Aus dem Institut für
Medizinische Biometrie und Epidemiologie
Direktor: Prof. Dr. Helmut Schäfer

des Fachbereichs Medizin der Philipps-Universität Marburg

Titel der Dissertation:

INTERIM DESIGN MODIFICATIONS IN TIME-TO-EVENT STUDIES

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
(Dr. rer. nat.)
dem Fachbereich Medizin
der Philipps-Universität Marburg
vorgelegt von

SEBASTIAN IRLE

aus Siegen

Marburg, 2012

Angenommen vom Fachbereich Medizin
der Philipps-Universität Marburg am: 20.01.2012

Gedruckt mit Genehmigung des Fachbereichs.

DEKAN: Prof. Dr. Matthias Rothmund
REFERENT: Prof. Dr. Helmut Schäfer
KORREFERENT: Prof. Dr. Hajo Holzmann

Ohana means family.

Family means nobody gets left behind, or forgotten.

— Lilo & Stitch

ABSTRACT

We propose a flexible method for interim design modifications in time-to-event studies. With this method, it is possible to inspect the data at any time during the course of the study, without the need for pre-specification of a learning phase, and to make certain types of design modifications depending on the interim data without compromizing the type I error risk. The method can be applied to studies designed with a conventional statistical test, fixed sample or group sequential, even when no adaptive interim analysis and no specific method for design adaptations (such as combination tests) had been foreseen in the protocol. Currently, the method supports design changes such as an extension of the recruitment or follow-up period, as well as certain modifications of the number and the schedule of interim analyses as well as changes of inclusion criteria. In contrast to existing methods offering the same flexibility, our approach allows to make use of the full interim information collected until the time of the adaptive data inspection. This includes time-to-event data from patients who have already experienced an event at the time of the data inspection, and preliminary information from patients

still alive, even if this information is predictive for survival, such as early treatment response in a cancer clinical trial.

Our method is an extension of the so-called conditional rejection probability (CRP) principle. It is based on the conditional distribution of the test statistic given the final value of the same test statistic from a subsample, namely the learning sample. It is developed in detail for the example of the logrank statistic, for which we derive this conditional distribution using martingale techniques.

Major parts of this work will be published in the Journal of the American Statistical Association, see Irle and Schäfer (2012).

ZUSAMMENFASSUNG

Wir stellen eine flexible Methode vor, mit der Design-Adaptionen in Überlebenszeitstudien vorgenommen werden können. Ein Vorteil dieses Ansatzes ist, dass eine adaptive Zwischenauswertung zu einem beliebigen Zeitpunkt während der Studiendauer vorgenommen werden kann, ohne dass dieser vor Studienbeginn festgelegt worden sein muss. Abhängig von den beobachteten Studiendaten können mit Hilfe unserer Methode verschiedene Design-Änderungen unter Wahrung des Signifiganzniveaus vorgenommen werden. Dabei kann unser Ansatz sogar in Studien mit optimalen fixed-sample bzw. gruppensequentiellen Designs

zur Anwendung kommen, in denen ursprünglich keine Zwischenauswertung vorgesehen gewesen ist. Derzeit unterstützt unsere Methode Verlängerungen der Nachbeobachtungszeit, bestimmte Veränderungen in der Anzahl und Terminierung der Zwischenauswertungen sowie Veränderungen der Einschlusskriterien.

Im Gegensatz zu bereits existierenden Methoden, die die selbe Flexibilität bieten, hat unser Ansatz den großen Vorteil, dass sämtliche Patienteninformationen, die zum Zeitpunkt der Dateninspektion erhoben werden, für Designveränderungen genutzt werden können. Dies umfasst genauso Überlebenszeitdaten von Patienten, die zum Zeitpunkt der Dateninspektion bereits verstorben sind, wie sämtliche Informationen zu Patienten, die zum Zeitpunkt der Dateninspektion noch leben - selbst, wenn diese Informationen mit der noch nicht beobachteten Überlebenszeit korrelieren. Mithin ist es beispielsweise in Krebsstudien möglich, die Tumor-Response von noch lebenden Patienten als Entscheidungsgrundlage für eventuelle Designänderungen zu verwenden. Die Möglichkeit, sämtliche Patienteninformationen für Designänderungen heranzuziehen, minimiert die Wahrscheinlichkeit suboptimaler oder gar falscher Designadaptionen.

Unser Ansatz ist eine Erweiterung des sogenannten Conditional Rejection Probability (CRP) Prinzips und basiert auf einer bedingten Verteilung der verwendeten Teststatistik. Wir leiten

diese mit Hilfe eleganter Martingalmethoden für die Logrank-Statistik her.

Große Teile dieser Arbeit werden in der hier vorliegenden Form im Journal of the American Statistical Association publiziert, vgl. Irle und Schäfer (2012).

CONTENTS

LIST OF FIGURES

1

INTRODUCTION

In the last two decades, a broad variety of adaptive test procedures has been proposed that allow the adaptation of design elements to the data collected earlier in the course of the trial, while conserving the type I error risk, see e.g. Bauer and Köhne (1994), Proschan and Hunsberger (1995), Lehmacher and Wassmer (1999), Müller and Schäfer (2004). Liu et al. (2002) developed a unified theoretical framework for adaptive designs that broadly established the validity of adaptation. Typical applications of adaptive designs include data dependent variance and sample size re-estimations. In the last decade, more complex designs and design changes have been considered such as dropping of treatment arms in multi-armed or dose finding trials, or focussing patient recruitment on most promising patient subgroups. In general, these so-called phase II/III combination designs combine multiple testing procedures such as closed testing with adaptive design methods (Brannath et al. 2009; Liu and Pledger 2005).

Schäfer and Müller (2001), Wassmer (2006), Brannath et al. (2009) and various other authors proposed adaptive design methods for studies that employ specific test statistics for censored

survival times or time-to-event endpoints. With several of these methods, alpha inflation may still arise when predictive information from patients still alive at the time of the adaptive interim data inspection is used to decide upon design modifications, see Bauer and Posch (2004) and Jahn-Eimermacher and Ingel (2009). This is due to the fact that design modifications then are not stochastically independent of survival times that have not been observed yet. Therefore, these methods do not allow to fully exploit preliminary information correlated with time-to-event of patients still alive at the time of the interim inspection, although this would be highly desirable in order to ensure that decisions upon an eventual modification of the future study design are as substantiated as possible. For example, researchers may wish to use interim information on tumor response observable in an early stage of a cancer clinical trial when only little survival information has been observed yet.

Jenkins et al. (2010) have recently developed a data-adaptive phase II/III seamless design for subgroup selection availing of the full interim information from the patients in the learning sample. Their method is based on a p-value combination test combining p-values from the learning stage patients and the validation stage patients at the end of the study. In their design, the number of events to be observed in each of the two stages (or the length of the follow-up periods) is pre-specified. Also Liu and Pledger (2005) proposed an adaptive procedure for phase

II/III combination designs that allows to use an early endpoint for deciding upon the further design at the end of the phase II part, and a long-term clinical endpoint for decision making at the end of the phase III part. Their method is also based on a p-value combination rule. It deals with continuous outcomes and the t-statistics, not with censored survival data.

In the present paper we present a method for data dependent design changes in survival studies, which also makes the full interim information available for design modifications. As an extension of the so-called conditional rejection probability (CRP) principle of Schäfer and Müller (2001), it is a lot more flexibile than possible approaches based on combination tests. At the current stage of development, the method supports a variety of design changes including the extension of the sample size, of the follow-up period, and/or of the number of events, modifications of the inclusion criteria and modifications of the number and time schedule of interim analyses, as long as the first adaptive interim analysis is not scheduled prior to the first one under the initial design. In sum, our method has the following features:

- The study can be designed and started with a conventional statistical test (e.g., an UMPU test) foreseen for the final analysis. It is not necessary to use special combination tests or combination functions. This implies that our method can be applied to make type I error conserving design changes in ongoing survival

studies in which no interim analysis or adaptive data inspection had been foreseen or planned in the protocol.

- It does not require a pre-specification of the end of the learning phase, i.e., of the time point of the adaptive interim data inspection. The decision about the end of the learning phase may be triggered by external information such as the end of a separate pilot study conducted in parallel, and/or it may be made depending on a more or less frequent inspection of the accumulated data. In any case, our method guarantees full control of the type I error rate after design changes.

- When deciding on possible design changes, the researcher is allowed to use all data collected up to the adaptive interim inspection. Specifically, one may use survival information from those patients who have already died before the interim inspection, as well as auxiliary information from patients who are still alive at this time point (censored cases), even if this information is correlated with survival time.

- The method can be applied repeatedly during the course of a study to change the design again if this is necessary.

- It can be applied to group sequential studies designed with a usual group sequential design and it allows for modifications of the number and the time schedule of interim analyses and of the alpha spending function.

To date there is no method which offers this degree of flexibility and the free use of all collected interim information for

the mentioned types of design changes. With the CRP approach of Schäfer and Müller (2001), only the survival status and the survival times observed at the end of the learning phase may be used. The combination test approach of Jenkins et al. (2010) could be extended to incorporate some of the design changes considered in this paper. However, as this approach is based on combination tests, such an extension will not offer the flexibility of our CRP method concerning the time point of the data inspection (end of learning phase), the applicability to studies in which no design adaptation had been pre-planned, and the flexibility to change the schedule and the procedure of interim analyses.

This increased flexibility comes at the price of higher mathematical complexity as compared to combination tests. In fact, it requires the calculation of the conditional distribution of the chosen test statistic, such as the logrank statistic, conditional upon the test statistic of a *subsample* of patients. Using martingale techniques, we derive an approximation of this conditional distribution for the example of the logrank statistic. This implies that our method is based on asymptotic distribution theory when applied with the logrank test. Existing asymptotic theory allows to calculate the distribution of the logrank statistic conditional upon the *interim value of the test statistic at some time point*, which is different from the problem encountered here.

In Section 2, we describe our basic statistical principle. Section 3 provides the asymptotic distribution theory to apply this prin-

ciple with two-armed studies using the logrank statistic. Section 4 contains a practical application of our method. In Section 5, we finally conduct a simulation study of the finite sample behavior of the logrank statistic conditional upon the full survival information of a subsample. Additionally, we estimate the type I error rate of the adapted procedure in the example given in Section 4.

2

AN EXTENDED CRP PRINCIPLE

2.1 METHOD

If φ denotes the statistical decision function defined by the initial design of the trial, the basic idea of the CRP principle (Müller and Schäfer, 2004) is to replace φ by a modified decision function ψ representing the modified study design, such that the conditional error probability condition $\mathbb{E}_0(\varphi|X) = \mathbb{E}_0(\psi|X)$ holds. $\mathbb{E}_0$ denotes the expectation under the null hypothesis and X denotes the set of all interim information used for making the design modification. In the present paper, we will develop and illustrate our method for the example of the logrank statistic and a one-sided test, for sake of simplicity. Let S_k denote the logrank statistic calculated at the time point of the k-th death, defined as the number of failures in group 1 minus the sum of proportions in group 1 of those at risk at observed failure times. By abuse of language, we will call k the *information time* of the study. Then, for the example of a study design with a one-sided fixed sample test, $\varphi = I(S_K > b)$ and $\psi = I(S_{K^*} > b^*)$, where K and K^* denote the total number of deaths to be observed in the initial and modified

study, respectively, and b and b^* denote the critical limits for rejecting the null hypothesis according to the initial and the modified study design, respectively.

Let P_{orig} denote the total sample of patients according to the original study design and define $P' \subset P_{orig}$ the subset of patients recruited up to the interim inspection, which takes place at the calendar time of the k_0-th death among patients in P_{orig}.

In earlier applications of the CRP principle to survival studies, Schäfer and Müller (2001) proposed to use $X = S'_{k_0}$ in the CRP condition mentioned above, where S'_{k_0} denotes the logrank statistic calculated in P' at the time of the adaptive interim inspection. This implies that no information other than S'_{k_0} may be used for design modifications. Strictly speaking, the researchers then have to be blinded with respect to all gathered individual information from the patients, except S'_{k_0}, when making the decision on the design modification.

To overcome this restriction, the basic idea of the extended CRP approach proposed in the present paper, in its full generality, is as follows: The full interim information observed in all the patients recruited up to the interim inspection, X, is made available to the researchers for design changes. However, it may be difficult to calculate the conditional rejection probabilities involved, $\mathbb{E}_0(\varphi|X)$ and $\mathbb{E}_0(\psi|X)$, because the joint distribution of X and φ or ψ cannot be determined. In this case, use a suitable random vector Y for which you can calculate the conditional expectations $\mathbb{E}_0(\varphi|Y)$

and $\mathbb{E}_0(\psi|Y)$ and for which $\mathbb{E}_0(\psi|Y,X) = \mathbb{E}_0(\psi|Y)$ holds true. This is the case when X is stochastically independent of ψ given Y. Then, after the design modification, determine the final decision function under the modified design, ψ, from the generalized CRP condition $\mathbb{E}_0(\varphi|Y) = \mathbb{E}_0(\psi|Y)$ instead of $\mathbb{E}_0(\varphi|X) = \mathbb{E}_0(\psi|X)$. The whole procedure will keep the type I error risk α, which is a consequence of the fundamental property of conditional expectations:

$$\begin{aligned}\mathbb{E}_0(\psi) &= \mathbb{E}_0\big(\mathbb{E}_0(\psi|Y,X)\big) = \mathbb{E}_0\big(\mathbb{E}_0(\psi|Y)\big)\\ &= \mathbb{E}_0\big(\mathbb{E}_0(\varphi|Y)\big) = \mathbb{E}_0(\varphi) = \alpha\end{aligned}$$

In the following, we will work out this extended CRP principle for the logrank statistic. We will show that in this case $Y = S'_{K^*}$ is suitable, i.e. the logrank statistic calculated in P' at the end of the modified design, and we will derive the asymptotic conditional distribution of S_k given the value of S'_k.

The stepwise procedure for logrank test studies is as follows:

1. Start the study with a conventional study design (called the original study design), specifying the total number of patients n to be recruited, the total number of deaths K to be observed, and the critical limit b for the final test statistic. Hence, $\varphi = I(S_K > b)$. The numbers n and K can be specified explicitly in the study protocol, or they can be defined implicitly by specifying the duration of the

recruitment phase and of the follow-up period of the last patient.

2. During the course of the study, the data may be inspected at any time. If a decision for a design modification is made, identify and document the subsample P' of patients recruited so far. Specify the modified design including the extended total number of deaths, K^*, to be observed. The calculation of the modified critical boundary b^* for the final test statistic under the modified design, S_{K^*}, is postponed to the final data analysis.

3. As soon as K^* deaths have been observed, stop the study and determine the generalized CRP $\mathbb{P}_0\left(S_K > b|S'_K\right)$. Then, determine the critical boundary b^* for the final test under the modified design from the CRP condition

$$\mathbb{P}_0\left(S_K > b|S'_K\right) = \mathbb{P}_0\left(S_{K^*} > b^*|S'_{K^*}\right).$$

Note that on the left-hand side, S'_K could formally be replaced by S'_{K^*} because of the well-known property of independent increments of time-sequential log-rank statistics.

4. Determine the observed value of S_{K^*} and reject H_0 if $S_{K^*} > b^*$.

The method can be applied again at a later point of time in exactly the same way. Then, one has to update the sample size K

and the subsample P' to incorporate all patients recruited up to this later time point and restart with Step 2 above.

Our method can also be applied to group sequential designs and allows for modifications of the number and time schedule of interim analyses. However, since the overall conditional rejection probability can only be determined at the end of the study, care has to be taken in order to avoid interim overspending of the conditional type I error. This will impose some limitations on the conditional alpha spending function of the modified design. Suppose that the original group sequential design has m interim analyses at information times (number of events) $k_1, ..., k_m$ with (one-sided, for sake of simplicity) stopping boundaries $b_1, ..., b_m$ for the test statistic S_{k_i} at time k_i. Define

$$CRP_{orig}(k) := \sum_{i=1,...,m} I(k_i \leqslant k) \cdot \mathbb{P}_0\left(S_{k_j} \leqslant b_j \; \forall \, j < i, S_{k_i} > b_i \,\middle|\, S'_{k_j}, j \leqslant i\right),$$

which is the sum of the CRPs at all interim analyses prior to the actual information time k under the initial design. Similarly, let

$$CRP_{mod}(k) := \sum_{i=1,...,m^*} I(k_i^* \leqslant k) \cdot \mathbb{P}_0\left(S_{k_j^*} \leqslant b_j^* \; \forall \, j < i, S_{k_i^*} > b_i^* \,\middle|\, S'_{k_j^*}, j \leqslant i\right)$$

for the modified design. Overspending of conditional type I error can be avoided by choosing the boundaries b_i^* for the time points k_i^* under the modified design such that

$$CRP_{mod}(k) \leqslant CRP_{orig}(k) \text{ for all } k \leqslant \max(k_m, k_{m^*}^*)$$

with $CRP_{mod}(k_{m^*}^*) = CRP_{orig}(k_m)$. In practice, this means that the critical boundaries under the modified design are determined such that the cumulative conditional rejection probability under the modified design is smaller or equal to the cumulative conditional rejection probability for each k.

2.2 PRACTICAL IMPLEMENTATION IN CLINICAL STUDIES

Several precautions are in place and additional measures of study organization and information flow must be taken for a save implementation of our method, in order to avoid bias. The method fundamentally relies on the definition of an initial design, which must be unambigiously specified in the study protocol. This includes a precise *inclusion rule* defining which events contribute to the original decision function φ in the sense that these events will be used to calculate the final value of the test statistic. To avoid bias, it is important that the same inclusion rule is applied whether or not the study design is changed. If a design modification is made, then the inclusion rule of the original design will specify how to calculate the conditional rejection probability. A

second rule (called *termination rule* hereafter) specifies the time point of the final data analysis under the original and under the adapted design.

The calculation of the CRP in case of a design extension requires a third rule (called *assignment rule* herafter) specifying which of the events occuring up to the end of the initial study belong to patients in P' and which belong to patients in P''. To this end, the calendar time of the design change has to be prospectively documented in an amendment to the study protocol, and the set of patients enrolled so far has to be identified (called the "learning set"). No grouped data except from this precisely defined set of patients may be made available to people who are involved in the decision about the design change.

Obviously, the inclusion, assignment and termination rule need to be defined and operated using ungrouped data only. In practical operation, this should be ensured by making the related decisions blinded with respect to treatment groups. The use of ungrouped data implies that the test statistics are stochastically independent from the stopping rule, which guarantees an unbiased estimation of the CRP.

The termination rule can be defined by a number of events. Alternatively, the termination rule may be defined by a calendar time, in which case all events occuring before this calendar time will be used for the calculation of S_K. Even though the number of events in P' and P'' up to the end of the initial study are

random variables in this case, blinding with respect to treatment (i.e., using ungrouped data only) as described above will again warrant stochastic independence of these random event numbers and the test statistics. Then, the CRP and the final test statistics can be unbiasedly calculated conditional upon the number of events observed until the respective calendar time, which turns the event numbers in P' and P'' into deterministic values. In other words, by conditioning upon the number of observed events, the calendar time model can thus be reduced to the usual information time model in which time is measured in terms of the total number of events observed.

3

APPLICATION WITH THE LOGRANK TEST

The CRP condition $\mathbb{P}_0\left(S_K > b \,\middle|\, S'_K\right) = \mathbb{P}_0\left(S_{K^*} > b^* \,\middle|\, S'_{K^*}\right)$ of Step 3 in Section 2.1 implies that the extended CRP principle requires the distribution of S_k conditional on S'_k, which will be derived in this section using martingale theory. Some of the techniques applied here have already been introduced by Sellke and Siegmund (1983) and Olschewski and Schumacher (1986).

3.1 MODEL AND NOTATION

In the remainder, we adapt the notation introduced by Tsiatis (1981), Tsiatis et al. (1985) and Sellke and Siegmund (1983). Let the positive random variable Y denote the time of a patient's entry into the study as measured from the study's onset and let X denote the latent failure time measured from entry, which is censored on the right by a possibly infinite random variable E. In the remainder, Z will denote the treatment indicator, which is either zero or one, representing one treatment or the other. We assume that Z has a mean μ_Z and variance $\sigma_Z^2 := 1/4$ (balanced design).

Using the proportional hazards regression model proposed by Cox (1972), we assume that the hazard rate $\lambda(x)$ for failure at time x is related to Z in a log linear fashion, that is $\lambda(x|z) = \lambda(x)e^{\theta z}$, where $\lambda(x|z)$ denotes the hazard rate at time x given that the treatment indicator Z is equal to z. Under the null hypothesis, $H_0 : \theta = 0$, the distribution function and density function of the failure time X is given by $F(x) := \mathbb{P}(X \leqslant x)$ and $f(x) = dF(x)/dx$, respectively. The hazard rate $\lambda(x)$ is equal to $f(x)/\big(1 - F(x)\big)$ and the cumulative hazard function is denoted by Λ.

During accrual, n individuals enter the study at times $Y_1, Y_2, ..., Y_n$, assumed identically and independently distributed with the distribution function $H(y) = \mathbb{P}(Y \leqslant y)$. This assumption guarantees an increased number of entries in a fixed accrual period as the sample size increases, see Tsiatis et al. (1985, p. 366). Associated with the i-th individual is the vector (X_i, E_i, Y_i, Z_i) $(i = 1, 2, ..., n)$, assumed identically and independently distributed under H_0 (Tsiatis et al. 1985, p. 366).

Define the set of patients at risk at time t and age s by

$$R(t, s) := \{i \,|\, Y_i \leqslant t - s, X_i \wedge E_i \geqslant s\}$$

and let

$$N_i(t, s) := I\{Y_i + X_i \leqslant t, X_i \leqslant E_i, X_i \leqslant s\}$$

describe whether or not patient i has arrived and died before time t, and that he was uncensored and of age $\leqslant s$ at the time of death.

To test the hypothesis

$$H_0 : \theta = 0$$

vs.

$$H_1^+ : \theta > 0, \quad H_1^- : \theta < 0,$$

we employ the sequential form of the logrank statistic

$$\begin{aligned} S_n(t) :&= \sum_{i=1}^{n} \int_0^t \left\{ Z_i - \frac{\sum_{j \in R(t,s)} Z_j}{|R(t,s)|} \right\} N_i(t, ds) \\ &= \sum_{i=1}^{n} \int_0^t \left\{ Z_i - \frac{\sum_{j \in R(t,s)} Z_j}{|R(t,s)|} \right\} M_i(t, ds), \end{aligned}$$

where

$$M_i(t,s) := N_i(t,s) - \Lambda_i \left\{ s \wedge X_i \wedge E_i \wedge (t - Y_i)^+ \right\}.$$

The logrank statistic after k deaths, S_k, as introduced in Section 2 is related to the stochastic process $S_n(t)$ by $S_k := S_n\left(t_{(k)}\right)$, where $t_{(k)}$ denotes the calendar time of the k-th death among patients in P_{orig}.

3.2 CONDITIONAL DISTRIBUTION OF THE LOGRANK STATISTIC

Let k_0, K and K^* denote the information time (number of deaths) of the adaptive interim inspection, the final data analysis under

the initial design and the final data analysis under the modified design, respectively, where $K^* \geqslant K$. Additionally, let T such that $\mathbb{P}\left(T \geqslant t_{(K^*)}\right) = 1$. Also define $P_{orig} := \{1, ..., n\}$, $P' := \left\{i \in P_{orig} \,\middle|\, Y_i \leqslant t_{(k_0)}\right\}$ and $P'' := \left\{i \in P_{orig} \,\middle|\, Y_i > t_{(k_0)}\right\}$.

Let $R'(t,s) := R(t,s) \cap P'$ and $R''(t,s) := R(t,s) \cap P''$ and note that it is

$$S_n(t) = S'_n(t) + S''_n(t) + S'''_n(t), \tag{3.1}$$

where

$$S'_n(t) := \sum_{i \in P'} \int_0^t \left\{ Z_i - \frac{\sum_{j \in R'(t,s)} Z_j}{|R'(t,s)|} \right\} M_i(t, ds),$$

$$S''_n(t) := \sum_{i \in P''} \int_0^t \left\{ Z_i - \frac{\sum_{j \in R''(t,s)} Z_j}{|R''(t,s)|} \right\} M_i(t, ds) \text{ and}$$

$$S'''_n(t) := \sum_{i \in P'} \int_0^t \left\{ \frac{\sum_{j \in R'(t,s)} Z_j}{|R'(t,s)|} - \frac{\sum_{j \in R(t,s)} Z_j}{|R(t,s)|} \right\} M_i(t, ds)$$
$$+ \sum_{i \in P''} \int_0^t \left\{ \frac{\sum_{j \in R''(t,s)} Z_j}{|R''(t,s)|} - \frac{\sum_{j \in R(t,s)} Z_j}{|R(t,s)|} \right\} M_i(t, ds).$$

This subdivision of $S_n(t)$ serves as the basis of the following theorem, which is the main result of this section and which is necessary for the concrete calculation of the CRP under the original and under the adapted design. In the remainder, we denote $D[a,b]$ the space of functions on $[a,b]$ that are right-continuous and have left-hand limits.

Theorem 1. *Let* $\Delta_i(t) := I\left\{X_i < \min\left(E_i, (t - Y_i)^+\right)\right\}$ *denote the censoring indicator and let* $\vartheta'(t) := \mathbb{P}(\Delta_i(t) = 1, i \in P')$ *and* $\vartheta''(t) :=$

$\mathbb{P}(\Delta_i(t) = 1, i \in P'')$ *be continuous as functions in* t *denoting the probability that a death can be observed before time* t *in* P' *and* P''*, respectively. Then under a sequence of local alternatives* $H_{1n} : \theta_n := \delta n^{-\frac{1}{2}}$*,*

$$\frac{1}{\sqrt{n}} \begin{pmatrix} S'_n(\cdot) \\ S_n(\cdot) - S'_n(\cdot) \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} \delta\vartheta'(\cdot)/4 + B'(\vartheta'(\cdot)/4) \\ \delta\vartheta''(\cdot)/4 + B''(\vartheta''(\cdot)/4) \end{pmatrix} \tag{3.2}$$

on $D[0, T]$ *under* $H_{1,n}$ *as* $n \to \infty$ *with independent standard Brownian motions* B' *and* B''*. Here,* $\xrightarrow{\mathcal{L}}$ *denotes convergence in distribution.*

For $t \in [0, T]$, the probabilities $\vartheta'(t)$ and $\vartheta''(t)$ can be consistently estimated by $\widehat{\vartheta}'(t) := \frac{1}{n}\sum_{i=1}^{n} I\{i \in P'\}\Delta_i(t)$ and $\widehat{\vartheta}''(t) := \frac{1}{n}\sum_{i=1}^{n} I\{i \in P''\}\Delta_i(t)$, respectively. Thus under H_0, the distribution of

$$\begin{pmatrix} S'_n(t) \\ S_n(t) - S'_n(t) \end{pmatrix}$$

can be approximated by

$$\mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} n\widehat{\vartheta}'(t)/4 & 0 \\ 0 & n\widehat{\vartheta}''(t)/4 \end{pmatrix}\right)$$

on $[0, T]$. This theorem can be used to calculate the conditional rejection probabilities involved in the generalized CRP equation in Step 3 of Section 2 with $S_k := S_n\left(t_{(k)}\right)$, $S'_k := S'_n\left(t_{(k)}\right)$,

$S''_k := S''_n\left(t_{(k)}\right)$, and $S'''_k := S'''_n\left(t_{(k)}\right)$, where $t_{(k)}$ as defined before denotes the calendar time of the k-th death among patients in P_{orig}.

For the extended CRP principle, it remains to be shown that $\mathbb{E}_0(\psi|Y,X) = \mathbb{E}_0(\psi|Y)$, where $Y = S'_{K^*}$ and X is the full interim information from the subsample P' observed at the time of the adaptive interim inspection. As outlined in paragraph 4 of Section 2, this will allow us to use the data in X to decide upon design adaptations. As $S_{K^*} = S'_{K^*} + S''_{K^*} + S'''_{K^*}$ according to equation (3.1) and $\frac{1}{\sqrt{n}} S'''_{K^*} \xrightarrow{\mathbb{P}} 0$ as shown in the Appendix, X is asymptotically independent of S_{K^*} given the value of S'_{K^*}, since S''_{K^*} does not depend on X. Therefore, $\left|\mathbb{E}_0(\psi(S_{K^*})|S'_{K^*}, X) - \mathbb{E}_0(\psi(S_{K^*})|S'_{K^*})\right| \longrightarrow 0$, implying that also X may be used for design changes without compromizing the type I error rate (assuming an infinitely large sample size). The second simulation experiment of Section 5 illustrates that the use of X does not cause a significant alpha inflation (if any at all) in the example of Section 4, which is based on a sample size that is of practical relevance.

Remark 1. *According to Step 4 of Section 2, the final statistical test under the modified design is of the form $S_{K^*} > s^*$, for some critical limit that depends on the observed value s'_{K^*} of S'_{K^*}. In other words, it is based on the logrank statistic in the complete study sample including the learning and the validation sample. According to Theorem 1, the test can asymptotically be re-written as $S''_{K^*} > \tilde{s}^*$ which is a usual logrank*

test in the subsample P'' *of patients recruited after the adaptive data inspection. Here,* $\tilde{s}^*$ *is a modified critical limit reflecting the survival data from the subsample recruited before the adaptive data inspection,* P'*, and which is determined from the asymptotic unconditional version of the CRP equation,* $\mathbb{P}_0\left(S''_{K^*} > \tilde{s}^*\right) = \mathbb{P}_0\left(S''_K > s - s'_K\right)$*, where* s *is the critical limit of the final statistical test under the initial design and* s'_K *denotes the observed value of* S'_K*.*

4

EXAMPLE

We illustrate our approach for an open-labelled randomized two-stage multicenter trial investigating the effect of combined radiotherapy plus chemotherapy treatment compared to chemotherapy treatment alone in patients with non-small-cell lung cancer (Schäfer and Müller, 2001).

If $\theta := \log(\lambda_C/\lambda_E)$ denotes the log hazard ratio of the control treatment (C) versus the experimental treatment (E), the null hypothesis to be tested is $H_0 : \theta = 0$. The sample size of the trial was calculated based on the assumption of 14 months median survival in the control group (chemotherapy) and 20 months in the experimental group (chemotherapy plus radiotherapy), resulting in a total number of $k_2 = 257$ deaths to be observed (one-sided α=0.025, $\beta = 0.2$, interim analysis at $k_1 = 193$ deaths). Assuming an exponential survival model with a recruitment period of 40 months and a follow-up of 20 months, this implies a sample size of $n = 328$. The design $\varphi = I\{S_{k_i} > b_{k_i} \text{ for one } i = 1, 2\}$ with $b_{k_1} = 16.25208$ and $b_{k_2} = 16.12500$ was found using numerical integration. In our fictitious example, the interim look is performed after $k_0 = 60$ deaths, observed 23 months after

randomization of the first patient. At this point of time, a total number of 190 patients have been recruited. The median survial times observed at the interim look are 13.29 months in the control group and 18.48 months in the experimental group. This results in an estimated log hazard ratio of $\log(0.719)$, which alone would suggest $k_2 = 300$ deaths to be observed. When analyzing the progression-free survival time (PFS) on the other hand, the study group estimated a log hazard ratio of $\log(0.752)$ and thus decides to increase the number of deaths to be observed by $\Delta k_2 = 143$, obtaining $k_1^* = 315$ and $k_2^* = 400$ under the modified design. This implies a sample size of $n^* = 504$ under the adjusted design, which for example might be achieved by recruiting one further participating center without extending the initial recruitment period of 40 months. Since the estimated median survival time in the experimental group is shorter than 20 months, the duration of the follow-up period is not extended either. Now, ψ needs to be determined, which here has the form

$$\psi = I\left\{S_{k_i^*} > b_{k_i^*} \text{ for some } i = 1, 2\right\}.$$

After k_1 deaths in P_{orig}, the study group determines $S'_{k_1} = 16.33873$ and $\widehat{\vartheta}''\left(t_{(k_1)}\right) = 33/n$. As defined before, S'_{k_1} denotes the logrank statistic in the subsample P' calculated at the time of the k_1-th death in P_{orig}. Similarly, $\widehat{\vartheta}''\left(t_{(k_1)}\right)$ is the unbiased estimator of the probability that a death can be observed in P'' before the calender time of the k_1-th death in P_{orig}, as defined in

Theorem 1. The new critical boundary $b_{k_1^*}$ of the first stage of the modified design is now calculated from the equation

$$\mathbb{P}_0\left(S_{k_1^*} > b_{k_1^*} \middle| S'_{k_1^*}\right) = \mathbb{P}_0\left(S_{k_1} > b_{k_1} | S'_{k_1}\right). \tag{4.1}$$

By Theorem 1, with $\xi \sim \mathcal{N}\left(0, \widehat{\vartheta}''\left(t_{(k_1)}\right)/4\right)$,

$$\begin{aligned}
&\mathbb{P}_0\left(S_{k_1} > b_{k_1} | S'_{k_1} = 16.33873\right) \\
&= \mathbb{P}_0\left(S_{k_1} - S'_{k_1} > b_{k_1} - 16.33873 | S'_{k_1} = 16.33873\right) \\
&\approx \mathbb{P}\left(\xi > \frac{b_{k_1} - 16.33873}{\sqrt{n}}\right) = 0.51203.
\end{aligned}$$

At the time of the k_1^*-th death in P_{orig}, the study group finds $S'_{k_1^*} = 22.03081$ and $\widehat{\vartheta}''\left(t_{(k_1^*)}\right) = 131/n^*$ and is now able to determine the new critical boundary $b_{k_1^*}$. By Theorem 1 and equation (4.1),

$$\begin{aligned}
&\mathbb{P}_0\left(S_{k_1^*} > b_{k_1^*} \middle| S'_{k_1^*} = 22.03081\right) \\
&\approx 1 - \Phi\left(\frac{2 \cdot (b_{k_1^*} - 22.03081)}{\sqrt{131}}\right) \\
&\stackrel{!}{=} 0.51203,
\end{aligned}$$

resulting in $b_{k_1^*} = 21.85822$.

The new critical boundary $b_{k_2^*}$ of the second stage of the modified design can be calculated at the time of the second data analysis from

$$\begin{aligned}&\mathbb{P}_0\left(S_{k_1^*} \leqslant b_{k_1^*}, S_{k_2^*} > b_{k_2^*} \middle| S'_{k_1^*}, S'_{k_2^*}\right)\\&= \mathbb{P}_0\left(S_{k_1} \leqslant b_{k_1}, S_{k_2} > b_{k_2} | S'_{k_1}, S'_{k_2}\right).\end{aligned} \tag{4.2}$$

At the time of the k_2-th death in P_{orig}, the study group finds $S'_{k_2} = 21.12618$ and $\widehat{\vartheta}''\left(t_{(k_2)}\right) = 78/n$. By Theorem 1 and the fact that $\operatorname{cov}\left(S_{k_1} - S'_{k_1}, S_{k_2} - S'_{k_2}\right) \approx n \cdot \widehat{\vartheta}''\left(t_{(k_1)}\right)/4$,

$$\begin{aligned}&\mathbb{P}_0\left(S_{k_1} \leqslant b_{k_1}, S_{k_2} > b_{k_2} | S'_{k_1} = 16.33873, S'_{k_2} = 21.12618\right)\\&\approx 0.37212.\end{aligned}$$

At the time of the k_2^*-th death in P_{orig}, the study group finally observes $S'_{k_2^*} = 22.09059$ and $\widehat{\vartheta}''\left(t_{(k_2^*)}\right) = 215/n^*$. It is now able to determine the new critical boundary $b_{k_2^*}$ of the second stage, as $b_{k_1^*}$ is already known (see above): By equation (4.2),

$$\begin{aligned}&\mathbb{P}_0\left(S_{k_1^*} \leqslant b_{k_1^*}, S_{k_2^*} > b_{k_2^*} \middle| S'_{k_1^*} = 22.03081, S'_{k_2^*} = 22.09059\right)\\&\stackrel{!}{=} 0.37212,\end{aligned}$$

resulting in $b_{k_2^*} = 13.46469$.

According to Remark 1, this example could have also been carried out equivalently using tests of the form $S''_k > \tilde{s}^*_k$.

5

SIMULATION STUDY

Using R Development Core Team (2009), we calculated the empirical distribution of $\frac{1}{\sqrt{n}}\left\{S_n\left(t_{(k)}\right) - S'_n\left(t_{(k)}\right)\right\}$. Simulations were performed under Weibull distributed survival times with two different baseline hazards defined by the Weibull parameters $(\mu, \sigma^2) = (400, 200^2)$ (model 1) and $(\mu, \sigma^2) = (600, 300^2)$ (model 2). In both models, we suppose a log hazard ratio $\theta = \log(14/17)$ and apply a uniform censoring distribution based on a recruitment period of 1,200 days and a follow-up period of 600 days. In model 1, we consider $\left(n, k_0, k, \widehat{\vartheta}''\left(t_{(k)}\right)\right) = (50, 10, 40, 13/50)$ and $\left(n, k_0, k, \widehat{\vartheta}''\left(t_{(k)}\right)\right) = (300, 60, 240, 81/300)$, where the two values 13 and 81 of $\widehat{\vartheta}''\left(t_{(k)}\right) = \frac{1}{n}\sum_{i \in P''} \Delta_i\left(t_{(k)}\right)$ have been pre-determined. As defined in Theorem 1, $\widehat{\vartheta}''\left(t_{(k)}\right)$ is the observed probability that a death can be observed in P'' before the time point of the k-th death in P_{orig}. In model 2, we consider $\left(n, k_0, k, \widehat{\vartheta}''\left(t_{(k)}\right)\right)$ $= (50, 10, 40, 12/50)$ and $\left(n, k_0, k, \widehat{\vartheta}''\left(t_{(k)}\right)\right) = (300, 60, 240, 48/300)$.

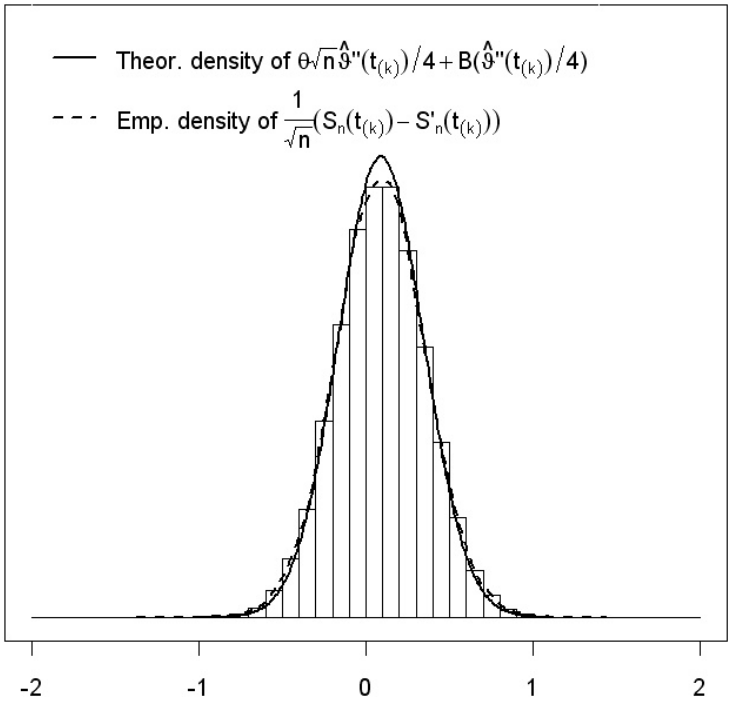
With the notation introduced in Table 1, simulations are performed as follows: We resample the survival information of patients in P_{orig} over and over again, discarding every replication of P_{orig} in which the observed value of $\widehat{\vartheta}''\left(t_{(k)}\right)$ does not

correspond to the pre-determined one (see above). A total of 100.000 suitable replications of P_{orig} is then used to estimate the empirical distribution of $\frac{1}{\sqrt{n}}\left\{S_n\left(t_{(k)}\right) - S'_n\left(t_{(k)}\right)\right\}$, which is then compared with the theoretical distribution of $\theta\sqrt{n}\widehat{\vartheta}''\left(t_{(k)}\right)/4 + B\left(\widehat{\vartheta}''\left(t_{(k)}\right)/4\right)$. Results are presented in figures 1 and 2. With $n = 50$, both distributions are already very close, whereas accuracy is even greater for $n = 300$.

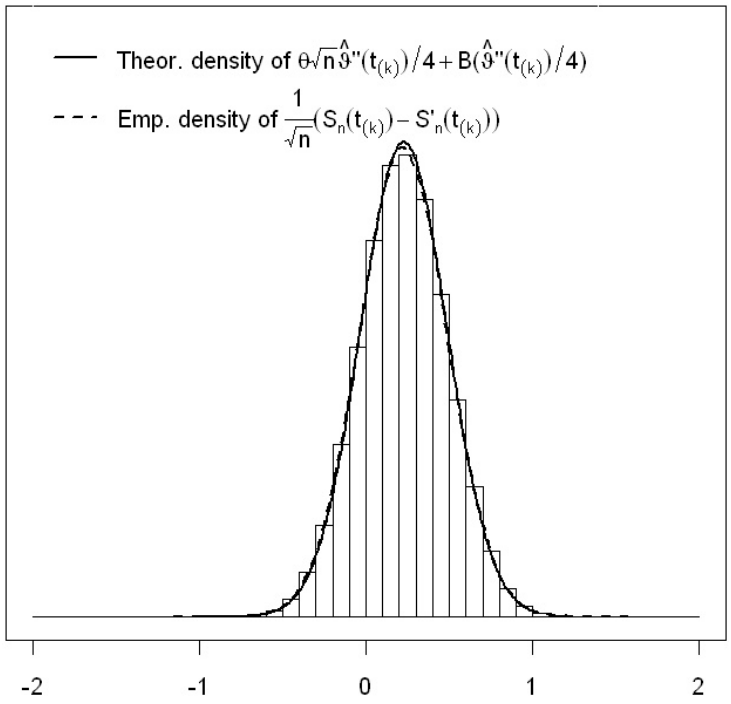
In a further experiment, we calculate the empirical type I error rate of the adjusted group sequential design for the example already studied in the previous Section 4. We assume that the effect size θ is estimated at the time of the interim data inspection after k_0 deaths based on the observed progression-free survival (PFS) time and that $\operatorname{corr}\left(\exp(\widehat{\theta}_{PFS}), \exp(\widehat{\theta}_{OS})\right) = 0.9990$, where $\widehat{\theta}_{OS}$ denotes the estimated effect size calculated at the time of the final data analysis under the adapted design based on the data from patients in P'. A sample size extension was only made if $\exp(\widehat{\theta}_{PFS})$ differed from the originally assumed hazard ratio of 14/20 by more than 30%, i.e. $\exp(\widehat{\theta}_{PFS}) > 1.3 \cdot 14/20$. In this case, the necessary number of additional patients was calculated on the basis of $\widehat{\theta}_{PFS}$. Based on 100.000 repetitions, we found $\widehat{\alpha} = 0.02527$ with a 95%-confidence interval of $[0.02433, 0.02623]$ under the adapted design.

Based on the second experiment described above, we additionally compare our extended CRP principle with the traditional one introduced by Schäfer and Müller (2001) in terms of power of

the adjusted design. As expected, the increased flexibility of the extended CRP principle is paid by a loss in power of the adjusted design compared to the one under the traditional CRP principle. However, this loss in power is very small and over-compensated in situations in which the treatment difference is relatively small and the additional patient information not available under the traditional CRP principle can be used to notably improve the estimation of the log hazard ratio θ. Results are presented in figure 3.

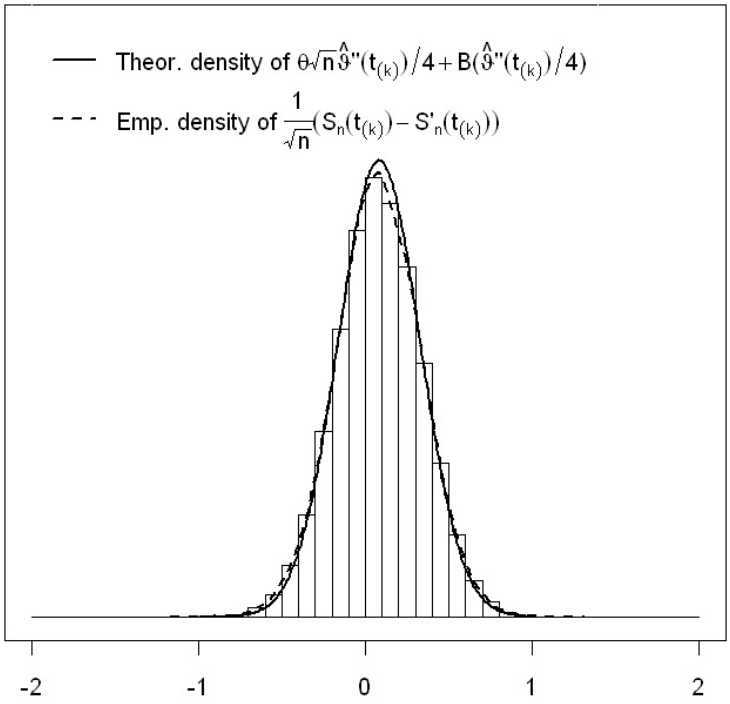


Model 1, $\nu = (50, 10, 40, 13/50)$.

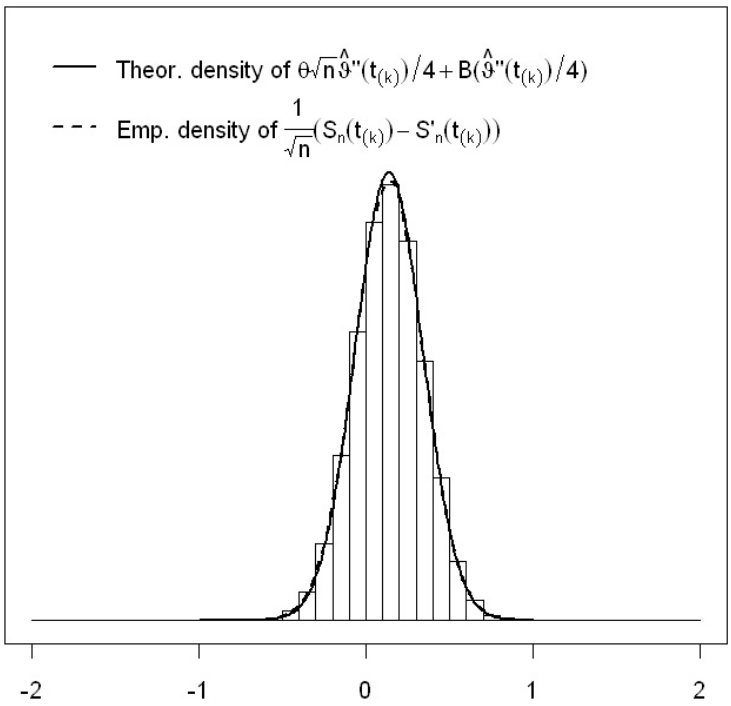


Model 1, $\nu = (300, 60, 240, 81/500)$.

Figure 1: Convergence in Model 1

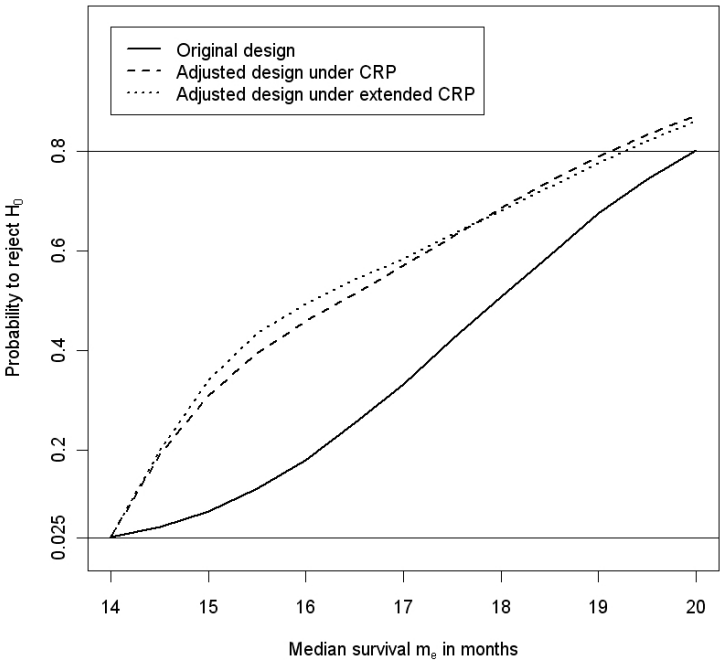


Model 2, $\nu = (50, 10, 40, 12/50)$.



Model 2, $\nu = (300, 60, 240, 48/300)$.

Figure 2: Convergence in Model 2



Power of the adjusted design for different hazard ratios $\exp(\theta) = 14/m_e$.

Figure 3: Comparison of the traditional and the extended CRP principle.

6

DISCUSSION

We here present an approach for design modification in survival studies, which makes the full interim patient information available for design changes. Based on the CRP principle, it is more flexible than possible combination test approaches. In particular, in contrast to combination tests, the study can be started with a conventional statistical test, and it does not require the time point of the interim inspection to be pre-specified, i.e. the length of the learning phase can be chosen dependent on the data. These advantages come at the price of higher mathematical complexity as the distribution of the employed test statistic conditional on the test statistic of a subsample must be known. We have derived this for the example of the logrank test.

In its current stage of development, our approach allows the extension of the recruitment or follow-up period, as well as certain modifications of the number and the schedule of interim analyses as well as changes of inclusion criteria. Further work is needed to extend the CRP method to more complex design changes like subgroup selection in phase II/III seamless designs,

for which Jenkins et al. (2010) have recently developed a method based on combination tests.

Although the CRP principle introduced here is highly flexible and can even be used to make type I error conserving design changes in ongoing studies for which no design adaptation has been foreseen or specified in the protocol, we nevertheless strongly recommend to make corresponding specifications in the protocol, such as the intended time points of adpative data inspections and the details of how the CRP principle will be applied in the study. Moreover, we emphasize that flexible design methods should not be a reason to reduce the efforts of a careful planning of the trial. Design changes in ongoing trials depending on data inspections will always imply some loss of power as compared to the optimal design if this had been chosen from the beginning. However, as every experienced clinical trialist knows, the practical conduct of clincial trials and especially of long term clinical trials rarely complies exactly with our plans. Thus, important reasons may arise during the course of the study to change the design due to unforeseen deviations from the assumptions made in the planning phase. In this case, our method offers the option to make the necessary design changes. Herby, it is a highly flexible method and it is simple, as it can be applied to usual statistical tests. In sum, our recommendation for a survival study under the proportional hazards assumption would be to plan and to start it with a sequential logrank test, at best fulfilling

defined statistical optimality properties such as minimal expected number of events for ethical reasons, and to implement the CRP principle in the protocol as a method to make design changes if this turns out to be necessary.

For an unbiased implementation of our method, unambigous rules for the end of the study according to the original and the modified must be documented. When the end of the study is defined by a certain calendar date, instead of the number of events to be observed, the set of events to be included in the final analysis according to the original and according to the modified design should be recorded using ungrouped data only, i.e., blinded with respect to treatment. Whenever a design change is made, the set of patients recruited up to this point in time (the "learning set" P') must be identified and documented for an unambiguous calculation of the condition S'_K at the end of the study.

A

APPENDIX A: GENERAL RESULTS

In the following, we will briefly provide a collection of key definitions and results on stochastic processes in the space $D[a,b]$ of functions on $[a,b]$ that are right-continuous and have left-hand limits. These results are well-known and necessary for the proof of Theorem 1 in Section 3, which will be formally shown in Appendix B.

Definition 1. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and M a metric space. Denote $\mathcal{B}(M)$ the Borel σ-algebra of M. Then, $X : \Omega \to M$ is called a random element in M, if $X^{-1}(\mathcal{B}(M)) \subset \mathcal{A}$, see Billingsley (1968, Chapter 1.4).*

Definition 2. *Denote $\mathbb{P}^{X_n}$ the probability measure of the random element X_n. The sequence $\left(\mathbb{P}^{X_n}\right)_{n\geqslant 1}$ is said to converge weakly against $\mathbb{P}^X$ on the metric space M, if*

$$\lim_{n\to\infty} \int_M f d\mathbb{P}^{X_n} \longrightarrow \int_M f d\mathbb{P}^X$$

for all bounded and continuous functions $f : M \to \mathbb{R}$.

Definition 3. *Define $D[a,b]$ the space of functions f on $[a,b]$ that are right-continuous and have left-hand limits:*

1. *For* $a \leqslant t < b, f(t+) = \lim_{s \searrow t} f(s)$ *exists and it is* $f(t+) = f(t)$.

2. *For* $a < t \leqslant b, f(t-) = \lim_{s \nearrow t} f(s)$ *exists,*

see Billingsley (1968, Chapter 3).

Definition 4. *Let* $\Lambda[a, b]$ *denote the class of strictly increasing, continuous mappings of* $[a, b]$ *onto itself with* $\lambda(a) = a$ *and* $\lambda(b) = b$. *For* $f, g \in D[a, b]$ *let*

$$s(f, g) := \inf_{\lambda \in \Lambda[a,b]} \{\|f - g \circ \lambda\| + \|\lambda - Id\|\},$$

where $\|\cdot\|$ *denotes the supremum norm. Then,* s *is a metric (Billingsley 1968, Chapter 3.14) and the pair* $(D[a, b], s)$ *is called* ***Skorokhod space****.*

Definition 5. *Let* $X_n, X \in D[a, b]$. *We say that* $(X_n)_{n \geqslant 1}$ *converges in distribution to* X *as* $n \to \infty$ *(we write* $X_n \xrightarrow{\mathcal{L}} X$*), if the sequence* $\left(\mathbb{P}^{X_n}\right)_{n \geqslant 1}$ *converges weakly towards* $\mathbb{P}^X$ *according to Definition 2.*

Lemma 1. *Let* X_n, $X \in D[a, b]$. *If*

- $X_n \xrightarrow{\mathcal{L}} X$ *and*

- X *has continuous paths,*

then $\left(X_n(t_1), ..., X_n(t_k)\right) \xrightarrow{\mathcal{L}} \left(X(t_1), ..., X(t_k)\right)$ *for* $a \leqslant t_1 \leqslant ... \leqslant t_k \leqslant b$.

Lemma 2. *Let* $X_n, X \in D[a, b]$ *such that*

- X *has continuous paths,*

- $\big(X_n(t_1),...,X_n(t_k)\big) \xrightarrow{\mathcal{L}} \big(X(t_1),...,X(t_k)\big)$ *for* $a \leqslant t_1 \leqslant ... \leqslant t_k \leqslant b$ *and*
- $\lim\limits_{\delta \searrow 0} \limsup\limits_{n\to\infty} \mathbb{P}\left(\sup\limits_{\substack{|t-t'|\leqslant\delta \\ t,t'\in[a,b]}} |X_n(t) - X_n(t')| > \varepsilon \right) = 0$ *for all* $\varepsilon > 0$.

Then, $X_n \xrightarrow{\mathcal{L}} X$ *in* $D[a,b]$.

The proof of Lemma 1 and Lemma 2 is omitted here.

Lemma 3. *Consider a sequence of processes* $(X_{1,n}, X_{2,n}, ..., X_{r,n})_{n\geqslant 1}$ *and a vector of limiting processes* $(X_1,...,X_r)$. *Then*

$$(X_{1,n}, X_{2,n}, ..., X_{r,n}) \xrightarrow{\mathcal{L}} (X_1,...,X_r) \text{ in } (D[a,b])^r$$

if and only if

$$\sum_{l=1}^{r} \int_0^{\cdot} c_l(s)\,dX_{l,n}(s) \xrightarrow{\mathcal{L}} \sum_{l=1}^{r} \int_0^{\cdot} c_l(s)\,dX_l(s) \text{ in } D[a,b]$$

for any bounded left-continuous step functions c_l *on* $[a,b]$, *where* $l = 1,...,r$.

Proof. See Fleming and Harrington (1991, Lemma C.3.1). ❑

In case that $X_{l,n}$ is a semimartingale for each n, the following Lemma gives sufficient conditions such that $\int_0^{\cdot} c_l(s)\,dX_{l,n}(s) \xrightarrow{\mathcal{L}} \int_0^{\cdot} c_l(s)\,dX_l(s)$:

Lemma 4. *Let* $(X_n)_{n\geqslant 1}$ *be a sequence of* $\mathbb{R}^d$*-valued semimartingales with decompositions* $X_n = M_n + A_n$ *such that*

$$\sup_{n\geqslant 1}\left\{\mathbb{E}_n\big([M_n, M_n](t)\big) + \mathbb{E}_n\left(\int_0^t |dA_n(s)|\right)\right\} < \infty,$$

each $t > 0$*. Further assume that* $(H_n, X_n) \xrightarrow{\mathcal{L}} (H, X)$*, where* $(H_n)_{n\geqslant 1}$ *is a* $d \times k$ *matrix process. Then,*

$$\int_0^{\cdot} H_n(s-)dX_n(s) \xrightarrow{\mathcal{L}} \int_0^{\cdot} H(s-)dX(s) \text{ in } D[a, b].$$

Proof. See Kurtz and Protter (1996), Definition 7.3, equation (7.12) and Theorem 7.10. ❑

B

APPENDIX B: PROOF OF THEOREM 1

Proof of Theorem 1. First of all, note that $S_n(t) - S_n'(t) = S_n''(t) + S_n'''(t)$ according to equation (3.1). Following Sellke and Siegmund (1983, p. 317), let $\mathcal{F}_s$ be the σ-algebra generated by

$$Y_i, E_i, Z_i, I\{X_i \leqslant s\}, X_i \cdot I\{X_i \leqslant s\} \text{ for } i = 1, 2, ...$$

for $s \geqslant 0$ and let $\mathcal{F}_{t,s}$ be the sub-σ-algebra of $\mathcal{F}_s$ generated by

$$I\{Y_i \leqslant t\}, Y_i \cdot I\{Y_i \leqslant t\}, I\left\{X_i \leqslant s \wedge E_i \wedge (t - Y_i)^+\right\},$$
$$X_i \cdot I\left\{X_i \leqslant s \wedge E_i \wedge (t - Y_i)^+\right\},$$
$$I\left\{E_i \leqslant s \wedge X_i \wedge (t - Y_i)^+\right\},$$
$$E_i \cdot I\left\{E_i \leqslant s \wedge X_i \wedge (t - Y_i)^+\right\}, Z_i, \ (i = 1, 2, ...).$$

Then according to Sellke and Siegmund (1983, p. 318), $M_i(t, s)$ is a $\mathcal{F}_{t,s}$-martingale in s for each fixed t. However, since the score processes $S_n'(t)$ and $S_n''(t)$ fail to be martingales in t due to the dependence of their integrands on t, we cannot apply standard martingale techniques directly to derive their asymptotic joint

distribution at different values of t, such as the martingale central limit theorem. Therefore, we define

$$S_n'^*(t) := \sum_{i \in P'} \int_0^t \{Z_i - \mathbb{E}(Z)\} M_i(t, ds)$$

and

$$S_n''^*(t) := \sum_{i \in P''} \int_0^t \{Z_i - \mathbb{E}(Z)\} M_i(t, ds),$$

which are both $\mathcal{F}_{t,t}$-martingales in t (Sellke and Siegmund 1983, p. 318), and show that under H_0 the score processes $S_n'(t)$ and $S_n''(t) + S_n'''(t)$ can be approximated uniformly in time by $S_n'^*(t)$ and $S_n''^*(t)$, respectively.

In order to apply the martingale central limit theorem to $S_n'^*(t)$ and $S_n''^*(t)$, we need to ensure that the quadratic variation of each process grows approximately linearly in t, see Sellke and Siegmund (1983, p. 319).

For $S_n'^*(t)$ and $S_n''^*(t)$ this linear growth does not occur in the time scale t, but under a data-dependent transformation of time:

Let $\vartheta'^{-1}(\tau) := \inf\{t\,|\vartheta'(t) = \tau\}$ and $\vartheta''^{-1}(\tau) := \inf\{t\,|\vartheta''(t) = \tau\}$ with $\vartheta'(t)$ and $\vartheta''(t)$ defined in Theorem 1. Then,

$$\begin{aligned}
&\left[\frac{1}{\sqrt{n}}S_n'^*\right]\left(\vartheta'^{-1}(\tau)\right)\\
&= \frac{1}{n}\sum_{i\in P'}\int_0^{\vartheta'^{-1}(\tau)}\{Z_i - \mathbb{E}(Z)\}^2\, N_i\left(\vartheta'^{-1}(\tau), ds\right)\\
&= \frac{1}{n}\sum_{i\in P'}\{Z_i - \mathbb{E}(Z)\}^2\,\Delta_i\left(\vartheta'^{-1}(\tau)\right)\\
&\xrightarrow{\mathbb{P}} \sigma_Z^2\tau = \frac{\tau}{4}
\end{aligned}$$

for all $\tau \in [0, \vartheta'(T)]$ and similar for $S_n''^*$ and the transformation $\vartheta''^{-1}(\tau)$ for all $\tau \in [0, \vartheta''(T)]$, where the first equation follows from Andersen et al. (1995, p. 84) and the second one from Tsiatis (1981, p. 312). Hence by Rebolledo (1980, p. 273, Proposition 1),

$$\frac{1}{\sqrt{n}}S_n'^*\left(\vartheta'^{-1}(\cdot)\right) \xrightarrow{\mathcal{L}} B'(\cdot/4)$$

and

$$\frac{1}{\sqrt{n}}S_n''^*\left(\vartheta''^{-1}(\cdot)\right) \xrightarrow{\mathcal{L}} B''(\cdot/4)$$

in $D[0, \vartheta'(T)]$ and $D[0, \vartheta''(T)]$, respectively. Thus by Lemma 1 of Appendix A,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left(S_n'^*\left(\vartheta'^{-1}(\tau_1')\right), ..., S_n'^*\left(\vartheta'^{-1}(\tau_k')\right)\right)\\
&\xrightarrow{\mathcal{L}} \left(B'\left(\tau_1'/4\right), ..., B'\left(\tau_k'/4\right)\right)
\end{aligned} \tag{B.1}$$

and

$$\begin{aligned}&\frac{1}{\sqrt{n}}\left(S_n''^{*}\left(\vartheta''^{-1}(\tau_1'')\right),...,S_n''^{*}\left(\vartheta''^{-1}(\tau_l'')\right)\right)\\&\xrightarrow{\mathcal{L}}\left(B''\left(\tau_1''/4\right),...,B''\left(\tau_l''/4\right)\right)\end{aligned}\tag{B.2}$$

for $0 \leqslant \tau_1' \leqslant ... \leqslant \tau_k' \leqslant \vartheta'(T)$ and $0 \leqslant \tau_1'' \leqslant ... \leqslant \tau_l'' \leqslant \vartheta''(T)$, where B' and B'' are standard Brownian motions.

Lemma 5. *Under a sequence of contiguous alternatives* $H_{1n} : \theta_n = \delta n^{-1/2}$,

$$\begin{aligned}&\frac{1}{\sqrt{n}}\Big(S_n'\big(\vartheta'^{-1}(\tau_1')\big)-S_n'^{*}\big(\vartheta'^{-1}(\tau_1')\big),...,\\&\qquad S_n'\big(\vartheta'^{-1}(\tau_k')\big)-S_n'^{*}\big(\vartheta'^{-1}(\tau_k')\big)\Big)\xrightarrow{\mathbb{P}}\Big(\frac{\delta\tau_1'}{4},...,\frac{\delta\tau_k'}{4}\Big),\end{aligned}\tag{B.3}$$

$$\begin{aligned}&\frac{1}{\sqrt{n}}\Big(S_n''\big(\vartheta''^{-1}(\tau_1'')\big)-S_n''^{*}\big(\vartheta''^{-1}(\tau_1'')\big),...,\\&\qquad S_n''\big(\vartheta''^{-1}(\tau_l'')\big)-S_n''^{*}\big(\vartheta'^{-1}(\tau_l'')\big)\Big)\xrightarrow{\mathbb{P}}\Big(\frac{\delta\tau_1''}{4},...,\frac{\delta\tau_l''}{4}\Big)\end{aligned}\tag{B.4}$$

and

$$\frac{1}{\sqrt{n}}\Big(S_n'''\big(\vartheta''^{-1}(\tau_1'')\big),...,S_n'''\big(\vartheta''^{-1}(\tau_l'')\big)\Big)\xrightarrow{\mathbb{P}}(0,...,0)\in\mathbb{R}^l\tag{B.5}$$

for $0 \leqslant \tau_1' \leqslant ... \leqslant \tau_k' \leqslant \vartheta'(T)$ *and* $0 \leqslant \tau_1'' \leqslant ... \leqslant \tau_l'' \leqslant \vartheta''(T)$.

Please refer to the proof of Lemma 5 below. Thus by Slutsky's theorem,

$$\begin{aligned}&\frac{1}{\sqrt{n}}\Big(S'_n\big(\vartheta'^{-1}(\tau'_1)\big),...,S'_n\big(\vartheta'^{-1}(\tau'_k)\big)\Big)\\&\xrightarrow{\mathcal{L}}\Big(\frac{\delta\tau'_1}{4}+B'\left(\tau'_1/4\right),...,\frac{\delta\tau'_k}{4}+B'\left(\tau'_k/4\right)\Big),\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{\sqrt{n}}\Big(S''_n\big(\vartheta''^{-1}(\tau''_1)\big)+S'''_n\big(\vartheta''^{-1}(\tau''_1)\big),...,\\&\qquad S''_n\big(\vartheta''^{-1}(\tau''_l)\big)+S'''_n\big(\vartheta''^{-1}(\tau''_l)\big)\Big)\\&\xrightarrow{\mathcal{L}}\Big(\frac{\delta\tau''_1}{4}+B''\left(\tau''_1/4\right),...,\frac{\delta\tau''_l}{4}+B''\left(\tau''_l/4\right)\Big)\end{aligned}$$

under H_{1n}. Note that $\forall\,\eta,\varepsilon>0\;\exists\,\delta>0\;\exists\,n_0\geqslant 1\;\forall\,n\geqslant n_0$:

$$\mathbb{P}\Big(\sup_{\substack{|\tau-\tilde{\tau}|<\delta\\ \tau,\tilde{\tau}\in[0,\vartheta'(T)]}}\left|S'_n\big(\vartheta'^{-1}(\tau)\big)-S'_n\big(\vartheta'^{-1}(\tilde{\tau})\big)\right|>\varepsilon\Big)\leqslant\eta$$

and

$$\begin{aligned}\mathbb{P}\Big(\sup_{\substack{|\tau-\tilde{\tau}|<\delta\\ \tau,\tilde{\tau}\in[0,\vartheta''(T)]}}\Big|&S''_n\big(\vartheta''^{-1}(\tau)\big)+S'''_n\big(\vartheta''^{-1}(\tau)\big)\\&-\Big\{S''_n\big(\vartheta''^{-1}(\tilde{\tau})\big)+S'''_n\big(\vartheta''^{-1}(\tilde{\tau})\big)\Big\}\Big|>\varepsilon\Big)\leqslant\eta\end{aligned}$$

and that $\delta\tau/4 + B'(\tau/4)$ has continuous paths. Hence by Lemma 2 of Appendix A,

$$\frac{1}{\sqrt{n}} S_n'\left(\vartheta'^{-1}(\cdot)\right) \xrightarrow{\mathcal{L}} \delta \cdot \mathrm{Id}_{[0,\vartheta'(T)]}(\cdot)/4 + B'(\cdot/4)$$

and

$$\frac{1}{\sqrt{n}} \left\{ S_n''\left(\vartheta''^{-1}(\cdot)\right) + S_n'''\left(\vartheta''^{-1}(\cdot)\right) \right\} \xrightarrow{\mathcal{L}} \delta \cdot \mathrm{Id}_{[0,\vartheta''(T)]}(\cdot)/4 + B''(\cdot/4)$$

in $D[0, \vartheta'(T)]$ and $D[0, \vartheta''(T)]$, respectively, where Id_A denotes the identity function on a set A.

Using Lemma 3 and Lemma 4 of Appendix A, it can finally be shown that

$$\frac{1}{\sqrt{n}} \begin{pmatrix} S_n'(\cdot) \\ S_n''(\cdot) + S_n'''(\cdot) \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} \delta\vartheta'(\cdot)/4 + B'(\vartheta'(\cdot)/4) \\ \delta\vartheta''(\cdot)/4 + B''(\vartheta''(\cdot)/4) \end{pmatrix}$$

in $D[0, T]$. ❑

Proof of Lemma 5. According to Schoenfeld (1981), $\frac{1}{\sqrt{n}}\{S_n(t) - S_n^*(t)\} \xrightarrow{\mathbb{P}} \frac{\delta\vartheta(t)}{4}$ pointwise for $t \in [0, T]$, where

$$\vartheta(t) = \mathbb{P}\left(\Delta_i(t) = 1, i \in P_{\mathrm{orig}}\right)$$

$S_n(t)$ as defined in Section 3.1 and

$$S_n^*(t) := \sum_{i \in P_{\mathrm{orig}}} \int_0^t \{Z_i - \mathbb{E}(Z)\} M_i(t, ds).$$

These results are also true for $S_n'(t) - S_n'^*(t)$ and $\vartheta'(t)$ as well as $S_n''(t) - S_n''^*(t)$ and $\vartheta''(t)$, which immediately implies (B.3) and (B.4). Finally, (B.5) follows from the weak law of large numbers.

❑

BIBLIOGRAPHY

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1995). *Statistical Models Based on Counting Processes (Springer Series in Statistics)*. Springer.

Bauer, P. and Köhne, K. (1994). Evaluation of experiments with adaptive interim analyses. *Biometrics*, 50(4):1029–1041.

Bauer, P. and Posch, M. (2004). Letter to the editor: Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections. *Statistics in Medicine*, 23:1333–1334.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley and Sons Inc.

Brannath, W., Zuber, E., Branson, M., Bretz, F., Gallo, P., Posch, M., and Racine-Poon, A. (2009). Confirmatory adaptive designs with bayesian decision tools for a targeted therapy in oncology. *Statistics in Medicine*, 28:1145–1463.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, 34(2):187–220.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. John Wiley & Sons.

Irle, S. and Schäfer, H. (2012). Interim design modifications in time-to-event studies. *Journal of the American Statistical Association*.

Jahn-Eimermacher, A. and Ingel, K. (2009). Adaptive trial design: A general methodology for censored time to event data. *Contemporary Clinical Trials*, 30(2):171 – 177.

Jenkins, M., Stone, A., and Jennison, C. (2010). An adaptive seamless phase ii/iii design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharmaceutical Statistics*.

Kurtz, T. and Protter, P. (1996). Weak convergence of stochastic integrals and differential equations. In *Probabilistic Models for Nonlinear Partial Differential Equations*, volume 1627 of *Lecture Notes in Mathematics*, pages 1–41. Springer Berlin / Heidelberg.

Lehmacher, W. and Wassmer, G. (1999). Adaptive sample size calculations in group sequential trials. *Biometrics*, 55(4):1286–1290.

Liu, Q. and Pledger, G. W. (2005). Phase 2 and 3 combination designs to accelerate drug development. *Journal of the American Statistical Association*, 100:493–502.

Liu, Q., Proschan, M. A., and Pledger, G. W. (2002). A unified theory of two-stage adaptive designs. *Journal of the American Statistical Association*, 97:1034–1041.

Müller, H.-H. and Schäfer, H. (2004). A general statistical principle for changing a design any time during the course of a trial. *Statistics in Medicine*, 23:2497–2508.

Olschewski, M. and Schumacher, M. (1986). Sequential analysis of survival times in clinical trials. *Biometrical Journal*, 28(3):273–293.

Proschan, M. A. and Hunsberger, S. A. (1995). Designed extension of studies based on conditional power. *Biometrics*, 51(4):1315–1324.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rebolledo, R. (1980). Central limit theorems for local martingales. *Probability Theory and Related Fields*, 51:269–286. 10.1007/BF00587353.

Schäfer, H. and Müller, H.-H. (2001). 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.

Schoenfeld, D. (1981). The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*, 68(1):316–319.

Sellke, T. and Siegmund, D. (1983). Sequential analysis of the proportional hazards model. *Biometrika*, 70(2):315–326.

Tsiatis, A. A. (1981). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika*, 68:311–315.

Tsiatis, A. A., Rosner, G., and Tritchler, D. L. (1985). Group sequential tests with censored survival data adjusting for covariates. *Biometrika*, 72(2):365–373.

Wassmer, G. (2006). Planning and analyzing adaptive group sequential survival trials. *Biometrical Journal*, 48(4):714–729.

C

ACADEMIC TEACHERS

My academic teachers at the University of Gießen have been:

- Prof. Dr. Bernd Baumann
- Prof. Dr. Wolfgang Bessler
- Prof. Dr. Rüdiger Kabst
- apl. Prof. Dr. Bernhard Lani-Wayda
- Prof. Dr. Jürgen Meckl
- Prof. Dr. Wolfgang Scherf
- Prof. Dr. Dr. h.c. Winfried Stute
- Prof. Dr. Barbara E. Weißenberger

At the University of Marburg, my academic teacher has been Prof. Dr. Helmut Schäfer.

D

ACKNOWLEDGEMENTS

This research project would not have been possible without the support of my supervisor, Prof. Dr. Helmut Schäfer, who was abundantly helpful and offered invaluable assistance, support and guidance.

My parents, Burkhard and Anne Irle, have always been my dear friends, a constant source of support and my personal guiding light, and this thesis would certainly not have existed without them.