NON-PARAMETRIC SEQUENTIAL AND ADAPTIVE DESIGNS FOR SURVIVAL TRIALS

von Matthias Brückner

DISSERTATION

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
- Dr. rer. nat. -

Vorgelegt im Fachbereich 3 (Mathematik & Informatik)
der Universität Bremen
im Juli 2014

Datum der Einreichung: 31.07.2014
Datum des Promotionskolloquiums: 01.10.2014

Gutachter: Prof. Dr. Werner Brannath (Universität Bremen)
Prof. Dr. Martin Posch (Medizinische Universität Wien)

ABSTRACT

This thesis deals with fixed samples size, sequential and adaptive survival trials and consists of two major parts. In the first part fixed sample size, sequential and adaptive testing methods are derived that utilize data from a survival as well as a categorical surrogate endpoint in a fully non-parametric way without the need to assume any type of proportional hazards. In the second part extensions to quality-adjusted survival endpoints are discussed.

In existing adaptive methods for confirmatory survival trials with flexible adaptation rules strict type-I-error control is only ensured if the interim decisions are based solely on the primary endpoint. In trials with long-term follow-up it is often desirable to base interim decisions also on correlated short-term endpoints, such as a surrogate marker. Surrogate information available at the interim analysis may be used to predict future event times. If interim decisions, such as selection of a subgroup or changes to the recruitment process, depend on this information, control of the type-I-error is no longer formally guaranteed for methods assuming an independent increments structure.

In this thesis the weighted Kaplan-Meier estimator, a modification of the classical Kaplan-Meier estimator incorporating discrete surrogate information, is used to construct a non-parametric test statistic for the comparison of survival distributions, a generalization of the average hazard ratio. It is shown in this thesis how this test statistic can be used in fixed design, group-sequential and adaptive trials, such that the type-I-error is controlled. Asymptotic normality of the multivariate average hazard ratio is first verified in the fixed sample size context and then applied to non-inferiority testing in a three-arm trial with non-proportional hazards survival data. In the next step the independent increments property is shown to hold asymptotically for the weighted Kaplan-Meier estimator. Consequently, for all test statistics based on it. Standard methods for the calculation of group-sequential rejection boundaries are applicable. For adaptive designs the weighted Kaplan-Meier estimator is modified to support stage-wise left-truncated and right-censored data to ensure independence of the stage-wise test statistics, even when interim decisions are based on surrogate information. Standard combination test methodology can then be used to ensure strict type-I-error control.

Quality-adjusted survival is an integrated measure of quality-of-life data, which has gained interest in recent years. In this thesis a novel non-parametric two-sample test for quality-adjusted survival distributions is developed, that allows adjustment for covariate-dependent censoring, whereby the censoring is assumed to follow a proportional hazards model. It is shown how this result can be used to design adaptive trials with a quality-adjusted survival endpoint.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

In randomized controlled clinical trials (RCTs) with a fixed design the data collected during the trial is only unblinded at the end of trial. In sequential RCTs the data is unblinded at one or more interim analyses and a decision is made to continue or to stop the trial, allowing early rejection of the null hypothesis or stopping for futility. No data-dependent changes to the trial design are allowed. Adaptive trial designs explicitly account for the possibility of data-dependent changes, such as sample size adjustment or subgroup selection. An important problem in these designs is the control of the type-I-error rate. Two approaches for the strict control of the the type-I-error are the conditional rejection probability principle by Müller and Schäfer [2004] and the combination test approach by Bauer [1989] and Bauer and Köhne [1994].

The conditional rejection probability principle requires one to calculate the conditional probability of rejection given all the information used in the interim decision. The second stage must then be planned with a significance level equal to this probability. The combination test approach combines two stochastically independent test statistics or p-values using a pre-specified combination function. In trials, where the outcome of interest is immediately available, stochastic independence of the test statistics can be easily achieved by splitting the patient population into those recruited before and those recruited after the interim analysis.

For survival data (or data with delayed response) the problem of overrunning patients arises, i.e. patients recruited before the interim analysis whose outcomes have not yet been observed at the time of the interim analysis. These patients are included as right-censored observations in the first stage and are followed-up in the second stage, since stopping follow-up of these patients at the interim analysis would be inefficient. In this case independence of the stage-wise test statistics for survival data is achieved by exploiting the (asymptotic) independent increments property of the efficient score (Schäfer and Müller [2001] and Wassmer [2006]). However, in general the independent increments property does not hold, when using secondary endpoint data, like from surrogate variables, in the interim decision from patients whose primary endpoint has not yet been observed. This can lead to an inflation of the type-I-error rate as illustrated in an extreme scenario by Bauer and Posch [2004]. The

basic idea in this example is, that the secondary endpoint information can be used to predict the primary endpoint opening up the possibility of trial manipulations to increase the probability of rejection.

Consider a two-sample comparison of overall survival using the log-rank test in a two-stage trial. The interim analysis is scheduled before any overall survival event has been observed, but after the surrogate variable has been observed for all patients. If the surrogate variable is a perfect predictor of the event time and there is no censoring, then the value of the log-rank test statistic and hence the future test decision can be exactly predicted. If the log-rank test would reject, then recruitment is stopped, else a huge number of additional patients is recruited, such that the influence of the first stage patients is negligible. If the significance level of the log-rank test is α, then with probability α recruitment is stopped at the interim analysis and the null hypothesis is rejected with probability one at the end of the trial. With probability $1-\alpha$ recruitment is increased, such that the rejection probability at the end is again α. The overall type-I-error is $\alpha + (1-\alpha)\alpha = 2\alpha - \alpha^2$ (see figure 1.1). For $\alpha = 0.05$ this equals 0.0975.

If the surrogate variable is binary the problem can be illustrated in the following way. Suppose the binary surrogate indicates non-response / response to the treatment. Suppose the response indicator is a perfect predictor of survival. Under the null hypothesis the response probability is the same in the experimental and the control group. However, by pure chance (with probability α) a difference in the response rates will be observed at the interim analysis, large enough to lead to a rejection of the null hypothesis. By stopping recruitment at this point, the difference in the observed response rates is retained and the null hypothesis eventually rejected. If the difference is small (with probability $1-\alpha$) recruitment is increased, such that again with probability α a large enough difference in the response rates is observed at the final analysis, leading to inflation of the overall type-I-error. Inflation will also occur in less extreme scenarios, i.e. when less patients than planned are recruited, but recruitment is not completely stopped. The stratified log-rank test, where the strata are defined according to the levels of the surrogate variable, would control the type-I-error in this scenario, since it completely ignores differences in the response rates. The stratified log-rank test only estimates the effect *conditional* on the response indicator. A higher response rate in the experimental group than in the control group, which would be likely, if the treatment were effective, would not contribute to the power of the stratified log-rank test. Hence, the stratified log-rank test controls the type-I-error rate, however it does not exploit treatment effects in the surrogate endpoint. The stratified

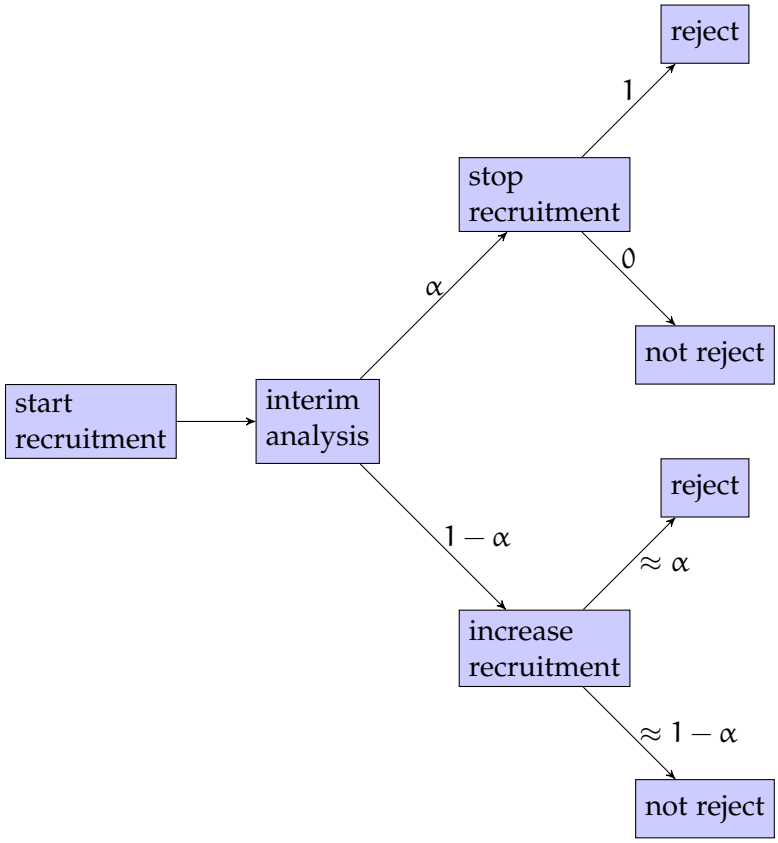

Figure 1.1: Type-I-error inflation strategy from Bauer and Posch [2004]. Overall type-I-error is $\approx 2\alpha - \alpha^2$.

log-rank test has been used by Brannath et al. [2009] in the context of an adaptive enrichment design with treatment selection based on the survival endpoint only.

Jenkins et al. [2011] propose a different solution to the problem of Bauer and Posch [2004], by splitting the patient population into two disjunct subsets, patients recruited before and patients recruited after interim analysis, like in the non-survival case (patient-wise splitting). Patients recruited before the interim analysis are followed-up after the interim analysis. The design for these patients must remain unchanged. This also means, that the pre-specified follow-up time must not be changed as noted by Magirr et al. [2014]. Choosing the follow-up time, such that the value of the conditional error rate is maximized inflates the probability of rejection. This problem can also be interpreted as an informative censoring problem for the overrunning patients. The censoring of the overrunning patients is manipulated, by changing the follow-up time. Censoring and survival time are no longer stochastically independent. How-

ever, they are independent conditional on the surrogate variable. The validity of the log-rank test and the Kaplan-Meier estimator depends critically on the assumption of non-informative censoring. If a discrete surrogate variable is used, the stratified log-rank test would also solve the informative censoring problem, since within each stratum censoring would be non-informative, but again only the conditional effect would be estimated.

For a discrete time-independent surrogate variable the informative censoring problem can be solved by a modification of the Kaplan-Meier estimator, called the weighed Kaplan-Meier estimator, which incorporates the surrogate variable information, and has been proposed by Malani [1995] and Murray and Tsiatis [1996]. The weighted Kaplan-Meier estimator estimates the marginal survival function, by combining estimates of the conditional survival functions (given the surrogate) weighted by the corresponding response probabilities. By using the surrogate variable information, information lost due to censoring can be recovered, increasing the efficiency of the estimator. This estimator is used in this thesis to construct an estimator for the average hazard ratio introduced by Kalbfleisch and Prentice [1981]. In group-sequential trials and in adaptive trials, when patient-wise splitting is used, the weighted Kaplan-Meier estimator can be used directly without modification. For adaptive trials, with stage-wise splitting, if adaptive design changes have been made at an interim analysis, then the maximum likelihood estimator of the response probability using all data, i.e. from patients recruited before and after the interim analysis, is biased (Brannath et al. [2006]). The weighted Kaplan-Meier estimator needs to be modified accordingly.

OUTLINE

Chapter 2 is a short introduction to the basic concepts of survival analysis, including the Kaplan-Meier estimator and the (stratified) proportional hazards model. In chapter 3 the weighted Kaplan-Meier estimator and the average hazard ratio are introduced and their large sample properties are derived for the k-sample case ($k \geqslant 0$) in a fixed sample size setting. Two interesting special cases are considered, the two sample and the three sample case. In the three sample case non-inferiority testing based on the average hazard ratio is considered. This provides a generalization of Kombrink et al. [2013] to non-proportional hazards data. Related two-sample methods, such as the test statistics proposed by Pepe and Fleming [1989] and the median are considered. Type-I-error and power of the two-sample methods are compared with

the log-rank test in simulations. In chapter 4 the sequential weighted Kaplan-Meier estimator and sequential average hazard ratio test statistics are defined. The independent increments property is shown to hold asymptotically for these test statistics and other test statistics based on the sequential weighted Kaplan-Meier estimator. In chapter 5 asymptotic results for the weighted Kaplan-Meier estimator for stage-wise left-truncated and right-censored data are derived. These results and the general results from chapter 3 can be used for adaptive designs, where interim decisions are based on discrete surrogate data. In a simulation study the method is applied to the problem of subgroup selection in a seamless adaptive phase II/III trial.

Inverse probability of censoring weighting (IPCW) methods for handling informative censoring in the context of quality-adjusted survival with a focus on two-sample testing are discussed in chapter 6. Asymptotic normality of a new IPCW Mann-Whitney U-statistic is proved, when censoring is assumed to follow a proportional hazards model. Adaptive designs for quality-adjusted survival using the patient-wise splitting approach based on this result are discussed.

All simulations are done with the R software package and programming environment (R Core Team [2014]). The newly developed methods are implemented in a new R package, described in the appendix. Furthermore, the appendix contains a summary of the mathematical background required for the asymptotic methods in survival analysis.

2

INTRODUCTION TO SURVIVAL ANALYSIS

This chapter introduces some basic concepts of survival analysis, including the Kaplan-Meier estimator, the Cox proportional hazards model and the (stratified) log-rank test.

2.1 BASIC CONCEPTS

In clinical with a time-to-event endpoint, such as overall survival, some patients may still be alive, when the trial ends (administrative censoring) or they are lost to follow-up during the trial (drop-out). This kind of missing data mechanism is called *(right-) censoring*. Only the minimum $Y = T \wedge C$ of the *survival time* T and the *censoring time* C is observed, as well as the censoring indicator $\delta = \mathbb{1}\{T \leqslant C\}$. The problem is to estimate the distribution of the survival time based on the independent and identically distributed observations $\{Y_i, \delta_i, i = 1, \ldots, n\}$ of (Y, δ). The distribution of T is completely determined by its *hazard rate*

$$\lambda(t) = \lim_{h \to 0} \frac{1}{h} P(t \leqslant T < t + h | T \geqslant t) \quad t \geqslant 0.$$

For continous survival times the relation between the *conditional survival function*

$$S(t) = P(T > t)$$

and the *cumulative hazard rate*

$$\Lambda(t) = \int_0^t \lambda(u) du$$

is given by

$$S(t) = e^{-\Lambda(t)}.$$

A basic assumption about the censoring mechanism is that the hazard rate of T is the same with and without censoring, i.e.

$$\lambda(u) = \lambda^{\#}(u), \tag{2.1.1}$$

where

$$\lambda^{\#}(u) = \lim_{h \downarrow 0} \frac{1}{h} P(t \leqslant T < t + h | T \geqslant t, C \geqslant t).$$

In competing risk terminology $\lambda^{\#}$ is also called *crude hazard rate* and λ is also called *net hazard rate*. The assumption in eq. (2.1.1) is called *independent censoring* or *non-informative censoring* (section III.2.2 in Andersen [1993]).

Remark 2.1. *The independent censoring assumption is, despite its name, weaker than stochastic independence of the survival and censoring time, which is also called* random censoring.

Besides (right-) censoring, observations may also be subject to *left-truncation*. This is most often the case in observational studies, when e.g. the patient is only observed some time after the diagnosis and the exact time of the diagnosis is unknown. Ignoring this fact, i.e. assuming, that the recruitment time is also the time of the first diagnosis may lead to *length bias*, since patients with longer survival have higher chance of being recruited (see e.g. example I.3.2 in Andersen [1993]). Only patients whose survival time is larger than the left-truncation time V are observed. V is set to 0 for patients, who are not left-truncated. Depending on the support of the left-truncation time V, the marginal distribution of the survival time may not be identifiable from the data. For example, denote the left endpoint of the support of V by τ. Then only the conditional probability $P(T > t | T \geqslant \tau)$ can be estimated from the data. Left-truncation is called *independent* or *non-informative*, when the hazard rate of the left-truncated survival time is the same as that of the original survival time (section III.3 in Andersen [1993]).

2.2 COUNTING PROCESS MARTINGALES

The derivation of the large sample properties of estimators and test statistics for survival data is usually based on the theory of counting processes and the associated martingales (the essential definitions and results are given in the appendix A). Most estimators and test statistics can be written as a sum of stochastic integrals with respect to a counting process martingale. For $i = 1, \ldots, n$, consider the stochastic process $N_i(s) = \delta_i \mathbb{1}\{Y_i \leqslant s\}$, which indicates, if patient i died in the interval $[0, s]$. Similarly, let $N_i^c(s) = (1 - \delta_i)\mathbb{1}\{Y_i \leqslant s\}$, which indicates, if patient i was censored

in the interval $[0, s]$. Moreover let $\mathcal{F} = (\mathcal{F}_t, t \geqslant 0)$ be the filtration generated by the processes N_i and N_i^c, $i = 1, \ldots, n$, where

$$\mathcal{F}_t = \sigma(N_i(s), N_i^c(s), s \leqslant t, i = 1, \ldots, n).$$

The compensator of N_i with respect to $\mathcal{F}$ exists. Under the independent censoring assumption it is given by the continuous predictable process

$$\int_0^t Y_i(u)\lambda(u)du,$$

i.e.

$$M_i(t) = N_i(t) - \int_0^t Y_i(u)\lambda(u)du$$

is a local square integrable martingale with respect to $\mathcal{F}$, where λ is the hazard rate of T_i (theorem 1.3.1 from Fleming and Harrington [2011]). If follows, that the optional quadratic variation process $[M_i]$ of M_i is equal to N_i and the predictable quadratic variation process $\langle M_i \rangle$ is given by

$$\langle M_i \rangle(t) = \int_0^t Y_i(u)\lambda(u)du \quad t \geqslant 0.$$

Intuitively, every counting process can be written as the sum of a systematic part (compensator) and mean zero noise (martingale). In this sense the properties of the counting process are determined by the hazard rate, explaining its importance in the mathematical treatment of time-to-event data.

If the observed times Y_i, $i = 1, \ldots, n$, are continuous random variables and stochastically independent, then the counting processes N_i, $i = 1, \ldots, n$, have no common jump discontinuities, since the probability of two independent continuous random variables having the same value is 0. Thus the sum $\bar{N}^{(n)} = N_1 + \ldots + N_n$ is again a counting process and the counting process martingales M_i, $i = 1, \ldots, n$, are independent. Denote the number at risk at time u by $\bar{Y}^{(n)}(u) = Y_1(u) + \ldots + Y_n(u)$. The sum $\bar{M}^{(n)} = M_1 + \ldots + M_n$ is a local square integrable martingale with predictable quadratic variation

$$\left\langle \bar{M}^{(n)} \right\rangle = \sum_{i=1}^{n} \langle M_i \rangle = \int_0^t \bar{Y}^{(n)}(u)\lambda(u)du.$$

2.3 KAPLAN-MEIER ESTIMATOR

The independent censoring assumption is now relaxed and the survival T and the censoring time C are assumed to be stochastically independent conditional on a discrete covariate $X \in \{1, \dots, J\}$, i.e. T and C are stochastically independent in each of the strata defined by the levels of X. The observed data is

$$\{(Y_i = T_i \wedge C_i, \delta_i = \mathbb{1}\{T_i \leqslant C_i\}, X_i), i = 1, \dots, n\}.$$

Denote the conditional survival function of T in stratum j by $S_j(t) = P(T > t|X = j)$, the conditional survival function for the censoring time C by $K_j(t) = P(C > t|X = j)$ and the conditional hazard rate for T in stratum j by

$$\lambda_j(t) = \lim_{h \to 0} \frac{1}{h} P(t \leqslant T < t + h | T \geqslant t, X = j).$$

The stratum-specific cumulative hazard rate is then

$$\Lambda_j(t) = \int_0^t \lambda_j(s) ds.$$

In each stratum the cumulative hazard rate and the survival function can be estimated by the Nelson-Aalen estimator (Andersen [1993, section IV.1]) and the Kaplan-Meier estimator (Andersen [1993, section IV.3]), respectively, using only observations in the specific stratum:

Definition 2.2 (stratum-specific Nelson-Aalen and Kaplan-Meier estimators)**.** *For* $j = 1, \dots, J$, $s > 0$, *the stratum-specific Nelson-Aalen estimator is given by*

$$\hat{\Lambda}_j(s) = \int_0^s J_j^{(n)}(u) \frac{\bar{N}_j^{(n)}(du)}{\bar{Y}_j^{(n)}(u)} \tag{2.3.1}$$

where

$$\bar{N}_j^{(n)}(u) = \sum_{i=1}^n \mathbb{1}\{X_i = j\} N_i(u),$$

is the number of events in stratum j *up to time* u*,*

$$\bar{Y}_j^{(n)}(u) = \sum_{i=1}^n \mathbb{1}\{X_i = j\} Y_i(u),$$

is the number at risk in stratum j *up to time* u *and*

$$J_j^{(n)}(u) = \mathbb{1}\left\{\bar{Y}_j^{(n)}(u) > 0\right\}.$$

The stratum-specific Kaplan-Meier estimator is given by

$$\hat{S}_j(s) = \prod_{(0,s]} \left\{1 - \hat{\Lambda}_j(ds)\right\} = \prod_{(0,s]} \left\{1 - J_j^{(n)}(u)\frac{\bar{N}_j^{(n)}(du)}{\bar{Y}_j^{(n)}(u)}\right\}, \tag{2.3.2}$$

where $\prod_{(0,s]}$ *denotes the product integral (section A.3), i.e.*

$$\hat{S}_j(s) = \prod_{i:Y_i \leqslant s} \left\{1 - \frac{\delta_i \mathbb{1}\{X_i = j\}}{\bar{Y}_j^{(n)}(Y_i)}\right\}.$$

2.4 COX PROPORTIONAL HAZARDS MODEL

The Cox proportional hazards model is the most widely used semi-parametric model for regression analysis of survival times. The large sample properties of the maximum partial likelihood estimator of the regression coefficients and the closely related stratified log-rank test are presented. The consequences, when the model assumptions are incorrect, are discussed.

2.4.1 *Stratified proportional hazards model*

This section considers the situation in which the survival and censoring times depend on a vector Z of time-independent covariates and a stratification variable $X_i \in \{1, \ldots, J\}$, such that T and C are stochastically independent conditional on (X, Z), and the conditional hazard rate in each stratum

$$\lambda_j(t|Z) = \lim_{h \to 0} \frac{1}{h} P(t < T \leqslant t + h | T \geqslant t, X = j, Z)$$

follows a stratified proportional hazards model

$$\lambda_j(t|Z) = \lambda_{0j}(t) e^{\beta_0^T Z}, \tag{2.4.1}$$

where λ_{0j}, $j = 1, \ldots, J$ are unknown baseline hazard functions and β_0 is a vector of unknown regression coefficients. Each stratum may have a different baseline hazard λ_{0j}, while the effect of the covariate Z is the same in each stratum. This allows for stratum-specific effects on the hazard, which violate the overall proportionality assumption. The stratified model is a special case of the more general multivariate model of Andersen [1993]. In this model, there is a multivariate counting process

$$N_i(t) = (N_{i1}(t), \ldots, N_{ik}(t)) \quad k \geqslant 1$$

associated with each subject $i = 1, \ldots, n$, i.e. a vector of counting processes where no two counting processes may jump at the same time. Moreover there are type-specific censoring indicators δ_{ji}, covariates Z_{ji}, survival times T_{ji}, censoring time C_{hi} and at-risk indicators Y_{ji} for $j = 1, \ldots, J$ and $i = 1, \ldots, n$. For the stratified model set $T_{ji} = \mathbb{1}\{X_i = j\}T_i$, $C_{ji} = \mathbb{1}\{X_i = j\}C_i$, $N_{ji}(t) = \mathbb{1}\{X_i = j\}N_i(t)$, $\delta_{ji} = \mathbb{1}\{X_i = j\}\delta_i$, $Z_{ji} = \mathbb{1}\{X_i = j\}Z_i$, $Y_{ji}(t) = \mathbb{1}\{X_i = j\}Y_i(t)$. Then we have a multivariate counting process model as in Andersen [1993].

The support of the censoring distribution imposes a limit on the time range on which the survival function can be consistently estimated, namely

$$L < \sup\{u : P(T \wedge C > u | X = j) > 0\} \quad \forall j = 1, \ldots, J.$$

The following results are from Andersen [1993]. The estimator $\hat{\beta}_{PH}$ of β_0 is obtained by maximizing the partial likelihood

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} \left(\frac{e^{\beta^T Z_i}}{S^{(0)}_{X_i}(\beta, T_i)} \right)^{\delta_i}$$

where

$$S^{(0)}_j(\beta, t) = \sum_{i=1}^{n} \mathbb{1}\{Y_i \geqslant t, X_i = j\} e^{\beta^T Z_i} \tag{2.4.2}$$

Equivalently, $\hat{\beta}$ is the unique solution of the score equation $U(\beta) = 0$, where

$$U(\beta) = \frac{\partial}{\partial \beta} \log \mathcal{L}(\beta) = \sum_{i=1}^{n} \int_0^L \left[Z_i - \frac{S^{(1)}_{X_i}(\beta, t)}{S^{(0)}_{X_i}(\beta, t)} \right] dN_i(t), \tag{2.4.3}$$

and

$$S_j^{(1)}(\beta, t) = \frac{\partial S_j^{(0)}(\beta, t)}{\partial \beta} = \sum_{i=1}^{n} \mathbb{1}\{Y_i \geqslant t, X_i = j\} Z_i e^{\beta^T Z_i}. \tag{2.4.4}$$

The partial derivative of the vector $S_j^{(1)}(\beta, t)$ is needed to express the asymptotic covariance matrix of the vector $\hat{\beta}$,

$$S_j^{(2)}(\beta, t) = \frac{\partial S_j^{(1)}(\beta, t)}{\partial \beta} = \sum_{i=1}^{n} \mathbb{1}\{Y_i \geqslant t, X_i = j\} Z_i^{\otimes 2} e^{\beta^T Z_i}. \tag{2.4.5}$$

($x^{\otimes 2} = xx^T$ where x is any column-vector). Consistency and asymptotic normality of $\hat{\beta}$ requires the following technical conditions:

Condition 2.3 (Condition VII.2.1, Andersen [1993]). *There exists a neighbourhood $\mathcal{G}$ of β_0 and scalar, vector and matrix functions $s_j^{(0)}$, $s_j^{(1)}$ and $s_j^{(2)}$, respectively, defined on $\mathcal{G} \times [0, L]$ such that for $l = 0, 1, 2$, $j = 1, \ldots, J$,*

(a)

$$\sup_{t \in [0,L], \beta \in \mathcal{G}} \left\| \frac{1}{n} S_j^{(l)}(\beta, t) - s_j^{(l)}(\beta, t) \right\| = o_p(1);$$

(b) $s_j^{(l)}(\cdot, t)$ is a continuous function of $\beta \in \mathcal{G}$ uniformly in $t \in [0, L]$ and bounded on $\mathcal{G} \times [0, L]$;

(c) $s_j^{(0)}(\beta_0, \cdot)$ is bounded away from 0 on $[0, L]$;

(d) for $\beta \in \mathcal{G}$, $t \in [0, L]$,

$$s_j^{(1)}(\beta, t) = \frac{\partial}{\partial \beta} s_j^{(0)}(\beta, t), \qquad s_j^{(2)}(\beta, t) = \frac{\partial^2}{\partial \beta^2} s_j^{(0)}(\beta, t)$$

(e) the matrix

$$\Omega = \sum_{j=1}^{J} \int_0^L \left[\frac{s_j^{(2)}(\beta_0, t)}{s_j^{(0)}(\beta_0, t)} - \frac{s_j^{(1)}(\beta_0, t)^{\otimes 2}}{s_j^{(0)}(\beta_0, t)^2} \right] s_j^{(0)}(\beta_0, t) \lambda_{0j}(t) dt \tag{2.4.6}$$

is positive definite;

(f) the cumulative baseline hazard is finite, i.e.

$$\int_0^L \lambda_{0j}(t) dt < \infty.$$

Theorem 2.4 (Theorem VII.2.1, Andersen [1993]). *Assume condition 2.3. Then the probability that the equation* $U(\beta) = 0$ *has a unique solution* $\hat{\beta}$ *tends to* 1 *and* $\hat{\beta} \xrightarrow{p} \beta_0$ *as* $n \to \infty$.

Theorem 2.5 (Theorem VII.2.2, Andersen [1993]). *Assume conditions 2.3. Then, as* $n \to \infty$,

$$\sqrt{n}\{\hat{\beta} - \beta_0\} \xrightarrow{\mathcal{L}} N(0, \Omega^{-1})$$

and

$$n^{-1}\hat{\Omega}(\hat{\beta}) \xrightarrow{p} \Omega \quad \textit{as } n \to \infty$$

where

$$\hat{\Omega}(\beta) = -\frac{\partial U(\beta)}{\partial \beta} = \sum_{i=1}^{n} \delta_i \left\{ \frac{S^{(2)}_{X_i}(\hat{\beta}, T_i)}{S^{(0)}_{X_i}(\hat{\beta}, T_i)} - \frac{S^{(1)}_{X_i}(\hat{\beta}, T_i)^{\otimes 2}}{S^{(0)}_{X_i}(\hat{\beta}, T_i)^2} \right\}. \tag{2.4.7}$$

2.4.2 *Stratified log-rank test*

The stratified log-rank test is the score test in the stratified proportional hazards model,

$$\lambda(t|X, Z) = \lambda_X(t) e^{\beta_0^T Z},$$

where Z is a discrete covariate and X is the discrete stratification variable. The log-rank test statistic is identical to the standardized score test statistic

$$U(\hat{\beta})^T \hat{\Omega}^{-1}(\hat{\beta}) U(\hat{\beta}).$$

This test statistic has asymptotically a χ^2-distribution with number of degrees equal to 1 minus the number of levels of Z. One-sided hypothesis tests are not possible with the χ^2-distribution If Z is binary, then the score test statistic can also be defined as

$$\frac{U(\hat{\beta})}{\sqrt{\hat{\Omega}}},$$

which under the null hypothesis $H_0 : \beta_0 = 0$ (and condition 2.3) converges in distribution to standard normal random variable.

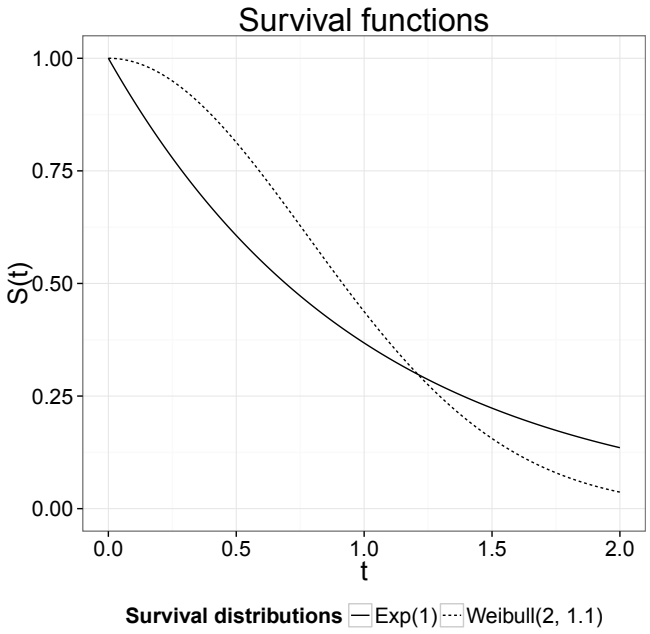

Figure 2.1: Example of two survival distributions with non-proportional hazards

2.4.3 *Misspecified proportional hazards model*

There exists extensive literature concerning the properties of the maximum partial likelihood estimator and the log-rank test, when the assumptions of the proportional hazard model are violated. Struthers and Kalbfleisch [1986] show, that the maximum partial likelihood estimator of the misspecified model still converges in probability to a limit β^* defined as the solution of the equation

$$h(\beta) = \int_0^\infty s^{(1)}(x) - \int_0^\infty \frac{s^{(1)}(\beta, x)}{s^{(0)}(\beta, x)} s^{(0)}(x) dx = 0, \tag{2.4.8}$$

where

$$\begin{aligned} s^{(l)}(x) &= E[P(T \wedge C > x \mid Z) Z^l \lambda(x|Z)], \\ s^{(l)}(\beta, x) &= E[P(T \wedge C > x \mid Z) Z^l e^{\beta Z}], \qquad l = 0, 1. \end{aligned}$$

The parameter β^* is also called the *least false parameter*. It corresponds to the proportional hazards model, which minimizes the distance to the true model with respect to the Kullback-Leibler divergence (Hjort [1992]).

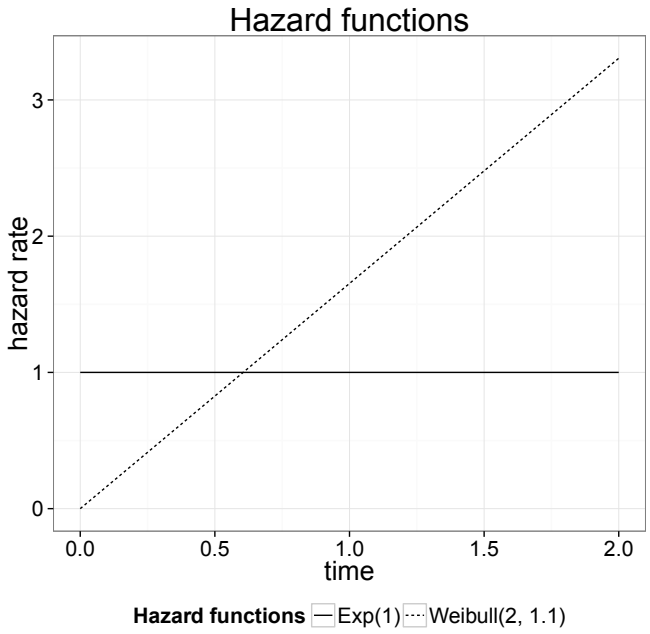

Figure 2.2: Example of two non-proportional hazard functions

The value of β^* will in general depend on the distribution of the censoring time (thus implicitly also on the recruitment process and the calendar time at which the analysis is made). Hence the power of the log-rank test will also depend on these nuisance parameters. This is demonstrated in a simple non-proportional hazards example with a single binary covariate (group indicator). The survival distributions are exponential with parameter 1 in group 1 and Weibull with shape parameter 2 and scale parameter 1.1 in group 2 (see figures 2.1 and 2.2). The log hazard ratio at time t is approximately $0.5 + \log(t)$. Three different censoring distributions are considered. Table 2.1 shows the results of 10^4 simulation runs with a sample size of 500 in each group at a significance level of 0.05. The maximum follow-up time was 2. The simulation results show that the power of the log-rank test for non-proportional hazards alternatives depends strongly on the shape of the censoring distribution. This is already a problem for fixed designs, but is even worse in adaptive designs, where the adaptations may lead to a manipulation of the censoring mechanism.

Table 2.1: Power of the log-rank test under a non-proportional hazards alternative for different censoring distributions[a].

Censoring	Censored obs. (%)	β^*	$\hat{\beta}_{PH}$	**Power**
Uniform[0,2]	46	0.77	0.77	**0.86**
Weibull(1, 2.24)	35	0.84	0.89	**0.32**
Weibull(2, 2.3)	22	0.96	0.94	**0.14**

[a] Results from 10^4 simulation runs for each scenario at a significance level of 0.05 and sample size 500 in each group.

2.5 INFORMATIVE CENSORING

All methods considered so far are sensitive to violations of the independent censoring assumption. This may be the case, when survival and censoring times are dependent via a common covariate. Specifically, the type-I-error of the log-rank test may be inflated, when censoring is dependent on treatment and related to survival via a covariate.

As an example consider the scenario described in table 2.2. Here the survival times are distributed identically in the control and treatment group, but censoring depends on a binary covariate (response / non-response). Responders in the treatment group and non-responders in the control group are not censored at all (except for administrative censoring at $t = 2$ (study end)), while non-responders in the treatment group and responders in the control group are censored with a probability up to 50%. The response rate is 0.5 in both groups. Since responders have a better survival than non-responders, this leads to an overestimation of the survival probability in the treatment group and an underestimation of the survival probability in the control group.

Table 2.2: Simulation scenario with informative censoring

	Control ($n = 200$)[a]		Treatment ($n = 200$)[a]	
	Non-Resp.	Responders	Non-Resp.	Responders
Survival	Weibull(2,1)	Weibull(1.5, 1.5)	Weibull(2,1)	Weibull(1.5, 1.5)
Drop-out	-	Weibull(1.5, 1.5)	Weibull(2,2)	-

[a] Response rate is 0.5.

In 10^5 simulations with a sample size of 200 in each group the log-rank test showed an inflated type-I-error of approx. 32.8%. The bias of the Kaplan-Meier estimators is shown in figure 2.3.

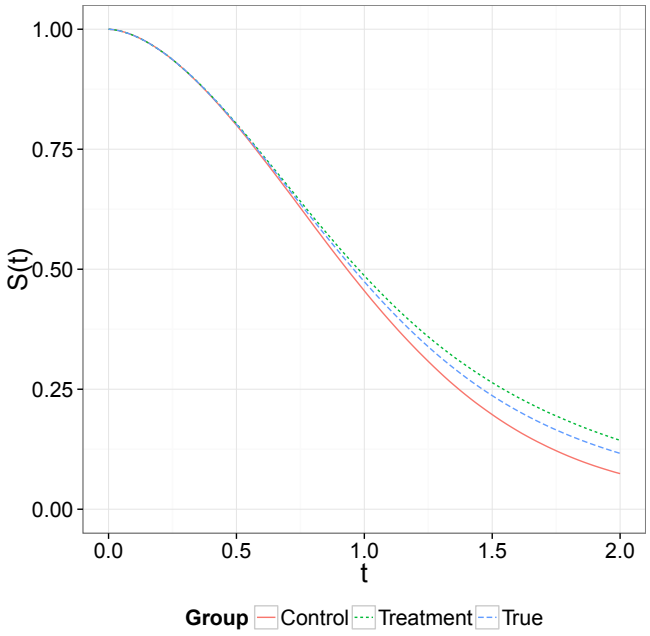

Figure 2.3: Bias of Kaplan-Meier estimators in treatment and control group in dependent censoring example of table 2.2. Mean of 10^5 replications. Sample size was 200 in each group.

Adjustment for covariate dependent censoring is easy within the proportional hazards model. Additional discrete and continuous covariates can be seamlessly included in the model. A stratified proportional hazards model leads to the stratified log-rank test. However, by including additional covariates, only the conditional treatment effect can be estimated. If the covariate is correlated with treatment, as is the case for surrogate variables, then the marginal treatment effect might differ substantially from the conditional treatment effect. Consider the example from above, but this time with a higher response rate in the treatment group than in the control group. Then clearly there is a treatment effect, since there is a higher proportion of responders in the treatment group than in the control group, who have a better survival than the non-responders. However, the stratified log-rank test will fail to detect any treatment effect, since the conditional survival functions are equal in the two groups.

3

FIXED SAMPLE SIZE METHODS

In this chapter the weighted Kaplan-Meier estimator is defined and its large sample properties are derived in a fixed design setting. Then the asymptotic distribution of the estimator of the average hazard ratio based on arbitrary estimators of the marginal survival function, which fulfill certain asymptotic properties, is derived. This generalizes the work of Kalbfleisch and Prentice [1981], who only consider the average ratio based on the Kaplan-Meier estimator. These results form the basis for the sequential and adaptive results in the following chapters. Kalbfleisch and Prentice [1981] give an incorrect formula for the asymptotic variance of the multivariate average hazard ratio. The correct formula is derived and the results are applied to the problem of testing non-inferiority in three-arm trials with survival data. This generalizes the results of Kombrink et al. [2013] to non-proportional hazards settings. The two-sample special case is treated in detail. Moreover, alternative two-sample methods based on the marginal survival functions, such as restricted mean survival and the estimation of the median and other quantiles are described. Finally the finite-sample performance of the two-sample methods is compared in a simulation study.

3.1 WEIGHTED KAPLAN-MEIER ESTIMATOR

In this section the *weighted Kaplan-Meier* (WKM) estimator of Malani [1995] and Murray and Tsiatis [1996] is defined and the necessary large sample properties required for the functional delta method are given in theorem 3.3. This extends the results of Murray and Tsiatis [1996], by proving uniform consistency and weak convergence of the WKM estimator in the space $D[0, L]$ of right-continuous functions with left limits (section A.2.1).

3.1.1 *Definition*

Assume that the survival time T and the censoring time C are stochastically independent given a discrete covariate $X \in \{1, \ldots, J\}$, i.e. T and C are stochastically indepen-

dent in each of the strata defined by the levels of X. The observed data consists of the n independent and identically distributed tuples

$$\{(Y_i = T_i \wedge C_i, \delta_i = \mathbb{1}\{T_i \leqslant C_i\}, X_i), i = 1, \ldots, n\},$$

The sample size in stratum j is

$$n_j = \sum_{i=1}^{n} \mathbb{1}\{X_i = j\},$$

and the total sample size is $n = n_1 + \ldots + n_J$. The *conditional* survival function given $X = j$,

$$S_j(t) = P(T > t|X = j),$$

can be consistently estimated by the stratum-specific Kaplan-Meier estimator $\hat{S}_j$ (see section 2.3). The *marginal* survival function can be written as

$$S(t) = E[P(T > t|X)] = \sum_{j=1}^{J} p_j S_j(t),$$

where $p_j = P(X = j)$ is the response probability for stratum j. This suggests an estimator of the marginal survival function.

Definition 3.1 (Murray and Tsiatis [1996])**.** *The weighted Kaplan-Meier (WKM) estimator for a single time-independent discrete covariate X is defined by*

$$\hat{S}_{WKM}(t) = \sum_{j=1}^{J} \hat{p}_j \hat{S}_j(t), \tag{3.1.1}$$

where $\hat{S}_j$ is the Kaplan-Meier estimator in the stratum defined by covariate value j and

$$\hat{p}_j = \frac{n_j}{n}$$

is the response rate *in stratum j.*

3.1.2 *Large sample properties*

Some technical conditions are required to derive the asymptotic distribution of the WKM estimator. The support of the censoring distribution imposes a limit on the

time range on which the survival function can be consistently estimated. Let $K_j(x) = P(C > x|X = j)$ for $x \geqslant 0$ and $j = 1, \ldots, J$. Then each $\hat{S}_j$ is only consistent on any interval $[0, L_j]$, where

$$L_j < \sup\{x : S_j(x)K_j(x) > 0\}.$$

Consequently, consistency of the WKM estimator is restricted to the interval $[0, L]$, where $L \leqslant \min\{L_1, \ldots L_J\}$. Furthermore, the number at risk $\bar{Y}_j^{(n)}(s)$ at any time $s \in [0, L]$ must diverge to infinity. These two conditions ensure, that $n^{-1}\bar{Y}_j^{(n)}(s)$ converges in probability to $p_j S_j(s)K_j(s) > 0$ for each $s \in [0, L]$. The following assumptions are required for theorem 3.3 below:

Assumption 3.2. *For* $j = 1, \ldots, J$ *and* $i = 1, \ldots, n$,

1. $p_j > 0$,
2. T_i *is stochastically independent of* C_i *given* X_i,
3. $L < \sup\{x : S_j(x)K_j(x) > 0\}$,
4. $\bar{Y}_j^{(n)}(s) \to \infty$, *as* $n \to \infty$ *for all* $s \in [0, L]$.

Theorem 3.3 states the result on the asymptotic distribution of the weighted Kaplan-Meier estimator and is a special case of the sequential result (theorem 4.10), which is proved in section 4.2.

Theorem 3.3. *Under assumption 3.2,*

1. $\hat{S}_{WKM}$ *is a uniformly consistent estimator of* S,
2. $\sqrt{n}\{\hat{S}_{WKM} - S\}$ *converges weakly in* $D[0, L]$ *to a mean-zero Gaussian process with covariance function* ρ *given by*

$$\begin{aligned}\rho_{WKM}(s,t) = \sum_{j=1}^{J} p_j S_j(s)S_j(t) \int_0^{s\wedge t} \frac{\lambda_j(u)du}{S_j(u)K_j(u)} \\ + \sum_{j=1}^{J} p_j S_j(s)S_j(t) - \sum_{j=1}^{J}\sum_{l=1}^{J} p_j p_l S_j(s)S_l(t),\end{aligned} \tag{3.1.2}$$

3. *a uniformly consistent estimator of the covariance function* ρ_{WKM} *is given by*

$$\hat{\rho}_{WKM}(s,t) = \sum_{j=1}^{J} \hat{p}_j \hat{S}_j(s)\hat{S}_j(t) n \int_0^{s\wedge t} \frac{J_j^{(n)}(t,u)\bar{N}_j^{(n)}(du)}{\bar{Y}_j^{(n)}(u)^2} + \sum_{j=1}^{J} \hat{p}_j \hat{S}_j(s)\hat{S}_j(t) - \sum_{j=1}^{J}\sum_{l=1}^{J} \hat{p}_j \hat{p}_l \hat{S}_j(s)\hat{S}_l(t),$$

where $J_j^{(n)}$, $\bar{N}_j^{(n)}$ *and* $\bar{Y}_j^{(n)}$ *are defined in section 2.3.*

Asymptotic relative efficiency

The asymptotic covariance function of the Kaplan-Meier estimator is

$$\rho_{KM}(s,t) = S(s)S(t)\int_0^{s\wedge t} \frac{\lambda(u)du}{S(u)K(u)}, \tag{3.1.3}$$

where $S(u) = \sum_{j=1}^{J} p_j S_j(u)$ and $K(u) = \sum_{j=1}^{J} p_j K_j(u)$ and $\lambda(u) = -\frac{d}{du}(\log S(u))$.

As noted by Malani [1995], the weighted Kaplan-Meier (WKM) estimator is more efficient, i.e. it has a smaller asymptotic variance, than the Kaplan-Meier (KM) estimator, whenever the stratification variable is *not* stochastically independent of the survival time. There are four possible ways how the stratification variable is related to the distributions of the survival and censoring times:

1. If $S_j = S$ and $K_j = K$ for all $j = 1, \ldots, J$, then $\rho_{WKM}(t,t) = \rho_{KM}(t,t)$ for all t.
2. If $S_j = S$ for all j, but $K_j \neq K$ for some j, then $\rho_{WKM}(t,t) > \rho_{KM}(t,t)$ for all t.
3. If $S_j \neq S$ for some j and $K_j = K$ for all j, then $\rho_{WKM}(t,t) < \rho_{KM}(t,t)$ for all t.
4. If $S_j \neq S$ and $K_j \neq K$ for all j, then censoring is informative and the KM estimator is not valid.

It easy to see, that the equality in (1.) actually holds for all pairs (s,t), i.e. $\rho_{WKM}(s,t) = \rho_{KM}(s,t)$ for all s,t. Numerical calculations suggest, that the inequalities in (2.) and (3.) also hold uniformly for all pairs (s,t). The WKM estimator is only worse than the KM estimator in cases where the stratification variable is independent from the survival time and at the same time related to the censoring time. It is safe to assume, that if a surrogate variable is used for stratification, then either the third or fourth case is valid. The interpretation for the efficiency gain in the third case is, that some

of the information lost due to censoring can be recovered by incorporating covariate information related to the survival endpoint. Thus it is expected, that the efficiency gains are the largest, when the proportion of censored observations is huge. Without any censoring, there is no difference between the stratified and the unstratified case, no matter how the stratification variable is related to the primary endpoint. For uncensored observations, the Kaplan-Meier estimator reduces to 1 minus the empirical distribution function. Suppose the observations $Y_1, \ldots, Y_n$ are uncensored, then the stratified and the unstratified estimates are identical:

$$\begin{aligned}\hat{S}_{WKM}(t) &= \sum_{j=1}^{J} \frac{n_j}{n} \left(\frac{1}{n_j} \sum_{i=1}^{n} \mathbb{1}\{Y_i > t, X_i = j\} \right) \\ &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{Y_i > t\} \underbrace{\sum_{j=1}^{J} \mathbb{1}\{X_i = j\}}_{=1} = \hat{S}_{KM}(t)\end{aligned}$$

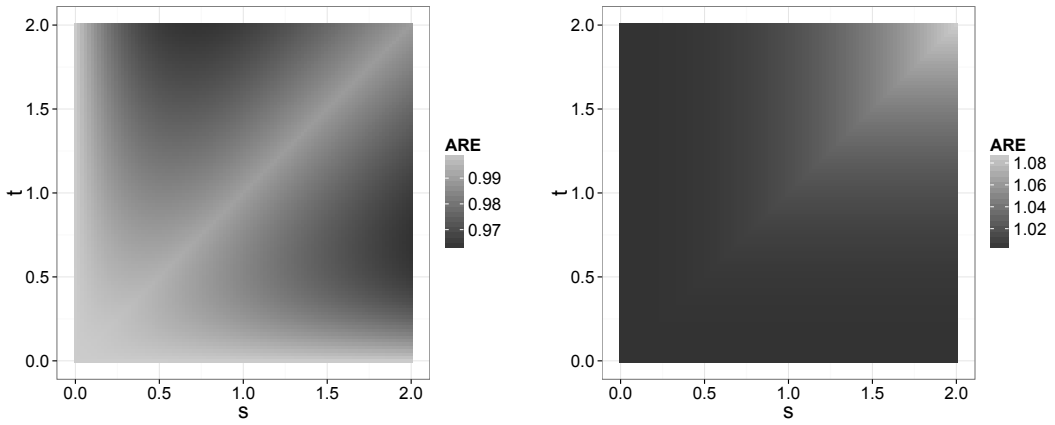

Figure 3.1: Asymptotic relative efficiency (ARE) of the weighted Kaplan-Meier estimator compared to the Kaplan-Meier estimator for exponential survival times and four strata. Left: Stratification independent of the censoring time. Right: Stratification independent of the survival time.

With censoring the reduction in the size of the asymptotic variance is small except in extreme cases. Consider for example a situation with four strata, each with response probability 0.25, with exponential survival times with means 1, 1.25, 1.5

and 1.75 respectively. Censoring times are exponential with mean 0.25 in all strata, resulting in $70-80\%$ censoring. The ratio

$$\frac{\rho_{WKM}(s,t)}{\rho_{KM}(s,t)}$$

was calculated for values of s and t ranging between 0 and 2. The result is shown in figure 3.1 (left panel). The reduction in the size of the asymptotic variance is at most 4%. Now reversing the roles of the censoring and survival times, i.e. the censoring times are now exponentially distributed with means 1, 1.25, 1.5 and 1.75, respectively, and the survival times are exponentially distributed with mean 0.25, results in figure 3.1 (right panel).

The use of the WKM estimator is still justified in the non-adaptive setting, since it can also reduce bias from informative censoring, when the stratification variable is also related to the censoring time.

3.1.3 *Missing surrogate values*

In the construction of the WKM estimate it was assumed, that the value of the surrogate variable is available immediately after treatment for all patients, i.e. at time $t = 0$. More realistically, the response status will only be known after a certain delay or is missing, e.g. assessement of tumor response after a fixed time span, and some patients die before. Ignoring the delay might result in a so-called *time-dependent bias* (Wolkewitz et al. [2012]) as illustrated in the following simulation. The scenario was that of table 2.2 in section 2.5 with an important modification. For about 50% of the responders, the response was set to missing, simulating patients, whose response status was missing at the analysis time. Censoring for patients with missing response status was non-informative. Clearly, the missingnes of the response variable depends on the unobserved value of the response variable. Three different strategies for handling the missing data with the WKM estimator were examined: Simply ignoring patients with missing surrogate variable, i.e. a complete case analysis (*CC*), treating all patients with missing surrogate variable as non-responders (*as.NR*), and adding a new surrogate variable category containing all patients with missing surrogate variable (*new.cat*).

Figure 3.2 clearly shows the substantial bias of the WKM estimates if patients with missing surrogate variable values are excluded from analysis or treated as non-

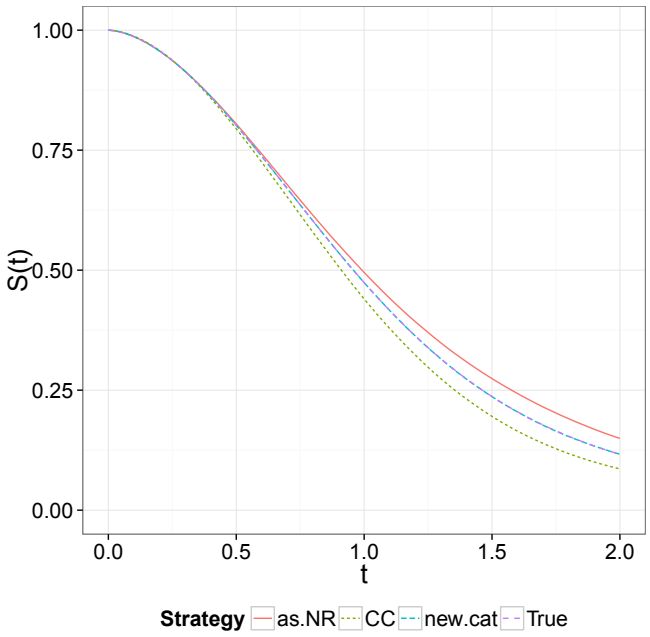

Figure 3.2: Bias of WKM estimator caused by missing surrogate. Mean of 10^5 replications. Sample size was 200.

responders. The curve corresponding to the WKM estimator with the additional stratum is indistinguishable from the true survival curve. This approach only works, if censoring for those patients with missing surrogate variable is non-informative, i.e. censoring does not depend on the unobserved value of the surrogate variable. This is plausible in the adaptive design setting, where informative censoring is only induced through the adaptive design changes, by using surrogate variable information in the interim decision. The censoring in the group of patients with missing surrogate variable is non-informative, since their surrogate variable information cannot possibly have been used in any interim decision. Note that using the information, that the surrogate variable was missing, in the interim decision does not induce informative censoring. In face of this simple strategy of handling missing surrogate variable data it is justified to assume for simplicity, that all surrogate variable data is available at time $t = 0$.

3.2 MULTIVARIATE AVERAGE HAZARD RATIO

In this section the multivariate average hazard ratio as introduced by Kalbfleisch and Prentice [1981] is defined and its large sample properties are derived based on $k+1$ sets of independent right-censored continuous failure time data. Denote the sample size of group i by n_i, the hazard rate by λ_i and the survival function by S_i, $i = 0, \ldots, k$. The total sample size is then $n = n_0 + \ldots + n_k$. Note that the existence of a hazard rate λ_i implies, that S_i is continuous. The observed data is

$$\{(Y_l = T_l \wedge C_l, \delta_l = \mathbb{1}\{T_l \leqslant C_l\}, X_l, Z_l), l = 1, \ldots, n\},$$

where $X_l \in \{1, \ldots, J\}$ is the surrogate and $Z_l \in \{0, \ldots, k\}$ is the group indicator.

3.2.1 *Average hazard ratio*

The multivariate average hazard ratio is defined by Kalbfleisch and Prentice [1981] as a weighted time-average of a hazard ratio.

Definition 3.4 (average sample i to total hazard ratio)**.** *Given a survival function* G *define the average sample* i *to 'total' hazard ratio as*

$$\theta_i(G) = -\int_0^\infty \frac{\lambda_i(t)}{\sum_{j=0}^k \lambda_j(t)} G(dt). \tag{3.2.1}$$

A natural choice for the weight function is the product of all survival functions

$$G(t) = \prod_{i=0}^k S_i(t)^\alpha,$$

where $\alpha > 0$ is a constant, which controls the weight given to the tails of the survival curve. For $\alpha = 1$, this leads to the relative risk

$$\theta_i(G) = P(\min\{T_0, \ldots, T_{i-1}, T_{i+1}, \ldots, T_k\} > T_i)$$

(see lemma 3.6). As mentioned in section 3.1.2, the true survival function can only be consistently estimated up to a finite time point $L < \infty$ depending on the support of the survival and censoring times (see assumption 3.2). Therefore the weight function G is truncated at a fixed point L.

Definition 3.5 (weight function). *Let* $\alpha > 0$ *and* $L > 0$. *For given survival functions* S_i, $i = 0, \dots, k$ *define*

$$G(t) = \prod_{i=0}^{k} S_i^{\alpha}(t),$$

and

$$\tilde{G}(t) = \frac{G(t)}{1 - G(L)} \mathbb{1}\{t \leqslant L\}.$$

The normalization factor $1 - G(L)$ in the definition of $\tilde{G}$ ensures, that

$$\sum_{i=0}^{k} \theta_i(\tilde{G}) = -\int_0^{\infty} \tilde{G}(dt) = 1.$$

Lemma 3.6.

$$\theta_i \equiv \theta_i(\tilde{G}) = -\int_0^{\infty} \frac{\lambda_i(t)}{\sum_{j=0}^{k} \lambda_j(t)} \tilde{G}(dt) = \frac{\varkappa_i}{1 - G(L)},$$

where

$$\varkappa_i = -\int_0^{L} G(t) \log S_i(dt) = -\int_0^{L} \xi_i(t) S_i(dt).$$

and

$$\xi_i(t) = \prod_{j \neq i} S_j(t) = \frac{G(t)}{S_i(t)}.$$

Proof.

$$G(dt) = G(t) \log G(dt) = G(t) \sum_{j=0}^{k} \log S_j(dt) = -G(t) \sum_{j=0}^{k} \lambda_j(t) dt,$$

and

$$\xi_i(t) S_i(dt) = G(t) \log S_i(dt) = -\lambda_i(t) G(t) dt.$$

Hence

$$\theta_i(\tilde{G}) = \frac{1}{1 - G(L)} \int_0^{L} \lambda_i(t) G(t) dt = -\frac{1}{1 - G(L)} \int_0^{L} \xi_i(t) S_i(dt).$$

□

Remark 3.7. *The value of* L *is fixed throughout this and the following chapters. Therefore the dependence on* L *of the average hazard ratio and other quantities is omitted in the notation.*

3.2.2 *Choice of the weight function*

The shape of the weight function in definition 3.5 is controlled by the parameter α.

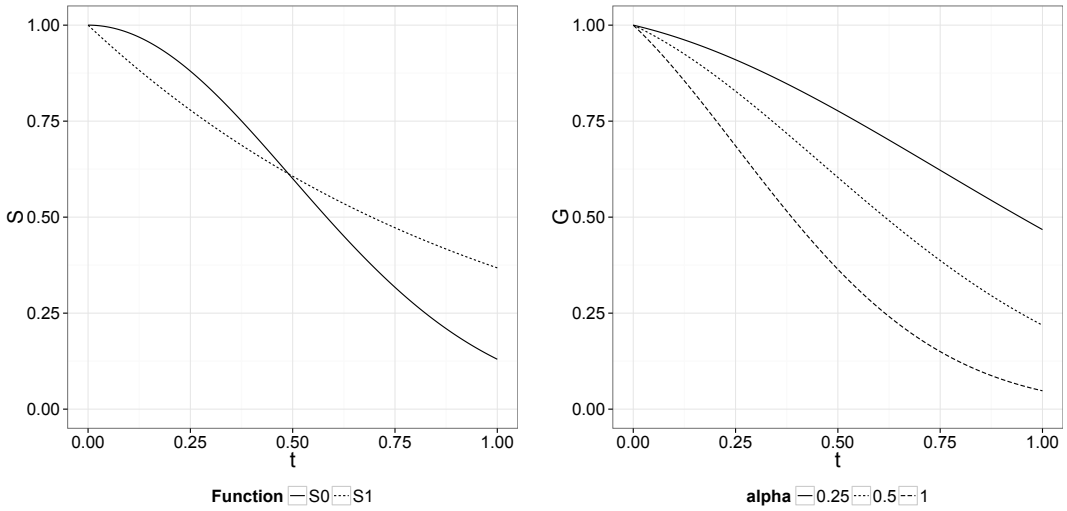

Figure 3.3: Left: Survival functions S_0 (= Weibull(2, 0.7)) and S_1 (= Exp(1)). Right: Weight function $G = S_0^\alpha S_1^\alpha$ for $\alpha \in \{0.25, 0.5, 1\}$.

The performance of the average hazard ratio based tests depends critically on the choice of the weight function, i.e. the parameter α. Figure 3.3 gives an example in the two-sample case. The left panel shows the two survival functions. The right panel shows the weight function for three different values of α. Figure 3.4 shows the values of the average hazard ratio $\theta_1(1)$ for different values of α between 0 and 2. The effect size $\theta_1(1) - 0.5$ becomes smaller with α getting larger, and it even changes sign at $\alpha = 1.5$.

In principle, weight functions, which are not of the form as in definition 3.5 are possible, but would require modification of the proofs of the large sample properties of the average hazard ratio. Optimal choice of the weight function, of course, requires knowledge of the true shape of the survival curves. For simplicity and comparability, in all the simulations the truncated weight function from definition 3.5 was used with $\alpha = 1$.

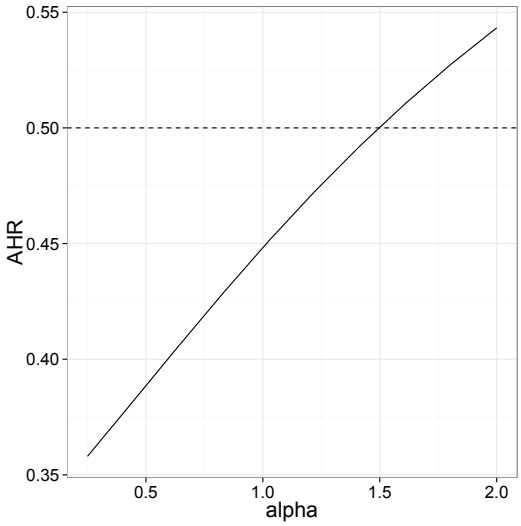

Figure 3.4: Two-sample average hazard ratio for different weight functions $G = S_0^{\alpha} S_1^{\alpha}$.

3.2.3 *Estimation of the average hazard ratios*

According to lemma 3.6, the parameters θ_i are functions of the survival functions $S_0, \ldots, S_k$, if the weight function from definition 3.5 is used. The estimation of the average hazard ratios is based on estimators of the marginal survival function in each sample. Any estimators can be used as long as the following assumptions hold:

Assumption 3.8. *There exists* $L > 0$, *such that for* $i = 0, 1, \ldots, k$,

1. $n_i/n \to \nu_i > 0$, *as* $n \to \infty$,
2. $\hat{S}_i$ *is a uniformly consistent estimator of* S_i *on* $[0, L]$,
3. $\sqrt{n_i}\{\hat{S}_i - S_i\}$ *converges weakly in* $D[0, L]$ *to a mean-zero Gaussian process with covariance function* ρ_i,
4. *there exists a uniformly consistent estimator* $\hat{\rho}_i$ *of the covariance function* ρ_i.

Assumption 3.8 holds, e.g. for the Kaplan-Meier estimator and the weighted Kaplan-Meier estimator (theorem 3.3).

Remark 3.9. *The estimators* $\hat{S}_i$, $i = 0, 1, \ldots, k$ *are always considered as random elements of the space* $D[0, L]$.

Remark 3.10. *The dependence on the parameter α in the weight function 3.5 is not made explicit in the rest of this chapter, since if assumption 3.8 holds for $\hat{S}_i$ and S_i it also holds for $\hat{S}_i^\alpha$ and S_i^α for any $\alpha > 0$.*

From assumption 3.8 and the independence of the samples it follows, that

$$\left(\sqrt{n}\{\hat{S}_i - S_i\}\right)_{0 \leqslant i \leqslant k} \xrightarrow{\mathcal{L}} \left(U_i\right)_{0 \leqslant i \leqslant k}$$

as $n \to \infty$, where U_i, $i = 0, \ldots, k$ are independent mean-zero Gaussian processs with covariance functions $\nu_i^{-1}\rho_i$. In the closed-form expressions for the asymptotic variance, it is convenient to use the asymptotic covariance function $\rho_i^{\log}$ of $\sqrt{n_i}\{\log \hat{S}_i - \log S_i\}$ instead of the covariance function ρ_i of $\sqrt{n_i}\{\hat{S}_i - S_i\}$. The relationship between ρ_i and $\rho_i^{\log}$ is

$$\rho_i(s,t) = S_i(s)S_i(t)\rho_i^{\log}(s,t),$$

because

$$\sqrt{n_i}\{\hat{S}_i - S_i\} = \sqrt{n_i}\{e^{\log \hat{S}_i} - e^{\log S_i}\} = e^{\log S_i}\sqrt{n_i}\{\log \hat{S}_i - \log S_i\} + o_p(1),$$

by the functional delta method (theorem A.50). Moreover, the asymptotic covariance function of $\sqrt{n}\{\hat{S}_i - S_i\}$ is $\nu_i^{-1}\rho_i$ and the asymptotic covariance function of $\sqrt{n}\{\log \hat{S}_i - \log S_i\}$ is $\nu_i^{-1}\rho_i^{\log}$. Estimators of the quantities ξ_i, G, $\varkappa_i$ and θ_i are obtained by replacing the survival functions S_i with their estimators $\hat{S}_i$.

Definition 3.11. *For any estimators $\hat{S}_i$ of S_i, $i = 0, \ldots, k$, define*

$$\hat{\xi}_i = \prod_{j \neq i} \hat{S}_j,$$

$$\hat{G}(L) = \prod_{i=0}^{k} \hat{S}_i(L),$$

$$\hat{\varkappa}_i = -\int_0^L \hat{\xi}_i(t)\hat{S}_i(dt),$$

$$\hat{\theta}_i = \frac{\hat{\varkappa}_i}{1 - \hat{G}(L)}.$$

The next lemma proves asymptotic expansions of the estimators $\hat{\xi}_i$, $\hat{G}(L)$ and $\hat{\varkappa}_i$, $i = 0, \ldots, k$, which are a key step in the proof of the main result of this chapter (theorem 3.15).

Lemma 3.12. *Under assumption 3.8, for* $i = 0, \dots, k$,

1.
$$\sqrt{n}\{\hat{\xi}_i - \xi_i\} = \xi_i \sum_{j \neq i} \frac{1}{S_j} \sqrt{n}\{\hat{S}_j - S_j\} + o_p(1),$$

2.
$$\sqrt{n}\{\hat{G}(L) - G(L)\} = \sum_{i=0}^{k} \xi_i(L) \sqrt{n}\{\hat{S}_i(L) - S_i(L)\} + o_p(1).$$

3.
$$\begin{aligned}\sqrt{n}\{\hat{x}_i - x_i\} = &\sum_{j \neq i} \int_0^L \sqrt{n}\{\hat{S}_i - S_i\} \xi_i \, d \log S_j \\ &- \sum_{j \neq i} \int_0^L \frac{\xi_i}{S_j} \sqrt{n}\{\hat{S}_j - S_j\} dS_i \\ &- \sqrt{n}\{\hat{S}_i(L) - S_i(L)\} \xi_i(L) + o_p(1)\end{aligned}$$

Proof. Let $\phi : B \to D[0, L]$ be the map defined by $S \mapsto \log S$, where $B \subset D[0, L]$ is the subset of functions in $D[0, L]$, which are strictly positive everywhere. Let $\psi : \mathbb{R}^+ \to \mathbb{R}$ be the function defined by $x \mapsto \log x$ for $x > 0$. Remember that $\hat{\xi}_i$ and $\hat{S}_i$ are random elements in the space $D[0, L]$, and are in B with probability tending to 1, since the truncation point L was chosen, such that $S_i(L) > 0$ for $i = 0, \dots, k$.

1. Twice application of the functional delta method with the map ϕ and its inverse ϕ^{-1},
$$\begin{aligned}\sqrt{n}\{\hat{\xi}_i - \xi_i\} &= \xi_i \sqrt{n}\{\log \hat{\xi}_i - \log \xi_i\} + o_p(1) \\ &= \xi_i \sum_{j \neq i} \sqrt{n}\{\log \hat{S}_i - \log S_i\} + o_p(1) \\ &= \xi_i \sum_{j \neq i} \frac{1}{S_i} \sqrt{n}\{\hat{S}_i - S_i\} + o_p(1).\end{aligned} \tag{3.2.2}$$

2. Again twice application of the delta method, this time with the map ψ and its inverse ψ^{-1},

$$\begin{aligned}\sqrt{n}\{\hat{G}(L) - G(L)\} &= G(L)\sqrt{n}\{\log \hat{G}(L) - \log G(L)\} + o_p(1) \\ &= G(L)\sum_{i=0}^{k}\sqrt{n}\{\log \hat{S}_i(L) - \log S_i(L)\} + o_p(1) \\ &= \sum_{i=0}^{k}\xi_i(L)\sqrt{n}\{\hat{S}_i(L) - S_i(L)\} + o_p(1).\end{aligned}$$

3. From the definitions of $\hat{\xi}_i$ and ξ_i it follows

$$\sqrt{n}\{\hat{x}_i - x_i\} = -\int_0^L \sqrt{n}\hat{\xi}_i(\cdot-)d\hat{S}_i + \int_0^L \sqrt{n}\xi_i dS_i,$$

where $\hat{\xi}_i(\cdot-)$ denotes the limit from the left, i.e. $\hat{\xi}_i(t-) = \lim_{s\uparrow t}\hat{\xi}_i(s)$. Adding and substracting the term $\sqrt{n}\int_0^L \xi_i d\hat{S}_i$ at the right hand side gives

$$\sqrt{n}\{\hat{x}_i - x_i\} = -\int_0^L \sqrt{n}\{\hat{\xi}_i(\cdot-) - \xi_i\}d\hat{S}_i - \int_0^L \xi_i d\sqrt{n}\{\hat{S}_i - S_i\}$$

Integration by parts (theorem A.28) of the second term on the right hand side,

$$\begin{aligned}\sqrt{n}\{\hat{x}_i - x_i\} = &-\int_0^L \sqrt{n}\{\hat{\xi}_i(\cdot-) - \xi_i\}d\hat{S}_i + \int_0^L \sqrt{n}\{\hat{S}_i(\cdot-) - S_i\}d\xi_i \\ &- \sqrt{n}\{\hat{S}_i(L) - S_i(L)\}\xi_i(L) - \left[\sqrt{n}\{\hat{S}_i - S_i\}, \xi_i\right](L).\end{aligned} \tag{3.2.3}$$

The last term on the right hand side vanishes, by lemma A.25, because the stochastic process ξ_i is of bounded variation (since it is monotone decreasing a.s.) and continuous. Adding and substracting the term

$$\int_0^L \sqrt{n}\{\hat{\xi}_i(\cdot-) - \xi_i\}dS_i$$

gives

$$\begin{aligned}\sqrt{n}\{\hat{x}_i - x_i\} = &\int_0^L \sqrt{n}\{\hat{S}_i(\cdot-) - S_i\}d\xi_i - \sqrt{n}\{\hat{S}_i(L) - S_i(L)\}\xi_i(L) \\ &- \int_0^L \sqrt{n}\{\hat{\xi}_i(\cdot-) - \xi_i\}dS_i - \int_0^L \sqrt{n}\{\hat{\xi}_i(\cdot-) - \xi_i\}d\{\hat{S}_i - S_i\}.\end{aligned} \tag{3.2.4}$$

The process $\sqrt{n}\{\hat{\xi}_i - \xi_i\}$ converges weakly, by (1.), and is independent of the process $\sqrt{n}\{\hat{S}_i - S_i\}$, such that the two processes converge jointly. By lemma A.49(2.) and the functional delta method, the sequence of (real-valued) random variables

$$\int_0^L \sqrt{n}\{\hat{\xi}_i(\cdot-) - \xi_i\} d(\sqrt{n}\{\hat{S}_i - S_i\})$$

converges in distribution and is therefore tight, i.e. stochastically bounded.This implies, that the last term on the right hand side in eq. (3.2.4) converges in probability to 0. Hence

$$\begin{aligned}\sqrt{n}\{\hat{x}_i - x_i\} = &\int_0^L \sqrt{n}\{\hat{S}_i(\cdot-) - S_i\} d\xi_i - \int_0^L \sqrt{n}\{\hat{\xi}_i(\cdot-) - \xi_i\} dS_i \\ &- \sqrt{n}\{\hat{S}_i(L) - S_i(L)\}\xi_i(L) + o_p(1)\end{aligned}$$

Now use

$$d\xi_i = \xi_i \sum_{j \neq i} d \log S_j$$

for the first term on the right hand side, and the result from (1.) for the second term on the right hand side to obtain

$$\begin{aligned}\sqrt{n}\{\hat{x}_i - x_i\} = &\sum_{j \neq i} \int_0^L \sqrt{n}\{\hat{S}_i(\cdot-) - S_i\}\xi_i \, d \log S_j - \sum_{j \neq i} \int_0^L \frac{\xi_i}{S_j} \sqrt{n}\{\hat{S}_j(\cdot-) - S_j\} dS_i \\ &- \sqrt{n}\{\hat{S}_i(L) - S_i(L)\}\xi_i(L) + o_p(1).\end{aligned}$$

□

The following abbreviations are used in the expression for the asymptotic variance in the next theorem:

Definition 3.13 (Variance components). *For* $0 \leqslant i, j \leqslant k$, $i \neq j$,

$$\begin{aligned}
V_x^{(i)} &= V_x^{(i,i)} = \sum_{j\neq i}\sum_{l\neq i} B_{ijl} + \sum_{j\neq i} B_{jii} + C_i - 2\sum_{j\neq i} A_{ij} \\
V_x^{(i,j)} &= -\sum_{l\neq i} B_{ilj} - \sum_{l\neq j} B_{jli} + \sum_{l\neq i,j} B_{lij} + A_{ij} + A_{ji} \\
V_{xG}^{(i)} &= \sum_{j\neq i} (A_{ij} - A_{ji}) - C_i \\
V_G &= \sum_{i=0}^{k} C_i
\end{aligned}$$

where

$$\begin{aligned}
A_{ij} &= \nu_i^{-1} G(L) \int_0^L \rho_i^{\log}(t, L) G(t) \mathrm{d} \log S_j(t) \\
B_{ijl} &= \nu_i^{-1} \int_0^L \int_0^L \rho_i^{\log}(s, t) G(s) G(t) \mathrm{d} \log S_j(s) \mathrm{d} \log S_l(t) \\
C_i &= \nu_i^{-1} G^2(L) \rho_i^{\log}(L, L)
\end{aligned}$$

The next lemma contains some rather tedious covariance calculations, which are used in the derivation of the asymptotic covariance in theorem 3.15.

Lemma 3.14. *For* $i = 0, \ldots, k$ *let*

$$f_i = \sum_{l\neq i} \int_0^L X_i \xi_i \mathrm{d} \log S_l - \sum_{l\neq i} \int_0^L \frac{\xi_i}{S_l} X_l \mathrm{d} S_i - X_i(L)\xi_i(L),$$

and

$$g = \sum_{i=0}^{k} \xi_i(L) X_i(L),$$

where $X_0, \ldots, X_k$ *are independent mean zero Gaussian processes with covariance functions* $\nu_l^{-1}\rho_l$, $l = 0, \ldots, k$. *Then*

1. $\mathrm{Cov}(f_i, f_j) = V_x^{(i,j)}$ *for* $i \neq j$,
2. $\mathrm{Var}(f_i) = V_x^{(i)}$,
3. $\mathrm{Cov}(f_i, g) = V_{xG}^{(i)}$.

Proof. 1. Note that independence of X_l, $l = 0, \dots, k$, together with $E[X_l(s)] = 0$ for all $s \geqslant 0$, $l = 0, \dots, k$, implies $E[X_i(s)X_j(t)] = 0$ for $i \neq j$ for all s, t. Then

$$\begin{aligned}
\operatorname{Cov}(f_i, f_j) = &- E\left[\sum_{l\neq i}\sum_{h\neq j}\int_0^L\int_0^L X_i(s)\xi_i(s)\frac{\xi_j(t)}{S_h(t)}X_h(t)\,d\log S_l(s)\,dS_j(t)\right] \\
&- E\left[\sum_{l\neq i}\sum_{h\neq j}\int_0^L\int_0^L X_j(t)\xi_j(t)\frac{\xi_i(s)}{S_l(s)}X_l(s)\,dS_i(s)\,d\log S_h(t)\right] \\
&+ E\left[\sum_{l\neq i}\int_0^L \frac{\xi_i(s)}{S_l(s)}X_l(s)X_j(L)\xi_j(L)\,dS_i(s)\right] \\
&+ E\left[\sum_{l\neq j}\int_0^L \frac{\xi_j(s)}{S_l(s)}X_l(s)X_i(L)\xi_i(L)\,dS_j(s)\right] \\
&+ E\left[\sum_{l\neq i,j}\int_0^L\int_0^L \frac{\xi_i(s)\xi_j(t)}{S_l(s)S_l(t)}X_l(s)X_l(t)\,dS_i(s)\,dS_j(t)\right]
\end{aligned}$$

a) First term:

$$\begin{aligned}
&- E\left[\sum_{l\neq i}\sum_{h\neq j}\int_0^L\int_0^L X_i(s)\xi_i(s)\frac{\xi_j(t)}{S_h(t)}X_h(t)\,d\log S_l(s)\,dS_j(t)\right] \\
&= -\sum_{l\neq i}\nu_i^{-1}\int_0^L\int_0^L \rho_i(s,t)\xi_i(s)\frac{\xi_j(t)}{S_i(t)}\,d\log S_l(s)\,dS_j(t) \\
&= -\sum_{l\neq i}\nu_i^{-1}\int_0^L\int_0^L \rho_i^{\log}(s,t)G(s)G(t)\,d\log S_l(s)\,d\log S_j(t) \\
&= -\sum_{l\neq i} B_{ilj}
\end{aligned}$$

b) Second term: Like first term, equals $-\sum_{l\neq j} B_{jli}$.

c) Third term:

$$\begin{aligned}
&E\left[\sum_{l\neq i}\int_0^L \frac{\xi_i(s)}{S_l(s)}X_l(s)X_j(L)\xi_j(L)dS_i(s)\right]\\
&=\nu_j^{-1}\int_0^L \rho_j(s,L)\frac{\xi_i(s)}{S_j(s)}\xi_j(L)dS_i(s)\\
&=\nu_j^{-1}\int_0^L \rho_j^{\log}(s,L)G(s)G(L)d\log S_i(s)\\
&=A_{ji}
\end{aligned}$$

d) Fourth term: Like third term, equals A_{ij}.

e) Fifth term:

$$\begin{aligned}
&E\left[\sum_{l\neq i,j}\int_0^L\int_0^L \frac{\xi_i(s)\xi_j(t)}{S_l(s)S_l(t)}X_l(s)X_l(t)dS_i(s)dS_j(t)\right]\\
&=\sum_{l\neq i,j}\nu_l^{-1}\int_0^L\int_0^L \frac{\xi_i(s)\xi_j(t)}{S_l(s)S_l(t)}\rho_l(s,t)dS_i(s)dS_j(t)\\
&=\sum_{l\neq i,j}\nu_l^{-1}\int_0^L\int_0^L G(s)G(t)\rho_l^{\log}(s,t)d\log S_i(s)d\log S_j(t)\\
&=\sum_{l\neq i,j}B_{lij}
\end{aligned}$$

Hence $Cov(f_i,f_j)=V_x^{(i,j)}$.

2.

$$\begin{aligned}
Var(f_i) =& E[f_i^2] \\
=& E\left[\sum_{j\neq i}\sum_{l\neq i}\int_0^L\int_0^L X_i(s)X_i(t)\xi_i(s)\xi_i(t)d\log S_j(s)d\log S_l(t)\right] \\
&+ E\left[\sum_{l\neq i}\int_0^L\int_0^L \frac{\xi_i(s)\xi_i(t)}{S_l(s)S_l(t)}X_l(s)X_l(t)dS_i(s)dS_i(t)\right] \\
&- 2\sum_{j\neq i} E\left[\int_0^L X_i(s)\frac{\xi_i(s)}{S_j(s)}X_i(L)\xi_i(L)dS_j(s)\right] \\
&+ E[X_i(L)^2]\xi_i(L)^2 \\
=& \sum_{j\neq i}\sum_{l\neq i} B_{ijl} + \sum_{j\neq i} B_{jii} - 2\sum_{j\neq i} A_{ij} + C_i
\end{aligned}$$

3.

$$\begin{aligned}
Cov(f_i, g) =& E\left[\sum_{j\neq i}\int_0^L X_i(s)X_i(L)\xi_i(s)\xi_i(L)d\log S_j(s)\right] \\
&- E\left[\sum_{j\neq i}\int_0^L \frac{\xi_i(s)}{S_j(s)}\xi_j(L)X_j(L)X_j(s)dS_i(s)\right] - \rho_i(L,L)\xi_i(L)^2 \\
=& \sum_{j\neq i} A_{ij} - \sum_{j\neq i} A_{ji} - C_i = V_{xG}^{(i)}
\end{aligned}$$

□

The next theorem gives the asymptotic distribution of the k-vector $\sqrt{n}\{\bar{\theta}^{(n)} - \bar{\theta}\}$ of average hazard ratios, where $\bar{\theta}^{(n)} = (\hat{\theta}_0, \ldots, \hat{\theta}_k)^T$ and $\bar{\theta} = (\theta_0, \ldots, \theta_k)^T$.

Theorem 3.15. *Under assumption 3.8, as* $n \to \infty$,

$$\sqrt{n}\{\bar{\theta}^{(n)} - \bar{\theta}\} \xrightarrow{\mathcal{L}} N(0, \Sigma_\theta),$$

where Σ_θ *is the* $(k+1)\times(k+1)$*-matrix with entries*

$$\sigma_{i+1,j+1} = \frac{1}{\{1-G(L)\}^2}\{V_x^{(i,j)} + \theta_j V_{xG}^{(i)} + \theta_i V_{xG}^{(j)} + \theta_i\theta_j V_G\} \quad \textit{for } 0 \leqslant i, j \leqslant k$$

The covariance matrix Σ_θ can be estimated consistently by replacing ν_i, S_i and ρ_i by their consistent estimators n_i/n, $\hat{S}_i$ and $\hat{\rho}_i$.

Proof. Define the linear map $\phi : D[0,L]^{k+1} \to \mathbb{R}^{k+2}$ by

$$\begin{pmatrix} X_0 \\ \vdots \\ X_k \end{pmatrix} \mapsto \begin{pmatrix} \sum_{j\neq 0} \int_0^L X_0 Z_0 \, d\log S_j - \sum_{j\neq 0} \int_0^L \frac{Z_i}{S_j} X_j \, dS_0 - X_0(L)Z_0(L) \\ \vdots \\ \sum_{j\neq k} \int_0^L X_k Z_k \, d\log S_j - \sum_{j\neq k} \int_0^L \frac{Z_i}{S_j} X_j \, dS_k - X_k(L)Z_k(L) \\ \sum_{i=0}^k Z_i(L)X_i(L) \end{pmatrix}$$

From the continuous mapping theorem together with lemma 3.12 it follows

$$\begin{aligned} \begin{pmatrix} \sqrt{n}\{\hat{x}_0 - x_0\} \\ \vdots \\ \sqrt{n}\{\hat{x}_k - x_k\} \\ \sqrt{n}\{\hat{G}(L) - G(L)\} \end{pmatrix} &= \sqrt{n}\{\phi(\hat{S}_0, \ldots, \hat{S}_k) - \phi(S_0, \ldots, S_k)\} + o_p(1) \\ &\xrightarrow{\mathcal{L}} \phi(U_0, \ldots, U_k) \end{aligned}$$

Since ϕ is linear, $\phi(U_0, \ldots, U_k)$ has a $(k+2)$-variate mean-zero normal distribution with covariance matrix

$$\Sigma = \begin{pmatrix} V_x^{(0)} & V_x^{(0,1)} & \ldots & V_x^{(0,k)} & V_{xG}^{(0)} \\ & V_x^{(1)} & \ldots & V_x^{(1,k)} & V_{xG}^{(1)} \\ & & \ddots & \vdots & \vdots \\ & & & V_x^{(k)} & V_{xG}^{(k)} \\ & & & & V_G \end{pmatrix}$$

The entries in the covariance matrix follow from lemma 3.14. Finally, consider the function $\psi : \mathbb{R}^{k+1} \times (0,1) \to \mathbb{R}^{k+1}$, defined by

$$(x_0, \ldots, x_k, G) \mapsto \left(\frac{x_0}{1-G}, \ldots, \frac{x_k}{1-G} \right)$$

The derivative of ψ at the point $(x_0, \ldots, x_k, G(L))$ is given by the $(k+1) \times (k+2)$-matrix

$$V = \frac{1}{1-G(L)} \begin{pmatrix} & & \theta_0 \\ & I_{k+1} & \vdots \\ & & \theta_k \end{pmatrix}$$

where I_{k+1} is the $(k+1) \times (k+1)$ identity matrix. Applying the delta method again,

$$\sqrt{n}\{\bar{\theta}^{(n)} - \bar{\theta}\} \xrightarrow{\mathcal{L}} N(0, V\Sigma V^T).$$

It remains to show $V\Sigma V^T = \Sigma_\theta$. To see this, let $\Sigma_\theta = (u_{ij})$, $V = (v_{ij})$, $\Sigma = (\sigma_{ij})$ and $W = (w_{ij})$, where $W = \Sigma V^T$. Then for $1 \leqslant i \leqslant k+2$ and $1 \leqslant j \leqslant k+1$,

$$v_{ij} = \begin{cases} 1 & i = j \leqslant k+1 \\ 0 & i \neq j, \quad i,j \leqslant k+1 \\ \theta_{i-1} & j = k+2 \end{cases}$$

and

$$\sigma_{ij} = \begin{cases} V_x^{(i-1)} & i = j \leqslant k+1 \\ V_x^{(i-1,j-1)} & i \neq j, \quad i,j \leqslant k+1 \\ V_{xG}^{(i-1)} & i \leqslant k+1, j = k+2 \\ V_{xG}^{(j-1)} & j \leqslant k+1, i = k+2 \\ V_G & i = j = k+2 \end{cases}$$

Using the matrix multiplication formula,

$$\begin{aligned}
\{1 - G(L)\}^2 u_{ij} &= \sum_{l=1}^{k+2} v_{il} w_{lj} \\
&= \sum_{l=1}^{k+2} v_{il} \sum_{m=1}^{k+2} \sigma_{lm} v_{jm} \\
&= \sum_{l=1}^{k+1} v_{il} \sum_{m=1}^{k+2} \sigma_{lm} v_{jm} + \underbrace{v_{i,k+2}}_{\theta_{i-1}} \left(\sum_{m=1}^{k+1} \underbrace{\sigma_{k+2,m}}_{V_{xG}^{(m-1)}} v_{jm} + \underbrace{\sigma_{k+2,k+2}}_{V_G} \underbrace{v_{j,k+2}}_{\theta_{j-1}} \right) \\
&= \sum_{m=1}^{k+2} \sigma_{im} v_{jm} + \theta_{i-1} V_{xG}^{j-1} + \theta_{i-1}\theta_{j-1} V_G \\
&= \sum_{m=1}^{k+1} \underbrace{\sigma_{im}}_{V_x^{(i-1,m-1)}} v_{jm} + \underbrace{\sigma_{i,k+2}}_{V_{xG}^{(i-1)}} \underbrace{v_{j,k+2}}_{\theta_{j-1}} + \theta_{i-1} V_{xG}^{j-1} + \theta_{i-1}\theta_{j-1} V_G \\
&= V_x^{(i-1,j-1)} + \theta_{j-1} V_{xG}^{(i-1)} + \theta_{i-1} V_{xG}^{j-1} + \theta_{i-1}\theta_{j-1} V_G,
\end{aligned}$$

□

Remark 3.16. *All rows and columns of the asymptotic covariance matrix* Σ_θ *sum to* 0*, because of the constraint*

$$\sum_{i=0}^{k} \theta_i = 1,$$

and the fact, that

$$V_G + \sum_{i=0}^{k} V_{xG}^{(i)} = 0,$$

and

$$V_x^{(i)} + \sum_{j \neq i} V_x^{(i,j)} + V_{xG}^{(i)} = 0 \quad 0 \leqslant i \leqslant k.$$

Therefore the $(k+1) \times (k+1)$*-matrix* Σ_θ *has only rank* k*.*

3.2.4 *Estimation of the log average hazard ratios*

For $i = 1, \ldots, k$ let

$$\beta_i = \log\left(\frac{\theta_i}{\theta_0}\right) = \log\theta_i - \log\theta_0$$

The parameters $\beta_1, \ldots, \beta_k$ indeed coincide with the usual log-hazard ratios, if the proportional hazards model is true.

Lemma 3.17. *For any* $i = 1, \ldots, k$*, if the hazards are proportional, i.e.*

$$\lambda_i(t) = \lambda_0(t) e^{\tilde{\beta}_i}, \tag{3.2.5}$$

then β_i *does not depend on the truncation point* L *or the weight function* G *and is identical to the log hazard ratio* $\tilde{\beta}_i$*.*

Proof. Under the proportional hazards assumption the ratio

$$\frac{\lambda_i(t)}{\sum_{j=0}^{k} \lambda_j(t)} = \frac{e^{\tilde{\beta}_i}}{\sum_{j=0}^{k} e^{\tilde{\beta}_j}}$$

is constant. Thus

$$\theta_i = -\int_0^\infty \frac{\lambda_i}{\sum_{i=0}^{k} \lambda_j} G(dt) = -\frac{\lambda_i}{\sum_{j=0}^{k} \lambda_j} \int_0^\infty G(dt) = \frac{e^{\tilde{\beta}_i}}{\sum_{j=0}^{k} e^{\tilde{\beta}_j}},$$

and

$$e^{\beta_i} = \frac{\theta_i}{\theta_0} = e^{\tilde{\beta}_i}.$$

□

Each β_i is consistently estimated by $\hat{\beta}_i = \log\hat{\theta}_i - \log\hat{\theta}_0$, because of theorem 3.15. Define the k-vectors $\bar{\beta} = (\beta_1, \ldots, \beta_k)^T$ and $\bar{\beta}^{(n)} = (\hat{\beta}_1, \ldots, \hat{\beta}_k)^T$.

Corollary 3.18. *Under assumption 3.8, as* $n \to \infty$,

$$\sqrt{n}\{\bar{\beta}^{(n)} - \bar{\beta}\} \xrightarrow{\mathcal{L}} N(0, \Sigma_\beta),$$

where $\Sigma_\beta = (u_{ij})_{1\leqslant i,j\leqslant k}$ *is the* $k \times k$*-matrix with entries*

$$u_{ij} = \frac{\sigma_{11}}{\theta_0^2} - \frac{\sigma_{1,j+1}}{\theta_0\theta_j} - \frac{\sigma_{i+1,1}}{\theta_0\theta_i} + \frac{\sigma_{i+1,j+1}}{\theta_j^2}.$$

Proof. Define the map $f : \mathbb{R}^{k+1} \to \mathbb{R}^k$ by

$$f(y_0, \ldots, y_k) = (\log y_1 - \log y_0, \ldots, \log y_k - \log y_0)^T,$$

and note that $f(\theta_0, \theta_1, \ldots, \theta_k) = \beta$. The derivative of f at the point $(\theta_0, \ldots, \theta_k)$ is given by the $k \times (k+1)$-matrix

$$\Gamma = \nabla f(\theta_0, \ldots, \theta_k) = \begin{pmatrix} -\theta_0^{-1} & \theta_1^{-1} & & \\ \vdots & & \ddots & \\ -\theta_0^{-1} & & & \theta_k^{-1} \end{pmatrix}$$

Thus by theorem 3.15 and the delta method

$$\sqrt{n}\{\hat{\beta} - \beta\} = \sqrt{n}\{f(\hat{\theta}_0, \ldots, \hat{\theta}_k) - f(\theta_0, \ldots, \theta_k)\} \xrightarrow{\mathcal{L}} N(0, \Gamma\Sigma_\theta\Gamma^T),$$

where $\Sigma_\theta = (\sigma_{ij})_{1\leqslant i,j\leqslant k+1}$ is the covariance matrix from theorem 3.15. It remains to show, that $\Sigma_\beta = \Gamma\Sigma_\theta\Gamma^T$. Now let $\Gamma\Sigma_\theta\Gamma^T = (u_{ij})$ and $\Gamma = (\Gamma_{ij})$, then for $1 \leqslant i \leqslant k$ and $1 \leqslant j \leqslant k+1$,

$$\Gamma_{ij} = \begin{cases} -\frac{1}{\theta_0} & j = 1 \\ \frac{1}{\theta_i} & j = i+1 \\ 0 & else \end{cases}$$

and for $1 \leqslant i, j \leqslant k$,

$$\begin{aligned} u_{ij} &= \sum_{h=1}^{k+1} \Gamma_{ih} \sum_{l=1}^{k+1} \sigma_{hl} \Gamma_{jl} \\ &= \sum_{h=1}^{k+1} \Gamma_{ih} \left(-\frac{\sigma_{h1}}{\theta_0} + \frac{\sigma_{h,j+1}}{\theta_j} \right) \\ &= \frac{\sigma_{11}}{\theta_0^2} - \frac{\sigma_{1,j+1}}{\theta_0 \theta_j} - \frac{\sigma_{i+1,1}}{\theta_0 \theta_i} + \frac{\sigma_{i+1,j+1}}{\theta_j^2} \end{aligned}$$

□

3.2.5 *Multivariate testing*

Theorem 3.15 can be used to directly derive the asymptotic distribution of various useful test statistics. In the following χ^2_k denotes the χ^2-distribution with k degress of freedom.

Corollary 3.19. *Under assumption 3.8, as* $n \to \infty$,

$$n\{\bar{\theta}^{(n)} - \bar{\theta}\}^T \hat{\Sigma}_\theta^- \{\bar{\theta}^{(n)} - \bar{\theta}\} \xrightarrow{\mathcal{L}} \chi^2_k,$$

where $\hat{\Sigma}_\theta^-$ *is any generalized inverse of* $\hat{\Sigma}_\theta$, *and for any* $c \in \mathbb{R}^{k+1}$,

$$\frac{c^T \sqrt{n}\{\bar{\theta}^{(n)} - \bar{\theta}\}}{\sqrt{c^T \hat{\Sigma}_\theta c}} \xrightarrow{\mathcal{L}} N(0,1).$$

A corresponding result holds for the k-vector $\bar{\beta}^{(n)}$. The limiting distribution is also χ^2 with k degrees of freedom, since the asymptotic covariance matrix Σ_β has full rank.

Corollary 3.20. *Under assumption 3.8, as* $n \to \infty$,

$$n\{\bar{\beta}^{(n)} - \bar{\beta}\}^T \hat{\Sigma}_\beta^{-1} \{\bar{\beta}^{(n)} - \bar{\beta}\} \xrightarrow{\mathcal{L}} \chi^2_k,$$

where $\hat{\Sigma}_\beta^{-1}$ *is the inverse of* $\hat{\Sigma}_\beta$, *and for any linear contrast* $c \in \mathbb{R}^k$,

$$\frac{c^T \sqrt{n}\{\bar{\beta}^{(n)} - \bar{\beta}\}}{\sqrt{c^T \hat{\Sigma}_\beta c}} \xrightarrow{\mathcal{L}} N(0,1).$$

3.2.6 *Special case: two samples*

In the two-sample case definition 3.4 reduces to

$$\theta_i = -\int_0^\infty \frac{\lambda_i(t)}{\lambda_0(t)+\lambda_1(t)}\,dG(t), \tag{3.2.6}$$

for $i = 0, 1$. Under the proportional hazards assumption,

$$\lambda_1(t) = \lambda_0(t)e^{\beta},$$

the average hazard ratio θ_1 is equal to

$$\frac{e^{\beta}}{1+e^{\beta}},$$

or equivalently

$$\beta = \log\left(\frac{\theta_1}{1-\theta_1}\right)$$

regardless of the weight function G. In the two-sample case it suffices to consider only x_1 and θ_1, since $\theta_0 = 1-\theta_1$ and $x_0 = 1-G(L)-x_1$. Correspondingly $\hat{\theta}_0 = 1-\hat{\theta}_1$ and $\hat{x}_0 = 1-\hat{G}(L)-\hat{x}_1$.

Corollary 3.21. *Under assumption 3.8, as* $n \to \infty$,

$$\sqrt{n}\{\hat{\theta}_1 - \theta_1\} \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2),$$

where

$$\sigma_\theta^2 = \frac{1}{\{1-G(L)\}^2}\{V_x + 2\theta_1 V_{xG} + \theta_1^2 V_G\},$$

and σ_θ^2 *can estimated consistently by* $\hat{\sigma}_\theta^2$*, which is obtained by replacing* ν_i*,* S_i *and* ρ_i *by* n_i/n*,* $\hat{S}_i$ *and* $\hat{\rho}_i$*.*

Proof. For the special case $k = 1$, theorem 3.15 reduces to

$$\sqrt{n}\left\{\begin{pmatrix}\hat{\theta}_0\\ \hat{\theta}_1\end{pmatrix} - \begin{pmatrix}\theta_0\\ \theta_1\end{pmatrix}\right\} \xrightarrow{\mathcal{L}} N(0, \Sigma_\theta),$$

where

$$\Sigma_\theta = \begin{pmatrix} \sigma_\theta^2 & -\sigma_\theta^2 \\ -\sigma_\theta^2 & \sigma_\theta^2 \end{pmatrix}.$$

□

Estimation of the log average hazard ratio

Define the estimator $\hat{\beta} = \log \hat{\theta}_1 - \log(1 - \hat{\theta}_1)$ for β, then

Corollary 3.22. *Under assumption 3.8, as* $n \to \infty$,

$$\sqrt{n}\{\hat{\beta} - \beta\} \xrightarrow{\mathcal{L}} N(0, \sigma_\beta^2),$$

where

$$\sigma_\beta^2 = \frac{\sigma_\theta^2}{\theta_1^2(1-\theta_1)^2}.$$

σ_β^2 *can be estimated consistently by using the consistent estimators* $\hat{\sigma}_\theta^2$ *and* $\hat{\theta}_1$.

Proof.

$$\begin{aligned}\sqrt{n}\{\hat{\beta} - \beta\} &= \sqrt{n}\{\log \hat{\theta}_1 - \log \theta_1\} - \sqrt{n}\{\log(1-\hat{\theta}_1) - \log(1-\theta_1)\} \\ &= \frac{1}{\theta_1(1-\theta_1)}\sqrt{n}\{\hat{\theta}_1 - \theta_1\} + o_p(1)\end{aligned}$$

By corollary 3.21, $\sqrt{n}\{\hat{\theta}_1 - \theta_1\}$ converges in distribution to normal distribution with mean zero and variance σ_θ^2. □

Remark 3.23. *Corollary 3.22 is indeed a special case of corollary 3.18 for* $k = 1$. *In the proof of corollary 3.18 it is shown, that* $\Sigma_\beta = \Gamma \Sigma_\theta \Gamma^T$, *where* Σ_θ *is the asymptotic covariance matrix of theorem 3.15 and* $\Gamma = (\theta_0^{-1}, \theta_1^{-1})^T$ *in the case* $k = 1$. *Thus*

$$\Sigma_\beta = \Gamma \Sigma_\theta \Gamma^T = \begin{pmatrix} -\frac{1}{\theta_0} & \frac{1}{\theta_1} \end{pmatrix} \begin{pmatrix} \sigma_\theta^2 & -\sigma_\theta^2 \\ -\sigma_\theta^2 & \sigma_\theta^2 \end{pmatrix} \begin{pmatrix} -\frac{1}{\theta_0} \\ \frac{1}{\theta_1} \end{pmatrix} = \frac{\sigma_\theta^2}{\theta_0^2\theta_1^2}(\underbrace{\theta_0 + \theta_1}_{=1})^2 = \sigma_\beta^2$$

Simulations

Table 3.1 shows the result of 10^5 simulations for several scenarios with exponentially distributed survival data. Censoring is also exponential with rate adjusted to achieve approx. 30% of censored observations. The truncation point L was set to 1 to prevent the number at risk to become too small at time L, which could negatively impact the

finite sample performance of the estimators. Sample size was 100 in both groups. Both estimators exhibit about the same finite sample bias, whereas the empirical standard deviation of the maximum partial likelihood estimator is smaller than that of the average hazard ratio based estimator. This is expected from the efficiency results for the maximum partial likelihood estimator for the proportional hazards model (Cox [1972]). The estimated log hazard ratio from the Cox model is denoted by $\hat{\beta}_{PH}$, the estimated log average hazard ratio by $\hat{\beta}_{AHR}$. The average hazard ratio is denoted by θ and its estimate by $\hat{\theta}$. The score test in the Cox model, i.e. the log-rank test is denoted by LR. The test based on the asymptotic distribution of $\hat{\beta}_{AHR}$ by AHR(β). Table 3.2 shows the estimated and true asymptotic variances of $\hat{\theta}$ and $\hat{\beta}$ for the same

Table 3.1: Empirical mean and standard deviation of the log average hazard ratio estimator $\hat{\beta}_{AHR}$ compared with the maximum partial likelihood estimator $\hat{\beta}_{PH}$ of the log-hazard ratio from the Cox proportional hazards model.

β [b]	$\hat{\beta}_{AHR}$	$\hat{\beta}_{PH}$	$sd(\hat{\beta}_{AHR})$	$sd(\hat{\beta}_{PH})$
0	0.000	0.001	0.199	0.173
-0.080	-0.081	-0.081	0.200	0.173
-0.154	-0.154	-0.155	0.201	0.172
-0.223	-0.225	-0.225	0.203	0.173
-0.288	-0.291	-0.290	0.206	0.174
-0.348	-0.351	-0.350	0.208	0.174
-0.405	-0.409	-0.409	0.210	0.175
-0.460	-0.463	-0.463	0.212	0.176
-0.511	-0.515	-0.515	0.214	0.177
-0.560	-0.564	-0.564	0.216	0.177

[a] Results of 10^5 simulation runs together with the true log hazard ratio. The survival times are exponential in both groups. Censoring was approx. 30% censoring. Sample size was 100 in both groups.
[b] log hazard ratio

simulation scenario.

Table 3.2: Estimated and true asymptotic variances of $\sqrt{n}\{\hat{\theta} - \theta\}$ and $\sqrt{n}\{\hat{\beta}_{AHR} - \beta\}$ and empirical bias of the asymptotic variance estimators for various proportional hazards alternatives at a fixed sample size[a].

	$\sqrt{n}\{\hat{\theta} - \theta\}$		$\sqrt{n}\{\hat{\beta}_{AHR} - \beta\}$		Bias (%)	
β [b]	Estimated	True	Estimated	True	$\hat{\theta}$	$\hat{\beta}_{AHR}$
0	0.479	0.480	7.819	7.684	-0.21	1.76
-0.080	0.486	0.487	7.957	7.820	-0.21	1.75
-0.154	0.491	0.492	8.105	7.965	-0.20	1.76
-0.223	0.493	0.495	8.258	8.114	-0.40	1.77
-0.288	0.495	0.496	8.421	8.269	-0.20	1.84
-0.348	0.495	0.496	8.582	8.426	-0.20	1.85
-0.405	0.493	0.495	8.746	8.585	-0.40	1.86
-0.460	0.492	0.492	8.916	8.749	-0.00	1.91
-0.511	0.489	0.490	9.087	8.912	-0.20	1.96
-0.560	0.485	0.486	9.254	9.077	-0.21	1.96

[a] Results of 10^5 simulation runs. The survival times are exponential in both groups. Censoring was approx. 30% censoring. Sample size was 100 in both groups.
[b] log hazard ratio

3.2.7 *Special case: three samples*

Simulations

Table 3.3 shows the result of 10^5 simulations of a three-sample setup. Survival was exponential with parameter 1 in all three groups with censoring also exponential, such that the censoring probability was about 30%. The truncation points was $L = 2$. Both test statistics should be approximately χ^2 distributed with 2 degrees of freedom, such that the mean should be close to 2 and the variance close to 4. The empirical distributions of both test statistics are seen to approach the asymptotic distribution as the sample size increases. While the LR method is clearly anti-conservative for small sample sizes, the AHR(β) method is slightly conservative even for the smallest group size $n = 50$.

Table 3.4 shows the result of 10^5 simulations of a scenario with proportional hazards alternative and a scenario with non-proportional hazards alternative. In the first

Table 3.3: Mean, variance and type-I-error of the multivariate AHR(β) and LR methods for three-samples for increasing group sizes n. The test statistics are asymptotically χ^2 distributed with two degress of freedom. [a]

	Mean		Variance		Type-I-error	
n	AHR(β)	LR	AHR(β)	LR	AHR(β)	LR
50	1.99	2.06	3.86	4.34	0.049	0.056
100	1.99	2.02	3.91	4.16	0.048	0.052
200	1.99	2.01	3.91	4.07	0.049	0.051

[a] Results of 10^5 simulations, $\alpha = 0.05$.

scenario the survival times are sampled from a Weibull distribution with shape parameter 1 in all three groups and scale parameters 1, 1.5 and 2 respectively. The true average hazard ratios are $\theta_0 = 0.46$, $\theta_1 = 0.31$, and $\theta_2 = 0.23$. In the second scenario the survival times are sampled from a Weibull distribution with shape parameters 1, 1.5 and 2 respectively, and scale parameter 1 in all three groups. The true average hazard ratios in this scenario are $\theta_0 = 0.42$, $\theta_1 = 0.32$, and $\theta_2 = 0.26$. Censoring in both scenarios is again exponential with rate parameter adjusted, such that the amount of censoring is about 30% in all three groups. The table shows the power of the multivariate tests as well as the power of the trend tests (i.e. contrast $c = (1, 2, 3)$). The performance of the AHR(θ) and AHR(β) methods relative to the LR method is similar to the two-sample case (cf. tables 3.7 and 3.8). The trend tests have larger power here, since they specifically test the alternative $\theta_0 > \theta_1 > \theta_2$. The trend test for the AHR(β) method in the three sample case is simply the one-sided test $H_0 : \beta_1 = \beta_2$ vs. $H_1 : \beta_1 > \beta_2$, whereas the multivariate test is the two-sided test of $H_0 : \beta_1 = \beta_2$ vs. $H_1 : \beta_1 \neq \beta_2$.

Three-arm trial

The results of the previous section can be directly applied to test non-inferiority and superiority in three-arm survival trials in the non-proportional hazards case. This generalizes the results of Mielke et al. [2008] for exponentially distributed survival times and the results of Kombrink et al. [2013] for survival times following a pro-

Table 3.4: Power of the multivariate and trend test of the AHR(β) and LR methods in a three-sample proportional and a non-proportional hazards scenario at a significance level of 0.05. [a]

					Power (mv.)		Power (trend)	
θ_0	θ_1	θ_2	n	L	AHR(β)	LR	AHR(β)	LR
0.46	0.31	0.23	100	3	0.89	0.96	0.96	0.98
0.42	0.32	0.26	200	1.5	0.81	0.07	0.92	0.07

[a] Results of 10^5 simulations, $\alpha = 0.05$.

portional hazards model. Mielke et al. [2008] consider the retention of control effect hypothesis

$$H_0 : \frac{\lambda_T}{\lambda_P} \geqslant \left(\frac{\lambda_R}{\lambda_P}\right)^{\Delta}$$

where $\Delta \in [0, \infty)$ is the non-inferiority margin and $\lambda_T, \lambda_P, \lambda_R$ are the parameters of the exponential distributions in the test (T), placebo (P) and reference (R) groups. Replace $\lambda_T, \lambda_P, \lambda_R$ with the corresponding average sample to total hazard ratios $\theta_T, \theta_P, \theta_R$:

$$H_0^{\theta} : \frac{\theta_T}{\theta_P} \geqslant \left(\frac{\theta_R}{\theta_P}\right)^{\Delta} \tag{3.2.7}$$

or equivalently

$$H_0^{\beta} : \beta_T \geqslant \Delta\beta_R, \tag{3.2.8}$$

where $\beta_T = \log\theta_T - \log\theta_P$ and $\beta_R = \log\theta_R - \log\theta_P$. Setting $\Delta = 0$ means testing for efficacy of test treatment over placebo, $\Delta \in (0, 1)$ means that the test treatment achieves at least $\Delta \times 100$ per cent of the reference effect, when both are compared to placebo, $\Delta \geqslant 1$ means testing for superiority of test over reference treatment. Let $\hat{\eta} = \hat{\beta}_T - \Delta\hat{\beta}_R$ and $\eta = \beta_T - \Delta\beta_R$. Using corollary 3.20 with the linear contrast $c = (-\Delta, 1)^T$,

Theorem 3.24. *Under assumption 3.8, as* $n \to \infty$,

$$\sqrt{n}\{\hat{\eta} - \eta\} = \sqrt{n}\{c^T\bar{\beta}^{(n)} - c^T\bar{\beta}\} \xrightarrow{\mathcal{L}} N(0, c^T\Sigma_\beta c).$$

Since β_T and β_R reduce to the usual hazard ratio in the proportional hazards case, this result can be seen as a direct generalization of the results of Kombrink et al. [2013].

3.3 RELATED TWO-SAMPLE METHODS

In this section two alternative two-sample methods are considered, which are also functions of the marginal survival functions.

3.3.1 *Restricted mean survival*

A totally different approach avoiding the concept of hazard rates is based on the survival functions directly. The mean of the survival time T is equal to the area under the survival curve

$$E[T] = \int_0^\infty S(t)dt.$$

Again the tails of the survival distribution cannot be consistently estimated, because of censoring. So instead the *restricted mean survival*

$$E[\min(T, L)] = \int_0^L S(t)dt,$$

is considered. Two distributions can be compared, by comparing their restricted (weighted) mean survival. This leads to the Pepe-Fleming class of test statistics (Pepe and Fleming [1989]),

$$\int_0^L \hat{w}(t)\{\hat{S}_0(t) - \hat{S}_1(t)\}dt,$$

where $\hat{S}_0$ and $\hat{S}_1$ are estimates of the survival functions S_0 and S_1 and $\hat{w}$ is a weight function. This kind of test statistics with the weighted Kaplan-Meier estimator has already been considered by Murray and Tsiatis [1996]. Consistency and asymptotic normality follow from the continuous mapping theorem under similar conditions as for the average hazard ratio, since the map $S \mapsto \int_0^L w(u)S(u)du$ is continuous.

Theorem 3.25. *If assumption 3.8 holds for $k = 2$ and $\hat{w}$ converges uniformly in probability to the bounded function w on $[0, L]$, then under the null hypothesis $S_0 = S_1$, as $n \to \infty$,*

$$\sqrt{n}\left(\int_0^L \hat{w}(u)\left\{\hat{S}_0(u) - \hat{S}_1(u)\right\}du\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2(w, L)),$$

where

$$\sigma^2(w, L) = \int_0^L \int_0^L w(s)w(t) \left\{ \nu_0^{-1} \rho_0(s,t) + \nu_1^{-1} \rho_1(s,t) \right\} ds dt.$$

A consistent estimator of $\sigma^2(w, L)$ *is given by*

$$\hat{\sigma}^2(w, L) = \int_0^L \int_0^L \hat{w}(s)\hat{w}(t) \left\{ \nu_0^{-1} \hat{\rho}_0(s,t) + \nu_1^{-1} \hat{\rho}_1(s,t) \right\} ds dt.$$

Proof. Since $S_0 = S_1$ under the null hypothesis,

$$\begin{aligned} &\sqrt{n} \int_0^L \hat{w}(s) \left\{ \hat{S}_0(s) - \hat{S}_1(s) \right\} ds \\ &= \sqrt{n} \int_0^L \hat{w}(s) \left\{ \hat{S}_0(s) - S_0(s) \right\} ds - \sqrt{n} \int_0^L \hat{w}(s) \left\{ \hat{S}_1(s) - S_1(s) \right\} ds. \end{aligned}$$

The $\hat{w}$ can be replaced by w, because

$$\begin{aligned} &\left| \sqrt{n} \int_0^L \{\hat{w}(s) - w(s)\} \left\{ \hat{S}_i(s) - S_i(s) \right\} ds \right| \\ &\leqslant L \sup_{s \in [0,L]} |\hat{w}(s) - w(s)| \sup_{s \in [0,L]} \sqrt{n} \left| \hat{S}_i(s) - S_i(s) \right| \end{aligned}$$

The right hand side converges to 0 in probability, because $\sqrt{n} \left\{ \hat{S}_i - S_i \right\}$ converges weakly in $D[0, L]$ to a continuous Gaussian process Z_i, $i = 0, 1$, by assumption 3.8. Thus

$$\sup_{s \in [0,L]} \sqrt{n} \left| \hat{S}_i(s) - S_i(s) \right| = O_p(1).$$

Moreover by assumption

$$\sup_{s \in [0,L]} |\hat{w}(s) - w(s)| = o_p(1).$$

This gives

$$\begin{aligned} &\sqrt{n} \int_0^L \hat{w}(s) \left\{ \hat{S}_0(s) - \hat{S}_1(s) \right\} ds \\ &= \sqrt{n} \int_0^L w(s) \left\{ \hat{S}_0(s) - S_0(s) \right\} du - \sqrt{n} \int_0^L w(s) \left\{ \hat{S}_1(s) - S_1(s) \right\} ds + o_p(1). \end{aligned}$$

The two terms are stochastically independent and, by the continuous mapping theorem each converges in distribution to a normal random variable

$$\int_0^L w(s)Z_i(s)ds,$$

which has mean 0 and variance

$$E\left[\left(\int_0^L w(s)Z_i(s)ds\right)^2\right] = \int_0^L\int_0^L w(s)w(t)\rho_i(s,t)du,$$

since $E[Z_i(s)Z_i(t)] = \rho_i(s,t)$. Consistency of the variance estimator follows immediately from the uniform consistency of $\hat{w}$ and ρ_i, $i = 0, 1$. □

It is clear, that the test from theorem 3.25 does have low power against alternatives with *crossing* survival curves. This could be fixed, e.g. by considering the integrated squared difference of the survival functions or by using an appropriate weight function. However, the optimal choice requires knowledge of the truth and might be difficult to interpret. There is no direct multivariate generalization of this method. More than two samples can only be compared by pairwise comparisons.

3.3.2 *Median survival*

The median (or any other quantile) is an attractive alternative, since it has a simple interpretation. A consistent estimate of the median can be derived from any consistent estimate of the survival function. Methods for construction of confidence intervals exist requiring only pointwise consistency and weak convergence of the survival function estimator and the corresponding asymptotic variance estimator (Brookmeyer and Crowley [1982]). Similarly, confidence intervals for the difference and the ratio of two medians can be constructed (Su and Wei [1993]). The true median can be consistently estimated, as long as the truncation point L is larger than the true median, since truncation right of the median does not change the median, in contrast to the mean (see restricted mean survival above). However, if in a finite sample the estimated survival curve does not drop below 0.5, then the median cannot be calculated without extrapolating the survival curve.

The p-th quantile of the survival distribution is defined as

$$z_p := \inf_{x \geqslant 0} \{S(x) \leqslant p\}.$$

Given an estimate $\hat{S}$ of the survival function S, the estimated quantile is

$$\hat{z}_p := \inf_{x \geqslant 0} \{\hat{S}(x) \leqslant p\}.$$

The asymptotic variance of the estimator $\hat{z}_p$ depends on the density of the survival time at the p-th quantile, which is difficult to estimate nonparametrically. Brookmeyer and Crowley [1982] describe how to obtain a nonparametric asymptotic confidence interval without the need to estimate the density. Suppose there is a consistent and asymptotically normal estimator $\hat{S}(x)$ of $S(x)$ for any $x \in [0, L]$, and a consistent estimator $\hat{\sigma}(x)$ of the asymptotic variance of $\sqrt{n}\{\hat{S}(x) - S(x)\}$. Then the following result follows immediately.

Theorem 3.26. *If $z_p < L$, then*

$$\frac{n\{\hat{S}(z_p) - p\}^2}{\hat{\sigma}^2(z_p)} \xrightarrow{\mathcal{L}} \chi^2_1.$$

An asymptotic level α test for the hypothesis $H_0 : z = z_p$ is to reject H_0 if

$$\frac{n\{\hat{S}(z_p) - p\}^2}{\hat{\sigma}^2(z_p)} > c_\alpha,$$

where c_α is the upper α-quantile of the χ^2_1-distribution. A confidence interval for z_p is given by

$$CI_\alpha = \left\{ z : \frac{n\{\hat{S}(z) - p\}^2}{\hat{\sigma}^2(z)} \leqslant c_\alpha \right\}.$$

Under similar conditions asymptotic confidence intervals for the difference and ratio of two median survival times in a two-sample setting can be constructed (cf. Su and Wei [1993]). This method is also implemented in the R package written for this thesis (see appendix B).

Theorem 3.26 requires only the pointwise consistency and asymptotic normality of the survival function estimator and its variance estimator (as compared with the much stronger conditions for the results in section 3.2.6). Moreover, as long as the true quantile is contained in the interval $[0, L]$, the quantile of the survival time and

not just the truncated survival time can be estimated as is the case for the mean (see section 3.3.1). However, one of the problems of this method is, that even if the true quantile is contained in the interval $[0, L]$, it may happen, that for a given finite sample the estimated survival function never falls below p. In this case the estimator would be undefined.

3.4 COMPARISON OF TWO-SAMPLE METHODS

The finite sample performance of the various two-sample methods is compared. The method based on the average hazard ratio is denoted by AHR(θ), the method based on the log-odds of the average hazard ratio by AHR(β), the restricted mean survival by RMS and the log-rank test by LR.

The survival times are sample from Weibull distributions with different shape and scale parameters. Two Weibull distributions have proportional hazards if and only if they have the same shape parameter. This makes it very easy to simulate proportional and non-proportional hazards scenarios.

3.4.1 *Type-I-error*

Independent censoring / Non-informative censoring

To assess the actual type-I-error rate of the various methods, survival times are sampled from an exponential distribution with rate parameter 1 in both groups. The censoring times have an exponential distribution with parameter 1/2.34, resulting in approx. 30% censoring.

Table 3.5 shows the actual type-I-error of the AHR(θ), AHR(β), RMS and LR methods for increasing group sizes $n = 50, 100, 200, 400$ under independent censoring for two-sided testing of the null hypothesis of no treatment effect. All methods control the type-I-error at the nominal level ($\alpha = 0.05$) as n increases. For very small group sizes (e.g. $n = 50$) all methods, except AHR(β), are anti-conservative. It is interesting to see that the AHR(β) method is slightly conservative for all sample sizes, and is more conservative for small than for larger sample sizes.

Table 3.5: Type-I-error of several two-sample methods for increasing group sizes n under independent censoring. [a]

	Type-I-error			
n	AHR(θ)	AHR(β)	RMS	LR
50	0.059	0.048	0.056	0.052
100	0.054	0.048	0.052	0.050
200	0.052	0.049	0.051	0.050
400	0.052	0.050	0.050	0.050

[a] Results of 10^5 simulation runs, standard error for all estimates is ≈ 0.0007, $\alpha = 0.05$.

Informative censoring

The methods based on the Kaplan-Meier estimator are now compared with the methods based on the weighted Kaplan-Meier estimator for the simulation scenario already described in section 2.5. The results are shown in table 3.6. The tests based on the Kaplan-Meier estimator and the the log-rank test show a huge inflation of the type-I-error rate, whereas the tests based on the weighted Kaplan-Meier estimator and the stratified log-rank test control the type-I-error rate at the nominal level (0.05). These results clearly show the sensitivity of the unstratified methods with respect to a violation of the independent censoring assumption.

3.4.2 *Power under proportional hazards alternatives*

In this simulation scenario various proportional hazards alternatives are considered. The survival times have a Weibull distribution with shape parameter 1 in both groups and scale parameter 1 in group 1 and different scale parameters b in group 2. The censoring times are exponentially distributed with the rate parameter adjusted, such that approx. 30% of censoring are achieved. The sample size is 100 in each group. The truncation point is $L = 1$. Table 3.7 shows the result of 10^5 simulations for each scenario. The results are also depicted in figure 3.5. As expected the log-rank test (LR) has the highest power of all methods. All other methods have up to 15% power loss compared to the log-rank test.

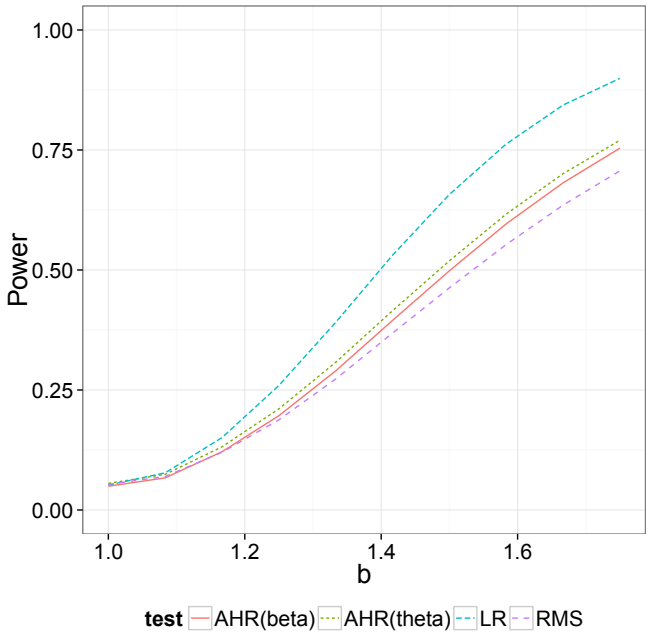

Figure 3.5: Power of several two-sample methods under proportional hazards alternatives

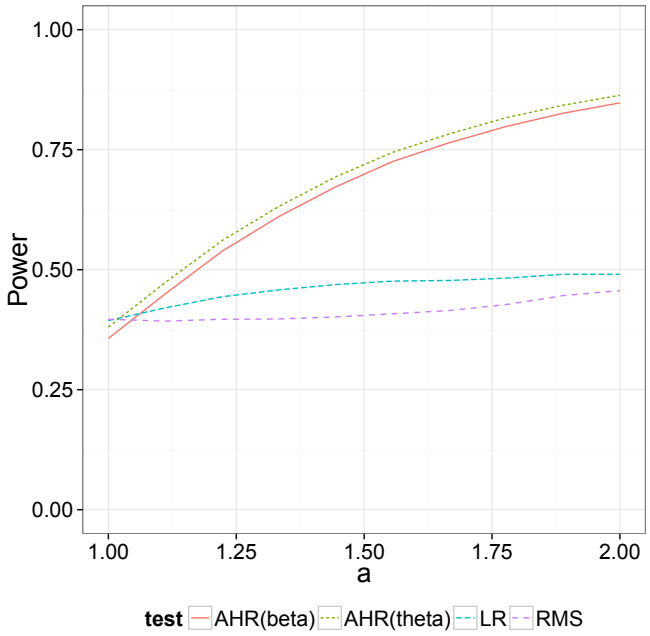

Figure 3.6: Power of several two-sample methods under non-proportional hazards alternatives

Table 3.6: Type-I-error of several two-sample methods under informative censoring [a].

	Type-I-error			
stratified	AHR(θ)	AHR(β)	RMS	LR
no [b]	0.163	0.159	0.214	0.327
yes [c]	0.052	0.050	0.052	0.050

[a] Results of 10^5 simulation runs, standard error of all estimates is ≈ 0.0007, $\alpha = 0.05$.
[b] log-rank test and methods based on the Kaplan-Meier estimator
[c] stratified log-rank test and methods based on the weighted Kaplan-Meier estimator

3.4.3 *Power under non-proportional hazards alternatives*

To assess the power under several non-proportional hazards alternatives survival times are sampled from a Weibull distribution with shape parameter 1 and scale parameter 0.3 in group 1 and different shape parameters a and scale parameter 0.4 in group 2. The censoring times are exponentially distributed with the rate parameter adjusted, such that approx. 30% of censoring are achieved. The sample size is 100 in each group. The truncation point is again $L = 1$.

Table 3.8 shows the results of 10^5 simulations for each scenario. The results are also depicted in figure 3.6. The power loss of the log-rank test compared to the relative risk based methods (AHR(θ) and AHR(β)) becomes substantial with increasing shape parameter. The RMS method performs even worse than the log-rank test in these scenarios.

Remark 3.27. *The power simulations were all done without stratification. In simulations not reported here, there was no difference in the actual power of the average hazard ratio methods, when a stratification variable was included, i.e. when the WKM estimator was used, although the WKM based average hazard ratio had a slightly smaller asymptotic variance.*

Table 3.7: Power of several two-sample methods under proportional hazards alternatives. Total sample size was 200 with 140 observed events on average. [a]

		Power			
b	$\frac{\theta_1}{\theta_0}$	AHR(θ)	AHR(β)	RMS	LR
1.00	1	0.06	0.05	0.05	0.05
1.08	0.92	0.07	0.07	0.07	0.08
1.17	0.86	0.13	0.12	0.12	0.15
1.25	0.80	0.21	0.20	0.19	0.26
1.33	0.75	0.31	0.29	0.27	0.39
1.42	0.71	0.42	0.39	0.37	0.53
1.50	0.67	0.52	0.50	0.46	0.66
1.58	0.63	0.62	0.60	0.55	0.76
1.67	0.60	0.70	0.68	0.64	0.84
1.75	0.57	0.77	0.75	0.71	0.90

[a] Results of 10^5 simulation runs, standard error of all estimates is ≈ 0.0007, $\alpha = 0.05$, $H_0 : \frac{\theta_1}{\theta_0} = 1$ vs. $H_1 : \frac{\theta_1}{\theta_0} \neq 1$

Table 3.8: Power of several two-sample methods under non-proportional hazards alternatives. Total sample size was 200 with 140 observed events on average. [a]

		Power			
a	$\frac{\theta_1}{\theta_0}$	AHR(θ)	AHR(β)	RMS	LR
1.00	0.76	0.38	0.36	0.40	0.39
1.11	0.73	0.47	0.45	0.39	0.42
1.22	0.70	0.56	0.54	0.40	0.44
1.33	0.67	0.63	0.61	0.40	0.46
1.44	0.65	0.69	0.67	0.40	0.47
1.56	0.63	0.74	0.73	0.41	0.48
1.67	0.62	0.78	0.76	0.42	0.48
1.78	0.60	0.82	0.80	0.43	0.48
1.89	0.59	0.84	0.83	0.45	0.49
2.00	0.57	0.86	0.85	0.46	0.49

[a] Results of 10^5 simulations, standard error of all estimates is ≈ 0.0007, $\alpha = 0.05$, $H_0 : \frac{\theta_1}{\theta_0} = 1$ vs. $H_1 : \frac{\theta_1}{\theta_0} \neq 1$

4

SEQUENTIAL TWO-SAMPLE METHODS

In group-sequential trials several (at least one) pre-planned interim analyses are performed. At each interim analysis it is decided wether to stop the trial by an early rejection of the null hypothesis, to continue the trial or to stop for futility. The interim analyses are performed either at pre-specified calendar times or after a certain (pre-specified) number of patients have been recruited or, in the case of survival trials, after a certain (pre-specified) number of events have been observed. The standardized test statistic at the k-th stage, Z_k, is calculated from all data available up to the time of k-th interim analysis. The null hypothesis is rejected and the trial stopped, if $Z_k < b_k$, for some critical value b_k. To control the familywise type-I-error, i.e. the probability of falsely rejecting the null hypothesis at any of the interim analyses without allowing futility stopping, at the level α, the rejection boundaries need to be chosen, such that

$$P_0(Z_1 < b_1, \ldots, Z_K < b_K) = 1 - \alpha.$$

The number and timing of interim analyses can be chosen flexibly, by using *error spending* functions. An error spending function is any montone increasing function $\alpha : [0, 1] \mapsto [0, 1]$, such that $\alpha(0) = 0$ and $\alpha(1) = \alpha$. If the critical values $b_1, \ldots, b_K$ are chosen, such that

$$P_0(Z_1 \leqslant b_1, \ldots, Z_{k-1} \leqslant b_{k-1}, Z_k > b_k) = \alpha(t_k) - \alpha(t_{k-1}),$$

then the familywise type-I-error is controlled at the level α. Note that it is straightforward to extend this approach to allow for early acceptance of the null hypothesis, an option which is not considered here explicitely. To determine the critical values $b_1, \ldots, b_K$, the joint distribution of the standardized test statistics $Z_1, \ldots, Z_K$ must be known. The rejection boundaries can be efficiently calculated, if the standardized

test statistics have the *canoncial joint distribution* (Jennison and Turnbull [1999]), i.e. $Z_1, \ldots, Z_K$ are jointly normal with

$$Z_k \sim N(0,1) \quad k = 1, \ldots, K$$
$$Cov(Z_i, Z_j) = \sqrt{\frac{\mathcal{I}_i}{\mathcal{I}_j}} \quad i < j,$$

where $\mathcal{I}_1, \ldots, \mathcal{I}_K$ are the *information levels*. Equivalently, this can be formulated in terms of the *score statistics* $S_k = \sqrt{\mathcal{I}_k} Z_k$, $S_1, \ldots, S_K$ are jointly normal with

$$S_k \sim N(0, \mathcal{I}_k)$$
$$Cov(S_i, S_j) = Var(S_i) \quad i < j,$$

i.e. the score statistics have the independent increments property. This special covariance structure allows for efficient recursive computation of the rejection boundaries, which can be calculated before the trial starts, since they only depend on the pre-specified error spending function and the pre-specified analysis times. Many test statistics have this property (at least asymptotically). In fact any semi-parametrically efficient test statistics will have asymptotically the independent increments property (Scharfstein et al. [1997]). This property also holds for the log-rank test statistics (Sellke and Siegmund [1983]) or more generally for the score test statistics in the proportional hazards model (Tsiatis et al. [1985]). For the log-rank test the information levels are proportional to the expected number of events (Tsiatis et al. [1985]).

The main result of this chapter is the derivation of the asymptotic joint distribution of the sequential weighted Kaplan-Meier estimator and the average hazard ratio test statistic at different calendar times (theorem 4.16). The independent increments structure of the weighted Kaplan-Meier estimator is inherited by the average hazard ratio test statistic. The result for the average hazard ratio is given in the two-sample case only, which is the most common, to keep the notation manageable.

As with the fixed sample size methods in chapter 3, any estimator of the marginal survival function can be used with theorem 4.16, given that it has certain large sample properties. This is the case for the weighted Kaplan-Meier estimator as will be proved in section 4.2. Since no adaptations are made, it is not necessary to use the weighted Kaplan-Meier to control the type-I-error.

4.1 STAGGERED ENTRY

In clinical trials patients do not enter all at the same time, but at different 'staggered' entry times, i.e. at independent and identically distributed entry or recruitment times $R_i, i = 1, \ldots, n$, which are assumed to be independent of all other data. At any calendar time t only patients with $R_i \leqslant t$ have entered the trial. The number of patients in the trial by time t is given by

$$n(t) = \sum_{i=1}^{n} \mathbb{1}\{R_i \leqslant t\}.$$

The sample size at calendar time t in stratum j is

$$n_j(t) = \sum_{i=1}^{n} \mathbb{1}\{X_i = j, R_i \leqslant t\} = \sum_{i=1}^{n(t)} \mathbb{1}\{X_i = j\}.$$

Clearly, $n(t) = n_1(t) + \ldots + n_J(t)$. Denote the proportion of patients recruited before calendar time t by

$$\hat{\pi}(t) = \frac{n(t)}{n},$$

and let $\pi(t) = P(R \leqslant t)$. It is assumed, that $n(t)$ tends to infinity for all t, such that as $n \to \infty$

$$\hat{\pi}(t) \xrightarrow{a.s.} \pi(t) > 0.$$

The observed data at calendar time t consists of the iid tuples

$$\{(Y_i \wedge (t - R_i)^+, \delta_i(t), X_i, R_i), i = 1, \ldots, n(t)\},$$

where

$$\delta_i(t) = \mathbb{1}\left\{T_i \leqslant C_i \wedge (t - R_i)^+\right\}$$

is the censoring indicator for patient i at calendar time t. If only one analysis is done at a fixed calendar time t, this is equivalent to all patients entering the trial at time 0 with censoring times $C_i' = C_i \wedge (t - R_i)^+$. The definitions of the counting processes, the at-risk indicators and the counting process martingales need to be adjusted to include the dependence on the calendar time t. Define for patient i at calendar time t and survival time s, the counting process

$$N_i(t, s) = \mathbb{1}\left\{Y_i \wedge (t - R_i)^+ \leqslant s, \delta_i(t) = 1\right\},$$

and the at-risk indicator

$$Y_i(t,s) = \mathbb{1}\big\{Y_i \wedge (t-R_i)^+ \geqslant s\big\}.$$

The number of events in stratum j at calendar time t and survival time s is then given by

$$\bar{N}_j^{(n)}(t,s) = \sum_{i=1}^{n} \mathbb{1}\{X_i = j, R_i \leqslant t\} N_i(t,s),$$

and the number at risk by

$$\bar{Y}_j^{(n)}(t,s) = \sum_{i=1}^{n} \mathbb{1}\{X_i = j, R_i \leqslant t\} Y_i(t,s).$$

As in the previous chapters, the conditional hazard rate of T_i given $X_i = j$ is denoted by λ_j, i.e.

$$\lambda_j(s) = \lim_{h\to 0} \frac{1}{h} P(s \leqslant T_i < t+h | T_i \geqslant t, X_i = j)$$

The hazard does not depend on the calendar time, which is also consistent with the assumption, that the recruitment time R_i is independent of the survival time T_i. Define the filtration $\mathcal{F}_R$ as in Tsiatis et al. [1995]:

Definition 4.1. *The filtration $\mathcal{F}_R$ as the increasing sequence of σ-algebras defined by*

$$\mathcal{F}_R(s) = \sigma(R_i, \mathbb{1}\{x \wedge Y_i \leqslant x, T_i \leqslant C_i\}, \mathbb{1}\{Y_i \leqslant x, T_i > C_i\}, X_i; x \leqslant s, i = 1, \ldots, n)$$

for $s \geqslant 0$

$\mathcal{F}_R(s)$ contains all observed survival and censoring information for all patients up to s time units on study, as well as the entry times and all the covariate information. This filtration does not depend on the calendar time, thus the counting processes $N_i(t,s)$ at any fixed calendar time t are adapted to the filtration $\mathcal{F}_R$ and, given $X_i = j$, the stochastic process

$$s \mapsto M_i(t,s) = N_i(t,s) - \int_0^s Y_i(t,s)\lambda_j(u)du \quad s \leqslant t,$$

is a local square integrable $\mathcal{F}_R$-martingale for $i = 1, \ldots, n$. Consequently the sum

$$\bar{M}_j^{(n)}(t,\cdot) = \sum_{i=1}^{n} \mathbb{1}\{X_i = j, R_i \leqslant t\} M_i(t,\cdot) = \bar{N}_j^{(n)}(t,s) - \int_0^s \bar{Y}_j^{(n)}(t,u)\lambda_j(u)du,$$

is, for any j and fixed t, also a local square integrable $\mathcal{F}_R$-martingale with

$$\left\langle \bar{M}_j^{(n)}(t,\cdot) \right\rangle (s) = \int_0^s \bar{Y}_j^{(n)}(t,u)\lambda_j(u)du.$$

The fact, that for any two fixed t and t', $\bar{M}_j^{(n)}(t,\cdot)$ and $\bar{M}_j^{(n)}(t',\cdot)$ are both martingales with respect to the same filtration makes it possible to define and calculate their predictable quadratic covariation process, which is useful for the covariance calculations in the next section:

$$\left\langle \bar{M}_j^{(n)}(t,\cdot), \bar{M}_j^{(n)}(t',\cdot) \right\rangle (s) = \int_0^s \bar{Y}_j^{(n)}(t \vee t',u)\lambda_j(u)du. \tag{4.1.1}$$

The counting process martingales $\bar{M}_j^{(n)}(t,\cdot)$ and $\bar{M}_{j'}^{(n)}(t,\cdot)$ associated with different strata $j \neq j'$ are orthogonal, i.e.

$$\left\langle \bar{M}_j^{(n)}(t,\cdot), \bar{M}_{j'}^{(n)}(t',\cdot) \right\rangle \equiv 0. \tag{4.1.2}$$

4.2 SEQUENTIAL WEIGHTED KAPLAN-MEIER ESTIMATOR

In this section the asymptotic joint distribution of the weighted Kaplan-Meier estimator at different calendar times is derived.

4.2.1 *Definition*

The conditional survival function of $C_i \wedge (t - R_i)^+$ given $X_i = j$ and $R_i \leqslant t$ is denoted by

$$K_j(t,s|R_i \leqslant t) = P(C_i \wedge (t - R_i)^+ > s | X_i = j, R_i \leqslant t).$$

The conditional survival function of $C_i \wedge (t - R_i)^+$ given $X_i = j$ is denoted by

$$\begin{aligned} K_j(t,s) &= P(C_i \wedge (t - R_i)^+ > s | X_i = j) \\ &= P(C_i \wedge (t - R_i)^+ > s, R_i \leqslant t | X_i = j) = K_j(t,s|R_i \leqslant t)\pi(t). \end{aligned}$$

The stratum-specific Nelson-Aalen and Kaplan-Meier estimators with staggered entry are defined as

Definition 4.2. *Define, for* $j = 1, \dots, J$, $s, t > 0$,

$$\hat{\Lambda}_j(t,s) = \int_0^s J_j^{(n)}(t,u) \frac{\bar{N}_j^{(n)}(t, du)}{\bar{Y}_j^{(n)}(t,u)}, \tag{4.2.1}$$

where $J_j^{(n)}(t,u) = \mathbb{1}\left\{\bar{Y}_j^{(n)}(t,u) > 0\right\}$, *and*

$$\hat{S}_j(t,s) = \prod_{[0,s]} \left(1 - J_j^{(n)}(t,u) \frac{\bar{N}_j^{(n)}(t, du)}{\bar{Y}_j^{(n)}(t,u)}\right). \tag{4.2.2}$$

The response probabilities at calendar time t are estimated by

$$\hat{p}_j(t) = \frac{n_j(t)}{n(t)} \quad \text{for } j = 1, \dots, J.$$

With these definitions the sequential weighted Kaplan-Meier estimator can be defined.

Definition 4.3. *The sequential weighted Kaplan-Meier estimator at calendar time* t *and survival time* $s < t$ *is defined by*

$$\hat{S}(t,s) = \sum_{j=1}^{J} \hat{p}_j(t) \hat{S}_j(t,s).$$

4.2.2 *Large sample properties*

The large sample properties of the sequential weighted Kaplan-Meier estimator are dervied under the following assumptions:

Assumption 4.4. *For* $j = 1, \dots, J$ *and* $i = 1, \dots, n$ *and given calendar time* t ,

1. $p_j = P(X_i = j) > 0$,
2. T_i *is stochastically independent of* C_i *given* X_i,
3. $L < \sup\{s \geqslant 0 : S_j(s) K_j(t,s) > 0\}$,
4. $\bar{Y}_j^{(n)}(t,s) \to \infty$, *as* $n \to \infty$ *for all* $s \in [0, L]$,
5. R_i *is stochastically independent of* T_i, C_i *and* X_i,

6. $\pi(t) = P(R_i \leqslant t) > 0$

The first four assumptions are essentially the same as in the fixed sample size case (assumption 3.2), where assumption 4.4(3) implies $L \leqslant t$. Assumption 4.4(5) implies, that $\hat{p}_j(t)$ is a consistent estimator of p_j, since

$$\hat{p}_j(t) = \frac{n_j(t)}{n}\frac{n}{n(t)} \xrightarrow{p} \frac{P(X_i = j, R_i \leqslant t)}{\pi(t)} = p_j,$$

as $n \to \infty$. Note that the response probability p_j does not depend on the calendar time t. The uniform convergence of the number at-risk in each stratum now follows from the Glivenko-Cantelli theorem.

Lemma 4.5. *Under assumption 4.4, as* $n \to \infty$,

$$\sup_{u \in [0,L]} \left| \frac{1}{n}\bar{Y}_j^{(n)}(t,u) - p_j S_j(u) K_j(t,u) \right| \xrightarrow{P} 0.$$

Proof. $n^{-1}\bar{Y}_j^{(n)}(t,u)$ converges uniformly in u to $P(T > u, C \wedge (t-R) > u, X = j)$, by the Glivenko-Cantelli theorem. Since T and C are stochastically independent given X,

$$\begin{aligned} &P(T > u, C \wedge (t-R) > u, X = j) \\ &= P(T > u | X = j) P(C \wedge (t-R) > u | X = j) P(X = j) \\ &= S_j(u) K_j(t,u) p_j. \end{aligned}$$

□

The next lemma proves some asymptotic results of the stratum-specific Kaplan-Meier estimator needed in the construction of the weighted Kaplan-Meier estimator. This result is similar to asymptotic results about the Kaplan-Meier estimator found in the literature (see e.g. Fleming and Harrington [2011]). The proof uses standard counting process methods of survival analysis (see chapter 2 and appendix A).

Lemma 4.6. *Under assumption 4.4, for any* $t > 0$ *and for* $j = 1, \ldots, J$,

1. *uniformly in* $s \in [0, L]$,

$$\sqrt{n}\{\hat{S}_j(t,s) - S_j(s)\} = -\frac{S_j(s)}{p_j}\frac{1}{\sqrt{n}}\int_0^s \frac{\bar{M}_j^{(n)}(t, du)}{S_j(u) K_j(t,u)} + o_P(1).$$

2. $\hat{S}_j(t,\cdot)$ *is a uniformly consistent estimator of* $S(t,\cdot)$ *on* $[0, L]$,

3. $\sqrt{n}\{\hat{S}_j(t,\cdot) - S_j(t,\cdot)\}$ *converges to a mean-zero Gaussian process with covariance function*

$$(t, s, s') \mapsto S_j(s)S_j(s')\rho_j(t, s, s'),$$

where

$$\rho_j(t, s, s') = \frac{1}{p_j}\int_0^{s\wedge s'} \frac{\lambda_j(u)du}{S_j(u)K_j(t,u)}.$$

4. *a uniformly consistent of the covariance function in (3) is given by*

$$(t, s, s') \mapsto \hat{S}_j(s)\hat{S}_j(s')\hat{\rho}_j(t, s, s'),$$

where

$$\hat{\rho}_j(t, s, s') = \frac{n}{\hat{p}_j(t)}\int_0^{s\wedge s'} \frac{J_j^{(n)}(t,u)\bar{N}_j^{(n)}(t, du)}{\bar{Y}_j^{(n)}(t,u)^2}.$$

Proof. 1. For each s write

$$\begin{aligned}\sqrt{n}\{\hat{\Lambda}_j(t,s) - \Lambda_j(t,s)\} = \sqrt{n}\int_0^s J_j^{(n)}(t,u)\left\{\frac{\bar{N}_j^{(n)}(t,du)}{\bar{Y}_j^{(n)}(t,u)} - \lambda_j(t,u)du\right\} \\ + \sqrt{n}\int_0^s \left\{J_j^{(n)}(t,u) - 1\right\}\lambda_j(t,u)du\end{aligned} \tag{4.2.3}$$

The second term converges to 0 in probability uniformly in s, as $n \to \infty$, since, by lemma 4.5

$$P\left(|\sqrt{n}\int_0^s \{J_j^{(n)}(t,u) - 1\}\lambda_j(t,u)du| > 0\right) \leqslant P\left(\inf_{u\in[0,L]} J_j^{(n)}(t,u) = 0\right) \to 0.$$

Since $S_j(t,u)K_j(t,u)$ is bounded away from 0 on $[0, L]$ (by the choice of L),

$$\sup_{u\in[0,L]}\left|\frac{nJ_j^{(n)}(t,u)}{\bar{Y}_j^{(n)}(t,u)} - \frac{1}{p_jS_j(u)K_j(t,u)}\right| \xrightarrow{P} 0,$$

as $n \to \infty$. It follows from lemma A.35, that

$$\sqrt{n}\{\hat{\Lambda}_j(t,s) - \Lambda_j(t,s)\} = \frac{1}{p_j}\frac{1}{\sqrt{n}}\int_0^s J_j^{(n)}(t,u)\frac{\bar{M}_j^{(n)}(t,du)}{S_j(u)K_j(t,u)} + o_p(1). \tag{4.2.4}$$

uniformly in s. The martingale central limit theorem (theorem A.45) implies weak convergence of the process $\sqrt{n}\{\hat{\Lambda}_j(t,\cdot) - \Lambda_j(t,\cdot)\}$ in $D[0,L]$. The functional delta method can now be applied with theorem A.52 to obtain

$$\sqrt{n}\{\hat{S}_j(t,s) - S_j(s)\} = -S_j(s)\sqrt{n}\{\hat{\Lambda}_j(t,s) - \Lambda_j(t,s)\} + o_p(1),$$

uniformly in s.

2. The uniform consistency of $\hat{\Lambda}_j$ follows from eq. (4.2.4) and Lenglart's inequality (lemma A.34) together with

$$\left\langle \frac{1}{n}\int_0^L \frac{\bar{M}_j^{(n)}(t,ds)}{S_j(s)K_j(t,s)} \right\rangle = \frac{1}{n}\int_0^L \frac{\bar{Y}_j^{(n)}(t,u)}{n}\frac{\lambda_j(t,u)du}{S_j(u)^2K_j(t,u)^2} \xrightarrow{p} 0.$$

3. The weak convergence result follows from the martingale central limit theorem and (1.). The value of the asymptotic covariance function at (t,s,s') is the limit of the predictable covariation process,

$$\begin{aligned}&\left\langle \frac{1}{\sqrt{n}}\int_0^s \frac{\bar{M}_j^{(n)}(t,du)}{S_j(t,u)K_j(t,u)}, \frac{1}{\sqrt{n}}\int_0^{s'} \frac{\bar{M}_j^{(n)}(t,du)}{S_j(t,u)K_j(t,u)} \right\rangle \\ &\quad = \int_0^{s\wedge s'} \frac{\bar{Y}_j^{(n)}(t,u)}{n}\frac{\lambda_j(t,u)du}{S_j(t,u)^2K_j(t,u)^2},\end{aligned}$$

which converges in probability to $\rho_j(t,s,s')$, by lemma 4.5.

4. The uniform consistency of $\hat{\rho}_j$ follows from theorem IV.1.2 in Andersen [1993] and the consistency of $\hat{p}_j(t)$.

□

The next two lemmas contain the covariance calculations necessary to determine the covariance structure of the asymptotic joint distribution of the sequential weighted Kaplan-Meier estimator at different calendar times.

Lemma 4.7. *For* $t \geqslant 0$ *and* $s, s' \in [0,L]$ *let*

$$\bar{S}(t,s) = \frac{1}{\sqrt{n}}\sum_{j=1}^{J} S_j(s)\int_0^s \frac{\bar{M}_j^{(n)}(t,du)}{S_j(u)K_j(t,u)},$$

and

$$\rho_S(t,s,s') = \frac{1}{\pi(t)} \sum_{j=1}^{J} p_j S_j(s) S_j(s') \int_0^{s\wedge s'} \frac{\lambda_j(u)}{S_j(u) K_j(t,u|R_i \leqslant t)} du.$$

Then

$$\rho_S(t \vee t', s, s') = \mathrm{Cov}(\bar{S}(t,s), \bar{S}(t',s')).$$

Under assumption 4.4 ρ_S is uniformly (in s and s') consistently estimated by

$$\hat{\rho}_S(t,s,s') = \frac{n}{n(t)} \sum_{j=1}^{J} \hat{p}_j(t) \hat{S}_j(t,s) \hat{S}_j(t,s') n(t) \int_0^{s\wedge s'} J_j^{(n)}(t,u) \frac{\bar{N}_j^{(n)}(t,du)}{\bar{Y}_j^{(n)}(t,u)^2}.$$

Proof. Define the processes $Z_j(x,\cdot)$ for $x = t$ and $x = t'$ and $j = 1,\ldots,J$ by

$$Z_j(x,s) = \int_0^s \frac{\bar{M}_j^{(n)}(x,du)}{S_j(u) K_j(x,u)} \quad s \in [0,L].$$

Then

$$\bar{S}(x,u) = \frac{1}{\sqrt{n}} \sum_{j=1}^{J} S_j(u) Z_j(x,u),$$

Since $Z_j(t,s)$ and $Z_{j'}(t',s')$ are independent random variables for any s, s',

$$\mathrm{Cov}(\bar{S}(t,s), \bar{S}(t',s')) = \frac{1}{n} \sum_{j=1}^{J} S_j(s) S_j(s') \mathrm{Cov}(Z_j(t,s), Z_j(t',s')).$$

The processes $Z_j(t,\cdot)$ and $Z_j(t',\cdot)$ are local square integrable $\mathcal{F}_R$-martingales for $j = 1,\ldots,J$. By the Ito isometry (theorem A.22) and theorem A.27,

$$\begin{aligned}
\mathrm{Cov}(Z_j(t,s), Z_{j'}(t',s')) &= E\left[\int_0^{s\wedge s'} \frac{\left\langle \bar{M}_j^{(n)}(t,du), \bar{M}_{j'}^{(n)}(t',du) \right\rangle}{S_j(u) S_{j'}(u) K_j(t,u) K_{j'}(t',u)} \right] \\
&= \mathbb{1}\{j = j'\} \sum_{i=1}^{n} E\left[\int_0^{s\wedge s'} \frac{\bar{Y}_j^{(n)}(t\wedge t', u) \lambda_j(u) du}{S_j(u)^2 K_j(t,u) K_j(t',u)} \right],
\end{aligned}$$

where the last equality follows from equations (4.1.1) and (4.1.2). Hence

$$\begin{aligned}\operatorname{Cov}(\bar{S}(t,s),\bar{S}(t',s')) &= \frac{1}{n}\sum_{j=1}^{J} S_j(s)S_j(s')E\left\{\int_0^{s\wedge s'} \frac{\bar{Y}_j^{(n)}(t\wedge t',u)\lambda_j(u)du}{S_j(u)^2(u)K_j(t,u)K_j(t',u)}\right\}\\ &= \sum_{j=1}^{J} p_j S_j(s)S_j(s')\int_0^{s\wedge s'} \frac{\lambda_j(u)}{S_j(u)K_j(t\vee t',u)}du,\end{aligned}$$

where the last equality follows from Fubini's theorem and

$$E[n^{-1}\bar{Y}_j(t\wedge t',u)] = p_j S_j(u)K_j(t\wedge t',u).$$

Finally, the covariance formula follows by noting that

$$K_j(t\vee t',u) = K_j(t\vee t',u|R\leqslant t\vee t')\pi(t\vee t').$$

The consistency of the covariance estimator follows from lemma 4.6 and the consistency of $\hat{p}_j(t)$. □

Lemma 4.8. *Let*

$$\bar{n}(t,s) = \sum_{j=1}^{J}\sqrt{n}\left\{\frac{n_j(t)}{n} - p_j\frac{n(t)}{n}\right\}\frac{S_j(s)}{\pi(t)},$$

and

$$\rho_{\bar{n}}(t,s,s') = \pi(t)^{-1}\sum_{j=1}^{J}\left[p_j S_j(s)S_j(s') - \sum_{l=1}^{J} p_j p_l S_j(s)S_l(s')\right].$$

Then

$$\rho_{\bar{n}}(t\vee t',s,s') = \operatorname{Cov}(\bar{n}(t,s),\bar{n}(t',s')).$$

Under assumption 4.4 $\rho_{\bar{n}}$ can be estimated uniformly consistently by

$$\hat{\rho}_{\bar{n}}(t,s,s') = \frac{n}{n(t)}\sum_{j=1}^{J}\left[\hat{p}_j(t)\hat{S}_j(s)\hat{S}_j(s') - \sum_{l=1}^{J}\hat{p}_j(t)\hat{p}_l(t)\hat{S}_j(s)\hat{S}_l(s')\right],$$

for $s,s'\leqslant L$.

Proof. Let $n_{ij}(t,s) = \mathbb{1}\{X_i = j, R_i\leqslant t\} - p_j\mathbb{1}\{R_i\leqslant t\}$, such that

$$\bar{n}(t,s) = \frac{1}{\sqrt{n}}\sum_{j=1}^{J}\sum_{i=1}^{n} n_{ij}(t,s)\frac{S_j(s)}{\pi(t)},$$

and note that $E[n_{ij}(t,s)] = 0$. The random variables $n_{ij}(t,s)$ and $n_{i'j'}(t',s')$ are uncorrelated for $i \neq i'$, such that the covariance of $\bar{n}(t,s)$ and $\bar{n}(t',s')$ is given by

$$\begin{aligned}
\mathrm{Cov}(\bar{n}(t,s),\bar{n}(t',s')) &= \sum_{j=1}^{J}\sum_{j'=1}^{J} \mathrm{Cov}(n_{ij}(t,s),n_{ij'}(t',s'))\frac{S_j(s)S_{j'}(s')}{\pi(t)\pi(t')} \\
&= \sum_{j=1}^{J} \mathrm{Cov}(n_{ij}(t,s),n_{ij}(t',s'))\frac{S_j(s)S_j(s')}{\pi(t)\pi(t')} \\
&\quad + \sum_{j=1}^{J}\sum_{j'=1,j'\neq j}^{J} \mathrm{Cov}(n_{ij}(t,s),n_{ij'}(t',s'))\frac{S_j(s)S_{j'}(s')}{\pi(t)\pi(t')}
\end{aligned}$$

Remember that X_i and R_i are stoch. independent, and that $E[\mathbb{1}\{X_i = j\}] = p_j$ and $E[\mathbb{1}\{R_i \leqslant t\}] = \pi(t)$. Hence

$$\begin{aligned}
\mathrm{Cov}(n_{ij}(t,s),n_{ij}(t',s')) &= E[\mathbb{1}\{X_i = j, R_i \leqslant t\wedge t'\}] - 2p_jE[\mathbb{1}\{X_i = j, R_i \leqslant t\wedge t'\}] \\
&\quad + p_j^2E[\mathbb{1}\{R_i \leqslant t\wedge t'\}] \\
&= p_j(1-p_j)\pi(t\wedge t')
\end{aligned}$$

For $j \neq j'$,

$$\begin{aligned}
\mathrm{Cov}(n_{ij}(t,s),n_{ij'}(t',s')) &= -p_jE[\mathbb{1}\{X_i = j', R_i \leqslant t\wedge t'\}] - p_{j'}E[\mathbb{1}\{X_i = j, R_i \leqslant t\wedge t'\}] \\
&\quad + p_jp_{j'}\mathbb{1}\{R_i \leqslant t\wedge t'\} \\
&= -p_jp_{j'}\pi(t\wedge t')
\end{aligned}$$

Consistency of the covariance estimator follows from the consistency of $\hat{p}_j(t)$ and $\hat{S}_j$. □

Definition 4.9. *For any s, s', t, t' with $s, s' \leqslant L \leqslant t, t'$, let*

$$\rho(t\vee t', s, s') = \rho_S(t\vee t', s, s') + \rho_{\bar{n}}(t\vee t', s, s'),$$

and

$$\hat{\rho}(t\vee t', s, s') = \hat{\rho}_S(t\vee t', s, s') + \hat{\rho}_{\bar{n}}(t\vee t', s, s'),$$

where $\rho_S, \hat{\rho}_S$ and $\rho_{\bar{n}}, \hat{\rho}_{\bar{n}}$ are defined in lemmas 4.7 and 4.8.

Finally the joint distribution of the sequential weighted Kaplan-Meier estimator at different calendar times can be derived.

Theorem 4.10. *Under assumption 4.4,*

1. $\hat{S}(t,\cdot)$ *is a uniformly consistent estimator of* $S(\cdot)$ *on* $[0, L]$.

2. $\sqrt{n}\{\hat{S}(t,\cdot) - S(\cdot)\}$ *converges weakly in* $D[0, L]$ *to a mean-zero Gaussian process with covariance function* $(s, s') \mapsto \rho(t, s, s')$,

3. $\sqrt{n}\left(\hat{S}(t_1,\cdot) - S(\cdot), \ldots, \hat{S}(t_K,\cdot) - S(\cdot)\right)$ *converges weakly (in* $(D[0, L])^K$*) to a multivariate Gaussian process with covariance function* $(t, t', s, s') \mapsto \rho(t \vee t', s, s')$.

4. *for fixed* t*, the covariance function* $\rho(t,\cdot,\cdot)$ *can be estimated consistently uniformly in* s *and* s' *by* $\hat{\rho}(t,\cdot,\cdot)$*, as defined in definition 4.9.*

Proof. 1. Uniform consistency of $\hat{S}(t,\cdot)$ follows immediately from the uniform consistency of $\hat{S}_j(t,\cdot)$ for each j.

2. $\sqrt{n}\{\hat{S}(t,s) - S(s)\}$ can be written as the sum of two terms,

$$\begin{aligned}\sqrt{n}\{\hat{S}(t,s) - S(s)\} &= \sqrt{n}\left\{\sum_{j=1}^{J} \hat{p}_j(t)\hat{S}_j(t,s) - \sum_{j=1}^{J} p_j S_j(s)\right\} \\ &= \sum_{j=1}^{J} p_j \sqrt{n}\{\hat{S}_j(t,s) - S_j(s)\} + \sum_{j=1}^{J} \sqrt{n}\{\hat{p}_j(t) - p_j\}\hat{S}_j(t,s).\end{aligned}$$

By lemma 4.6,

$$\sum_{j=1}^{J} p_j \sqrt{n}\{\hat{S}_j(t,s) - S_j(s)\} = -\sum_{j=1}^{J} S_j(s)\frac{1}{\sqrt{n}}\int_0^s \frac{\bar{M}_j^{(n)}(t, du)}{S_j(u)K_j(t,u)} + o_P(1).$$

The second term can be written as

$$\begin{aligned}&\sum_{j=1}^{J} \sqrt{n}\left\{\frac{n_j(t)}{n} - p_j\frac{n(t)}{n}\right\}\frac{n}{n(t)}\hat{S}_j(t,s) \\ &= \sum_{j=1}^{J} \sqrt{n}\left\{\frac{n_j(t)}{n} - p_j\frac{n(t)}{n}\right\}\frac{S_j(s)}{\pi(t)} + o_p(1),\end{aligned}$$

since $\hat{S}_j(t,s) = S_j(s) + o_P(1)$ uniformly in s and $n(t)/n \to \pi(t) > 0$. This proves, that the process $\sqrt{n}\{\hat{S}(t,\cdot) - S(\cdot)\}$ is asymptotically equivalent to the sum of two processes:

$$\begin{aligned}\sqrt{n}\{\hat{S}(t,s) - S(s)\} = &-\sum_{j=1}^{J} S_j(s)\frac{1}{\sqrt{n}}\int_0^s \frac{\bar{M}_j^{(n)}(t,du)}{S_j(u)K_j(t,u)} \\ &+\sum_{j=1}^{J}\sqrt{n}\left\{\frac{n_j(t)}{n} - p_j\frac{n(t)}{n}\right\}\frac{S_j(s)}{\pi(t)} + o_P(1)\end{aligned} \tag{4.2.5}$$

The two processes in eq. (4.2.5) are uncorrelated since the first process is a $\mathcal{F}_R$-martingale and the second process is $\mathcal{F}_R(0)$-measurable. For each s, both processes are sums of iid random variables, by (2.), such that the limiting finite dimensional distributions are normal, by the multivariate central limit theorem, and the covariance functions are given by lemma 4.7 and 4.8 with $t = t'$. Tightness of the first process follows from the martingale central limit theorem (weak convergence implies tightness by theorem A.42). Tightness of the second process follows from the uniform continuity of each S_j on the compact interval $[0, L]$ and by the stochastic equicontinuity criterion for tightness (theorem A.41). Hence the sum of the two processes is also tight.

3. Tightness of the vector is clear, since each component is tight according to (3.) (see lemma A.40). Convergence of the finite-dimensional distributions follows immediately from the multivariate central limit theorem, since by eq. (4.2.5) each component of the vector is asymptotically linear. The covariance formula follows immediately from lemmas 4.7 and 4.8.

4. The uniform consistency follows from the consistency results in lemmas 4.7 and 4.8.

□

4.3 SEQUENTIAL TWO-SAMPLE METHODS

4.3.1 *Average hazard ratio*

In this section the results of the previous section are applied to derive the asymptotic joint distribution of the sequentially computed estimators of the average hazard ratio.

For group-sequential designs it is necessary to derive the asymptotic joint distribution of

$$\left(\sqrt{n(t_1)}\{\hat{\theta}_1(t_1)-\theta_1\},\ldots,\sqrt{n(t_K)}\{\hat{\theta}_1(t_K)-\theta_1\}\right)$$

for any finite number $K \geqslant 1$ of calendar times $t_1,\ldots,t_K$, where $\hat{\theta}_1(t)$ is the estimator of θ_1 based on all data available up to calendar time t. The following assumptions about the asymptotic properties of the sequentially computed estimators of the marginal survival functions S_0 and S_1 are made:

Assumption 4.11. *There exists* $L > 0$, *such that for* $i = 0, 1$, *and* $t, t' = t_1,\ldots,t_K$,

1. $n_i/n \to \nu_i > 0$, *as* $n \to \infty$,
2. $S_i(\cdot)$ *is continuous,*
3. $\hat{S}_i(t,\cdot)$ *is a uniformly consistent estimator of* $S_i(\cdot)$ *on* $[0, L]$,
4. $\sqrt{n}\left(\hat{S}_i(t_1,\cdot)-S_i(\cdot),\ldots,\hat{S}_i(t_K,\cdot)-S_i(\cdot)\right)$ *converges weakly in* $D[0, L]$ *to a multivariate Gaussian process with covariance function* $(t, t', s, s') \mapsto \rho_i(t \vee t', s, s')$,
5. *there exists a uniformly consistent estimator* $\hat{\rho}_i(t,\cdot,\cdot)$ *of the covariance function* $\rho_i(t,\cdot,\cdot)$.

Assumption 4.11 holds e.g. for the weighted Kaplan-Meier estimator (theorem 4.10).

Remark 4.12. *In assumption 4.11(1.) the allocation ratio, i.e. the limit* ν_i *of* $n_i(t)/n(t)$, *does not depend on the calendar time. Changes of the allocation ratio are not allowed in a group-sequential trial, because this would destroy the independent increments structure. For the adaptive designs in the next chapter, where the independent increments structure is not needed, calendar time-dependence of the allocation ratio is allowed.*

From assumption 4.11 and the independence of the samples it follows, that

$$\begin{pmatrix} \sqrt{n}\{\hat{S}_0(t_1,\cdot)-S_0(\cdot)\} \\ \vdots \\ \sqrt{n}\{\hat{S}_0(t_K,\cdot)-S_0(\cdot)\} \\ \sqrt{n}\{\hat{S}_1(t_1,\cdot)-S_1(\cdot)\} \\ \vdots \\ \sqrt{n}\{\hat{S}_1(t_K,\cdot)-S_1(\cdot)\} \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} U_0(t_1,\cdot) \\ \vdots \\ U_0(t_K,\cdot) \\ U_1(t_1,\cdot) \\ \vdots \\ U_1(t_K,\cdot) \end{pmatrix}$$

as $n \to \infty$, where $U_i(t_l, \cdot)$, $l = 1, \dots, K$, $i = 0, \dots, k$ are mean-zero Gaussian processes with covariance functions $\nu_i^{-1}\rho_i(t_l, \cdot, \cdot)$ and

$$\mathrm{Cov}(U_i(t,s), U_j(t',s')) = \mathbb{1}\{i=j\}\nu_i^{-1}\rho_i(t \wedge t', s, s').$$

As in the fixed design case of chapter 3, the asymptotic covariance function $\sqrt{n}\{\log \hat{S}_i(t_l, \cdot) - \log S_i(\cdot)\}$ is

$$(s,s') \mapsto \rho_i^{\log}(t_l, s, s') = \frac{\rho_i(t_l, s, s')}{S_i(s)S_i(s')}.$$

Estimators of χ_i, G and θ_i are obtained by replacing the survival functions S_i with their estimators $\hat{S}_i$. The estimator of the average hazard ratio from definition 3.11 is defined in the two-sample case with explicit dependence on the calendar time.

Definition 4.13. *For $i = 0, 1$ and $t = t_1, \dots, t_K$ define the estimate of the weight function $G(L)$ by*

$$\hat{G}(t, L) = \hat{S}_0(t, L)\hat{S}_1(t, L),$$

the estimate of the relative risk $\chi_i = P(\min(T_j, L) > \min(T_i, L))$ by

$$\hat{\chi}_i(t) = -\int_0^L \hat{S}_j(t, s)\hat{S}_i(t, ds) \quad i \neq j,$$

and the estimate of the average hazard ratio $\theta_i(L)$ by

$$\hat{\theta}_i(t) = \frac{\hat{\chi}_i(t)}{1 - \hat{G}(t, L)}.$$

Definition 4.14 (Variance components). *For any $t \geqslant 0$ and $i, j, l = 0, 1$, let*

$$\begin{aligned} V_\chi(t) &= B_{011}(t) + B_{100}(t) + C_1(t) - 2A_{10}(t), \\ V_{\chi G}(t) &= A_{10}(t) - A_{01}(t) - C_1(t), \\ V_G(t) &= C_0(t) + C_1(t). \end{aligned}$$

where

$$\begin{aligned} A_{ij}(t) &= \nu_i^{-1} G(L) \int_0^L \rho_i^{\log}(t, s, L) G(s) \log S_j(ds), \\ B_{ijl}(t) &= \nu_i^{-1} \int_0^L \int_0^L \rho_i^{\log}(t, s, u) G(s) G(u) \log S_j(ds) \log S_l(du), \\ C_i(t) &= \nu_i^{-1} G^2(L) \rho_i^{\log}(t, L, L), \end{aligned}$$

The next lemma is the two-sample special case of lemma 3.12:

Lemma 4.15. *Under assumption 4.11, for any* $t = t_1, \ldots, t_K$, *as* $n \to \infty$,

$$\begin{aligned}\sqrt{n}\{\hat{x}_1(t) - x_1\} = &-\int_0^L \sqrt{n}\{\hat{S}_0(t,s) - S_0(s)\}S_1(ds) \\ &+\int_0^L \sqrt{n}\{\hat{S}_1(t,s) - S_1(s)\}S_0(ds) \\ &-\sqrt{n}\{\hat{S}_1(t,L) - S_1(L)\}S_0(L) + o_p(1),\end{aligned}$$

and

$$\begin{aligned}\sqrt{n}\{\hat{G}(t,L) - G(L)\} = &S_1(L)\sqrt{n}\{\hat{S}_0(t,L) - S_0(L)\} \\ &+ S_0(L)\sqrt{n}\{\hat{S}_1(t,L) - S_1(L)\} + o_p(1).\end{aligned}$$

Theorem 4.16. *Under assumption 4.11, as* $n \to \infty$,

$$\begin{pmatrix} \sqrt{n(t_1)}\{\hat{\theta}_1(t_1) - \theta_1\} \\ \vdots \\ \sqrt{n(t_K)}\{\hat{\theta}_1(t_K) - \theta_1\} \end{pmatrix} \xrightarrow{\mathcal{L}} N(0, \Sigma_\theta), \tag{4.3.1}$$

where Σ_θ *is the* $m \times m$*-matrix with entries*

$$\sigma_\theta(t_l, t_j) = \sqrt{\pi(t_l)\pi(t_j)}\tilde{\sigma}(t_l \vee t_j, L) \quad 1 \leqslant l, j \leqslant K,$$

and

$$\tilde{\sigma}(t) = \frac{1}{\{1 - G(L)\}^2}\{V_x(t) + 2\theta_1 V_{xG}(t) + \theta_1^2 V_G(t)\}.$$

$\sigma_\theta(t_l, t_j)$ *can be estimated consistently by replacing* S_i, ρ_i, ν_i, $\pi(t_l \vee t_j)$, *and* G *in the definition of* $\sigma_\theta(s,t)$ *by* $\hat{S}_i$, $\hat{\rho}_i$, $n_i(t_l \vee t_j)/n(t_l \vee t_j)$, $n(t_l \vee t_j)/n$, *and* $\hat{G}$. *Note that the resulting estimator uses only data observed up to calendar time* t.

Proof. Define the linear map $\phi : (D[0,L])^{2m} \to \mathbb{R}^{2m}$ by

$$\begin{pmatrix} X_{01} \\ \vdots \\ X_{0m} \\ X_{11} \\ \vdots \\ X_{1m} \end{pmatrix} \mapsto \begin{pmatrix} -\int_0^L X_{01}(s)S_1(ds) + \int_0^L X_{11}(s)S_0(ds) - X_{11}(L)S_0(L) \\ S_1(L)X_{01}(L)) + S_0(L)X_{11}(L) \\ \vdots \\ -\int_0^L X_{0m}(s)S_1(ds) + \int_0^L X_{0m}(s)S_0(ds) - X_{0m}(L)S_0(L) \\ S_1(L)X_{0m}(L) + S_0(L)X_{1m}(L) \end{pmatrix}$$

Define the vectors

$$\begin{aligned} \bar{U} &= (U_0(t_1,\cdot),\ldots,U_0(t_K,\cdot),U_1(t_1,\cdot),\ldots,U_1(t_K,\cdot)), \\ \bar{S}^{(n)} &= (\hat{S}_0(t_1,\cdot),\ldots,\hat{S}_0(t_K,\cdot),\hat{S}_1(t_1,\cdot),\ldots,\hat{S}_1(t_K,\cdot)), \\ \bar{S} &= (S_0(\cdot),\ldots,S_0(\cdot),S_1(\cdot),\ldots,S_1(\cdot)). \end{aligned}$$

It then follows from lemma 4.15 and the continuous mapping theorem, that as $n \to \infty$

$$\begin{pmatrix} \sqrt{n}\{\hat{x}_1(t_1) - x_1\} \\ \sqrt{n}\{\hat{G}(t_1,L) - G(L)\} \\ \vdots \\ \sqrt{n}\{\hat{x}_1(t_K) - x_1\} \\ \sqrt{n}\{\hat{G}(t_K,L) - G(L)\} \end{pmatrix} = \sqrt{n}\left\{\phi(\bar{S}^{(n)}) - \phi(\bar{S})\right\} \xrightarrow{\mathcal{L}} \phi(\bar{U})$$

Since ϕ is linear, $\phi(\bar{U})$ has a multivariate normal distribution with mean zero and covariance matrix

$$\Sigma_K = \begin{pmatrix} \Sigma(t_1) & \Sigma(t_2) & \cdots & \Sigma(t_K) \\ & \Sigma(t_2) & & \vdots \\ & & \ddots & \\ & & & \Sigma(t_K) \end{pmatrix}$$

where

$$\Sigma(t) = \begin{pmatrix} V_x(t) & V_{xG}(t) \\ V_{xG}(t) & V_G(t) \end{pmatrix}$$

The derivation of $\Sigma(t)$ is similar to the calculations in lemma 3.14. Denote the i-th component of the vector $\phi(\bar{U})$ by f_i and remember that

$$E[U_i(t,s)U_i(t',s')] = \nu_i^{-1}\rho_i(t \vee t', s, s') = \nu_i^{-1}\rho_i^{\log}(t \vee t', s \wedge s')S_i(s)S_i(s')$$

Then, by Fubini's theorem,

$$\begin{aligned}
&\mathrm{Cov}(f_{2i-1}, f_{2j-1}) = \mathrm{E}[f_{2i-1}f_{2j-1}] \\
&= \mathrm{E}[\int_0^L \int_0^L U_0(t_i, s)U_0(t_j, u)\mathrm{d}S_1(s)\mathrm{d}S_1(u)] + \mathrm{E}[\int_0^L \int_0^L U_1(t_i, s)U_1(t_j, u)\mathrm{d}S_0(s)\mathrm{d}S_0(u)] \\
&\quad + \mathrm{E}[U_1(t_i, L)U_1(t_j, L)]S_0^2(L) - \mathrm{E}[\int_0^L U_1(t_i, s)U_1(t_j, L)\mathrm{d}S_0(L)\mathrm{d}S_0(s)] \\
&\quad - \mathrm{E}[\int_0^L U_1(t_j, s)U_1(t_i, L)S_0(L)\mathrm{d}S_0(s)] \\
&= B_{011}(t_i \vee t_j) + B_{100}(t_i \vee t_j) + C_1(t_i \vee t_j) - 2A_{10}(t_i \vee t_j) = V_x(t_i \vee t_j)
\end{aligned}$$

Similarly,

$$\mathrm{Cov}(f_{2i}, f_{2j}) = \mathrm{E}[f_{2i}f_{2j}] = V_G(t_i \vee t_j),$$

$$\mathrm{Cov}(f_{2i-1}, f_{2j}) = \mathrm{E}[f_{2i-1}f_{2j}] = V_{xG}(t_i \vee t_j).$$

Finally, consider the function $\psi : \mathbb{R}^K \times (0,1)^K \to \mathbb{R}^K$, defined by

$$(a_1, b_1, \ldots, a_K, b_K) \mapsto \left(\frac{a_1}{1-b_1}, \ldots, \frac{a_K}{1-b_K}\right)^T$$

Define the vectors

$$x_G = (x_1, G(L), \ldots, x_1, G(L))^T,$$

$$\hat{x}_G = (\hat{x}_1(t_1), \hat{G}(t_1, L), \ldots, \hat{x}_1(t_K), \hat{G}(t_K, L))^T.$$

Note that $\psi(x_G) = (\theta_1, \ldots, \theta_1)^T$. The derivative of ψ at the point x_G is given by the $K \times 2K$-matrix

$$V_K = \frac{1}{1-G(L)} \begin{pmatrix} 1 & \theta_1 & & & & \\ & & 1 & \theta_1 & & \\ & & & \ddots & \ddots & \\ & & & & 1 & \theta_1 \end{pmatrix}$$

Applying the delta method again,

$$\begin{pmatrix} \sqrt{n}\{\hat{\theta}_1(t_1) - \theta_1\} \\ \vdots \\ \sqrt{n}\{\hat{\theta}_1(t_K) - \theta_1\} \end{pmatrix} = \sqrt{n}\{\psi(\hat{x}_G) - \psi(x_G)\} \xrightarrow{\mathcal{L}} N(0, V_K \Sigma_K V_K^T) \tag{4.3.2}$$

Now define the matrix the $K \times K$-diagonal matrices

$$V_\pi^{(n)} = \begin{pmatrix} \sqrt{\frac{n(t_1)}{n}} & & \\ & \ddots & \\ & & \sqrt{\frac{n(t_K)}{n}} \end{pmatrix}, \quad V_\pi = \begin{pmatrix} \sqrt{\pi(t_1)} & & \\ & \ddots & \\ & & \sqrt{\pi(t_K)} \end{pmatrix}$$

Clearly $V_\pi^{(n)}$ converges to V_π a.s., thus by the lemma of Slutsky,

$$\begin{pmatrix} \sqrt{n(t_1)}\{\hat{\theta}_1(t_1) - \theta_1\} \\ \vdots \\ \sqrt{n(t_K)}\{\hat{\theta}_1(t_K) - \theta_1\} \end{pmatrix} = V_\pi^{(n)} \sqrt{n}\{\psi(\hat{x}_G) - \psi(x_G)\} \xrightarrow{\mathcal{L}} N(0, V_\pi V_K \Sigma_K V_K^T V_\pi).$$

The result now follows, since $\Sigma_\theta = V_\pi V_K \Sigma_K V_K^T V_\pi$. □

A similar result holds for the sequentially computed log-odds of the average hazard ratio:

Corollary 4.17. *Under assumption 4.11, as* $n \to \infty$,

$$\begin{pmatrix} \sqrt{n(t_1)}\{\hat{\beta}(t_1) - \beta\} \\ \vdots \\ \sqrt{n(t_K)}\{\hat{\beta}(t_K) - \beta\} \end{pmatrix} \xrightarrow{\mathcal{L}} N(0, \Sigma_\beta), \tag{4.3.3}$$

where $\Sigma_\beta(L)$ *is the* $m \times m$*-matrix with entries*

$$\sigma_\beta(t_l, t_j) = \frac{\sigma_\theta(t_l, t_j)}{\theta_1^2(1-\theta_1)^2} \quad 1 \leqslant l, j \leqslant K.$$

$\sigma_\beta(t_l, t_j)$ *can be estimated consistently by*

$$\frac{\hat{\sigma}_\theta(t_l, t_j)}{\hat{\theta}_1(t_l \vee t_j)^2\{1 - \hat{\theta}_1(t_l \vee t_j)\}^2},$$

where $\hat{\sigma}_\theta$ *is the consistent estimator of* σ_θ *from theorem 4.16.*

Corollary 4.17 is a consequence of theorem 4.16 and the delta method, since $\hat{\beta}(t) = \log(\hat{\theta}_1(t)) - \log(1 - \hat{\theta}_1(t))$ (confer the proof of corollary 3.22).

Independent increments

Consider the vector of standardized test statistics $(\hat{Z}(t_1), \ldots, \hat{Z}(t_K))^T$, where

$$\hat{Z}(t) = \frac{\sqrt{n(t)}\{\hat{\theta}_1(t) - \theta_1\}}{\sqrt{\hat{\sigma}^2_\theta(t)}} \quad t = t_1, \ldots, t_K,$$

and let $\sigma^2_\theta(t) = \sigma_\theta(t,t)$. Then from theorem 4.16 it follows, that this vector converges in distribution to a vector $\bar{Z} = (Z_1, \ldots, Z_K)^T$, which has a multivariate normal distribution with mean zero and and covariances given by

$$Cov(Z_i, Z_j) = \frac{\sigma_\theta(t_i, t_j)}{\sqrt{\sigma^2_\theta(t_i)}\sqrt{\sigma^2_\theta(t_j)}} = \sqrt{\frac{\tilde{\sigma}(t_i \vee t_j)}{\tilde{\sigma}(t_i \wedge t_j)}}.$$

Thus $\bar{Z}$ has the canonical joint distribution with information levels $\{\mathcal{I}_1, \ldots, \mathcal{I}_K\}$, where

$$\mathcal{I}_j = \frac{1}{\tilde{\sigma}(t_j)} = \frac{\pi(t_j)}{\sigma^2_\theta(t_j)}.$$

The information levels can be estimated by

$$\hat{\mathcal{I}}_j = \frac{\hat{\pi}(t_j)}{\hat{\sigma}^2_\theta(t_j)}.$$

Equivalently, the vector

$$\left(Z_1\sqrt{\mathcal{I}_1}, \ldots, Z_K\sqrt{\mathcal{I}_K}\right)^T$$

is multivariate normal with mean zero and covariances

$$Cov\left(Z_i\sqrt{\mathcal{I}_i}, Z_j\sqrt{\mathcal{I}_j}\right) = \mathcal{I}_{i\wedge j},$$

i.e. it has the independent increments property. Thus group-sequential rejection boundaries can be derived with standard methods as described in Jennison and Turnbull [1999]. The variance of the increments is

$$Var\left(Z_k\sqrt{\mathcal{I}_k} - Z_{k-1}\sqrt{\mathcal{I}_{k-1}}\right) = \mathcal{I}_k - \mathcal{I}_{k-1}.$$

4.3.2 *Restricted mean survival*

The next result is a sequential version of theorem 3.25, which can be easily derived using the existing theory of the previous chapters. Such a result has already been proved by Murray and Tsiatis [1999] based on the Kaplan-Meier estimator only. The weight function $\hat{w}$ and its limit w may now depend on the calendar time t.

Theorem 4.18. *If for each* $t = t_1, \ldots, t_K$, $\hat{w}(t, \cdot)$ *converges uniformly in probability to a bounded function* $w(t, \cdot)$ *on* $[0, L]$, *then, under assumption 4.11, as* $n \to \infty$,

$$\begin{pmatrix} \sqrt{n(t_1)} \int_0^L \hat{w}(t_1, u) \left\{ \hat{S}_0(t_1, u) - \hat{S}_1(t_1, u) \right\} du \\ \vdots \\ \sqrt{n(t_K)} \int_0^L \hat{w}(t_1, u) \left\{ \hat{S}_0(t_K, u) - \hat{S}_1(t_K, u) \right\} du \end{pmatrix} \xrightarrow{\mathcal{L}} N(0, \Sigma_\mu),$$

where $\Sigma_\mu = (\sigma(t_l, t_j))_{1 \leqslant l, j \leqslant K}$ *is the* $K \times K$*-matrix with entries*

$$\begin{aligned} \sigma(t_l, t_j) = \sqrt{\pi(t_l)\pi(t_j)} \int_0^L \int_0^L & w(t_l, u) w(t_j, v) \{ \nu_0^{-1} \rho_0(t_l \vee t_j, u, v) \\ & + \nu_1^{-1} \rho_1(t_l \vee t_j, u, v) \} du dv. \end{aligned}$$

The proof is another application of the functional delta method, which is trivial here, since the mapping is linear, and is similar to the proof of theorem 4.16. For a proof of the special case with the Kaplan-Meier estimator only without the delta method see Murray and Tsiatis [1999].

Remark 4.19. *The test statistics from theorem 4.18 have the independent increments property if and only if the weight function* w *does not depend on the calendar time* t.

4.4 SIMULATIONS

All simulations of group-sequential trials in this section were done with O'Brien-Fleming boundaries (O'Brien and Fleming [1979]).

4.4.1 *Type-I-error*

Table 4.1 shows the type-I-error of the AHR(β), RMS and LR methods for different maximum number of interim analyses under the null hypothesis of no treatment effect. The survival times were exponentially distributed with rate parameter 1 in both

groups. The censoring times were also exponentially distributed with rate parameter 1/2.34 in both groups truncated at 10. The maximum sample size was 500, the maximum information was 250 events and the maximum calendar time was 10. Interim analyses were performed every 50 events. Recruitment was uniform on $[0, 10]$. The truncation point L for the AHR and RMS methods was 2. The nominal significance level 0.025 is maintained closely for all methods. It is interesting, that the AHR(β) method is conservative in all scenarios.

Table 4.1: One-sided type-I-error of the AHR(β), RMS and LR methods for different maximum number of interim analyses.[a]

	Type-I-error		
K	AHR(β)	RMS	LR
1	0.0248	0.0259	0.0255
2	0.0252	0.0270	0.0249
3	0.0244	0.0262	0.0255
4	0.0241	0.0269	0.0248
5	0.0243	0.0264	0.0254

[a] 10^5 replications, maximum information = 250 events, equally spaced interim analyses, $\alpha = 0.025$

4.4.2 *Power*

The same simulation scenarios as in Murray and Tsiatis [1999] are used. The first scenario is a proportional hazards alternative with exponential survival times (rate 1) in the control group and exponential survival times (rate 0.655) in the experimental group. The median survival time is $\log(2)/0.655 \approx 1$ in the experimental group and $\log(2) \approx 0.7$ in the control group. In the second scenario the experimental group is now Weibull distributed with shape parameter 1.5 and scale parameter 1/0.737, whereas the survival times in the control group have the same distribution as in the first scenario. This results in non-proportional hazards. The median survival time is

again approx. 1 in the experimental group. The corresponding survival functions are shown in figure 4.1.

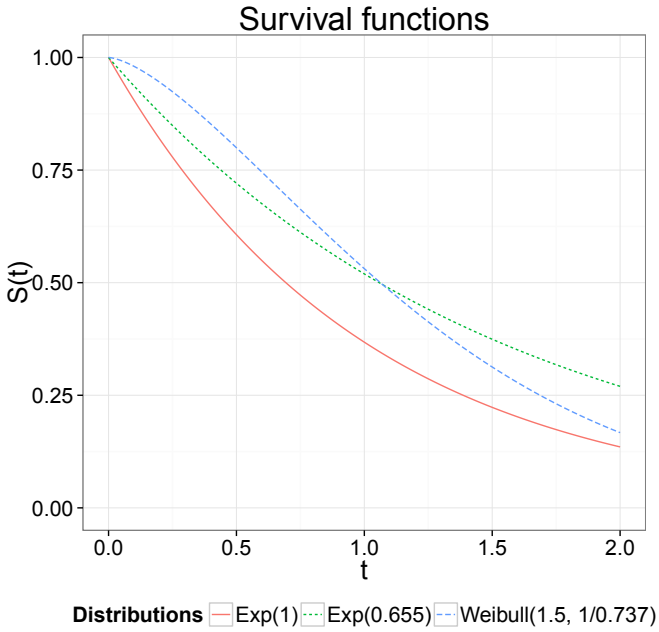

Figure 4.1: Survival functions of the control and the experimental group in both power scenarios

The maximum information was set to 239 events, such that the log-rank test achieves approx. 90% power. The truncation point L was set to 2 in both scenarios. The results of 10^5 simulation runs for scenario 1 are shown in table 4.2, the results for scenario 2 are shown in table 4.3. In scenario 1 all methods are almost identical with respect to power and average number of events. The log-rank outperforms the other methods slightly, as is expected under proportional hazards. However, in scenario 2 the power of the log-rank drops to about 80%, whereas the power of the other two methods stays above 90%.

Table 4.2: Power and average number of events of the AHR(β), RMS and LR methods for different maximum number of interim analyses in power scenario 1 (proportional hazards alternative).[a]

	Power			ANE [c]		
K [b]	AHR(β)	RMS	LR	AHR(β)	RMS	LR
1	0.87	0.87	0.90	239	239	239
2	0.86	0.87	0.90	214	209	210
3	0.86	0.87	0.89	197	193	193
4	0.86	0.87	0.89	189	184	185
5	0.86	0.87	0.89	184	180	179

[a] 10^5 replications, maximum information = 239 events, equally spaced interim analyses, $\alpha = 0.025$
[b] maximum number of interim analyses
[c] Average Number of Events

Table 4.3: Power and average number of events of the AHR(β), RMS and LR methods for different maximum number of interim analyses in power scenario 2 (non-proportional hazards alternative).[a]

	Power			ANE [c]		
K [b]	AHR(β)	RMS	LR	AHR(β)	RMS	LR
1	0.94	0.91	0.79	239	239	239
2	0.94	0.90	0.79	202	205	210
3	0.94	0.90	0.79	184	188	197
4	0.94	0.90	0.79	174	180	190
5	0.94	0.90	0.80	169	175	185

[a] 10^5 replications, maximum information = 239 events, equally spaced interim analyses, $\alpha = 0.025$
[b] maximum number of interim analyses
[c] Average Number of Events

5

ADAPTIVE DESIGN METHODS

This chapter concerns methods for trials with data-dependent interim adaptations, such as sample size reassessement or enrichment designs with subgroup selection, where information from a discrete surrogate variable is used in the interim decision. If data-dependent changes are made, the sequential methods of chapter 4 do not control the type-I-error in general. The approach used in this chapter is based on *combination functions*, which combine independent stage-wise test statistics to an overall test statistic. Independence of the stage-wise test statistics is a critical assumption here for the control of the type-I-error rate.

Suppose a two-stage trial is planned with sample sizes n_1 in the first and n_2 in the second stage and that the data is normally distributed with mean 0 and variance 1 under the null hypothesis. After n_1 patients have been recruited an interim analysis is performed and the pre-planned second stage sample size n_2 is changed to $\tilde{n}_2$ based on the unblinded data of the first stage. After $\tilde{n}_2$ new patients have been recruited in the second stage the final analysis is performed. Denote the standardized mean of stage i by Z_i, $i = 1, 2$. Z_i is standard normal distributed for $i = 1, 2$ under the null hypothesis. The test statistics Z_1 and Z_2 are independent, since they are calculated from disjoint sets of independent observations. The overall standardized mean is

$$\tilde{Z} = \frac{1}{\sqrt{n_1 + \tilde{n}_2}} \sum_{i=1}^{\tilde{n}_2} X_i = \frac{\sqrt{n_1}}{\sqrt{n_1 + \tilde{n}_2}} Z_1 + \frac{\sqrt{n_2}}{\sqrt{n_1 + \tilde{n}_2}} Z_2$$

In general $\tilde{Z}$ will not be standard normally distributed for data dependent choices of $\tilde{n}_2$, leading to an inflation of the type-I-error. If instead Z_1 and Z_2 are combined with pre-fixed weights w_1 and w_2, with $w_1^2 + w_2^2 = 1$,

$$Z = w_1 Z_1 + w_2 Z_2,$$

then the overall Z-score Z is standard normally distributed. The test can be written in terms of a p-value combination function $C(p_1, p_2)$, where p_i is the p-value from stage i. The Z-test above corresponds to the inverse normal combination test, where

$$C(p_1, p_2) = 1 - \Phi(w_1 \Phi^{-1}(1 - p_1) + w_2 \Phi^{-1}(1 - p_2)),$$

where Φ denotes the cumulative distribution function of the standard normal distribution. The trial is stopped for futility, if $p_1 > \alpha_0$ and the null hypothesis is rejected early, if $p_1 < \alpha_1$. For arbitrary combination functions the critical value c_α is chosen, such that

$$\alpha_1 + \int_{\alpha_1}^{\alpha_0} \int_0^1 \mathbb{1}\{C(p_1, p_2) \leqslant c_\alpha\} dp_2 dp_1 = \alpha.$$

Strictly speaking, stochastic independence of the stage-wise p-values is sufficient, but not necessary for type-I-error control of the combination test. Brannath et al. [2002] has shown, that the type-I-error of the combination test is controlled, when the p-values are only *p-clud*, i.e. the distribution of p_1 and the conditional distribution of p_2 given p_1 are stochastically larger than or equal to the uniform distribution on $[0, 1]$. In trials, where the response is immediate, stochastic independence of the p-values can be achieved by separating the observations into disjunct sets, those recruited before and those recruited after the interim analysis. In survival trials the situation is more complicated, because of overrunning patients, i.e. patients recruited before the interim analysis and still followed-up after interim analysis. Adaptive group-sequential survival trials can be designed with the inverse normal method as described by Wassmer [2006]. At stage k the standardized test statistic is

$$Z_k^{**} = \left(\sum_{l=1}^{k} w_l^2 \right)^{-1/2} \sum_{l=1}^{k} w_l Z_l^*,$$

where Z_k^* is the standardized test statistic from stage k and $w_k > 0$ are prefixed weights, $k = 1, \ldots, K$. The test statistics $Z_1^{**}, \ldots, Z_K^{**}$ have (asymptotically) the canonical joint distribution with information levels

$$\mathcal{I}_k = \left(\sum_{l=1}^{k} w_l^2 \right)^{-1/2},$$

if the stage-wise test statistics Z_k^* are jointly stochastically independent and (asymptotically) standard normal. This means, that the standard group-sequential rejection

boundaries can be used. To ensure control of the type-I-error, the weights must remain fixed. Independence of the stage-wise test statistics Z_k^* can be achieved by exploiting the independent increments property of the sequential test statistics $Z(t)$ (confer theorem 4.16 and 4.18). The stage-wise test statistics are then defined as

$$\begin{aligned} Z_1^* &= Z(t_1) \\ Z_k^* &= \frac{\sqrt{\mathcal{I}(t_k)}Z(t_k) - \sqrt{\mathcal{I}(t_{k-1})}Z(t_{k-1})}{\sqrt{\mathcal{I}(t_k) - \mathcal{I}(t_{k-1})}} \quad k \geqslant 2, \end{aligned}$$

where $\mathcal{I}(t)$ is the information at calendar time t.

This approach works, as long as modifications of the design are based only on the value of the test statistic and the data of patients, who are no longer at risk at the time of the interim analysis. Using information from secondary endpoints from patients still at risk at the time of interim analysis is used may inflate the type-I-error as demonstrated by Bauer and Posch [2004] (confer figure 1.1 in chapter 1). The type-I-error in this scenario is approximately $2\alpha - \alpha^2$, where α is the nominal significance level. This problem is not restricted to time-to-event data and the log-rank test. A similar type-I-error inflation can also be observed e.g. for the t-test with a normal endpoint and a binary surrogate variable.

For discrete surrogate markers two approaches are possible to ensure strict type-I-error control in combination with the weighted Kaplan-Meier estimator. Section 5.1 describes the *patient-wise splitting* approach by Jenkins et al. [2011]. Section 5.2 describes the *stage-wise splitting* approach, which uses stage-wise left-truncated and right-censored data.

5.1 PATIENT-WISE SPLITTING

Jenkins et al. [2011] avoid the problem of type-I-error inflation, caused by prediction of future events based on surrogate data, by splitting the patient population into two disjunct sets, those recruited before the interim analysis and those recruited after the interim analysis (figure 5.1). The interim analysis decision is based on PFS only. The test statistics used in the combination test for the primary endpoint (OS) are calculated only at the end of the study. Hence, no efficacy testing at interim is possible in this design. The test statistics are stochastically independent, since they are calculated from disjunct patient populations.

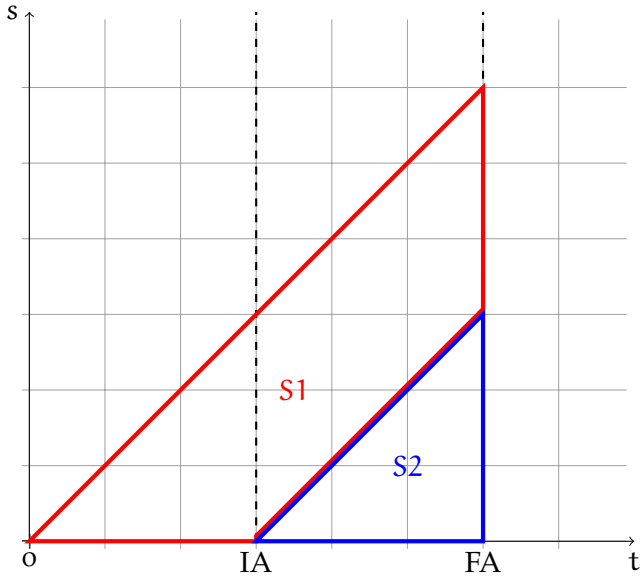

Figure 5.1: Lexis diagram illustrating patient-wise splitting. x-axis: calendar time, y-axis: survival time. S1: cohort of all patients recruited before the interim analysis (IA). S2: cohort of all patients recruited after the interim analysis. All patients in the cohorts S1 and S2 are followed-up until the final analysis (FA).

However, follow-up of stage 1 patients must not be changed in this design after the interim analysis (Jenkins et al. [2011] and Magirr et al. [2014]). The observed PFS events of patient still at risk at the time of the interim analysis can be used to predict future OS events, which in turn allows to predict the value of the stage 1 test statistic calculated at the end of the follow-up. Now changing the follow-up changes the number of observed events, i.e. the follow-up can be chosen to maximize the value of the conditional error, inflating the type-I-error. This can also be interpreted as an informative censoring problem. Changing follow-up changes the actual censoring time $\tilde{C}$, since $\tilde{C} = C \wedge (t - R)^+$, where t is the calendar time of the analysis, R is the entry time and C is the drop-out time. Since t is chosen using the surrogate information of the overrunning patients, the censoring time depends on the surrogate variable. Hence the survival and the censoring time are no longer stochastically independent, but stochastically independent conditional on the surrogate variable.

Covariate-dependent censoring can be handled in the Cox proportional hazards model by including the surrogate variables as additional covariates (besides the group indicator) in the model. For a discrete covariate this results in the stratified log-rank

test. However, as explained in chapter 1, this estimates only the conditional effect (conditional on the surrogate). The test statistics constructed with the weighted Kaplan-Meier estimator (chapter 3) do not suffer from this problem, while still controlling the type-I-error.

5.2 STAGE-WISE SPLITTING

If adaptive design changes have been made at an interim analysis, then the maximum likelihood estimator of the response probability $p_j = P(X = j)$ using all data, i.e. from patients recruited before and after the interim analysis, is biased (Brannath et al. [2006]). An unbiased estimator of the response rate is obtained by using only data from patients recruited after the adaptations have been made. Suppose adaptations are made at calendar time t_1 and the data is subsequently analyzed at calendar time t_2. Define the new response probability estimator by

$$\hat{p}_j(t_1, t_2) = \frac{n_j(t_1, t_2)}{n(t_1, t_2)}, \tag{5.2.1}$$

where

$$n_j(t_1, t_2) = \sum_{i=1}^{n} \mathbb{1}\{X_i = j, t_1 < R_i \leqslant t_2\},$$

is the number of patients recruited in stratum j in the interval $(t_1, t_2]$ and

$$n(t_1, t_2) = \sum_{j=1}^{J} n_j(t_1, t_2),$$

is the total number of patients recruited in the interval $(t_1, t_2]$. The recruitment rate in the interval $(t_1, t_2]$, $\pi_j(t_1, t_2) = P(t_1 < R \leqslant t_2 | X = j)$, is estimated by

$$\hat{\pi}(t_1, t_2) = \frac{n(t_1, t_2)}{n}.$$

It is tempting to define the weighted Kaplan-Meier estimator as

$$\tilde{S}(t_1, t_2, s) = \sum_{j=1}^{J} \hat{p}_j(t_1, t_2)\hat{S}_j(t_2, s), \tag{5.2.2}$$

where $\hat{S}_j(t_2, \cdot)$ is the stratum-specific Kaplan-Meier estimator in stratum j at calendar time t, i.e. the Kaplan-Meier estimator calculated from all in stratum j accrued up to calendar time t_2. However, the estimator in eq. (5.2.2) does not have the independent increments property. This makes it impossible to define (asymptotically) independent test statistics based on the (asymptotically) independent increments required for the combination test approach.

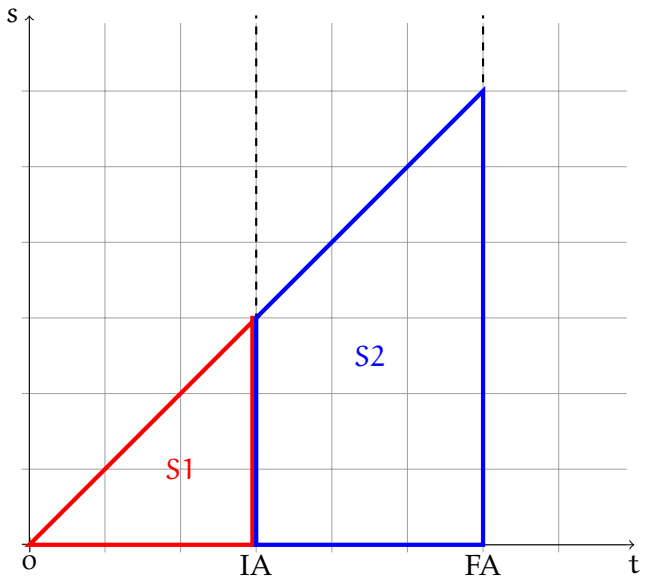

Figure 5.2: Lexis diagram illustrating stage-wise splitting. x-axis: calendar time, y-axis: survival time. S1: cohort of all patients recruited before the interim analysis (IA). Patients in this cohort are followed-up until the interim analysis. Patients still at-risk at time of the interim analysis are right-censored. S2: cohort of all patients recruited after the interim analysis up to the final analysis (FA) and all patients recruited before the interim analysis and still at-risk at the time of the interim analysis. Patients recruited before the interim analysis are left-truncated in this cohort. Patients in this cohort are followed-up until the final analysis.

The approach of Keiding et al. [1987] for two-stage trials is an alternative possibility to obtain (asymptotically) independent test statistics. Overrunning patients of the first stage are included in the first stage as right-censored observations (right-censored at the time of the interim analysis) and included in the second stage as left-truncated observations (figure 5.2). The left-truncation times are the calendar time of the interim analysis minus the recruitment times. This is in contrast to the usual group-sequential splitting, where the second stage data consists of all data accrued up to

the final analysis (cohort S1 *and* S2 of figure 5.1). Keiding et al. [1987] shows, that the partial likelihood is the product of two factors. The first factor contains the likelihood contribution of all patients recruited before the interim analysis, right-censored at the interim analysis. The second factor contains the likelihood contribution of the overrunnning patients, left-truncated at the time of the interim analysis, and all patients recruited after the interim analysis. This gives a heuristic argument for the independence of the stage-wise right-censored and left-truncated data. The weighted Kaplan-Meier estimator based on the stage-wise data for any two calendar times $t_1 < t_2$ and $s \in [0, L]$ is defined as

$$\hat{S}(t_1, t_2, s) = \sum_{j=1}^{J} \hat{p}_j(t_1, t_2)\hat{S}_j(t_1, t_2, s), \tag{5.2.3}$$

where $\hat{S}_j(t_1, t_2, s)$ is the stratum-specific Kaplan-Meier estimator in stratum j calculated from the data in stratum j, where each patient is possibly left-truncated at $(t_1 - R_i)^+$ and right-censored at $(t_2 - R_i)^+$. A formal proof of the asymptotic independence of the estimator in eq. (5.2.3) for subsequent pairs of calendar times (t_1, t_2) and (t_3, t_4) with $t_1 < t_2 \leqslant t_3 < t_4$, is based on the martingale central limit theorem (theorem A.45), by showing the weak convergence to a vector of independent Gaussian processes (theorem 5.5).

5.2.1 *Right-censoring and left-truncation*

Before theorem 5.5 can be proved, the counting process methods need to be extended to handle left-truncated data. Left-truncation is seamlessly handled by the counting process approach, with only minor modifications of the definitions of the counting process and the risk set. The counting process for patient i at calendar time t_2 started at $(t_1 - R_i)^+$ is defined as

$$N_i(t_1, t_2, s) = \mathbb{1}\{(t_1 - R_i)^+ < Y_i \wedge (t_2 - R_i)^+ \leqslant s, \delta_i(t_2) = 1\},$$

for $s \in [0, L]$. Moreover let

$$\bar{N}_j^{(n)}(t_1, t_2, s) = \sum_{i=1}^{n} \mathbb{1}\{X_i = j\}N_i(t_1, t_2, s).$$

A patient with observed survival time left-truncated at $(t_1 - R)^+$ is known to survive at least until $(t_1 - R)^+$, and hence is *not* at risk prior to $(t_1 - R)^+$. The risk set needs to be adjusted accordingly. Define the at-risk indicator for patient i at calendar times t_2 and survival time s with left-truncation at calendar time t_1 as

$$Y_i(t_1, t_2, s) = \mathbb{1}\{Y_i \wedge (t_2 - R_i)^+ \geqslant s > (t_1 - R_i)^+\}.$$

The number at risk is then given by

$$\bar{Y}_j^{(n)}(t_1, t_2, s) = \sum_{i=1}^{n} \mathbb{1}\{X_i = j, R_i \leqslant t_2\} Y_i(t_1, t_2, s).$$

The multiplicative structure of the intensity is preserved by random left-truncation (Andersen [1993]). This means, that the counting process $\mathbb{1}\{X_i = j\}N_i(t_1, t_2, s)$ has compensator

$$\int_0^s \mathbb{1}\{X_i = j\} Y_i(t_1, t_2, u) \lambda_j(u) du$$

with respect to the filtration $\mathcal{F}_R$ (definition 4.1). The process defined by

$$M_i(t_1, t_2, s) = \mathbb{1}\{X_i = j\}N_i(t_1, t_2, s) - \int_0^s \mathbb{1}\{X_i = j\} Y_i(t_1, t_2, u) \lambda_j(u) du$$

is a local square integrable martingale. Consequently, the process defined by

$$\bar{M}_j^{(n)}(t_1, t_2, s) = \sum_{i=1}^{n} \mathbb{1}\{X_j = j\} M_i(t_1, t_2, s) = \bar{N}_j^{(n)}(t_1, t_2, s) - \int_0^s \bar{Y}_j^{(n)}(t_1, t_2, u) \lambda_j(u) du,$$

is also a local square integrable martingale. The associated counting process $\bar{N}_j^{(n)}(t_1, t_2, \cdot)$ has continuous compensator

$$s \mapsto \int_0^s \bar{Y}_j^{(n)}(t_1, t_2, u) \lambda_j(u) du.$$

With the modified counting process and risk set the definitions of the Nelson-Aalen estimator is

$$\hat{\Lambda}_j(t_{l-1}, t_l, s) = \int_0^s J_j^{(n)}(t_{l-1}, t_l, s) \frac{\bar{N}_j^{(n)}(t_{l-1}, t_l, du)}{\bar{Y}_j^{(n)}(t_{l-1}, t_l, u)},$$

where

$$J_j^{(n)}(t_1, t_2, s) = \mathbb{1}\left\{\bar{Y}_j^{(n)}(t_1, t_2, s) > 0\right\}.$$

As before, the Kaplan-Meier estimator is the product integral of the Nelson-Aalen estimator

$$\hat{S}(t_{l-1}, t_l, s) = \prod_{(0,s]} \left\{1 - \hat{\Lambda}_j(t_{l-1}, t_l, ds)\right\}.$$

Suppose interim analyses are planned at calendar times $t_1 < \ldots < t_K$ and the trial starts at time $t_0 = 0$. The the following assumptions, which are similar to those made in chapters 3 and 4, are sufficient to derive the required large sample properties.

Assumption 5.1. *1.* T *and* C *are stochastically independent conditional on* X*,*

2. R *is independent of* T*,* C *and* X*,*

3. $P((t_l - R)^+ = 0) > 0$ *for* $l = 1, \ldots, K-1$*,*

4. $L < \sup\{s : \gamma_j(t_{l-1}, t_l, s) > 0\}$ *for all* $l = 1, \ldots, K$*, where*

$$\gamma_j(t_1, t_2, s) = P\left((t_1 - R)^+ < s \leqslant Y \wedge (t_2 - R)^+ | X = j\right),$$

5. $\bar{Y}_j^{(n)}(t_1, t_2, s) \to \infty$ *as* $n \to \infty$ *for any two calendar times* $t_1 < t_2$ *and all* $s \in [0, L]$*.*

The first two assumptions correspond to random right-censoring and random left-trunctation, respectively. The third and fourth assumptions imply, that the function $s \mapsto \gamma_j(t_{l-1}, t_l, s)$ is bounded away from 0 on the whole interval $[0, L]$ ensuring uniform consistency and weak convergence of the Nelson-Aalen and Kaplan-Meier estimators (example IV.1.7 in Andersen [1993]). This assumption holds as long as recruitment continues after any interim analysis, which is always the case for group-sequential designs, whenever the trial is not stopped at the interim analysis.

5.2.2 *Weighted Kaplan-Meier estimator*

The weighted Kaplan-Meier estimator for stage-wise right-censored and left-truncated data has been defined in eq. (5.2.3). Its large sample properties are derived in three steps. First unbiased estimation of the response probabilities is considered, then the stratum-specific Kaplan-Meier estimator based on left-truncated and right-censored data, and finally the weighted Kaplan-Meier estimator itself.

Estimation of the response probabilities

The large sample properties of the estimator $\hat{p}_j(t_1, t_2)$ follow directly from the law of large numbers and the central limit theorem.

Lemma 5.2. *Under assumption 5.1, for each* $j = 1, \ldots, J$ *and any two calendar times* $t_1 < t_2$,

1. $\hat{\pi}(t_1, t_2)$ *is a consistent estimator of* $\pi(t_1, t_2) = P(t_1 < R \leqslant t_2) > 0$,

2. $\hat{p}_j(t_1, t_2)$ *is a consistent estimator of* p_j *if no adaptations have been made within the interval* $(t_1, t_2]$,

3. $\hat{p}_j(t_1, t_2)$ *is asymptotically linear, i.e.*

$$\sqrt{n}\{\hat{p}_j(t_1, t_2) - p_j\} = \frac{1}{\pi(t_1, t_2)} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} p_{ij}(t_1, t_2) + o_p(1),$$

where $p_{ij}(t_1, t_2) = \mathbb{1}\{X_i = j, t_1 < R_i \leqslant t_2\} - p_j \mathbb{1}\{t_1 < R_i \leqslant t_2\}$,

4.

$$\sqrt{n}\{\hat{p}_j(t_1, t_2) - p_j\} \xrightarrow{\mathcal{L}} N(0, \sigma_j^2(t_1, t_2)),$$

where

$$\sigma_j^2(t_1, t_2) = \frac{p_j(1 - p_j)}{\pi(t_1, t_2)},$$

5. $\sigma_j^2(t_1, t_2)$ *is consistently estimated by*

$$\frac{\hat{p}_j(t_1, t_2)\{1 - \hat{p}_j(t_1, t_2)\}}{\hat{\pi}(t_1, t_2)},$$

6. *for any calendar times* $t_1 < t_2 < t_3$, $\hat{p}_j(t_1, t_2)$ *and* $\hat{p}_j(t_2, t_3)$ *are stochastically independent.*

Proof. 1. Follows immediately from the law of large numbers.

2. $E[n_j(t_1, t_2)] = P(X = j, t_1 < R \leqslant t_2) = p_j \pi(t_1, t_2)$, since X and R are independent. Thus, by the law of large numbers, (1.) and the lemma of Slutzky (lemma A.37).

$$\hat{p}_j(t_1, t_2) = \frac{n_j(t_1, t_2)}{n} \frac{1}{\hat{\pi}(t_1, t_2)} \xrightarrow{p} p_j$$

3.

$$\sqrt{n}\{\hat{p}_j(t_1, t_2) - p_j\} = \frac{1}{\hat{\pi}(t_1, t_2)} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} p_{ij}(t_1, t_2).$$

The result now follows from (1) and the lemma of Slutsky, by noting, that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} p_{ij}(t_1, t_2) = O_p(1),$$

by the central limit theorem.

4. Asymptotic normality of $\sqrt{n}\{\hat{p}_j(t_1,t_2) - p_j\}$ follows from (3.) and the central limit theorem. The asymptotic variance is

$$Var\left(\frac{1}{\pi(t_1,t_2)}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} p_{ij}(t_1,t_2)\right) = \frac{1}{\pi(t_1,t_2)^2}E\left[p_{ij}(t_1,t_2)^2\right] = \sigma_j^2(t_1,t_2).$$

5. Consistency of the variance estimator follows from the consistency of $\hat{p}_j(t_1,t_2)$ and the consistency of $\hat{\pi}(t_1,t_2)$.

6. Independence of $\hat{p}_j(t_1,t_2)$ and $\hat{p}_j(t_2,t_3)$ is clear, since the first uses only data from patients recruited in the interval $(t_1,t_2]$ and the second uses only data from patients recruited in the interval $(t_2,t_3]$.

□

Stratum-specific Kaplan-Meier estimator

The result from lemma 4.5 needs to be slightly modified to account for left-truncation of the survival times.

Lemma 5.3. *As* $n \to \infty$,

$$\sup_{s\in[0,L]}\left|\frac{1}{n}\bar{Y}_j^{(n)}(t_1,t_2,s) - p_j\gamma_j(t_1,t_2,s)\right| \xrightarrow{p} 0.$$

Proof. Write

$$\frac{1}{n}\bar{Y}_j^{(n)}(t_1,t_2,s) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{(t_1-R_i)^+ < s, X_i = j\} - \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{Y_i \wedge (t_2-R_i)^+ < s, X_i = j\},$$

and apply the Glivenko-Cantelli theorem to each term on the right hand side to obtain

$$\sup_{s\in[0,L]}\left|\frac{1}{n}\bar{Y}_j^{(n)}(t_1,t_2,s) - P\left((t_1-R)^+ < s < Y\wedge(t_2-R)^+, X = j\right)\right| \xrightarrow{p} 0,$$

as $n \to \infty$. The result now follows by noting, that

$$P\left((t_1-R)^+ < s \leqslant Y\wedge(t_2-R)^+, X = j\right) = \gamma_j(t_1,t_2,s)p_j,$$

□

The next lemma is a direct generalization of lemma 4.6. The proof is almost identical and is omitted. The main difference is, that now lemma 5.3 is used in the proof instead of lemma 4.5.

Lemma 5.4. *For* $j = 1, \dots, J$ *and* $l = 1, \dots, K$,

1. *uniformly in* $s \in [0, L]$,

$$\sqrt{n}\{\hat{S}_j(t_{l-1}, t_l, s) - S_j(s)\} = -\frac{S_j(s)}{p_j} \frac{1}{\sqrt{n}} \int_0^s \frac{\bar{M}_j^{(n)}(t_{l-1}, t_l, du)}{\gamma_j(t_{l-1}, t_l, u)} + o_P(1).$$

2. $\hat{S}_j(t_{l-1}, t_l, \cdot)$ *is a uniformly consistent estimator of* $S(\cdot)$ *on* $[0, L]$,

3. $\sqrt{n}\{\hat{S}_j(t_{l-1}, t_l, \cdot) - S_j(\cdot)\}$ *converges to a mean-zero Gaussian process with covariance function*

$$(s, s') \mapsto S_j(s) S_j(s') \rho_j(t_{l-1}, t_l, s, s'),$$

where

$$\rho_j(t_{l-1}, t_l, s \wedge s') = \frac{1}{p_j} \int_0^{s \wedge s'} \frac{\lambda_j(u) du}{\gamma_j(t_{l-1}, t_l, u)},$$

4. *a uniformly consistent of the covariance function in (3) is given by*

$$(s, s') \mapsto \hat{S}_j(t_{l-1}, t_l, s) \hat{S}_j(t_{l-1}, t_l, s') \hat{\rho}_j(t_{l-1}, t_l, s, s'),$$

where

$$\hat{\rho}_j(t_{l-1}, t_l, s) = \frac{n}{\hat{p}_j(t_{l-1}, t_l)} \int_0^s J_j^{(n)}(t_{l-1}, t_l, u) \frac{\bar{N}_j^{(n)}(t_l, du)}{\bar{Y}_j^{(n)}(t_{l-1}, t_l, u)^2}.$$

Large sample properties of the weighted Kaplan-Meier estimator

The next theorem is the main result of this section and proves uniform consistency and weak convergence of the weighted Kaplan-Meier estimator based on stage-wise left-truncated and right-censored data and gives a uniformly consistent estimator of the asymptotic covariance function, as well as a prove of the asymptotic independence of the stage-wise weighted Kaplan-Meier estimators.

Theorem 5.5. *Under assumption 5.1, for* $l = 1, \dots, K$,

1. $\hat{S}(t_{l-1}, t_l, \cdot)$ *is a uniformly consistent estimator of* $S(\cdot)$ *on* $[0, L]$,

2. $\sqrt{n}\{\hat{S}(t_{l-1}, t_l, \cdot) - S(\cdot)\}$ *converges weakly in* $D[0, L]$ *to a mean-zero Gaussian process with covariance function* $(s, s') \mapsto \rho_{sw}(t_{l-1}, t_l, s, s')$*, where*

$$\rho_{sw}(t_{l-1}, t_l, s, s') = \sum_{j=1}^{J} p_j S_j(s) S_j(s') \int_0^{s \wedge s'} \frac{\lambda(u)du}{\gamma_j(t_{l-1}, t_l, u)} + \sum_{j=1}^{J} \left\{ p_j S_j(s) S_j(s') - \sum_{j'=1}^{J} p_j p_{j'} S_j(s) S_{j'}(s') \right\},$$

3. *a uniformly consistent estimator of* $(s, s') \mapsto \rho_{sw}(t_{l-1}, t_l, s, s')$ *is given by*

$$\begin{aligned}
&\hat{\rho}_{sw}(t_{l-1}, t_l, s, s') \\
&= \sum_{j=1}^{J} \hat{p}_j(t_{l-1}, t_l) \hat{S}_j(t_{l-1}, t_l, s) \hat{S}_j(t_{l-1}, t_l, s') \int_0^{s \wedge s'} J_j^{(n)}(t_{l-1}, t_l, u) \frac{\bar{N}_j^{(n)}(t_l, du)}{\bar{Y}_j^{(n)}(t_{l-1}, t_l, u)} \\
&+ \sum_{j=1}^{J} \hat{p}_j(t_{l-1}, t_l) \hat{S}_j(t_{l-1}, t_l, s) \hat{S}_j(t_{l-1}, t_l, s') \\
&- \sum_{j=1}^{J} \sum_{j'=1}^{J} \hat{p}_j(t_{l-1}, t_l) \hat{p}_{j'}(t_{l-1}, t_l) \hat{S}_j(t_{l-1}, t_l, s) \hat{S}_{j'}(t_{l-1}, t_l, s'),
\end{aligned}$$

4. *the processes* $\sqrt{n}\{\hat{S}(t_{l-1}, t_l, \cdot) - S(\cdot)\}$, $l = 1, \ldots, K$ *are asymptotically jointly independent.*

Proof. The proof of the first three claims is very similar to the proof of theorem 4.10 in the previous chapter and is therefore omitted. The important result here is the asymptotic independence of the stage-wise Kaplan-Meier estimators.

The weighted Kaplan-Meier estimator is a function of the response rate estimators and the stratum-specific Kaplan-Meier estimators. The response rate estimators $\hat{p}_j(t_{l-1}, t_l)$, $j = 1, \ldots, J$, $l = 2, \ldots, K$ are stochastically independent, since the response rate $\hat{p}_j(t_{l-1}, t_l)$ in stratum j and stage l is estimated using only data from patients recruited in this stratum and stage. The stratum-specific Kaplan-Meier estimators of different strata are also stochastically independent. It remains to show, that the stratum-specific processes $\hat{S}_j(t_{l-1}, t_l, \cdot)$, $l = 1, \ldots, K$ within the same stratum j across different stages $l = 1, \ldots, K$ are asymptotically independent. This is accomplished by

proving, that these processes converge weakly to independent Gaussian processes. From lemma 5.4 it follows, that for $l = 1, \dots, K$ uniformly in $s \in [0, L]$

$$\frac{\sqrt{n}\{\hat{S}_j(t_{l-1}, t_l, s) - S_j(s)\}}{S_j(s)} = -\frac{1}{p_j}\frac{1}{\sqrt{n}}\int_0^s \frac{\bar{M}_j^{(n)}(t_{l-1}, t_l, du)}{\gamma_j(t_{l-1}, t_l, u)} + o_P(1).$$

Note that $S_j(s) > 0$ for $s \in [0, L]$, by the choice of L (see assumption 5.1). Let

$$X_l^{(n)}(s) = -\frac{1}{p_j}\frac{1}{\sqrt{n}}\int_0^s \frac{\bar{M}_j^{(n)}(t_{l-1}, t_l, du)}{\gamma_j(t_{l-1}, t_l, u)}.$$

The stochastic integrals $X^{(n)}$ are local square integrable martingales. From the multivariate martingale central limit theorem (theorem A.45) follows the joint asymptotic normality of $(X_1^{(n)}, \dots, X_K^{(n)})$. The Lindeberg condition eq. (A.2.2) of theorem A.45 is satisified, since $\gamma_j(t_{l-1}, t_l, u)$ is bounded away from 0 (see remark A.46). It remains to prove convergence of the quadratic covariation processes

$$\left\langle X_a^{(n)}, X_b^{(n)} \right\rangle(s) = \frac{1}{p_j^2}\frac{1}{n}\int_0^s \frac{\left\langle \bar{M}_j^{(n)}(t_{a-1,a}, \cdot), \bar{M}_j^{(n)}(t_{b-1,b}, \cdot) \right\rangle(du)}{\gamma_j(t_{a-1}, t_a, u)\gamma_j(t_{b-1}, t_b, u)}$$

For $a = b$,

$$\left\langle \bar{M}_j^{(n)}(t_{a-1}, t_a, \cdot) \right\rangle(s) = \int_0^s \bar{Y}^{(n)}(t_{a-1}, t_a, u)\lambda_j(u)du.$$

Thus,

$$\left\langle X_a^{(n)}, X_b^{(n)} \right\rangle(s) = \left\langle X_a^{(n)} \right\rangle(s) = \frac{1}{p_j^2}\frac{1}{n}\int_0^s \frac{\bar{Y}^{(n)}(t_{a-1}, t_a, u)\lambda_j(u)}{\gamma_j(t_{a-1}, t_a, u)^2}du,$$

which converges uniformly in s in probability to

$$\frac{1}{p_j}\int_0^s \frac{\lambda_j(u)}{\gamma_j(t_{a-1}, t_a, u)}du,$$

by lemma 5.3. For $a \neq b$,

$$\begin{aligned}&\left\langle \bar{M}_j^{(n)}(t_{a-1}, t_a, \cdot), \bar{M}_j^{(n)}(t_{b-1}, t_b, \cdot) \right\rangle(s) \\ &= \sum_{i=1}^n \mathbb{1}\{X_i = j\}\int_0^s Y_i(t_{a-1}, t_a, u)Y_i(t_{b-1}, t_b, u)\lambda_j(u)du.\end{aligned} \tag{5.2.4}$$

The right hand side of eq. (5.2.4), because a patient cannot be at risk in the two disjunct calendar time intervals $(t_{a-1}, t_a]$ and $(t_{b-1}, t_b]$ at the same survival time. Formally, for all $u \geqslant 0$,

$$\begin{aligned}
&Y_i(t_{a-1}, t_a, u) Y_i(t_{b-1}, t_b, u) \\
&= \mathbb{1}\{(t_{a-1} - R_i)^+ < u \leqslant Y_i \wedge (t_a - R_i)^+\} \mathbb{1}\{(t_{b-1} - R_i)^+ < u \leqslant Y_i \wedge (t_b - R_i)^+\} \\
&= \mathbb{1}\{(t_{b-1} - R_i)^+ < u \leqslant (t_a - R_i)^+\} \mathbb{1}\{(t_{a-1} - R_i)^+ < u \leqslant Y_i\} \\
&\quad \times \mathbb{1}\{(t_{b-1} - R_i)^+ < u \leqslant Y_i\} \\
&= 0,
\end{aligned}$$

because $(t_{b-1} - R_i)^+ \geqslant (t_a - R_i)^+$.

□

5.2.3 *Application to testing*

Any test statistics, which are (measurable) functions of the weighted Kaplan-Meier estimator (eq. (5.2.3)), such as the test statistics in chapter 3, will also be asymptotically independent and can be used in the combination test approach. An example is given in the next section, where the two-sample log average hazard ratio test statistic (corollary 3.22) is used in an adaptive enrichment design with subgroup selection.

5.3 ADAPTIVE ENRICHMENT DESIGN WITH SUBGROUP SELECTION

In this section a two-stage adaptive seamless phase II/III trial with a binary surrogate and overall survival as primary endpoint is simulated with both the patient-wise and stage-wise splitting approaches. The average hazard ratio based test is compared with the stratified log-rank test. The methodological aspects of adaptive seamless phase II/III designs are described in detail in Bretz et al. [2006], Schmidli et al. [2006] and Brannath et al. [2009]. Such adaptive designs are attractive, since they are more powerful, than a standalone phase III trial and there is no time lag between the end of a phase II trial and the start of a consecutive phase III trial. The role of the first stage (phase II) is to identify a promising subgroup, the role of the second stage (phase III) is to confirm the finding of the first stage.

Suppose the experimental treatment group (E) and the control group (C) are each partitioned into a biomarker-positive (B) and the complementary biomarker-negative

subgroup (B^c). The biomarker-positive subgroup is expected to have a larger benefit from the treatment, than the biomarker-negative subgroup.

In the first stage recruitment starts with the full population. At the interim analysis the first stage data is analyzed and it is decided, whether to continue recruitment in full population or the biomarker-positiv subgroup only or to stop the trial for futility. At the final analysis efficacy is tested in the selected population, i.e. in the full population and the biomarker-positive subgroup or in the biomarker-positive subgroup only. The null hypotheses of equal survival distributions in the control and the experimental group in the full population and the biomarker-positive subgroup are denoted by H_F and H_B, respectively. Denote the stage 1 p-values by p_F and p_B and the stage 2 p-values by q_F and q_B. The p-values are calculated in the following way:

1. Patient-wise splitting:
   a) Stage 1 p-values: Calculated from all data accrued up to final analysis from all patients recruited *before* interim analysis.
   b) Stage 2 p-values: Calculated from all data accrued up to final analysis from all patients recruited *after* interim analysis.
2. Stage-wise splitting:
   a) Stage 1 p-values: Calculated from all data accrued up to interim analysis from all patients. Overrunning patients are right-censored at the interim analysis.
   b) Stage 2 p-values: Calculate from all data accrued up to final analysis from all patients. Survival times are left-truncated at the interim analysis and right-censored at the final analysis.

Testing is done using combination tests for each hypothesis (to adjust for the adaptations) in conjunction with the closed testing principle (Bretz et al. [2006] and Brannath et al. [2009]). In the first stage the p-value p_{BF} of the intersection hypothesis $H_{BF} = H_B \cap H_F$ is calculated according to Simes' procedure:

$$p_{BF} = \min\{2\min(p_B, p_F), \max(p_B, p_F)\}.$$

In the second stage the p-value q_{BF} of the intersection hypothesis H_{BF} is calculated as $q_{BF} = \min\{2\min(q_B, q_F), \max(q_B, q_F)\}$ if recruitment was continued in the full population and both null hypotheses, H_B and H_F, are tested in the final analysis. If

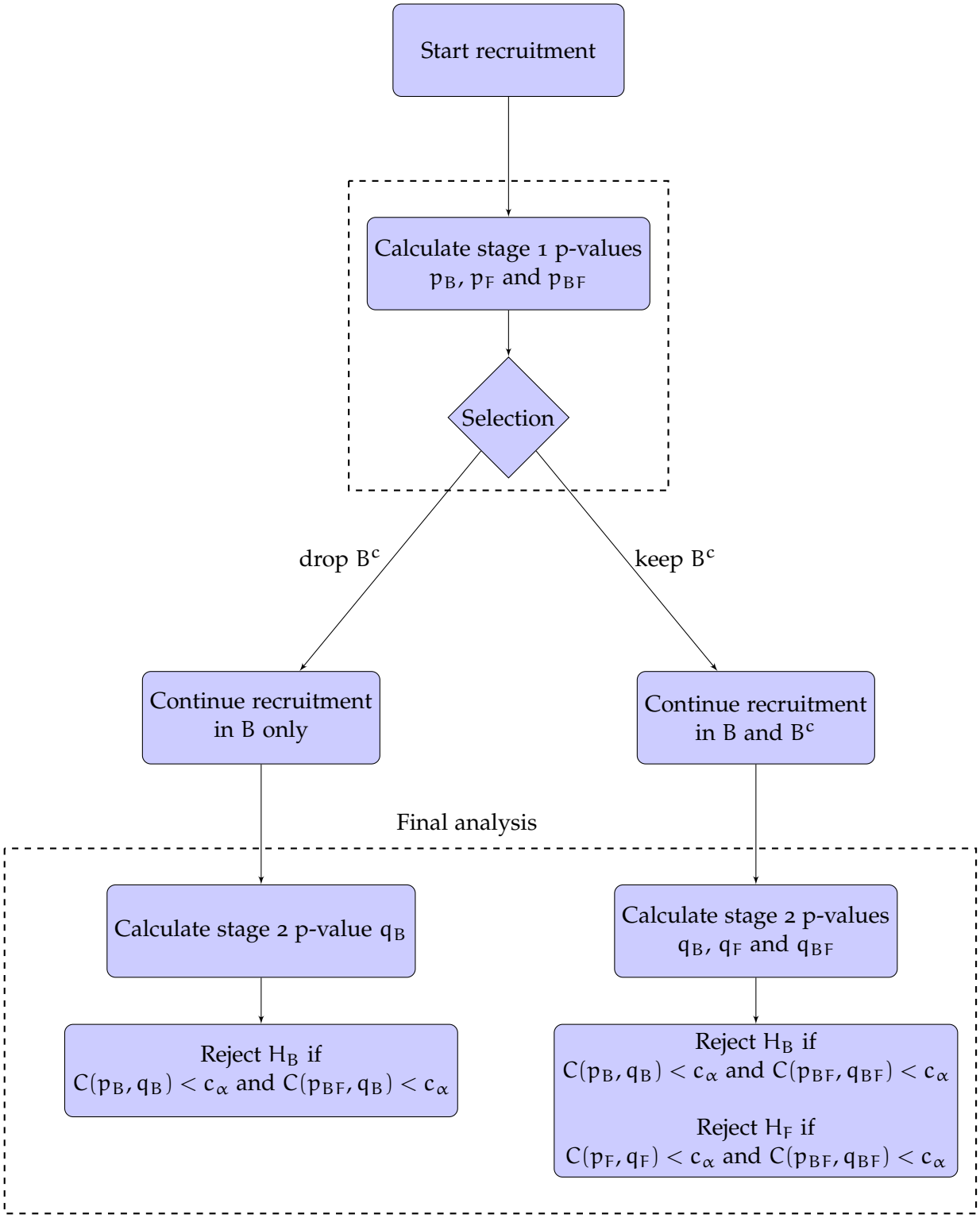

Figure 5.3: Flowchart of an adaptive enrichment design with subgroup selection. C is a combination function and c_α is the corresponding critical value.

only the biomarker-positive subgroup was selected, then $q_{BF} = q_B$. A flowchart of the complete procedure is shown in figure 5.3.

Simulations

The simulation study aims to compare the log average hazard ratio based test (corollary 3.22) to the stratified log-rank test for the patient-wise and the stage-wise splitting approach. The surrogate used is a binary variable indicating response (R) to the treatment or non-response (N), where responders have a higher survival probability, than the non-responders. Within each biomarker group $k \in \{B, B^c\}$ assume the following simple but realistic model for the conditional hazard rates:

$\lambda_{C,N,k}(t) = \lambda_k(t)$	baseline hazard
$\lambda_{C,R,k}(t) = r_k\lambda_k(t)$	R vs. N in control group
$\lambda_{E,N,k}(t) = c_k\lambda_k(t)$	E vs. C for non-responders
$\lambda_{E,R,k}(t) = c_k r_k \lambda_k(t)$	multiplicative model (no-interaction)

The baseline hazard $\lambda_k(t)$ was taken to be equal to 0.0462. Biomarker-positive subgroup prevalence was 50%.

Remark 5.6. *Stratification is done according to response / non-response* and *biomarker group membership, i.e. the stratification variable has 4 levels in the full population and 2 levels in the subgroup.*

The truncation point was $L = 12$. Recruitment was uniform on $[0, 60]$. The final analysis was done after 500 observed events. The interim analysis was conducted after 125 observed events, i.e. after 25% of the maximum number of events have been observed. The weights for the inverse normal combination test were set accordingly to $\sqrt{0.25}$ for the first stage and $\sqrt{0.75}$ for the second stage. The selection was based on the response rate. If the response rate in the experimental biomarker positive subgroup was higher than in the experimental full population, then the biomarker positive subgroup was selected, else the trial continued with the full population. The selection was based on the response rate only, such that the exactly same selection rule can be used for the average hazard ratio based test and the stratified log-rank test.

Four scenarios were considered in total, the null hypothesis and 3 power scenarios. In the power scenarios, the parameters were chosen, such that the average hazard ratio was always approx. 0.7. The first scenario is the null hypothesis. The second

Table 5.1: Overall power of the log average hazard ratio test and stratified log-rank test in an adaptive enrichment trial with subgroup selection with patient-wise splitting. [a]

			Response rates			Power	
r	c_B	c_{B^c}	$\pi_{E,B}$	π_{E,B^c}	π_C	AHR(β)	SLR
0.7	1	1	0.4	0.4	0.4	0.019	0.019
0.7	0.7	0.7	0.4	0.4	0.4	0.770	0.851
0.7	0.7	0.8	0.5	0.3	0.2	0.887	0.798
0.5	1	1	0.8	0.65	0.2	0.870	0.019

[a] Results of 10^5 simulation runs, $\alpha = 0.025$.

scenario ($c_B = c_{B^c} = 0.7$, $\pi_{E,B} = \pi_{E,B^c} = \pi_C = 0.4$) has only an effect conditional on the response, but no difference in the response rates. The third scenario ($c_B = 0.7$, $c_{B^c} = 0.8$, $\pi_{E,B} = 0.5$, $\pi_{E,B^c} = 0.3$, $\pi_C = 0.2$) has a slightly smaller conditional effect than in the second scenario, but now also has a difference in the response rates. The fourth scenario ($c_B = c_{B^c} = 1$, $\pi_{E,B} = 0.8$, $\pi_{E,B^c} = 0.65$, $\pi_C = 0.2$) has no effect conditional on the response, but a large difference in the response rates. It is expected, that the stratified log-rank test will fail to detect any treatment effect in this scenario. The null hypothesis $\beta_E = \beta_C$ was tested against one-sided alternatives $\beta_E \geqslant \beta_C$ with a significance level of $\alpha = 0.025$, where β_E and β_C is the log (average) hazard ratio in the experimental and the control group respectively. In the power simulations, the overall power, i.e. the power to reject any hypothesis (H_B or H_F or both) was considered.

Table 5.1 shows the result of 10^5 simulations of four different scenarios with patient-wise splitting. Both methods control the type-I-error rate and are somewhat conservative. In the second scenario the stratified log-rank test has slightly more power to reject H_B or H_F, than the average hazard ratio based test. This is expected, since the deviation from the proportional hazards is small and the (stratified) log-rank test is optimal for proportional hazards alternatives. In the third scenario, the stratified log-rank test has less power than the average hazard ratio based test, since it only draws power from the conditional effect, and ignores differences in the response rates. In the fourth scenario the stratified log-rank test has no power at all. Table 5.1 shows the result of 10^5 simulations of four different scenarios with stage-wise splitting. The results for stage-wise splitting are very similar to those for patient-wise splitting.

Table 5.2: Overall power of the log average hazard ratio test and stratified log-rank test in an adaptive enrichment trial with subgroup selection with stage-wise splitting. [a]

			Response rates			Power	
r	c_B	c_{B^c}	$\pi_{E,B}$	π_{E,B^c}	π_C	AHR(β)	SLR
0.7	1	1	0.4	0.4	0.4	0.019	0.019
0.7	0.7	0.7	0.4	0.4	0.4	0.771	0.877
0.7	0.7	0.8	0.5	0.3	0.2	0.889	0.834
0.5	1	1	0.8	0.65	0.2	0.868	0.019

[a] Results of 10^5 simulation runs, $\alpha = 0.025$.

6

QUALITY-ADJUSTED SURVIVAL

In this chapter the results for fixed sample size, sequential and adaptive survival trials of the previous chapters are extended to the more general quality-adjusted survival endpoints. Quality-adjusted survival is an integrated measure of clinical benefit for an individual patient, defined as the AUC of repeated health utility score measurements up to the death time of the patient or a pre-defined time limit L (see Gelber et al. [1989], Glasziou et al. [1998], and Gelber et al. [1995]). Overall survival is a special case of quality-adjusted survival, when the utility scores are set equal to 1 for all times. The same methods to the much more general setting, when the observed data are from an increasing stochastic process, which is observed at a censored stopping time (Strawderman [2000]). An example would be the number of recurrent events up to a terminal event (Cook and Lawless [1997]).

Numerous methods have been proposed in the literature for this kind of data. Among them are estimation of the mean (Bang and Tsiatis [2000] and Zhao and Tsiatis [2000]), the survival function (Zhao and Tsiatis [1997] and Zhao and Tsiatis [1999]), the median (Zhao et al. [2012]), and general one-sample U-statistics (Datta et al. [2010]), as well as regression methods including linear regression (Lin [2000]) and Cox proportional hazards regression (Cole et al. [1993]), and two-sample methods for testing equality of distributions (Zhao and Tsiatis [2001] and Huang [1999]).

In order to apply the methods of chapter 3 suitable estimators of the marginal survival functions of QAS are required. The problem is complicated by the fact, that censoring on the scale of the summary measure is informative even in cases, where censoring is stochastically independent of the underlying survival time. Estimators of the survival function of QAS in the presence of induced informative censoring based on inverse probability of censoring weighting (IPCW) methods are discussed in section 6.3. With these estimators, the methods of chapter 3 can be used to obtain test statistics for the comparison of the distribution of QAS in two or more samples. This is dicussed in section 6.4 together with existing two-sample methods. It is shown that the tests proposed by Zhao and Tsiatis [2001] and Huang [1999] are in fact identical. A new two-sample IPCW test statistic is proposed, which adjusts for covariate-dependent censoring, when censoring follows a (stratified) proportional

hazards model. Moreover it is shown, how the results for QAS can be used with the patient-wise approach of 5.

6.1 INVERSE PROBABILITY OF CENSORING WEIGHTING

The analysis of QAS is made difficult by the problem of induced dependent censoring, which is intrinsic to the definition of this endpoint. The censoring of the summary measure is dependent even if the censoring times for the underlying time-to-event data is stochastically independent of the survival times as noted by Lin [2003]. Suppose two patients are censored at the same time and one of the patients is in better health and therefore accrues quality-adjusted survival faster than the other. The observed QAS at the time of the censoring allows to predict the QAS at the future death time. Hence censoring is informative on the QAS scale. This problem prevents the use of the usual survival analysis methods, such as the Kaplan-Meier estimator or proportional hazards regression.

Inverse probability weighting of censoring (IPCW) can be used to derive unbiased estimators and tests. These methods have first been used by Horvitz and Thompson [1952] in the context of sample surveys, and have later been applied to survival problems by Koul et al. [1981] and Robins and Rotnitzky [1992]. Each patients health history is represented by the stochastic process $\{H_i(u), u \geqslant 0\}$. The quality-adjusted survival time is defined as

$$Q_i = \int_0^{T_i} H_i(u)du.$$

Consider the data

$$\left\{Q_i, T_i, C_i, \delta_i = \mathbb{1}\{T_i \leqslant C_i\}, V_i, W_i; \bar{H}_i = \{H_i(u), u \leqslant T_i\}, i = 1, \ldots, n\right\},$$

consisting of iid replicates of the QAS time Q, the survival time T, the health history $\bar{H}_i$ up to time T_i, the censoring time C, the censoring indicator δ, the discrete covariate V and the continuous covariate W, such that T and C are independent given (V, W). The censoring time is assumed to be independent of the health history process given the covariates. Note that $E[\delta_i|T_i, V_i, W_i] = K(T_i|V_i, W_i)$, where $K(t|v, w) = P(C >$

$t|V = v, W = w)$ is the conditional survival function of the censoring time. Thus for any measurable function f,

$$\begin{aligned} E\left[\frac{\delta_i f(Q_i)}{K(T_i \mid V_i, W_i)}\right] &= E\left[E\left[\frac{\delta_i f(Q_i)}{K(T_i|V_i, W_i)}|T_i, V_i, W_i, \bar{H}_i\right]\right] \\ &= E\left[\frac{f(Q_i)}{K(T_i \mid V_i, W_i)} E[\delta_i|T_i, V_i, W_i]\right] \\ &= E[f(Q_i)]. \end{aligned}$$

Usually the censoring probabilities $K(T_i \mid V_i, W_i)$ are unknown and must be estimated. Without covariates this can be done with the Kaplan-Meier estimator. In the presence of covariate-dependent censoring an estimator based on a model, such as the proportional hazards model, must be used. Estimation of censoring probabilities is discussed in section 6.2. IPCW methods can be avoided, if additional modelling assumptions about the accumulation process are made. Gelber et al. [1995] assume a three-state model with known entry and exit times and constant utility in each state (Q-TWiST: Time Without Symptoms or Toxicitiy, Toxicity, Relapse). Pan and Zeng [2011] use a kernel based approach with additional covariates to model the underlying survival time and the accumulation process in the context of medical cost analysis. Such modelling assumptions and additional covariates can also be used to increase the semiparametric efficiency of the IPCW estimators (e.g. Zhao and Tsiatis [1999], van der Laan and Hubbard [1999], Bang and Tsiatis [2000], Strawderman [2000], and Datta et al. [2010]).

6.2 ESTIMATION OF CENSORING PROBABILITIES

Inverse probability of censoring weighting (IPCW) methods require estimates of the censoring probabilities conditional on the covariates. Censoring probabilities are estimated from the same data and by the same methods as survival probabilities (chapter 2) by changing the censoring indicator from δ to $1 - \delta$, although different covariates may be used to model the censoring distribution. Here the conditional hazard rate of the censoring time C given the covariates (V, W)

$$\lambda^c(t|V, W) = \lim_{h \to 0} \frac{1}{h} P\left(t \leqslant C < t + h|T \geqslant t, C \geqslant t, V, W\right)$$

is assumed to be of the form

$$\lambda^c(t|V,W) = \lambda_0^c(t|V)e^{\gamma_0^\top W}, \tag{6.2.1}$$

where λ_0^c is an unknown baseline hazard function (depending on V) and γ_0 is a vector of regression coefficients. The data is stratified according to the levels of V. The censoring counting process is defined as

$$\bar{N}^c(u|v) = \sum_{i=1}^n N_i^c(u|v) = \sum_{i=1}^n \mathbb{1}\{V_i = v\}\mathbb{1}\{Y_i \leqslant t, \delta_i = 0\}.$$

The stratum-specific cumulative baseline hazard

$$\Lambda_0^c(t|v) = \int_0^t \lambda_0^c(u|v)du$$

is estimated by the *Breslow estimator* (Andersen [1993, p. 483])

$$\hat{\Lambda}_0^c(t,v) = \int_0^t \frac{J^{(n)}(u,v)}{S^{(0)}(\hat{\gamma},u,v)} d\bar{N}^c(u,v) = \sum_{j=1}^n \frac{J^{(n)}(u,v)(1-\delta_j)\mathbb{1}\{C_j \leqslant t, V_j = v\}}{S^{(0)}(\hat{\gamma},C_j,v)}, \tag{6.2.2}$$

where $\hat{\gamma}$ is the maximum partial likelihood estimator of model (6.2.1),

$$J^{(n)}(u,v) = \mathbb{1}\left\{\sum_{i=1}^n \mathbb{1}\{V_i = v\}Y_i(u) > 0\right\},$$

and

$$S^{(0)}(\gamma,t,v) = \sum_{i=1}^n \mathbb{1}\{Y_i \geqslant t, V_i = v\}e^{\gamma^\top W_i}.$$

The conditional cumulative hazard rate

$$\Lambda^c(t|v,w) = \int_0^t e^{\gamma_0^\top w}\lambda_0^c(u|v)du = \int_0^t e^{\gamma_0^\top w}d\Lambda_0^c(u|v)$$

is then estimated by

$$\hat{\Lambda}^c(t|v,w) = \int_0^t e^{\hat{\gamma}^\top w}d\hat{\Lambda}_0^c(t,v) \tag{6.2.3}$$

The conditional censoring probability

$$K(t|v,w) = P(C > t|V = v, W = w) = e^{-\Lambda^c(t,v,w)}$$

is estimated by

$$\hat{K}(t|v,w) = e^{-\hat{\Lambda}^c(t,v,w)}. \tag{6.2.4}$$

The following technical assumptions are necessary for the asymptotic results in this chapter:

Assumption 6.1. *1. The conditional censoring hazard rate follows the model in eq. (6.2.1),*

2. $V \in \mathcal{V}$ *a.s., where* $\mathcal{V}$ *is finite,*

3. $W \in \mathcal{W}$ *a.s., where* $\mathcal{W}$ *is a bounded subset of* $\mathbb{R}^d$, $d \geqslant 1$.

4. there exists $L > 0$, *such that* $\inf_{v\in\mathcal{V},w\in\mathcal{W}} K(L|v,w) > 0$.

The last assumption ensures, that the censoring probabilities are bounded away from 0 uniformly in the covariates. This is a technical condition needed in the proofs. In the following $\sup_t$ and $\sup_{t,v,w}$ are defined by $\sup_{t\in[0,L]}$ and $\sup_{t\in[0,L],v\in\mathcal{V},w\in\mathcal{W}}$, respectively. The next lemma proves some asymptotic properties of the estimators $\hat{\Lambda}_0^c$ and $\hat{K}$.

Lemma 6.2. *Under assumption 6.1,*

1.
$$\sup_t |\hat{\Lambda}_0^c(t) - \Lambda_0^c(t)| = o_p(1),$$

2.
$$\sup_t \sqrt{n}|\hat{\Lambda}_0^c(t) - \Lambda_0^c(t)| = O_p(1).$$

3.
$$\sup_{t,v,w} \sqrt{n}|\hat{\Lambda}^c(t|v,w) - \Lambda^c(t|v,w)| = O_p(1),$$

4.
$$\sup_{t,v,w} \sqrt{n}|\hat{K}(t|v,w) - K(t|v,w)| = O_p(1),$$

5.
$$\sup_{t,v,w} \hat{K}^{-1}(t|v,w) = O_p(1).$$

Proof. (1.) and (2.) are a direct consequence of corollary VII.2.4 in Andersen [1993].

3. Adding and substracting the term $e^{\gamma_0 w}\hat{\Lambda}_0(t|v)$,

$$\begin{aligned}&\sqrt{n}\{\hat{\Lambda}(t|v,w) - \Lambda(t|v,w)\}\\ &= \sqrt{n}\{\hat{\Lambda}_0(t|v) - \Lambda_0(t|v)\}e^{\gamma_0 w} + \sqrt{n}\{e^{\hat{\gamma}w} - e^{\gamma_0 w}\}\hat{\Lambda}_0(t|v)\end{aligned} \tag{6.2.5}$$

By a Taylor expansion of the second term around $e^{\gamma_0 w}$, this equals

$$\sqrt{n}\{\hat{\Lambda}_0(t|v) - \Lambda_0(t|v)\}e^{\gamma_0 w} + we^{\gamma^* w}\sqrt{n}\{\hat{\gamma} - \gamma_0\}\hat{\Lambda}_0(t|v), \tag{6.2.6}$$

where γ^* is a point on the line segment between γ_0 and $\hat{\gamma}$. The first term on the right hand side in eq. (6.2.6) is uniformly stochastically bounded, since $\sup_w e^{\gamma_0 w} < \infty$ a.s. and by (2.)

$$\sup_{t|v} \sqrt{n}|\hat{\Lambda}_0(t|v) - \Lambda_0(t|v)| = O_p(1).$$

The second term on the right hand side in eq. (6.2.6) is uniformly stochastically bounded, since $\sqrt{n}\{\hat{\gamma} - \gamma_0\} = O_p(1)$ by theorem 2.4,

$$\sup_w we^{\gamma^* w} = O_p(1),$$

since $\mathcal{W}$ is bounded and $\gamma^* \to \gamma_0 + o_p(1)$, and by (1.)

$$\sup_{t|v} \hat{\Lambda}_0(t|v) = O_p(1).$$

4. Follows from (1.) and a Taylor expansion of $\hat{K}(t|v, w)$ around $\Lambda(t|v, w)$:

$$\sqrt{n}\{\hat{K}(t|v, w) - K(t|v, w)\} = -\Lambda^*(t|v, w)\sqrt{n}\{\hat{\Lambda}(t|v, w) - \Lambda(t|v, w)\},$$

where $\Lambda^*(t|v, w)$ is some point on the line segment between $\hat{\Lambda}(t|v, w)$ and $\Lambda(t|v, w)$.

5. Let $\epsilon > 0$ and $C = \inf_{t,v,w} K(t|v, w) - \epsilon$.

$$\begin{aligned}
P\left(\sup_{t,v,w} \frac{1}{\hat{K}(t|v, w)} < \frac{1}{C}\right) &= P\left(\frac{1}{\inf_{t,v,w} \hat{K}(t|v, w)} < \frac{1}{C}\right) \\
&= P\left(\inf_{t,v,w} \hat{K}(t|v, w) > C\right) \\
&\geqslant P\left(\forall t, v, w : \hat{K}(t|v, w) > K(t|v, w) - \epsilon\right) \\
&\geqslant P\left(\forall t, v, w : |\hat{K}(t|v, w) - K(t|v, w)| < \epsilon\right) \\
&= P\left(\sup_{t,v,w} |\hat{K}(t|v, w) - K(t|v, w)| < \epsilon\right)
\end{aligned}$$

The right hand side converges to 1, because of (4.).

□

6.3 ESTIMATION OF THE SURVIVAL FUNCTION

The survival function $S_Q(x) = P(Q > x)$ of quality-adjusted survival or medical cost can be estimated with the IPCW approach, when either censoring of the underlying survival time is independent or conditionally independent, i.e. independent conditional on some covariates.

For technical reasons it is necessary, that the conditional censoring probabilities are uniformly bounded away from 0. The survival times are truncated by the constant τ, which is chosen, such that assumption 6.1 holds. The analysis is then based on iid replications of the quality-adjusted survival time Q, the censored survival time $Y = T \wedge C$, the truncated censored survival time $Y^* = T^* \wedge C$, where $T^* = T \wedge \tau$, the censoring indicator $\delta = \mathbb{1}\{T \leqslant C\}$, the censoring indicator of the truncated survival time $\delta^* = \mathbb{1}\{T^* \leqslant C\}$, a discrete covariate V and a continous covariate vector W. Note that the quality-adjusted survival time is calculated only up to the truncated survival time.

6.3.1 *IPCW without covariates*

In Zhao and Tsiatis [1997] an IPCW estimator of the survival function S_Q is defined as an IPCW version of the empirical distribution function:

$$\hat{S}_Q(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i^* \mathbb{1}\{Q_i > x\}}{\hat{K}(T_i^*)}. \tag{6.3.1}$$

Remark 6.3. *The independent censoring case is contained in the proportional hazards model, by setting* $V_i = W_i = 0$ *for all* $i = 1, \ldots, n$.

Huang and Louis [1998] derives an estimator of S_Q as the non-parametric maximum likelihood estimator. It can be shown that this estimator is identical to the Zhao-Tsiatis estimator from eq. (6.3.1) (appendix 3 in Bang and Tsiatis [2000]). $\hat{S}_Q(x)$ can also be seen as a special case of the one-sample IPCW U-statistics of Datta et al. [2010]. Pointwise consistency of $\hat{S}_Q(x)$ and asymptotic normality of $\sqrt{n}\{\hat{S}_Q(x) - S_Q(x)\}$ for all $x \geqslant 0$ is proved by Zhao and Tsiatis [1997] and Datta et al. [2010]. Zhao et al.

[2012] use these results in the context of medical cost to construct point estimators and confidence intervals for the median and the difference and ratio of two medians using the method described in section 3.3.2. Uniform consistency of $\hat{S}_Q$ and weak convergence in $D[0, L]$ of $\sqrt{n}\{\hat{S}_Q - S_Q\}$ is proved by Huang and Louis [1998].

6.3.2 *IPCW with covariates*

Lin [2000] considers the linear regression model

$$Y = \beta^T Z + \epsilon, \tag{6.3.2}$$

where Y is a metric response variable and Z is a $p \times 1$-vector of covariates and β is a $p \times 1$-vector of unknown regression parameters, and ϵ is mean-zero error term with an unspecified distribution. The first component of Z is set to 1, such that the first component of β is the intercept. The estimator of β is defined as the solution to the estimating equation

$$\sum_{i=1}^{n} \frac{\delta_i^*}{\hat{K}(T_i^*|V_i, W_i)}(Y_i - \beta^T Z_i)Z_i = 0, \tag{6.3.3}$$

where $\hat{K}$ is the estimator of the conditional survival function of the censoring time based on the stratified proportional hazards model

$$\lambda(t|V, W) = \lambda_0(t|V)e^{\gamma^T W}, \tag{6.3.4}$$

i.e.

$$\hat{K}(t|V, W) = e^{-e^{\hat{\gamma} W}\hat{\Lambda}_0(t|V)},$$

where

$$\hat{\Lambda}_0(t|V) = \sum_{C_i \leqslant t} \frac{(1 - \delta_i)N_i^c(C_i)}{Y_i(C_i)}$$

is the Breslow estimator of the censoring baseline hazard. The estimating equation (6.3.3) has the closed-form solution

$$\hat{\beta} = \left\{ \sum_{i=1}^{n} \frac{\delta_i^*}{\hat{K}(T_i^*|V_i, W_i)} Z_i^{\otimes 2} \right\}^{-1} \sum_{i=1}^{n} \frac{\delta_i^*}{\hat{K}(T_i^*|V_i, W_i)} Y_i Z_i \quad . \tag{6.3.5}$$

Consistency and asymptotic normality of the estimator in eq. (6.3.5) is proved by Lin [2000]. If in the linear model from eq. (6.3.2) the outcome variable is $Y = \mathbb{1}\{Q > x\}$ for fixed x, and the covariate Z is equal 1, then

$$\beta \equiv \beta(x) = E[Y] = P(Q > x)$$

and the estimator $\hat{\beta}$ reduces to

$$\hat{\beta}(x) = \left(\sum_{i=1}^{n} \frac{\delta_i^*}{\hat{K}(T_i^* | V_i, W_i)} \right)^{-1} \sum_{i=1}^{n} \frac{\delta_i^* \mathbb{1}\{Q_i > x\}}{\hat{K}(T_i^* | V_i, W_i)}, \tag{6.3.6}$$

Note that, as $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i^*}{\hat{K}(T_i^* | V_i, W_i)} \xrightarrow{p} E\left[\frac{\delta_i^*}{K(T_i^* \mid V_i, W_i)} \right] = 1,$$

such that asymptotically $\hat{\beta}(x)$ coincides with $\hat{S}_Q(x)$ from eq. (6.3.1). The results of Lin [2000] imply the pointwise consistency of $\hat{\beta}(x)$ and asymptotic normality of $\sqrt{n}\{\hat{\beta}(x) - \beta(x)\}$. Instead of the stratified proportional hazards model eq. (6.3.4), the estimate $\hat{K}$ can also be based on an additive hazard model

$$\lambda(t|W) = \beta_0 + \sum_{i=1}^{k} \beta_i \phi_i(W_i), \tag{6.3.7}$$

where ϕ_i are known functions and W_i denotes the i-the component of the covariate vector W. The resulting estimate is a special case of the one-sample IPCW U-statistics with an additive hazard model for the censoring times introduced by Datta et al. [2010]. The estimators of Lin [2000] and Datta et al. [2010] can be seen as a generalization of the estimator of Zhao and Tsiatis [1997] from eq. (6.3.1). Their results are sufficient for the methods based on the median (section 3.3.2), but not for average hazard ratio methods, which are based on the functional delta method, which requires uniform consistency and weak convergence in $D[0, L]$.

6.4 TWO-SAMPLE TESTS

In the two-sample setting the data is

$$\{Q_{ki}, T_{ki}, C_{ki}, \delta_{ki} = \mathbb{1}\{T_{ki} \leqslant C_{ki}\}, V_{ki}, W_{ki}, \bar{H}_{ki} = \{H_{ki}(u), u \leqslant T_{ki}\}, i = 1, \ldots, n_k, k = 1, 2\}.$$

The conditional censoring survival functions in each sample are denoted by K_k and the corresponding Kaplan-Meier estimators by $\hat{K}_k$, $k = 1, 2$. Equivalently the data can be represented as one-sample with a binary group indicator Z

$$\left\{Q_i, T_i, C_i, \delta_i = \mathbb{1}\{T_i \leqslant C_i\}, V_i, W_i, \bar{H}_i = \{H_i(u), u \leqslant T_i\}, Z_i, i = 1, \ldots, n\right\}.$$

6.4.1 *IPCW Log-rank test*

Zhao and Tsiatis [2001] have generalized the log-rank test for comparing the distributions of two survival times. They consider the IPCW test statistic

$$LR(\hat{w}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i^*}{\hat{K}(T_i|Z_i)} \hat{w}(Q_i) \left\{ Z_i - \frac{\sum_{j=1}^{n} \frac{\delta_j^*}{\hat{K}(T_j|Z_j)} Z_j \mathbb{1}\{Q_j \geqslant Q_i\}}{\sum_{j=1}^{n} \frac{\delta_j}{\hat{K}(T_j|Z_j)} \mathbb{1}\{Q_j \geqslant Q_i\}} \right\} \tag{6.4.1}$$

where Z is a binary group indicator and $\hat{K}(\cdot|Z)$ is the group-specific Kaplan-Meier estimator of the censoring times. Let

$$\hat{w}(u) = \frac{\hat{K}(u, 1)\hat{K}(u, 0)}{\hat{K}(u)},$$

where $\hat{K}(\cdot)$ is the Kaplan-Meier estimator of the pooled sample. With this choice of the weight function the test statistic $LR(\hat{w})$ reduces to the standard log-rank test statistic in the special case $Q_i = T_i$ (appendix in Zhao and Tsiatis [2001]).

6.4.2 *Weighted Hazard and Weighted Survival Statistics*

Huang [1999] proposes two classes of test statistics for comparing two independent samples. Let n_i be the sample size and $\hat{S}_i$ be the estimator from eq. (6.3.1), $i = 1, 2$. The *weighted hazard statistics* (WHS) are defined as

$$WHS(\hat{W}) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \int_0^L \hat{W}(u) \left\{ \frac{\hat{S}_2(du)}{\hat{S}_2(u-)} - \frac{\hat{S}_1(du)}{\hat{S}_1(u-)} \right\} \tag{6.4.2}$$

where $\hat{W}$ is a uniformly consistent weight function, which converges uniformly a.s. to a bounded function on $[0, L]$, and

$$\hat{S}_k(u) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_{ki}^* \mathbb{1}\{Q_{ki} > u\}}{\hat{K}_k(T_{ki}^*)}.$$

Huang [1999] proves asymptotic normality of the WHS-type statistics by the functional delta method. Intuitively the IPCW log-rank test and the weighted hazard statistics are very similar. Indeed every weighted hazard statistic eq. (6.4.2) can be written as IPCW log-rank test eq. (6.4.1) and vice versa:

Theorem 6.4. *Let $\hat{w}$ be a weight function which converges uniformly in probability to a bounded function w on $[0, L]$ and*

$$\hat{W}(u) = \sqrt{n_1 n_2} \frac{\hat{S}_1(u-)\hat{S}_2(u-)}{n_1 \hat{S}_1(u-) + n_2 \hat{S}_2(u-)} \hat{w}(u)$$

If $n_1/(n_1 + n_2) \to c$, then $\hat{W}$ converges uniformly to a bounded function on $[0, L]$ and

$$WHS(\hat{W}) = LR(\hat{w})$$

Proof. Since $n_1/(n_1 + n_2) \to c$ and $\hat{S}_i$ is strongly uniformly consistent (see Huang and Louis [1998]) for $i = 1, 2$, the weight function $\hat{W}$ converges uniformly a.s. to

$$\sqrt{c(1-c)} \frac{S_1(u-)S_2(u-)}{cS_1(u-) + (1-c)S_2(u-)} w(u),$$

which is a bounded function on $[0, L]$. Thus $\hat{W}$ is a valid weight function for the weighted hazard statistics. $\hat{S}_1$ and $\hat{S}_2$ can be written as

$$\hat{S}_1(u) = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{\delta_{1i}^* \mathbb{1}\{Q_{1i} > u\}}{\hat{K}_1(T_{1i}^*)} = \frac{1}{n_1} \sum_{i=1}^{n} \frac{\delta_i^* \mathbb{1}\{Q_i > u\}(1 - Z_i)}{\hat{K}(T_i^* | Z_i)}$$

$$\hat{S}_2(u) = \frac{1}{n_2} \sum_{i=1}^{n_2} \frac{\delta_{2i}^* \mathbb{1}\{Q_{2i} > u\}}{\hat{K}_2(T_{2i}^*)} = \frac{1}{n_2} \sum_{i=1}^{n} \frac{\delta_i^* \mathbb{1}\{Q_i > u\} Z_i}{\hat{K}(T_i^* | Z_i)},$$

where $Z_i = 0$ in sample 1 and $Z_i = 1$ in sample 2. Inserting this into equation (6.4.2) yields

$$\begin{aligned}
WHS(\hat{W}) &= \sqrt{\frac{n_1 n_2}{n}} \sum_{i=1}^{n} \frac{\delta_i^* \hat{W}(Q_i)}{\hat{K}(T_i^* | Z_i)} \left\{ \frac{Z_i}{n_2 \hat{S}_2(Q_i -)} - \frac{1 - Z_i}{n_1 \hat{S}_1(Q_i -)} \right\} \\
&= \sqrt{\frac{n_1 n_2}{n}} \sum_{i=1}^{n} \frac{\delta_i^* \hat{W}(Q_i)}{\hat{K}(T_i^* | Z_i)} \frac{[Z_i \{n_1 \hat{S}_1(Q_i -) + n_2 \hat{S}_2(Q_i -)\} - n_2 \hat{S}_2(Q_i -)]}{n_1 n_2 \hat{S}_1(Q_i -) \hat{S}_2(Q_i -)} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i^* \hat{w}(Q_i)}{\hat{K}(T_i^* \mid Z_i)} \left\{ Z_i - \frac{n_2 \hat{S}_2(Q_i -)}{n_1 \hat{S}_1(Q_i -) + n_2 \hat{S}_2(Q_i -)} \right\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i^* \hat{w}(Q_i)}{\hat{K}(T_i^* \mid Z_i)} \left\{ Z_i - \frac{\sum_{j=1}^{n} \frac{\delta_j^*}{\hat{K}(T_j^* | Z_j)} Z_j \mathbb{1}\{Q_j \geqslant Q_i\}}{\sum_{j=1}^{n} \frac{\delta_j^*}{\hat{K}(T_j^* | Z_j)} \mathbb{1}\{Q_j \geqslant Q_i\}} \right\} \\
&= LR(\hat{w}) \quad .
\end{aligned}$$

□

Another class of two-sample statistics defined by Huang [1999] are the *weighted survival statistics* (WHS), which are similar to the Pepe-Fleming statistics for ordinary survival data (section 3.3.1), but based on IPCW survival function estimators. Weighted survival statistics (WSS) are defined as

$$WSS(\hat{W}) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \int_0^L \hat{W}(u) \left\{ \hat{S}_1(u) - \hat{S}_2(u) \right\} du. \tag{6.4.3}$$

Asymptotic normality of the weighted survival statistics is also proved by the functional delta method (Huang [1999]).

6.4.3 *Relative risk*

Independent censoring

The large sample properties of $\hat{S}_Q$ eq. (6.3.1) proved by Huang and Louis [1998] are sufficient for the methods from chapter 3. However, the true survival function S_Q is in general not continuous, since it may have a jump discontinuity at the point τ. For example in the special case $Q = T^* = T \wedge \tau$,

$$\Delta S(\tau) = S(\tau) - S(\tau-) = P(T \wedge \tau > \tau) - P(T \wedge \tau \geqslant \tau) = -P(T \geqslant \tau).$$

The general definition of the average hazard ratio (definition 3.4) requires the existence of hazard rates, which imply continuous survival functions. This no problem as long as $L < \tau$, where L is the truncation point used in the definition of the average hazard ratio (see definitions 3.4 and 3.5 and assumption 3.8), since only continuity of the survival functions on the interval $[0, L]$ is required. Thus, all methods from chapter 3 can be applied to a quality-adjusted survival endpoint, including the results for the multivariate average hazard ratio.

Covariate-dependent censoring

If censoring is covariate-dependent, then it is very difficult to prove large sample properties required for the results in chapter 3 (see assumption 3.8). For any uncensored observations Y_{ki}, $i = 1, \ldots, n_k$, $k = 1, 2$, the Mann-Whitney U-statistic, given by

$$U = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} h(Y_{1i}, Y_{2j}),$$

where

$$h(x, y) = \begin{cases} 1 & x > y \\ 0.5 & x = y \\ 0 & x < y \end{cases},$$

is an unbiased estimator of the relative risk

$$E[h(Y_{1i}, Y_{2j})] = P(Y_{1i} > Y_{2j}).$$

Fan and Datta [2013] have already considered IPCW Mann-Whitney U-statistic under the assumption of independent censoring, hereby extending the one-sample

IPCW U-statistics results of Datta et al. [2010]. For the case of covariate-dependent assume, that the censoring times follow a proportional hazards model in each sample, i.e.

$$\lambda_k^c(t|v,w) = \lambda_{0k}^c(t|v)e^{\gamma_{0k}w} \quad k = 1,2 \tag{6.4.4}$$

and define the IPCW Mann-Whitney U-statistic as

$$\hat{U} = \frac{1}{n_1}\frac{1}{n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{\delta_{1i}\delta_{2j}h(Q_{1i},Q_{2j})}{\hat{K}_1(T_{1i}|V_{1i},W_{1i})\hat{K}_2(T_{2j}|V_{2j},W_{2j})},$$

where $\delta_{ki} = \mathbb{1}\{T_{ki} \leqslant C_{ki}\}$ is the censoring indicator and $\hat{K}_k$ are the estimators of the conditional survival function of the censoring times in each sample k (see section 6.2). If the model in eq. (6.4.4) is correct and the censoring times are stochastically independent from the survival times conditional on the covariates, then $\hat{U}$ is consistent and asymptotically normal as proved in theorem 6.7. The statement and proof of theorem 6.7 requires some additional notation and two technical lemmas. Let $F_k(x) = P(Q_{ki} \leqslant x)$ and $S_k(x) = 1 - F_k(x)$, $k = 1,2$, and

$$f_k(x) = \begin{cases} F_2(x-) & k = 1 \\ S_1(x) & k = 2 \end{cases}$$

and

$$\begin{aligned} q_k(t,v) &= E\left[\frac{f_k(Q_{k1})e^{\beta_k^T W_{k1}}\mathbb{1}\{T_{k1} \geqslant t\}}{s_k^{(0)}(\beta_{k0},t,v)}\right] \\ \bar{w}_k(t,v) &= \frac{s_k^{(1)}(\gamma_{0k},t,v)}{s_k^{(0)}(\gamma_{0k},t,v)} \\ H_k(t,v,w) &= \int_0^t e^{\gamma_{0k}^T w}\{w - \bar{w}_k(u,v)\}\lambda_k^c(u \mid v,w)du \\ R_k &= E\left[f_k(Q_{k1})H_k(T_{k1},V_{k1},W_{k1})^T\right] \end{aligned}$$

Moreover, denote by Ω_k the asymptotic covariance matrix of the regression parameter vector γ_k in the proportional hazards model eq. (6.4.4), and by M_{ki}^c the counting process martingale of the censoring counting process

$$M_{ki}^c(t) = \mathbb{1}\{Y_{ki} \leqslant t, \delta_{ki} = 0\} - \int_0^t \lambda_{0k}^c(s|V)e^{-\gamma_{0k}^T W_{ki}(s)}Y_{ki}(s)ds.$$

Lemma 6.5. *For $n \geqslant 1$ consider a triangular array R_{ni}, $i = 1, \ldots, n$. If $R_{n1}, \ldots, R_{nn}$ are identically distributed for each n and $R_{n1} \xrightarrow{P} 0$ as $n \to \infty$, and*

$$\max_{1 \leqslant i \leqslant n} |R_{ni}| = O_p(1), \tag{6.4.5}$$

then

$$\frac{1}{n} \sum_{i=1}^{n} |R_{ni}| \xrightarrow{P} 0$$

Proof. Write $R_{ni} = R_{ni} \mathbb{1}\{|R_{ni}| \leqslant C\} + R_{ni} \mathbb{1}\{|R_{ni}| > C\}$ for any $C \geqslant 0$. Let $\epsilon > 0$. Then

$$\begin{aligned} P\left(\left|\frac{1}{n} \sum_{i=1}^{n} R_{ni} > \epsilon\right|\right) \leqslant & P\left(\frac{1}{n} \sum_{i=1}^{n} |R_{ni}| \mathbb{1}\{|R_{ni}| > C\} > \frac{\epsilon}{2}\right) \\ & + P\left(\frac{1}{n} \sum_{i=1}^{n} |R_{ni}| \mathbb{1}\{|R_{ni}| \leqslant C\} > \frac{\epsilon}{2}\right) \end{aligned}$$

If the term

$$\frac{1}{n} \sum_{i=1}^{n} |R_{ni}| \mathbb{1}\{|R_{ni}| > C\}$$

is larger than $\epsilon/2$, then $|R_{ni}| > C$ for at least one i, i.e. $\max_i |R_{ni}| > C$. Thus

$$P\left(\frac{1}{n} \sum_{i=1}^{n} |R_{ni}| \mathbb{1}\{|R_{ni}| > C\} > \frac{\epsilon}{2}\right) \leqslant P(\max_i |R_{ni}| > C)$$

The probability $P(\max_i |R_{ni}| > C)$ can be made arbitrarily small for large n and C, because of assumption (6.4.5). By the Markov inequality, the second term on the right hand side is bounded by

$$2\epsilon^{-1} E\left[|R_{n1}| \mathbb{1}\{|R_{n1}| \leqslant C\}\right]$$

which converges to 0 as $n \to \infty$, because of $|R_{n1}| \mathbb{1}\{|R_{n1}| \leqslant C\} \xrightarrow{p} 0$ and the dominated convergence theorem. □

The next lemma is a special case of equation (A2) of Lin [2000].

Lemma 6.6.

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n} \frac{f_k(Q_{ki}) \delta_{ki}}{K(T_{ki}|V_{ki}, W_{ki})} \frac{\sqrt{n} \left\{\hat{K}_k(T_{ki}|V_{ki}, W_{ki}) - K_k(T_{ki}|V_{ki}, W_{ki})\right\}}{\hat{K}_k(T_{ki}|V_{ki}, W_{ki})} \\ & = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^L \left[q_k(t, V_{ki}) + R_k \Omega_k^{-1} \{W_{ki} - \bar{w}_k(V_{ki})\}\right] M_{ki}^c(dt) + o_p(1), \end{aligned}$$

The next theorem gives the asymptotic distribution of the IPCW Mann-Whitney U-statistic. The key step is to separate the U-statistics into two stochastically independent terms (one for each sample).

Theorem 6.7. *Let* $\theta = P(Q_{1i} > Q_{2i})$ *and* $c_k = \lim_{n\to\infty} \frac{n_k}{n}$, $k = 1, 2$. *Then, as* $n \to \infty$,

$$\sqrt{n}(\hat{U} - \theta) \xrightarrow{\mathcal{L}} N\left(0, \frac{\sigma_1^2}{c_1} + \frac{\sigma_2^2}{c_2}\right),$$

where

$$\sigma_k^2 = Var\left(\frac{f_k(Q_{ki})\delta_{ki}}{K_k(T_{ki}|V_{ki}, W_{ki})} - \int_0^L \left[q_k(t, V_{ki}) + R_k\Omega_k^{-1}\{W_{ki} - \bar{w}_k(V_{ki})\}\right] M_{ki}^c(dt)\right).$$

A consistent estimator of σ_k^2 *is given in lemma 6.8.*

Proof. Let $K_{ki} = K(T_{ki}|V_{ki}, W_{ki})$ and $\hat{K}_{ki} = \hat{K}_{ki}(T_{ki}|V_{ki}, W_{ki})$ for $k = 1, 2$ and $i = 1, \ldots, n_k$. Write

$$\begin{aligned}\sqrt{n}\{\hat{U} - \theta\} = \sqrt{n}\{U - \theta\} \\ - \frac{\sqrt{n}}{\sqrt{n_1}} \frac{1}{n_1 n_2} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \frac{h(Q_{1i}, Q_{2j})\delta_{1i}\delta_{2j}}{K_{1i}\hat{K}_{2j}} \frac{\sqrt{n_1}\{\hat{K}_{1i} - K_{1i}\}}{\hat{K}_{1i}} \\ - \frac{\sqrt{n}}{\sqrt{n_2}} \frac{1}{n_1 n_2} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \frac{h(Q_{1i}, Q_{2j})\delta_{1i}\delta_{2j}}{\hat{K}_{1i}K_{2j}} \frac{\sqrt{n_2}\{\hat{K}_{2j} - K_{2j}\}}{\hat{K}_{2j}},\end{aligned} \tag{6.4.6}$$

where

$$U = \frac{1}{n_1}\frac{1}{n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \frac{h(Q_{1i}, Q_{2j})\delta_{1i}\delta_{2j}}{K_{1i}K_{2j}}.$$

The orthogonal projection of $U - \theta$ onto the set of all statistics of the form

$$\sum_{i=1}^{n_1} f_i(Q_{1i}, T_{1i}, \delta_{1i}, V_{1i}, W_{1i}) + \sum_{j=1}^{n_2} g_j(Q_{2j}, T_{2j}, \delta_{2j}, V_{2j}, W_{2j})$$

is given by the conditional expectations

$$\begin{aligned}\tilde{U} &= \sum_{i=1}^{n_1} E\left[U-\theta|Q_{1i},T_{1i},\delta_{1i},V_{1i},W_{1i}\right]+\sum_{j=1}^{n_2} E\left[U-\theta|Q_{2j},T_{2j},\delta_{2j},V_{2j},W_{2j}\right]\\ &= \frac{1}{n_1}\sum_{i=1}^{n_1}\left\{\frac{F_2(Q_{1i}-)\delta_{1i}}{K_{1i}}-\theta\right\}+\frac{1}{n_2}\sum_{j=1}^{n_2}\left\{\frac{S_1(Q_{2j})\delta_{2j}}{K_{2j}}-\theta\right\}\end{aligned}$$

See van der Vaart [2000]. Therefore

$$\sqrt{n}\{U-\theta\}=\frac{\sqrt{n}}{n_1}\sum_{i=1}^{n_1}\left\{\frac{F_2(Q_{1i}-)\delta_{1i}}{K_{1i}}-\theta\right\}+\frac{\sqrt{n}}{n_2}\sum_{j=1}^{n_2}\left\{\frac{S_1(Q_{2j})\delta_{2j}}{K_{2j}}-\theta\right\}+o_p(1)$$

now consider the second term of eq. (6.4.6). The $\hat{K}_{2j}$ in the denominator can be replaced by K_{2j}, because of lemma 6.2 and lemma 6.5. The term can be written as

$$-\frac{\sqrt{n}}{\sqrt{n_1}}\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\hat{\mu}_{2i}\frac{\sqrt{n_1}\{\hat{K}_{1i}-K_{1i}\}}{\hat{K}_{1i}}, \tag{6.4.7}$$

where

$$\hat{\mu}_{2i}=\frac{1}{n_2}\sum_{j=1}^{n_2}\frac{h(Q_{1i},Q_{2j})\delta_{1i}\delta_{2j}}{K_{1i}K_{2j}}.$$

Note that,

$$\frac{\sqrt{n}}{\sqrt{n_1}}\to\frac{1}{\sqrt{c_1}}\quad \text{a.s.},$$

and

$$\max_{i=1,\ldots,n_1}\frac{\sqrt{n_1}|\hat{K}_{1i}-K_{1i}|}{\hat{K}_{1i}}=O_p(1),$$

as $n\to\infty$, by lemma 6.2. For fixed i, $\hat{\mu}_{2i}$ converges in probability to μ_{2i} conditional on $\{Q_{1i},T_{1i},\delta_{1i},V_{1i},W_{1i}\}$, where

$$\mu_{2i}=E\left[\hat{\mu}_{2i}\mid Q_{1i},T_{1i},\delta_{1i},V_{1i},W_{1i}\right]=\frac{F_2(Q_{1i}-)\delta_{1i}}{K_{1i}},$$

by the law of large numbers. The conditional expectation

$$E\left[|\hat{\mu}_{2i}-\mu_{2i}|\mid Q_{1i},T_{1i},\delta_{1i},V_{1i},W_{1i}\right]$$

is bounded by some constant not depending on n, since $|\hat{\mu}_{2i} - \mu_{2i}|$ is bounded. The dominated convergence theorem implies

$$E\left[|\hat{\mu}_{2i} - \mu_{2i}| \mid Q_{1i}, T_{1i}, \delta_{1i}, V_{1i}, W_{1i}\right] \to 0,$$

and

$$E\left[|\hat{\mu}_{2i} - \mu_{2i}|\right] = E\left[E\left[|\hat{\mu}_{2i} - \mu_{2i}| Q_{1i}, T_{1i}, \delta_{1i}, V_{1i}, W_{1i}\right]\right] \to 0,$$

as $n \to \infty$. Therefore

$$\frac{\sqrt{n}}{\sqrt{n_1}} \frac{1}{n_1} \sum_{i=1}^{n_1} |\hat{\mu}_{2i} - \mu_{2i}| \max_{i=1,\ldots,n_1} \frac{\sqrt{n_1}|\hat{K}_{1i} - K_{1i}|}{\hat{K}_{1i}} = o_p(1),$$

and

$$-\frac{\sqrt{n}}{\sqrt{n_1}} \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \hat{\mu}_{2i} \frac{\sqrt{n_1}\{\hat{K}_{1i} - K_{1i}\}}{\hat{K}_{1i}} = -\frac{1}{\sqrt{c_1}} \frac{1}{n_1} \sum_{i=1}^{n_1} \mu_{2i} \frac{\sqrt{n_1}\{\hat{K}_{1i} - K_{1i}\}}{\hat{K}_{1i}} + o_p(1).$$

The same argument works for the third term of eq. (6.4.6), too. Hence

$$\begin{aligned}
\sqrt{n}\{\hat{U} - \theta\} = {} & \frac{1}{\sqrt{c_1}} \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} \left\{ \frac{F_2(Q_{1i}-)\delta_{1i}}{K_{1i}} - \theta \right\} + \frac{1}{\sqrt{c_2}} \frac{1}{\sqrt{n_2}} \sum_{j=1}^{n_2} \left\{ \frac{S_1(Q_{2j})\delta_{2j}}{K_{2j}} - \theta \right\} \\
& - \frac{1}{\sqrt{c_1}} \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{F_2(Q_{1i}-)\delta_{1i}}{K_{1i}} \frac{\sqrt{n_1}\{\hat{K}_{1i} - K_{1i}\}}{\hat{K}_{1i}} \\
& - \frac{1}{\sqrt{c_2}} \frac{1}{n_2} \sum_{j=1}^{n_2} \frac{S_1(Q_{2j})\delta_{2j}}{K_{2j}} \frac{\sqrt{n_2}\{\hat{K}_{2j} - K_{2j}\}}{\hat{K}_{2j}} + o_p(1)
\end{aligned}$$

This can be written as a sum of iid random variables, by lemma 6.6. Thus

$$\begin{aligned}\sqrt{n}\{\hat{U}-\theta\} = &\frac{1}{\sqrt{c_1}}\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\left\{\frac{F_2(Q_{1i}-)\delta_{1i}}{K_{1i}}-\theta\right\}\\ &+\frac{1}{\sqrt{c_2}}\frac{1}{\sqrt{n_2}}\sum_{j=1}^{n_2}\left\{\frac{S_1(Q_{2j})\delta_{2j}}{K_{2j}}-\theta\right\}\\ &-\frac{1}{\sqrt{c_1}}\frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\left[q_1(t,V_{1i})+R_1\Omega_1^{-1}\{W_{1i}-\bar{w}_1(V_{1i})\}\right]M_{1i}(dt)\\ &-\frac{1}{\sqrt{c_2}}\frac{1}{\sqrt{n_2}}\sum_{j=1}^{n_2}\left[q_2(t,V_{2j})+R_2\Omega_2^{-1}\{W_{2j}-\bar{w}_2(V_{2j})\}\right]M_{2j}(dt)\\ &+o_p(1)\end{aligned} \tag{6.4.8}$$

Asymptotic normality now follows from the central limit theorem. □

Variance estimation

In this section a consistent estimator of the asymptotic variance of the IPCW Mann-Whitney U-test of theorem 6.7 is given. Let

$$h_k(x) = \begin{cases}\mathbb{1}\{Q_{21} < x\} & k = 1\\ \mathbb{1}\{Q_{11} > x\} & k = 2\end{cases}$$

Note that $E[h_k(x)] = f_k(x)$. As in section 2.4, define the estimator of the asymptotic variance of the maxium partial likelihood estimator $\hat{\gamma}_k$, $k = 1,2$ by

$$\hat{\Omega}_k(\gamma) = \sum_{i=1}^{n_k}\delta_{ki}\left\{\frac{S_k^{(2)}(\hat{\gamma}_k,T_{ki},V_{ki})}{S_k^{(0)}(\hat{\gamma}_k,T_{ki},V_{ki})}-\frac{S_k^{(1)}(\hat{\gamma}_k,T_{ki},V_{ki})^{\otimes 2}}{S_k^{(0)}(\hat{\gamma}_k,T_{ki},V_{ki})^2}\right\},$$

where

$$\begin{aligned}S_k^{(0)}(\gamma,t,v) &= \sum_{i=1}^{n_k}\mathbb{1}\{Y_{ki}\geqslant t, V_{ki}=v\}e^{\gamma^T W_{ki}}\\ S_k^{(1)}(\gamma,t,v) &= \frac{\partial S_k^{(0)}(\gamma,t,v)}{\partial\gamma}\\ S_k^{(2)}(\gamma,t,v) &= \frac{\partial S_k^{(1)}(\gamma,t,v)}{\partial\gamma}.\end{aligned}$$

Moreover for $k = 1, 2$ let

$$\begin{aligned}
\hat{q}_k(t, v) &= \frac{1}{S_k^{(0)}(\hat{\gamma}_k, t, v)} \sum_{i=1}^{n} \frac{h_k(Q_{ki})\delta_{ki}\mathbb{1}\{T_{ki} \geqslant t, V_{ki} = v\} e^{\hat{\gamma}_k^T W_{ki}}}{\hat{K}_k(T_{ki} \mid V_{ki}, W_{ki})} \\
\hat{w}_k(t, v) &= \frac{S_k^{(1)}(\hat{\gamma}_k, t, v)}{S_k^{(0)}(\hat{\gamma}, t, v)} \\
\hat{H}_k(t, v, w) &= \sum_{i=1}^{n_k} \frac{e^{\hat{\gamma}_k^T w}(1 - \delta_{ki})\mathbb{1}\{C_{ki} \leqslant t, V_{ki} = v\}}{S_k^{(0)}(\hat{\gamma}_k, C_{ki}, v)} \{w - \hat{w}_k(C_{ki}, v)\} \\
\hat{R}_k &= \frac{1}{n_k} \sum_{i=1}^{n_k} \frac{h_k(Q_{ki})\delta_{ki}\hat{H}_k(T_{ki}, V_{ki}, W_{ki})}{\hat{K}_k(T_{ki} \mid V_{ki}, W_{ki})} \\
\hat{D}_{ki}(t) &= \hat{q}_k(t, V_{ki}) + \hat{R}_k n_k \hat{\Omega}_k^{-1} \{W_{ki} - \hat{w}_k(t, V_{ki})\} \\
\hat{S}_{ki} &= \frac{h_k(Q_{ki})\delta_{ki}}{\hat{K}_k(T_{ki} \mid V_{ki}, W_{ki})} + (1 - \delta_{ki})\hat{D}_{ki}(C_{ki}) \\
&\quad - e^{\hat{\gamma}_k^T W_{ki}} \sum_{j=1}^{n_k} \frac{(1 - \delta_{kj})\mathbb{1}\{C_{kj} \leqslant T_{ki}, V_{kj} = V_{ki}\}\hat{D}_{ki}(C_{kj})}{S_k^{(0)}(\hat{\gamma}_k, C_{kj}, V_{kj})},
\end{aligned}$$

The sample variance of $\hat{S}_{ki}$, $i = 1, \ldots, n_k$ is a consistent estimator of σ_k^2:

Lemma 6.8. *As* $n_k \to \infty$,

$$\hat{\sigma}_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} \left(\hat{S}_{ki} - n_k^{-1} \sum_{j=1}^{n_k} \hat{S}_{kj} \right)^2 \xrightarrow{p} \sigma_k^2$$

Proof. The result follows from the law of large numbers, the lemma of Slutzky (lemma A.37) and lemma 6.5. □

Simulation results

Table 6.1 shows the results of 10^5 simulations for the IPCW Mann-Whitney U-test and the average hazard ratio from 3, which treats the QAS observations as ordinary survival data. The data was simulated according to the Q-TWiST model described in Zhao and Tsiatis [2001]. First the patients experience toxicity (TOX), then they experience time without symptoms or toxicity (TWIST) until disease replapse. Here TWIST has an exponential distribution with rate $\lambda = 1/52$ truncated at 104. Toxicity (TOX) is uniformly distributed on $[TWIST/4, TWIST/2]$ in group 1 and is equal to 52 minus a uniform random variable on $[TWIST/4, TWIST/2]$ in group 2. Censoring is uniform

on $[10, 208]$ in both groups. Quality-adjusted survival is set equal to TWIST. In this model the time to relapse (= TOX + TWiST) is different in the two groups, whereas quality-adjusted survival (= TWiST) has the same distribution in both groups. For a group size of 50, i.e. a total sample size of 100, the actual type-I-error of the IPCW methods is 0.056 and approaches the nominal level 0.05 as the sample size increases. The bias of the IPCW method is close to 0. This is a substantial improvement over the AHR method, which is biased and has an inflated type-I-error rate, since it does not account for the informative censoring inherent to quality-adjusted survival data.

Table 6.1: Type-I-error under the null hypothesis $\theta = 0.5$ for the IPCW Mann-Whitney U-test and the average hazard ratio for QAS data. [a]

	$\hat{\theta}$		Type-I-error	
n	IPCW	AHR	IPCW	AHR
50	0.497	0.454	0.056	0.190
100	0.499	0.455	0.052	0.256
200	0.500	0.455	0.051	0.381

[a] Results of 10^5 simulation runs for different group sizes n, $\alpha = 0.05$.

6.5 SEQUENTIAL AND ADAPTIVE DESIGNS

While the previous result was only derived for right-censored data, it is straightforward to extend to left-truncated and right-censored data, since the proportional hazards model can also be fit to left-truncated data. This can be used together with the inverse normal method for non-adaptive sequential trials or adaptive trials, where no surrogate information is used in the interim decisions. When interim decisions are also based on surrogate information, then the stratification approach of chapter 5 can be combined with inverse probablitity weighting. Suppose the survival function of QAS in each stratum is estimated by the IPCW estimator

$$\hat{S}_j(t) = \frac{1}{n_j} \sum_{l=1}^{n} \frac{\mathbb{1}\{Q_l > t\}\delta_l \mathbb{1}\{X_l = j\}}{\hat{K}(T_l|V_l, W_l, X_l)},$$

where X_l is a discrete surrogate variable as in chapter 5 and V_l and W_l are the covariates in the (stratified) proportional hazard model for the censoring times. The estimator of the marginal survival function of QAS is then given by

$$\begin{aligned}\hat{S}(t) = \sum_{j=1}^{J} \frac{n_j}{n} \hat{S}_j(t) &= \sum_{j=1}^{J} \frac{n_j}{n} \frac{1}{n_j} \sum_{l=1}^{n} \frac{\mathbb{1}\{Q_l > t\}\delta_l \mathbb{1}\{X_l = j\}}{\hat{K}(T_l|V_l, W_l, X_l)} \\ &= \frac{1}{n} \sum_{l=1}^{n} \frac{\delta_l \mathbb{1}\{Q_l > t\}}{\hat{K}(T_l|V_l, W_l, X_l)} \underbrace{\sum_{j=1}^{J} \mathbb{1}\{X_l = j\}}_{=1} \\ &= \frac{1}{n} \sum_{l=1}^{n} \frac{\delta_l \mathbb{1}\{Q_l > t\}}{\hat{K}(T_l|V_l, W_l, X_l)}.\end{aligned}$$

This is identical to the IPCW estimator in the whole sample, but with the additional stratification variable X in the censoring model. Thus, the IPCW survival function estimators can be used with the patient-wise splitting approach, by adding the surrogate variable as a stratification variable to the censoring model.

6.6 DISCUSSION

QAS is an interesting integrated measure of clinical benefit, combining overall survival and longitudinal quality-of-life data. However, the definition of this endpoint creates methodological problems. The large body of established methods for the analysis of time-to-event and longitudinal data, cannot be used for the analysis of QAS. The repeated measures are conflated to a single number, preventing the use of methods for longitudinal data. The induced dependent censoring prevents the use of the standard survival methods. Moreover, when using summary measures of longitudinal data, careful planning is required, since the assumption of equal variances may not be plausible, if the number and/or timing of repeated measures is different between individuals and/or groups.

The IPCW approach for time-to-event data is very general and flexible. Arbitrary censoring patterns can be handled, including left-truncation and right-censoring. The censoring mechanism can be modelled with discrete and continuous covariates. However, usually the censoring mechanism is considered a nuisance parameter.

Under independent censoring of the underlying survival times, the IPCW estimator of the survival function of QAS can be used with the methods of chapter 3. For

covariate-dependent censoring this is not possible, but the IPCW Mann-Whitney U-test provides a direct approach to the estimation of the relative risk in this situation. The IPCW methods can be used with the patient-wise approach of chapter 5, when the surrogate variable is used as stratification variable in the censoring model. Thus provding the novel possibility of adaptive designs QAS data.

7

CONCLUSION

In confirmatory phase III trials strict control of the type-I-error is important. There exists a large literature on adaptive and group-sequential designs for time-to-event data. Most of it focuses on the log-rank test and the Cox proportional hazards model. Much less results are available for non-proportional hazards data. In this work several novel non-parametric sequential and adaptive methods for survival data have been developed.

As opposed to the proportional hazards situation, where the log-rank test is the most efficient test, there is no "best" test in the non-proportional hazards situation. Alternative tests for comparing distributions in the non-proportional hazards case exist, such as that of Pepe and Fleming [1989]. The average hazard ratio has been proposed by Kalbfleisch and Prentice [1981] as a generalization of the usual hazard ratio with a useful interpretation in situation with non-proportional hazards. If the proportional hazards assumption holds, then the average hazard ratio reduces to the usual hazard ratio. In this work it was proved, that the average hazard ratio based test statistics do have (asymptotically) the independent increments structure, allowing group-sequential rejection boundaries to be calculated with standard methods. The average hazard ratio can now be used as an endpoint in group-sequential trials. In fact, the result holds for many other non-parametric test statistics, which are functions of the marginal survival function.

In adaptive designs for survival trials usually the independent increments property of the test statistics is exploited to achieve independent stage-wise test statistics (Wassmer [2006]). When surrogate information, e.g. short-term endpoints, is used in the interim decision, then the independent increments property does no longer hold in general, as noted by Bauer and Posch [2004] and Magirr et al. [2014]. This problem is not restricted to time-to-event data, but is present in any data with delayed response or long-term follow-up.

For a discrete surrogate two solutions to this problem have been proposed in this work, which are both based on (modifications of) the weighted Kaplan-Meier estimator (Malani [1995] and Murray and Tsiatis [1996]). In the patient-wise splitting approach of Jenkins et al. [2011] the weighted Kaplan-Meier estimator is used to ac-

count for possible informative censoring induced by modification of the follow-up of overrunning patients. In the stage-wise splitting approach a modified weighted Kaplan-Meier estimator is used to account for a possible biased response rate. These approaches can be applied to all test statistics derived from the weighted Kaplan-Meier estimator.

The patient-wise splitting approach, where the test statistics for the primary endpoint are only calculated at the end of the trial, allows for early interim analyses and decision making based on surrogate information, even before the primary endpoint has been observed for most patients. The drawback is, that stopping for efficacy based on the primary endpoint is not possible at interim. The stage-wise splitting approach, on the other hand, is more similar to the standard group-sequential approach. Combined with the inverse normal method, the usual rejection boundaries for group-sequential trial can be used.

The patient-wise splitting approach can also be applied to the somewhat more general quality-adjusted survival endpoint. The major difficulty with the quality-adjusted endpoint is, that censoring of the quality-adjusted survival times is informative even when the censoring of the underlying survival times is non-informative. This is solved by inverse probability of censoring weighting. The inverse probability weighting approach even allows for the inclusion of continuous covariates in order to account for informative censoring cases, provided one is willing to accept a semi-parametric model for the conditional censoring distribution. These are the first proposed adaptive designs for a quality-adjusted survival endpoint.

The main limitation of this work is the restriction to time-independent discrete covariates. This excludes the important cases of normal and time-to-event surrogates. Inclusion of continuous or even censored time-to-event endpoints like progression-free survival is not directly possible with the weighted Kaplan-Meier approach.

The restriction to a fixed time interval $[0, L]$ is mostly a technical limitation without much practical relevance, which is required for the asymptotic results (this could actually be relaxed at the cost of additional regularity assumptions). In the non-parametric case however, the time point L is a part of the definition of the endpoint, e.g. the average hazard ratio over 1 year is a different endpoint, than the average hazard ratio over 2 years. In principle, L could be changed at the interim analysis without compromising the overall type-I-error rate, but this would amount to a selection of a different endpoint, since e.g. the average hazard ratio over 1 year is not the same endpoint as the average hazard ratio over 2 years.

Another limitation is, that the asymptotic results only hold under the null hypothesis, hence cannot be used for power considerations. Sample size and other design parameters depending on a specific alternative hypothesis need to be evaluated in simulation studies.

Furthermore, the completely non-parametric approach in the construction of the test statistics, i.e. no assumption about the form the hazard rate, sacrifices some power compared to a correctly specified parametric model.

Extension of the method to continuous and especially time-to-event surrogate variables is a possible route for future research. A direct extension of the weighted Kaplan-Meier approach to continuous covariates is possible by discretization of the continuous covariate. However this introduces additional problems, such as choosing number and location of the breakpoints. Imposing a semi-parametric model for the conditional hazard, e.g. a proportional hazard model, is another possibility (Shen and Fleming [1997]), which, however, requires model assumptions. How to apply such methods in sequential and adaptive designs needs to be inverstigated.

Another direction of future research is how surrogate information can be used in the interim decision process. It is expected, that the inclusion of surrogate information improves the interim decision process, e.g. reduce the probability of wrongly selecting a subgroup or treatment in adaptive enrichment trials. This has to be examined by extensive simulation studies.

A

MATHEMATICAL BACKGROUND

This chapter presents the mathematical / probability theoretic background related to stochastic processes required for the asymptotic results in this thesis. These are all standard results and can be found in any textbook on measure theoretic probability theory (see e.g. Billingsley [2009] and Andersen [1993]).

A.1 STOCHASTIC PROCESSES

In this section let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{B}(\mathbb{R})$ be the Borel σ-algebra on $\mathbb{R}$ and $I \subset \mathbb{R}$ an arbitrary subset of the real numbers. Usually I is the set $[0, \infty)$ and is interpreted as time. The Borel σ-algebra on I is denoted by $\mathcal{B}(I)$.

Definition A.1 (Stochastic process)**.** *A collection of random variables* $X = \{X(t), t \in I\}$ *on* $(\Omega, \mathcal{F}, P)$ *with values in* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is called a* stochastic process *with index set* I *and state space* $\mathbb{R}$*.*

Definition A.2. *A stochastic process* $\{X(t), t \in I\}$ *with values in* $\mathbb{R}$ *is called*

1. integrable, *if* $E[|X(t)|] < \infty$ *for all* t,
2. square integrable, *if* $E[|X(t)|^2] < \infty$ *for all* t.

A stochastic process $\{X(t), t \in I\}$ *with values in* E *is called*

1. process with independent increments, *if for all* $n \in \mathbb{N}$ *and all* $t_0, \ldots, t_n \in I$ *with* $t_0 < \ldots < t_n$ *the random variables* $X(t_i) - X(t_{i-1})$, $i = 1, \ldots, n$ *are (jointly) independent.*
2. (left-/right-)continuous, *if its paths are (left-/right-)continuous a.s., i.e. if for almost every* $\omega \in \Omega$ *the function* $t \mapsto Z(t, \omega)$ *is (left-/right-)continuous,*
3. jointly measurable, *if the map* $X : \Omega \times I \to E$, $X(\omega, t) = X(t)(\omega)$ *is* $\mathcal{F} \otimes \mathcal{B}(I) - \mathcal{B}(\mathbb{R})$*-measureable.*

Definition A.3 (Filtration). *A family* $\{\mathcal{F}_t, t \in I\}$ *of* σ*-algebras on* $(\Omega, \mathcal{F}, P)$ *with* $\mathcal{F}_t \subseteq \mathcal{F}$ *for each* t *is called* filtration, *if* $\mathcal{F}_s \subseteq \mathcal{F}_t$ *for all* $s, t \in I$ *with* $s \leqslant t$. *The filtration is called* complete, *if* $\mathcal{F}_0$ *contains all* P*-null sets. The filtration is called* right-continuous, *if*

$$\mathcal{F}_s = \bigcap_{s<t} \mathcal{F}_t.$$

Definition A.4 (adapted). *A stochastic process* $\{X(t), t \in I\}$ *is called* adapted *to the filtration* $(\mathcal{F}_t, t \in I)$, *if* $X(t)$ *is* $\mathcal{F}_t$*-measurable for all* $t \in I$.

Definition A.5. *Let* $\mathbb{R}^I$ *be the set of functions from* I *to* $\mathbb{R}$. *Define the product* σ*-algebra* $\mathcal{B}(\mathbb{R})^{\otimes I}$ *on* $\mathbb{R}^I$ *as the smallest* σ*-algebra, which contains all sets of the form*

$$\{f \in \mathbb{R}^I, f(t_1) \in A_1, \ldots, f(t_k) \in A_k\}$$

for $k \in \mathbb{N}$, $t_1, \ldots, t_k \in I$, $A_1, \ldots, A_k \in \mathbb{R}$.

Lemma A.6. *Let* $\{X(t), t \in I\}$ *be a stochastic process on* $(\Omega, \mathcal{F}, P)$ *with state space* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, *then the map* $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^I, \mathcal{B}(\mathbb{R})^{\otimes I})$, *defined by*

$$X(\omega)(t) := X(t)(\omega) \quad \omega \in \Omega, t \in I$$

is measurable, i.e. $X^{-1}(B) \in \mathcal{F}$ *for all* $B \in \mathcal{B}(\mathbb{R})^{\otimes I}$.

Definition A.7 (Distribution of a stochastic process). *Let* $\{X(t), t \in I\}$ *be a jointly measurable process on* $(\Omega, \mathcal{F}, P)$ *with state space* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *Define the map* X *as in the preceding lemma. The probability measure* P^X *on* $(\mathbb{R}^I, \mathcal{B}(\mathbb{R})^{\otimes I})$ *defined as the image measure of the map* X,

$$P^X(B) := P\left(X^{-1}(B)\right) \quad B \in \mathcal{B}(\mathbb{R})^{\otimes I}$$

is called the distribution *of* $\{X(t), t \in I\}$.

Definition A.8 (finite dimensional distributions). *Let* $\{X(t), t \in I\}$ *be a stochastic process and denote the distribution of* $X(t)$ *by* $P_t, t \in I$. *The family of probability measures* $\{P_t, t \in J \subseteq I, J$ *finite*$\}$ *is called* family of finite-dimensional distributions *of* $\{X(t), t \in I\}$.

Theorem A.9. *The distribution of a stochastic process is uniquely determined by its finite-dimensional distributions.*

Definition A.10 (Martingale). *A stochastic process* M *is called a* martingale *with respect to the filtration* $\{\mathcal{F}_t, t \in I\}$ *if*

1. M *is adapted to* $\{\mathcal{F}_t, t \in I\}$,
2. M *is integrable,*
3. $E[M(t)|\mathcal{F}_s] = M(s)$ P*-a.s. for* $s \leqslant t$.

If $E[M(t)|\mathcal{F}_s] \geqslant M(s)$ P*-a.s. for* $s \leqslant t$ *holds instead of (3.), then the* M *is called* submartingale.

Definition A.11 (stopping time)**.** *A random variable* τ *with values in* $I \cup \{\infty\}$ *is called* stopping time *with respect to* $\{\mathcal{F}_t, t \in I\}$*, if for all* $t \in I$

$$\{\tau \leqslant t\} \in \mathcal{F}_t.$$

Definition A.12 (stopped process)**.** *Given a stochastic process* $\{X(t), t \in I\}$ *and a stopping time* τ*, the process* $X^\tau = \{X^\tau(t), t \in I\}$ *defined by*

$$X^\tau(t)(\omega) = X(t \wedge \tau(\omega))(\omega) \quad \forall \omega \in \Omega$$

is called stopped process*.*

Definition A.13 (local martingale)**.** *A stochastic process* M *is called a* local martingale*, if there exists an increasing sequence of stopping times* T_n*, with* $T_n \to \infty$ *a.s., such that for each* n *the stopped process* M^{T_n} *is a martingale.*

Definition A.14 (Gaussian process)**.** *A Gaussian process is a stochastic process, whose finite-dimensional distributions are multivariate normal.*

Lemma A.15. *A Gaussian process* $\{X(t), t \in I\}$ *is uniquely determined by its mean function* $t \mapsto E[X(t)]$ *and its covariance function* $\rho(s,t) = \mathrm{Cov}(X(s), X(t))$.

Example A.16 (Brownian motion)**.** *Brownian motion is a continuous square integrable Gaussian martingale with mean zero and covariance function* $s \wedge t$.

A.1.1 *Quadratic variation*

Definition A.17. *A stochastic process* $\{X(t), t \in I\}$ *is called*

1. finite variation process *if its paths are right-continuous with left-hand limits everywhere and are almost surely of finite variation on all bounded subsets of* I,

2. predictable *with respect to the filtration* $\{\mathcal{F}_t, t \in I\}$*, if it is measurable with respect to the predictable* σ*-algebra, i.e. the* σ*-algebra generated by all left-continuous adapted processes,*

3. locally bounded*, if there exists a sequence of stopping times* $(T_n)_{n\in\mathbb{N}}$*, such that the stopped process* $X_{t\wedge T_n}$ *is bounded.*

Definition A.18 (Compensator)**.** *A finite-variation predictable process* A *with* $A(0) = 0$ *is called* compensator *of the process* N *with respect to the filtration* $\{\mathcal{F}_t, t \in I\}$*, if* $N - A$ *is a local martingale with respect to* $\{\mathcal{F}_t, t \in I\}$*.*

The Doob-Meyer decomposition asserts the existence and uniqueness of a compensator of a non-negative right-continuous local submartingale.

Theorem A.19 (Doob-Meyer decomposition, theorem 2.2.3 in Fleming and Harrington [2011])**.** *Let* $X = \{X(t), t \in [0, \infty)\}$ *be a non-negative right-continuous local submartingale with respect to the right-continuous filtration* $\{\mathcal{F}_t, t \in I\}$*. Then there exists a unique increasing right-continuous predictable process* A *such that* $A(0) = 0$ *a.s.,* $P(A(t) < \infty) = 1$ *for* $t > 0$*, and* $X - A$ *is a right-continuous local martingale.*

Definition A.20 (Quadratic variation)**.** 1. *The* optional quadratic covariation *of two local martingales* X *and* Y *is defined as the unique increasing process* $[X, Y]$ *such that*

$$XY - [X, Y]$$

is a local martingale. The optional quadratic variation *of a local martingale* X *is defined as* $[X] := [X, X]$*.*

2. *The* predictable quadratic covariation *of two stochastic processes* X *and* Y *is defined as the unique increasing predictable process* $\langle X, Y\rangle$ *such that*

$$XY - \langle X, Y\rangle$$

is a local martingale. The predictable quadratic variation *of a local square integrable martingale* X *is defined as* $\langle X\rangle := \langle X, X\rangle$*.*

Theorem A.21 (Bilinearity of quadratic covariations)**.** *Let* $X_1, \ldots, X_n$ *and* $Y_1, \ldots, Y_n$ *be local martingales. Then*

$$\left[\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} [X_i, Y_j].$$

If the predictable quadratic covariations $\langle X_i, Y_j \rangle$ exist, then

$$\left\langle \sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \langle X_i, Y_j \rangle.$$

Theorem A.22 (Ito isometry)**.** *Let M_1 and M_2 be two local martingales. If $E[[M_1]] < \infty$ and $E[[M_2]] < \infty$, then M_1 and M_2 are square integrable and*

$$\mathrm{Cov}(M_1(t), M_2(t)) = E\left[\langle M_1, M_2 \rangle(t)\right] = E\left[[M_1, M_2](t)\right].$$

Definition A.23 (Orthogonal martingales)**.** *Two local martingales X and Y are called* orthogonal, *if $[X, Y] \equiv 0$.*

Remark A.24. *For independent martingales X and Y the product XY is already a martingale. Hence $[X, Y] \equiv 0$. The converse is not true in general, but orthogonal martingales are uncorrelated, by the Ito isometry (theorem A.22).*

Lemma A.25 (Quadratic variation of finite variation processes)**.** *Let X be a local martingale and Y a finite variation local martingale. Then*

$$[X, Y](t) = \sum_{s \leqslant t} \Delta X(s) \Delta Y(s),$$

where $\Delta X(s) = X(s) - X(s-)$ and $\Delta Y(s) = Y(s) - Y(s-)$. If X or Y is continuous, then $[X, Y] \equiv 0$.

Theorem A.26. *Let $X = \{X(t), t \in I\}$ be a finite variation process and $Y = \{Y(t), t \in I\}$ a locally bounded stochastic process. The integral*

$$Z(t) = \int_0^t Y dX \quad t \in I$$

is interpreted as pathwise Lebesgue-Stieltjes integral, i.e. the process $Z = \{Z(t), t \in I\}$ is defined by

$$Z(t)(\omega) = \int_0^t Y(s)(\omega) X(ds)(\omega) \quad \omega \in \Omega.$$

If Y is predictable, then this definition coincides with the Ito-integral of Y with respect to X. If X is a local (square integrable) martingale, then Z is also a local (square integrable) martingale.

Theorem A.27. *Let M_1 and M_2 be finite variation local square integrable martingales and H_1 and H_2 be predictable locally bounded processes. Then $\int H_1 \, dM_1$ and $\int H_2 \, dM_2$ are local square integrable martingales and*

$$\left[\int H_1 \, dM_1, \int H_2 \, dM_2\right] = \int H_1 H_2 \, d[M_1, M_2]$$

$$\left\langle \int H_1 \, dM_1, \int H_2 \, dM_2 \right\rangle = \int H_1 H_2 \, d\langle M_1, M_2 \rangle.$$

Theorem A.28 (Integration by parts formula)**.** *Let X and Y be local martingales. Then*

$$X(t)Y(t) = X(0)Y(0) + \int_0^t X(s-)Y(ds) + \int_0^t Y(s-)X(ds) + [X,Y](t),$$

where $X(s-) = \lim_{t \uparrow s} X(t)$ is the left limit of X at the point s.

Corollary A.29. *Let X and Y be mean zero local martingales, such that $\langle X, Y \rangle$ exists. Then $\langle X, Y \rangle$ is the compensator of $[X, Y]$, i.e.*

$$[X,Y] - \langle X, Y \rangle$$

is a local martingale.

A.1.2 *Counting processes*

Definition A.30 (Counting process)**.** *A* counting process *is a stochastic process $\{N(t), t \in [0,\infty)\}$ with values in $\mathbb{N}_0$ adapted to a filtration $\{\mathcal{F}_t, t \in [0,\infty)\}$ with $N(0) = 0$ and $N(t) < \infty$ a.s. whose paths are with probability one right-continuous, piecewise constant and have only jump discontinuities, with jumps of size $+1$.*

Remark A.31. *Counting processes are finite variation processes, because they have a.s. monotone increasing paths.*

By the Doob-Meyer decomposition (theorem A.19), unique predictable compensators for counting processes always exists.

Definition A.32 (Counting process martingales)**.** *Let N be a counting process and A its unique predictable compensator. The local square integrable martingale $N - A$ is called* counting process martingale*.*

Assume that A is the continuous compensator of the counting process N and let $M = N - A$ be the associated counting process martingale, then $[M] = [N - A] = [N] = N$. Since $\langle M \rangle$ is the compensator of $[M] = N$ it follows $\langle M \rangle = A$, i.e. $M^2 - A$ is square integrable local martingale and $Var(M) = E[M^2] = E[A]$, by the Ito isometry.

Lemma A.33. *Let $N_1, \dots, N_n$ ($n \geqslant 1$) be counting processes, which do not have any common jump discontinuities, i.e. $[N_i, N_j] \equiv 0$ for $i \neq j$. Let A_i be the compensator of N_i and $M_i = N_i - A_i$ the associated counting process martingales. Assume that A_i is continuous for $i = 1, \dots, n$. Then the martingales $M_1, \dots, M_n$ are orthogonal and the sum $N_1 + \dots + N_n$ is itself a counting process with*

$$\left[\sum_{i=1}^{n} M_i\right] = \left[\sum_{i=1}^{n} N_i\right] = \sum_{i=1}^{n} N_i$$

and

$$\left\langle \sum_{i=1}^{n} M_i \right\rangle = \left\langle \sum_{i=1}^{n} N_i \right\rangle = \sum_{i=1}^{n} A_i.$$

Lemma A.34 (Lenglart's inequality for counting processes, Corollary 3.4.1, Fleming and Harrington [2011])**.** *Let N be a counting process, and $M = N - A$ the corresponding local square integrable martingale. Suppose H is an adapted left-continuous process with right-hand limits or, more generally, a predictable and locally bounded process. Then for any stopping time T such that $P(T < \infty) = 1$, and any $\epsilon, \eta > 0$,*

$$P\left(\sup_{t \leqslant T} \left\{ \int_0^t H(s) M(ds) \right\}^2 \geqslant \epsilon \right) \leqslant \frac{\eta}{\epsilon} + P\left(\int_0^T H^2(s) \langle M, M \rangle (ds) \geqslant \eta \right)$$

The next lemma is a simple consequence of Lenglart's inequality.

Lemma A.35. *Consider the integrals*

$$\frac{1}{\sqrt{n}} \int_0^t H^{(n)}(u) M^{(n)}(du) \quad t \leqslant \tau$$

with respect to the (local) martingales $M^{(n)}$ with increasing compensators $\Lambda^{(n)}$, such that all conditions of Lenglart's inequality hold. Moreover suppose that there exists a function H, such that

$$\sup_{u \in [0,\tau]} |H^{(n)}(u) - H(u)| = o_p(1),$$

and for each $t \leqslant \tau$,

$$\Lambda^{(n)}(t) = O_p(n)$$

then

$$\sup_{u\in[0,\tau]} \left| \frac{1}{\sqrt{n}} \int_0^t H^{(n)}(u) M^{(n)}(du) - \frac{1}{\sqrt{n}} \int_0^t H(u) M^{(n)}(du) \right| = o_p(1)$$

Proof. By Lenglart's inequality, for each $\epsilon > 0$ and $\eta > 0$,

$$\begin{aligned} &P\left(\sup_{u\in[0,\tau]} \left| \frac{1}{\sqrt{n}} \int_0^t (H^{(n)}(u) - H(u)) M^{(n)}(du) \right|^2 > \epsilon \right) \\ &\leqslant \frac{\eta}{\epsilon} + P\left(\left| n^{-1} \int_0^\tau (H^{(n)}(u) - H(u))^2 \Lambda^{(n)}(du) \right| > \eta \right) \end{aligned} \tag{A.1.1}$$

The second term on the right hand side of A.1.1 converges to zero for each η, since

$$\begin{aligned} &\left| \int_0^\tau (H^{(n)}(u) - H(u))^2 n^{-1} \Lambda_n(du) \right| \\ &\quad \leqslant \sup_{u\in[0,\tau]} |H^{(n)}(u) - H(u)|^2 \frac{\Lambda^{(n)}(\tau)}{n} = o_p(1) \end{aligned} \tag{A.1.2}$$

Hence the right-hand side of A.1.1 can be made arbitrarily small for each ϵ and the conclusion follows. □

A.2 WEAK CONVERGENCE OF MEASURES

A.2.1 *The space* $D[0, L]$

Denote by $D[0, L]$ the space of all right-continuous functions on $[0, L]$ with left limits. This space is complete but not separable with respect to the uniform metric. Therefore $D[0, L]$ is equipped with the *Skorohod metric* (Billingsley [2009]) to make it a separable and complete metric space. The standard theory for stochastic and weak convergence in metric spaces can be applied. The Skorohod metric restricted to the subspace $C([0, L])$ of continuous functions on $[0, L]$ coincides with usual uniform metric.

A.2.2 *Weak convergence in metric spaces*

Definition A.36 (Weak convergence). *A sequence of finite measures $(\mu_n)_{n\in\mathbb{N}}$ on a metric space (E, d) converges weakly to a measure μ, if*

$$\int f d\mu_n \to \int f d\mu$$

as $n \to \infty$ for all continuous bounded functions from E to $\mathbb{R}$. Weak convergence of $(\mu_n)_{n\in\mathbb{N}}$ to μ is denoted by $\mu_n \xrightarrow{\mathcal{L}} \mu$. A sequence of random variables with values in (E, d) converges weakly if its distributions converge weakly.

Lemma A.37 (Slutzky). *Let $X, X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be random variables with values in the metric space (E, d) and $X_n \xrightarrow{\mathcal{L}} X$ and $d(X_n, Y_n) \to 0$ in probability as $n \to \infty$. Then $Y_n \xrightarrow{\mathcal{L}} X$.*

Theorem A.38 (Continous mapping theorem). *Let (E_1, d_1) and (E_2, d_2) be two metric spaces and $\phi : E_1 \to E_2$ measurable and U_ϕ the set of discontinuity points of ϕ.*

1. *If $\mu, \mu_1, \mu_2, \ldots$ are probability measures with $\mu(U_\phi) = 0$ and $\mu_n \to \mu$ weakly, then $\mu_n \circ \phi^{-1} \to \mu \circ \phi^{-1}$ weakly.*
2. *If $X, X_1, X_2, \ldots$ are random variables with values in E_1 with $P(X \in U_\phi) = 0$ and $X_n \xrightarrow{\mathcal{L}} X$, then $\phi(X_n) \xrightarrow{\mathcal{L}} \phi(X)$.*

Definition A.39 (Tightness). *The collection $\mathcal{M}$ of finite measures on the topological space (E, τ) is called* tight, *if for every $\epsilon > 0$ there exists a compact set $K \subset \Omega$, such that*

$$\sup\{\mu(E \setminus K) : \mu \in \mathcal{M}\} < \epsilon.$$

A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ on a proability space $(\Omega, \mathcal{F}, P)$ is called tight, *if the collection $\{P \circ X_n^{-1} : n \in \mathbb{N}\}$ is tight.*

Lemma A.40. *For each $i = 1, \ldots, k$ let $(X_{in})_{n\in\mathbb{N}}$ be a tight sequence of random variables. Then the sequences*

$$(X_{1n} + \ldots + X_{kn})_{n\in\mathbb{N}}$$

and

$$((X_{1n}, \ldots, X_{kn}))_{n\in\mathbb{N}}$$

are tight.

Theorem A.41 (Stochastic equicontinuity, Billingsley [2009]). *A sequence of random variables $(Z_n)_{n\in\mathbb{N}}$ with values in a metric space (E,d) is tight, if for all $\epsilon > 0$ and $\eta > 0$ there exists $\delta > 0$, such that*

$$\limsup_{n\to\infty} P\left(\sup_{s,t\in E:d(s,t)<\delta} |Z_n(s) - Z_n(t)| < \epsilon\right) < \eta$$

as $n \to \infty$.

Theorem A.42 (Prohorov). *Let (E,d) be a metric space and $\mathcal{M}$ a collection of probability measures on E.*

1. *$\mathcal{M}$ is relatively compact, i.e. its closure is compact, if it is tight.*
2. *If E is polish, then $\mathcal{M}$ is relatively compact if and only if it is tight.*

Theorem A.43. *A sequence of random variables $(Z_n)_{n\in\mathbb{N}}$ with values in $D[0,L]$ converges weakly to random variable Z with values in $D[0,L]$, if*

1. *the finite-dimensional distributions of Z_n converge in distribution to the finite-dimensional distributions of Z, i.e. as $n \to \infty$, $(Z_n(t_1),\dots,Z_n(t_m))$ converges in distribution to $(Z(t_1),\dots,Z(t_m))$, for all $t_1,\dots,t_m \in [0,L]$, $m \geqslant 1$,*
2. *the sequence $(Z_n)_{n\in\mathbb{N}}$ is tight.*

Definition A.44 (Asymptotic covariance function). *Let $(Z_n)_{n\in\mathbb{N}}$ be sequence of stochastic processes with state space $D[0,L]$, which converges weakly to a stochastic process Z as $n \to \infty$. The covariance function of Z, $(s,t) \mapsto \mathrm{Cov}(Z(s),Z(t))$, is called* asymptotic covariance function *of $(Z_n)_{n\in\mathbb{N}}$.*

A.2.3 *Martingale central limit theorem*

Theorem A.45 (Andersen and Gill [1982]). *For each $n = 1,2,\dots$ let $N^{(n)}$ be a multivariate counting process with n components. Let $H^{(n)}$ be a $p \times n$ ($p \leqslant 1$ is fixed) matrix of locally bounded predictable processes. Suppose that $N^{(n)}$ has an intensity process $\lambda^{(n)}$, and define local square integrable martingales $W^{(n)} = (W_1^{(n)},\dots,W_p^{(n)})$ by*

$$W_i^{(n)}(t) = \int_0^t \sum_{l=1}^{n} H_{il}^{(n)}(u)\{dN_l^{(n)}(u) - \lambda_l^{(n)}(u)du\}$$

Let A *be a* $p \times p$ *matrix of continuous functions on* $[0, L]$ *which form the covariance functions of a continuous* p*-variate Gaussian martingale* $W^{(\infty)}$*, with* $W^{(\infty)}(0) = 0$*; i.e.*

$$\mathrm{Cov}(W_i^{(\infty)}(t), W_j^{(\infty)}(u)) = A_{ij}(t \wedge u)$$

for all i, j, t *and* u*. Suppose that for all* i, j *and* t

$$\left\langle W_i^{(n)}, W_j^{(n)} \right\rangle (t) = \int_0^t \sum_{l=1}^{n} H_{il}^{(n)}(s) H_{jl}^{(n)}(s) \lambda_l^{(n)}(s) ds \xrightarrow{p} A_{ij}(t) \tag{A.2.1}$$

as $n \to \infty$ *and that for all* i *and* $\epsilon > 0$

$$\int_0^L \sum_{l=1}^{n} H_{il}^{(n)}(t)^2 \lambda_l^{(n)}(t) \mathbb{1}\left\{ \left| H_{il}^{(n)}(t) \right| > \epsilon \right\} dt \xrightarrow{p} 0 \quad \text{as } n \to \infty. \tag{A.2.2}$$

Then $W^{(n)} \xrightarrow{\mathcal{L}} W^{(\infty)}$ *as* $n \to \infty$ *in* $D([0, L]^p)$*.*

Remark A.46 (Lindeberg condition)**.** *Assume the processes* $H^{(n)}$ *of theorem A.45 are of the form*

$$\frac{1}{\sqrt{n}} \tilde{H}^{(n)},$$

where $\tilde{H}^{(n)}$ *is a almost surely bounded predictable process. Then the Lindeberg condition (eq. (A.2.2)) is trivially satisfied, since the indicator function* $\mathbb{1}\left\{ \left| H_{il}^{(n)}(t) \right| > \epsilon \right\}$ *is eventually* 0 *a.s. for all* s *and any* $\epsilon > 0$ *for large* n*.*

Remark A.47. *Independent martingales are orthogonal. The converse is not true in general, but orthogonal martingales are uncorrelated, by the Ito isometry, and asymptotically independent, by the martingale central limit theorem.*

A.2.4 *Functional delta method*

Definition A.48 (Hadamard differentiability, definition II.8.1 of Andersen [1993])**.** *Let* B, B' *be Banach spaces.* $\phi : B \to B'$ *is* compactly *or* Hadamard differentiable *at a point* $\theta \in B$ *if and only if a continuous, linear map*

$$d\phi(\theta) : B \to B'$$

exists (called the derivative of ϕ at the point θ) such that for all real sequences $a_n \to \infty$ and all convergent sequences $h_n \to h \in B$,

$$a_n(\phi(\theta + a_n^{-1}h_n) - \phi(\theta)) \to d\phi(\theta) \cdot h \quad \textit{as } n \to \infty.$$

Lemma A.49. *1. A continuous, linear map $\phi : B \to B'$ is Hadamard differentiable at any point $\theta \in B$ with $d\phi(\theta) = \phi$.*

2. Denote the set of real-valued functions on $[0, L]$ whose total variation is bounded by the constant M by $BV_M[0, L]$. The map $\phi : D[0, L] \times BV_M[0, L] \to \mathbb{R}$ defined by

$$(F, G) \mapsto \int F dG$$

is Hadamard differentiable with derivative at the point (F, G) defined by

$$(f, g) \mapsto d\phi(F, G)(f, g) = \int F dg + \int f dG.$$

Proof. 1. For any $h \in B$ and $h_n \to h$, by linearity and continuity of ϕ,

$$a_n(\phi(\theta + a_n^{-1}h_n) - \phi(\theta)) = \phi(h_n) \to \phi(h) = d\phi(\theta) \cdot h.$$

2. Follows from lemma 3.9.17 in Vaart and Wellner [1996].

□

Theorem A.50 (Functional delta method, theorem II.8.1 of Andersen [1993])**.** *Let T_n be a sequence of random variables with values in a Banach space B, $a_n \to \infty$ a real sequence, such that*

$$a_n\{T_n - \theta\} \xrightarrow{\mathcal{L}} Z$$

for some fixed point $\theta \in B$ and a random variable Z with values in B. Suppose $\phi : B \to B'$ is Hadamard differentiable at θ. Then

$$a_n\{\phi(T_n) - \phi(\theta)\} \xrightarrow{\mathcal{L}} d\phi(\theta) \cdot Z$$

and, moreover,

$$a_n\{\phi(T_n) - \phi(\theta)\} \quad \textit{and} \quad d\phi(\theta) \cdot a_n\{T_n - \theta\}$$

are asymptotically equivalent.

A.3 PRODUCT INTEGRAL

The survival function can also be written as the product integral (Gill and Johansen [1990]) of the cumulative hazard rate. For our special case of univariate real-valued functions the product integral can be defined in the following way:

Definition A.51 (Definition 4 of Gill and Johansen [1990])**.** *Let* X *be any real-valued signed-measure. Define*

$$\prod_{(0,t]}(1-dX) := \prod_{s\in(0,t]}(1+X(\{s\}))e^{X^c(t)}$$

where $X^c(t) = X(t) - \sum_{s\leqslant t} X(\{t\})$ *is the continuous part of* X*.*

Since the cumulative hazard function Λ is continuous, we get

$$\prod_{(0,t]}(1-d\Lambda) = e^{-\Lambda(t)} = S(t).$$

Theorem A.52 (Compact differentiability of the product integral with respect to the supremum norm, Theorem 8 in Gill and Johansen [1990])**.** *Consider the product integral as a mapping* $\mathcal{P}$ *from the space of additive interval functions on* $(0,\tau]$ *with variation bounded by the constant* c *to the space of interval functions on* $(0,\tau]$*, both domain and range endowed with the supremum norm. Let* α *be given and define* $\mu = \mathcal{P}(\alpha) = \prod(1+d\alpha)$*. Then* $\mathcal{P}$ *is compactly differentiable at* α *with derivative* $d\mathcal{P}(\alpha)$ *given by*

$$(d\mathcal{P}(\alpha)\cdot h)(s,t) = \int_{(s,t]} \mu(s,u-)\mu(u,t)h(du),$$

where the integral with respect to h *is defined by the integration by parts formula.*

Remark A.53. *If* Λ *is continuous, then the Hadamard derivative of the product integral at* Λ *is simply* $-\exp(-\Lambda) = -S$*.*

B

DESCRIPTION OF THE R PACKAGE AHR

The package `AHR` (Average Hazard Ratio) implements the weighted Kaplan-Meier estimator and estimation and testing of the (log-) average hazard ratio, as well as estimation and testing of the restricted mean and arbitrary quantiles of the survival curves estimated with the weighted Kaplan-Meier estimator. All methods are implemented in pure `R` (R Core Team [2014]), except for the weighted Kaplan-Meier estimator, which is based on a custom `C++` implementation of the Kaplan-Meier estimator for reasons of speed, since the standard `survfit` (package `survival`) and `prodlim` (package `prodlim`) functions contain a lot of overhead, because of many additional features not needed in the package `AHR`.

FASTKM

Description

This function calculates the Kaplan-Meier estimator for right-censored survival data, at arbirtrary time points. It can handle left-truncated and/or right-censored data with ties. Avoids the overhead of the `survfit` or `prodlim` functions by stripping away most of the features not needed here. This function simply passes the data to the `C++` code, which does the real work.

Usage

```
fastkm(time, status, ltrunc = rep.int(0, length(time)),
       left.limit = FALSE, eval = time)
```

Arguments

`time`	vector of right-censored survival times

`status` censoring indicator for each element of `time` (0 = right-censored, 1 = event)

`ltrunc` vector of left-truncation times

`left.limit` indicates wether estimated survival function is left continuous

`eval` points at which the estimated survival function should be evaluated

Value

A list containing the elements

`time` vector of evaluation times (equal to the `eval` argument or the `time` argument if `eval=NULL`)

`surv` survival estimated at each element of `eval` (or `time` if `eval=NULL`)

`variance` variance estimate (Greenwood) at each element of `time`

`n.atrisk` number at risk at each event time (only if `eval=NULL`)

WKM

Description

Weighted Kaplan-Meier estimator with discrete time-independent covariate.

Usage

```
wkm(times, data, start = 0, alpha = 1)
```

Arguments

`times:` a vector of evaluation times

`data:` a data frame or list containing the vectors `Y` (time), `D` (censoring indicator), `W` (stratification variable), `R` (recruitment times) and `V` (left truncation times)

`start`: time of interim analysis (estimation of response rates is based only on data accruing after time `start`)

`alpha`: fractional parameter

Value

An object of class "wkm", i.e. a list containing the elements

`times`: the `times` argument supplied to the function

`start`: the `start` argument supplied to the function

`p`: the estimated response rates

`S`: vector of survival probabilities

`COV`: (asymptotic) covariance matrix of the vector `S`

`logCOV`: (asymptotic) covariance matrix of the vector `log S`

`V`: (asymptotic) variances of the vector `S`

`logV`: (asymptotic) variances of the vector `log S`

WKM.QUANTILE

Description

Estimate arbitrary quantiles of a survival distribution based on the weighted Kaplan-Meier estimator.

Usage

```
wkm.quantile(tau, data = NULL, conf.level = 0.95, null.value = NULL,
             start = 0)
```

Arguments

`tau:`	number between 0 and 1 specifiying quantile to estimate
`data:`	a data frame of list containing the variables in the model (see `wkm`)
`conf.level:`	confidence level at which to calculate the confidence interval
`null.value:`	true value of quantile or `NULL` if no p-value should be calculated
`start:`	time of interim analysis (estimation of response rates is based only on data accruing after time `start` (passed on to `wkm`))

Value

A list containing the elements

`quantile`	the estimated quantile
`p.value`	the p-value for the null hypothesis of the true quantile being equal to `tau`
`conf.int`	the confidence interval for the quantile

WKM.COMPARE.QUANTILES

Description

Compare quantiles of two independent samples (ratio or difference) based on the weighted Kaplan-Meier estimator.

Usage

```
wkm.compare.quantiles(p, data, conf.level = 0.95, null.value = 1,
                      method = "ratio", p.value = FALSE, start=0)
```

Arguments

`p:`	number between 0 and 1 specifying the quantile

`data1`: a data frame or list containing the variables in model `cf1`

`data2`: a data frame or list containing the variables in model `cf2`

`conf.level`: confidence level at which to calculate the confidence interval

`null.value`: true value of quantile ratio or difference

`method`: either "ratio" or "difference"

`p.value`: if `TRUE` `p.value` will be calculated (requires `null.value`)

`start`: time of interim analysis (estimation of response rates is based only on data accruing after time `start` (passed on to `wkm`))

Value

A list containing the elements

`quantile1` the estimated quantile of the first sample

`quantile2` the estimated quantile of the second sample

`p.value` the p-value of the null hypothesis of equal quantiles

`conf.int` confidence interval for the difference / ratio of the quantiles

AHR

Description

Estimate average hazard ratios in a k-sample trial. See theorem 3.15.

Usage

```
ahr(L, formula, data, groups, strata = NULL, null.theta = NULL,
    contrast = NULL, start = 0, alpha = 1, multi.test = FALSE)
```

Arguments

`L`:	time-limit specifying time-interval $[0, L]$ over which average hazard ratios will be calculated
`formula`:	an object of class "formula" the conditional survival model
`data`:	data frame containing the variables in formula
`groups`:	a factor specifying the k groups (at least two)
`strata`:	a factor specifying the strata (if any) or `NULL`
`null.theta`:	vector specifying the null hypothesis for the average hazard ratios
`start`:	time of interim analysis (estimation of response rates is based only on data accruing after time `start` (passed on to `wkm`))
`alpha`:	parameter of the weight function
`multi.test`:	if `TRUE` calculate multivariate test statistic

Value

Returns an object of class "ahr", i.e. a list containing the elements.

`k`:	the number of groups, i.e. number of levels of the factor `Trt`
`n`:	vector of length `k` of sample sizes of each group
`p`:	vector of allocation ratios
`fit`:	list of "wkm" objects, one for each group
`times`:	vector of times at which each WKM estimator was evaluated
`n.times`:	length of `times`
`null.theta`:	vector of length `k` of null values of the average hazard ratios
`contrast`:	vector of length `k`
`groups`:	factor specifying the groups

`strata`: factor specifying the strata

`multi.test`: `TRUE` if multivariate test statistic has been calculated

`theta`: vector of length `k` of estimated average hazard ratios

`cov.theta`: (asymptotic) covariance matrix of vector `theta`

`Z.theta`: vector of standardized univariate test statistics for testing H_0: `theta[i] = null.theta[i]` for each $i = 1, \ldots, k$

`Z.multi`: standardized multivariate test statistic for testing H_0: `theta = null.theta` (only if `multi.test` is `TRUE`)

`Z.contrast`: standardized test statistic for testing H_0: `theta * contrast = 0` (only if `contrast` is not `NULL`)

AHR.BETA

Description

Estimate log average hazard ratios. See corollary 3.18.

Usage

```
ahr.beta(ahr.obj, null.beta = 0, contrast = NULL)
```

Arguments

`ahr.obj`: an object of class "ahr" as returned by `ahr`

`null.beta`: vector specifying the null hypothesis for the log average hazard ratios

`contrast`: vector of contrasts to test the null hypothesis `contrast * (beta - null.beta) = 0`

Value

A list containing:

`beta`: vector of estimated generalized log-hazard ratios

`cov.beta`: an estimated of the asymptotic covariance matrix

`Z.beta`: the standardized test statistic for testing beta=null.beta

`null.beta`: the argument `null.beta` passed to the function

RMEAN.DIFF.AHR

Description

Test difference of restricted means. See theorem 3.25.

Usage

```
rmean.diff.ahr(ahr.obj)
```

Arguments

`ahr.obj`: object of class "ahr" as returned by `ahr`

Value

A list containing:

`rmean.diff`: estimated restricted mean difference

`var.rmean.diff`: an estimate of the asymptotic variance of the restricted mean difference

`Z.rmean`: the standardized test statistic for testing `rmean.diff=0`

`p.value`: the p-value corresponding to `Z.rmean`

BIBLIOGRAPHY

P. K. Andersen. *Statistical models based on counting processes*. Springer, 1993.

P. K. Andersen and R. D. Gill. Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4):1100–1120, Dec. 1982.

H. Bang and A. A. Tsiatis. Estimating medical costs with censored data. *Biometrika*, 87(2):329–343, June 2000.

P. Bauer. Multistage testing with adaptive designs (with discussion). *Biometrie und Informatik in Medizin und Biologie*, 230:130–148, 1989.

P. Bauer and K. Köhne. Evaluation of experiments with adaptive interim analyses. *Biometrics*, 50(4):1029–1041, Dec. 1994.

P. Bauer and M. Posch. Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections by h. schf̈er and h.-h. müller, statistics in medicine 2001; 20: 3741-3751. *Statistics in Medicine*, 23(8):1333–1334, 2004.

P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, Sept. 2009.

W. Brannath, M. Posch, and P. Bauer. Recursive combination tests. *Journal of the American Statistical Association*, 97(457):236–244, 2002.

W. Brannath, F. König, and P. Bauer. Estimation in flexible two stage designs. *Statistics in Medicine*, 25(19):3366–3381, Oct. 2006.

W. Brannath, E. Zuber, M. Branson, F. Bretz, P. Gallo, M. Posch, and A. Racine-Poon. Confirmatory adaptive designs with bayesian decision tools for a targeted therapy in oncology. *Statistics in Medicine*, 28(10):1445–1463, May 2009.

F. Bretz, H. Schmidli, F. König, A. Racine, and W. Maurer. Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: General concepts. *Biometrical Journal*, 48(4):623–634, 2006.

R. Brookmeyer and J. Crowley. A confidence interval for the median survival time. *Biometrics*, 38(1):29–41, Mar. 1982.

B. F. Cole, R. D. Gelber, and A. Goldhirsch. Cox regression models for quality adjusted survival analysis. *Statistics in Medicine*, 12(10):975–987, 1993.

R. J. Cook and J. F. Lawless. Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine*, 16(8):911–924, Apr. 1997.

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

S. Datta, D. Bandyopadhyay, and G. A. Satten. Inverse probability of censoring weighted u-statistics for right-censored data with an application to testing hypotheses. *Scandinavian Journal of Statistics*, 37(4):680–700, 2010.

J. Fan and S. Datta. On Mann-Whitney tests for comparing sojourn time distributions when the transition times are right censored. *Annals of the Institute of Statistical Mathematics*, 65(1):149–166, Feb. 2013.

T. R. Fleming and D. P. Harrington. *Counting Processes and Survival Analysis*. John Wiley & Sons, Sept. 2011.

R. D. Gelber, R. S. Gelman, and A. Goldhirsch. A quality-of-life-oriented endpoint for comparing therapies. *Biometrics*, 45(3):781–795, Sept. 1989.

R. D. Gelber, B. F. Cole, S. Gelber, and A. Goldhirsch. Comparing treatments using quality-adjusted survival: The q-twist method. *The American Statistician*, 49(2):161–169, May 1995.

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.

P. P. Glasziou, B. F. Cole, R. D. Gelber, J. Hilden, and R. J. Simes. Quality adjusted survival analysis with repeated quality of life measures. *Statistics in Medicine*, 17 (11):1215–1229, 1998.

N. L. Hjort. On inference in parametric survival data models. *International Statistical Review / Revue Internationale de Statistique*, 60(3):355, Dec. 1992.

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, Dec. 1952.

Y. Huang. The Two-Sample problem with induced dependent censorship. *Biometrics*, 55(4):1108–1113, Dec. 1999.

Y. Huang and T. A. Louis. Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika*, 85(4):785–798, Dec. 1998.

M. Jenkins, A. Stone, and C. Jennison. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharmaceutical Statistics*, 10(4):347–356, 2011.

C. Jennison and B. W. Turnbull. *Group Sequential Methods with Applications to Clinical Trials*. CRC Press, Sept. 1999.

J. D. Kalbfleisch and R. L. Prentice. Estimation of the average hazard ratio. *Biometrika*, 68(1):105–112, Apr. 1981.

N. Keiding, T. Bayer, and S. Watt-Boolsen. Confirmatory analysis of survival data using left truncation of the life times of primary survivors. *Statistics in Medicine*, 6 (8):939–944, 1987.

K. Kombrink, A. Munk, and T. Friede. Design and semiparametric analysis of non-inferiority trials with active and placebo control for censored time-to-event data. *Statistics in Medicine*, 2013.

H. Koul, V. Susarla, and J. V. Ryzin. Regression analysis with randomly right-censored data. *The Annals of Statistics*, 9(6):1276–1288, Nov. 1981.

D. Y. Lin. Linear regression analysis of censored medical costs. *Biostatistics*, 1(1): 35–47, Mar. 2000.

D. Y. Lin. Regression analysis of incomplete medical cost data. *Statistics in Medicine*, 22(7):1181–1200, Apr. 2003.

D. Magirr, T. Jaki, F. Koenig, and M. Posch. Adaptive survival trials. *arXiv:1405.1569 [stat]*, May 2014. URL `http://arxiv.org/abs/1405.1569`.

H. M. Malani. A modification of the redistribution to the right algorithm using disease markers. *Biometrika*, 82(3):515–526, Sept. 1995.

M. Mielke, A. Munk, and A. Schacht. The assessment of non-inferiority in a gold standard design with censored, exponentially distributed endpoints. *Statistics in Medicine*, 27(25):5093–5110, 2008.

H.-H. Müller and H. Schäfer. A general statistical principle for changing a design any time during the course of a trial. *Statistics in Medicine*, 23(16):2497–2508, 2004.

S. Murray and A. A. Tsiatis. Nonparametric survival estimation using prognostic longitudinal covariates. *Biometrics*, 52(1):137–151, Mar. 1996.

S. Murray and A. A. Tsiatis. Sequential methods for comparing years of life saved in the two-sample censored data problem. *Biometrics*, 55(4):1085–1092, 1999.

P. C. O'Brien and T. R. Fleming. A multiple testing procedure for clinical trials. *Biometrics*, 35(3):549–556, Sept. 1979.

W. Pan and D. Zeng. Estimating mean cost using auxiliary covariates. *Biometrics*, 67 (3):996–1006, 2011.

M. S. Pepe and T. R. Fleming. Weighted kaplan-meier statistics: A class of distance tests for censored survival data. *Biometrics*, 45(2):497, June 1989.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL `http://www.R-project.org/`.

J. M. Robins and A. Rotnitzky. Recovery of information and adjustment for dependent censoring using surrogate markers. *Aids Epidemiology, Methodological issues*, pages 297–331, 1992.

H. Schäfer and H.-H. Müller. Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections. *Statistics in Medicine*, 20(24):3741–3751, 2001.

D. O. Scharfstein, A. A. Tsiatis, and J. M. Robins. Semiparametric efficiency and its implication on the design and analysis of group-sequential studies. *Journal of the American Statistical Association*, 92(440):1342–1350, 1997.

H. Schmidli, F. Bretz, A. Racine, and W. Maurer. Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: Applications and practical considerations. *Biometrical Journal*, 48(4):635–643, Aug. 2006.

T. Sellke and D. Siegmund. Sequential analysis of the proportional hazards model. *Biometrika*, 70(2):315 –326, 1983.

Y. Shen and T. R. Fleming. Large sample properties of some survival estimators in heterogeneous samples. *Journal of Statistical Planning and Inference*, 60(1):123–138, Apr. 1997.

R. L. Strawderman. Estimating the mean of an increasing stochastic process at a censored stopping time. *Journal of the American Statistical Association*, 95(452):1192–1208, 2000.

C. A. Struthers and J. D. Kalbfleisch. Misspecified proportional hazard models. *Biometrika*, 73(2):363–369, Aug. 1986.

J. Q. Su and L. J. Wei. Nonparametric estimation for the difference or ratio of median failure times. *Biometrics*, 49(2):603–607, June 1993.

A. A. Tsiatis, G. L. Rosner, and D. L. Tritchler. Group sequential tests with censored survival data adjusting for covariates. *Biometrika*, 72(2):365–373, 1985.

A. A. Tsiatis, H. Boucher, and K. Kim. Sequential methods for parametric survival models. *Biometrika*, 82(1):165–173, Mar. 1995.

A. v. d. Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, Mar. 1996.

M. J. van der Laan and A. Hubbard. Locally efficient estimation of the quality-adjusted lifetime distribution with right-censored data and covariates. *Biometrics*, 55(2):530–536, 1999.

A. van der Vaart. *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, June 2000.

G. Wassmer. Planning and analyzing adaptive group sequential survival trials. *Biometrical Journal*, 48(4):714–729, Aug. 2006.

M. Wolkewitz, A. Allignol, S. Harbarth, G. de Angelis, M. Schumacher, and J. Beyersmann. Time-dependent study entries and exposures in cohort studies can easily be sources of different and avoidable types of bias. *Journal of clinical epidemiology*, 65 (11):1171–1180, Nov. 2012.

H. Zhao and A. A. Tsiatis. A consistent estimator for the distribution of quality adjusted survival time. *Biometrika*, 84(2):339 –348, June 1997.

H. Zhao and A. A. Tsiatis. A consistent estimator for the distribution of quality adjusted survival time. *Biometrika*, 84(2):339–348, June 1997.

H. Zhao and A. A. Tsiatis. Efficient estimation of the distribution of quality-adjusted survival time. *Biometrics*, 55(4):1101–1107, Dec. 1999.

H. Zhao and A. A. Tsiatis. Estimating mean quality adjusted lifetime with censored data. *Sankhya: The Indian Journal of Statistics, Series B (1960-2002)*, 62(1):175–188, Apr. 2000.

H. Zhao and A. A. Tsiatis. Testing equality of survival functions of quality-adjusted lifetime. *Biometrics*, 57(3):861–867, Sept. 2001.

H. Zhao, C. Zuo, S. Chen, and H. Bang. Nonparametric inference for median costs with censored data. *Biometrics*, 68(3):717–725, 2012.

Die vorliegende veröffentlichte Version der Dissertation unterscheidet sich von der eingereichten Dissertation an folgenden Stellen:

Seite 26:
For $\alpha = 1$, this leads to the relative risk

$$\theta_i(G) = P(\min\{T_0, \ldots, T_{i-1}, T_{i+1}, \ldots, T_k\} > T_i)$$

(see lemma 3.6).

Seite 30, Definition 3.11:

$$\hat{\theta}_i = \frac{\hat{\kappa}_i}{1 - \hat{G}(L)}$$

Seite 80:
Table 4.1 shows the type-I-error of the AHR(β), RMS and LR methods for different maximum number of interim analyses under the null hypothesis of no treatment effect.

Seite 81, Tabelle 4.1: One-sided type-I-error of the AHR(β), RMS and LR methods for different maximum number of interim analyses.

Seite 96, Lemma 5.4:

$$\hat{\rho}_j(t_{l-1}, t_l, s) = \frac{n}{\hat{p}_j(t_{l-1}, t_l)} \int_0^s J_j^{(n)}(t_{l-1}, t_l, u) \frac{\bar{N}_j^{(n)}(t_l, du)}{\bar{Y}_j^{(n)}(t_{l-1}, t_l, u)^2}.$$

gez. Prof. Dr. Werner Brannath (Vorsitzender des Prüfungsausschusses)