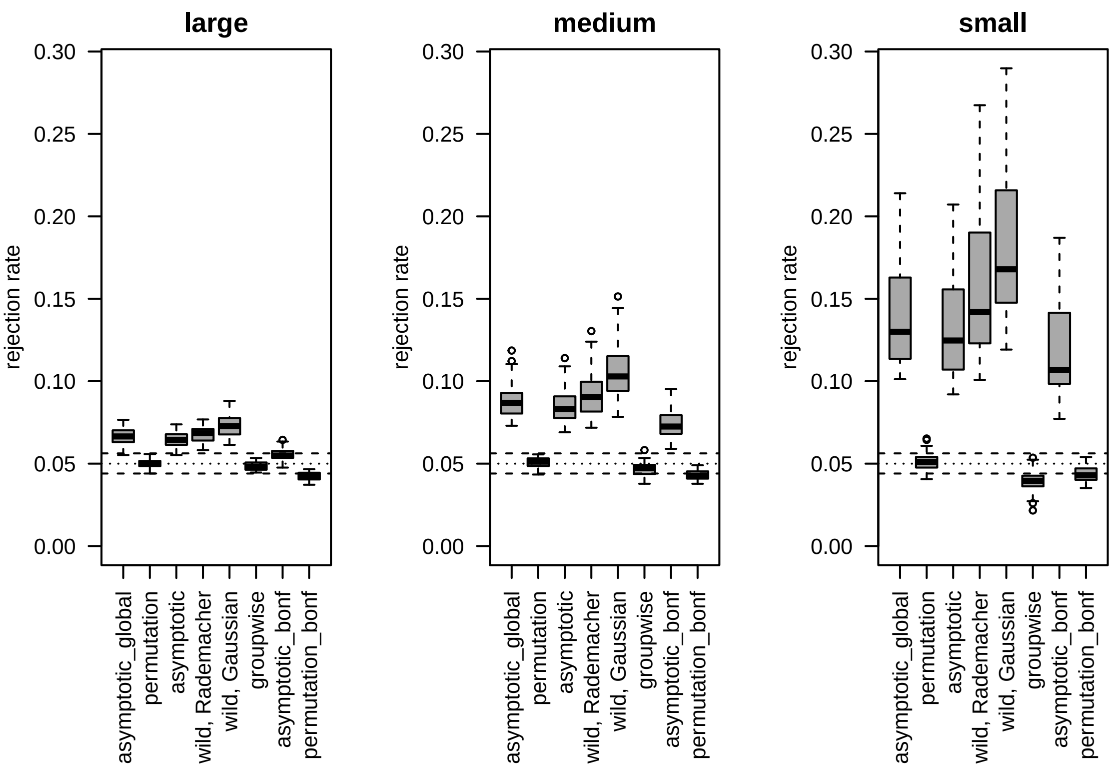


Figure 16: Rejection rates over all settings under the null hypothesis for the Tukey-type contrast matrix. The dashed lines represent the borders of the binomial confidence interval $[0.044, 0.0562]$.

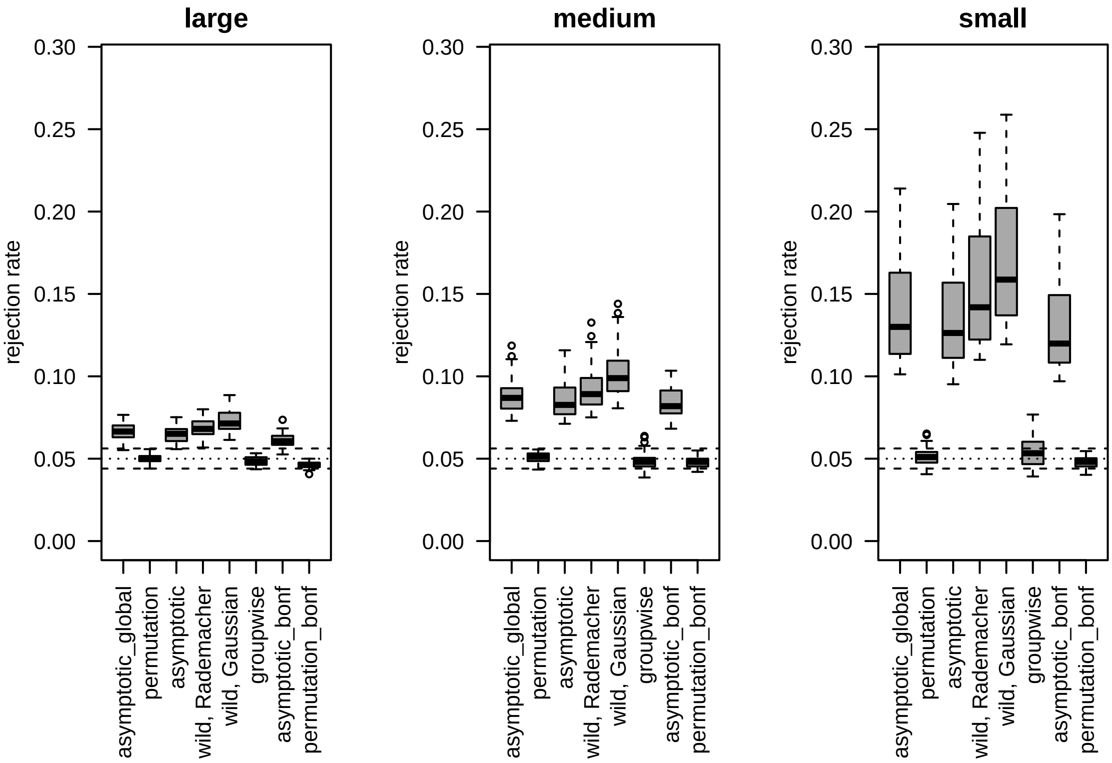


Figure 17: Rejection rates over all settings under the null hypothesis for the Grand-mean-type contrast matrix. The dashed lines represent the borders of the binomial confidence interval $[0.044, 0.0562]$.

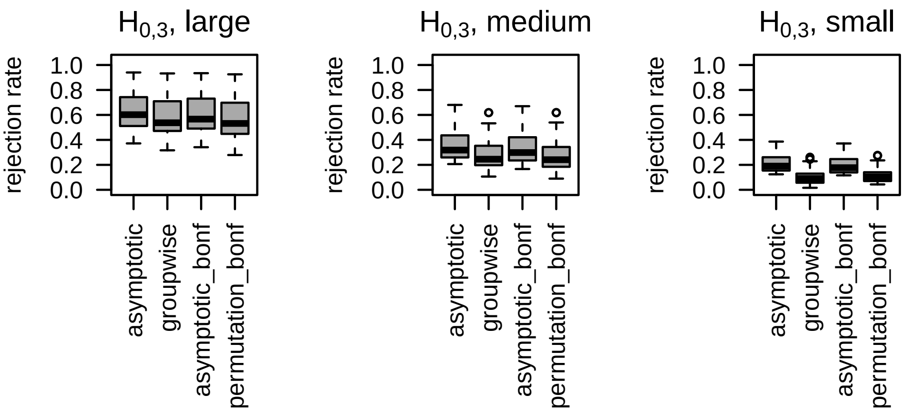


Figure 18: Rejection rates of the false local hypothesis over all settings under the alternative hypothesis for the Dunnett-type contrast matrix.

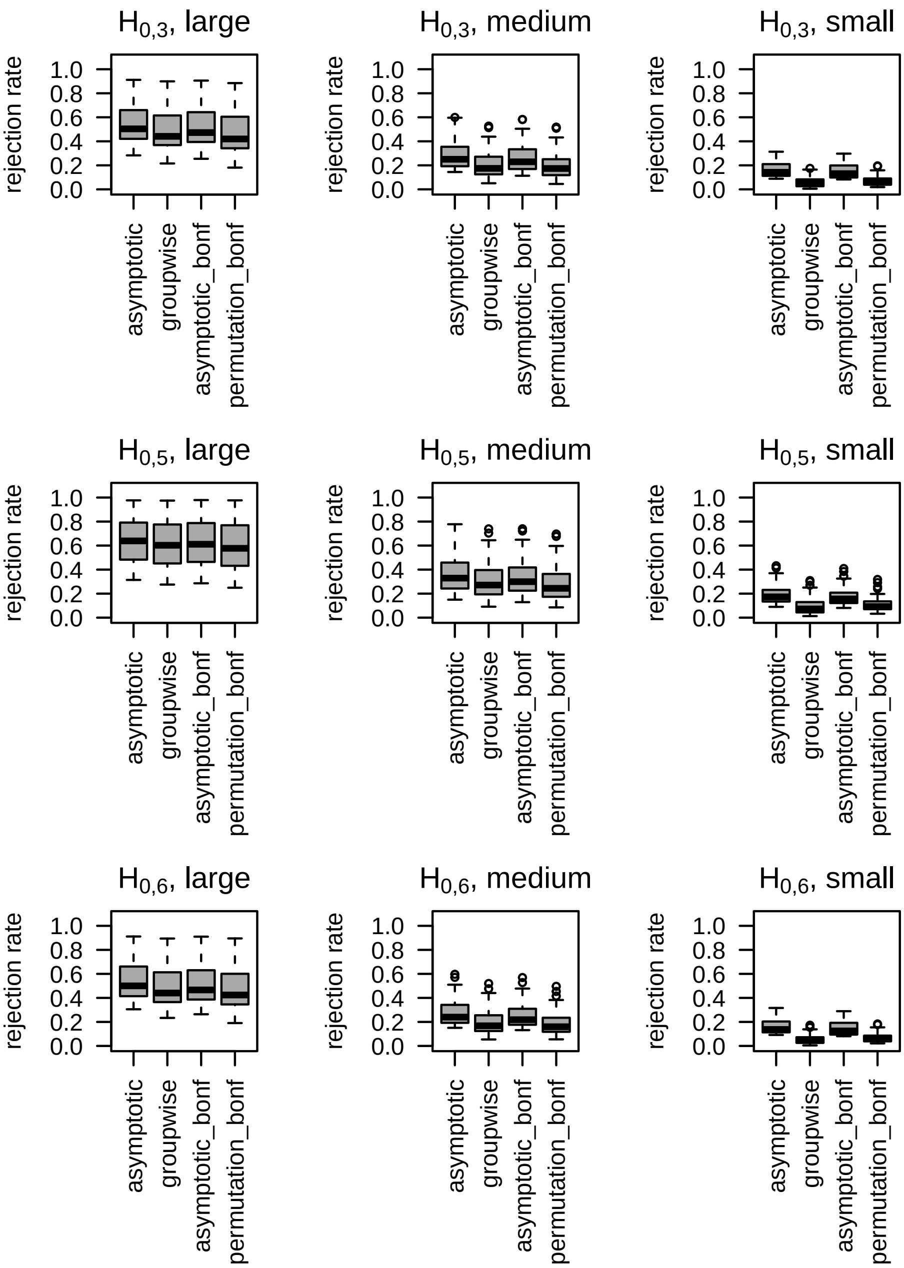


Figure 19: Rejection rates of all false local hypotheses over all settings under the alternative hypothesis for the Tukey-type contrast matrix.

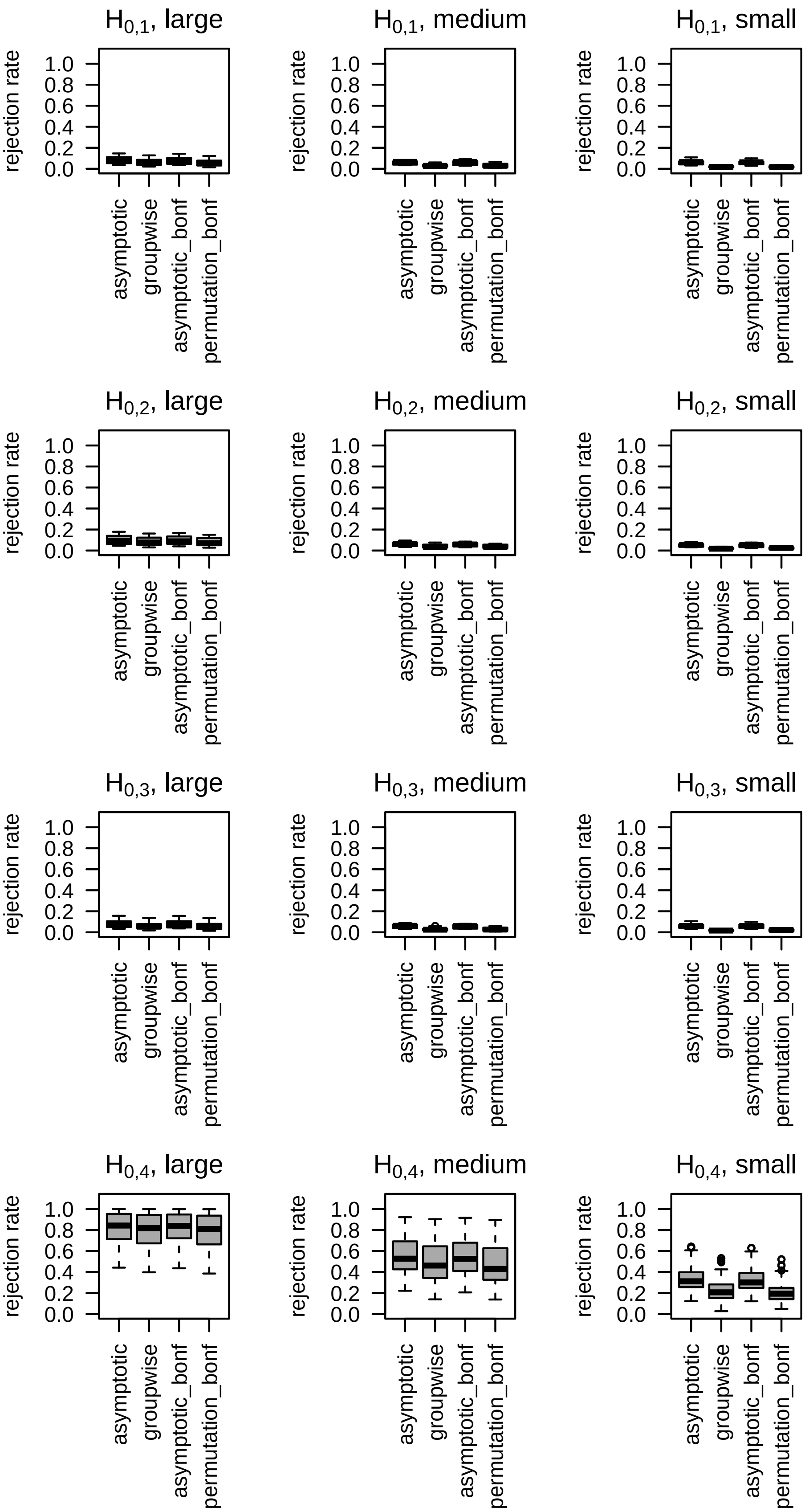


Figure 20: Rejection rates of the false local hypothesis over all settings under the alternative hypothesis for the Grand-mean-type contrast matrix.

C.2 Simulations for Analyzing the Asymptotic Behaviour

We have seen in Section 4.4.2 that the three asymptotic approaches (*asymptotic_global, asymptotic, asymptotic_bonf*) do not lead to a good type I error control. Thus, one may be interested in the sample sizes needed to obtain a good control of the type I error for these naive methods. Therefore, in this section we consider the simulation setup from Section 4.4.1 again with an increased factor for the scaling of the sample sizes, that is $K \in \{6, 8, 10\}$, resulting in sample sizes from 60 up to 200 in the groups. Furthermore, only the three asymptotic approaches (*asymptotic_global, asymptotic, asymptotic_bonf*) are considered under the null hypothesis. The performance of these methods regarding the power was already quite good for small and medium sample sizes, see Section 4.4.3 for details. This is why we did not analyze the power for larger sample sizes. Note that the censoring rates for the different scenarios are as shown in Table C.1.
In Figures 21 to 23, the rejection rates across all settings are illustrated for the three different contrast matrices. It can be seen that the empirical type I error rates are quite close to the desired level of significance of 0.05 for large sample sizes in all scenarios. Moreover, the rejection rates seem to tend more and more to 0.05 as the sample sizes increase. However, the difference between the rejection rates for different values of $K \in \{6, 8, 10\}$ is rather small, indicating that the convergence is relatively slow.
It can be observed that quite large sample sizes are needed to obtain a good type I error control for the multiple asymptotic and the global asymptotic test without Bonferroni-correction. Even for $K = 10$, i.e. sample sizes between 100 and 200 in each group, these tests are still slightly liberal. The empirical type I error rates for the multiple asymptotic and the global asymptotic test without Bonferroni-correction reach up to 0.0702.
By using a Bonferroni-correction, the asymptotic test does not need very large sample sizes to control the level of significance. Here, $K = 6$, i.e. sample sizes between 60 and 120 in each group, or even $K = 4$ seems to be enough as can be seen in Figures 15 to 17.

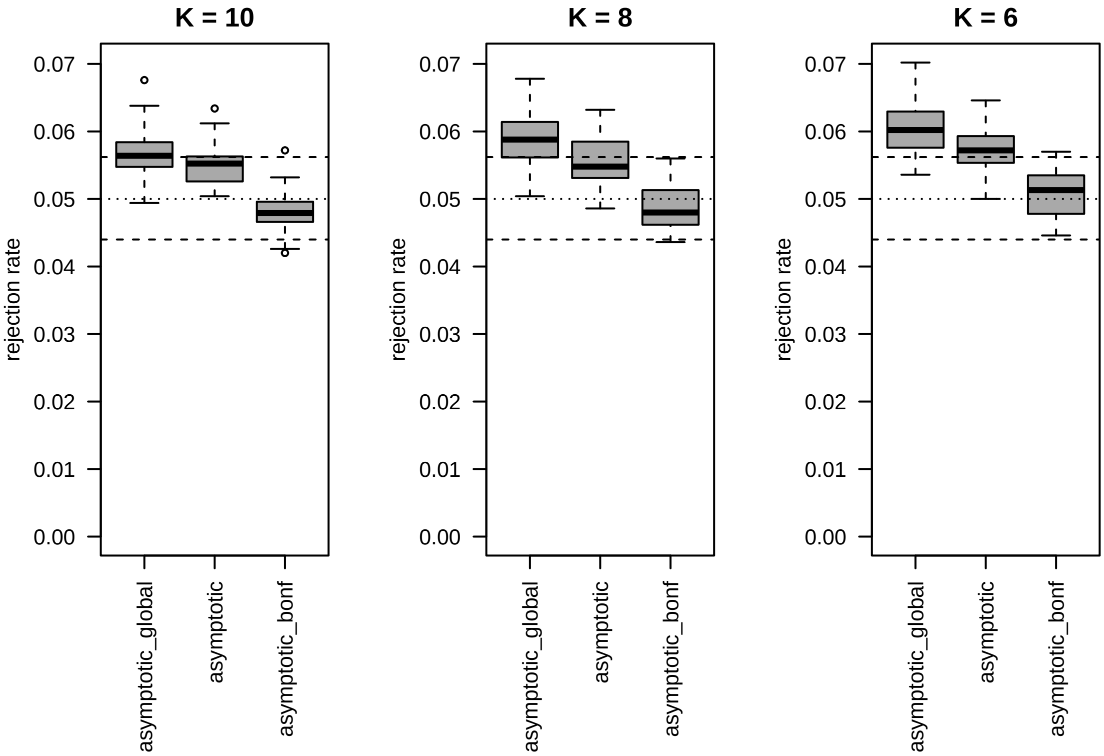


Figure 21: Rejection rates over all settings under the null hypothesis for the Dunnett-type contrast matrix. The dashed lines represent the borders of the binomial confidence interval $[0.044, 0.0562]$.

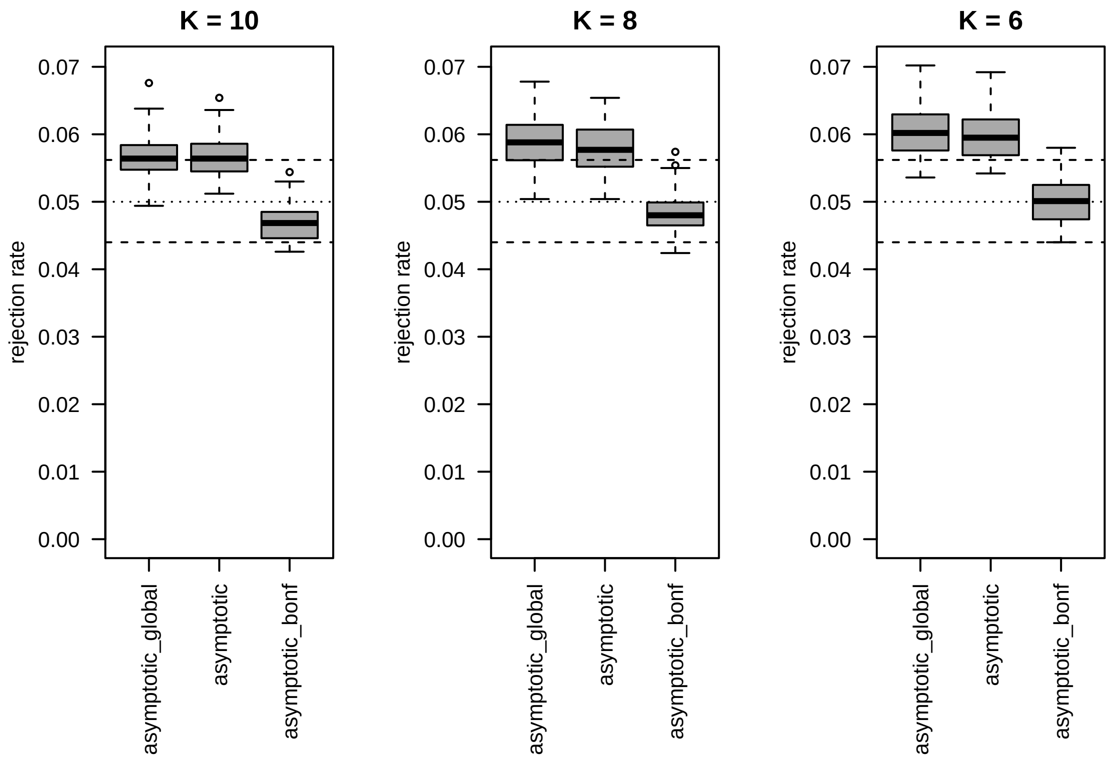


Figure 22: Rejection rates over all settings under the null hypothesis for the Tukey-type contrast matrix. The dashed lines represent the borders of the binomial confidence interval $[0.044, 0.0562]$.

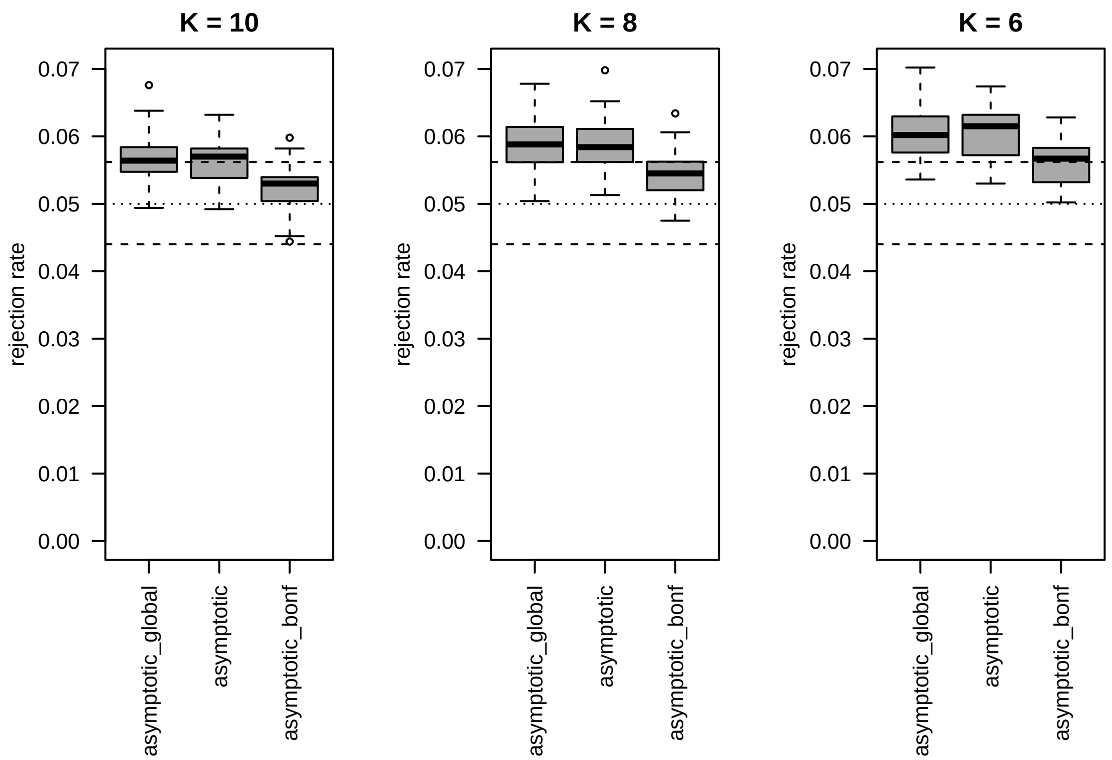


Figure 23: Rejection rates over all settings under the null hypothesis for the Grand-mean-type contrast matrix. The dashed lines represent the borders of the binomial confidence interval $[0.044, 0.0562]$.

C.3 Additional Simulations under Non-Exchangeability

In Section 4.4, the empirical power of the groupwise approach and the permutation approach with Bonferroni-correction seems comparable over all simulation setups. However, the Bonferroni-correction is known to have low power for a large number of hypotheses. Thus, we aim to motivate that the groupwise bootstrap approach may perform better than the permutation approach with Bonferroni-correction in specific setups in this section. Therefore, we consider again $k = 4$ groups with sample sizes $\mathbf{n} = (40, 80, 40, 80)$, hypotheses matrices as in Section 4.4 and $\alpha = 0.05$. Furthermore, we generated $N_{sim} = 5000$ simulation runs with $B = 1999$ resampling iterations. In contrast to the simulation study in Section 4.4, the survival times are drawn from different distributions for all groups as follows:

- Different piece-wise exponential distributions (*pwExp diff*): $T_{11} \sim Exp(0.2)$,
  T_{21} with hazard function $t \mapsto 0.3 \cdot \mathbb{1}\{t \leqslant \lambda_{10}\} + 0.1 \cdot \mathbb{1}\{t > \lambda_{10}\}$,
  T_{31} with hazard function $t \mapsto 1.5 \cdot \mathbb{1}\{t \leqslant \lambda_{11}\} + 0.01 \cdot \mathbb{1}\{t > \lambda_{11}\}$ and
  T_{41} with hazard function $t \mapsto 0.5 \cdot \mathbb{1}\{t \leqslant \lambda_{\delta,5}\} + 0.05 \cdot \mathbb{1}\{t > \lambda_{\delta,5}\}$,
- Different Weibull distributions (*Weib diff*): $T_{11} \sim Weib(3, 8)$, $T_{21} \sim Weib(1.5, \lambda_{0,8})$, $T_{31} \sim Weib(\lambda_{0,9}, 14)$ and $T_{41} \sim Weib(\lambda_{\delta,9}, 14)$.

Here, the parameters λ_{10} and λ_{11} are determined such that the RMST equals μ_1. Hence, note that only μ_4 differs under the alternative hypothesis but the distributions of the survival times differ across the groups under the null and alternative hypothesis. In Figure 24, the different survival functions are illustrated. For the censoring times, the same distributions as in Section 4.4 are considered, i.e. *equal; unequal, high* and *unequal, low.* The resulting censoring rates can be found in Table 5 and reach from 11 up to 62%.
In Figure 25, the rejection rates over all settings under the null hypothesis are presented. Here, it is observable that the groupwise bootstrap and the permutation approach with Bonferroni-correction perform well in terms of type I error control for the multiple testing problem. The permutation approach with Bonferroni-correction tends to be too conservative for the Tukey-type contrast matrix. Furthermore, the asymptotic approaches and the wild bootstrap are too liberal and, thus, they do not seem to control the family-wise type I error. However, the empirical power of the groupwise bootstrap is slightly higher than of the permutation approach with Bonferroni-correction in most of the scenarios which can be seen in Table 6 to 8. Only for hypothesis $\mathcal{H}_{0,3}$ for the Grand-mean-type contrast matrix, the permutation approach with Bonferroni-correction has a higher power than the groupwise bootstrap in some scenarios. The empirical powers of the false hypotheses are also illustrated in Figure 26 to 28, where it is observable that the groupwise bootstrap tends to have a higher empirical power than the permutation approach with Bonferroni-correction, particularly in Figure 27. The asymptotic approaches even have higher empirical powers in several scenarios but, however, they can not control the family-wise error adequately which can be seen in Figure 25.

δ	distribution	censoring distribution	group 1	group 2	group 3	group 4
0.0	pwExp diff	equal	0.21	0.27	0.40	0.33
0.0	pwExp diff	unequal, high	0.38	0.44	0.42	0.37
0.0	pwExp diff	unequal, low	0.20	0.27	0.39	0.22
0.0	Weib diff	equal	0.34	0.45	0.53	0.53
0.0	Weib diff	unequal, high	0.49	0.57	0.62	0.57
0.0	Weib diff	unequal, low	0.29	0.45	0.48	0.31
1.5	pwExp diff	equal	0.21	0.27	0.40	0.16
1.5	pwExp diff	unequal, high	0.38	0.44	0.42	0.25
1.5	pwExp diff	unequal, low	0.20	0.27	0.39	0.11
1.5	Weib diff	equal	0.34	0.45	0.53	0.44
1.5	Weib diff	unequal, high	0.49	0.57	0.62	0.48
1.5	Weib diff	unequal, low	0.29	0.45	0.48	0.34

Table 5: Censoring rates for the additional simulation

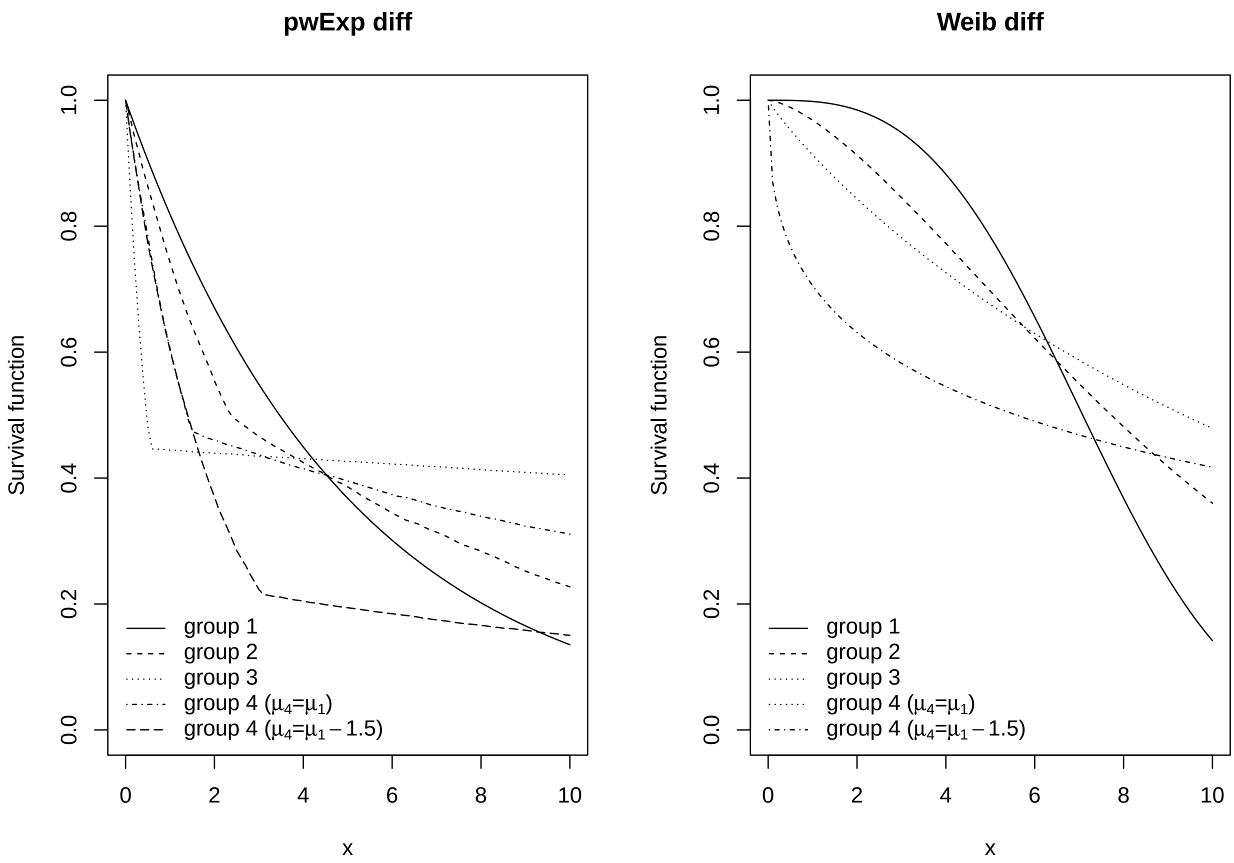


Figure 24: The survival functions of the two settings under the null hypothesis as well as under the alternative $\mu_4 = \mu_1 - 1.5$. Note that the survival functions of group 3 and 4 coincide under the null hypothesis for the setting *Weib diff*.

hypothesis	distribution	censoring distribution	asymptotic	groupwise	asymptotic bonf	permutation bonf
$\mathcal{H}_{0,3}$	pwExp diff	equal	0.495	0.461	0.474	0.437
	pwExp diff	unequal, high	0.376	0.324	0.357	0.299
	pwExp diff	unequal, low	0.474	0.435	0.451	0.412
	Weib diff	equal	0.507	0.479	0.495	0.455
	Weib diff	unequal, high	0.405	0.368	0.393	0.345
	Weib diff	unequal, low	0.512	0.477	0.497	0.458

Table 6: Rejection rates of the false hypothesis for the Dunnett-type contrast matrix with $\delta = 1.5$

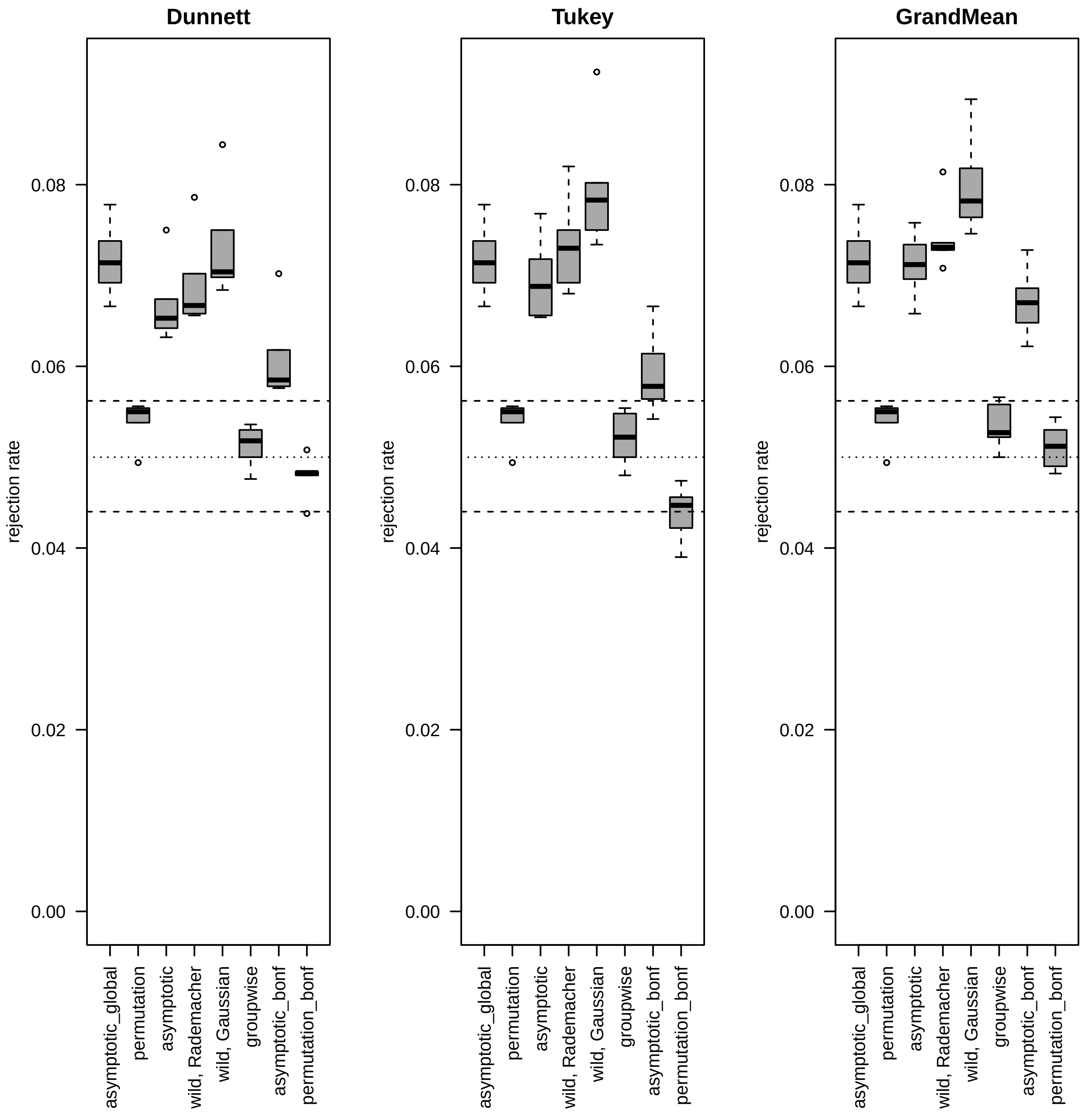


Figure 25: Rejection rates over all settings under the null hypothesis. The dashed lines represent the borders of the binomial confidence interval [0.044, 0.0562].

hypothesis	distribution	censoring distribution	asymptotic	groupwise	asymptotic bonf	permutation bonf
$\mathcal{H}_{0,3}$	pwExp diff	equal	0.414	0.370	0.378	0.331
	pwExp diff	unequal, high	0.298	0.231	0.270	0.201
	pwExp diff	unequal, low	0.392	0.345	0.356	0.306
	Weib diff	equal	0.430	0.397	0.399	0.350
	Weib diff	unequal, high	0.331	0.288	0.308	0.245
	Weib diff	unequal, low	0.436	0.408	0.409	0.358
$\mathcal{H}_{0,5}$	pwExp diff	equal	0.532	0.513	0.503	0.481
	pwExp diff	unequal, high	0.382	0.350	0.352	0.325
	pwExp diff	unequal, low	0.530	0.513	0.500	0.478
	Weib diff	equal	0.444	0.423	0.412	0.391
	Weib diff	unequal, high	0.348	0.319	0.318	0.288
	Weib diff	unequal, low	0.460	0.440	0.428	0.407
$\mathcal{H}_{0,6}$	pwExp diff	equal	0.215	0.174	0.190	0.156
	pwExp diff	unequal, high	0.198	0.159	0.177	0.127
	pwExp diff	unequal, low	0.208	0.169	0.185	0.149
	Weib diff	equal	0.290	0.256	0.262	0.226
	Weib diff	unequal, high	0.231	0.188	0.211	0.171
	Weib diff	unequal, low	0.277	0.235	0.252	0.209

Table 7: Rejection rates of the false hypotheses for the Tukey-type contrast matrix with $\delta = 1.5$

hypothesis	distribution	censoring distribution	asymptotic	groupwise	asymptotic bonf	permutation bonf
$\mathcal{H}_{0,1}$	pwExp diff	equal	0.047	0.032	0.044	0.026
	pwExp diff	unequal, high	0.037	0.024	0.036	0.017
	pwExp diff	unequal, low	0.041	0.031	0.039	0.025
	Weib diff	equal	0.087	0.073	0.081	0.060
	Weib diff	unequal, high	0.070	0.050	0.065	0.042
	Weib diff	unequal, low	0.078	0.062	0.075	0.053
$\mathcal{H}_{0,2}$	pwExp diff	equal	0.071	0.065	0.067	0.059
	pwExp diff	unequal, high	0.048	0.038	0.045	0.038
	pwExp diff	unequal, low	0.071	0.063	0.067	0.058
	Weib diff	equal	0.094	0.085	0.090	0.079
	Weib diff	unequal, high	0.072	0.060	0.070	0.058
	Weib diff	unequal, low	0.094	0.086	0.090	0.082
$\mathcal{H}_{0,3}$	pwExp diff	equal	0.035	0.024	0.033	0.021
	pwExp diff	unequal, high	0.038	0.025	0.036	0.018
	pwExp diff	unequal, low	0.038	0.025	0.035	0.022
	Weib diff	equal	0.071	0.046	0.068	0.051
	Weib diff	unequal, high	0.062	0.032	0.059	0.039
	Weib diff	unequal, low	0.069	0.039	0.064	0.048
$\mathcal{H}_{0,4}$	pwExp diff	equal	0.678	0.653	0.667	0.642
	pwExp diff	unequal, high	0.590	0.556	0.580	0.540
	pwExp diff	unequal, low	0.674	0.650	0.662	0.636
	Weib diff	equal	0.585	0.558	0.570	0.543
	Weib diff	unequal, high	0.492	0.460	0.479	0.442
	Weib diff	unequal, low	0.590	0.560	0.579	0.553

Table 8: Rejection rates of the false hypotheses for the Grand-mean-type contrast matrix with $\delta = 1.5$

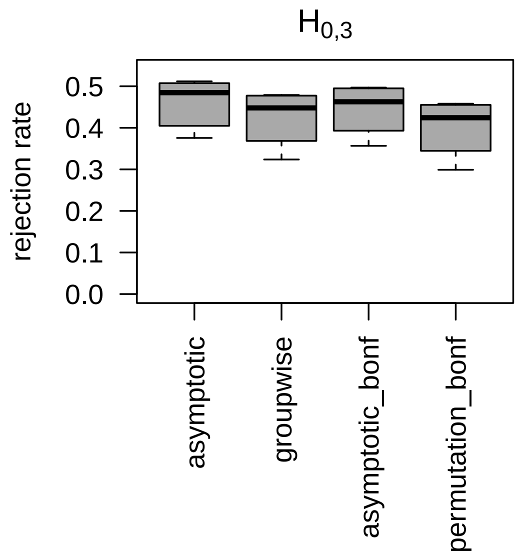


Figure 26: Rejection rates of the false local hypothesis over all settings under the alternative hypothesis for the Dunnett-type contrast matrix.

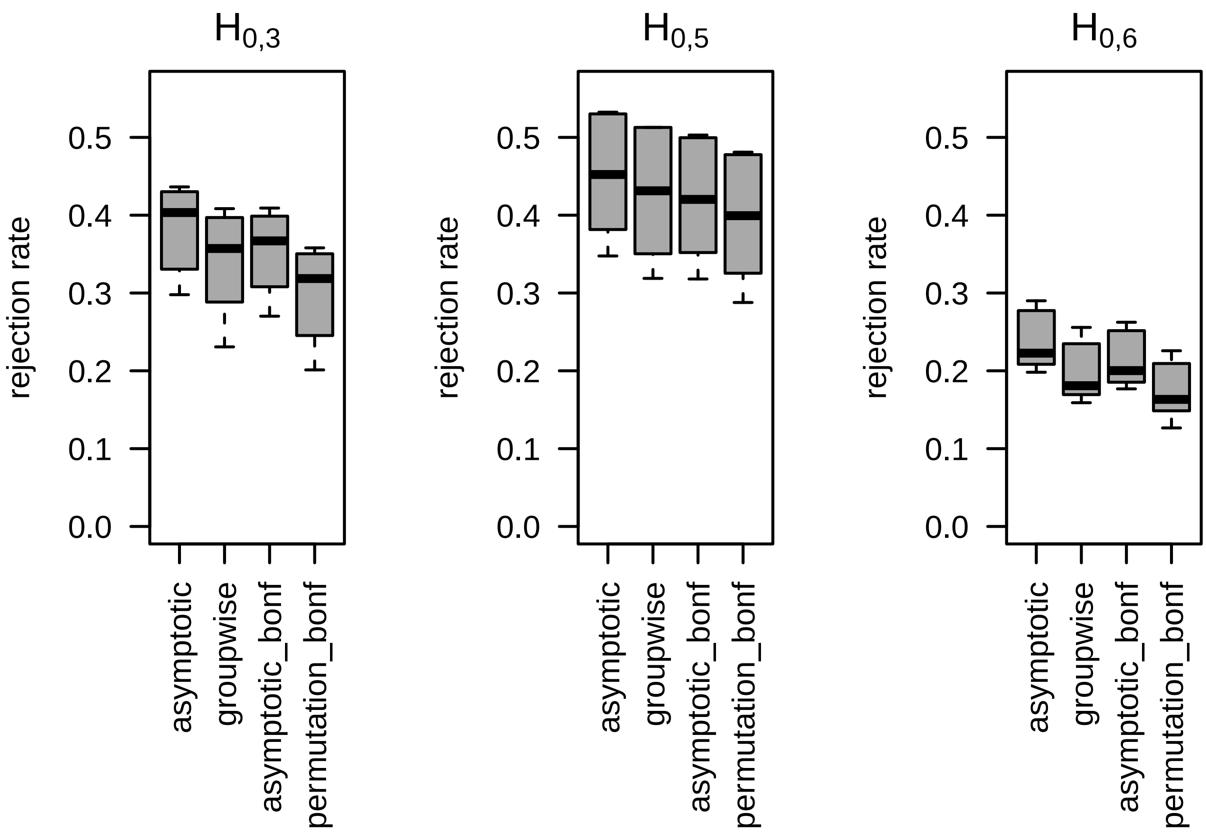


Figure 27: Rejection rates of all false local hypotheses over all settings under the alternative hypothesis for the Tukey-type contrast matrix.

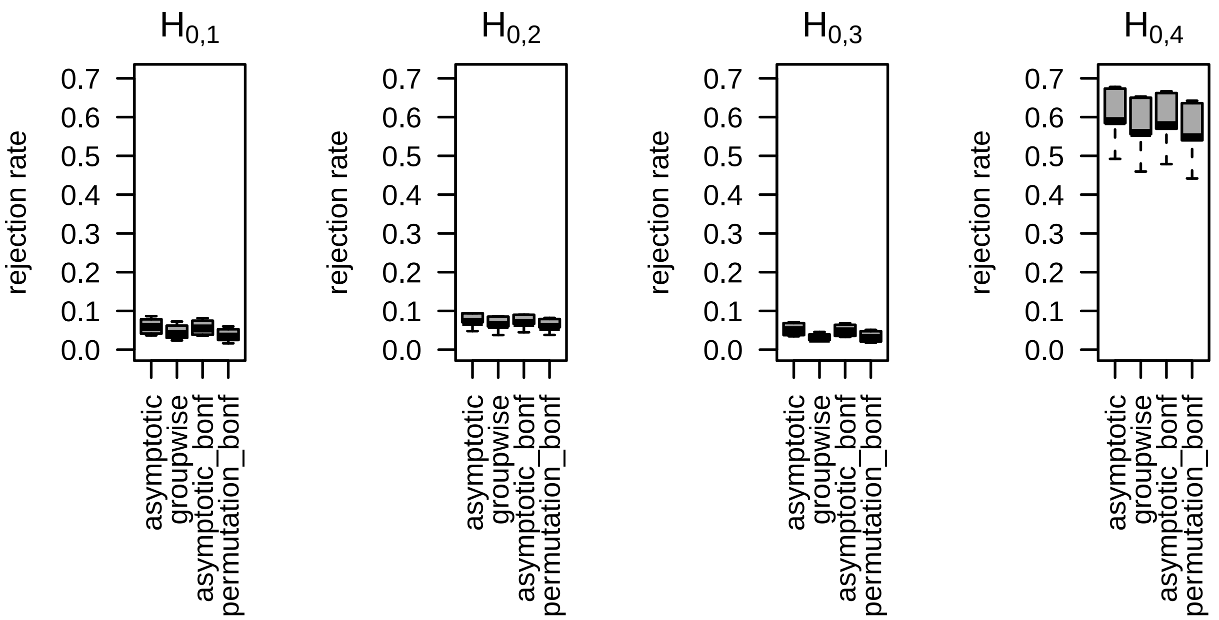


Figure 28: Rejection rates of the false local hypothesis over all settings under the alternative hypothesis for the Grand-mean-type contrast matrix.

C.4 Simulation inspired by the Data Example

Since the Simulation study in Section 4.4 does not fit perfectly to the data example about the occurrence of hay fever in Section 4.5, we also considered a small simulation setup inspired by the data example. Therefore, we considered $k = 4$ groups with sample sizes $\mathbf{n} = (450, 481, 654, 649)$, the hypotheses matrices as in Section 4.5, i.e. $\mathbf{H} := [\mathbf{H}'_A, \mathbf{H}'_B, \mathbf{H}'_{AB}]'$, and $\alpha = 0.05$. Moreover, $N_{sim} = 5000$ simulation runs with $B = 19999$ resampling iterations were generated. The survival times of group i were simulated from a distribution with the Kaplan-Meier estimator of the pooled sample under the null and of the ith sample under the alternative hypothesis as distribution function. Analogously, the Kaplan-Meier estimators for the censoring times of the different samples are used for the data generation of the censoring times. Proceeding as described leads to a censoring rate of 82% in all groups under the null hypothesis and censoring rates from 72% up to 89% under the alternative hypothesis.
In Table 9, the resulting rejection rates are shown. It is observable that all methods seem to control the global level of significance of 5% quite accurately under the given scenario. Furthermore, all methods have a quite high empirical power under the alternative hypothesis. The power of the global approaches is around 90% while all methods for the multiple testing problem have a power of 100%. In Figure 29, it is shown how the rejection rates of the multiple testing procedures result from the local decisions. Here, it can be seen that the methods detect both of the main effects simultaneously in around 70% of the simulation runs, only the main effect of factor A in 20% and all main and interaction effects in 6% under the alternative hypothesis. Furthermore, the methods seem to yield very similar local test decisions.

	asymptotic global	permutation	asymptotic	wild Rademacher	wild Gaussian	groupwise	asymptotic bonf	permutation bonf
$\mathcal{H}_0$	0.0568	0.0566	0.0506	0.0514	0.0522	0.0516	0.0506	0.0498
$\mathcal{H}_1$	0.8984	0.8984	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 9: Rejection rates for the simulation inspired by the data example under the null ($\mathcal{H}_0$) and alternative ($\mathcal{H}_1$) hypothesis.

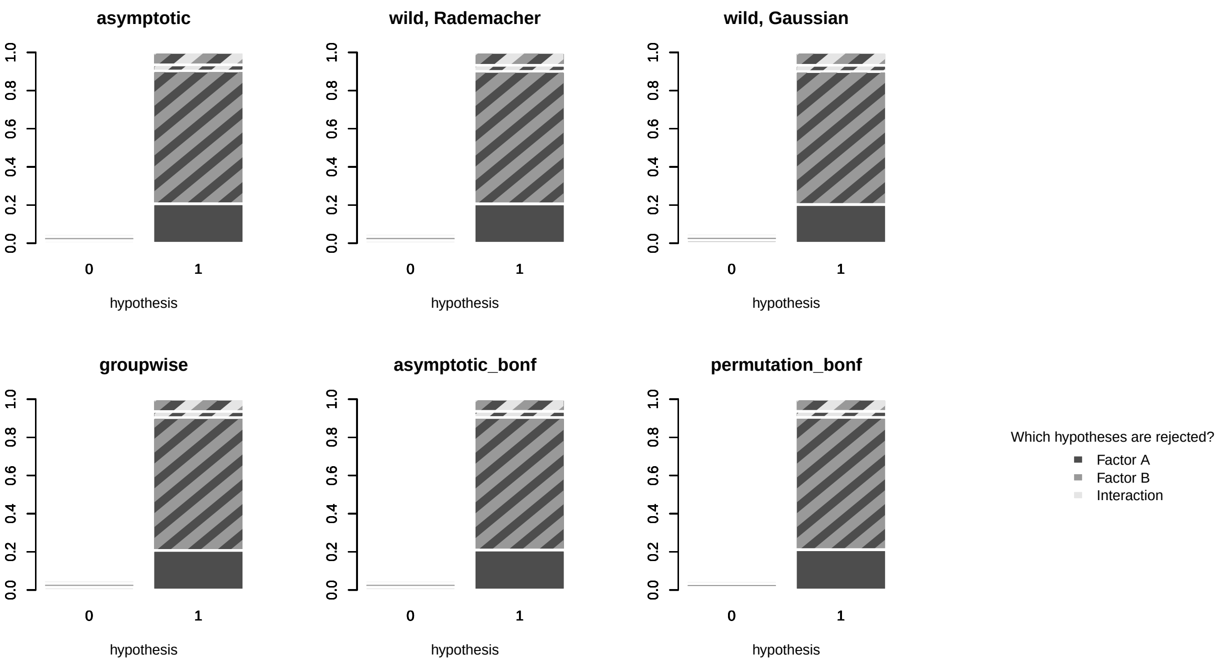


Figure 29: Rejection rates for the simulation inspired by the data example under the null (0) and under the alternative (1) hypothesis. The heights of the bars represent the rates of the rejections caused by the corresponding hypotheses. Two- and Three-colored bars indicate that the corresponding two or three hypotheses are rejected simultaneously. The overall height represents the rate of global rejections.

D Additional Figures of the Simulation Results in Section 5.4

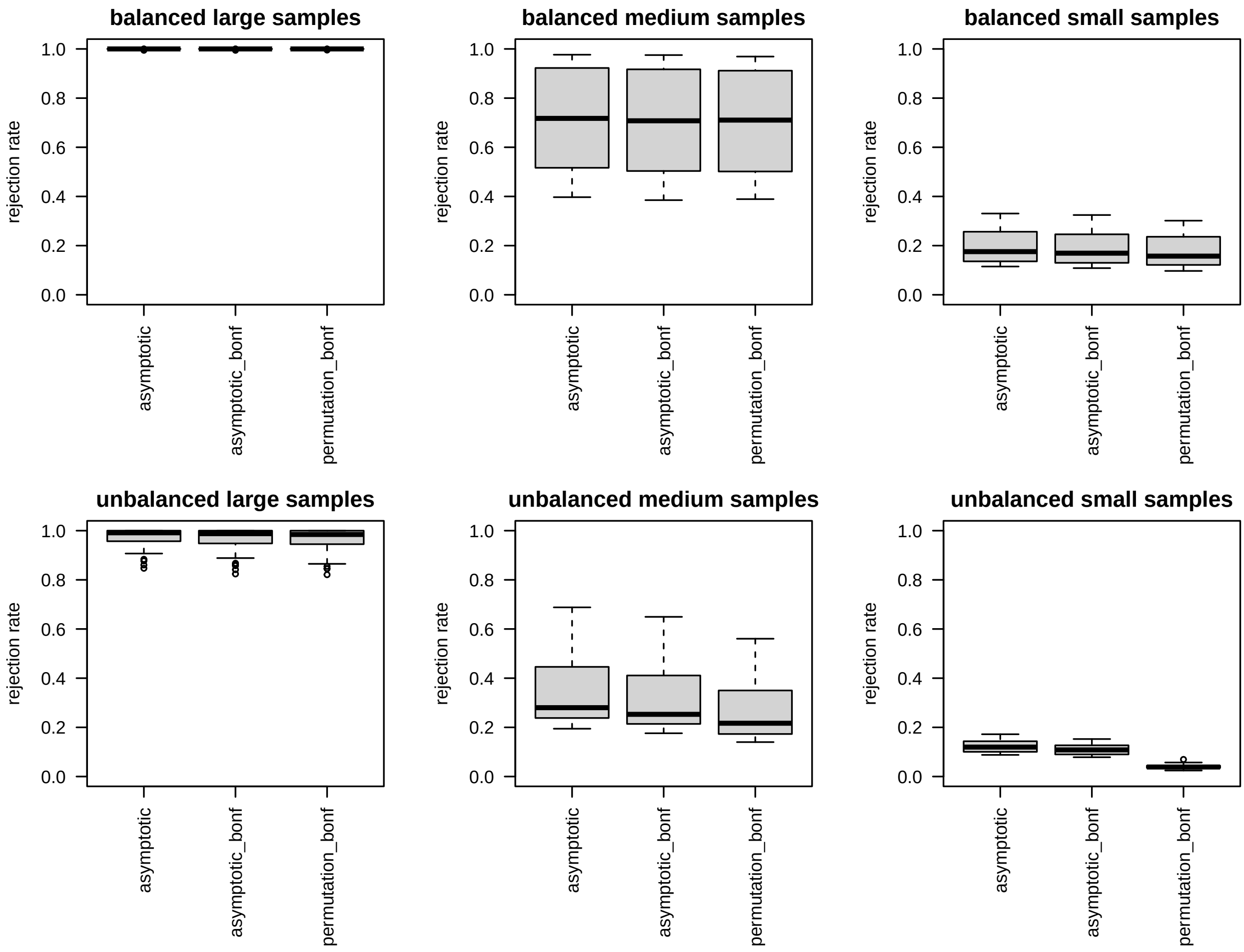


Figure 30: Empirical powers for the 2-by-2 design across all scenarios under the alternative hypothesis.

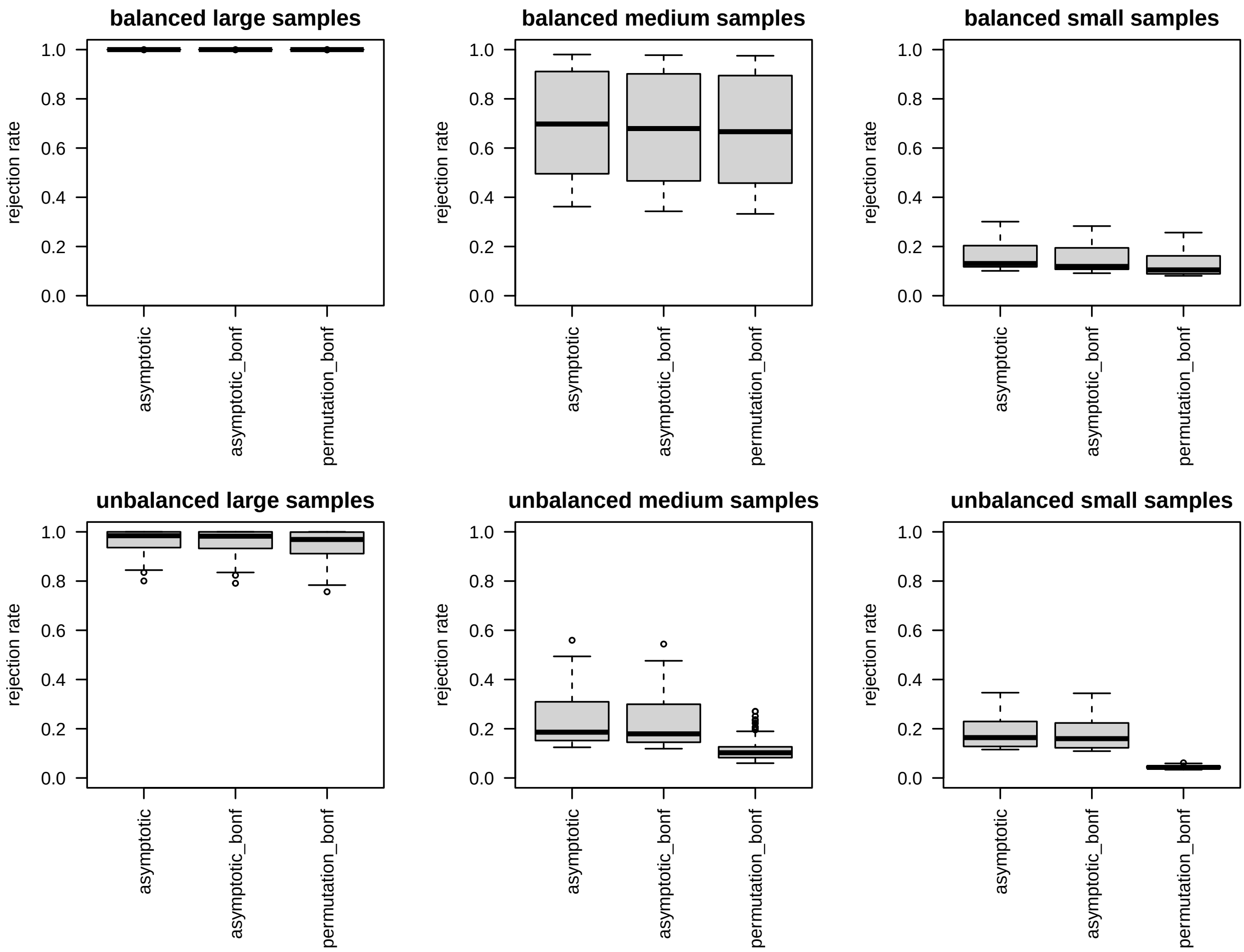


Figure 31: Empirical powers for the Dunnett-type contrast hypotheses across all scenarios under the alternative hypothesis.

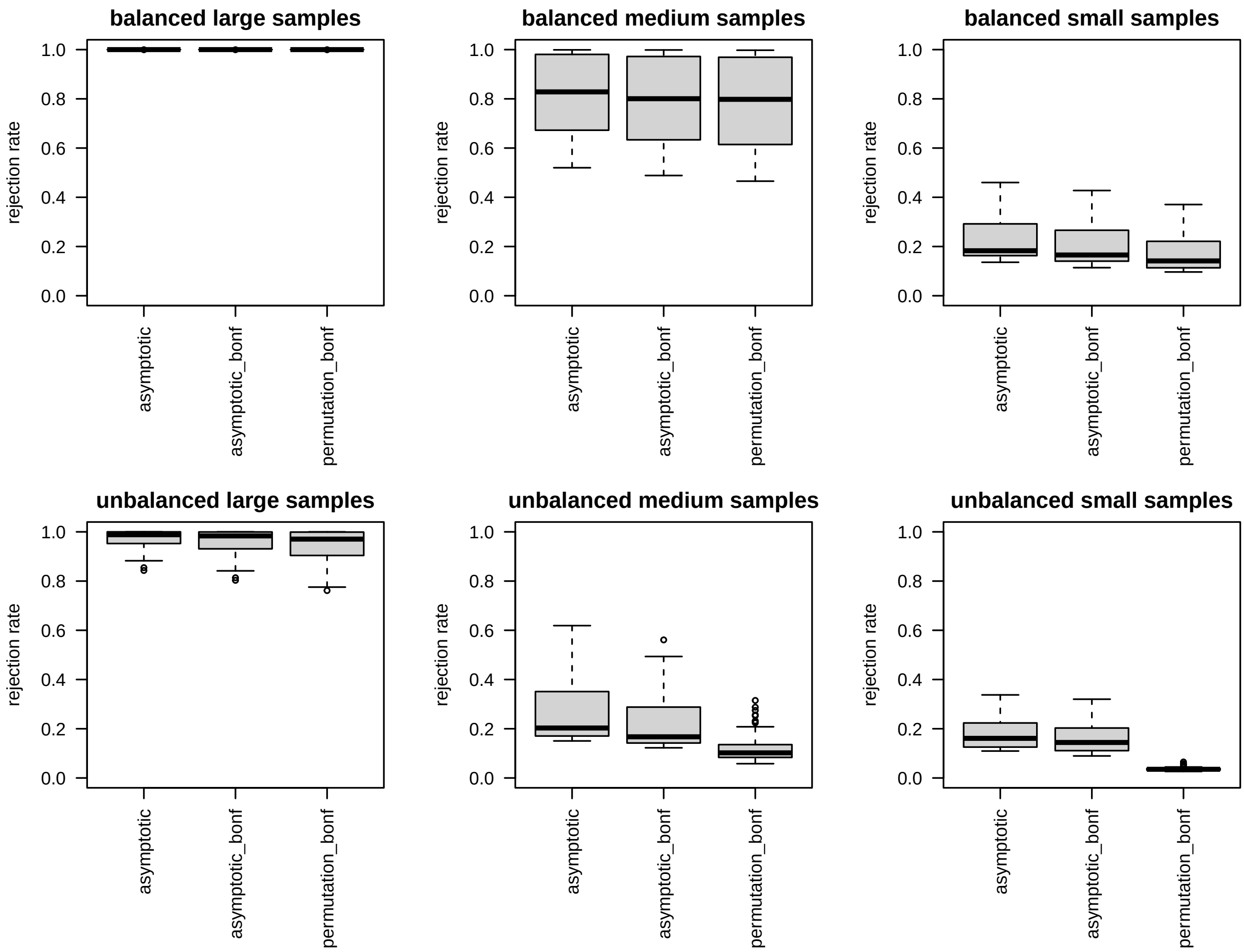


Figure 32: Empirical powers for the Tukey-type contrast hypotheses across all scenarios under the alternative hypothesis.

E R Packages: `GFDrmst` and `GFDrmtl`

As part of this work, two R packages were published on CRAN, that are `GFDrmst` [20] and `GFDrmtl` [21]. These packages contain the implementations of the methodology described in Sections 4 and 5, respectively. In this section, we describe the included functions and their usage in detail.

E.1 `GFDrmst`

The R package `GFDrmst` exports the following four functions:

- `RMST.test` to perform the multiple RMST-based tests as described in Section 4. The output of this function is a list of class `GFDrmst`.
- `summary.GFDrmst` to summarize an object of class `GFDrmst`.
- `plot.GFDrmst` to plot simultaneous confidence intervals for an object of class `GFDrmst`.
- `GFDrmstGUI` to perform the multiple RMST-based tests as described in Section 4 on an interactive user interface.

In the following, we explain their arguments and usage.

`RMST.test` This function contains the implementation of the asymptotic multiple tests of Section 4.3.1, the multiple groupwise bootstrap tests of Section 4.3.2, and the adjusted permutation tests of Section 4.2.2. Furthermore, confidence intervals for RMST contrasts can be calculated as in Remark 2.6 and the stepwise extension of Remark 2.5, which can improve the power of the multiple tests, is available.
The function `RMST.test` has the following arguments.

- `time`
  A vector containing the observed event times X_{ij}.
- `status`
  A vector containing the corresponding censoring status indicator $\delta_{ij} = \mathbb{1}\{X_{ij} = T_{ij}\}$.
- `group`
  A vector containing the corresponding group labels $i \in \{1, ..., k\}$.
- `formula`
  A model `formula` object. The left hand side contains the time variable and the right hand side contains the factor variables of interest.
- `event`
  The name of censoring status indicator.
- `data`
  A data.frame or list containing the variables in `formula` and the censoring status indicator.
- `hyp_mat`
  A list containing all the hypothesis matrices $\mathbf{H}_1, ..., \mathbf{H}_L$ for the multiple tests or one of the options `"Tukey", "Dunnett", "center", "crossed factorial"` for the matrices described in Example 4.1 or a matrix $\mathbf{H}$ if only one hypothesis is of interest. The option `"crossed factorial"` is only available if `formula` is specified. For the permutation test, all matrices need to be contrast matrices.
- `hyp_vec`
  A list containing all the hypothesis vectors $\mathbf{c}_1, ..., \mathbf{c}_L$ for the multiple tests or a vector $\mathbf{c}$ if only one hypothesis is of interest. By default, all hypothesis vectors are set to zero vectors of suitable length.
- `tau`
  A numeric value $\tau > 0$ specifying the end of the relevant time window for the analysis.
- `method`
  One of the methods `"groupwise", "permutation"` and `"asymptotic"` that should be used for calculating the critical values, cf. Sections 4.3.2, 4.2.2, and 4.3.1, respectively. Default option is `"groupwise"`.
- `stepwise`
  A logical vector indicating whether the stepwise extension of Remark 2.5 should be performed. If `TRUE`, no confidence intervals can be computed but it may be that more tests can reject. Default option is `FALSE`.

- `alpha`
  A numeric value specifying the global level of significance α. Default option is `0.05`.

- `Nres`
  The number B of random variables to approximate the limiting distribution. The default option is `4999`.

- `seed`
  A single value, interpreted as an integer, for providing reproducibility of the results or `NULL` if reproducibility is not wanted. Default option is `1`.

The output of this function is a list of class `GFDrmst` containing the following components:

- `method`
  A character containing the method which has been used.

- `test_stat`
  A numeric vector containing the calculated Wald-type test statistics $W_n(\mathbf{H}_1, \mathbf{c}_1), ..., W_n(\mathbf{H}_L, \mathbf{c}_L)$ for the local hypotheses.

- `p.value`
  A numeric vector containing the adjusted p-values $p_1, ..., p_L$ as in Section 2.3.2 for the local hypotheses.

- `res`
  A list containing the results of the multiple tests including the hypothesis matrices, estimators of the linear combinations of RMSTs, potentially confidence intervals for the linear combinations (if all matrices are row vectors and `stepwise = FALSE`), Wald-type test statistics, critical values and the test decisions.

- `alpha`
  A numeric value containing the global level of significance α.

Note that either `time, status, group` or `formula, event, data` needs to be specified. The following example illustrates the usage of both versions.

```
> # load the package and the data
> library(GFDrmst)
> data(colonCS, package = "condSURV")
>
> # multiple asymptotic tests
> out <- RMST.test(formula = "Stime ~ rx",
+                  event = "event",
+                  data = colonCS,
+                  hyp_mat = "Tukey",
+                  tau = 3000,
+                  method = "asymptotic")
> ## or, equivalently,
> out <- RMST.test(time = colonCS$Stime,
+                  status = colonCS$event,
+                  group = colonCS$rx,
+                  hyp_mat = "Tukey",
+                  tau = 3000,
+                  method = "asymptotic")
```

The output looks like this:

```
> out
$method
[1] "Multiple asymptotic RMST Wald-type tests"

$test_stat
W_n(H_1, c_1) W_n(H_2, c_2) W_n(H_3, c_3)
  0.004007639   9.142532300   8.536680422

$p.value
[1] 0.997792822 0.007029102 0.009736379

$res
     hyp_matrix estimator lwr_conf  upr_conf test_stat   critical value adj_pvalue  decision
[1,] numeric,3  5.663569  -204.0409 215.3681 0.004007639 5.494443       0.9977928   "not significant"
[2,] numeric,3  267.2749  60.07632  474.4735 9.142532    5.494443       0.007029102 "H1"
[3,] numeric,3  261.6113  51.72996  471.4927 8.53668     5.494443       0.009736379 "H1"

$alpha
[1] 0.05
```

```
attr(,"class")
[1] "GFDrmst"
```

summary.GFDrmst This is the `summary` method for class `"GFDrmst"`. It takes the arguments

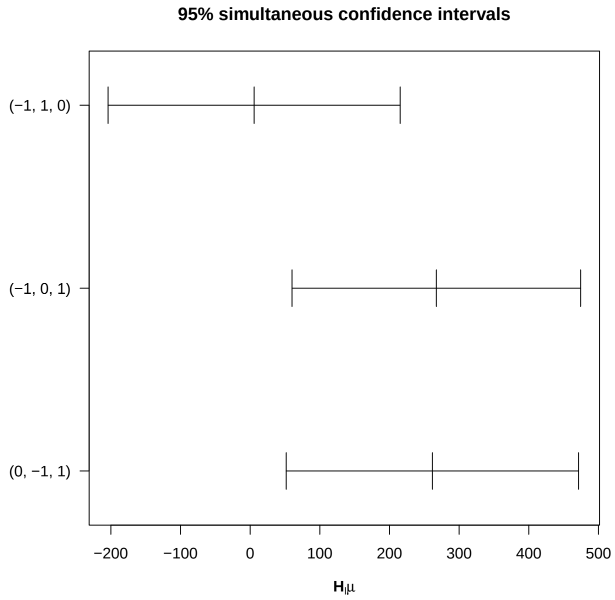
- `object`
  An object of class `"GFDrmst"`.
- `digits`
  An integer indicating the number of decimal places to be used. Default option is `8`.

All further arguments passed to this function are ignored. The function prints the information about the used method, significance level, hypothesis matrices, Wald-type test statistics, adjusted p-values and the overall results of the tests. It does not have a return value, but is called for side effects.
In our example, the summary looks like this:

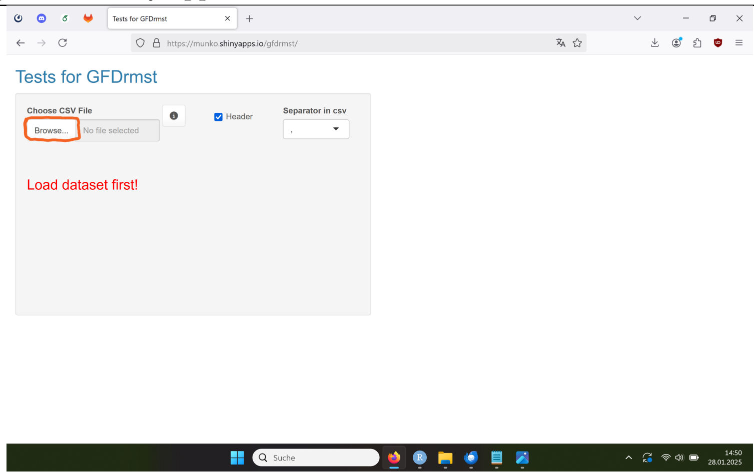
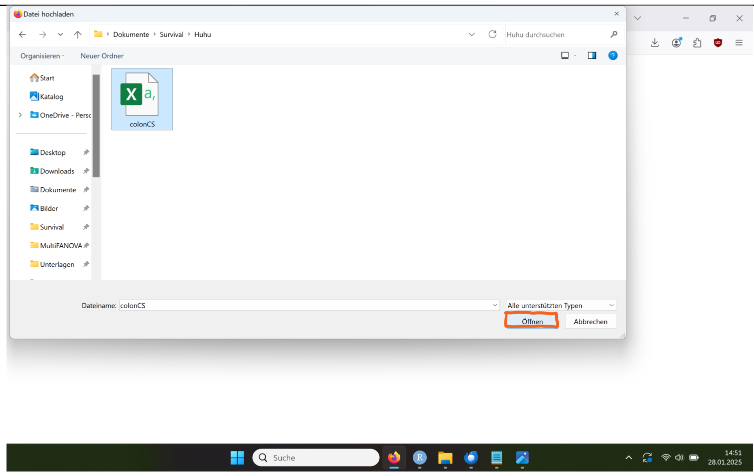
```
> summary(out, digits = 3)
# - - - - - - - - - - - Multiple asymptotic RMST Wald-type tests - - - - - - - - - - #

- Significance level: 0.05

#--- Hypothesis matrices ---------------------------------------------------------#
[[1]]
     [,1] [,2] [,3]
[1,]   -1    1    0

[[2]]
     [,1] [,2] [,3]
[1,]   -1    0    1

[[3]]
     [,1] [,2] [,3]
[1,]    0   -1    1


#--- Wald-type test statistics --------------------------------------------------#
W_n(H_1, c_1) W_n(H_2, c_2) W_n(H_3, c_3)
  0.004007639   9.142532300   8.536680422

#--- Overall p-values -----------------------------------------------------------#
[1] 0.998 0.007 0.010

#--- Overall results ------------------------------------------------------------#
     hyp_matrix estimator lwr_conf upr_conf test_stat critical value adj_pvalue decision
[1,] numeric,3  5.664     -204.041 215.368  0.004     5.494          0.998      "not significant"
[2,] numeric,3  267.275   60.076   474.473  9.143     5.494          0.007      "H1"
[3,] numeric,3  261.611   51.73    471.493  8.537     5.494          0.01       "H1"
#---------------------------------------------------------------------------------#
```

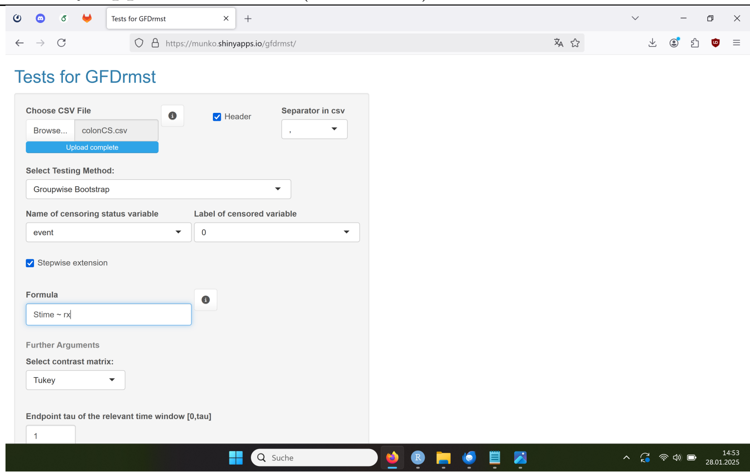
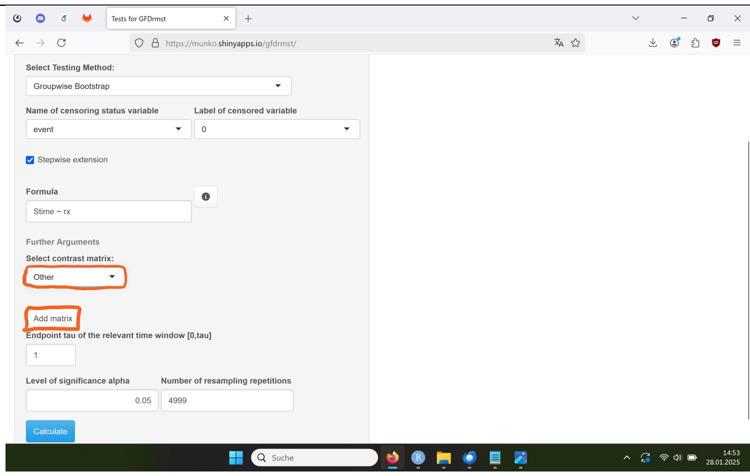
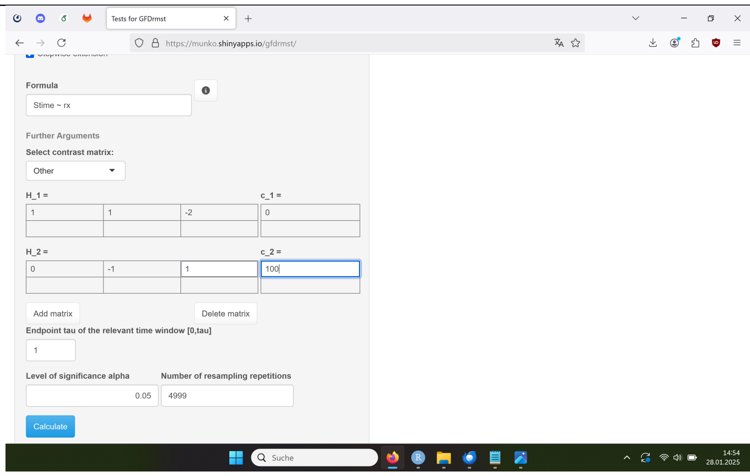
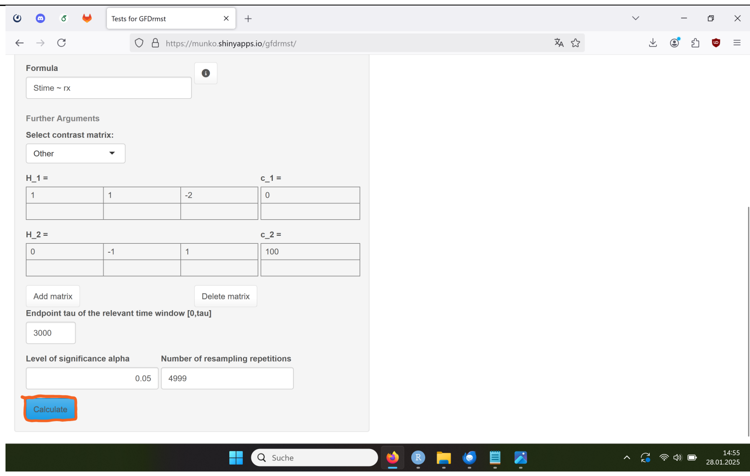
plot.GFDrmst With the function `plot.GFDrmst`, simultaneous confidence intervals of an object of class `"GFDrmst"` can be plotted. The function only takes the argument `x`, which is an object of class `"GFDrmst"`. All further arguments passed to this function are ignored. The displayed vectors on the y-axis are the coefficients $\mathbf{H}_\ell$ of the linear combinations. The function has no return value, but is called for side effects.
The corresponding R code and plot for our example is the following.

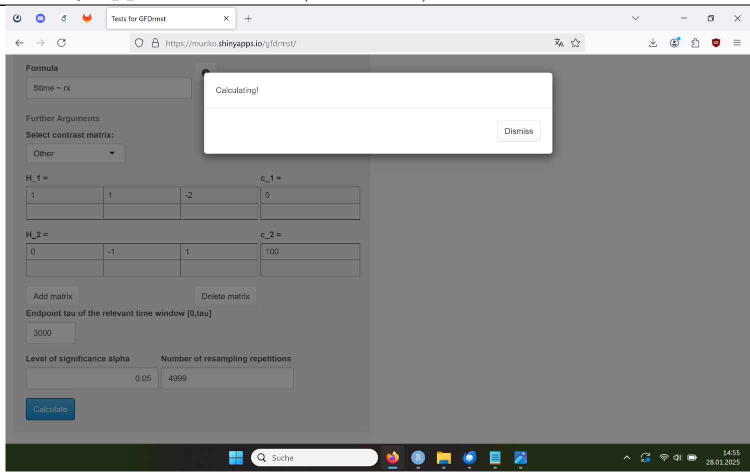
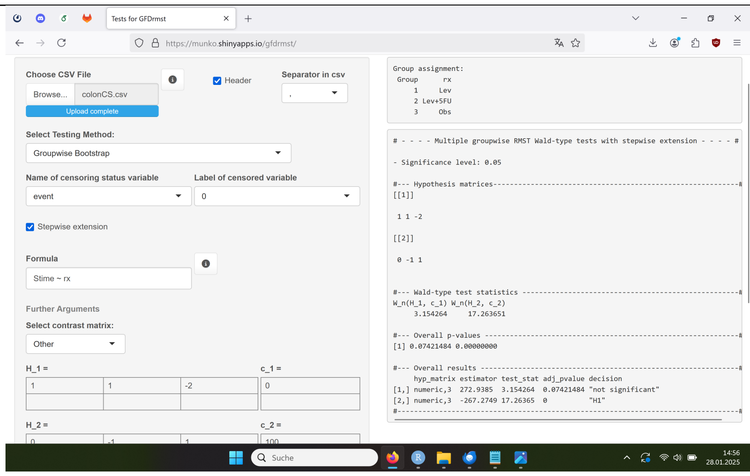
```
> plot(out)
```



GFDrmstGUI This function provides a shiny app for performing the multiple RMST-based tests. To explain its usage, we provide several screenshots in the following.

Example usage of the shiny app **GFDrmstGUI**.	
1. First of all, the shiny app needs to be started with the command `GFDrmstGUI()` in R or via `https://munko.shinyapps.io/gfdrmst/`. Depending on the csv file, (un)check "Header" if the csv (does not) contain headers and choose the correct separator. Next, the csv file containing the data can be loaded by clicking on "Browse...".	
2. Choose and open the csv file containing the data.	

Example usage of the shiny app `GFDrmstGUI` (continued).	
3. Now, further options become visible. Select the testing method that should be used and make sure that the correct censoring status variable (here: event) and the correct label of a censored variable (here: 0) are selected. If "Stepwise extension" is unchecked, the option for plotting simultaneous confidence intervals becomes possible. The "Formula" field allows user-specific formulas to model the event types by the different factors (e.g. here: "Stime ∼ rx" for one factor or "Stime ∼ rx * sex" for two factors).	
4. For the contrast matrix, options for the Tukey-type, Dunnett-type and for the centering matrix are available. Moreover, it is possible to define user-specific matrices, which we show in the following. Therefore, choose the option "Other". To add a user-specific matrix, click on "Add matrix".	
5. Fill in the desired hypothesis matrices $\mathbf{H}_\ell$. If you want to perform a permutation test, make sure that all matrices are contrast matrices. Further, fill in the hypothesis vectors $\mathbf{c}_\ell$. Further matrices can be added by clicking on "Add matrix". Accidentally added matrices can be deleted by clicking on "Delete matrix".	
6. Specify the endpoint τ (here: 3000), the global level of significance α (here: 0.05) and the number of resampling repetitions B (here: 4999). When everything is in place, click on "Calculate" to start the calculations.	

<table>
<tr><th colspan="2">Example usage of the shiny app <code>GFDrmstGUI</code> (continued).</th></tr>
<tr><td>7. Wait until the calculations are done. This might take a while.</td><td>
</td></tr>
<tr><td>8. When the calculations are done, the results appear immediately. In the upper field, the group assignment to the considered factors (here: rx) is clarified. Then, the results of the chosen testing method are shown. In this example, we observe that the second hypothesis can be rejected, while the first can not be rejected (see the last column of "Overall results").</td><td>
</td></tr>
</table>

E.2 GFDrmtl

The R package `GFDrmtl` works rather similar to the R package `GFDrmst`. It exports only two functions:
The function `GFDrmtlGUI` provides a shiny app for performing the multiple RMTL-based tests as described in Section 5 on an interactive user interface. Its usage closely mirrors the usage of the function `GFDrmstGUI` as described in Section E.1. Due to this similarity, we have chosen to omit a detailed description here to avoid redundancy.
The second function is called `RMTL.test` and can be used to perform the multiple RMTL-based tests as described in Section 5. The output of this function is a list of class `GFDrmst`. A detailed description can be found below.

`RMTL.test` This function contains the implementation of the asymptotic multiple tests of Section 5.3.1 and the adjusted permutation tests of Section 5.2.2. Furthermore, confidence intervals for RMTL contrasts can be calculated as in Remark 2.6 and the stepwise extension of Remark 2.5, which can improve the power of the multiple tests, is available.
The arguments of the function `RMTL.test` are more or less the same than of the function `RMST.test`, except for

- `status`
  A vector containing the corresponding censoring status indicator $\delta_{ij} = D_{ij}\mathbb{1}\{X_{ij} = T_{ij}\}$ with values $0 =$ censored and $1, ..., M$ for the M different competing events.

- `hyp_mat`
  A list containing all the hypothesis matrices $\mathbf{H}_1, ..., \mathbf{H}_L$ for the multiple tests or one of the options `"Tukey"`, `"Dunnett"`, `"center"`, cf. Example 5.2, or `"2by2"`, `"2by2 cause-wisely"` for tests on main and interaction effects in a 2-by-2 design without or with cause-wise results, cf. Example 5.7, or a matrix $\mathbf{H}$ if only one hypothesis is of interest. For the permutation test, all matrices need to fulfill the contrast property in Section 5.2.

- `M`
  An integer specifying the number of competing risks. By default, the maximum of the values in `status` or `event` is chosen.

- `method`
  One of the methods `"permutation"` or `"asymptotic"` that should be used for calculating the critical values, cf. Sections 5.2.2 and 5.3.1, respectively. Default option is `"permutation"`.

The output of this function is a list of class `GFDrmst`. For a detailed description of the components see Section E.1. As in Section E.1, either `time, status, group` or `formula, event, data` needs to be specified. The following example illustrates the usage of both versions. Note that since we use the same class `GFDrmst` as in the package `GFDrmst`, we can still use the already defined functions `summary.GFDrmst` and `plot.GFDrmst` to summarize and plot the results of the output of `RMTL.test`, respectively.

```
> library(GFDrmtl)
> library(mstate)
> data("ebmt2")
>
> # multiple asymptotic tests
> out <- RMTL.test(time = ebmt2$time,
+                  status = ebmt2$status,
+                  group = ebmt2$match,
+                  hyp_mat = "Dunnett",
+                  tau = 120,
+                  method = "asymptotic")
> summary(out)
# - - - - - - - - - - - Multiple asymptotic RMTL Wald-type tests - - - - - - - - - - - #

- Significance level: 0.05

#--- Hypothesis matrices-----------------------------------------------------------#
[[1]]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]   -1    0    0    0    0    0    1    0    0     0     0     0

[[2]]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]    0   -1    0    0    0    0    0    1    0     0     0     0

[[3]]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]    0    0   -1    0    0    0    0    0    1     0     0     0

[[4]]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]    0    0    0   -1    0    0    0    0    0     1     0     0

[[5]]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]    0    0    0    0   -1    0    0    0    0     0     1     0

[[6]]
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]    0    0    0    0    0   -1    0    0    0     0     0     1


#--- Wald-type test statistics ----------------------------------------------------#
W_n(H_1, c_1) W_n(H_2, c_2) W_n(H_3, c_3) W_n(H_4, c_4) W_n(H_5, c_5) W_n(H_6, c_6)
   0.05327241   16.65959825    0.02229563    0.09266721    0.72453873    5.25291896

#--- Overall p-values -------------------------------------------------------------#
[1] 0.99996141 0.00026849 0.99999708 0.99980490 0.94952365 0.12387207

#--- Overall results --------------------------------------------------------------#
     hyp_matrix estimator   lwr_conf   upr_conf  test_stat  critical value adj_pvalue
[1,] numeric,12 -0.2039252  -2.527734  2.119884  0.05327241 6.917694       0.9999614
[2,] numeric,12 3.624833    1.289029   5.960638  16.6596    6.917694       0.00026849
[3,] numeric,12 -0.05441319 -1.012875  0.904049  0.02229563 6.917694       0.9999971
[4,] numeric,12 0.1122358   -0.8574905 1.081962  0.09266721 6.917694       0.9998049
[5,] numeric,12 -0.3092148  -1.26467   0.6462401 0.7245387  6.917694       0.9495237
[6,] numeric,12 2.047158    -0.3021053 4.396421  5.252919   6.917694       0.1238721
     decision
[1,] "not significant"
[2,] "H1"
[3,] "not significant"
[4,] "not significant"
[5,] "not significant"
[6,] "not significant"
#----------------------------------------------------------------------------------#
>
```

```
> ## or, equivalently,
> out <- RMTL.test(formula = "time ~ match",
+                  event = "status",
+                  data = ebmt2,
+                  hyp_mat = "Dunnett",
+                  tau = 120,
+                  method = "asymptotic")
```

List of Symbols

$\mathbf{0}_r$	r-dimensional vector of zeros
$\mathbf{0}_{r\times k}$	$r \times k$ dimensional matrix of zeros
$\mathbf{1}_k$	k-dimensional vector of ones
$\mathbb{1}$	indicator function
$\mathbf{A}^+$	Moore–Penrose inverse of a matrix $\mathbf{A}$
$\xrightarrow{a.s.}$	almost sure convergence
$BL_1(\mathbb{E})$	set of all real functions on $\mathbb{E}$ with a Lipschitz norm bounded by 1
$BV_M[a,b]$	subset of càdlàg functions $[a,b] \to \mathbb{R}$ with total variation bounded by M
$\tilde{BV}_M[a,b]$	subset of all functions in $\tilde{D}[a,b]$ with total variation bounded by M and $F \in D[a,b]$
$BV_M^{>-1}[a,b]$	subset of functions in $BV_M[a,b]$ whose jumps are contained in $(-1,\infty)$ and bounded away from -1
χ_r^2	χ^2-distribution with r degrees of freedom
$\chi_{r,1-\alpha}^2$	$(1-\alpha)$-quantile of the χ_r^2-distribution
$\mathbb{C}ov$	covariance
$\xrightarrow{d}$, $\rightsquigarrow$	weak convergence in the sense of [74]
$\xrightarrow{d*}$	conditional weak convergence in outer probability in the sense of Definition 2.1
$D[a,b]$	set of all càdlàg functions $[a,b] \subset \mathbb{R} \to \mathbb{R}$, equipped with the sup-norm
$\tilde{D}[a,b]$	set of all functions $[a,b] \to \mathbb{R}$ that are everywhere left-continuous and have right limits everywhere
E	expectation
E^*	outer expectation
$\mathrm{E}_1, \mathrm{E}_2$	expectations regarding $(\Omega_1, \mathcal{A}_1, Q_1), (\Omega_2, \mathcal{A}_2, Q_2)$, respectively
F_-	left-continuous version of a monotone or a càdlàg function F
F^{-1}	inverse of a monotone increasing and right-continuous function F, that is $\inf\{x \mid . \leqslant F(x)\}$
ΔF	increment of a monotone function F, that is $F - F_-$
$\mathbf{I}_k$	$k \times k$ dimensional unit matrix
i.i.d.	independent identically distributed
$\ell^\infty(\mathcal{F})$	set of all bounded real-valued functions on $\mathcal{F}$
$\mathcal{N}(\mu, \sigma^2)$	normal distribution with mean μ and variance $\sigma^2 \geqslant 0$
$\mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	k-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\bigoplus$	direct sum
$\times$	Cartesian product
$\otimes$	Kronecker product
P	probability
P^*	outer probability
$\xrightarrow{P}$	convergence in outer probability
$\prod$	product integral as in [39]
$\mathbb{V}ar$	variance