Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Sequential tests of promise with discrete time-to-event data

Bruce Levin[a,*], Louise Kuhn[b], Cheng-Shiun Leu[a], Wei-Yann Tsai[a]

[a] Department of Biostatistics, Columbia University, Mailman School of Public Health, 722 West 168th Street, New York, NY 10032, USA
[b] Department of Epidemiology, Columbia University, Mailman School of Public Health, 722 West 168th Street, New York, NY 10032, USA

ARTICLE INFO

Keywords:
ART cessation trials
Futility design
Non-superiority design
Simon's two-stage design
Survival data

ABSTRACT

We introduce a family of sequential test procedures in the context of a futility study design, or as we prefer to call it, a formal test of promise, suitable for use with time-to-event data. The procedures are motivated by an actual trial that was undertaken to test the promise of very early antiretroviral therapy to achieve viral remission in infants with perinatally-acquired HIV. Important gains in efficiency are illustrated in terms of early stopping and statistical power compared with other methods such as Simon's two-stage design with binary outcomes. We show how to calculate the operating characteristics of the proposed sequential tests of promise and provide optimal or near-optimal boundaries for small or medium size samples which provide the typical context for the tests under consideration. The design features discussed in this article are also of immediate pertinence to trials designed to test disease cures which may require treatment interruption and small numbers of participants.

1. Introduction

Early phase clinical studies can be used to screen new treatments or interventions for promise of efficacy or lack thereof. A prototype example is Simon's two-stage design for testing whether a new drug holds promise for continuing into large-scale confirmatory trials or whether, in the alternative, the treatment lacks promise [12]. Study designs of this type, these days often called *futility designs*, are arguably best suited for such screening purposes. We do not refer here to stopping a trial "for futility" during interim monitoring of a large-scale confirmatory trial. In the screening context of an early phase study, for reasons explained below, the futility design posits a pre-specified degree of positive or superior efficacy for the experimental intervention and adopts it as the null hypothesis, allowing the data possibly to reject that hypothesis, suggesting that further experimentation in the face of such unpromising evidence would be "futile". See Levin [4] and literature cited therein for a review of the futility design. Because study designs are conventionally named after the alternative hypothesis rather than the null hypothesis which is tested, it would be correct to call the futility design a *non-superiority* study (as distinct from the non-inferiority study, which in a sense is the mirror image of the design we are discussing here). To avoid such terminological confusions and to emphasize the prime motivation when screening interventions, we shall prefer to think of this study, whatever we choose to call it, as a formal *test of promise*.

We shall return to Simon's two-stage design below after beginning with an interesting example in the context of HIV research which will nicely serve to motivate our discussion. One feature of the test of promise formulation that is attractive in any screening context and especially in the motivating example is that the burden of evidence is put on the observations to reject the hypothesis of promise. This offers an important logistical advantage in early phase research as explained below. A second objective of this article is to delineate how to test for promise in a *sequential* manner with *time-to-failure* data. These two features allow for early stopping given sufficient accumulating evidence of lack-of-promise, while allowing full follow-up information to accrue otherwise. These are attractive features because they minimize the number of failures one would need to observe before concluding the intervention lacks promise, while allowing watchful follow-up if sufficiently many subjects do well. Additionally, as discussed in Section 13, by utilizing time-to-failure data, the test of promise achieves somewhat greater power for given type I error control than a corresponding test based only on binary outcomes such as the "response/no response" outcomes used in the Simon two-stage design.

In technical statistical terms, then, in what follows we specify a sequential, single-arm, one-sided test for the discrete hazard constant in a sample of geometric time-to-event observations. After stating assumptions in Section 3 and formulating hypotheses in Section 4, we introduce the sequential test boundaries in Section 5 and discuss immediate entry in Section 6 and staggered entry in Section 7. Section 8 discusses computational details of the procedure and Section 9

[*] Corresponding author.
E-mail addresses: Bruce.Levin@Columbia.edu (B. Levin), LK24@cumc.columbia.edu (L. Kuhn), CL94@cumc.columbia.edu (C.-S. Leu), WT5@cumc.columbia.edu (W.-Y. Tsai).

https://doi.org/10.1016/j.cct.2019.105818
Received 16 April 2019; Received in revised form 25 July 2019; Accepted 4 August 2019
Available online 21 August 2019

introduces a useful duality between monitoring a sequence of ordered geometric failure times on the one hand and monitoring cumulative binomial failure counts over follow-up time on the other. Sections 10 and 11 discuss how to calculate median-unbiased estimates of the hazard constant, *P*-values, and confidence intervals given the sequential design, and Section 12 describes how to calculate the relevant stopping time distributions. In Section 13 we compare some operating characterisitcs of the sequential test with those of Simon's two-stage design. Section 14 explains how to calculate optimal or near-optimal stopping boundaries, and large-sample approximations for such boundaries are provided in an appendix. Section 15 closes with a discussion of the methods.

2. Motivating example—an ART cessation trial

Studies undertaken in adults with primary HIV infection have observed that if antiretroviral therapy (ART) is started very soon after infection occurred, ART can be stopped in a small proportion (5–15%) and viral control will be retained in the absence of ART [6,7,9]. In infants with perinatally-acquired HIV, age at ART initiation is roughly equivalent to time after infection. However, observations of children stopping ART are rare. One child in Mississippi who started ART within 30 h of birth and who, after defaulting treatment, remained virally suppressed for 27 months without ART, drew the attention of the scientific community to this special population [8]. Moreover, consideration of immunological differences between newborns and adults led to the hypothesis that very early ART may be especially beneficial in this group [10,11].

The LEOPARD (Latency Early neOnatal Provision of AntiRetroviral Drugs) study was a single-arm clinical trial in South Africa undertaken to test the promise of very early ART to achieve remission (*clinicaltrials.gov* NCT02431975). The study recruited HIV-infected newborns and initiated ART within 48 h of birth. The investigators expected that ART would lead to rapid decline of plasma HIV RNA to undetectable levels by 24 weeks of treatment and that plasma HIV RNA would remain below detection through 104 weeks. At that point ART would be stopped and the proportion who remained below a cut-off plasma HIV RNA threshold through 48 weeks after ART interruption would be calculated as the success endpoint of HIV remission. The original power calculation made the argument that the benefit of early ART would need to be substantial in order to justify the efforts of starting ART so early. An estimate of 10–25% achieving remission was used in the sample size calculation and a sample size of 40 children stopping ART was chosen. A Data Safety and Monitoring Board was established that deemed the design safe with the monitoring schedule in place.

Recruitment of the HIV-infected newborns proceeded and ART was initiated within 48 h of birth as originally planned. However, preliminary analysis revealed that the proportion of infants achieving the minimum eligibility criteria for ART interruption—a rapid decline and sustained HIV RNA below detection of the assay—was much lower than expected [3]. An initial cohort more than four times the size of that originally planned would be necessary to generate the target sample size of 40 early-treated children to interrupt ART.

These preliminary results raised the question of whether, with the very much smaller number of children eligible to stop ART (currently about ten children), would there still be scientific utility of stopping ART in these few?

We considered this question from the perspective of the futility trial design with an appropriate formal test of promise of whether or not stopping ART in these few children holds promise as a clinical strategy. Testing promise as a null hypothesis seems especially well-suited for addressing the question of whether cessation of ART could be desirable in very young HIV infected children for whom early viral suppression with ART may (or may not) have prevented establishment of a viral reservoir. As mentioned in the Introduction, the burden of evidence is put on the observations to reject the hypothesis of promise. If we fail to reject that hypothesis, an inference which is statistically legitimate to draw is that *the possibility of a promising intervention cannot be ruled out* at the given rates of error control. If one does reject the null hypothesis, however, one can legitimately infer that there is statistically significant evidence of lack of promise. This is an appropriate strategic position to aim for in screening trials, especially with small samples, because there is better negative predictive value when the null hypothesis of promise is rejected (in which case suspending further research seems justified) than there is positive predictive value when one cannot rule out the promise of a proposed intervention (in which case further confirmatory testing seems justified). Contrast this position with the more conventional formulation in which one tests the null hypothesis of no promise and hopes to reject that in favor of the alternative of promising efficacy. With the conventional formulation, failure to find a statistically significant benefit (possibly due to low power in small samples) leaves one allowed to conclude *only* that one hasn't demonstrated efficacy whilst one cannot rule out that the intervention is unpromising. Whereupon enthusiasm for further confirmatory research immediately evaporates. These ideas are further discussed in Levin [4].

3. Notation and assumptions

Consider a sample of K subjects who are followed in discrete time for observation of a failure event for a total of M time units. In our application, we have $K = 10$ children for whom ART could be suspended and who would be followed for $M = 12$ months for the failure event of a viral rebound which could be observed with monthly viral load blood tests. Henceforth we adopt "month" as the unit of time. The K failure times will be denoted by $X_1, \ldots, X_K$. The event that subject i has an observed failure at month m will be denoted by $[X_i = m]$; the censoring event $[X_i > M]$ signifies that no failure is observed within the first M months of observation. We shall assume that each failure time is independent and identically distributed as a geometric distribution with probability function $P[X_i = m | \theta] = \theta(1 - \theta)^{m-1}$ for $m = 1, 2, 3, \ldots$ The discrete-time survival function is $S(m) = P[X_i > m \mid \theta] = (1- \theta)^m$. The parameter θ is the *discrete-time hazard constant*, representing the monthly risk of failing given no prior failure. The assumption of a constant discrete-time hazard function will be discussed below. As an aid to interpretation, we also introduce the *cumulative annual failure probability* P_a given by $P_a = 1 - (1- \theta)^M$. The inverse relation is given by $\theta = 1 - (1-P_a)^{1/M}$.

4. Formulation of hypotheses

We adopt the criterion of promise to be $P_a = 0.50$. That is, if the chances of a viral rebound within 12 months were truly no greater than one-half in the population of children under study, we would consider the cessation of ART a promising intervention. On the other hand, an annual breakthrough probability greater than one-half would be unpromising. Therefore we propose to test the *null hypothesis of promise* H_0: $P_a \leq P_0 = 0.50$ against the *alternative hypothesis of lack of promise* H_1: $P_a > P_0$. Under the constant hazard assumption, this is equivalent to testing H_0: $\theta \le \theta_0 = 1 - (1 - P_0)^{1/M} = 1 - 0.50^{1/12} \approx 0.0561$versus H_1: $\theta > \theta_0$. Let us control the probability of a type I error at the one-sided $\alpha = 0.05$ level. Here, a type I error corresponds to declaring a truly promising intervention lacking in promise. A type II error corresponds to not declaring a truly unpromising intervention lacking in promise. We shall adopt as the *design alternative* a cumulative annual failure probability of $P_a = P_1 = 0.90$, lacking almost all promise, and we will require our test to have power of at least 90% to reject the null hypothesis of promise under this design alternative. The corresponding hazard constant is $\theta_1 = 1 – (1–P_1)^{1/M} = 1–0.10^{1/12} \approx 0.1746$.

5. Sequential testing

To test the hypothesis of promise sequentially, we introduce the *ordered failure times* $X_{(1)} \le X_{(2)} \le \cdots \le X_{(K)}$ and will refer to these times as corresponding to the first observed failure, second observed failure, etc. Censored failure times will be grouped at the end of the chain of inequalities with no implied ordering among the group of failures with censored failure times. In particular, let D denote the total number of observed failures, $D = \sum_{i=1}^{K} I[X_i \le M]$. The sequential test will only utilize the sufficient statistics D and $X_{(1)} \le X_{(2)} \le \cdots \le X_{(D)}$, together with the fact that, by definition of D, the $K–D$ censored observations have $X_{(D+1)} > M, \ldots, X_{(K)} > M$.

At this point we consider two design possibilities, namely, *immediate entry* versus *staggered entry*. As the name suggests, with immediate entry we assume that all K subjects begin their follow-up simultaneously in calendar time. In this case, there is no difference between *follow-up times*, as denoted by the X_i, and when failures occur in *calendar time* after the start of the study. With staggered entry, we assume that for practical reasons it is not possible to intervene simultaneously with all K subjects. In the present case, for example, resources at the clinic to closely monitor all ten children with simultaneous suspension of ART may not be available. Instead we assume there will be a pre-specified schedule of start times s_i for subjects to receive the intervention and begin follow-up. The start times will be measured in calendar time after the first intervention. Now there is a distinction between follow-up times X_i and *calendar event times*, which are given by $X_i + s_i$. For convenience of presentation in what follows, we shall begin by assuming simultaneous entry and later consider the effect of staggered entry. As we shall see, the sequential test will have the interesting and useful property that the null hypothesis of promise will be rejected with staggered entry *if and only if* the hypothesis is rejected with simultaneous entry. Therefore, the type I and type II error rates and power curves are identical under either design. The only effect of staggered entry is to lengthen the duration of the study and possibly to delay the calendar time at which H_0 might be rejected.

6. Stopping boundaries with simultaneous entry

We will refer to an ordered vector of non-negative integers $(b_1, \ldots, b_K)$ as a *stopping boundary*. The *stopping index* will be defined as the smallest $k \ge 1$ such that $X_{(k)} \le b_k$. If the stopping index is less than or equal to K we *stop and reject* H_0. By “stop” we mean that further follow-up of subjects is optional but not required for testing purposes because sufficient evidence of lack of promise is at hand to reject the null hypothesis, irrespective of subsequent outcomes. In the event that the stopping boundary is not crossed after all K outcomes have been observed (including both observed failure times and censored failure times after M months), we *do not reject* H_0. Specifically, we propose the following stopping boundary: $(b_1, \ldots, b_{10}) = (0, 0, 0, 1, 2, 4, 7, 11, 12, 12)$. This boundary implies that with simultaneous entry, four failures must accrue before stopping can occur. (This must be so in order to control the type I error rate at ≤ 0.05.) If all four failures occur after 1 month of follow-up, we stop and reject H_0. If the fourth failure occurs after month 1, we check to see if the fifth failure occurs at month 2; if so, we stop and reject H_0, else we check to see if the sixth failure occurs at or before month 4, and so on. If the eleventh failure is observed (occurring at or before month 12), we stop and reject H_0. Given no prior stopping, if both the ninth and tenth failures are censored, having not occurred at or before month $M = 12$, we do not reject H_0. As explained below, the above stopping boundary has type I error rate ≤ 0.0498 under H_0 and maximizes the power to reject H_0 at the design alternative $P_a = P_1 = 0.90$, namely, with power 0.9275.

7. Sequential testing with staggered entry

The test procedure with staggered entry proceeds with the same stopping boundary as above, only now the follow-up times must be repeatedly reordered as failures occur in calendar time. The first time the boundary is crossed with the *currently re-ordered failure times*, stop and reject H_0. If this does not occur after K outcomes (observed or censored), do not reject H_0. As an illustration, suppose the K subjects enter the study once per month, as follows.

Month of entry	Months to failure[a]	Calendar event time
Study start 0	5	5
1	1	2
2	2	4
3	1	4
4	1	5
5	1	6
6	C	C
7	1	8
8	2	10
Last entry 9	3	12

[a] C = censored ($X_i > 12$).

Extend the notation for ordered failure times by letting$X_{(k,c)}$ now denote the time-to-failure for the k^{th} earliest failure time after observing c events in calendar event time. Three failures occur within the first four calendar months, and then the fourth and fifth failures observed in calendar time occur at calendar month 5, with reordered failure times 1, 1, 1, 2, 5. Because $X_{(4,5)} = 2 > 1 = b_4$ and also $X_{(5,5)} = 5 > 2 = b_5$ we do not stop. However, the sixth failure observed at calendar month 6 occurred after 1 month of follow-up. The reordered failure times are 1, 1, 1, 1, 2, 5 which path crosses the boundary value b_4, so we stop at calendar month 6 and reject H_0. Note that had there been simultaneous entry, the study could have terminated after the first month with a boundary crossing at b_4. With staggered entry, rejection of H_0 is delayed until the sixth failure observed (coincidentally) at month 6. One positive feature of the delay, though, is that if the last three or even four subjects had not yet begun their follow-up, stopping those subjects' ART can be avoided.

It should be clear from the foregoing that if the boundary is crossed with staggered entry, the same boundary crossing must occur with simultaneous entry times (assuming identical sets of failure times). Conversely, if a boundary crossing occurs at b_k, say, with simultaneous entry, crossing must *eventually* occur at the same stopping index with staggered entry, albeit at a later calendar month and with possibly additional failures with longer waiting times already observed. Therefore, as previously asserted, the entry schedule does not affect the type I and type II error rates.

To further illustrate these points, the tables below display the distribution of stopping times with simultaneous and delayed entry, first under the null hypothesis value $P_a = 0.50$ and then under the design alternative $P_a = 0.90$. In each set of tables the left-hand panel displays results for simultaneous entry while the right-hand panel displays results for staggered entry, assuming the same uniform schedule as above in which one subject enters follow-up each month for ten consecutive months. For simultaneous entry, the left-hand column reports the stopping index k for instances in which H_0 is rejected while the right-hand column reports the month m of such rejections. With simultaneous entry, the maximum duration of the study is 12 months (whether the null hypothesis is rejected or not) and must last 12 months in those instances in which H_0 is not rejected. For staggered entry, the left-hand column reports the failure number f at which a rejection occurs (which may be greater than or equal to the stopping index). The right-hand column reports the calendar-time month of stopping from the start of the trial. With staggered entry, the study duration can be as long as 21 months (when the last subject enters 9 months after the first and has a follow-up time of 12 months) and the duration of the trial varies even when the null hypothesis is not rejected. The smaller tables below display the duration of the trial in those instances.

Stopping time distributions when testing under H_0: $P_a = 0.50$ using boundary (0, 0, 0, 1, 2, 4, 7, 11, 12, 12) (100,000 simulations; Monte Carlo error ≤ 0.0016)

Simultaneous entry						Staggered entry (one subject per month)					
By stopping index k			By month number m			By failure number f			By month number m		
k	Point exit prob.'s	Cum. exit prob.'s	m	Point exit prob.'s	Cum. exit prob.'s	f	Point exit prob.'s	Cum. exit prob.'s	m	Point exit prob.'s	Cum. exit prob.'s
1	0	0	1	0.0017	0.0017	1	0	0	1	0	0
2	0	0	2	0.0019	0.0035	2	0	0	2	0	0
3	0	0	3	0.0011	0.0047	3	0	0	3	0	0
4	0.0017	0.0017	4	0.0049	0.0096	4	0.0007	0.0007	4	0.0000	0.0000
5	0.0019	0.0035	5	0.0011	0.0107	5	0.0015	0.0022	5	0.0000	0.0001
6	0.0061	0.0096	6	0.0043	0.0150	6	0.0048	0.0070	6	0.0001	0.0002
7	0.0144	0.0240	7	0.0089	0.0240	7	0.0114	0.0184	7	0.0003	0.0005
8	0.0243	0.0483	8	0.0012	0.0252	8	0.0149	0.0333	8	0.0006	0.0011
9	0.0007	0.0490	9	0.0039	0.0291	9	0.0111	0.0444	9	0.0017	0.0028
10[a]	0	0.0490	10	0.0073	0.0364	10	0.0046	0.0490	10	0.0032	0.0059
			11	0.0119	0.0483				11	0.0038	0.0097
			12	0.0007	0.0490				12	0.0045	0.0142
									13	0.0053	0.0195
									14	0.0050	0.0245
									15	0.0057	0.0302
									16	0.0054	0.0355
									17	0.0047	0.0402
									18	0.0041	0.0444
									19	0.0027	0.0471
									20	0.0018	0.0489
									21	0.0001	0.0490
No rej	0.9510	1.0000		0.9510	1.0000	No rej	0.9510	1.0000	No rej[b]	0.9510	1.0000

[a] If a boundary crossing were to occur at the 10th failure, it would already have occurred at the 9th.

[b] The exit distribution for instances in which H_0 was not rejected is given in the table below.

Trial duration among instances in which H_0 is not rejected

Staggered entry		
By month number m		
m	Point exit prob.'s	Cum. exit prob.'s
14	0.0000	0.0000
15	0.0008	0.0008
16	0.0042	0.0049
17	0.0153	0.0202
18	0.0458	0.0661
19	0.1142	0.1802
20	0.2536	0.4338
21	0.5172	0.9510

Stopping time distributions when testing under H_1: $P_a = 0.90$ using boundary (0, 0, 0, 1, 2, 4, 7, 11, 12, 12) (100,000 simulations; Monte Carlo error ≤ 0.0016)

Simultaneous entry						Staggered entry (one subject per month)					
By stopping index k			By month number m			By failure number f			By month number m		
k	Point exit prob.'s	Cum. exit prob.'s	m	Point exit prob.'s	Cum. exit prob.'s	f	Point exit prob.'s	Cum. exit prob.'s	m	Point exit prob.'s	Cum. exit prob.'s
1	0	0	1	0.0817	0.0817	1	0	0	1	0	0
2	0	0	2	0.1246	0.2063	2	0	0	2	0	0
3	0	0	3	0.0924	0.2987	3	0	0	3	0	0
4	0.0817	0.0817	4	0.1940	0.4927	4	0.0124	0.0124	4	0.0009	0.0009
5	0.1246	0.2063	5	0.0588	0.5515	5	0.0508	0.0633	5	0.0041	0.0050
6	0.2864	0.4927	6	0.1161	0.6677	6	0.2005	0.2637	6	0.0134	0.0184
7	0.2882	0.7809	7	0.1133	0.7809	7	0.3700	0.6337	7	0.0344	0.0528
8	0.1406	0.9215	8	0.0214	0.8023	8	0.2224	0.8561	8	0.0708	0.1236
9	0.0047	0.9262	9	0.0412	0.8435	9	0.0550	0.9111	9	0.1197	0.2434
10[a]	0	0.9262	10	0.0430	0.8865	10	0.0151	0.9262	10	0.1653	0.4087
			11	0.0350	0.9215				11	0.1392	0.5478
			12	0.0047	0.9262				12	0.1083	0.6561
									13	0.0850	0.7411
									14	0.0618	0.8029
									15	0.0449	0.8479
									16	0.0311	0.8790
									17	0.0201	0.8990
									18	0.0135	0.9125
									19	0.0086	0.9211
									20	0.0043	0.9253
									21	0.0009	0.9262
No rej	0.0738	1.0000	12	0.0738	1.0000	No rej	0.0738	1.0000	No rej[b]	0.0738	1.0000

[a] If a boundary crossing were to occur at the 10th failure, it would already have occurred at the 9th.

[b] The exit distribution for instances in which H_0 was not rejected is given in the table below.

Trial duration among instances in which H_0 is not rejected

Staggered entry		
By month number m		
m	Point exit prob.'s	Cum. exit prob.'s
14	0.0001	0.0001
15	0.0005	0.0006
16	0.0019	0.0025
17	0.0046	0.0071
18	0.0090	0.0161
19	0.0138	0.0300
20	0.0188	0.0487
21	0.0251	0.0738

The table contents come from a simulation experiment using 100,000 replications per table. In the next section we discuss exact computation of error rates and show that the exact type I error rate for the optimal boundary is 0.0498 and the exact power at $P_1 = 0.90$ is 0.9275. The values from the simulation experiment (0.0490 and 0.9262, respectively) are within Monte Carlo error of the exact values. The standard error for a given estimated probability p is $\sqrt{p(1-p)/10^5}$ which in all cases is ≤0.0016.

8. Exact calculation of error probabilities

The following algorithm can be used to calculate the exact hypothesis test error probabilities in the case of simultaneous entry. Assume a given integer value of $M > 1$, a pre-specified number of subjects $K \geq 1$, and a given boundary $(b_1, \ldots, b_K)$ with integer values $0 \leq b_1 \leq \cdots \leq b_K \leq M$. Assume a given cumulative annual failure probability P_a with $0 < P_a < 1$ and corresponding discrete-time hazard constant $\theta = 1 - (1-P_a)^{1/M}$. The goal is to calculate two failure-time tail distributions for each failure number $k = 1, \ldots, K$:

$$S_k(x_k) = P[X_{(k)} > x_k \mid X_{(k-1)} > b_{k-1},\ldots, X_{(1)} > b_1] \ \text{ for } \ x_k = b_{k-1}, \ \ldots, M, \tag{1}$$

which is the conditional probability that the k^{th} failure time exceeds x_k months given no *previous* boundary crossing; and

$$T_k(x_k) = P[X_{(k)} > x_k \mid X_{(k)} > b_k, X_{(k-1)} > b_{k-1},\ldots, X_{(1)} > b_1] \ \text{ for } \ x_k = b_k, \ \ldots, M, \tag{2}$$

which is the conditional probability that the k^{th} failure time exceeds x_k months given no *current or previous* boundary crossing.

8.1. Initialization

Though we are interested in the given sample size K, we will need to work recursively with each smaller sample size, which will be indexed by the letter κ (kappa) and used as a superscript in the symbols for ordered failure times and tail distributions. When there is no ambiguity, those symbols without superscripts will refer to sample size K, as in Eqs. (1) and (2) above.

For each sample size $\kappa = 1, \ldots, K$, set

$$S_1^{(\kappa)}(x_1) = P[X_{(1)}^{(\kappa)} > x_1] = (1-\theta)^{\kappa x_1} \ \text{ for } \ x_1 = 1, \ \ldots, M, \tag{3}$$

reflecting the fact that the earliest of κ independent failure times with tail distribution $S(x)$ has tail distribution $S(x)^{\kappa}$. Also set

$$T_1^{(\kappa)}(x_1) = P\left[X_{(1)}^{(\kappa)} > x_1 \mid X_{(1)}^{(\kappa)} > b_1\right] = S_1^{(\kappa)}(x_1)/S_1^{(\kappa)}(b_1) = (1-\theta)^{\kappa(x_1-b_1)} \text{for} x_1 = b_1, \ \ldots, M. \tag{4}$$

Note that Eq. (3) provides the *continuation probability at first failure*, $P[X_{(1)}{}^{(\kappa)} > b_1] = S_1{}^{(\kappa)}(b_1)$ among κ competing independent failure times; the *exit probability at first failure*, $P[X_{(1)}{}^{(\kappa)} \leq b_1] = 1 - S_1{}^{(\kappa)}(b_1)$;

and the *lower tail probability* $P[X_{(1)}{}^{(\kappa)} \le x_1] = 1 - S_1{}^{(\kappa)}(x_1)$ for $x_1 = 1, \ldots, b_1$, which will be used to obtain p-values upon rejecting H_0. Note also that Eq. (4) provides the conditional failure time distribution at first failure given no boundary crossing then, namely, for $x_1 = b_1 + 1, \ldots, M$,

$$\begin{aligned} P[X_{(1)}^{(\kappa)} = x_1 \mid X_{(1)}^{(\kappa)} > b_1] &= T_1^{(\kappa)}(x_1 - 1) - T_1^{(\kappa)}(x_1) \\ &= (1-\theta)^{\kappa(x_1 - 1 - b_1)} - (1-\theta)^{\kappa(x_1 - b_1)} \\ &= \{1 - (1-\theta)^{\kappa}\}(1-\theta)^{\kappa(x_1 - b_1 - 1)}. \end{aligned}$$

8.2. Inductive step

We proceed inductively on both the sample size κ and failure number k. For values $x_k = b_{k-1} + 1, \ldots, M$, let

$$S_k^{(\kappa)}(x_k) = P[X_{(k)}^{(\kappa)} > x_k \mid X_{(k-1)}^{(\kappa)} > b_{k-1}, \ldots, X_{(1)}^{(\kappa)} > b_1] \tag{5}$$

denote the tail distribution for the k^{th} failure time given no previous stopping with κ competing failure times. Note that when evaluated at $\kappa = K$, the values of $S_k(x_k) = S_k{}^{(K)}(x_k)$ provide the following additional quantitites:

(a) tail probabilities at the k^{th} failure given no previous or current stopping,

$$T_k(x_k) = P[X_{(k)} > x_k \mid X_{(k)} > b_k, \ldots, X_{(1)} > b_1] = S_k(x_k)/S_k(b_k) \quad \text{for} \quad x_k = b_k, \ldots, M; \tag{6}$$

(b) point probilities for the k^{th} failure time given no previous or current stopping,

$$P[X_{(k)} = x_k \mid X_{(k)} > b_k, \ldots, X_{(1)} > b_1] = T_k(x_k - 1) - T_k(x_k) \quad \text{for} \quad x_k = b_k + 1, \ldots, M; \tag{7}$$

(c) the continuation probability at the k^{th} failure given no prior stopping,

$$P[X_{(k)} > b_k \mid X_{(k-1)} > b_{k-1}, \ldots, X_{(1)} > b_1] = S_k(b_k); \tag{8}$$

(d) the stopping probability at the k^{th} failure given no prior stopping,

$$P[X_{(k)} \le b_k \mid X_{(k-1)} > b_{k-1}, \ldots, X_{(1)} > b_1] = 1 - S_k(b_k); \tag{9}$$

(e) the probability of stopping at the k^{th} failure,

$$S_1(b_1) \cdots S_{k-1}(b_{k-1})\{1 - S_k(b_k)\}; \tag{10}$$

(f) the unconditional continuation probabilities beyond the k^{th} failure,

$$S_1(b_1) \cdots S_k(b_k); \text{and} \tag{11}$$

(g) the cumulative probability of stopping at or before the k^{th} failure,

$$1 - S_1(b_1) \ldots S_k(b_k). \tag{12}$$

For the inductive step, assume that we have already calculated all of the $S_k{}^{(\kappa)}(x_k)$ quantities in Eq. (5) for each failure number $k = 1, \ldots, \kappa$ for each sample size κ up to some value which, to simplify the notation and without loss of generality, we can assume is K–1. Assume further that we have calculated the first k values for sample size K, namely, $S_1(x_1) = S_1{}^{(K)}(x_1)$ through $S_k(x_k) = S_k{}^{(K)}(x_k)$. To complete the inductive step, we need to calculate $S_{k+1}(x_{k+1}) = S_{k+1}{}^{(K)}(x_{k+1})$ for $x_{k+1} = b_k + 1, \ldots, M$. To do this, we need the following proposition.

8.3. Proposition

For each $k = 1, \ldots, K$–1 and any $x_k = b_k + 1, \ldots, M$,

$$P[X_{(k+1)} > x_{k+1} \mid X_{(k)} = x_k, \quad X_{(k)} > b_k, \ \ldots \ , X_{(1)} > b_1]$$

$$= \frac{\binom{K}{k}(1-\theta)^{(K-k)x_{k+1}}\{T_k^{(k)}(x_k - 1) - T_k^{(k)}(x_k)\}S_k^{(k)}(b_k) \cdots S_1^{(k)}(b_1)}{\{T_k^{(K)}(x_k - 1) - T_k^{(K)}(x_k)\}S_k^{(K)}(b_k) \cdots S_1^{(K)}(b_1)} \quad \text{if} x_{k+1} \ge x_k,$$

and

$$= 1 \ \text{if} x_{k+1} < x_k. \tag{13}$$

8.4. Proof

The term $X_{(k)} > b_k$ is of course redundant with $X_{(k)} = x_k$, but we retain it to remind us there was no stopping after the k^{th} failure. If $x_{k+1} < x_k$ the event $[X_{(k+1)} > x_{k+1}]$ is conditionally certain because $X_{(k+1)} \ge X_{(k)} = x_k > x_{k+1}$. So assume $x_{k+1} \ge x_k$. The numerator of the conditional probability on the left-hand side of Eq. (13) is the probability of the event $[X_{(k+1)} > x_{k+1}, \cdots X_{(k)} = x_k, \quad X_{(k)} > b_k, \ldots, X_{(1)} > b_1]$, which event implies that K–k out of K subjects fail strictly after time x_{k+1} and that the failure times for the k other subjects occur no later than $x_k \le x_{k+1}$ whilst staying above the boundary values $b_1, \ldots, b_k$. For any of the $\binom{K}{k}$ partitions of subjects into two such subsets of subjects, say, $\{i_1, \ldots, i_k\}$ and $\{j_1, \ldots, j_{K-k}\} = \{1, \ldots, K\} \backslash \{i_1, \ldots, i_k\}$, the first subset can have events such as

$$(X_{i_1}, \ldots, X_{i_k}) \in [X_{(k)} = x_k, \quad X_{(k)} > b_k, \ldots, \quad X_{(1)} > b_1]$$

which, jointly with the independent event $[X_{j_1} > x_{k+1}, \ \ldots, \ X_{j_{K-k}} > x_{k+1}]$ for the complementary subset, has *the same probability* as

$$P[X_{(k)}^{(k)} = x_k, \quad X_{(k)}^{(k)} > b_k \quad , \ldots, \quad X_{(1)}^{(k)} > b_1](1-\theta)^{(K-k)x_{k+1}},$$

where in the first factor we consider the sequential procedure with only subjects $\{X_{i_1}, \ldots, X_{i_k}\}$ participating and with the boundary restricted to $(b_1, \ldots, b_k)$. That first factor, however, is

$$\begin{aligned} &P[X_{(k)}^{(k)} = x_k \mid X_{(k)}^{(k)} > b_k \quad , \ldots, \quad X_{(1)}^{(k)} > b_1]P[X_{(k)}^{(k)} > b_k \quad , \ldots, \quad X_{(1)}^{(k)} > b_1] \\ &\quad = \{T_k^{(k)}(x_k - 1) - T_k^{(k)}(x_k)\}S_k^{(k)}(b_k) \cdots S_1^{(k)}(b_1). \end{aligned}$$

Similarly, the denominator of the conditional probability in Eq. (13) is

$$\begin{aligned} &P[X_{(k)}^{(K)} = x_k \mid X_{(k)}^{(K)} > b_k \quad , \ldots, \quad X_{(1)}^{(K)} > b_1]P[X_{(k)}^{(K)} > b_k \quad , \ldots, \quad X_{(1)}^{(K)} > b_1] \\ &\quad = \{T_k^{(K)}(x_k - 1) - T_k^{(K)}(x_k)\}S_k^{(K)}(b_k) \cdots S_1^{(K)}(b_1). \end{aligned}$$

This proves the proposition.

To complete the inductive step, then, suppose $1 \le k \le K$–1. For $x_{k+1} = b_k + 1, \ldots, M$,

$$\begin{aligned} &S_{k+1}(x_{k+1}) = P[X_{(k+1)} > x_{k+1} \mid X_{(k)} > b_k, \quad \ldots, \quad X_{(1)} > b_1] \\ &= \sum_{x_k = b_k + 1}^{\infty} P[X_{(k+1)} > x_{k+1} \mid X_{(k)} = x_k, \quad X_{(k)} > b_k, \quad \ldots, \quad X_{(1)} > b_1]P[X_{(k)} = x_k \mid X_{(k)} > b_k, \quad \ldots, \quad X_{(1)} > b_1] \\ &= T_k(x_{k+1}) + \sum_{x_k = b_k + 1}^{x_{k+1}} P[X_{(k+1)} > x_{k+1} \mid X_{(k)} = x_k, \quad X_{(k)} > b_k, \quad \ldots, \quad X_{(1)} > b_1]\{T_k(x_k - 1) - T_k(x_k)\} \\ &= T_k(x_{k+1}) + \sum_{x_k = b_k + 1}^{x_{k+1}} \frac{\binom{K}{k}(1-\theta)^{(K-k)x_{k+1}}\{T_k^{(k)}(x_k - 1) - T_k^{(k)}(x_k)\}S_k^{(k)}(b_k) \cdots S_1^{(k)}(b_1)}{\{T_k(x_k - 1) - T_k(x_k)\}S_k(b_k) \cdots S_1(b_1)}\{T_k(x_k - 1) - T_k(x_k)\} \\ &= T_k(x_{k+1}) + \binom{K}{k}(1-\theta)^{(K-k)x_{k+1}}\frac{S_k^{(k)}(b_k) \cdots S_1^{(k)}(b_1)}{S_k(b_k) \cdots S_1(b_1)} \sum_{x_k = b_k + 1}^{x_{k+1}} \{T_k^{(k)}(x_k - 1) - T_k^{(k)}(x_k)\} \\ &= T_k(x_{k+1}) + \binom{K}{k}(1-\theta)^{(K-k)x_{k+1}}\frac{S_k^{(k)}(b_k) \cdots S_1^{(k)}(b_1)}{S_k(b_k) \cdots S_1(b_1)}\{T_k^{(k)}(b_k) - T_k^{(k)}(x_{k+1})\} \\ &= T_k(x_{k+1}) + \binom{K}{k}(1-\theta)^{(K-k)x_{k+1}}\frac{S_k^{(k)}(b_k) \cdots S_1^{(k)}(b_1)}{S_k(b_k) \cdots S_1(b_1)}\{1 - T_k^{(k)}(x_{k+1})\} \end{aligned} \tag{14}$$

since $T_k^{(k)}(b_k) = 1$ by definition. Because each of the terms in Eq. (14) have already been computed, the inductive step is complete.

The above algorithm was used to calculate the following exact continuation and exit probabilities by stopping index in the previous illustration with simultaneous entry.

Exact exit and continuation probabilities by stopping index for boundary (0, 0, 0, 1, 2, 4, 7, 11, 12, 12) under the null and design alternative hypotheses

Stopping index k	Under H_0: $P_a = 0.50$				Under H_1: $P_a = 0.90$			
	Exit probabilities given no previous stopping	Continuation probabilities given no previous stopping	Point exit prob.'s	Cum. exit prob.'s	Exit probabilities given no previous stopping	Continuation probabilities given no previous stopping	Point exit prob.'s	Cum. exit prob.'s
1	0	1	0	0	0	1	0	0
2	0	1	0	0	0	1	0	0
3	0	1	0	0	0	1	0	0
4	0.0016	0.9984	0.0016	0.0016	0.0805	0.9195	0.0805	0.0805
5	0.0019	0.9981	0.0019	0.0035	0.1340	0.8660	0.1232	0.2037
6	0.0062	0.9938	0.0061	0.0096	0.3633	0.6367	0.2893	0.4930
7	0.0150	0.9850	0.0148	0.0244	0.5661	0.4339	0.2870	0.7800
8	0.0251	0.9749	0.0245	0.0489	0.6489	0.3511	0.1428	0.9228
9	0.0009	0.9991	0.0009	0.0498	0.0614	0.9386	0.0047	0.9275

Values of $S_k(m) = P[X_{(k)} > m \mid \text{no exit prior to } k^{th} \text{ failure}]$ for $P_a = 0.50$

	Stopping index k									
m	1	2	3	4	5	6	7	8	9	10
12	0.0010	0.0107	0.0547	0.1719	0.3775	0.6249	0.8342	0.9581	*0.9991*	*1.0000*
11	0.0017	0.0172	0.0789	0.2250	0.4526	0.6957	0.8786	*0.9749*	1.0000	
10	0.0031	0.0273	0.1126	0.2904	0.5344	0.7639	0.9164	0.9869		
9	0.0055	0.0432	0.1587	0.3688	0.6204	0.8268	0.9467	0.9947		
8	0.0098	0.0677	0.2205	0.4599	0.7070	0.8817	0.9695	0.9988		
7	0.0175	0.1049	0.3009	0.5613	0.7895	0.9263	*0.9850*	1.0000		
6	0.0313	0.1607	0.4020	0.6685	0.8628	0.9596	0.9944			
5	0.0557	0.2421	0.5231	0.7739	0.9220	0.9816	0.9988			
4	0.0992	0.3571	0.6587	0.8678	0.9639	*0.9938*	1.0000			
3	0.1768	0.5113	0.7960	0.9397	0.9881	0.9989				
2	0.3150	0.7007	0.9133	0.9827	*0.9981*	1.0000				
1	0.5612	0.8950	0.9843	*0.9984*	1.0000					
0	*1.0000*	*1.0000*	*1.0000*	1.0000						

Notes: Cells at boundary values are italicized and correspond to continuation probabilities in the table above. Values of $S_k(m)$ for $m < b_k$ are used for P-values upon rejection of H_0 (see below).

Values of $S_k(m) = P[X_{(k)} > m \mid \text{no exit prior to } k^{th} \text{ failure}]$ for $P_a = 0.90$

	Stopping index k									
m	1	2	3	4	5	6	7	8	9	10
12	0.0000	0.0000	0.0000	0.0000	0.0002	0.0020	0.0238	0.2302	*0.9386*	*1.0000*
11	0.0000	0.0000	0.0000	0.0000	0.0005	0.0048	0.0456	*0.3511*	1.0000	
10	0.0000	0.0000	0.0000	0.0001	0.0013	0.0112	0.0853	0.5136		
9	0.0000	0.0000	0.0000	0.0004	0.0037	0.0251	0.1541	0.7098		
8	0.0000	0.0000	0.0001	0.0014	0.0101	0.0546	0.2662	0.9024		
7	0.0000	0.0000	0.0006	0.0046	0.0264	0.1130	*0.4339*	1.0000		
6	0.0000	0.0002	0.0023	0.0145	0.0654	0.2202	0.6540			
5	0.0001	0.0012	0.0091	0.0432	0.1509	0.3954	0.8821			
4	0.0005	0.0058	0.0337	0.1194	0.3160	*0.6367*	1.0000			
3	0.0032	0.0278	0.1140	0.2929	0.5768	0.8834				
2	0.0215	0.1223	0.3345	0.5992	*0.8660*	1.0000				
1	0.1468	0.4573	0.7528	*0.9195*	1.0000					
0	*1.0000*	*1.0000*	*1.0000*	1.0000						

Notes: Cells at boundary values are italicized and correspond to continuation probabilities in the table above. Values of $S_k(m)$ for $m < b_k$ are used for P-values upon rejection of H_0 (see below).

9. Dual monitoring of event counts

Formula (14) provides the wherewithal to calculate the conditional tail probabilities $S_k(m)$ but roundoff error due to floating-point arithmetic in the computation tends to accumulate, so that even with double-precision floating-point arithmetic and care in organizing operations to minimize roundoff error, numerical inaccuracies become noticeable for large values of K (in excess of 30, say). Happily, there is another way to calculate $S_k(m)$ that we shall call *dual monitoring* which has negligible roundoff error. Dual monitoring essentially transposes the chart that could be used for monitoring the failure times (on the vertical axis) for failure index $k = 1, \ldots, K$ (on the horizontal axis) into a chart that monitors the *cumulative number of failures*, say $Y_{(m)}$, (on the vertical axis) for each month $m = 1, \ldots, M$ (on the horizontal axis). Thus a point (k, m) in the original monitoring scheme becomes a point (m, k) in the dual monitoring scheme and instead of monitoring whether the k^{th} observed failure down-crosses boundary value b_k (i.e., occurs at or earlier than month b_k), we monitor whether the cumulative number of failures $Y_{(m)}$ observed at a given month m up-crosses a suitably defined dual boundary value b_m' (i.e., whether $Y_{(m)} \geq b_m'$). We next present some isomorphisms between these two monitoring schemes.

Given any original boundary $\boldsymbol{b} = (b_1, \ldots, b_K)$, the *dual boundary* $\boldsymbol{b}' = (b_1', \ldots, b_M')$ is defined as follows. For each month $m = 1, \ldots, M$ that appears among the elements of $\boldsymbol{b}$ as $m = b_k$, say, set $b_m' = k$. If month $m = b_k$ appears for several consecutive values of k, use the smallest such k for b_m'. For any value of m that does not appear in $\boldsymbol{b}$, "back-fill" the corresponding values of b_m' using "last value carried backward", i.e., if $b_{k-1} = m_1 < m_2 < \cdots < m_{j-1} < m_j = b_k$, for some $j \geq 3$ say, set the dual values $b_{m_2}', \ldots, b_{m_j}'$ all equal to $k + 1$. See Fig. 1 for the chart that can be used to monitor with original boundary $\boldsymbol{b} = (0, 0, 0, 1, 2, 4, 7, 11, 12, 12)$ and the dual chart that can be used to monitor with dual boundary $\boldsymbol{b}' = (4, 5, 6, 6, 7, 7, 7, 8, 8, 8, 8, 9)$. The boundary values are plotted with circles; the dotted circles in the dual chart represent back-filled boundary values. Arbitrarily we have omitted the circles corresponding to $b_1 = b_2 = b_3 = 0$ in the original boundary that could have been plotted at coordinates (0,1), (0,2), and (0,3) in the dual boundary while we have included both circles at coordinates (12,9) and (12,10) corresponding to the repeated boundary values $b_9 = b_{10} = 12$.

It is evident from the definitions that a down-crossing of the boundary in the left-hand chart corresponds to an up-crossing of the dual boundary in the right-hand chart. It is also evident that a sample path which first touches an original boundary value at coordinates (k, b_k) will also first touch the dual boundary at month m at coordinates (m, b_m') where $m = b_k$ and $b_m' = k$.

Let $Y_1, \ldots, Y_M$ denote the frequency count of observed failures in month $m = 1, \ldots, M$, and as above, let $Y_{(m)} = Y_1 + \cdots + Y_m$ denote the cumulative number of failures observed at or before month m. Given the constant hazard assumption, the conditional distribution of $Y_{(m+1)} - Y_{(m)}$ given $Y_{(m)}$ is binomial with index $K–Y_{(m)}$ and parameter θ, because each of the $K–Y_{(m)}$ subjects still in follow-up has probability θ of failing the next month, independently of the others. Therefore, given a sample path which visits (m–1, $Y_{(m-1)}$) in the dual chart with no current or previous boundary crossing (dual or original), the conditional probability that it will visit (m, $Y_{(m)}$) the following month is the binomial probability $\binom{K - Y_{(m-1)}}{Y_{(m)} - Y_{(m-1)}}\theta^{Y_{(m)} - Y_{(m-1)}}(1-\theta)^{K - Y_{(m)}}$. Furthermore, to each lattice point (m, k) in the dual monitoring chart (marked by small squares in the diagram) we can assign the value of the conditional *lower* tail probability given no previous boundary crossing, $S_m'(k) = P[Y_{(m)} < k | Y_{(m-1)} < b_{m-1}', \ldots, Y_{(1)} < b_1']$. For $m = 1$, this is simply the lower binomial tail probability

$$S_1'(k) = \sum_{i=0}^{k-1}\binom{K}{i}\theta^i(1-\theta)^{K-i}$$

and for $m > 1$ we use

$$S_m'(k) = \sum_{i=0}^{k-1}\sum_{j=0}^{\min(i,\ b_{m-1}'-1)} P[Y_{(m-1)} = j \mid Y_{(m-1)} < b_{m-1}', \ldots, Y_{(1)} < b_1'] \cdot \binom{K-j}{i-j}\theta^{i-j}(1-\theta)^{K-i}. \tag{15}$$

Note that Eq. (15) does not require recursive evaluation of $S_m'(k)$ for sample sizes $k < K$ as Eq. (14) requires for $S_k(m)$. The conditional lower tail probability given no previous or current boundary crossing is given by $T_m'(k) = S_m'(k)/S_m'(b_m')$, so that we can also express Eq. (15) as

$$S_m'(k) = \sum_{i=0}^{k-1}\sum_{j=0}^{\min(i,\ b_{m-1}'-1)} \{T_{m-1}'(j+1) - T_{m-1}'(j)\} \cdot \binom{K-j}{i-j}\theta^{i-j}(1-\theta)^{K-i}. \tag{16}$$

Next, we obtain values of $S_k(m)$ from those of $S_m'(k)$ as follows for given $1 \leq k \leq K$ and any m satisfying $q = b_{k-1} + 1 \leq m \leq M$. (Of course $S_k(m) = 1$ for values of $m \leq b_{k-1}$.) If $m = q$, then set $S_k(m) = S_m'(k)$. If $m > q$, then multiply $S_m'(k)$ by the dual conditional continuation probabilities from month q through month m–1:

$$S_k(m) = S_m'(k) \cdot S_{m-1}'(b_{m-1}') \cdots S_q'(b_q'). \tag{17}$$

The logic of Eq. (17) is that for the k^{th} failure to occur strictly after month m (given no previous boundary crossing), there must be no more than k–1 failures at or before month m, the conditional probability of which is $S_m'(k)$, assuming the previous boundary value b_{k-1} was month m–1. If "no previous boundary crossing" only implies $X_{(k-1)} > b_{k-1}$ with $b_{k-1} < m$–1, however, we need to insure that there are no intermediate dual boundary crossings between month $b_{k-1} + 1 = q$ and month m–1 in order to satisfy the condition that $Y_{(m-1)} < b_{m-1}'$ which is implicit in the definition of $S_m'(k)$.

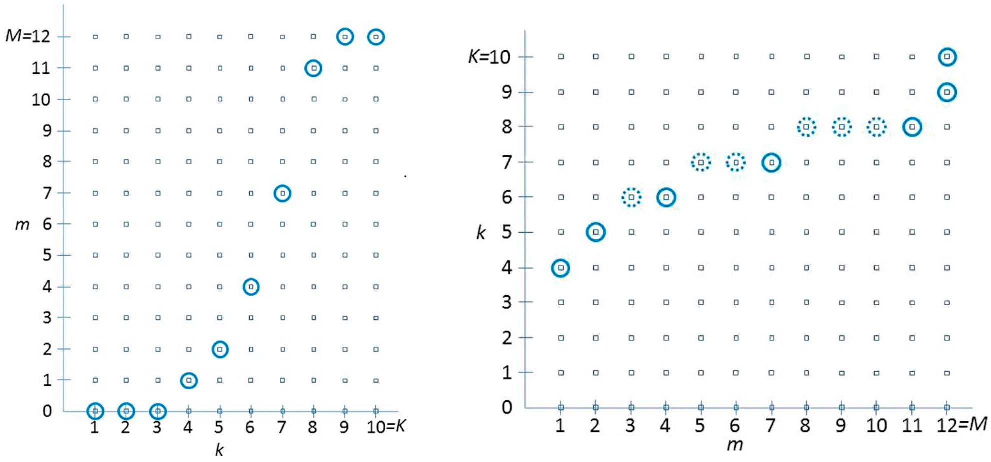

Fig. 1. Original and dual monitoring boundaries.
Left-hand chart is for monitoring failure times by failure index.
Right-hand chart is for monitoring cumulative failure frequencies by month.
Dotted circles in dual boundary are back-filled values where original boundary skips months.

Because all of the quantities in Eq. (5) through Eq. (12) depend only on $S_k(m)$, they can all be calculated from the dual tail probabilities $S_{m}'(k)$ with high accuracy.

10. *P*-values upon rejection or non-rejection of H_0

Sequential *P*-values for a completed experiment are based on establishing an ordering of outcomes along the stopping boundary. We define two *P*-value functions, one for cases in which the boundary is crossed and H_0 is rejected, the other for cases in which there is no boundary crossing and H_0 is not rejected. The first is based upon the number of failures observed and the month in which the dual boundary is crossed. The second is based upon the number of failures observed and the longest time-to-failure among them (with all longer failure times censored).

10.1. Definition of rejection P-value

Suppose a boundary crossing occurs for the first time at month m with a total of $k_m \geq b_m'$ failures observed when monitoring using the dual boundary. We define the *rejection P-value* as the prospective probability that an exit would occur for the first time either strictly before month m or at month m with at least k_m failures. Letting m^* denote the random stopping index, the rejection *P*-value, which will be denoted by $Pval_1(k_m, m|P_a)$, is

$$Pval_1(k_m, m \mid P_a) = P[m^* < m] + P[m^* = m \text{ and } Y_{(k)} \geq k_m].$$

Some equivalent expressions are as follows.

$$\begin{aligned} Pval_1(k_m, m \mid P_a) &= P[m^* < m] + P[m^* \geq m]P[Y_{(k)} \geq k_m \mid m^* \geq m] \\ &= 1–P[m^* \geq k]\{1–P[Y_{(k)} \geq k_m \mid m^* \geq m]\} \\ &= 1–P[m^* \geq k]P[Y_{(k)} < k_m \mid m^* \geq m] \\ &= 1 - S_1'(b_1')\cdots S_{m-1}'(b_{m-1}')S_m'(k_m) \end{aligned} \tag{18}$$

We note that while it would be possible to define the rejection *P*-value differently, as follows, using the original monitoring of ordered failure times, it is not advisable to do so. Suppose a boundary crossing occurs for the first time at stopping index k with time-to-failure $m_k \leq b_k$. Then the rejection *P*-value could be defined as the prospective probability that an exit would occur for the first time either strictly before index k or at index k with a time-to-failure no later than m_k. Letting k^* denote the random stopping index, the rejection *P*-value would be

$$\begin{aligned} &P[k* < k] + P[k* = k \text{ and } X_{(k)} \leq m_k] \\ &=P[k* < k]+P[k* \geq k]P[X_{(k)} \leq m_k \mid k* \geq k] \\ &= P[k* < k] + P[k* \geq k]\{1–S_k(m_k)\} \\ &= 1 - S_k(m_k)S_{k-1}(b_{k-1})\cdots S_1(b_1). \end{aligned}$$

This definition agrees with $Pval_1(k, m_k|P_a)$ in cases where only one failure occurs at month m_k, but it doesn't account for any additional events that might occur as boundary b_k is crossed at month m_k, i.e., it doesn't account for “horizontal overshoot”. For this reason we prefer to define $Pval_1(k, m_k|P_a)$ based on the dual boundary.

10.2. Definition of continuation P-value

Here we use monitoring by ordered failure times. Suppose the study ends with no boundary crossing after $k \geq 0$ failures have been observed (the remaining $K–k$ failure times having been censored), with the longest time-to-failure among the uncensored observations equal to $m_k > b_k$. Then the *continuation P-value* is defined to be one minus the prospective probability that either strictly fewer than k failures would occur or that exactly k failures would occur with a longest time to failure not less than m_k, in either case with no boundary crossing. To calculate the continuation *P*-value, denoted by $Pval_0(k, m_k|P_a)$, we express the complement of the event of interest as a disjoint union of events, namely,

$$\begin{aligned} &[\text{no events observed}] \cup [\text{one event observed and } k* > 1] \cup \cdots \cup \\ &[k–1 \text{ events observed and } k* > k–1] \cup [k \text{ events observed}, k* > k, \text{and } m_k \leq X_{(k)} \leq M]. \end{aligned}$$

If no events are observed ($k = 0$), only the first event is included, in which case the continuation *P*-value is simply $1 – P[X_{(1)} > M] = 1 – (1–\theta)^{MK} = 1 – (1–P_a)^K$. Otherwise the continuation *P*-value is

$$\begin{aligned} Pval_0(k,\ m_k \mid P_a) = 1 - &\sum_{j=0}^{k-1} P[X_{(j+1)} > M,\ b_j < X_{(j)} \leq M,\ X_{(j)} > b_j,\ldots,X_{(1)} > b_1] \\ &- P[X_{(k+1)} > M,\ (m_k - 1) < X_{(k)} \leq M,\ X_{(k)} > b_k,\ldots,X_{(1)} > b_1] \end{aligned}$$

By the same derivation leading up to Eq. (14), only here with $k = j$, $x_{k+1} = M$, and $b_j < X_{(j)} \leq M$ instead of $b_j < X_{(j)}$, each term in the sum can be written as

$$\begin{aligned} &P[X_{(j+1)} > M,\ b_j < X_{(j)} \leq M,\ X_{(j)} > b_j,\ldots,X_{(1)} > b_1] \\ &\quad =\binom{K}{j}(1 - \theta)^{M(K-j)}S_j^{(j)}(b_j)\cdots S_1^{(j)}(b_1)\{T_j^{(j)}(b_j) - T_j^{(j)}(M)\} \\ &\quad = \{S_{j+1}(M) - T_j(M)\}S_j(b_j)\cdots S_1(b_1) \end{aligned} \tag{19}$$

where the last inequality follows from Eq. (14) itself. For the final term in the continuation *P*-value, the derivation leading up to Eq. (14) also yields

$$\binom{K}{k}(1 - \theta)^{M(K-k)}S_k^{(k)}(b_k)\cdots S_1^{(k)}(b_1)\{T_k^{(k)}(m_k - 1) - T_k^{(k)}(M)\}.$$

Therefore we have

$$\begin{aligned} Pval_0(k, m_k \mid P_a) = 1 - &\sum_{j=0}^{k-1} \{S_{j+1}(M) - T_j(M)\}S_j(b_j)\cdots S_1(b_1) \\ &-\binom{K}{k}(1 - \theta)^{M(K-k)}S_k^{(k)}(b_k)\cdots S_1^{(k)}(b_1)\{T_k^{(k)}(m_k - 1) - T_k^{(k)}(M)\}. \end{aligned} \tag{20}$$

We can further express Eq. (20) solely in terms of tail probabilities and without use of the intermediate values of ${S_k}^{(k)}$ and ${T_k}^{(k)}$ by writing $\{{T_k}^{(k)}(m_k - 1) - {T_k}^{(k)}(M)\}$ as $\{1 - {T_k}^{(k)}(M)\} - \{1 - {T_k}^{(k)}(m_k - 1)\}$, so that upon another application of Eq. (14), the final term of Eq. (20) becomes

$$\begin{aligned} &S_k(b_k)\cdots S_1(b_1)[\{S_{k+1}(M) - T_k(M)\} - (1 - \theta)^{(M-m_k+1)(K-k)} \\ &\quad \{S_{k+1}(m_k - 1) - T_k(m_k - 1)\}] \end{aligned}$$

Thus, if $k < K$ we have

$$\begin{aligned} Pval_0(k, m_k \mid P_a) = 1 - &\sum_{j=0}^{k} \{S_{j+1}(M) - T_j(M)\}S_j(b_j)\cdots S_1(b_1) \\ &+ (1 - \theta)^{(M-m_k+1)(K-k)}\{S_{k+1}(m_k - 1) - T_k(m_k - 1)\}S_k(b_k)\cdots S_1(b_1). \end{aligned} \tag{21}$$

Note that if $m_k = b_k + 1$, the final term in Eq. (21) is 0 since $S_{k+1}(b_k) = T_k(b_k) = 1$. If $k = K$, we have directly from Eq. (20),

$$\begin{aligned} Pval_0(K, m_K \mid P_a) = 1 - \sum_{j=0}^{K-1} \{S_{j+1}(M) - T_j(M)\} S_j(b_j) \cdots S_1(b_1) \\ - \{T_K(m_K - 1) - T_K(M)\} S_K(b_K) \cdots S_1(b_1). \end{aligned}$$

The continuation *P*-value monotonically decreases from its largest value 1 – $(1–P_a)^K$ as the number of observed failures increases from zero and as the last observed time-to-failure decreases from M to $b_k + 1$. The smallest continuation *P*-value occurs with the largest number of observed failures k such that $b_k < M$ and with the last time-to-failure equal to $b_k + 1$. Similarly, the rejection *P*-value monotonically increases as the stopping index k^* increases from the smallest possible stopping index (namely, the smallest k with $b_k > 0$) and as the time-to-failure increases from $b_{k^*-1} + 1$ to b_{k^*}. The largest rejection *P*-value occurs with the largest k such that $b_k < M$ and with the last time to failure equal to b_k. Because this *P*-value is simply the total probability of all possible paths leading to a boundary crossing, the largest rejection *P*-value equals α. Furthermore, because the smallest continuation *P*-value equals one minus the probability of all possible sample paths which do not cross the boundary, which itself is the complement of the probability of all paths that do cross the boundary, the smallest continuation *P*-value also equals α. To avoid any ambiguity, then, we shall say that failure to reject the null hypothesis of promise occurs if and only if the continuation *P*-value is *strictly* greater than α, whereas rejection of the null hypothesis of promise occurs if and only if the rejection *P*-value is less than *or equal* to α.

Analogous to the case of the rejection *P*-value, it would be possible to define the continuation *P*-value in terms of the dual binomial monitoring scheme, but it is more direct to determine the conditional probabilities for the month of the last observed failure together with possibly later occurrences of that failure in the vertical direction of the geometric monitoring scheme than to do so in the horizontal direction of the binomial scheme.

11. Median-unbiased estimates of P_a and confidence intervals

The above definitions of *P*-value allow point and interval estimation for P_a. For example, upon a promising result with k failures and maximum time to failure m_k, one may solve the equation $Pval_0(k, m_k|P_a) = 0.50$ for P_a to obtain a median-unbiased estimate of P_a. Solving the equation $Pval_0(k, m_k|P_a) = 0.05$ provides a lower one-sided 95% confidence limit for P_a; solving the equation $Pval_0(k, m_k|P_a) = 0.95$ provides an upper one-sided 95% confidence limit for P_a; and both of those limits provide a two-sided 90% confidence interval. For example, with the boundary (0,0,0,1,2,4,7,11,12,12) illustrated above, suppose one observed 7 failures and 3 censored failure times without a boundary crossing, with the seventh failure occurring at month 10. The *P*-value for this event is $Pval_0(7,10|0.50) = 0.0848$. The median unbiased point estimate of P_a is 0.7083. A two-sided 90% confidence interval for P_a is (0.4572, 0.8892), with lower one-sided 95% confidence limit 0.4572 and upper one-sided 95% confidence limit 0.8892. Suppose on the other hand that the first 7 failures occur at month 1, 2, 3, 3, 3, 5, and 5, such that the boundary is crossed for the first time at stopping index $k^* = 7$. The *P*-value for this event is now $Pval_1(7,5|0.50) = 0.0108$. The median unbiased point estimate of P_a is 0.8870, which solves the equation $Pval_1(7,5|P_a) = 0.50$. A two-sided 90% confidence interval for P_a is (0.6339, 0.9823), with lower one-sided 95% confidence limit 0.6339 and upper one-sided 95% confidence limit 0.9823. These confidence intervals are considerably narrower than simple binomial confidence intervals. For 7 observed failures out of 10 subjects, a two-sided 90% confidence interval by the point-probability method (see [1], section 2.7) is (0.3979, 0.8842), which is not only wider than the above intervals but also loses the nuance of the sequential testing outcomes.

12. Distribution of the stopping index, number of observed events, and month of stopping

The calculations of *P*-value given above also provide the distribution of the stopping index k^* when there is lack of promise or the number of observed failures D and censored outcomes $K–D$ when there is no boundary crossing. The probability of the first event is given by

$$P[\text{cross boundary at}\, k * = k] = S_1(b_1) \cdots S_{k-1}(b_{k-1})\{1 - S_k(b_k)\}$$

while the probability of the second is given by

$$\begin{aligned} P[D = d] &= P[X_{(d+1)} > M, b_d < X_{(d)} \leq M, X_{(d-1)} > b_{d-1}, \ldots, X_{(1)} > b_1] \\ &= S_1(b_1) \cdots S_d(b_d)\{S_{d+1}(M) - T_d(M)\} \\ &= S_1(b_1) \cdots S_{d-1}(b_{d-1})[\{S_d(b_d) S_{d+1}(M)\} - S_d(M)]. \end{aligned}$$

With simultaneous entry, we can state the stopping time distribution for the month of first exit m^* using the dual boundary. (The stopping time is obviously M when there is no stopping.) We have

$$P[\text{stop at month}\, m^* = m] = S_1'(b_1') \cdots S_{m-1}'(b_{m-1}')\{1 - S_m'(b_m')\}$$

13. Comparison with Simon's two-stage design

The two-stage design of Simon [12] is a classic single-arm futility design or as we prefer to call it, a test of promise. It mightn't appear so because Simon originally formulated his one-sided null hypothesis in the unpromising region of the parameter space, as in a conventional superiority formulation (or even as in a non-inferiority design). However, his intention was clear ([12], p.2, with our paraphrasing in brackets):

> If the null hypothesis is true, then we require that the probability should be [small] of concluding that the drug is sufficiently promising that it should be accepted for further study in other clinical trials. We also require that if a specified alternative hypothesis [of promising performance] is true, then the probability of rejecting the drug for further study should be [small].

In other words, in Simon's formulation, if the *alternative* hypothesis (of promising performance) is true, rather than *wanting* to reject the null hypothesis (of unpromising performance) and declaring statistical significance, Simon *does not want to "reject"* ***the drug*** for further study. Indeed, in his discussion, Simon appropriately expresses reluctance to declare significant efficacy as potentially misleading in the context of a single-arm study with small sample size. Similarly, if the null hypothesis (of unpromising performance) is true, rather than expressing a type I error as rejecting the null *hypothesis*, Simon chooses to phrase this error as "concluding that the drug is sufficiently promising that it should be accepted for further study in other clinical trials." With these considerations, reformulating the design as a test of the *null hypothesis of promising performance* versus the *alternative hypothesis of lack of promise*, i.e., a futility design, truly helps to clarify the actual goals motivating the study. All that is needed to translate Simon's error rates into those of the futility design is to interchange α with β.

Simon's two-stage design uses binary endpoints, in which case it does not matter if the outcome is a positive one, such as response to chemotherapy as in Simon's original presentation, or a negative one, such as disease progression or death, or failure to respond within a given period. For consistency with the present discussion, we prefer to keep our focus on failure-type endpoints. We will return to this distinction in the Discussion. Simon's design is also a simple sequential procedure with two stages (for counts of binary outcomes), but we note that it is not a special case of ours which uses time-to-failure data. Even though it is possible to represent either stage of the two-stage design with binary outcomes individually as a corresponding sequential test with time-to-failure data—using a boundary of the form $\boldsymbol{b} = (0, \ldots, 0, M, \ldots, M)$ with $(r_1–1)$ zeros and $(K–r_1 + 1)$ M's, for example, where r_1 is Simon's criterion number of observed failures to stop after the first stage, because in that case early stopping occurs if and only if at least r_1 failures are observed—the two different decision criteria in the two-stage design would require two different sequential stopping boundaries with time-to-failure data. While one could contemplate implementing two distinct stages of sequential futility testing, there seems little point to doing so as compared to a single sequential plan. We shall therefore discuss only the latter.

Given that time-to-failure data contain more information per subject than do binary failure data, it is reasonable to ask to what extent utilization of *time-to-failure* in a fully sequential design such as we have been considering can improve over Simon's design. A partial answer is given by considering the Fisher information per observation with a binary endpoint and comparing it with the Fisher information per observation with a geometric endpoint with censoring after M months. To make the two functions comparable, we calculate each using the discrete-hazard constant $\theta = 1 - (1–P_a)^{1/M}$. A straightforward calculation shows that the Fisher information per binary observation is given by

$$I_B(\theta) = I_B(P_a)\left(\frac{dP_a}{d\theta}\right)^2 = \frac{M^2(1-\theta)^{2(M-1)}}{P_a(1-P_a)} = \frac{M^2(1-P_a)}{(1-\theta)^2 P_a},$$

and the Fisher information per geometric time-to-failure observation with censoring after M months is

$$I_G(\theta) = \frac{P_a}{\theta^2(1-\theta)}.$$

The relative efficiency $I_B(\theta)/I_G(\theta) = \{M^2\theta^{\,2}(1–\theta)^{M-1}\}/\{1 - (1–\theta)^M\}^2$ equals 1 for $M = 1$ and can easily be shown to be strictly < 1 for all $M > 1$, and noticeably so for $P_a > 0.75$. Therefore we may expect some efficiency gain by utilizing the time-to-failure data as well as by allowing early stopping with the sequential test procedure.

A simple numerical illustration should suffice. The first row of table 1 of Simon [12], re-expressed here in equivalent terms of cumulative annual failure probabilities, tests H_0: $P_a \le P_0 = 0.75$ versus H_1: $P_a > P_0$ with design alternative $P_a = P_1 = 0.95$ with type I and type II error rates each equal to 0.10. The right-hand panel gives the minimax design, which minimizes the maximum sample size (20 in this case) among all two-stage designs satisfying the error rate constraints. The procedure rejects H_0 if there are 13 failures among the first 13 patients tested; if fewer than 13 failures occur testing continues with an additional 7 patients, and H_0 is rejected if there are 18 or more failures in total among the 20 patients (including those from the first stage). Under the design alternative, the probability of stopping at stage 1 is $0.95^{13} = 0.51$ (this is Simon's PET or probability of early termination) and his table shows the expected number of subjects is $16.4 = (13 \times 0.51) + (20 \times 0.49)$. Of course, the actual sample size in any instantiation is either 13 or 20 barring any second-stage curtailment. With a failure probability of 0.95, the expected number of failures is $0.95 \times 16.4 = 15.6$ by Wald's lemma.

A sequential geometric time-to-failure test of promise can be implemented with a maximum of only 15 patients on treatment, which is even less than the two-stage design's expected number of subjects under the design alternative, only 2 more than the minimum in Simon's minimax two-stage design, but 5 fewer than the maximum sample size of 20. The boundary (0, 0, 0, 0, 0, 1, 1, 1, 3, 4, 5, 6, 9, 11, 12) yields type I error rate (for the test of promise) of $\alpha = 0.0993$ and power over 90% ($\beta = 0.0965$). The expected number of failures under the design alternative is 10.4, about a one-third reduction in expected failures compared to the two-stage design. The probability that the sequential test will stop sometime at *or before* 13 failures is 0.8654, substantially greater than the PET of 0.51 in the two-stage design. In fact, the cumulative probability of stopping is about 40% by the 9th failure, 55% by the 10th failure, 65% by the 11th failure, and 71% by the 12th failure.

The above illustration may somewhat overstate the advantages of the sequential time-to-failure test because we assumed simultaneous entry by design. As explained above in Section 7, staggered entry would increase the expected number of observed failures and lengthen the duration of the trial. On the other hand, with staggered entry, the possibility exists of enrolling fewer than 15 patients if rejection of promise occurs before all patients have been enrolled. Staggered entry may also be desirable if the treatment has its own serious adverse risks. This is a benefit of the two-stage design, which naturally staggers entry into two groups. Depending on circumstances, reducing the expected number of failures may outweigh the risk of starting all patients on treatment as soon as they become available. When this is not the case, staggered entry may be used to progress more cautiously.

14. Optimal and near-optimal boundaries

The boundaries discussed in Sections 6 and 13 are *optimal* boundaries in the sense that they maximize the power of the test of promise at the design alternative among all sequential tests defined in terms of boundary values $(b_1, \ldots, b_K)$ or $(b_1', \ldots, b_M')$ applied to ordered time-to-failure data and which limit the type I error probability to no more than α at the superiority margin. For small K and small M (e.g., $K = 10$ or 15 and $M = 12$), it is feasible to identify such optimal boundaries by a systematic search algorithm; we present such an algorithm in Appendix 1. For large K or M, however, exhaustive searches become infeasible. Moreover, the problem is somewhat ill-posed because of discreteness in the data, which makes the actual type I error probability somewhat smaller than α by a variable amount. In particular, two different boundaries can have very close power values, where one boundary has a slightly greater power than the other but also comes closer to the type I error bound α; but the reverse also occurs. For example, the optimal boundary $\boldsymbol{b} = (0, 0, 0, 1, 2, 4, 7, 11, 12, 12)$ has type I error probability 0.0498 at the margin of superiority $P_a = 0.75$ and power 0.9275 at the design alternative $P_a = 0.90$. By way of comparison, the boundary $\boldsymbol{b} = (0, 0, 0, 0, 1, 5, 5, 11, 12, 12)$ has, respectively, type I error probability 0.0497 and power 0.9252 (with slightly smaller power and smaller type I error rate than the optimal boundary), whereas the boundary $\boldsymbol{b} = (0, 0, 0, 0, 3, 4, 6, 11, 12, 12)$ has, respectively, type I error probability 0.04998 and power 0.9253 (also with slightly smaller power than the optimal boundary but slightly larger type I error rate).

Because of this feature, we shall be content merely to identify *near-optimal* boundaries, whose resulting power values are only negligibly below optimal for practical purposes. We describe three such methods here plus another in Appendix 2. One method we have found practicable in larger samples is to begin with boundaries in the form of step functions with constant boundary values over subintervals of {1, …, K} of length r, say, with $K = rK'$ for some integer divisor r of K. The algorithm for searching for the optimal boundary assuming sample size K' can be applied with such "cloned" boundary values to rapidly produce a feasible starting boundary, followed by manual adjustment of the jump discontinuities to "smooth" out the boundary, each time increasing the power of the test. We omit further details. Note, though, that any near-optimal boundary can be characterized by its alpha-spending function (the cumulative probability of rejection as a function of the stopping index). Therefore, another strategy to design a boundary for large K is as follows. For given α, β, M, K, θ_0, and θ_1, find a near-optimal boundary for smaller K' and larger θ_1', such that the square root of K' times the log odds ratio of θ_1' versus θ_0 equals the square root of K times the log odds ratio of θ_1 versus θ_0, on the heuristic grounds that the power of the test for larger K and smaller θ_1 ought to be approximately equal to that of the smaller K' and larger θ_1'. Calculate the alpha-spending function for the near-optimal dual boundary for the smaller K' and then locate the boundary values for the larger K for $m = 1, …, M$ which approximate to the interpolated spending function. Minor adjustments may be necessary to constrain the type I error probability less than or equal to α. The *RUNUP* algorithm in Appendix 1 can be used for this purpose.

A third method to produce near-optimal boundaries, perhaps the easiest to implement and therefore our preferred method, is motivated by a repeated likelihood ratio test or *RLRT*. The likelihood ratio test for testing a simple null hypothesis H_0: $\theta = \theta_0$ versus a simple alternative hypothesis H_1: $\theta = \theta_1$, where $\theta_1 > \theta_0$, based on either geometric time-to-event data in direct monitoring or cumulative monthly failure counts in dual monitoring, rejects H_0 when the log likelihood ratio exceeds a pre-specified constant chosen to control the type I error probability. Suppose at any point during the trial with either immediate or staggered entry, the largest current observed follow-up time is m months ($m = 1, …, M$) with cumulative monthly failure counts $Y_{(1)}, …, Y_{(m)}$ as defined in Section 9. The log-likelihood ratio LLR_m, is given by

$$\begin{aligned} LLR_m &= \beta_1 Y_{(m)} + \beta_0 \left\{ mK - \sum_{i=1}^{m} Y_{(i)} \right\} \\ &= (\beta_1 - \beta_0) Y_{(m)} + \beta_0 \left\{ mK - \sum_{i=1}^{m-1} Y_{(i)} \right\}, \end{aligned}$$

where $\beta_1 = \log(\theta_1/\theta_0) > 0$ and $\beta_0 = \log\{(1-\theta_1)/(1-\theta_0)\} < 0$. Note that $\beta_1 - \beta_0$ is the log odds ratio comparing the two monthly failure probabilities. Then the *RLRT* stops and rejects H_0 at the first follow-up month m where an event occurs such that $LLR_m \geq C_\alpha$, where C_α is chosen to limit the overall type I error probability at level α. Equivalently, *RLRT* stops at the first $m = 1, …, M$ such that

$$Y_{(m)} \geq \frac{C_\alpha - \beta_0 \left\{ Km - \sum_{i=1}^{m-1} Y_{(i)} \right\}}{\beta_1 - \beta_0}. \tag{22}$$

Note that the stopping criteria in Eq. (22) depend on the *random* past trajectory values of the cumulative monthly failure counts and so the *RLRT* is not expressible in terms of a *fixed* dual monitoring boundary $\boldsymbol{b}'$. However, if we replace the random sums in Eq. (22) with their largest possible values, conditional on no prior stopping, we can recursively define the following useful fixed dual monitoring boundary. First, set $b_1' = \left\langle \frac{C_\alpha - \beta_0 K}{\beta_1 - \beta_0} \right\rangle$, where $\langle\cdot\rangle$ denotes rounding to the nearest integer. Then, recursively for $m = 1, …, M$, define

$$b_m' = \left\langle \frac{C_\alpha - \beta_0 \left\{ Km - \sum_{i=1}^{m-1} (b_i' - 1) \right\}}{\beta_1 - \beta_0} \right\rangle. \tag{23}$$

The constant C_α may be found by trial and error or an iterative procedure to limit the type I error to a value close to α. We have found that an application of the *RUNUP* algorithm described in Appendix 1 can be used to reduce the exact type I error probability to a value below α if it is not already so, and tends to produce near optimal boundaries.

As an illustration, we calculated boundary Eq. (23) for the case $K = 20$, $M = 12$, $P_a = 0.75$ under H_0 and $P_a = 0.95$ under H_1 with $\alpha = 0.10$ and requiring power at least 95%. Using $C_\alpha = 3.0$ in Eq. (23) produced the dual boundary $\boldsymbol{b}' = (7, 9, 11, 13, 14, 15, 16, 17, 17, 18, 18, 19)$ whose type I error probability is 0.09768 and whose power at $P_a = 0.95$ is 0.9576. This is a near-optimal boundary; the optimal boundary is $\boldsymbol{b}' = (8, 10, 12, 13, 14, 15, 16, 16, 17, 18, 18, 19)$ with type I error probability 0.09959 and power at $P_a = 0.95$ is 0.9589.

Note that whereas the Neyman-Pearson lemma shows that the likelihood ratio test maximizes power for testing simple versus simple hypotheses, there is no claim that a *repeated* likelihood ratio test such as Eq. (22) with fixed C_α maximizes power. In fact, simulations show that Eq. (22) tends to have *lower* power than the optimal or even near-optimal fixed boundary tests such as Eq. (23). Presumably a similar assertion holds for a repeated score test procedure which, like the *RLRT*, is not expressible as a sequential test with fixed boundaries. This will be discussed elsewhere.

Appendix 2 contains another method for calculating near-optimal boundaries for large K based on asymptotic theory.

15. Discussion

The foregoing development focused exclusively on time-to-*failure* data. Here we point out that for certain *different* hypothesis test formulations, the above sequential tests may be applied directly to time-to-*success* data with no other changes. Suppose we redesignate θ as the discrete "hazard" constant for a *positive* outcome, such as a timely response to immunotherapy ("hazard" meaning "chance" here rather than "danger"). Let us revisit Simon's [12] original hypothesis test formulation, only now we monitor time-to-success data rather than only binary responses, in which case for a pre-specified value of $\theta = \theta_0$ the null hypothesis is H_0: $\theta \leq \theta_0$ which is to be tested against the alternative H_1: $\theta > \theta_0$. Note that this interchanges the hypotheses of the test of promise as discussed earlier for failure-type outcomes. Here θ_0 could specify a complete lack of clinical activity, or that of a placebo effect, or even that of a standard active treatment, in which case this formulation is a conventional one-sided "superiority" design. In between these two cases, θ_0 could also represent a positive clinical effect, but the largest such effect which is either not meaningful or not worthwhile pursuing or both, and typically less efficacious than a standard active treatment. The alternative H_1: $\theta > \theta_0$ then specifies an effect that is both meaningful and worthwhile pursuing, even if it may be less effective than a standard active treatment. In this case the formulation would be called a single-sample non-inferiority design. To avoid confusion we will not distinguish between these various interpretations terminologically but will just refer to this hypothesis test formulation as a one-sided "conventional" design, wherein we test the null hypothesis of a “null” effect (meaning no clinical effect, inferiority, or no excess effect, as the case may be) against the alternative of “benefit” (a clinical effect, non-inferiority, or superiority, respectively, as the case may be). In this sense, then, Simon's original two-stage design with positive outcomes can be viewed as a conventional design, although as pointed out in Section 13, his primary goal was not to declare significant benefit but rather to discard the *treatment* if the null hypothesis cannot be ruled out. Indeed, because one does *not* reject the null hypothesis when the two-stage procedure stops after the first stage, this is a type of “futility” stopping rule akin to those sometimes used in phase III trials.

Unlike Simon's design, however, we may want to stop early with a signficant *rejection* of the null hypothesis if there are sufficiently many rapid successes in order to proceed to further confirmatory testing, or, lacking such a signal, we may wish to let the trial continue to its planned maximum sample size, giving the treatment its full due. These are precisely the conditions under which sequential boundaries are appealing to use with time-to-success data. Though we switch from time-to-failure to time-to-success, exactly the same boundaries and monitoring schemes presented in Sections 5 through 7 can be used with no changes needed and all of the operating characteristics previously discussed carry over directly to this context.

In addition to the test formulations presented thus far, there are two others we have not yet discussed, namely, a futility design (test of promise) with time-to-success outcomes (testing H_0: $\theta \geq \theta_0$ versus H_1: $\theta < \theta_0$) and a conventional design (test of null effect) with time-to-failure outcomes (testing H_0: $\theta \leq \theta_0$ versus H_1: $\theta > \theta_0$). Here the sequential tests work as expected with simultaneous entry, but they have less appeal with staggered entry. This is because a boundary crossing may occur based on observations from those *currently* enrolled, but early events from *future* subjects could overturn the rejection after reordering of the event times. For example, in the test of promise with time-to-success outcomes, given a boundary $\boldsymbol{b} = (b_1, \ldots, b_K)$, we would look to reject H_0 upon boundary upcrossings (rather than downcrossings for failure-time monitoring) due to long waiting times to success or too few successes due to censored waiting times. With staggered entry, we might cross a boundary with $k \leq K$ subjects under follow-up, but once enrolled, future subjects could have short waiting times to success which would overturn the futility declaration because there would be no boundary upcrossing. With staggered entry, then, a decision to reject H_0 at a given boundary crossing must be deferred until one observes all current and not-yet enrolled subjects exceed the waiting time corresponding to that boundary crossing before a rejection can be made. This detracts from the appeal of the sequential test procedure. These issues do not arise in the case of a test of promise with time-to-failure data or the conventional design with time-to-success data because, as explained in Sections 5 and 7, rejection of H_0 occurs with staggered entry if and only if it occurs with simultaneous entry. This is a consequence of rejection arising from *downcrossings* of a non-decreasing boundary.

The schematic diagram below summarizes the various comparisons discussed above.

Finally, we note that tests of promise have application beyond the specific example provided within the broader field of pediatric HIV. While the results of LEOPARD suggest that early treatment on its own is insufficient to lead to a sizable number of children achieving remission [3], early treatment has been clearly demonstrated to reduce the size of the viral reservoir [13]. There remains optimism in the field that a smaller viral reservoir combined with the plasticity of the young immune system makes this group of children uniquely amenable to novel, immune-based interventions designed to attain HIV cure [2]. In designing these trials, many of the design features discussed here are of relevance. Since treatment interruption is necessary to directly test for evidence of remission, the sequential aspects of the design are particularly pertinent [5]. The number of children likely to be eligible for these trials is also not large, raising the importance of designs that maximize what can be learned from small numbers of subjects.

Outcome type	Test of promise (futility design)	Test of no promise (conventional design)
Failure	H_0: $P \leq P_0$ vs. H_1: $P > P_0$ (promise vs. lack of promise) *Simon two-stage design:* Stop and do NOT reject H_0 if $X_1 \leq r_1$ or $X_1 > r_1$ and $X_1 + X_2 \leq r$, else reject H_0 and declare lack of promise. *Sequential test with time-to-event data:* Stop and REJECT H_0 if sample path DOWNcrosses boundary with sufficiently many early failures and declare lack of promise.	H_0: $P \geq P_0$ vs. H_1: $P < P_0$ (lack of promise vs. promise) *Simon two-stage design:* Stop and do NOT reject H_0 if $X_1 \geq r_1$ or $X_1 < r_1$ and $X_1 + X_2 \geq r$, else reject H_0 and declare promise. *Sequential test with time-to-event data:* Stop and REJECT H_0 if sample path UPcrosses boundary with sufficiently few late failures and declare promise.
Success	H_0: $P \geq P_0$ vs. H_1: $P < P_0$ (promise vs. lack of promise) *Simon two-stage design:* Stop and do NOT reject H_0 if $X_1 \geq r_1$ or $X_1 < r_1$ and $X_1 + X_2 \geq r$, else reject H_0 and declare lack of promise. *Sequential test with time-to-event data:* Stop and REJECT H_0 if sample path UPcrosses boundary with sufficiently few late successes and declare lack of promise.	H_0: $P \leq P_0$ vs. H_1: $P > P_0$ (lack of promise vs. promise) *Simon two-stage design:* Stop and do NOT reject H_0 if $X_1 \leq r_1$ or $X_1 > r_1$ and $X_1 + X_2 \leq r$, else reject H_0 and declare promise. *Sequential test with time-to-event data:* Stop and REJECT H_0 if sample path DOWNcrosses boundary with sufficiently many early successes and declare promise.

Acknowledgement

LEOPARD was supported by the Eunice Kennedy Shriver National Institutes of Child Health and Human Development (NICHD) and the National Institutes of Allergy and Infectious Diseases (NIAID) U01 HD080441.

Appendix 1. Computing algorithms

We assume the availability of a computer program to calculate, for any given boundary $\boldsymbol{b} = (b_1, \ldots, b_K)$, the type I error at the superiority/non-superiority margin and the power of the test of promise at the design alternative, using the methods given in Sections 8 and 9. Here we consider a subset of boundaries of a pre-specified form, as follows. Let ζ be a given integer with $0 < \zeta < K$ which specifies the number of leading boundary values which shall be fixed equal to 0. There will typically be at least one leading zero required to limit the type I error probability to no more than α (and in any case we might not wish to stop the trial after only a single failure). Boundary values to the right of ζ are permitted to equal 0 so long as the boundary is non-decreasing. Also let κ be a given integer with $\zeta < \kappa \leq K$ such that the final $K-\kappa + 1$ boundary values must equal M. We impose this constraint as an external design feature, to require a rejection of H_0 if κ or more failures are observed. Below we state an algorithm to search for the most powerful boundary subject to the design features specified by ζ and κ and the type I error constraint.

We first present an intermediate algorithm named *RUNUP* which, given an input boundary, returns a new, more powerful boundary. It is obtained by "running up" the given boundary values, i.e., incrementing them within the subset of boundaries under consideration as far as possible while maintaining the type I error constraint. The result is not necessarily an optimal boundary, but *RUNUP* will be used in the search to improve power at various stages of the main algorithm.

RUNUP

Input: Any given boundary of the form $\boldsymbol{b} = (0,\dots,0,b_{\varsigma+1},\dots,b_{\kappa-1},M,\dots,M)$ with type I error probability $\leq \alpha$; and a pointer index, π, with $\zeta \leq \pi \leq \kappa–1$.

Output: A new boundary $\boldsymbol{b'}$ with $b_j' = b_j$ for $1 \leq j \leq \pi$ and $\kappa \leq j \leq K$, and $b_j' \geq b_j$ for $\pi + 1 \leq j \leq \kappa–1$, with type I error $\leq \alpha$ and power no less than that of $\boldsymbol{b}$.

Start: Evaluate the type I error probability of the current boundary $\boldsymbol{b}$; call that error probability e.

If $e \leq \alpha$, check if $b_{\pi+1} = M$. If so, go to *Exit*, returning current $\boldsymbol{b}$ as output $\boldsymbol{b'}$. If not, then identify the positions $j = \pi + 1, \dots, \kappa–1$ for which $b_j < b_{\pi+1} + 1$ and for such j, set $b_j \leftarrow b_{\pi+1} + 1$. Go to *Start*.

If $e > \alpha$, reduce the boundary value to the right of the pointer in position $\pi + 1$ by 1 and increment the pointer by 1. That is, set $b_{\pi+1} \leftarrow b_{\pi+1}– 1$ and $\pi \leftarrow \pi + 1$. Go to *Start*.

EXIT: END RUNUP

The main algorithm *OPTIMIZE* recursively examines given boundaries, decreasing certain component boundary values in hopes of being able to increase subsequent boundary values using *RUNUP* to increase power while still satisfying the constraint on the type I error probability. We shall use certain lists and arrays of varying length, as follows. Operating characteristics will be recorded in a two-column array *OC*; boundaries will be recorded in a *K*-column array named *BDRYS*; and certain lists will be collected in a (K–ζ) column array named *DCRS* (short for decrements). A (K–ζ)-vector variable *LIST* will be used to denote given rows of *DCRS*.

The interpretation of a *LIST* vector is that for a given row of *BDRYS* (call it *BDRY*), *LIST* encodes a new boundary to consider wherein after the first ζ zeros, each element of *BDRY* is reduced by the corresponding component of *LIST*. It will be convenient to speak of the *truncated* version of *LIST*, which drops all trailing zeros from the right of *LIST* (so that the truncated version will have varying lengths depending on the number of trailing zeros in *LIST*); and the *expanded* version of *LIST*, which appends ζ zeros on the left (so that the expanded version has length *K*). Thus given a boundary *BDRY* of length *K* and a (K–ζ)-vector *LIST*, the new boundary to consider is given by *BDRY*–expanded *LIST*.

For example, suppose $K = 10$, $M = 12$, $\zeta = 3$, and $\kappa = 10$, and furthermore, suppose the first row of *DCRS* is (0, 0, 0, 0, 0, 0, 0) and the first row of *BDRYS* is *BDRY* = (0, 0, 0, 2, 4, 5, 5, 6, 6, 12). Now suppose *LIST* is the 7-vector (0, 1, 0, 0, 0, 0, 0) or (0, 1) after truncation. The next boundary considered will be (0, 0, 0, 2, 4, 5, 5, 6, 6, 12)–(0, 0, 0, *LIST*) = (0, 0, 0, 2, 3, 5, 5, 6, 6, 12).

For any given *LIST* vector, we define the *parent list* as follows: drop the last component of the truncated list and then pad the result with trailing zeros to restore length K–ζ. The parent list will correspond to a unique boundary in *BDRYS*, which we call the *parent boundary*, in the same row position as occupied by the parent list appearing in *DCRS*. In the above example, the truncated *LIST* (0, 1) has truncated parent (0) which appears in the first row of *DCRS*, and correspondingly the parent of boundary (0, 0, 0, 2, 3, 5, 5, 6, 6, 12) is (0, 0, 0, 2, 4, 5, 5, 6, 6, 12) in the first row of *BDRYS*.

Initialization step for *OPTIMIZE*: Initialize *OC* as a 1×2 array, *BDRYS* as a $1 \times K$ array, and *DCRS* as a $1 \times (K–\zeta)$ array. Set $\boldsymbol{b} = (0, \dots, 0, M, \dots, M)$ with $K–\kappa + 1$ *M*'s. Execute *RUNUP* with boundary $\boldsymbol{b}$ and pointer index $\pi = \zeta$ as input. Record the boundary output from *RUNUP* in the first row of *BDRYS* and its type I error and power in the first row of *OC*. Assign the zero vector to the first row of *DCRS*. Then set the (K–ζ)-vector *LIST* to (0,1, 0…, 0), so that truncated *LIST* = (0,1).

OPTIMIZE (a recursive algorithm)

Input: A truncated *LIST* vector. Parameters K, M, P_0, P_1, α, ζ, and κ are global fixed parameters.

Output: Arrays *BDRYS*, *DCRS*, and *OC*. When the recursive algorithm terminates, *BDRYS* contains an optimal boundary $\boldsymbol{b} = (b_1, \dots, b_K)$, i.e., one having greatest power among all boundaries with type I error probability $\leq \alpha$ among the subset of boundaries with $b_1 = \cdots = b_\zeta = 0$ and $b_\kappa = \cdots = b_K = M$.

Step 1: Let β equal the component in position $\zeta + 1$ of the boundary in the first row of *BDRYS*. If *LIST* equals (β + 1,0, …,0), then exit out of all recursive levels. *OPTIMIZE* terminates.

If not, then find the parent boundary for the current value of *LIST*. Call it *PARENT*. Set the pointer π to ζ plus the length of the truncated *LIST*. Go to *Step 4* if *PARENT* in the pointer position π is either equal to *M* or strictly less than the last element of truncated *LIST* plus *PARENT* in position π–1. In symbols, let λ denote the last element of the truncated *LIST*. Then go to *Step 4* if either *PARENT*[π] = *M* or *PARENT*[π] < *PARENT*[π–1] + λ.

If neither condition holds, check whether *LIST* is already in *DCRS*. If it is, *Exit* the current recursive level of *OPTIMIZE*. Otherwise create a new variable *BDRY*, set it equal to *PARENT*, and then reduce the component in position π by λ: *BDRY* ← *PARENT*, *BDRY*[π] ← *PARENT*[π]–λ. Apply *RUNUP* to *BDRY* with pointer π. Store the output boundary in *BDRY* and append it after the bottom row of *BDRYS*; append *LIST* after the bottom row of *DCRS*; and append the type I error probability and power after the bottom row *OC*.

Step 2 (going to a deeper LIST level): If *BDRY*[π + 1] = *M*, go to *Step 3*. Otherwise, invoke a recursive call to *OPTIMIZE* with truncated input list given by ((truncated *LIST*),1). After the recursive call returns to this point, *Exit* the current recursive level of *OPTIMIZE*.

Step 3 (going to a deeper decrement): If *PARENT*[π + 1] = *M*, then go to *Step 4*. Otherwise, increment the last component of the truncated *LIST* by 1: *LIST*[λ] ← *LIST*[λ] + 1 where λ is the right-most element of truncated *LIST*. If *PARENT*[π] < *PARENT*[π–1] + λ then go to *Step 4*. Otherwise, invoke a recursive call to *OPTIMIZE* with the current value of truncated *LIST* as input. After the recursive call returns to this point, go back to the beginning of this step (*Step 3*).

Step 4 (going to a shallower LIST level): Drop the right-most component of truncated *LIST*. If this would result in an empty *LIST*, then *Exit* the current recursive level of *OPTIMIZE*. Otherwise, increment the last element of truncated *LIST* by 1. In symbols, if *LIST* = (x_1, …, x_λ, $x_{\lambda+1}$, 0, …,0) at the start of this step, such that truncated *LIST* = (x_1, …, x_λ, $x_{\lambda+1}$), then assign (x_1, …, x_λ + 1) as the new truncated *LIST* so that *LIST* ←(x_1, …, x_λ + 1, 0, …,0) with an additional 0 component on the right. Then go to *Step 1*.

EXIT: END OPTIMIZE

The table below provides the arrays *DCRS*, *BDRYS*, and *OC* for the example discussed in Section 6 with $K = 10$, $M = 12$, $P_0 = 0.50$, $P_1 = 0.90$, $\alpha = 0.05$, $\zeta = 3$, and $\kappa = 10$. The optimal boundary is highlighted in yellow.

DCRS	*BDRYS*	*Type I*	*Power*
0 0 0 0 0 0 0	0 0 0 2 4 5 5 6 6 12	0.0499865121	0.822699699
0 1 0 0 0 0 0	0 0 0 2 3 5 7 9 10 12	0.0488959257	0.889193828
0 1 1 0 0 0 0	0 0 0 2 3 4 7 9 12 12	0.047563054	0.906439746
0 1 1 1 0 0 0	0 0 0 2 3 4 6 10 12 12	0.0485624861	0.911441795
0 1 1 2 0 0 0	0 0 0 2 3 4 5 10 12 12	0.0468956233	0.906528631
0 1 1 3 0 0 0	0 0 0 2 3 4 4 10 12 12	0.0465577187	0.905150803
0 1 2 0 0 0 0	0 0 0 2 3 3 7 10 11 12	0.0499994679	0.907247794
0 1 2 1 0 0 0	0 0 0 2 3 3 6 10 12 12	0.0474261872	0.908994457
0 1 2 2 0 0 0	0 0 0 2 3 3 5 10 12 12	0.045594742	0.903312625
0 1 2 3 0 0 0	0 0 0 2 3 3 4 10 12 12	0.0450824684	0.901003533
0 1 2 4 0 0 0	0 0 0 2 3 3 3 10 12 12	0.0450276398	0.900661542
0 2 0 0 0 0 0	0 0 0 2 2 5 7 9 12 12	0.0499343666	0.911304234
0 2 1 0 0 0 0	0 0 0 2 2 4 7 10 12 12	0.0498989324	0.916734236
0 2 1 1 0 0 0	0 0 0 2 2 4 6 10 12 12	0.0453453194	0.907054991
0 2 1 2 0 0 0	0 0 0 2 2 4 5 10 12 12	0.0435571569	0.901526539
0 2 1 3 0 0 0	0 0 0 2 2 4 4 10 12 12	0.0431755981	0.899884876
0 2 2 0 0 0 0	0 0 0 2 2 3 7 10 12 12	0.0483975979	0.913676872
0 2 2 1 0 0 0	0 0 0 2 2 3 6 10 12 12	0.0436062315	0.902927625
0 2 2 2 0 0 0	0 0 0 2 2 3 5 10 12 12	0.0415661782	0.896102835
0 2 2 3 0 0 0	0 0 0 2 2 3 4 10 12 12	0.0409177503	0.892890623
0 2 2 4 0 0 0	0 0 0 2 2 3 3 10 12 12	0.0408139215	0.892161376
0 2 3 0 0 0 0	0 0 0 2 2 2 7 10 12 12	0.0482326802	0.91321725
0 2 3 1 0 0 0	0 0 0 2 2 2 6 10 12 12	0.0434151972	0.902307146
0 2 3 2 0 0 0	0 0 0 2 2 2 5 10 12 12	0.0413474743	0.895287474
0 2 3 3 0 0 0	0 0 0 2 2 2 4 10 12 12	0.0406697315	0.891839157
0 2 3 4 0 0 0	0 0 0 2 2 2 3 10 12 12	0.0405348447	0.890823862
0 2 3 5 0 0 0	0 0 0 2 2 2 2 10 12 12	0.0405271536	0.89074187
1 0 0 0 0 0 0	0 0 0 1 4 5 6 8 11 12	0.0493031357	0.883687581
1 1 0 0 0 0 0	0 0 0 1 3 5 7 10 11 12	0.0481141355	0.910799203
1 1 1 0 0 0 0	0 0 0 1 3 4 7 10 12 12	0.044860056	0.914088332
1 1 1 1 0 0 0	0 0 0 1 3 4 6 10 12 12	0.0402500875	0.904159059
1 1 1 2 0 0 0	0 0 0 1 3 4 5 11 12 12	0.0490381738	0.921991297
1 1 1 3 0 0 0	0 0 0 1 3 4 4 11 12 12	0.0487168645	0.920895235
1 1 2 0 0 0 0	0 0 0 1 3 3 7 10 12 12	0.0437206996	0.9118431
1 1 2 1 0 0 0	0 0 0 1 3 3 6 11 12 12	0.049365786	0.923596673
1 1 2 2 0 0 0	0 0 0 1 3 3 5 11 12 12	0.0476552129	0.919142595
1 1 2 3 0 0 0	0 0 0 1 3 3 4 11 12 12	0.0471593704	0.917262116
1 1 2 4 0 0 0	0 0 0 1 3 3 3 11 12 12	0.0471035508	0.916968522
1 2 0 0 0 0 0	0 0 0 1 2 6 6 8 11 12	0.0476154897	0.889749778
1 2 1 0 0 0 0	0 0 0 1 2 5 8 9 11 12	0.0483665592	0.906893837
1 2 1 1 0 0 0	0 0 0 1 2 5 7 10 12 12	0.0458586308	0.916760543
1 2 1 2 0 0 0	0 0 0 1 2 5 6 10 12 12	0.0418907472	0.908658033
1 2 1 3 0 0 0	0 0 0 1 2 5 5 10 12 12	0.0409016089	0.905878494
1 2 2 0 0 0 0	0 0 0 1 2 4 8 10 12 12	0.0496017927	0.922722274
1 2 2 1 0 0 0	0 0 0 1 2 4 7 11 12 12	0.0497991133	0.927510559
1 2 2 2 0 0 0	0 0 0 1 2 4 6 11 12 12	0.0455628591	0.919782113
1 2 2 3 0 0 0	0 0 0 1 2 4 5 11 12 12	0.0438450655	0.915204761
1 2 2 4 0 0 0	0 0 0 1 2 4 4 11 12 12	0.0434618707	0.913778209
1 2 3 0 0 0 0	0 0 0 1 2 3 8 10 12 12	0.0479046677	0.91949564
1 2 3 1 0 0 0	0 0 0 1 2 3 7 11 12 12	0.0479379263	0.924142108
1 2 3 2 0 0 0	0 0 0 1 2 3 6 11 12 12	0.0434293565	0.915333671
1 2 3 3 0 0 0	0 0 0 1 2 3 5 11 12 12	0.0414230544	0.909447879
1 2 3 4 0 0 0	0 0 0 1 2 3 4 11 12 12	0.0407341957	0.906436116
1 2 3 5 0 0 0	0 0 0 1 2 3 3 11 12 12	0.040608912	0.90565743
1 2 4 0 0 0 0	0 0 0 1 2 2 8 10 12 12	0.0476131761	0.918717499
1 2 4 1 0 0 0	0 0 0 1 2 2 7 11 12 12	0.0476182561	0.923329767
1 2 4 2 0 0 0	0 0 0 1 2 2 6 11 12 12	0.0430629144	0.914260876
1 2 4 3 0 0 0	0 0 0 1 2 2 5 11 12 12	0.0410070593	0.908059538
1 2 4 4 0 0 0	0 0 0 1 2 2 4 11 12 12	0.040265701	0.904665483
1 2 4 5 0 0 0	0 0 0 1 2 2 3 11 12 12	0.0400847958	0.90342364
1 2 4 6 0 0 0	0 0 0 1 2 2 2 11 12 12	0.040067977	0.903259399
1 3 0 0 0 0 0	0 0 0 1 1 6 7 8 11 12	0.0495765385	0.896131235
1 3 1 0 0 0 0	0 0 0 1 1 5 8 9 11 12	0.0477788351	0.905859173
1 3 1 1 0 0 0	0 0 0 1 1 5 7 10 12 12	0.0452482457	0.915804736
1 3 1 2 0 0 0	0 0 0 1 1 5 6 10 12 12	0.0412543945	0.907546983
1 3 1 3 0 0 0	0 0 0 1 1 5 5 11 11 12	0.0496297324	0.921525937
1 3 1 3 1 0 0	0 0 0 1 1 5 5 10 12 12	0.0402559106	0.904700877
1 3 1 3 2 0 0	0 0 0 1 1 5 5 9 12 12	0.0341180077	0.885882018
1 3 1 3 3 0 0	0 0 0 1 1 5 5 8 12 12	0.0308399793	0.87111795
1 3 1 3 4 0 0	0 0 0 1 1 5 5 7 12 12	0.0293789873	0.861468404
1 3 1 3 5 0 0	0 0 0 1 1 5 5 6 12 12	0.0289164739	0.857043297
1 3 1 3 6 0 0	0 0 0 1 1 5 5 5 12 12	0.028859296	0.856287149
1 3 2 0 0 0 0	0 0 0 1 1 4 8 10 12 12	0.0489308089	0.921692413
1 3 2 1 0 0 0	0 0 0 1 1 4 7 11 12 12	0.0491041268	0.926497075
1 3 2 2 0 0 0	0 0 0 1 1 4 6 11 12 12	0.0448293866	0.9185725
1 3 2 3 0 0 0	0 0 0 1 1 4 5 11 12 12	0.0430864739	0.913840174
1 3 2 4 0 0 0	0 0 0 1 1 4 4 11 12 12	0.0426942391	0.91234717
1 3 3 0 0 0 0	0 0 0 1 1 3 8 10 12 12	0.0471273296	0.918138034
1 3 3 1 0 0 0	0 0 0 1 1 3 7 11 12 12	0.0471263042	0.922786474
1 3 3 2 0 0 0	0 0 0 1 1 3 6 11 12 12	0.0425621831	0.913672207
1 3 3 3 0 0 0	0 0 0 1 1 3 5 11 12 12	0.040512682	0.907498537
1 3 3 4 0 0 0	0 0 0 1 1 3 4 11 12 12	0.0397956281	0.904259304
1 3 3 5 0 0 0	0 0 0 1 1 3 3 11 12 12	0.0396601973	0.903383079
1 3 4 0 0 0 0	0 0 0 1 1 2 8 10 12 12	0.0467231597	0.91696282
1 3 4 1 0 0 0	0 0 0 1 1 2 7 11 12 12	0.0466830629	0.921559608
1 3 4 2 0 0 0	0 0 0 1 1 2 6 11 12 12	0.0420540899	0.912051983
1 3 4 3 0 0 0	0 0 0 1 1 2 5 11 12 12	0.0399358806	0.905401749
1 3 4 4 0 0 0	0 0 0 1 1 2 4 11 12 12	0.0391460329	0.901585146
1 3 4 5 0 0 0	0 0 0 1 1 2 3 11 12 12	0.0389334798	0.900009422
1 3 4 6 0 0 0	0 0 0 1 1 2 2 11 12 12	0.0389052713	0.899702014
1 3 5 0 0 0 0	0 0 0 1 1 1 8 10 12 12	0.0466907061	0.916829216
1 3 5 1 0 0 0	0 0 0 1 1 1 7 11 12 12	0.0466474721	0.921420132
1 3 5 2 0 0 0	0 0 0 1 1 1 6 11 12 12	0.0420132917	0.911867788
1 3 5 3 0 0 0	0 0 0 1 1 1 5 11 12 12	0.0398895654	0.905163376
1 3 5 4 0 0 0	0 0 0 1 1 1 4 11 12 12	0.0390938726	0.901281135
1 3 5 5 0 0 0	0 0 0 1 1 1 3 11 12 12	0.0388751267	0.899625888
1 3 5 6 0 0 0	0 0 0 1 1 1 2 11 12 12	0.0388403574	0.899222136
1 3 5 7 0 0 0	0 0 0 1 1 1 1 11 12 12	0.0388388338	0.89919617
2 0 0 0 0 0 0	0 0 0 0 4 5 6 8 11 12	0.0488181881	0.883056227
2 1 0 0 0 0 0	0 0 0 0 3 5 7 10 12 12	0.0494625276	0.920994858
2 1 1 0 0 0 0	0 0 0 0 3 4 7 10 12 12	0.0442522772	0.913469299
2 1 1 1 0 0 0	0 0 0 0 3 4 6 11 12 12	0.0499800247	0.925252595
2 1 1 2 0 0 0	0 0 0 0 3 4 5 11 12 12	0.0484259527	0.92139994
2 1 1 3 0 0 0	0 0 0 0 3 4 4 11 12 12	0.0481023758	0.920288142
2 1 2 0 0 0 0	0 0 0 0 3 3 7 10 12 12	0.0431009912	0.911181272
2 1 2 1 0 0 0	0 0 0 0 3 3 6 11 12 12	0.0487490458	0.923009397
2 1 2 2 0 0 0	0 0 0 0 3 3 5 11 12 12	0.0470285115	0.91849694
2 1 2 3 0 0 0	0 0 0 0 3 3 4 11 12 12	0.0465285741	0.916585774
2 1 2 4 0 0 0	0 0 0 0 3 3 3 11 12 12	0.0464721091	0.916286142
2 2 0 0 0 0 0	0 0 0 0 2 6 7 8 11 12	0.0494525179	0.896312022
2 2 1 0 0 0 0	0 0 0 0 2 5 8 9 12 12	0.0499192727	0.919220931
2 2 1 1 0 0 0	0 0 0 0 2 5 7 10 12 12	0.0451606107	0.916026263
2 2 1 2 0 0 0	0 0 0 0 2 5 6 10 12 12	0.0411744864	0.907821928
2 2 1 3 0 0 0	0 0 0 0 2 5 5 10 12 12	0.0401794631	0.905002039
2 2 2 0 0 0 0	0 0 0 0 2 4 8 10 12 12	0.0488862594	0.921994105
2 2 2 1 0 0 0	0 0 0 0 2 4 7 11 12 12	0.0490703566	0.926806464
2 2 2 2 0 0 0	0 0 0 0 2 4 6 11 12 12	0.044810603	0.918964638
2 2 2 3 0 0 0	0 0 0 0 2 4 5 11 12 12	0.0430794952	0.914306381
2 2 2 4 0 0 0	0 0 0 0 2 4 4 11 12 12	0.0426922054	0.912849141
2 2 3 0 0 0 0	0 0 0 0 2 3 8 10 12 12	0.0471574884	0.918663298
2 2 3 1 0 0 0	0 0 0 0 2 3 7 11 12 12	0.0471744645	0.923329261
2 2 3 2 0 0 0	0 0 0 0 2 3 6 11 12 12	0.0426373173	0.914372576
2 2 3 3 0 0 0	0 0 0 0 2 3 5 11 12 12	0.0406123215	0.908363635
2 2 3 4 0 0 0	0 0 0 0 2 3 4 11 12 12	0.0399136681	0.905270005
2 2 3 5 0 0 0	0 0 0 0 2 3 3 11 12 12	0.0397856878	0.904463335
2 2 4 0 0 0 0	0 0 0 0 2 2 8 10 12 12	0.0468546079	0.917840487
2 2 4 1 0 0 0	0 0 0 0 2 2 7 11 12 12	0.0468423044	0.922470287
2 2 4 2 0 0 0	0 0 0 0 2 2 6 11 12 12	0.0422565579	0.913238197
2 2 4 3 0 0 0	0 0 0 0 2 2 5 11 12 12	0.0401800729	0.906895596
2 2 4 4 0 0 0	0 0 0 0 2 2 4 11 12 12	0.0394268687	0.903397729
2 2 4 5 0 0 0	0 0 0 0 2 2 3 11 12 12	0.0392410937	0.902101314
2 2 4 6 0 0 0	0 0 0 0 2 2 2 11 12 12	0.0392235074	0.901926335
2 3 0 0 0 0 0	0 0 0 0 1 6 7 8 11 12	0.0488180267	0.895017559
2 3 1 0 0 0 0	0 0 0 0 1 5 8 9 12 12	0.0492121755	0.918102543
2 3 1 1 0 0 0	0 0 0 0 1 5 7 10 12 12	0.0444194993	0.914809915
2 3 1 2 0 0 0	0 0 0 0 1 5 6 10 12 12	0.0404018458	0.906408022
2 3 1 3 0 0 0	0 0 0 0 1 5 5 11 12 12	0.0496663214	0.92524134
2 3 2 0 0 0 0	0 0 0 0 1 4 8 10 12 12	0.0480715708	0.920683519
2 3 2 1 0 0 0	0 0 0 0 1 4 7 11 12 12	0.0482265247	0.925516719
2 3 2 2 0 0 0	0 0 0 0 1 4 6 11 12 12	0.0439200424	0.917425302
2 3 2 3 0 0 0	0 0 0 0 1 4 5 11 12 12	0.0421584359	0.912569827
2 3 2 4 0 0 0	0 0 0 0 1 4 4 11 12 12	0.04176017	0.911028021
2 3 3 0 0 0 0	0 0 0 0 1 3 8 10 12 12	0.0462136676	0.916935627
2 3 3 1 0 0 0	0 0 0 0 1 3 7 11 12 12	0.0461890171	0.9216041
2 3 3 2 0 0 0	0 0 0 0 1 3 6 11 12 12	0.0415844212	0.91225822
2 3 3 3 0 0 0	0 0 0 0 1 3 5 11 12 12	0.0395069743	0.905882929
2 3 3 4 0 0 0	0 0 0 0 1 3 4 11 12 12	0.0387740872	0.902499824
2 3 3 5 0 0 0	0 0 0 0 1 3 3 11 12 12	0.0386337866	0.901569028
2 3 4 0 0 0 0	0 0 0 0 1 2 8 10 12 12	0.0457739764	0.915607508
2 3 4 1 0 0 0	0 0 0 0 1 2 7 11 12 12	0.0457068206	0.920217607
2 3 4 2 0 0 0	0 0 0 0 1 2 6 11 12 12	0.0410316732	0.910427191
2 3 4 3 0 0 0	0 0 0 0 1 2 5 11 12 12	0.0388794796	0.903513331
2 3 4 4 0 0 0	0 0 0 0 1 2 4 11 12 12	0.0380674011	0.899477735
2 3 4 5 0 0 0	0 0 0 0 1 2 3 11 12 12	0.0378432	0.897756428
2 3 4 6 0 0 0	0 0 0 0 1 2 2 11 12 12	0.0378117847	0.897399257
2 3 5 0 0 0 0	0 0 0 0 1 1 8 10 12 12	0.0457287392	0.915408338
2 3 5 1 0 0 0	0 0 0 0 1 1 7 11 12 12	0.0456572103	0.920009684
2 3 5 2 0 0 0	0 0 0 0 1 1 6 11 12 12	0.0409748043	0.910152602
2 3 5 3 0 0 0	0 0 0 0 1 1 5 11 12 12	0.0388149204	0.903157977
2 3 5 4 0 0 0	0 0 0 0 1 1 4 11 12 12	0.0379946943	0.899024531
2 3 5 5 0 0 0	0 0 0 0 1 1 3 11 12 12	0.0377618613	0.897184676
2 3 5 6 0 0 0	0 0 0 0 1 1 2 11 12 12	0.0377213007	0.896683881
2 3 5 7 0 0 0	0 0 0 0 1 1 1 11 12 12	0.0377188645	0.896638821
2 4 0 0 0 0 0	0 0 0 0 0 6 7 8 11 12	0.0487943118	0.894950375
2 4 1 0 0 0 0	0 0 0 0 0 5 8 9 12 12	0.0491857469	0.918044497
2 4 1 1 0 0 0	0 0 0 0 0 5 7 10 12 12	0.0443917994	0.914746785
2 4 1 2 0 0 0	0 0 0 0 0 5 6 10 12 12	0.0403729674	0.906334638
2 4 1 3 0 0 0	0 0 0 0 0 5 5 11 12 12	0.049638397	0.925183089
2 4 2 0 0 0 0	0 0 0 0 0 4 8 10 12 12	0.0480411208	0.920615497
2 4 2 1 0 0 0	0 0 0 0 0 4 7 11 12 12	0.0481949854	0.925449779
2 4 2 2 0 0 0	0 0 0 0 0 4 6 11 12 12	0.0438867566	0.917345408
2 4 2 3 0 0 0	0 0 0 0 0 4 5 11 12 12	0.0421240102	0.912479697
2 4 2 4 0 0 0	0 0 0 0 0 4 4 11 12 12	0.041725334	0.910933502
2 4 3 0 0 0 0	0 0 0 0 0 3 8 10 12 12	0.0461783912	0.916845958
2 4 3 1 0 0 0	0 0 0 0 0 3 7 11 12 12	0.0461521848	0.921514562
2 4 3 2 0 0 0	0 0 0 0 0 3 6 11 12 12	0.0415450679	0.912148482
2 4 3 3 0 0 0	0 0 0 0 0 3 5 11 12 12	0.0394656606	0.905754177
2 4 3 4 0 0 0	0 0 0 0 0 3 4 11 12 12	0.038731494	0.902356047
2 4 3 5 0 0 0	0 0 0 0 0 3 3 11 12 12	0.0385907329	0.901418809
2 4 4 0 0 0 0	0 0 0 0 0 2 8 10 12 12	0.0457335865	0.915491613
2 4 4 1 0 0 0	0 0 0 0 0 2 7 11 12 12	0.0456643806	0.92010069
2 4 4 2 0 0 0	0 0 0 0 0 2 6 11 12 12	0.0409858917	0.910281295
2 4 4 3 0 0 0	0 0 0 0 0 2 5 11 12 12	0.0388308683	0.903337786
2 4 4 4 0 0 0	0 0 0 0 0 2 4 11 12 12	0.0380165893	0.899274281
2 4 4 5 0 0 0	0 0 0 0 0 2 3 11 12 12	0.0377909521	0.897530922
2 4 4 6 0 0 0	0 0 0 0 0 2 2 11 12 12	0.0377590199	0.897164295
2 4 5 0 0 0 0	0 0 0 0 0 1 8 10 12 12	0.0456829317	0.915260669
2 4 5 1 0 0 0	0 0 0 0 0 1 7 11 12 12	0.045608829	0.919859596
2 4 5 2 0 0 0	0 0 0 0 0 1 6 11 12 12	0.0409222122	0.909962902
2 4 5 3 0 0 0	0 0 0 0 0 1 5 11 12 12	0.0387585776	0.902925742
2 4 5 4 0 0 0	0 0 0 0 0 1 4 11 12 12	0.0379351753	0.898748777
2 4 5 5 0 0 0	0 0 0 0 0 1 3 11 12 12	0.0376998723	0.896867957
2 4 5 6 0 0 0	0 0 0 0 0 1 2 11 12 12	0.0376576997	0.896334794
2 4 5 7 0 0 0	0 0 0 0 0 1 1 11 12 12	0.0376546833	0.896275854
2 4 6 0 0 0 0	0 0 0 0 0 0 8 10 12 12	0.0456819751	0.915254253
2 4 6 1 0 0 0	0 0 0 0 0 0 7 11 12 12	0.0456077799	0.919852898
2 4 6 2 0 0 0	0 0 0 0 0 0 6 11 12 12	0.0409210096	0.909954056
2 4 6 3 0 0 0	0 0 0 0 0 0 5 11 12 12	0.0387572124	0.902914295
2 4 6 4 0 0 0	0 0 0 0 0 0 4 11 12 12	0.0379336379	0.898734178
2 4 6 5 0 0 0	0 0 0 0 0 0 3 11 12 12	0.0376981523	0.89684954
2 4 6 6 0 0 0	0 0 0 0 0 0 2 11 12 12	0.0376557862	0.896311749
2 4 6 7 0 0 0	0 0 0 0 0 0 1 11 12 12	0.0376525649	0.896247205
2 4 6 8 0 0 0	0 0 0 0 0 0 0 11 12 12	0.0376525339	0.896246235

Appendix 2. Asymptotic considerations

When K is large, boundaries can also be set by using the asymptotic theory of empirical survival functions. Consider dual monitoring of the cumulative number of failures $Y_{(m)}$ with dual boundary $\boldsymbol{b}'$. Write $Y_{(m)} = \sum_{i=1}^{K} I[X_i \le m] = K\{1 - S_K^e(m)\}$, where $S_K{}^e(m)$ is the empirical survival function $S_K^e(m) = K^{-1}\sum_{i=1}^{K} I[X_i > m]$ for K independent and identically distributed event times with common survival function $S(m) = P[X_i > m \mid \theta] = (1\text{–}\theta)^m$. Then the event of a boundary crossing at some time $m = 1, \ldots, M$ can be written.

$$[Y_{(m)} \ge b'_m \quad \text{for some} \;\; m = 1,\ldots,M] = [K\{1 - S_K^e(m)\} \ge b'_m \quad \text{for some} \;\; m = 1,\ldots,M]$$
$$= \left[K\left\{1 - S(m)\exp\left(-K^{-1/2}\left\{\frac{1-S(m)}{S(m)}\right\}^{1/2} Z_K(m)\right)\right\} \ge b'_m \quad \text{for some} \;\; m = 1,\ldots,M\right]$$
$$= \left[Z_K(m) \ge K^{1/2}\frac{-\log\{(1-b'_m/K)/S(m)\}}{\left\{\frac{1-S(m)}{S(m)}\right\}^{1/2}} \quad \text{for some} \;\; m = 1,\ldots,M\right],$$

where in the second line we have used the transformation.

$$Z_K(m) = K^{1/2}\frac{-\log\{S_K^e/S(m)\}}{\left\{\frac{1-S(m)}{S(m)}\right\}^{1/2}} \text{for } m = 1, \ldots, M.$$

An application of the delta method shows that as $K \to \infty$, the random vector $Z_K(1), \ldots, Z_K(M)$ converges in distribution to a multivariate normal random vector $Z(1), \ldots, Z(M)$ with mean $(0, \ldots, 0)$ and covariance matrix $Cov\{Z(m), Z(m')\} = \left(\frac{S(m')\{1-S(m)\}}{S(m)\{1-S(m')\}}\right)^{1/2}$ for $1 \le m \le m' \le M$. It follows that the rejection probability can be approximated as

$$P\left[Z(m) \ge K^{1/2}\frac{-\log\{(1-b'_m/K)/S(m)\}}{\left\{\frac{1-S(m)}{S(m)}\right\}^{1/2}} \quad \text{for some} \;\; m = 1,\ldots,M\right].$$

This is generally tedious to calculate, but one simple class of boundaries can be obtained by setting the expression on the right-hand side of the inequality to a constant, chosen such that the approximate rejection probability is α, say $P[Z(m) \ge C_\alpha$ for some $m = 1,\ldots,M] = P[\max\{Z(1),\ldots,Z(M)\} \ge C_\alpha]$. We can then solve for the dual boundary. In practice, we have found that rounding the solutions up to the next largest integer and even adding 1 unit comes closest to achieving the desired control of type I error probability. Thus we have found a useful dual boundary in the large sample case to be

$$b'_m = 1 + \left\lceil K\left\{1 - S(m)\exp\left(-K^{-1/2}C_\alpha\left\{\frac{1-S(m)}{S(m)}\right\}^{1/2}\right)\right\}\right\rceil$$
$$= 1 + \left\lceil K\left\{1 - (1-\theta)^m\exp\left(-K^{-1/2}C_\alpha\left\{\frac{1-(1-\theta)^m}{(1-\theta)^m}\right\}^{1/2}\right)\right\}\right\rceil \quad \text{for} \;\; m = 1,\ldots,M. \tag{A2.1}$$

Given the computing algorithms described in Section 8, we don't actually need to evaluate the constant C_α from the normal distribution which in any case only provides approximate control of the type I error rate. Instead, we can try several different values of the constant C_α in Eq. (A2.1) and then evaluate the exact rejection probability for any given θ, such that with a bit of trial and error we can quickly find C_α and the corresponding dual boundary such that the exact type I error rate is limited to α.

We make no claim that the boundary Eq. (A2.1) is an optimal boundary or even a near optimal boundary as discussed in Section 14. However, we have found as a practical matter that using Eq. (A2.1) with a constant C_α providing type I error *greater* than α by a few percentage points, followed by two applications of the *RUNUP* algorithm described in Appendix 1 tends to produce near optimal boundaries. The first application of *RUNUP* produces a working boundary that reduces the type I error rate below α and then applying *RUNUP* to the working boundary improves the power. As an illustration, we calculated boundary Eq. (A2.1) for the case $K = 20$, $M = 12$, $P_a = 0.75$ under H_0 and $P_a = 0.95$ under H_1 with $\alpha = 0.10$ and power at least 95% discussed in the previous section. We used $C_\alpha = 1.7$ (whereas $C_\alpha = 2.04$ is the normal value) to find the dual boundary (6, 8, 10, 12, 13, 15, 16, 17, 17, 18, 18, 19) corresponding to geometric boundary (0, 0, 0, 0, 0, 1, 1, 2, 2, 3, 3, 4, 5, 5, 6, 7, 9, 11, 12, 12). This boundary has type I error probability 0.1427, but an application of *RUNUP* produced the boundary (0, 0, 0, 0, 0, 0, 0, 1, 1, 2, 2, 3, 4, 5, 6, 8, 9, 11, 12, 12) with type I error probability 0.0996 and power under $P_a = 0.95$ of 0.9589. This happens to be the optimal boundary in this case. The dual boundary is (8, 10, 12, 13, 14, 15, 16, 16, 17, 18, 18, 19).

References

[1] J.L. Fleiss, B. Levin, M.C. Paik, Statistical Methods for Rates and Proportions, 3rd ed., John Wiley & Sons, New York, 2003.

[2] N. Klein, P. Palma, K. Luzuriaga, S. Pahwa, E. Nastouli, D.M. Gibb, P. Rojo, W. Borkowsky, S. Bernardi, P. Zangari, V. Calvez, A. Compagnucci, B. Wahren, C. Foster, M.Á. Munoz-Fernández, A. De Rossi, J. Ananworanich, D. Pillay, C. Giaquinto, P. Rossi, Early antiretroviral therapy in children perinatally infected with HIV: a unique opportunity to implement immunotherapeutic approaches to prolong viral remission, Lancet Infect. Dis. 15 (2015) 1108–1114.

[3] L. Kuhn, K. Technau, R. Strehlau, S. Shiau, F. Patel, G. Sherman, C. Tiemessen, G. Aldrovandi, A. Coovadia, E.J. Abrams, Treatment of acute HIV infection in neonates, Conference on Retroviruses and Opportunistic Infections (CROI), 2017 February 13–16, 2017, Seattle, WA. Abstract number 27.

[4] B. Levin, The futility study—progress over the last decade, Contemp. Clin. Trials 45 (2015) 69–75 Pt A.

[5] J.Z. Li, D.M. Smith, J.W. Mellors, The need for treatment interruption studies and biomarker identification in the search for an HIV cure, AIDS 29 (2017) 1429–1432.

[6] G.E. Martin, M. Gossez, J.P. Williams, W. Stöhr, J. Meyerowitz, E.M. Leitman,

P. Goulder, K. Porter, S. Fidler, J. Frater, the SPARTAC Trial Investigators, Post-treatment control or treated controllers? Viral remission in treated and untreated primary HIV infection, AIDS (London, England) 31 (4) (2017) 477–484.

[7] G. Namazi, J.M. Fajnzylber, E. Aga, R.J. Bosch, E.P. Acosta, R. Sharaf, W. Hartogensis, J.M. Jacobson, E. Connick, P. Volberding, D. Skiest, D. Margolis, M.C. Sneller, S.J. Little, S. Gianella, D.M. Smith, D.R. Kuritzkes, R.M. Gulick, J.W. Mellors, V. Mehraj, R.T. Gandhi, R. Mitsuyasu, R.T. Schooley, K. Henry, P. Tebas, S.G. Deeks, T.W. Chun, A.C. Collier, J.P. Routy, F.M. Hecht, B.D. Walker, J.Z. Li, The control of HIV after antiretroviral medication pause (CHAMP) study: posttreatment controllers identified from 14 clinical studies, J. Infect. Dis. 218 (12) (2018) 1954–1963.

[8] D. Persaud, H. Gay, C. Ziemniak, Y.H. Chen, M. Piatak Jr., T.W. Chun, M. Strain, D. Richman, K. Luzuriaga, Absence of detectable HIV-1 viremia after treatment cessation in an infant, N. Engl. J. Med. 369 (19) (2013) 1828–1835.

[9] A. Sáez-Cirión, C. Bacchus, L. Hocqueloux, V. Avettand-Fenoel, I. Girault, C. Lecuroux, V. Potard, P. Versmisse, A. Melard, T. Prazuck, B. Descours, J. Guergnon, J.P. Viard, F. Boufassa, O. Lambotte, C. Goujard, L. Meyer, D. Costagliola, A. Venet, G. Pancino, B. Autran, C. Rouzioux, the ANRS VISCONTI Study Group, Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI study, PLoS Pathog. 9 (3) (2013) e1003211.

[10] S. Shiau, E.J. Abrams, S.M. Arpadi, L. Kuhn, Early antiretroviral therapy in HIV-infected infants: can it lead to HIV remission? Lancet HIV 5 (5) (2018) e250–e258.

[11] S. Shiau, L. Kuhn, Antiretroviral treatment in HIV-infected infants and young children: novel issues raised by the Mississippi baby, Expert Rev. Anti-Infect. Ther. 12 (3) (2014) 307–318.

[12] R. Simon, Optimal two-stage designs for phase II clinical trials, Control. Clin. Trials 10 (1989) 1–10.

[13] K.A. Veldsman, J. Maritz, S. Isaacs, M.G. Katusiime, A. Janse van Rensburg, B. Laughton, J.W. Mellors, M.F. Cotton, G.U. van Zyl, Rapid decline of HIV-1 DNA and RNA in infants starting very early antiretroviral therapy may pose a diagnostic challenge, AIDS 32 (2018) 629–634.