

Adaptive group sequential survival comparisons based on log-rank and pointwise test statistics

Statistical Methods in Medical Research

2021, Vol. 30(12) 2562–2581

© The Author(s) 2021



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09622802211043262

journals.sagepub.com/home/smm



Jannik Feld , Andreas Faldum  and Rene Schmidt

Abstract

Whereas the theory of confirmatory adaptive designs is well understood for uncensored data, implementation of adaptive designs in the context of survival trials remains challenging. Commonly used adaptive survival tests are based on the independent increments structure of the log-rank statistic. This implies some relevant limitations: On the one hand, essentially only the interim log-rank statistic may be used for design modifications (such as data-dependent sample size recalculation). Furthermore, the treatment arm allocation ratio in these classical methods is assumed to be constant throughout the trial period. Here, we propose an extension of the independent increments approach to adaptive survival tests that addresses some of these limitations. We present a confirmatory adaptive two-sample log-rank test that allows rejection regions and sample size recalculation rules to be based not only on the interim log-rank statistic, but also on point-wise survival rate estimates, simultaneously. In addition, the possibility is opened to adapt the treatment arm allocation ratio after each interim analysis in a data-dependent way. The ability to include point-wise survival rate estimators in the rejection region of a test for comparing survival curves might be attractive, e.g., for seamless phase II/III designs. Data-dependent adaptation of the allocation ratio could be helpful in multi-arm trials in order to successively steer recruitment into the study arms with the greatest chances of success. The methodology is motivated by the LOGGIC Europe Trial from pediatric oncology. Distributional properties are derived using martingale techniques in the large sample limit. Small sample properties are studied by simulation.

Keywords

Adaptive design, phase II trial, phase III trial, sample size recalculation, survival analysis, log-rank, Nelson–Aalen, bivariate, seamless design

Introduction

The log-rank test¹ is presently the gold standard method for analysing differences in survival data in randomised clinical trials. For this reason adaptive survival tests are commonly based upon the log-rank test statistic and its independent increments structure.^{2,3} However, these designs suffer from some limitations we want to address. One limitation is that effectively only the interim log-rank statistic may be used for design modifications (such as data-dependent sample size recalculation).⁴ Moreover, the treatment arm allocation ratio in these classical methods is assumed to be constant throughout the whole trial period. However, in the context of seamless phase II/III designs or early phase trials it may be desirable to include point-wise survival rates (e.g. 1 year survival rates) in the decision making, since survival rates at a given time-point of interest are chosen as a primary endpoint regularly in such trials. Likewise, data-dependent adaptations of the treatment arm allocation ratio could be helpful in multi-arm trials in order to successively steer recruitment into the study arms with the greatest chances of success. Therefore we propose an extension of the independent increments approach to adaptive survival tests, which can rely on both: (i)

Institute of Biostatistics and Clinical Research, University of Münster, Muenster, Germany

Corresponding author:

Jannik Feld, Institute of Biostatistics and Clinical Research, University of Münster, Schmeddingstraße 56, 48149 Münster, DE.

Email: Jannik.Feld@ukmuenster.de

the pointwise Nelson Aalen type survival rates estimator and (ii) the log rank test statistic. More specifically our approach extends the commonly used methodology by Wassmer,³ which neither supports the use of point-wise survival rate estimates nor foresees data-dependent adaptations of the treatment arm allocation ratio. In doing so, our approach avoids those difficulties associated with alternative methods based on the patient-wise separation principle, which have the common disadvantage that the test procedure may either neglect part of the observed survival data or tend to be conservative. We will show by simulation that our extended methodology maintains the performance of the current standard methodology while offering various new design possibilities.

The methodology presented here is motivated by the LOGGIC Europe trial (Eudra-CT: 2018-000636-10). LOGGIC Europe is a randomized, international multicentre phase III therapy optimization trial for children and adolescents with low-grade glioma. Primary endpoints of the trial are the progression-free survival (PFS) and the disease control rate (DCR). PFS addresses long-term efficacy of treatment and is defined as time from randomization up to progression of disease or death for all reasons whatever occurs first. DCR addresses short-term efficacy of treatment and is essentially defined as the PFS-rate at some early timepoint.

The paper is organized as follows. We start by settling notation and stochastic assumptions. Section ‘Joint martingale representation of the log-rank statistic and cumulative hazard difference’ presents briefly the bivariate representation of the two test statistics and its distributional properties. The design algorithm and corresponding planning methodology is presented in section ‘Adaptive log-rank test with simultaneous use of interim log-rank statistic and cumulative hazard rate difference’. In section ‘Example: A two-step log-rank test with futility criterion based on short-term survival rate’ we present some example use-case in order to illustrate practical implementation of our method. Small sample properties are studied by simulation in section ‘Simulation’. We conclude with a discussion of our findings and prospects for future research. Mathematical proofs are shifted to the supplemental material.

Notation and stochastic assumptions

Let (Ω, \mathcal{F}, P) denote the probability space upon which all random variables are defined. Unless otherwise specified, random variables are denoted by capital Latin letters, whereas realizations of random variables are denoted by the corresponding lower case Latin letters. We set $0/0 := 0$ whenever formal division of zero by zero occurs in sequel.

We consider the problem of testing the equality of survival distributions for two treatments A and B , say, based on accumulating survival data across several stages of a sequential design. After each stage a confirmatory (interim) analysis is performed with the possibility for interim decisions (e.g. binding futility stop or sample size recalculation) based on (i) the observed interim log-rank statistic and (ii) interim estimates of s_0 -years survival rate differences for some prefixed time-point $s_0 > 0$.

In this context we will assume an initial trial design with l stages. The stages will recruit patients successively, i.e. patients from stage k are recruited between calendar times $\sum_{i=1}^{k-1} a_i$ and $\sum_{i=1}^k a_i$ where $a_i > 0$ are the recruitment period lengths of the stages. We set $a := \sum_{i=1}^l a_i$ as the overall recruitment period length. The final analysis will be performed at calendar time $a + f$. Patients from stage k will therefore have at least a follow-up period length of $f_k = \sum_{i=k+1}^l a_i + f$. An example timeline for $l = 2$ stages is given in Figure 1. The planned annual recruitment rate is denoted with r .

For this purpose, let $\mathcal{N}_{x,k}$ denote the set of patients from treatment group $x = A, B$, who entered the trial at stage k (i.e. between calendar time $\sum_{i=1}^{k-1} a_i$ and $\sum_{i=1}^k a_i$), and let $n_{x,k} := \#\mathcal{N}_{x,k}$ denote the number of such patients. Let $\mathcal{N}_k := \bigcup_x \mathcal{N}_{x,k}$ denote the set of all patients from stage k pooled over both treatment groups, and $\mathcal{N} := \bigcup_{x,k} \mathcal{N}_{x,k}$ the overall set of trial patients. Let $n_k := \sum_x n_{x,k}$ and $n := \sum_{x,k} n_{x,k}$. The parameter n will index the arrival process and asymptotic results will be derived in the limit $n \rightarrow \infty$. Accordingly, we assume that group sizes grow uniformly as total sample size increases, i.e. we assume there exist constants $v_k > 0$ such that $\#\mathcal{N}_{A,k}/n_k \rightarrow \frac{1}{1+v_k}$ and $\#\mathcal{N}_{B,k}/n_k \rightarrow \frac{v_k}{1+v_k}$ in probability as $n \rightarrow \infty$. Thus the constants v_k are the asymptotic, stagewise allocation ratio between the treatment groups. We furthermore assume that the stages also grow uniformly as total sample size increases, i.e. $\#\mathcal{N}_k/n = n_k/n \rightarrow a_k/a$ in probability as $n \rightarrow \infty$.

To patient i is associated a random triplet $\{E_i, C_i, T_i\}$. E_i is the entry time into the study, the possibly infinite random variable C_i is the time of censoring after entry, and T_i is the survival time after entry. Our stochastic assumptions are as follows: (1) T_i, C_i and E_i are mutually independent for fixed i , and (2) data from different patients are independent and identical distributed within treatment groups.

Based on the observed data, we will calculate the *number of events* in stage k from treatment group $x = A, B$ up to study time $s \geq 0$ as

$$D_{x,k}(s) := \sum_{i \in \mathcal{N}_{x,k}} D_i(s), \quad D_i(s) := I(T_i \leq s, T_i \leq C_i), \tag{1}$$

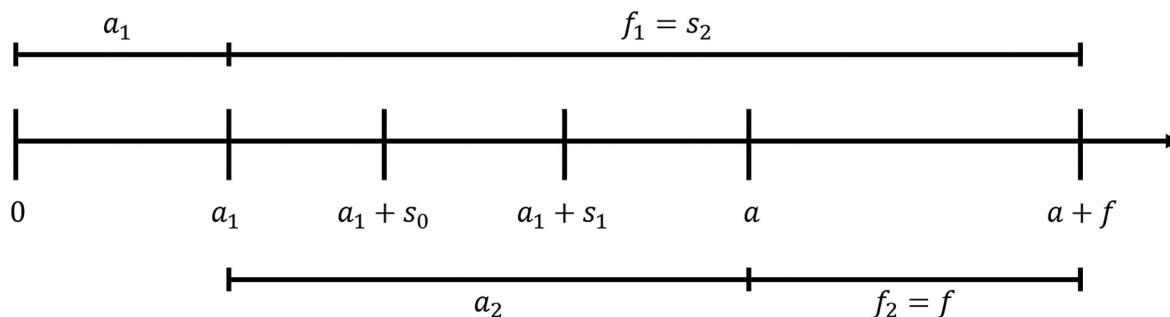


Figure 1. Initial time schedule. At time of the final analysis, first stage patients would have a minimum follow-up of $f_1 = a_2 + f > s_1$ years under the initial time schedule. Second stage patients would have a minimum follow-up of $f_2 = f$ at time of the final analysis.

and the number at risk by study time $s \geq 0$ in stage k and treatment group $x = A, B$ as

$$Y_{x,k}(s) := \sum_{i \in \mathcal{N}_{x,k}} I(T_i \wedge C_i \geq s). \tag{2}$$

Finally, let $J_{x,k}(s) := I(Y_{x,k}(s) > 0)$ and $L_k(s)$ the log-rank weight factor

$$L_k(s) := \frac{Y_{A,k}(s) \cdot Y_{B,k}(s)}{Y_{A,k}(s) + Y_{B,k}(s)}. \tag{3}$$

For each $s \geq 0$, let \mathcal{F}_s be the σ -algebra generated by

$$\begin{aligned} &I\{T_i \leq s \wedge C_i\}, \quad T_i \cdot I\{T_i \leq s \wedge C_i\}, \\ &I\{C_i \leq s \wedge T_i\}, \quad C_i \cdot I\{C_i \leq s \wedge T_i\}, \end{aligned} \tag{4}$$

for $i \in \mathcal{N}$. We consider $D_{x,k}, Y_{x,k}, J_{x,k}, L_k$ as stochastic process in study time $s \geq 0$, adapted to the filtration $(\mathcal{F}_s)_{s \geq 0}$. The filtration $(\mathcal{F}_s)_{s \geq 0}$ comprises the information that is observed in the study. Whenever we want to emphasize the dependence of above processes on $n \in \mathbb{N}$, we will index them additionally by n e.g. $D_{x,k}^n$ instead of $D_{x,k}$.

As usual, we let $\lambda_x(s) := \lim_{\delta \rightarrow 0} P(s \leq T_i < s + \delta | T_i \geq s) / \delta$ denote the hazard of a patient i from treatment group $x = A, B$. We denote by $\Lambda_x(s) := \int_0^s \lambda_x(u) du$ and $S_x(s) := \exp(-\Lambda_x(s)) \equiv P(T_i > s)$ the corresponding cumulative hazard and survival functions for treatment group $x = A, B$, respectively.

In this context, we consider testing the two-sided null hypothesis

$$H_0 : S_A(s) = S_B(s) \quad \text{for all } 0 \leq s \leq s_{\max} \tag{5}$$

that the survival functions in the two treatment arms coincide within some prefixed interval $[0, s_{\max}]$.

We proceed as follows to test H_0 . Using martingale techniques, we will first derive the joint distribution of (i) the stage-wise log-rank test statistics and (ii) the stage-wise difference in the Nelsen–Aalen estimates between the two treatment arms evaluated at some prefixed study time s_0 . On this basis, we provide a confirmatory adaptive two-sample log-rank test where provision is made for interim decision making and design modifications based on both (i) the interim log-rank statistic and (ii) interim estimates of the cumulative hazard rate differences at timepoint s_0 . With a view to practical application, sample size recalculation is one of the most common design modifications. Therefore, sample size recalculation based on conditional power will be elaborated and studied in detail, analytically and by simulation.

Joint martingale representation of the log-rank statistic and cumulative hazard difference

The weighted two-sample log-rank statistic in stage k is defined as

$$LR_k(s) := \sum_{i \in \mathcal{N}_{A,k} D_i(s)=1} \frac{L_k(T_i)}{Y_{A,k}(T_i)} - \sum_{i \in \mathcal{N}_{B,k} D_i(s)=1} \frac{L_k(T_i)}{Y_{B,k}(T_i)}, \tag{6}$$

where L_k is the weight from equation (3). The difference of the group-wise Nelson–Aalen estimates in stage k is given as

$$\Delta_k(s) := \sum_{i \in \mathcal{N}_{A,k} D_i(s)=1} \frac{J_{A,k}(T_i)}{Y_{A,k}(T_i)} - \sum_{i \in \mathcal{N}_{B,k} D_i(s)=1} \frac{J_{B,k}(T_i)}{Y_{B,k}(T_i)}, \tag{7}$$

which are both \mathcal{F}_s -adapted processes. It follows from theorem A2, that under mild regularity assumptions and the proportional hazards assumption $\lambda_B = \omega \lambda_A$ for some $\omega > 0$, the following distributional approximation holds:

$$\begin{pmatrix} \frac{n_k^{-1/2} L_{R,k}(s)}{\sqrt{[M_k^{LR}](s)}} \\ \frac{n_k^{1/2} \Delta_k(s)}{\sqrt{[M_k^\Delta](s)}} \end{pmatrix} D \approx \mathcal{N} \left(\begin{pmatrix} -\sqrt{n_k} \log(\omega) \cdot \sigma_{LR,k}(s) \\ -\sqrt{n_k} \log(\omega) \cdot \frac{\Lambda_A(s)}{\sigma_{\Delta,k}(s)} \end{pmatrix}, \begin{pmatrix} 1 & \frac{\Lambda_A(s)}{\sigma_{LR,k}(s)\sigma_{\Delta,k}(s)} \\ \frac{\Lambda_A(s)}{\sigma_{LR,k}(s)\sigma_{\Delta,k}(s)} & 1 \end{pmatrix} \right), \tag{8}$$

where $\sigma_{LR,k}^2(s) := \text{plim}_{n_k \rightarrow \infty} D_k(s)/n_k \cdot \frac{v_k}{(1-v_k)^2}$ and $\sigma_{\Delta,k}^2(s)$ are some deterministic functions (see equations (14) and (13) below) and

$$\begin{aligned} [M_k^{LR}](s) &:= n_k^{-1} \sum_{i \in \mathcal{N}_{A,k} D_i(s)=1} \frac{L_k(T_i)^2}{Y_{A,k}(T_i)^2} + n_k^{-1} \sum_{i \in \mathcal{N}_{B,k} D_i(s)=1} \frac{L_k(T_i)^2}{Y_{B,k}(T_i)^2}, \\ [M_k^\Delta](s) &:= n_k \sum_{i \in \mathcal{N}_{A,k} D_i(s)=1} \frac{J_{A,k}(T_i)}{Y_{A,k}(T_i)^2} + n_k \sum_{i \in \mathcal{N}_{B,k} D_i(s)=1} \frac{J_{B,k}(T_i)}{Y_{B,k}(T_i)^2}. \end{aligned} \tag{9}$$

The left hand side of (8) has also approximately independent, bivariate normal distributed increments as stated in theorem A2.

For given $\omega > 0$ we set

$$\mu_k := -\sqrt{n_k} \log(\omega). \tag{10}$$

In practice the time-dependant correlation parameter on the right hand side of (8) is unknown. However, for a fixed time point $s_0 > 0$ it can be consistently estimated at time of the interim analysis (see (24)). Under further planing assumptions it is possible to deduce closed formulas for the functions $\sigma_{LR,k}$ and $\sigma_{\Delta,k}$. Assuming (in addition to above mentioned mild regularity conditions of theorem A2):

- No loss to follow-up:

$$\forall i \in \mathcal{N}_k : E_i + C_i \equiv a + f \tag{11}$$

- Uniform recruitment:

$$\forall i \in \mathcal{N}_k : E_i \sim \mathcal{U} \left(\left[\sum_{i=1}^{k-1} a_i, \sum_{i=1}^k a_i \right] \right) \tag{12}$$

the following two equations hold (see appendix for proofs):

$$\sigma_{\Delta,k}^2(s) = \begin{cases} \frac{(1+v_k)^2}{v_k} \left(\frac{1}{S_A(s)} - 1 \right), & \text{if } s \leq f_k \\ \frac{(1+v_k)^2}{v_k} \left(\frac{1}{S_A(f_k)} - 1 + \int_{f_k}^s \frac{\lambda_A(u) \cdot a_k}{S_A(u) \cdot (a_k + f_k - u)} du \right), & \text{if } f_k < s < a_k + f_k \end{cases} \tag{13}$$

$$\sigma_{LR,k}^2(s) = \begin{cases} \frac{v_k}{(1+v_k)^2} (1 - S_A(s)), & \text{if } s \leq f_k \\ \frac{v_k}{(1+v_k)^2} \left(1 - \frac{a_k + f_k - s}{a_k} S_A(s) - \frac{1}{a_k} \int_{f_k}^s S_A(u) du \right), & \text{if } f_k < s < a_k + f_k \\ \frac{v_k}{(1+v_k)^2} \left(1 - \frac{1}{a_k} \int_{f_k}^{a_k + f_k} S_A(u) du \right), & \text{if } s \geq a_k + f_k \end{cases} \tag{14}$$

Adaptive log-rank test with simultaneous use of interim log-rank statistic and cumulative hazard rate difference

The design algorithm

For the sake of notational simplicity we will focus on two-step designs in the sequel (i.e. $l = 2$). The two-step adaptive design will proceed as follows: Assume an initial design with accrual of patients between calendar time 0 and a years,

and a final analysis at calendar time $a + f$ (corresponding to minimum follow-up period of f years). We assume that the value of f is prefixed by clinical consideration. Choice of a will be detailed in section 'Initial sample size calculation' based on power arguments. Patients recruited prior to calendar time $a > a_1 > 0$ define the set of first stage patients \mathcal{N}_1 , and patients recruited between calendar time a_1 and $a := a_1 + a_2$ define the set of second stage patients \mathcal{N}_2 . The interim analysis will take place at time $a_1 + s_1$ for some $0 < s_1 < a_2$ and will include the patients of stage one with their first s_1 years of follow-up.

At the interim analysis the log-rank statistic in stage 1 patients based on information up to study time s_1

$$Z_1^* := Z_{11} := n_1^{-1/2} \cdot \frac{LR_1(s_1)}{\sqrt{[M_1^{LR}](s_1)}} \quad (15)$$

and the standardized cumulative hazard rate difference at some prefixed (early) study time $0 < s_0 \leq s_1$

$$B_1 := n_1^{1/2} \cdot \frac{\Delta_1(s_0)}{\sqrt{[M_1^\Delta](s_0)}} \quad (16)$$

will be calculated. B_1 is an interim estimate of the difference in short-term response. More specifically, $\Delta_1(s_0)$ is an interim estimate of $\log(S_B(s_0)/S_A(s_0))$. The design algorithm is as follows: The design stops at the interim analysis with rejection of H_0 if the observed value z_1 for Z_1^* exceeds some critical value u_1 . The design stops for futility if either z_1 falls below some futility bound u_0 or if the observed value b_1 for B_1 drops below some prefixed boundary b_0 . Otherwise, if $u_0 \leq z_1 < u_1$ and $B_1 > b_0$, the design continues to stage two. At this time, the recruitment period length of stage two a_2 can be data-dependent recalculated. The recalculated recruitment period length $a'_2 := a'_2(Z_1^*, B_1)$ of stage two is chosen in dependence of the observed values for Z_1^* and B_1 subject to the constraint $s_1 < a'_2 \leq a_{\max} - a_1$. Here, $a_{\max} > a_1$ denotes a maximum trial recruitment period length that is fixed in advance in order to avoid unrealistic or unfeasible trial duration. We set $a' := a_1 + a'_2$ and $f'_1 := a'_2 + f$.

The final analysis will take place at calendar time $a' + f$ and will include both, patients of stage one \mathcal{N}_1 with their full follow-up data of at least f'_1 years and the set of second stage patients \mathcal{N}_2 with their follow-up time of at least f years. At the time of the final analysis, the increment of the log-rank statistic in stage one patients beyond study time s_1 will be calculated

$$Z_{12} := n_1^{-1/2} \cdot \frac{LR_1(a' + f) - LR_1(s_1)}{\sqrt{[M_1^{LR}](a' + f) - [M_1^{LR}](s_1)}}, \quad (17)$$

as well as the log-rank statistic of stage two patients

$$Z_{22} := n_2^{-1/2} \cdot \frac{LR_2(f'_1)}{\sqrt{[M_2^{LR}](f'_1)}}. \quad (18)$$

Notice that Z_{12} and Z_{22} are conditionally independent given Z_1^* and B_1 .

The null hypothesis H_0 will be rejected at the final analysis if the second stage test statistic

$$Z_2^* := \frac{\sqrt{\eta_{11}} \cdot Z_{11} + \sqrt{\eta_{12} - \eta_{11}} \cdot Z_{12} + \sqrt{\eta_{22}} \cdot Z_{22}}{\sqrt{\eta_{12} + \eta_{22}}} \quad (19)$$

exceeds some critical value u_2 , where the prefixed weight factors

$$\eta_{11} = \sigma_{LR,1}^2(s_1), \quad \eta_{12} = \sigma_{LR,1}^2(a_1 + f_1), \quad \eta_{22} = \sigma_{LR,2}^2(f_1), \quad (20)$$

amount to the expected variance of the log-rank statistics under some initial planning alternative K_1 (see section 'Calculation of the critical bounds'). Their values are given in (13) and (14). The weights η_{ij} have to be fixed in advance and remain unchanged while the trial is ongoing.

The rejection region

The design algorithm described in section 4.1 corresponds to the rejection region

$$\mathcal{R} := \{Z_1^* \geq u_1\} \cup \{B_1 > b_0, u_0 \leq Z_1^* < u_1, Z_2^* \geq u_2\} \quad (21)$$

of the null hypothesis H_0 . It is crucial that the design parameters $b_0, \eta_{11}, \eta_{12}, \eta_{22}$ and $0 < s_0 \leq s_1 < f$ as well as the critical bounds u_0, u_1 are prefixed and remain unchanged during the trial. Note that the critical bound u_2 will be calculated at the interim analysis according to formula 22 when the correlation ρ of Z_1 and B_1 can be estimated to obtain a rejection region which exhausts the full significance level. The calculation of the critical bounds b_0, u_0, u_1, u_2 is elaborated next.

Calculation of the critical bounds

The rejection region \mathcal{R} defines a level α test of the null hypothesis H_0 if the critical bounds b_0, u_0, u_1, u_2 are chosen according to the proviso that $P_{H_0}(\mathcal{R}) = \alpha$, i.e. such that

$$\alpha = \Phi(-u_1) + \int_{u_0}^{u_1} \Phi\left(\frac{\sqrt{\eta_{11}}z_1 - \sqrt{\eta_{12} + \eta_{22}u_2}}{\sqrt{\eta_{12} - \eta_{11} + \eta_{22}}}\right) \Phi\left(\frac{\rho z_1 - b_0}{\sqrt{1 - \rho^2}}\right) \phi(z_1) dz_1. \tag{22}$$

Notice that the critical bounds depend on the nuisance parameter

$$\rho = \frac{\Lambda_A(s_0)}{\sigma_{\Delta,1}(s_0)\sigma_{LR,1}(s_1)}, \tag{23}$$

which is in fact unknown during the trial if one does not know the true hazard function λ_A . However, it may be estimated consistently at time of the interim analysis via

$$\hat{\rho} := \frac{[M_1^{LR}, M_1^\Delta](s_0)}{\sqrt{[M_1^\Delta](s_0) \cdot [M_1^{LR}](s_1)}}. \tag{24}$$

Nevertheless there are infinite parameter combinations of the critical bounds which satisfy equation (22). It is therefore crucial, that one parameter constellation (u_0, b_0, u_1) is chosen in advance and remains unchanged during the trial. The critical bound u_2 will then be calculated at the interim analysis as the unique solution to (22) with $\hat{\rho}$ plugged in for ρ .

Initial sample size calculation

Initial sample size calculation is performed under the planning alternative hypothesis

$$K_1 : S_B = S_A^{a_0} \tag{25}$$

and under the assumption, that no sample size recalculation is performed i.e. $a'_2 = a_2$. For the initial sample size calculation we need to fix the proportion $\pi := a_1/a \in (0, 1)$ of accrual to stage 1. Note that the weights η_{ij} are fixed in advance and must not be changed while the trial is ongoing. In fact they have to be calculated simultaneously with the sample size. For given weight factors η_{ij} , the condition to reject null hypothesis H_0 with probability $1 - \beta$ under planning alternative K_1 is $P_{K_1}(\mathcal{R}) = 1 - \beta$. Using the distribution approximation (8) this proviso is tantamount to

$$1 - \beta = P_{K_1}(Z_1^* \geq u_1) + P_{K_1}(B_1 > b_0, u_0 \leq Z_1^* < u_1, Z_2^* \geq u_2). \tag{26}$$

Notice that B_1 and Z_2^* are independent given Z_1^* . Thus the right hand side of (26) equals

$$\Phi(\mu_1\sigma_{LR,1}(s_1) - u_1) + \int_{u_0}^{u_1} \phi(z_1 - \mu_1\sigma_{LR,1}(s_1)) \cdot P_{K_1}(B_1 > b_0 | Z_1^* = z_1) \cdot P_{K_1}(Z_2^* \geq u_2 | Z_1^* = z_1) dz_1. \tag{27}$$

Using again the distribution approximation (8) we get the identities

$$P_{K_1}(B_1 > b_0 | Z_1^* = z_1) = \Phi\left(\frac{b_0 - \mu_k \frac{\Lambda_A(s_0)}{\sigma_{\Delta,1}(s_0)} - \rho(z_1 - \mu_1\sigma_{LR,1}(s_1))}{\sqrt{1 - \rho^2}}\right) \tag{28}$$

and

$$\begin{aligned}
 &P_{K_1}(Z_2^* \geq u_2 | Z_1^* = z_1) \\
 &= \Phi \left(- \frac{u_2 \cdot \sqrt{\eta_{12} + \eta_{22}} - \sqrt{\eta_{11}}z_1 - \sqrt{\eta_{12} - \eta_{11}} \cdot (\mu_1 \sqrt{\sigma_{LR,1}^2(a_1 + f_1) - \sigma_{LR,1}^2(s_1)} - \sqrt{\eta_{22}}\mu_2\sigma_{LR,2}(f_1))}{\sqrt{\eta_{12} - \eta_{11} + \eta_{22}}} \right) \quad (29)
 \end{aligned}$$

Using the identities (28) and (29), formulas (13) and (14) for $\sigma_{LR,k}(s)$ and $\sigma_{\Delta,k}(s)$ and the identity $n_k = r \cdot a_k$ in equation (26), one can solve (26) and (22) numerically to obtain the critical bound u_2 and the needed recruitment period lengths a_1 and a_2 . We provide R syntax in the supplemental material to do so.

At the interim analysis, a_2 and thus n_2 may be modified in a data-dependent way to keep up adequate power performance of the trial, as will be detailed in the next section.

Data-dependent sample size recalculation at the interim analysis based on conditional power

At the interim analysis, we are free to revise the length of the stage two accrual period a_2 in the light of Z_1^* (interim log-rank statistic) and B_1 (observed difference in short-term response) without compromising type I error rate control. This is a consequence of the independent increments structure of the bivariate process given by the left hand side of (8). For this purpose, we will first calculate the required length of the accrual period a_2^{CP} to achieve a desired conditional power. To avoid unrealistic large trial duration, the revised length of the accrual period will finally be chosen as

$$a'_2 := \max(\min(a_2^{CP}, a_{\max} - a_1), s_1). \quad (30)$$

Recall that $a_1 + s_1$ is the calendar time of the interim analysis and a_{\max} is a prefixed maximum trial recruitment period length.

Likewise, we are free to revise the allocation ratio between treatment groups in the light of Z_1^* and B_1 . Let v'_2 denote the revised allocation ratio of stage two patients to treatment group B as referred to treatment group A. Furthermore we may use an updated recruitment rate r' to adjust for new experience.

To calculate a_2^{CP} , we estimate the true hazard ratio ω via

$$\hat{\omega} := \exp\left(-\frac{1}{n} \frac{LR_1(s_1)}{[M_1^{LR}](s_1)}\right). \quad (31)$$

Notice that $LR_1(s_1)$ and $N_1(s_1)$ are observed at the interim analysis. We can also estimate $\sigma_{LR,1}(s_1)$ consistently at the interim analysis through the estimator $\hat{\sigma}_{LR,1}^2(s_1) := [M_1^{LR}](s_1)$. Sample size recalculation will be performed under the revised planning alternative

$$K'_1 : S_B = \hat{S}_A^\omega \quad (32)$$

suggested by the observed interim estimate $\hat{\omega}$ of the true hazard ratio. The condition to achieve a conditional power $P_{K'_1}(Z_2^* \geq u_2 | Z_1^* = z_1, B_1 = b_1)$ of $1 - \beta_2$ under the revised planning alternative K'_1 is equivalent to

$$\begin{aligned}
 &\Phi^{-1}(\beta_2) = \\
 &\frac{u_2 \sqrt{\eta_{12} + \eta_{22}} - \sqrt{\eta_{11}}z_1 - \sqrt{\eta_{12} - \eta_{11}} \cdot \hat{\mu}_1 \sqrt{\sigma_{LR,1}^2(a_1 + a_2^{CP} + f) - \hat{\sigma}_{LR,1}^2(s_1)} - \sqrt{\eta_{22}} \cdot \hat{\mu}_2 \sigma_{LR,2}(a_2^{CP} + f)}{\sqrt{\eta_{12} - \eta_{11} + \eta_{22}}}, \quad (33)
 \end{aligned}$$

where $\hat{\mu}_k = -\sqrt{n_k} \cdot \log(\hat{\omega})$ is the estimated drift. Plugging in the identities $n_2 = a_2^{CP} \cdot r'$ and the formulas for $\sigma_{LR,2}(a_2^{CP} + f)$ and $\sigma_{LR,1}(a_1 + a_2^{CP} + f)$ given by (14) with updated values $v_2 \rightarrow v'_2$ and $r \rightarrow r'$, we can solve above equation (33) to obtain a_2^{CP} . Note that the equation can not be solved if $\hat{\omega} \geq 1$ holds. In this case we define $a_2^{CP} := \infty$.

The revised length of accrual a'_2 is finally chosen according to (30). We will provide R syntax to do so in the supplementary material.

Example: A two-step log-rank test with futility criterion based on short-term survival rate

In this section, we illustrate application of our methodology using the example of a two-step log-rank test with binding futility criterion based on a short-term survival rates and sample size recalculation based on conditional power. Recall that the underlying null hypothesis is $H_0 : S_A(s) = S_B(s)$ for all $0 \leq s \leq s_{\max}$ for some prefixed $s_{\max} > 0$. The underlying physical units of s will be "years".

In general, our two-step test of H_0 depends on a set of design parameters that have to be fixed in advance:

- (a) parameters b_0, u_0, u_1, u_2 defining the rejection region acc. to (22),
- (b) parameters s_0 and s_1 steering the amount of follow-up included into interim decision making,
- (c) parameters a_1, a_2, f defining the initially planned lengths of stage one accrual, stage two accrual, and follow-up period
- (d) parameters r, v_1, v_2 defining the initial accrual rate, and stage-wise treatment arm allocation ratios
- (e) weights η_{11}, η_{12} and η_{22} of the stage-wise log-rank increments acc. to (20) and (14).

More specifically, let us assume that we aim for a two-step, Pocock-type log-rank test of H_0 with binding stopping for futility if the observed 6 months survival rate in the experimental arm is worse than in the standard arm. This futility condition is realized by choosing $b_0 = 0, u_0 = -\infty$, and $s_0 = 0.5$. The Pocock condition means choosing $u_1 = u_2$.⁵ Note that an uncountable number of alternative functional relationships between u_1 and u_2 could have been chosen. The difference $s_1 - s_0$ is the interval between the time when the short-term endpoint B_1 becomes known and the date of the interim analysis. For practical reasons, $s_1 - s_0 \geq 0$ should not be chosen too large. On the other hand, s_1 should be sufficiently large such that the interim log-rank statistic Z_1 is informative. In our exemplary setting, we consider $s_1 = 1$ as sensible choice. The parameters f and r are determined by the clinical frame conditions. Let us assume a desired follow-up period of $f = 2$ years, and an annual overall accrual rate of $r = 75$. Also assume that we aim for equal randomization to both arm (i.e. $v_1 = v_2 = 1$) as well as an interim analysis after half of the planned overall accrual period, i.e. $\pi = a_1/(a_1 + a_2) = 0.5$. Finally, assume that we set a significance level of 5%, that we aim for a power of 80% if the true hazard ratio ω_0 equals $2/3$ (planning alternative hypothesis), and that there are exponentially distributed survival times with scale parameter of $\lambda = 1$ to a good approximation in the standard therapy arm.

With these specifications, the parameters u_1 and a_1 remain as the only unknown from the parameters listed under a)-d). Whereas the weight η_{11} is also fixed by above specifications, the weights η_{12} and η_{22} remain as functions in a_1 acc. to equation (20), since $s_1 = 1, a_1 + f_1 = 2 \cdot a_1 + f, f_1 = a_1 + f$. We are now in a position to determine the rejection region (see Section 'The rejection region') and to perform the initial sample-size calculation (see Section 'Calculation of the critical bounds'). Using $b_0 = 0, u_0 = -\infty, u_1 = u_2$ and ρ acc to (8) the equations (22) and (26) may be solved simultaneously for the two remaining free parameters u_1 and a_1 . Doing so, yields a stage-one recruitment period length of $a_1 = 1.7$ years (corresponding to $n_1 = r \cdot a_1 = 125$ patients), together with a stage-one critical boundary $u_1 = 2.18$. On this basis, the weights may be calculated as $\eta_{11} = 0.158, \eta_{12} = 0.247, \eta_{22} = 0.233$ using (20) and (14). To ensure that the rejection region does not depend on our initial planning assumptions regarding ρ , the value of the critical bound u_2 will be updated and ultimately fixed as described below at the time of the interim analysis, when an estimate of ρ becomes available.

After $1.7 + 0.5 = 2.3$ years, instead of B_1 can be evaluated. Assume that we find a value of instead of $B_1 = 1.08 > 0 = b_0$, which concludes that the trial can continue (no stopping for futility). After $1.7 + 1.0 = 2.7$ years, the interim log-rank statistic Z_{11} becomes known and the interim analysis has to be performed. Let us assume that a test statistic of $Z_{11} = 1.34 < 2.18 = u_1$ is observed as well as an estimated hazard ratio of $\hat{\omega} = 0.731$. In this case the trial continuous to stage two and the sample-size can be adapted in the light of this new information.

In a first step we now estimate the covariance parameter ρ according to (24) in the light of the interim data. Assume that we find an estimated value of $\hat{\rho} = 0.733$. With this estimate we calculate the final value of the stage-two critical boundary u_2 by solving (25) with our estimate plugged in as ρ , and all remaining parameters as specified as above. Doing so yields in the value $u_2 = 2.17$ and ensures that the rejection region does not depend on our initial planing assumptions regarding ρ .

Having determined the final rejection region, let us now recalculate the sample-size such that a conditional power of $1 - \beta_2 = 0.8$ is achieved, say, under the constraint that the overall accrual period is at least $a_1 + s_1 = 2.7$ years, but must not exceed $a_{\max} = 5$ years. Notice that it is principally possible to adapt the recruitment rate r or the allocation ratio v_2 depending on B_1 or Z_{11} at time of the interim analysis. For simplicity, we here assume that neither accrual rate nor allocation ratio shall be adapted, i.e. we choose $r' = r$ and $v_2' = v_2$. In order to carry out sample size recalculation according to these specification, we first calculate the required length a_2^{CP} of the second stage accrual period to realize the desired conditional power of $1 - \beta_2 = 0.8$. This can be done by solving equation (33) for the only remaining indeterminate a_2^{CP} , which in our case yields $a_2^{CP} = 3.0$. To implement the constraint on the minimum and maximum length of accrual, the revised

length a'_2 of the second stage accrual period is finally chosen according to (30). With $a_{\max} = 5$, $a_1 = 1.7$, $s_1 = 1$ equation (30) yields $a'_2 = a_2^{CP} = 3.0$, corresponding to $n_2 = 226$ patients in stage two.

Finally after $1.7 + 3.0 + 2.0 = 6.7$ years after start of the trial, the final analysis is due. At this time the test statistics Z_{12} and Z_{22} become known. Assuming that $Z_{12} = 1.67$ and $Z_{22} = 3.14$ are observed, we finally obtain the final test statistic Z_2^* according to (19)

$$Z_2^* = \frac{\sqrt{0.158} \cdot 1.34 + \sqrt{0.247 - 0.158} \cdot 1.67 + \sqrt{0.233} \cdot 3.14}{\sqrt{0.247 + 0.233}} = 3.67 > 2.17 = u'_2,$$

which concludes a successful trial with rejection of H_0 after stage two.

We will present an example design for a seamless phase II/III trial in detail in the supplemental material.

Simulation

Design of the main scenario

We consider testing the hypothesis formulated in equation (5) $H_0: S_A(s) = S_B(s)$ for all $0 \leq s \leq s_{\max}$ using the two-step adaptive design presented in section 4.1.

In the context of the LOGGIC Europe trial, it was of interest to show a positive effect on the short term PFS-rate at an interim analysis to obtain the preliminary conditional marketing authorisation. Only with this conditional marketing authorization it was desired to continue recruitment of patients and to additionally test the effect on the long-term progression free survival.

More specifically, a design with rejection region of the form

$$\mathcal{R} := \{Z_1^* \geq u_1\} \cup \{B_1 > 0, Z_1^* < u_1, Z_2^* \geq u_2\}$$

would have been of interest. Notice that we set the critical boundaries $b_0 = 0$ and $u_0 = -\infty$. We set $s_0 = s_1 = 1.5$, $f = 2$ and $\pi = 0.5$. This corresponds to a two-step log-rank test with binding futility criterion based on the 18-months response rate.

The following frame conditions were chosen as the main scenario for this simulation study: Patients are allocated equally to both treatment arms (allocation ratio $v_1 = v_2 = 1$). Survival times are Weibull distributed with scale parameter of $m = 1/\log(2)$ and shape parameter $k = 1$, which corresponds to a scaled exponential distribution with median survival of 1 year. To study the performance of our algorithm we ran also simulations with shape parameters $k = 0.5$ and $k = 2$. Planing was done under the planing alternative $S_B(s) = S_A(s)^{\omega_1}$, where $\omega_1 = 2/3$. We also ran simulations with $\omega_1 = 4/5$. We let the true hazard ratio ω range between 0.5 and 1 in steps of 1/15. The one sided type 1 error rate was set to $\alpha = 0.025$ and the desired power was set to $1 - \beta = 0.8$. We set the conditional power parameter β_2 such that it satisfies the equation

$$P_{K_1}(Z_1^* \geq u_1) + P_{K_1}(Z_1^* < u_1, B_1 \geq b_0) \cdot (1 - \beta_2) = 1 - \beta. \quad (34)$$

This choice tries to stabilize the power of the whole trial despite the adaptation. The recruitment rate was set to $r = 60$. The maximal trial duration a_{\max} was set as $PF = 1.5$ times the duration of a corresponding single-step two-sample log-rank test.¹ In some of our scenarios (Figure 2) we let the parameter PF vary in the set $[1.25, 1.75]$ as a fine-tuning parameter.

No loss to follow-up was assumed as well as block-randomization and uniform recruitment assumptions as required by theorem A2.

For each simulation the required recruitment period lengths of stage one a_1 and stage two a_2 were calculated according to section 'Initial sample size calculation'. In our simulations we additionally distinguished between (i) a Pocock-type design with $u_1 = u_2$ and (ii) a design without early stopping where $u_1 := \infty$. Note that the critical bounds b_0 and u_1 have to be fixed in advance, whereas the value for u_2 is calculated according to equation (22) at the interim analysis, when the estimator $\hat{\rho}$ for ρ becomes available. Thus the theoretical equality " $u_1 = u_2$ " in the Pocock setting is effectively only realized approximately.

With above values for r , a_1 , a_2 and f , the weights η_{11} , η_{12} , η_{22} were calculated according to equations (20).

Then $r \cdot a_1$ patients were simulated as first stage patients, with preliminary censoring at study time s_1 , which represents the data we are allowed to use at the interim analysis. Based on this simulated data the interim statistics Z_1^* , B_1 , N_1 , $\hat{\rho}$ and $\hat{\omega}$ were calculated.

The test statistics Z_1^* and B_1 were then compared to the prefixed critical bounds b_0 and u_1 to determine whether early successful stopping or stopping for futility has occurred.

In the case of an ongoing trial, i.e. $B_1 > b_0$ and $Z_1^* < u_1$, the critical bound u_2 is obtained by solving equation (22) with the estimator $\hat{\rho}$ plugged in. Additionally, the required recruitment period length of stage two a_2^{CP} was calculated such that a conditional power of $1 - \beta_2$ is achieved under the revised planing alternative hypothesis $K'_1: S_B = S_A^{\omega}$ corresponding to the

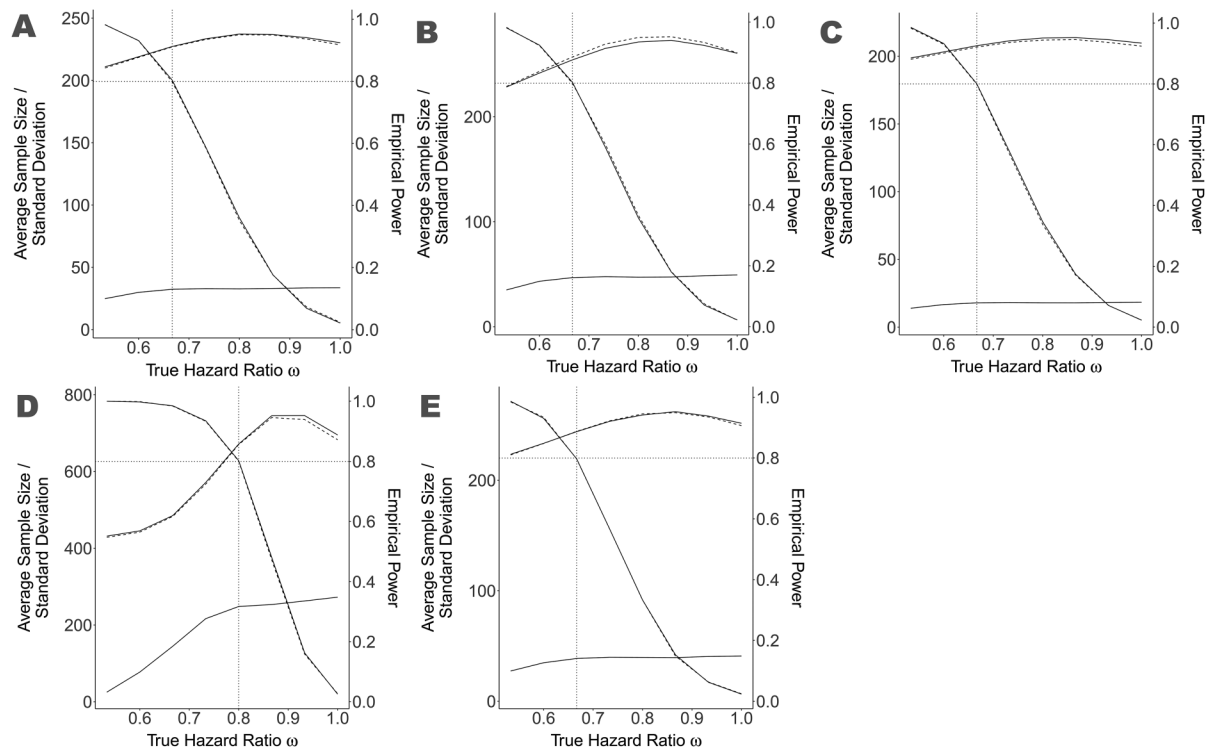


Figure 2. Average sample size, standard deviation of sample size and empirical power of the main scenario ($k = 1, \omega_1 = 2/3, r = 60, b_0 = 0, u_0 = -\infty, u_1 = \infty$) and some variations true hazard ratio ω ranging between 0.5 and 1.0, compared to a standard adaptive design with stop for futility. The solid lines represent our new methodology and the dashed lines the standard methodology, where the monotone decreasing lines starting at nearly 1 and ending by 0.025 represent the empirical power. The remaining upper lines show the average sample size and the lower lines the standard deviation of the sample size. Notice that the latter lines overlap considerably and are therefore difficult to distinguish. The vertical dotted line represents the value for ω used as planing alternative. The dotted, horizontal line represents the aimed power of 80%. Figure A is the main scenario, Figure B the variation with $k = 0.5$, Figure C the variation with $k = 2.0$, Figure D the variation with $\omega = 0.8$ and Figure E is the variation Pocock boundaries. The value of the fine-tune parameter PF is presented in the table on the bottom right for each scenario variation.

observed hazard ratio $\hat{\omega}$. The actual recruitment period length of stage two patients a'_2 was then updated as stated in (30), to stipulate the boundary conditions.

We then proceeded (i) to simulate $a'_2 \cdot r$ patients of stage two and (ii) to update the censoring date of stage one patients to calendar time $a' + f$.

Finally the test-statistic Z_2^* was calculated according to (19) and compared to the critical bound u_2 derived at the interim analysis to obtain the final test decision.

The above presented simulation algorithm was run 10,000 times for each scenario.

Results

The simulation results are presented briefly in Table 1. Reassuringly the designs hold the aimed significance level of 2.5%, even in the small sample size case. Note that with 10,000 simulations per scenario the estimated accuracy of our type one error rate estimator given through 95%-confidence intervals is $\pm 0.31\%$. Accordingly in no scenario the empirical type I error rate exceeded the aimed significance level of 2.5% in a statistically noticeable way.

The empirical power however shows a little more variation. This is due to the fact, that the initial sample size calculation does not factor in the randomness introduced by $\hat{\omega}$, which effects the sample size recalculation based on conditional power. This is a well-known effect of such adaptive designs.

Main simulation scenario. The strength of adaptive designs is undoubtedly the possibility for correction when the initial planning assumptions seem to be wrong. When the treatment effect is small one can stop for futility or increase the sample size to hold on the desired power. Conversely when the treatment effect is larger than expected one can decrease the sample size while still holding the desired power.

Table 1. Empirical type I error rate and power in the simulation scenarios. Empirical type I error rate (TOE) was obtained from simulations where the true hazard ratio $\omega = 1$. Empirical power was obtained by simulations where the true hazard ratio equals the planing hazard ratio ($\omega = \omega_1$). For further simulation details see section 6.

k	ω	u_1	Average n	Emp. TOE	Emp. power
0.5	2/3	∞	279.453	0.027	0.838
		Pocock	283.680	0.024	0.817
	4/5	∞	738.387	0.024	0.798
		Pocock	755.432	0.025	0.804
1.0	2/3	∞	249.870	0.026	0.839
		Pocock	256.583	0.025	0.815
	4/5	∞	671.721	0.025	0.793
		Pocock	690.295	0.026	0.799
2.0	2/3	∞	232.490	0.024	0.847
		Pocock	239.399	0.024	0.814
	4/5	∞	651.614	0.026	0.793
		Pocock	672.480	0.026	0.796

We simulated our main scenario ($k = 1, \omega_1 = 2/3, r = 60, u_1 = \infty$) with some variations. We used the parameter $PF \in [1.25, 1.75]$ as a fine-tuning parameter to level out the variation introduced by $\hat{\omega}$ and to match the aimed 80% power quite exactly. The choice of these fine-tuning parameter is presented in the table within Figure 2.

We compare our test algorithm with a standard adaptive design based on the standard methodology by Wassmer.³ To assure comparability we implemented a futility stop, when the short term log-rank test Z_1 based on the first half of patients shows a negative result. More specifically in the non Pocock designs, we compared our design to an adaptive design with rejection region

$$\mathcal{R}_{\text{simple},1} := \{Z_1 \geq 0, Z_2^* \geq \Gamma_1\}, \tag{35}$$

where Γ_1 is chosen such that $P_{H_0}(\mathcal{R}_{\text{simple},1}) = \alpha$. In the Pocock scenario we compared our design to a design with rejection region

$$\mathcal{R}_{\text{simple},2} := \{Z_1 \geq \Gamma_2\} \cup \{0 \leq Z_1 < \Gamma_2, Z_2^* \geq \Gamma_2\}, \tag{36}$$

where Γ_2 is chosen such that $P_{H_0}(\mathcal{R}_{\text{simple},2}) = \alpha$. These are rejection regions, which can be used within the methodology of Wassmer and are included in our methodology.

We set the required sample size such that the standard design also holds the desired power of $1 - \beta = 80\%$ under the planing hypothesis K_1 .

The operating characteristic of our test algorithm in the main simulation scenario ($k = 1, \omega_1 = 2/3, u_1 = \infty$) is presented in Figure 2 together with some variations of the scenario.

Across all scenario variations, the power and sample-size performance of our test statistic fits the performance of the standard methodology quite well.

In the main scenario the mean sample-size difference between the standard methodology and our new methodology is 0.74% at maximum. Under the planing hypothesis the maximal increase of the mean sample-size across all scenario variations was 0.5%, while in some cases the new design reduced the mean sample-size about 1.0%.

This suggests the use of easily interpretable survival rate differences as an interesting option for interim decision making in survival trials.

By using various Weibull shape parameters, planning hypothesis and design types we assured that the performance consistency is not dependant on our specific scenario assumptions.

Discussion

The confirmatory adaptive two-step log-rank test proposed here extends the one proposed by Wassmer.³ Whereas the test proposed by Wassmer essentially only allows the use of the interim log-rank statistic for data-dependent design modifications, our approach allows simultaneous use of the interim log-rank statistic and observed differences in cumulative hazard rates at time s_0 for interim decision making, while avoiding those problems arising with methods based on patient wise separation.⁶⁻⁸ Next to an adaptation of sample size, our approach also allows modification of the allocation ratio between the treatment arms or the recruitment rate, which neither has been described by

Wassmer³ nor Jenkins.⁶ This is of importance when thinking about application of our methodology in a multiarm, multistage setting. Even though the focus of this paper was on a trial design with two treatment arms and two analyses, the generalization to more than two arms and more than two analyses is straightforward using the methodology described by Hommel et al.⁹

Our adaptive two-step log-rank test exploits the independent increments structure of the limiting Gaussian process of the joint bivariate process defined by the log-rank statistic and the Nelson–Aalen difference at some time s_0 . Therefore, we emphasize that the full use of arbitrary interim data for design modifications is still not admissible here.⁴ However, our approach makes provision for the simultaneous use of (i) the interim log-rank statistic and (ii) differences in cumulative hazard rates at an arbitrary time s_0 .

The calculation of rejection regions and sample size formulas were based on distributional approximation of the bivariate test statistic in the large sample limit. Our methodology used mild regularity assumptions as well as the proportional hazards assumption. It is well known, that the log-rank test is less efficient and its distribution depends on the distribution of censoring times, when the proportional hazards assumption is violated.¹⁰ This is likely to be inherited by our method. The small sample properties were studied by simulations. The validity of the proposed design does not depend on specific model assumptions underlying these simulations such as exponentially distributed survival times. In view of the flexibility offered by our approach, however, applicants are recommended to assess different choices of design parameters in order to identify those parameter constellations with best operating characteristics as compared to a standard single-step two-sample log-rank test. For this purpose, we provide an R program in the supplemental material that enables easy assessment of operating characteristics and thus optimal calibration of the design parameters in a specific trial setting. The R program also underlies our simulation.

Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Jannik Feld  <https://orcid.org/0000-0003-1555-9192>

Andreas Faldum  <https://orcid.org/0000-0002-2674-8098>

Supplemental material

Supplemental material for this article is available online.

References

1. Peto R and Peto J. Asymptotically efficient rank invariant test procedures. *J R Statist Soc A* 1972; **135**: 185–207.
2. Schäfer H and Müller HH. Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections. *Stat Med* 2001; **20**: 3741–3751.
3. Wassmer G. Planning and analyzing adaptive group sequential survival trials. *Biom J* 2006; **48**: 714–729.
4. Bauer P and Posch M. Letter to the editor: modification of the sample size and the schedule of interim analyses in survival trials based on data inspections. *Stat Med* 2004; **23**: 1333–1335.
5. Pocock S. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977; **64**: 191–199.
6. Jenkins M, Stone A and Jennison C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharm Stat* 2011; **10**: 347–356.
7. Irle S and Schäfer HH. Interim design modifications in time-to-event studies. *J Am Stat Assoc* 2012; **107**: 341–348.
8. Magirr D, Jaki T, Koenig F, et al. Sample size reassessment and hypothesis testing in adaptive survival trials. *PLoS ONE* 2016; **11**: e0146465.
9. Hommel G. Adaptive modifications of hypotheses after an interim analysis. *Biom J* 2001; **43**: 581–589.
10. Brueckner M and Brannath W. Sequential tests for non-proportional hazards data. *Lifetime Data Anal* 2017; **23**: 339–352.
11. Sellke T and Siegmund S. Sequential analysis of the proportional hazards model. *Biometrika* 1982; **70**: 315–326.
12. Andersen PK, Borgan O, Gill RD, et al. *Statistical Models Based on Counting Processes*. New York: Springer, 1993.
13. Aalen O. *Borgan O and Gjessing H. Survival and Event History Analysis*. New York: Springer, 2008.

Appendix A

We will now deduce the distributional approximation presented in (8). The proofs presented here are formulated for a single-step design. However, the extension to a multi-step design is straight forward using the independent increments structure. We therefore drop the stage indices k for notational simplicity.

It is well known that for a patient i from treatment group $x = A, B$,

$$M_i(s) := D_i(s) - \int_0^s I(T_i \wedge C_i \geq u) \lambda_x(u) du \quad (37)$$

is an \mathcal{F}_s -martingale.¹¹ In particular, with $M_x(s) := \sum_{i \in N_x} M_i(s)$ and for any \mathcal{F}_s -adapted left-continuous process $H(s)$,

$$\begin{aligned} (H \bullet M_x)(s) &:= \int_0^s H(u) dM_x(u) \\ &= \int_0^s H(u) dD_x(u) - \int_0^s H(u) Y_x(u) \lambda_x(u) du \end{aligned}$$

is an \mathcal{F}_s -martingale with optional and predictable covariation process¹²

$$\begin{aligned} [H \bullet M_x](s) &= \int_0^s H^2(u) dD_x(u), \\ \langle H \bullet M_x \rangle(s) &= \int_0^s H^2(u) Y_x(u) \lambda_x(u) du. \end{aligned}$$

We aim for the joint distribution of the weighted two-sample log-rank statistic, which has the integral representation

$$LR(s) = \int_0^s \frac{L(u)}{Y_A(u)} dN_A(u) - \int_0^s \frac{L(u)}{Y_B(u)} dN_B(u)$$

and the difference of the group-wise Nelson-Aalen estimates

$$\Delta(s) = \int_0^s \frac{J_A(u)}{Y_A(u)} dN_A(u) - \int_0^s \frac{J_B(u)}{Y_B(u)} dN_B(u)$$

as \mathcal{F}_s -adapted processes, i.e. we aim for the distribution of the bivariate process

$$\Psi(s) := \begin{pmatrix} n^{-1/2} LR(s) \\ n^{1/2} \Delta(s) \end{pmatrix}.$$

Introducing the bivariate drift process

$$\mu(s) := \begin{pmatrix} n^{-1/2} \int_0^s L(u) \{ \lambda_A(u) - \lambda_B(u) \} du \\ n^{1/2} \int_0^s \{ J_A(u) \lambda_A(u) - J_B(u) \lambda_B(u) \} du \end{pmatrix},$$

it follows from (37) that

$$\begin{pmatrix} M^{LR}(s) \\ M^\Delta(s) \end{pmatrix} := \mathbf{M}(s) := \Psi(s) - \mu(s)$$

is a bivariate mean-zero \mathcal{F}_s -martingale. Since $[M_x] = N_x$ and $[M_A, M_B] = 0$, the optional covariation matrix $[\mathbf{M}](s)$ of \mathbf{M} has components

$$\begin{aligned}
 [M^{LR}](s) &= n^{-1} \int_0^s \frac{L(u)^2}{Y_A(u)^2} dN_A(u) + n^{-1} \int_0^s \frac{L(u)^2}{Y_B(u)^2} dN_B(u), \\
 [M^{LR}, M^\Delta](s) &= \int_0^s \frac{L(u)}{Y_A(u)^2} dN_A(u) + \int_0^s \frac{L(u)}{Y_B(u)^2} dN_B(u), \\
 [M^\Delta](s) &= n \int_0^s \frac{J_A(u)}{Y_A(u)^2} dN_A(u) + n \int_0^s \frac{J_B(u)}{Y_B(u)^2} dN_B(u).
 \end{aligned}$$

Since $\langle M_x \rangle(s) = \int_0^s Y_x(u) \lambda_x(u) du$ and $\langle M_A, M_B \rangle = 0$, the predictable covariation matrix $\langle \mathbf{M} \rangle(s)$ of \mathbf{M} has components

$$\begin{aligned}
 \langle M^{LR} \rangle(s) &= n^{-1} \int_0^s \frac{L(u)^2}{Y_A(u)} \lambda_A(u) du + n^{-1} \int_0^s \frac{L(u)^2}{Y_B(u)} \lambda_B(u) du, \\
 \langle M^{LR}, M^\Delta \rangle(s) &= \int_0^s \frac{L(u)}{Y_A(u)} \lambda_A(u) du + \int_0^s \frac{L(u)}{Y_B(u)} \lambda_B(u) du, \\
 \langle M^\Delta \rangle(s) &= n \int_0^s \frac{J_A(u)}{Y_A(u)} \lambda_A(u) du + n \int_0^s \frac{J_B(u)}{Y_B(u)} \lambda_B(u) du.
 \end{aligned}$$

Above equations are easily checked (see Aalen et al.¹³ Sec. 2.2.5). On this basis we may deduce the distributional properties of the bivariate process $\Psi(s) = \mathbf{M}(s) + \mu(s)$ in the large sample limit, as stated in the following theorems. The proofs of the theorems A1, A2 and equations (13) and (14) are presented after some additional results, which we need.

Theorem A1 Fix $s_{\max} > 0$ and assume that the hazard functions λ_A and λ_B are bounded on the interval $[0, s_{\max}]$ and $P(C_i > s_{\max}) > 0$. Assume furthermore that the treatment group allocation is done by block randomisation i.e. there exists a constant $BL \in \mathbb{N}$, that for all $n \in \mathbb{N}$ it holds

$$\left\| \#\mathcal{N}_A^n - \left[\frac{1}{1+v} \cdot n \right] \right\| \vee \left\| \#\mathcal{N}_B^n - \left[\frac{v}{1+v} \cdot n \right] \right\| < BL.$$

Then for all $s \in [0, s_{\max}]$ the limit $\Sigma(s) := plim_{n \rightarrow \infty} [\mathbf{M}](s)$ exists and there is a Gaussian mean zero martingale $\mathbf{M}^\infty(s)$ with covariance matrix $\Sigma(s)$ s.t.

$$\mathbf{M} \xrightarrow[n \rightarrow \infty]{D} \mathbf{M}^\infty.$$

In particular, the following distributional approximation holds:

- (i) $\mathbf{M}(s_2) - \mathbf{M}(s_1)$ and $\mathbf{M}(s_1)$ are independent for all $0 \leq s_1 \leq s_2 \leq s_{\max}$.
- (ii) $\mathbf{M}(s) \sim \mathcal{N}(\mathbf{0}, \Sigma(s))$ for all $s \in [0, s_{\max}]$.
- (iii) $\mathbf{M}(s_2) - \mathbf{M}(s_1) \sim \mathcal{N}(\mathbf{0}, \Sigma(s_2) - \Sigma(s_1))$ for all $0 \leq s_1 \leq s_2 \leq s_{\max}$.

Theorem A2 Assume the conditions from theorem 1. Then, under the contiguous alternatives $\Lambda_B(s) = \omega_n \Lambda_A(s)$ with $\omega_n := e^{-n^{-1/2}\gamma}$ for $\gamma \geq 0$, we have

$$\Sigma(s) = \begin{pmatrix} \sigma_{LR}^2(s) & \Lambda_A(s) \\ \Lambda_A(s) & \sigma_\Delta^2(s) \end{pmatrix} \text{ and } \lim_{n \rightarrow \infty} \mu(s) = \begin{pmatrix} \gamma \cdot \sigma_{LR}^2(s) \\ \gamma \cdot \Lambda_A(s) \end{pmatrix},$$

for some continuous variance functions $\sigma_{LR}^2(s)$ and $\sigma_\Delta^2(s)$. More explicit we have $\sigma_{LR}(s) = \sqrt{\pi(s)} \frac{\sqrt{v}}{1+v}$, where $\pi(s) := plim_{n \rightarrow \infty} N(s)/n$. In particular the limit $\pi(s)$ exists.

Since $\Psi(s) = \mathbf{M}(s) + \mu(s)$, we conclude that the processes in (38) below have independent increments with the following distributional large sample approximation:

$$\begin{aligned} \begin{pmatrix} n^{-1/2}LR(s) \\ n^{1/2}\Delta(s) \end{pmatrix} D \approx \mathcal{N} \left(\begin{pmatrix} \gamma \cdot \sigma_{LR}^2(s) \\ \gamma \cdot \Lambda_A(s) \end{pmatrix}, \begin{pmatrix} \sigma_{LR}^2(s) & \Lambda_A(s) \\ \Lambda_A(s) & \sigma_{\Delta}^2(s) \end{pmatrix} \right), \\ \begin{pmatrix} \frac{n^{-1/2}LR(s)}{\sqrt{[M^{LR}](s)}} \\ \frac{n^{1/2}\Delta(s)}{\sqrt{[M^{\Delta}](s)}} \end{pmatrix} D \approx \mathcal{N} \left(\begin{pmatrix} -\sqrt{n} \log(\omega) \cdot \sigma_{LR}(s) \\ -\sqrt{n} \log(\omega) \cdot \frac{\Lambda_A(s)}{\sigma_{\Delta}(s)} \end{pmatrix}, \begin{pmatrix} 1 & \frac{\Lambda_A(s)}{\sigma_{LR}(s)\sigma_{\Delta}(s)} \\ \frac{\Lambda_A(s)}{\sigma_{LR}(s)\sigma_{\Delta}(s)} & 1 \end{pmatrix} \right). \end{aligned} \tag{38}$$

Using the independence of increments we get for $0 < s_0 \leq s_1 \leq s_{\max}$

$$\begin{pmatrix} \frac{n^{-1/2}LR(s_1)}{\sqrt{[M^{LR}](s_1)}} \\ \frac{n^{1/2}\Delta(s_0)}{\sqrt{[M^{\Delta}](s_0)}} \end{pmatrix} D \approx \mathcal{N} \left(\begin{pmatrix} -\sqrt{n} \log(\omega) \cdot \sigma_{LR}(s_1) \\ -\sqrt{n} \log(\omega) \cdot \frac{\Lambda_A(s_0)}{\sigma_{\Delta}(s_0)} \end{pmatrix}, \begin{pmatrix} 1 & \frac{\Lambda_A(s_0)}{\sigma_{LR}(s_1)\sigma_{\Delta}(s_0)} \\ \frac{\Lambda_A(s_0)}{\sigma_{LR}(s_1)\sigma_{\Delta}(s_0)} & 1 \end{pmatrix} \right),$$

which is the distributional approximation we stated in (8).

To prove our theorems we need some additional results.

Proposition A3 Let $(X_n(s))_{n \in \mathbb{N}}$ be a sequence of stochastic processes, f a borel measurable function and $s_{\max} > 0$ with

- (1) $X_n(s) \xrightarrow{P} f(s)$,
- (2) $\int_0^{s_{\max}} |f(s)| < \infty$,

which also satisfy one of the following conditions.

- (a) It exists a constant $c > 0$, that for all $n \in \mathbb{N}$ and for all $s \in [0, s_{\max}]$ it holds

$$|X_n(s)| < c \quad \text{a.s.}$$

- (b) It exists a sequence of stochastic processes $(\tilde{X}_n(s))_{n \in \mathbb{N}}$, a function h and an integrable function g with

- (i) $|\tilde{X}_n(s)|$ is monotone in s for all $n \in \mathbb{N}$ a.s.,
- (ii) $|\tilde{X}_n(s)| \xrightarrow{P} h(s)$ for all $s \in [0, s_{\max}]$,
- (iii) $|X_n(s)| \leq |\tilde{X}_n(s) \cdot g(s)|$ for all $s \in [0, s_{\max}]$ a.s..

Then it holds

$$\int_0^s X_n(u)du \xrightarrow{P} \int_0^s f(u)du \quad \text{for all } s \in [0, s_{\max}].$$

Proof. We will show that conditions (1),(2) and (a) are sufficient to satisfy the preconditions of Hellands proposition¹² Prop. II.5.2 and that conditions (1),(2) and (b) are sufficient to satisfy the preconditions of Gills proposition¹² Prop. II.5.3 .

Assume conditions (1),(2) and (a) are satisfied. Then for any $c' > c$ it holds

$$1_{\{|X_n(s)| > c'\}} = 0 \text{ a.s. and furthermore}$$

$$\sup_{n \in \mathbb{N}} \mathbb{E}[|X_n(s)| 1_{\{|X_n(s)| > c'\}}] = 0.$$

With $k(s) = c$ it holds

$$\mathbb{E}[|X_n(s)|] \leq \mathbb{E}[c] = c = k(s).$$

Therefore the preconditions of Hellands proposition are satisfied.

Assume now that conditions (1),(2) and (b) are satisfied. Without loss of generality we can assume, that $|\tilde{X}_n(s)|$ is monotone increasing in s . Otherwise we transition to $|\tilde{X}_n(s_{\max} - s)|$. For arbitrary $\delta > 0$ we choose some arbitrary but fixed $\epsilon > 0$

and define $k_\delta(s) := |\{h(s_{\max}) + \epsilon\} \cdot g(s)|$. It holds

$$\begin{aligned} &P(|X_n(s)| \leq k_\delta(s) \text{ for all } s \in [0, s_{\max}]) \\ &= P(|X_n(s)| \leq |\{h(s_{\max}) + \epsilon\} \cdot g(s)| \text{ for all } s \in [0, s_{\max}]) \\ &\geq P(|\tilde{X}_n(s) \cdot g(s)| \leq |\{h(s_{\max}) + \epsilon\} \cdot g(s)| \text{ for all } s \in [0, s_{\max}]) \\ &\geq P(|\tilde{X}_n(s)| \leq |\{h(s_{\max}) + \epsilon\}| \text{ for all } s \in [0, s_{\max}]) \\ &= P(|\tilde{X}_n(s_{\max})| \leq |\{h(s_{\max}) + \epsilon\}|) \xrightarrow{n \rightarrow \infty} 1 > 1 - \delta. \end{aligned}$$

In the last equality we used, that $|\tilde{X}_n(s)|$ is monotone increasing in s . The convergence holds because of the convergence of $|\tilde{X}_n(s)|$ in probability. Above inequality yields in the preconditions of Gills proposition. \square

Lemma A4 (Simple weak law of large numbers) Let $(\mu_n)_{n \in \mathbb{N}}$ be a sequence in $[0, 1]$ with $\lim_{n \rightarrow \infty} \mu_n = \mu \in [0, 1]$ and let $(X_i^{(n)})_{i=1, \dots, n}$ be a sequence of independent $Ber(\mu_n)$ distributed random variables.

Then it holds

$$\frac{\sum_{i=1}^n X_i^{(n)}}{n} \xrightarrow[n \rightarrow \infty]{P} \mu.$$

Proof. We will proof the convergence in probability by showing, that the convergence holds in L^2 . Define $S_n := \sum_{i=1}^n X_i^{(n)}$, then it holds

$$\begin{aligned} \mathbb{E} \left[\left(\frac{S_n}{n} - \mu \right)^2 \right] &= \mathbb{E} \left[\left(\frac{S_n}{n} - \mu_n + \mu_n - \mu \right)^2 \right] \\ &= \mathbb{E} \left[\left(\frac{S_n}{n} - \mu_n \right)^2 \right] + 2 \mathbb{E} \left[\left(\frac{S_n}{n} - \mu_n \right) (\mu_n - \mu) \right] + \mathbb{E} \left[(\mu_n - \mu)^2 \right] \\ &= \frac{1}{n^2} \text{Var}(S_n) + 2 \mathbb{E} \left[\frac{S_n}{n} - \mu_n \right] \cdot (\mu_n - \mu) + (\mu_n - \mu)^2 \\ &= \frac{1}{n^2} n \mu_n (1 - \mu_n) + (\mu_n - \mu)^2 \rightarrow 0 \quad \square \end{aligned}$$

Lemma A5 (Weak law of large numbers) Let $(\#^n)_{n \in \mathbb{N}}$ be an infinite sequence of N -valued, monotone in $n \in \mathbb{N}$ increasing random variables, which satisfy the following conditions

- $\forall n \in \mathbb{N} : \#^n \leq n,$
- $\exists c \in [0, 1] : \frac{\#^n}{n} \xrightarrow[n \rightarrow \infty]{P} c,$
- $\exists BL \in \mathbb{N} : \forall n \in \mathbb{N} : \|\#^n - \lfloor cn \rfloor\| < BL.$

Let furthermore $(\mu_n)_{n \in \mathbb{N}}$ be a sequence in $[0, 1]$ with $\lim_{n \rightarrow \infty} \mu_n = \mu \in [0, 1]$ and $(X_i^{(n)})_{i=1, \dots, n}$ a sequence of independent $Ber(\mu_n)$ distributed random variables. Then it holds

$$\frac{\sum_{i=1}^{\#^n} X_i^{(n)}}{n} \xrightarrow[n \rightarrow \infty]{P} c \cdot \mu.$$

Proof. We will show the convergence in L^2 . Define $Z_n := \sum_{i=1}^{\#^n} X_i^{(n)}$ and $S_n := \sum_{i=1}^{\lfloor cn \rfloor} X_i^{(n)}$. Then it holds

$$\begin{aligned} \mathbb{E}\left[\left(\frac{Z_n}{cn} - \mu\right)^2\right] &= \mathbb{E}\left[\left(\frac{Z_n}{cn} - \frac{S_n}{cn} + \frac{S_n}{cn} - \mu\right)^2\right] \\ &= \mathbb{E}\left[\left(\frac{Z_n - S_n}{cn}\right)^2\right] + 2\mathbb{E}\left[\left(\frac{Z_n - S_n}{cn}\right)\left(\frac{S_n}{cn} - \mu\right)\right] + \mathbb{E}\left[\left(\frac{S_n}{cn} - \mu\right)^2\right]. \end{aligned}$$

The first summand satisfies the inequality

$$\left|\frac{Z_n - S_n}{cn}\right|^2 \leq \left|\frac{\#^n - \lfloor cn \rfloor}{cn}\right|^2 \leq \left|\frac{BL}{cn}\right|^2 \xrightarrow{n \rightarrow \infty} 0.$$

For the second summand it holds

$$\begin{aligned} 2\mathbb{E}\left[\left(\frac{Z_n - S_n}{cn}\right)\left(\frac{S_n}{cn} - \mu\right)\right] &\leq 2\mathbb{E}\left[\left|\frac{Z_n - S_n}{cn}\right|\left|\frac{S_n}{cn} - \mu\right|\right] \\ &\leq 2\frac{BL}{cn}\mathbb{E}\left[\left|\frac{S_n}{cn} - \mu\right|\right] \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

For the convergence we used, that the second factor is bounded. The last summand vanishes analogously to the proof of prior lemma in the limit. Therefore the whole sum converges to 0. Multiplying both sides with c concludes the assertion. \square

Proof of theorem A1. We want to show, that $\text{plim}_{n \rightarrow \infty} \langle \mathbf{M} \rangle(s)$ exists. For this purpose we first take a closer look at the random variables $Y_A^n(u)$ and $Y_B^n(u)$ for some fixed $u \in \mathbb{R}^+$. If patient $i \in \mathbb{N}$ is under treatment A we use the notation $P_A := P$ (analogously $P_B := P$) to emphasize the stochastic influence of the treatment. It holds

$$Y_A^n(u) = \sum_{i \in \mathcal{N}_A^n} 1_{\{T_i > u, C_i > u\}}$$

and furthermore

$$P_A(T_i > u, C_i > u) = P_A(T_i > u) \cdot P(C_i > u) = S_A(u) \cdot P(C_i > u).$$

In the first equation we used the independence of T_i and C_i . With the independence of patients, it follows

$$Y_A^n(u) \sim \text{Bin}(\#\mathcal{N}_A^n, S_A(u) \cdot P(C_i > u)),$$

analogously it follows, that

$$Y_B^n(u) \sim \text{Bin}(\#\mathcal{N}_B^n, S_B(u) \cdot P(C_i > u)). \tag{39}$$

Because of block randomisation, the convergences $\frac{\#\mathcal{N}_A^n}{n} \xrightarrow[n \rightarrow \infty]{P} \frac{1}{1+v}$ and $\frac{\#\mathcal{N}_B^n}{n} \xrightarrow[n \rightarrow \infty]{P} \frac{v}{1+v}$ hold and we can apply the weak law of large numbers to get the convergences

$$\begin{aligned} \frac{Y_A^n(u)}{n} &\xrightarrow[n \rightarrow \infty]{} \frac{1}{1+v} \cdot S_A(u) \cdot P(C_i > u) = : y_A(u), \\ \frac{Y_B^n(u)}{n} &\xrightarrow[n \rightarrow \infty]{} \frac{v}{1+v} \cdot S_B(u) \cdot P(C_i > u) = : y_B(u). \end{aligned} \tag{40}$$

For the sake of readability we introduce the Notation $-A := B$ and $-B := A$. Then $\langle \mathbf{M} \rangle(s)$ has components

$$\langle M_n^\Delta \rangle(s) = \sum_{x=A,B} n \int_0^s \frac{J_x^n(u)}{Y_x^n(u)} \lambda_x(u) du, \tag{41}$$

$$\langle M_n^{LR} \rangle(s) = \sum_{x=A,B} n^{-1} \int_0^s \frac{Y_x^n(u) \cdot Y_{-x}^n(u)^2}{\{Y_x^n(u) + Y_{-x}^n(u)\}^2} \lambda_x(u) du, \tag{42}$$

$$\langle M_n^\Delta, M_n^{LR} \rangle(s) = \sum_{x=A,B} \int_0^s \frac{Y_{-x}^n(u)}{Y_x^n(u) + Y_{-x}^n(u)} \lambda_x(u) du. \tag{43}$$

The limit in probability of the integrals can be computed with use of proposition A3. The integrands in (41) satisfy the preconditions (b) via

$$\tilde{X}_x^n(u) = n \cdot \frac{J_x^n(u)}{Y_x^n(u)} + n \cdot 1_{\{J_x^n(u)=0\}} \xrightarrow[n \rightarrow \infty]{P} c_x \frac{1}{y_x(u)} \text{ and } g(u) = \lambda_x(u),$$

with $c_A = (1 + \nu)$ and $c_B = (1 + \nu)/\nu$. Note that the convergence holds because of Slutsky’s theorem and equation (40). The integrands in (42) and (43) satisfy the precondition (a) of proposition A3, because they are bounded by $\sup_{u \in [0, s_{\max}]} \lambda_x(u) < \infty$. It follows in probability as $n \rightarrow \infty$

- $\langle M_n^{LR} \rangle(s) \rightarrow \sum_{x=A,B} \int_0^s \frac{y_x(u)y_{-x}(u)^2}{\{y_x(u)+y_{-x}(u)\}^2} \lambda_x(u) du,$
- $\langle M_n^\Delta, M_n^{LR} \rangle(s) \rightarrow \sum_{x=A,B} \int_0^s \frac{y_x(u)}{y_x(u)+y_{-x}(u)} \lambda_x(u) du,$
- $\langle M_n^\Delta \rangle(s) \rightarrow \sum_{x=A,B} \int_0^s \frac{1}{y_x(u)} \lambda_x(u) du.$

Moreover, the jumpsize of $\mathbf{M}(s)$ is of order $n^{-1/2}$ (and thus vanishes in the limit $n \rightarrow \infty$), because N_x has jumpsize 1 and Y_x is of order n . Thus, by a multivariate version of Rebolledo’s martingale central limit theorem¹² (theorem II.5.1), $\text{plim}_{n \rightarrow \infty} [\mathbf{M}](s)$ exists and coincides with $\text{plim}_{n \rightarrow \infty} \langle \mathbf{M} \rangle(s)$, and $[\mathbf{M}(s)]_{s \geq 0}$ converges to a mean zero Gaussian process $[\mathbf{M}^{(\infty)}(t)]_{t \geq 0}$ with independent increments and covariance matrix $\Sigma(s) = : \text{plim}_{n \rightarrow \infty} [\mathbf{M}](s)$. \square

Proof of theorem A2. Let $y(u) := (1 + \nu) \cdot y_A(u)$, with y_A as in (40). Analogously to (39) it follows, that

$$Y_B^n(u) \sim \text{Bin}(\#N_B^n, S_A^{o_n}(u) \cdot P(C_i > u)).$$

Using the weak law of large numbers, we therefore conclude

$$\frac{Y_B^n(u)}{n} \xrightarrow[n \rightarrow \infty]{} \frac{\nu}{1 + \nu} \cdot S_A(u) \cdot P(C_i > u) = \nu \cdot y_A(u). \tag{45}$$

We will now calculate the limit π and thus show its existence. Recalling that $n^{-1/2}M_A(s)D \rightarrow \mathcal{N}(0, 1)$, we have in the limit $n \rightarrow \infty$ in probability

$$\begin{aligned} \frac{N(s)}{n} &= \frac{1}{n}M(s) + \frac{1}{n} \int_0^s Y_A(u)\lambda_A(u)du + \frac{1}{n} \int_0^s Y_B(u)\lambda_B(u)du \\ &= \frac{1}{n}M(s) + \frac{1}{n} \int_0^s Y_A(u)\lambda_A(u)du + e^{-n^{-1/2}\gamma} \cdot \frac{1}{n} \int_0^s Y_B(u)\lambda_A(u)du \\ &\rightarrow 0 + \int_0^s y_A(u)\lambda_A(u)du + \int_0^s y_B(u)\lambda_A(u)du = \int_0^s y(u)\lambda_A(u)du. \end{aligned} \tag{46}$$

The convergence of the integrals holds using (40), Slutsky theorem and proposition A3 (a). To prove the assertion regarding $\mu_k(s)$, notice that in the limit $n \rightarrow \infty$ in probability

$$\begin{aligned} &n^{-1/2} \int_0^s L(u)\{\lambda_A(u) - \lambda_B(u)\} du \\ &= n^{1/2}(1 - e^{-n^{-1/2}\gamma}) \int_0^s \frac{Y_B(u)/n}{Y_A(u)/n + Y_B(u)/n} \frac{Y_A(u)}{n} \lambda_A(u) du \\ &\rightarrow \frac{\gamma \cdot \nu}{(1 + \nu)^2} \int_0^s y(u)\lambda_A(u) du = \frac{\gamma \cdot \nu}{(1 + \nu)^2} \cdot \pi(s). \end{aligned}$$

The convergence of the integrals holds again using (40), Slutsky theorem and proposition A3 (a). In the last equality we made use of (46). Since $J_A(u) \rightarrow 1$ and $J_B(u) \rightarrow 1$ as $n \rightarrow \infty$, we likewise conclude

$$\begin{aligned} & \lim_{n \rightarrow \infty} n^{1/2} \int_0^s \{J_A(u)\lambda_A(u) - J_B(u)\lambda_B(u)\} du \\ &= \lim_{n \rightarrow \infty} \{n^{1/2}(1 - e^{-n^{-1/2}\gamma})\} \int_0^s \lambda_A(u) du = \gamma \Lambda_A(s). \end{aligned}$$

With similar arguments, we used to show (44), we can conclude

$$\begin{aligned} \langle M^{LR}, M^\Delta \rangle(s) &= \sum_{x=A,B} \int_0^s \frac{Y_{-x}(u)/n}{Y_A(u)/n + Y_B(u)/n} \lambda_x(u) du \\ &\rightarrow \frac{v}{1+v} \int_0^s \lambda_A(u) du + \frac{1}{1+v} \int_0^s \lambda_A(u) du = \Lambda_A(s) \end{aligned}$$

as well as

$$\begin{aligned} \langle M^{LR} \rangle(s) &= \sum_{x=A,B} \int_0^s \left\{ \frac{Y_{-x}(u)/n}{Y_A(u)/n + Y_B(u)/n} \right\}^2 \{Y_x(u)/n\} \cdot \lambda_x(u) du \\ &\rightarrow \left(\frac{v}{1+v} \right)^2 \int_0^s \frac{y(u)}{1+v} \lambda_A(u) du + \left(\frac{1}{1+v} \right)^2 \int_0^s \frac{v \cdot y(u)}{1+v} \lambda_A(u) du \\ &= \frac{v}{(1+v)^2} \int_0^s y(u) \lambda_A(u) du = \frac{v}{(1+v)^2} \pi(s). \end{aligned} \tag{47}$$

In the last equality, we made use of (46). \square

Proof of equations (13) and (14). Our preliminary work enables a quick calculation of $\sigma_{LR}(s)$ and $\sigma_\Delta(s)$. For this purpose we first take a closer look at $y_A(u)$. Under the no loss to follow up and uniform recruitment assumptions, it holds

$$P(C_i > u) = P(C_i + E_i > u + E_i) = P(a + f - u > E_i) = \begin{cases} 1 & u \leq f \\ 0 & u \geq a + f \\ \frac{a+f-u}{a} & \text{else.} \end{cases} \tag{48}$$

Analogue to proof of theorem 1 and with use of (45) it holds

$$\langle M_n^\Delta \rangle(s) \xrightarrow{n \rightarrow \infty} \frac{P}{v} \frac{(1+v)^2}{v} \int_0^s \frac{\lambda_A(u)}{S_A(u) \cdot P(C_i > u)} du \tag{49}$$

Using equation (48), the identity $\lambda_A(u) = -S'_A(u)/S_A(u)$ and the substitution $z = S_A(u)$, we get equation (13).

With (47) and (48), we also conclude

$$\langle M_n^{LR} \rangle(s) \xrightarrow{n \rightarrow \infty} \frac{P}{(1+v)^2} \int_0^s S_A(u) \cdot P(C_i > u) \lambda_A(u) du$$

Using the same identities and substitution as in (49), we get equation (14). \square

A seamless phase II/III design

In this section we elaborate application of our design algorithm in the context of a two-armed randomized seamless phase II/III survival trial. In the phase II part, we assume that the two treatments are compared regarding the short-term endpoint *survival rate at time s_0* . I.e. as phase II part, we consider a local level α test of the confirmatory null hypothesis $H_{0,1} : S_A(s_0) = S_B(s_0)$ on the s_0 survival rates using the rejection region

$$\mathcal{R}_1 := \{B_1 > \Phi^{-1}(1 - \alpha)\}. \tag{50}$$

\mathcal{R}_1 realizes a single step test of $H_{0,1}$. Only in the case of rejection of $H_{0,1}$ (i.e. $B_1 > \Phi^{-1}(1 - \alpha)$), we continue the trial in order to compare the two treatments also regarding long-term survival. I.e. as phase III part, we consider a local level α test of the confirmatory null hypothesis $H_{0,2} : S_A(s) = S_B(s)$ for all $0 \leq s \leq s_{\max}$ for some prefixed $s_{\max} > 0$ using the rejection region

$$\mathcal{R}_2 := \{Z_1^* \geq u_1\} \cup \{Z_1^* < u_1, Z_2^* \geq u_2\}. \quad (51)$$

\mathcal{R}_2 realizes a two-step test of $H_{0,2}$. It makes sense to synchronize the analysis of $H_{0,1}$ with the interim analysis of $H_{0,2}$. Notice that, we may choose $u_1 := \infty$ if we wish to refrain from testing $H_{0,2}$ already at the interim analysis and that adjustment to multiple testing is done by hierarchical testing in the order $H_{0,1}$ followed by $H_{0,2}$, i.e. we reject $H_{0,2}$ to the multiple level α if and only if $H_{0,1}$ and $H_{0,2}$ are both rejected by their local level α tests. $H_{0,1}$ can be rejected to the multiple level α if $H_{0,1}$ is rejected locally.

At the interim analysis, we are free to perform a data-dependent sample size recalculation based on the observed interim log-rank statistic Z_1^* and the observed difference in the short term response B_1 .

For a sample size calculation algorithm we have to apply the methodology presented in section ‘Initial sample size calculation’ to the rejection region

$$\mathcal{R}_3 := \mathcal{R}_1 \cap \mathcal{R}_2 = \{B_1 > \Phi^{-1}(1 - \alpha), Z_1^* \geq u_1\} \cup \{B_1 > \Phi^{-1}(1 - \alpha), Z_1^* < u_1, Z_2^* \geq u_2\} \quad (52)$$

with power defined by the probability $P_K(\mathcal{R}_3)$ under some planing alternative K .