
Adaptive Designs for Group Sequential Clinical Survival Experiments

Eric V. Slud

Statistics Program, University of Maryland College Park, USA

Abstract: Randomized two-group clinical survival experiments now commonly allow at least one interim look, enabling possible early stopping to meet ethical concerns. Various authors have also studied the possibility of interim design modifications to adapt to unexpected accrual or control-group mortality rates. This paper formulates trial design as a decision theoretic problem with a finite number of interim looks and a large class of loss functions, in the setting of a statistic with the asymptotic behavior of Brownian motion with drift, as in Leifer and Slud (2002). A more general action space can specify adaptive designs allowing the option of continued followup without new accrual past an interim look, as was introduced in Koutsoukos, Rubinstein and Slud (1999). An optimal two-look design is displayed in the first formulation, and a seven-look design in the second, and both types of adaptation are given a unified decision-theoretic motivation.

Keywords and phrases: Accrual stopping; Bayesian decision theory; Nonrandomized decision rule; Loss function; Nuisance parameter; Stopping boundary

1.1 Introduction

Group sequential designs are designs in which experimental data on two-group treatment comparisons can be scrutinized at a finite number of interim look-times with the possibility of early termination of the experiment in such a way as to maintain a prescribed experimentwise significance level and power against a fixed alternative. Such designs first appeared for two-group randomized clinical trials with normally distributed quantitative responses in the mid-1960's. By the late 1970's, methods had appeared which took explicit account of the staggered entry, followup time, and delayed response of clinical trials

with survival-time endpoints. By the early 1980's, such methods were firmly established theoretically. Tsiatis (1982) showed that the repeatedly computed logrank-numerator statistic at a series of fixed scheduled interim look-times would under standard conditions behave in large two-sample trials as a sequence of independent-increment Gaussian variables, with mean 0 under the null hypothesis \mathbf{H}_0 of no treatment effect and with steady positive drift proportional to variance under local proportional-hazard alternatives. As the theory developed, look-times were allowed to be random (stopping-times for the observed information process), and additional classes of statistics including weighted logrank statistics (with weight functions also estimated from pooled two-group Kaplan-Meier survival function estimators) were justified to be usable in the same way as the logrank, although the logrank is the heavy practical favorite.

Slud and Wei (1982) showed how variance increments could be progressively estimated while allowing early stopping by means of an α -*spending schedule*. In a (one-sided) trial of sample size n , with the statistic S_k/\sqrt{n} calculated at the k 'th look-time t_k , a threshold or boundary b_k is used to stop the trial early with rejection of \mathbf{H}_0 if $S_k/\sqrt{n} \geq b_k$, where b_k is found inductively, in terms of the estimated large-sample variance V_k of S_k/\sqrt{n} , to satisfy

$$\alpha_k = Pr(S_j/\sqrt{n} < b_j \quad \text{for } 1 \leq j < k, \quad S_k/\sqrt{n} \geq b_k) \quad (1.1)$$

where the values $\alpha_1, \dots, \alpha_K$ are prescribed and sum to the experimentwise significance level α . The times at which interim looks might be taken can be allowed to be random stopping-times, e.g., to be level-crossing times for the proportional-hazard parameter's *information*, which is proportional to the logrank variance and thus also to the number of observed failure events. Moreover, the choice of the specific value α_k need not be made until the $k-1$ th look-time (Lan and DeMets 1983). The asymptotic theory underlying this extension was given by Slud (1984) and other authors, establishing that under local (contiguous) proportional-hazard alternatives the repeatedly computed logrank statistic considered as a stochastic process behaves asymptotically in large samples like a time-changed Brownian motion with drift. The history of these developments from the viewpoint of trial design, along with practical recommendations on the choice among early-stopping designs as of 1984, can be found in Fleming et al (1984). The context of these results in the setting of repeated significance testing within exponential families can be found in Siegmund (1985).

Later progress on the specification of early-stopping boundaries included generalizations beyond our scope here (more general statistics, adjustment for covariates, modified formulations of repeated significance testing, etc.), but also developed optimization methods: Tsiatis and co-authors restricted attention to parametrically restricted families of boundaries and computed the ones which minimized expected trial duration over boundaries with prescribed size and average power against specified alternatives; while Jennison (1987) undertook

a brute-force (grid-search) computation of optimal boundaries in the sense of minimizing a weighted linear combination of type-II error probabilities and expected sample sizes over specified alternatives, for given significance level.

Clinical investigators often find at the times of interim looks in clinical trials that planned accrual goals have not been met, or that due to noncompliance, lower than expected tolerated doses, or better survival than expected in the control group, power will be less than planned for against clinically meaningful alternatives. For this and other, ethical, reasons, there has been a perceived need for *adaptive* (group-) sequential trial designs accommodating flexibility in accrual rates through the spacing of look times. However, the designs must explicitly take account of such flexibility: Proschan et al. (1992) nicely illustrate the adverse effects on significance level of modifying look-time definitions and other trial assumptions in mid-trial. Various authors (Bauer and Köhne 1994, Proschan and Hunsberger 1995, and others cited in Burman and Sonesson 2006) have proposed methods of accommodating design changes (usually in sample size) after an interim look, resulting in procedures with valid experimentwise significance levels based on weighted combinations of statistic-increments calculated up to and after the design changes. But there is active controversy (see Burman and Sonesson 2006, with discussion) concerning whether such adaptations are a good idea, or are even ethical, considering the loss of power they entail against the originally envisioned alternatives.

The accelerating pace of biomedical discovery due to the genomics revolution, discussed by P.K. Sen (2006) in this volume, highlights the dramatic opportunity costs from protracted clinical trials and from incorrect decisions based on them. A principled statistical response should take account of those costs, as well as the important ethical costs that arise from clinical trial errors. The approach followed in this paper is decision-theoretic. We consider clinical trial designs which ‘adapt’ to interim results subject to experimentwise type I and II error probability constraints, in such a way as to minimize the expected values of realistically structured loss functions.

This paper has three objectives: first, in Section 1.2, to describe the Bayesian decision problem of Leifer and Slud (2002) incorporating multi-look trials with general loss components, penalizing trial length and incorrect decisions as a function of the treatment-group difference parameter ϑ ; second (Section 1.3), to describe how optimal decision procedures require later look-times and stopping-boundaries to depend on earlier observed statistic values, especially in the two-look case; and third, to describe a decision problem (in Sec. 1.2.2) motivating new design elements including those of Koutsoukos et al. (1999) described in Section 1.4, allowing group-sequential trials an option to stop accrual with or without early stopping, while maintaining experimentwise significance level.

1.2 Decision Theoretic Formulation

Many theoretical results (Tsiatis 1982, Slud 1984, Siegmund 1985) justify that the sequence two-sample (weighted-)logrank statistics calculated at interim looks of a multi-look staggered-accrual clinical trial with survival endpoints under local proportional-hazard alternatives (and also more general classes of alternatives) is asymptotically equivalent in large datasets to sampled values of a Wiener process with drift, $X(t) = W(t) + \vartheta t$. Here ϑ is an unknown real parameter quantifying positive or negative relative prognosis for treatment-versus control-group patients in the trial. The natural time-scale for estimation of the treatment difference parameter ϑ is *information time* (Siegmund 1985, Andersen et al. 1993), i.e. the information about ϑ in the data up to time t . Increments of time are transformed by this statistical time-scale, regarded as a function of nuisance parameters under near-null alternatives (i.e., those with $\vartheta \approx 0$). The nuisance parameters — all statistical parameters of the accrual, censoring, and survival mechanisms of the trial other than ϑ — are assumed to be consistently estimated at times of interim analysis of the data.

The objective of the trial is inference on ϑ to distinguish the null hypothesis $\vartheta \leq 0$ against alternatives with $\vartheta > 0$: process data $X(\tau_j)$ may be observed (only) at an increasing sequence of discrete times τ_j , $1 \leq j \leq K$, with τ_j allowed to be determined from $(\tau_i, X(\tau_i))$, $i < j$ (and, possibly, auxiliary randomizations independent of the data). The upper-bound K on the number of look times is generally nonrandom and fixed, and the trial ends at the first time τ_ν for which either $\nu = K$ or $\tau_{\nu+1} = \tau_\nu$, at which time a binary decision $\chi \in \{0, 1\}$ is made as a function of all observable data $(\tau_i, X(\tau_i))$, $i \leq \nu$. When actions $(\tau_i, 1 \leq i \leq \nu)$ and χ have been taken, losses are measured in terms of $\tau_\nu = t$ and $\chi = z \in \{0, 1\}$, when ϑ is the correct alternative (drift) parameter assumed distributed according to a prior distribution π on \mathbf{R} , by

$$L(t, z, \vartheta) = \begin{cases} c_1(t, z, \vartheta) + z c_2(t, \vartheta) + (1 - z) c_3(t, z, \vartheta), & \text{if } \vartheta \leq 0, \\ c_1(t, z, \vartheta) + (1 - z) c_2(t, \vartheta) + z c_3(t, z, \vartheta), & \text{if } \vartheta > 0. \end{cases} \quad (1.2)$$

Here z denotes the indicator of rejection of the null hypothesis $\mathbf{H}_0 : \vartheta \leq 0$. The functions c_1 , c_2 , and c_3 represent respectively the costs of trial duration; of incorrect terminal decision; and of correct, but late, terminal decision. These costs are general enough to apply to realistic clinical trial scenarios, both from the point of view of public health and of the drug developer. The interim looks are not assigned direct costs, because data-monitoring committees do in any case monitor the interim results of clinical trials for treatment-safety issues and ethically driven early stopping.

The cost functions are assumed to be π -integrable for each (t, z) , nondecreasing and piecewise smooth in t , and to satisfy for all (t, z, ϑ) :

$$c_1(0, \vartheta) = c_3(0, \vartheta) = 0 \quad , \quad c_3(t, z, \vartheta) < c_2(t, \vartheta). \quad (1.3)$$

In addition, π is assumed to place positive mass in small neighborhoods of $\vartheta = 0$ and $\vartheta = \vartheta_1 > 0$, and $c_1(\cdot, z, \vartheta)$ is assumed to grow to ∞ for $z = 0, 1$ and π -almost all ϑ .

In this setting, the decision problem is to choose decision rules

$$\delta = (\{\tau_j\}_{j=1}^K, \nu, \chi) \quad (1.4)$$

to minimize the expected loss or *risk function*

$$r(\delta) = \int E_{\vartheta}(L(\tau_{\nu}, \chi, \vartheta)) d\pi(\vartheta) \quad (1.5)$$

subject for fixed $\alpha, \beta > 0$ to the type I and II error probability constraints

$$E_{\vartheta=0}(\chi) \leq \alpha \quad , \quad E_{\vartheta=\vartheta_1}(1 - \chi) \leq \beta \quad (1.6)$$

where $\vartheta_1 > 0$ is a fixed alternative deemed to be sufficiently distant from the null-hypothesis value $\vartheta = 0$ to be a medically significant treatment difference.

This decision theoretic problem is the one defined by Leifer and Slud (2002). It can be analyzed, standardly, in terms of Lagrange multipliers (Berger 1985) so that the constraints (1.6) are omitted and the loss-function is replaced (after a reduction showing there is no loss of generality in assuming $\pi_0 \equiv \pi(\{0\}) > 0$ and $\pi_1 \equiv \pi(\{\vartheta_1\}) > 0$) by

$$L_{\lambda_0, \lambda_1}(t, z, \vartheta) \equiv L(t, z, \vartheta) + \frac{\lambda_0}{\pi_0} I_{[\vartheta=0]} + \frac{\lambda_1}{\pi_1} I_{[\vartheta=\vartheta_1]}. \quad (1.7)$$

Up to this point, ‘adaptivity’ of the clinical trial design is embodied in the flexibility of actions $(\{\tau_j\}_{j=1}^K, \chi)$: since data are re-examined at all of the look-times $\tau_1, \dots, \tau_{\nu}$, especially good or bad performance of the treatment group can lead to early decision (rejection or acceptance of H_0 with $\nu < K$), and nuisance parameters such as accrual rates and control group survival distribution can be re-estimated. Flexibility of clinical trial design has two aspects: first, that action-space coordinates permit decision at many possible times, but second, that investigators’ actions defining the times of later interim looks at the data may depend functionally on aspects of the nuisance parameters which do not directly appear in the reward or cost functions driving the trial, but which do affect expected costs.

While the interim look-times $\{\tau_i, 1 \leq i \leq K\}$ that are designed into a clinical trial add logistical complexity, they can be justified not only because of the economy in expected sample size and other costs common to all sequential methods (Siegmund 1985), but also because of the range of surprises — lower than expected accrual, or higher than expected treatment-group survival, reflected in a lower than expected rate of increase of statistical information about ϑ as a function of calendar time — under which the trial can still achieve desired experimentwise type I and type II error probabilities.

This last feature of group-sequential trials is underappreciated. In the starkest comparison, that between sequential and fixed-sample trials, the group-sequential trial is not only more economical on average under expected rates of accrual, but more robust in maintaining acceptable power under a variety of erratic accrual rates and other unexpected trial characteristics leading to slow increase of information with calendar time. We formulate the issue mathematically in the following brief subsection.

1.2.1 Inference in a Random-Information Environment

As before, let the action-space consist of elements $(\{\tau_j\}_{j=1}^K, \chi)$ — stopping-times and final binary decision — based upon data available only at times τ_j from a Wiener process with drift $X(A(t)) = W(A(t)) + \vartheta A(t)$, conditionally given the variance function $A(t)$. However, we now view $A(\cdot)$ itself as a smoothly increasing random function of time, with $A(t)$ observable at time t , and with a known or conjectured probability law μ_A but with trajectory definitely *not* known in advance. Note that the conditional statistical information about ϑ given $A(t)$, based on any subset of the data-history $(X(s), s \leq t)$ which includes the observation $X(t)$, coincides with $A(t)$. This is our idealized model of all of the surprises in a clinical trial which may have an impact on statistical information about ϑ . The observability of $A(t)$ at look-time t corresponds to the large-sample estimability of nuisance parameters.

Assume that the loss-function for the inference problem is exactly as given in (1.2)–(1.3) and either (1.6) or (1.7), but even a very simple cost-structure such as $c_1(t, \vartheta) = t$, $c_2(\vartheta) = 0 = c_3(t, \vartheta)$ can be used to convey the main idea. The prior probability $\pi(\cdot)$ law for ϑ must now be coupled with a probability law μ_A for $A(\cdot)$ regarded as an independent ‘parameter’ of the decision problem. Then risks (1.5) must be replaced by expectations taken jointly over $(\vartheta, A(\cdot))$ with respect to the product prior measure $\pi \times \mu_A$, in order to define the problem anew as a Bayesian decision problem, and again the constraints (1.6) are replaced by Lagrange multipliers when L in (1.5) is replaced by (1.7). A slightly different form of the decision-theoretic problem would average in Bayesian fashion over $d\pi(\vartheta)$ but treat unknown $A(\cdot)$ in minimax fashion, i.e., to replace (1.5) by its maximum or supremum over information functions $A(\cdot)$ in some class.

The point of this subsection is that an essentially nondeterministic and unknown $A(\cdot)$ makes even an otherwise ‘fixed-sample’ procedure — one with $K = 1$ and τ_1 deterministic — depend on an uncertain amount of statistical information. Evidently a group-sequential procedure which makes use of one or more looks at the data to estimate features of a very uncertain $A(\cdot)$ can be found which will outperform such a fixed-sample procedure: the logistical costs of interim looks should be borne in order that adequate power be available under an ensemble of possible trajectories for $A(\cdot)$. The desirability of a multi-look

procedure would derive from the requirement for robustness against alternative models $A(\cdot)$, even if we would be satisfied, scientifically and ethically, with a fixed-sample one-look procedure under the assumptions of uniform accrual and approximately constant withdrawal-rates and control-group failure rate often used to derive the power of clinical trials.

It seems likely that the observability of $A(\cdot)$ at look-times t will often force optimal Bayes decision procedures to allow more interim look-times than they would under the same cost-structures with deterministic $A(\cdot)$. Finding simple examples in which this can be proved is an attractive current research problem. In any case, it is intuitively clear that in problems with random $A(\cdot)$, independent of ϑ and observable at interim look-times, all optimal Bayes decisions must necessarily, under mild restrictions on μ_A , have number ν of interim looks equal to the upper bound K with positive probability. (If not, we could reason by contradiction and find a procedure which alters an existing decision rule δ by including an interim look before τ_1 , thereby narrowing the class of possible $A(\cdot)$ trajectories, and later employs this information to improve on δ .)

1.2.2 Extended Actions affecting Information Growth

We can imagine other sorts of violations of standard clinical-trial assumptions that still other aspects of design flexibility might overcome. For example, there might be time-trends in patient prognosis, occurring in random fashion but with overall effect estimable at interim look-times. This kind of violation of the usual *iid* assumptions about accrued patients will again have the effect of randomizing the information process $A(t)$: if these trends have no effect on treatment difference, then nothing new is needed beyond the formulation of the previous subsection. However, we can also imagine that at interim looks, a pattern of nonconstant treatment-to-control group hazard ratios might begin to emerge, such as an indication of treatment-group differences occurring only at later times-on-test. In that case, a new degree of design freedom might be desirable: to prolong followup of already-accrued patients without allowing any new patients to be accrued. Here again the motivation might be not primarily power under standard conditions but robustness of trial size and power characteristics under nonstandard ones.

One obstacle to the exercise of this kind of design freedom is to show how a group-sequential design might allow an option to terminate accrual but extend followup, while maintaining a fixed experimentwise significance level. This kind of ‘proof of concept’ was offered by the design of Koutsoukos et al. (1999) described in Section 1.4 below.

1.3 2-Look Optimal Decision Rules

Leifer and Slud (2002, revised 2006) show that optimal Bayesian decision rules for (1.7), in the standard setting of Section 1.2 without randomized information function or observable data other than X , have the following properties:

1. There is a finite, nonrandom constant $t_* > 0$, which may be made uniform with respect to compact sets of pairs $(\lambda_0, \lambda_1) \in \mathbf{R}_+^2$, such that $\tau_\nu \leq t_*$.
2. For each triple (α, β, r) lying on the (closed) lower boundary of the 3-dimensional convex set of triples

$$\left(E_{\vartheta=0}(\chi), E_{\vartheta=\vartheta_1}(\chi), \int E_{\vartheta}(L(\tau_\nu, \chi, \vartheta)) \pi(d\vartheta) \right) \quad (1.8)$$

of randomized decision rules, there exists a possibly randomized decision rule for which (α, β, r) is exactly equal to the triple (1.8).

3. Every minimum-risk, possibly randomized, decision rule for the decision problem with the loss-function (1.7) has a terminal decision χ which is a.s. equal to a nonrandom function of the form $\chi = I_{[X(\tau_\nu) \geq w(\tau_\nu)]}$, with $w(\cdot)$ uniquely defined implicitly through the equation

$$\int a_1(y, \lambda_0, \lambda_1, \theta) e^{\theta w(y) - \theta^2 y/2} \pi(d\theta) = 0$$

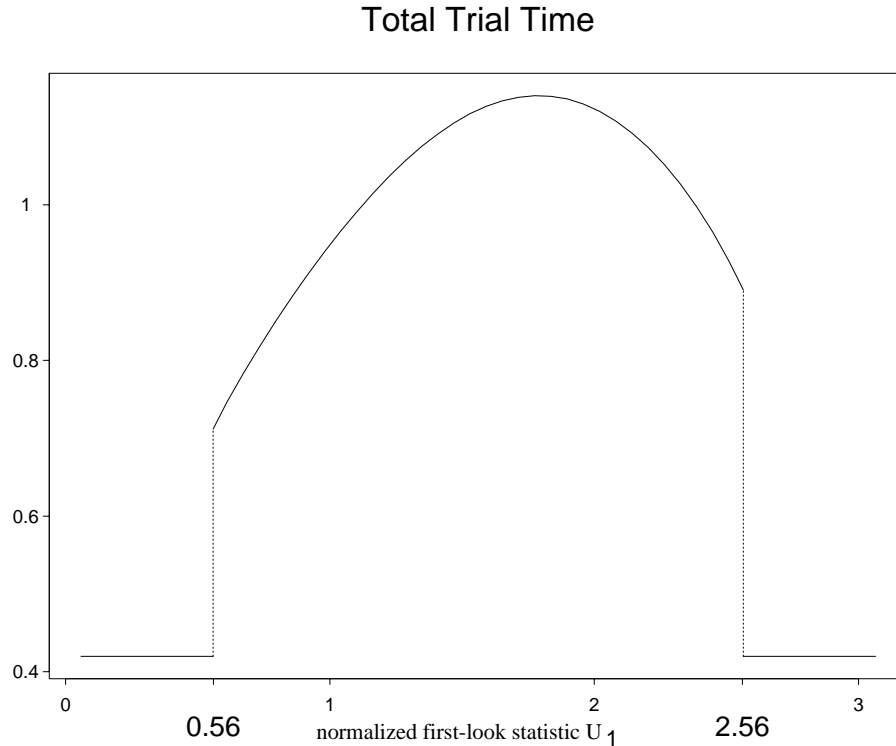
where

$$a_1(z, \lambda_0, \lambda_1, \theta) = (c_2(\theta) - c_3(z, \theta)) \cdot (2I_{[\theta \leq 0]} - 1) + \frac{\lambda_0}{\pi_0} I_{[\theta=0]} - \frac{\lambda_1}{\pi_1} I_{[\theta=\theta_1]}$$

4. Generically for the loss-function (1.7), that is, after a small random perturbation of the cost-function c_1 preserving the Assumptions, for almost every pair (λ_0, λ_1) , the optimal decision rule minimizing the Bayesian risk for loss function (1.7) is unique and nonrandomized and can be computed by backward induction.

We provide an example of such an optimized nonrandomized Bayes 2-look decision rule, taken from Leifer and Slud (2002). Consider $\alpha = .025$, $\beta = .1$, and $\vartheta_1 = \log(1.5)$, with time scaled so that a fixed-sample ($K = 1$) trial with this size and type II error probability has duration $\tau_1 = 1$. We exhibit an optimal rule, with $K = 2$, for the discrete prior and loss-function defined (after taking $c_3(t, \vartheta) \equiv 0$) through the Table

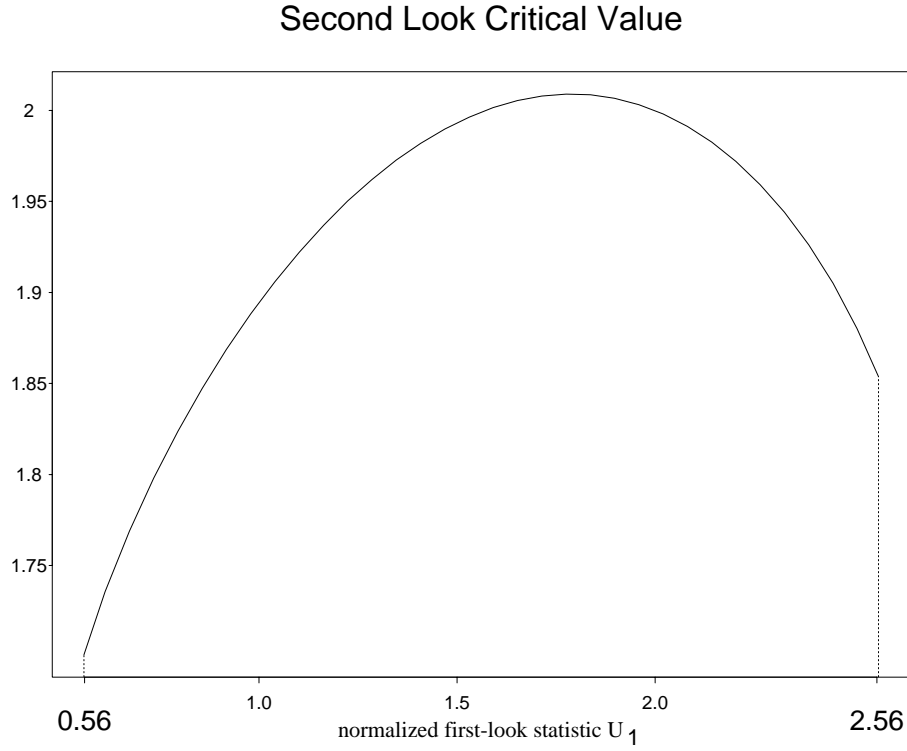
Figure 1.1: Second look time τ_2 in Example of Sec. 1.3, for fixed $\tau_1 = 0.42$, as a function of normalized statistic $U_1 = X(\tau_1)/\sqrt{\tau_1}$.



$e^\vartheta = \text{hazard ratio}$	0.9	1.0	1.25	1.5	1.75
$1.51 \cdot \pi(\{\vartheta\})$	0.2	1.0	0.2	0.1	0.01
$c_1(t, \vartheta)$	t	t	t	t	t
$c_2(\vartheta)$	200	100	50	250	500

The optimized (nonrandomized) procedure has three elements: an initial look-time at $\tau_1 = .42$, a second look-time $\tau_2(U_1)$ defined as a function of $U_1 \equiv X(\tau_1)/\sqrt{\tau_1}$ and displayed in Figure 1.1, and a final rejection-boundary $b(U_1)$ displayed in Figure 1.2 defining the rejection indicator as $\chi = I_{[X(\tau_2)/\sqrt{\tau_2} \geq b(U_1)]}$. These functions do completely describe the group-sequential procedure: the time- τ_1 rejection and acceptance boundaries are determined in Fig. 1.1 through the observation that termination with $\nu = 1$ occurs whenever $\tau_2 = \tau_1$, i.e. when $U_1 < 0.56$ or $U_1 > 2.56$, and the time- τ_1 decision (rejection-indicator) χ is 1 on $[U_1 > 2.56]$ and 0 on $[U_1 < 0.56]$.

Figure 1.2: Rejection boundary at τ_2 as a function of $U_1 = X(\tau_1)/\sqrt{\tau_1}$ in optimized 2-look procedure of Sec. 1.3.



1.4 Modified Trial Designs with Accrual-Stopping

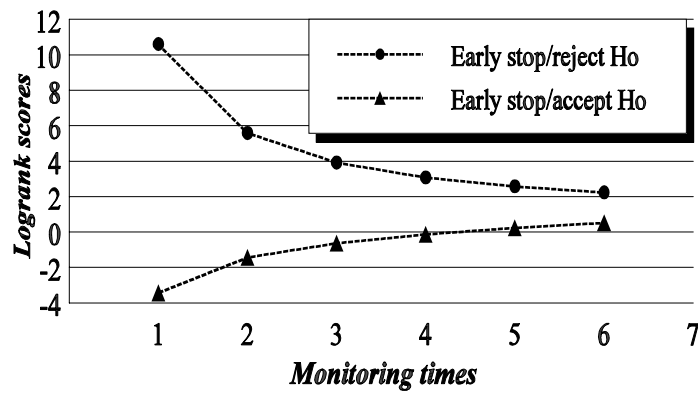
We conclude by describing a clinical trial design of Koutsoukos et al. (1999) extending that of Section 1.2, which allows the flexibility of modifying accrual without stopping followup, effectively reducing, but not to 0, the rate at which information about the survival difference parameter ϑ unfolds. (Also, as we have seen in our motivating discussion in Section 1.2.2, the continuing increment of information relates to later times on test, which may be particularly valuable information under some circumstances.) The notation concerning the repeatedly calculated statistic S_j/\sqrt{n} with (estimated) variance V_j is as in the Introduction. In this design, the look-times $\tau_j = j$ are evenly spaced, since at most one time-unit of further followup is allowed when accrual is stopped, and at the end of such a followup period the trial is stopped. Immediate termination of the trial, respectively with acceptance or rejection, is determined by extreme boundaries $C_{U,j}$ and $C_{L,j}$ of fixed shape (here, $C_{U,j} = c_1$ is constant and $C_{L,j}$ of the form $C_L + c_0 V_j$ for suitably chosen constants C_L, c_0); but accrual is stopped when S_j/\sqrt{n} crosses a less extreme boundary $C_{A,j} < C_U$ (or in any case, when $j = K - 1$), and rejection or acceptance is determined at the look-time following accrual-termination by a different boundary $C_{R,j}$.

The trial is stopped outright at j , with Rejection, if $S_j/\sqrt{n} \geq C_U$, and with Acceptance of H_0 , if $S_j/\sqrt{n} \leq C_{L,j}$.

The accrual (i.e. entry) of new patients is disallowed at time j if $C_{A,j} \leq S_j/\sqrt{n} < C_U$, in which case the trial is stopped at time $j + 1$, with final Rejection if $S_{j+1}/\sqrt{n} \geq C_{R,j+1}$ and Acceptance otherwise.

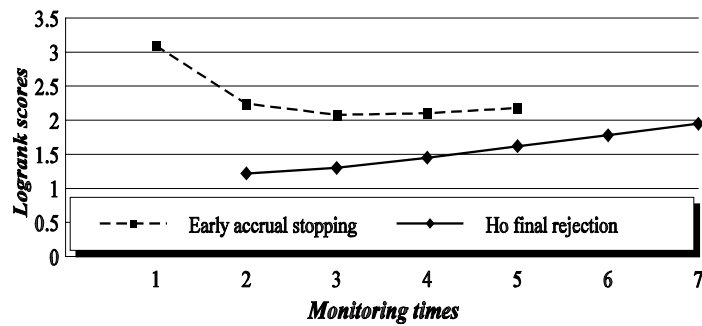
Boundaries of this type can be computed to have fixed size and power against a fixed alternative, and the free parameters in $C_U, C_{A,j}, C_{R,j+1}$ can be optimized with respect to a loss function containing costs for wrong decisions and trial durations under a range of alternatives weighted by a prior π . Details of calculation of such optimized boundaries can be found in Koutsoukos et al. (1999). An example of the resulting boundaries, in a trial with experimentwise significance level $\alpha = .025$ and power .8 against a hazard-ratio alternative of 1.4, are exhibited in Figures 1.3 and 1.4, which are taken from slides prepared by L. Rubinstein. In these Figures, the boundaries plotted are those corresponding to the normalized statistics $S_j/\sqrt{nV_j}$, and the variances V_j for logrank statistics S_j/\sqrt{n} were calculated under an assumption of constant failure-rates for a two-armed clinical trial with patient arrivals following a homogeneous Poisson process.

Figure 1.3: Immediate rejection and acceptance boundaries, respectively $C_U/\sqrt{V_j}$ (plotted with filled dots) and $C_{L,j}/\sqrt{V_j}$ (filled triangles) for normalized logrank statistics $S_j/\sqrt{nV_j}$ in a particular case of the Koutsoukos et al. (1998) boundaries described in Sec. 1.4.



- The upper bound is a standard O'Brien-Fleming upper bound for a 7 look design, with $\alpha=.025$.
- The lower bound is an asymmetric lower bound taken from an O'Brien-Fleming 6 look design, yielding .05 probability of crossing for $\Delta=1.4$ (for which the trial has power 80%).

Figure 1.4: Accrual-stopping and final rejection boundaries, respectively $C_{A,j}/\sqrt{V_j}$ (plotted with filled squares) and $C_{R,j}/\sqrt{V_j}$ (filled diamonds) for normalized logrank statistics $S_j/\sqrt{nV_j}$ in the same example of the Koutsoukos et al. (1998) boundaries as in Fig. 1.3.



- The H_0 final rejection (1-sided) bound is optimized with respect to power against a defined mix of alternative hypotheses.
- The early accrual stopping bound and rejection bound, together, yield .05 conditional probability of failure to reject H_0 after stopping accrual at the boundary (assuming a defined mix of alternative hypotheses).

References

1. Andersen, P., Borgan, O. Gill, R., and Keiding, N. (1993). *Statistical Methods Based on Counting Processes*, Springer-Verlag, New York.
2. Bauer, P. and Köhne, K. (1994). Evaluation of experiments with adaptive interim analyses, *Biometrics*, **50**, 1029-1041.
3. Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York.
4. Burman, C.-F. and Sonesson, C. (2006). Are flexible designs sound ?, *Biometrics*. **62**, 664-669, including Discussion.
5. Fleming, T., Harrington, D. and O'Brien, P. (1984). Designs for group sequential tests, *Controlled Clinical Trials*, **5**, 348-361.
6. Hald, A. (1975). Optimum double sampling tests of given strength I: the normal distribution, *Jour. Amer. Statist. Assoc.* **70**, 451-456.
7. Jennison, C. (1987). Efficient group sequential tests with unpredictable group sizes, *Biometrika*, **77**, 577-513.
8. Koutsoukos, A., Rubinstein, L. and Slud, E. (1999). Early accrual-stopping sequential designs for clinical trials. *US National Cancer Institute*, preprint.
9. Lan, G. and DeMets, D. (1983). Discrete sequential boundaries for clinical trials, *Biometrika*, **70**, 659-663.
10. Leifer, E. and Slud, E. (2002, rev. 2006). Optimal time-adaptive repeated significance tests. Preprint.
11. Liu, Q., Proschan, M. and Pledger, G. (2002). A unified theory of two-stage adaptive designs, *Jour. Amer. Statist. Assoc.* **97**, 1034-1041.
12. Proschan, M., Follmann, D. and Waclawiw, M. (1992). Effects of assumption violations on Type I error rate in group sequential monitoring, *Biometrics*, **51**, 1315-1324.
13. Proschan, M. and Hunsberger, S. (1995). Designed extension of studies based on conditional power, *Biometrics* **51**, 1315-1324.
14. Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*, Springer-Verlag, New York.
15. Slud, E. (1984). Sequential linear rank tests for two-sample censored survival data, *Annals of Statistics*, **12**, 551-571.

16. Slud, E. and Wei, L. J. (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic, *Journal of the American Statistical Association*, **77**, 862-868.
17. Tsiatis, A. (1982). Repeated significance testing for a general class of statistics used in censored survival analysis, *Journal of the American Statistical Association*, **77**, 855-861.