A Simple Low-Bias Estimate Following a Sequential Test With Linear Boundaries

By Aiyi Liu

National Institute of Child Health and Human Development

We suppose a Brownian motion with drift and linear stopping boundaries, possibly truncated, is observed. The maximum likelihood estimate of the drift, upon reaching a boundary, is known to be badly biased, while the uniformly minimum variance unbiased estimate of Liu and Hall (1998, *Sequential Analysis* **17**: 91-107) is difficult to compute. A bias-adjusted estimate of Whitehead (1986, *Biometrika* **73**: 573-581) is also complex computationally and may still have substantial bias. We propose a modification of the maximum likelihood estimate, a *segmented estimate*, which has a simple explicit formula. Computation shows the estimate to have little bias and to have a competitive mean-square-error. The results apply to various sequential testing problems through asymptotic approximation and adjustment for discrete-time observation.

1. Introduction. We consider observation of a Brownian motion $X(t)$, observed in continuous time t, with a stopping boundary. The boundary is typically chosen to provide a test of hypotheses about the drift parameter. But the test is unimportant here; we focus on estimating the drift once a boundary has been reached, and without regard to the conclusion of the sequential test.

We assume a *linear stopping boundary*, consisting of an upper boundary (a line with a positive intercept), a lower boundary (a line with a negative intercept), and possibly a vertical boundary. The boundary forms a closed region with one possible exception: parallel upper and lower boundaries without a vertical boundary are allowed. These boundaries include Wald's *sequential probability ratio tests* (SPRTs) [20], Anderson's *triangular designs* [1] (see also [17, 10, 22]), and Armitage's [2] *restricted designs*; these are the only fully-sequential designs (continuously distributed stopping times) for which the distribution of the stopped Brownian motion along the boundary is explicitly available. Formulas were derived by Anderson [1]; we use the simpler versions of [11].

This Brownian motion paradigm serves as an asymptotic approximation to many sequential analysis problems, including random sampling from a parametric model and a proportional-hazards survival analysis staggered-entry two-arm clinical trial model; see [22, 13]. Whitehead [22] includes a boundary adjustment to allow for observation in discrete-time. Recently, sequential tests with triangular boundaries have been utilized in clinical trials, e.g., [18], [19], and [3], in AIDS, cardiology and pediatrics, respectively.

Whitehead [21] considered estimation of the drift upon hitting the stopping boundary, at (t, x), say, with $x = X(t)$. The maximum likelihood estimate (MLE) is x/t. He found it to have considerable bias (noted earlier by Cox [5]; see his Figure 1, showing the bias function for a symmetric SPRT and a symmetric triangular test (2-SPRT).

He proposed a *bias-adjusted estimate* (defined in Section 2 below), designed to reduce the bias, and he provided some computational comparison of his estimate with the MLE. However, his estimate still retains some bias (see Section 4), and its computation is rather complex.

Ferebee [7] showed how to construct an unbiased estimate of the drift. Liu and Hall [16] applied Ferebee's general method to the case of linear boundaries, and went on to prove that the resulting estimate is the uniformly minimum variance unbiased estimate (UMVUE)—that is, it has uniformly (in the drift) minimum variance among all unbiased estimates based on the Brownian path from the origin to the time of hitting the boundary. However, the UMVUE is computationally somewhat complex, and hence there is need for a simple estimate that will remove much of the bias. One specific use could be by a data monitoring committee while a sequential trial progresses.

Here (Section 2) we first approximate the bias function by a *segmented curve*, starting and ending with horizontal linear segments with a slanted linear segment in between (see Figure 1). We then apply Whitehead's bias-adjustment method, with this approximate bias function replacing the true bias function, yielding the *segmented estimate.* This not only results in a simple estimate compared with Whitehead's; computations (Section 4) show that it may do a much better job of reducing the bias! More accurately stated, some complex computation may be required to find an optimal choice among segmented estimates for any given design, but application of the formula, once obtained, is simple and explicit. Moreover, we provide an optimal (*minimax absolute bias*) choice for many popular linear designs (Section 3)—including those provided by PEST software of Brunier and Whitehead [4]. A variety of alternative optimality criteria are suggested; however, a sub-optimal choice may still be useful.

In Section 3, we consider in detail two special cases: symmetric SPRTs and symmetric triangular designs (2-SPRTs). These are the designs featured in PEST software. Computations for these cases (Section 4) show that the precision of the new estimate, as measured by the root-mean-square (RMS) error, is comparable to that of Whitehead's and to that of the UMVUE; all three estimates are much more precise than the MLE.

One of these triangular designs was used in the MADIT clinical trial [19]; data from that trial are used for illustration in Section 5.

We recognize that, in practice, observation is not done in continuous time. Simulations (not detailed here) confirm, however, that the segmented estimate (5) continues to have substantially reduced bias when using Whitehead's 'Christmas tree' [22] boundary adjustment for discrete-time observation.

For the popular group-sequential designs—with relatively few well-spaced possible stopping times in contrast to a discrete-time modification of a continuous-boundary design—a segmented approximation may not be so appropriate or efficient; see [9] for a simple low-bias method for this case.

2. Derivation of the segmented estimate. We consider sequential tests based on observation of a Brownian motion $X(t)$ with drift θ, determined by stopping boundaries

$$(1) \qquad \begin{array}{rl} a_1 + b_1 t \text{ for } t < t_0 & \text{upper} \\ a_2 + b_2 t \text{ for } t < t_0 & \text{lower} \\ \text{for } t = t_0 & \text{vertical} \end{array}$$

with $a_2 < 0 < a_1$ and $0 < t_0 \leq \infty$; however, $t_0 \leq (a_1 - a_2)/(b_2 - b_1)$ if $b_2 > b_1$, and $t_0 < \infty$ if $b_2 < b_1$. These assure that the stopping time T is finite a.s.

The MLE of θ, upon termination of the Brownian motion, is

$$\hat{\theta}_{ML} = \frac{X(T)}{T} = \begin{cases} a_1/T + b_1 & \text{if the upper boundary is reached,} \\ a_2/T + b_2 & \text{if the lower boundary is reached,} \\ X(t_0)/t_0 & \text{if the vertical boundary is reached.} \end{cases}$$

This estimate is biased and often substantially so. There is a tendency for it to over-estimate the drift when reaching the upper boundary and under-estimate when reaching the lower boundary; bias tends to be small when reaching a vertical boundary or when stopping near the apex of a triangular boundary. Whitehead [21] and Liu [15] investigated the behavior of its bias function $b(\theta)$, which may be computed using the density formulas in Hall [11]; see Appendix A. Except in the case of the untruncated SPRT, the formulas involve infinite series with integrals in each term. For a symmetric horizontal SPRT, the formula is quite simple; see Appendix A ([15]).

Figure 1 (similar to Figure 1 in Whitehead [21]) reveals the general pattern of the bias function: approximately linear for θ in an intermediate range (between the hypothesized values) and approaching asymptotes quickly in either direction. The asymptotes are known to be the reciprocals of the upper and lower boundary intercepts, respectively ([5]; see also [15]). Such bias curves can be well-approximated by a segmented curve, as illustrated in Figure 1, using the asymptotes for extreme θ-values and connecting linearly over a suitable intermediate range:

$$(2) \qquad b_s(\theta) = \begin{cases} 1/a_2 & \text{for } \theta \leq \theta_2, \\ c\theta - d & \text{for } \theta_2 < \theta < \theta_1, \\ 1/a_1 & \text{for } \theta \geq \theta_1 \end{cases}$$

for some θ_2 and θ_1 and with

$$(3) \qquad c = \frac{a_1 + |a_2|}{(\theta_1 - \theta_2)a_1|a_2|} \qquad \text{and} \qquad d = \frac{a_1\theta_1 + |a_2|\theta_2}{(\theta_1 - \theta_2)a_1|a_2|}$$

to maintain continuity at θ_2 and θ_1. We call $b_s(\theta)$ a *segmented approximation* to the bias function $b(\theta)$ of the MLE. It is determined by θ_2 and θ_1 or by c and d.

When the boundaries are symmetric around the t-axis (possibly after shifting the drift parameter), $b(0) = 0$. Setting $b_s(0) = 0$ likewise, $d = 0$ and $\theta_2 = -\theta_1$; b_s is then determined by a single parameter θ_1 or $c = 1/(\theta_1 a_1)$. For an example, see Figure 1.

Whitehead [21] defined an adjusted estimate $\hat{\theta}_a$ as the solution (in θ) to the equation

$$(4) \qquad b(\theta) + \theta = \hat{\theta}_{ML}.$$

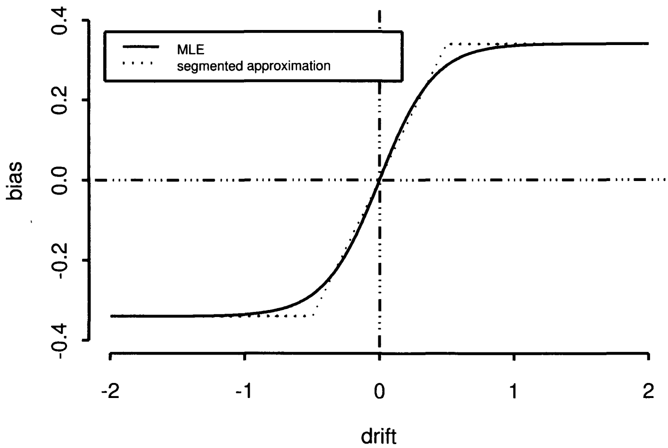

Figure 1. Bias of MLE and a segmented approximation thereto, for symmetric SPRT of drift $= \mp\frac{1}{2}$ with $\alpha = \beta = 5\%$.

If the bias function were linear—which it clearly is not—this would eliminate the bias. We now define a *segmented estimate* $\hat{\theta}_s$ similarly, simply replacing b in (4) by the approximation b_s in (2). While Whitehead needed to solve (4) for $\hat{\theta}_a$ iteratively, we find a simple explicit solution:

$$\hat{\theta}_s = \begin{cases} \hat{\theta}_{ML} - \frac{1}{a_2} & \text{for } \hat{\theta}_{ML} \le \theta_2 + \frac{1}{a_2} \\ r \cdot (\hat{\theta}_{ML} + d) & \text{for } \theta_2 + \frac{1}{a_2} < \hat{\theta}_{ML} < \theta_1 + \frac{1}{a_1} \\ \hat{\theta}_{ML} - \frac{1}{a_1} & \text{for } \hat{\theta}_{ML} \ge \theta_1 + \frac{1}{a_1} \end{cases} \tag{5}$$

with $r = 1/(1+c)$. For symmetric boundaries, note that $\hat{\theta}_s$ shrinks $\hat{\theta}_{ML}$ towards the origin, and otherwise towards $-d$.

There is yet need to choose θ_1 and θ_2, or equivalently choose d and c or r. In the symmetric case, there is but one constant to choose, namely θ_1 or c or r. We define an optimal choice as one for which the maximum (over θ) of the absolute bias of $\hat{\theta}_s$ is minimal—a *minimax bias* criterion. This requires computation of the bias of $\hat{\theta}_s$, as a function of the free parameter(s); this can be evaluated numerically using density formulas in Appendix A, and the minimax task carried out numerically. We provide the solution for some important cases in Section 3.

Other optimization criteria could be applied, informally or precisely: (i) Sketch the bias function $b(\theta)$ and choose a good-fitting $b_s(\theta)$ empirically; (ii) solve $b_s(\theta^*) = b(\theta^*)$ and $b'_s(\theta^*) = b'(\theta^*)$ for some intermediate θ^*-value; (iii) minimize (formally) some distance measure between the functions $b(\theta)$ and $b_s(\theta)$, or (iv) between the UMVUE and $\hat{\theta}_s$, or (v) between the bias function of $\hat{\theta}_s$ (see Appendix A) and 0, as in the minimax criterion.

3. Symmetric SPRT and 2-SPRT designs. First consider horizontal parallel boundaries with intercepts $\pm a$, and untruncated. For testing $\theta = \mp\frac{1}{2}$ with equal error probabilities α and $a = \log[(1-\alpha)/\alpha]$, these are SPRT boundaries. Since $\hat{\theta}_{ML} = \pm a/T$, segmented estimates have the form

$$\pm\left(\frac{a}{T} - \frac{1}{a}\right) \quad \text{for } T \le t_s \qquad \text{and} \qquad \pm\, r\frac{a}{T} \quad \text{for } T \ge t_s \tag{6}$$

with $t_s = a^2(1-r)$ $(0 < r < 1)$. Thus t_s (or r) defines a segmented estimate $\hat{\theta}_s$.

Next consider triangular boundaries with intercepts $\pm a$ and slopes $\mp\frac{1}{4}$, and untruncated. These are 2-SPRT boundaries for testing $\theta = \mp\frac{1}{2}$ with equal error probabilities α and $a = -2\log(2\alpha)$ ([17, 10, 22]). Now $\hat{\theta}_{ML} = \pm(a/T - \frac{1}{4})$, and segmented estimates have the form

$$\pm\left(\frac{a}{T} - \frac{1}{4} - \frac{1}{a}\right) \quad \text{for } T \le t_s \qquad \text{and} \qquad \pm\, r\left(\frac{a}{T} - \frac{1}{4}\right) \quad \text{for } T \ge t_s \tag{7}$$

with $t_s = a^2(1-r)/[1 + a(1-r)/4]$ $(0 < r < 1)$. Again, t_s (or r) defines a segmented estimate $\hat{\theta}_s$.

We consider these as *canonical versions* of equal-error-probability tests of two simple hypotheses about θ. To test $\theta = \theta_1$ versus θ_2, first transform from $X(t)$ to $X'(t') = \Delta[X(t) - \bar{\theta}t]$ with $\Delta = \theta_2 - \theta_1$, $\bar{\theta} = (\theta_2 + \theta_1)/2$ and $t = t'/\Delta^2$; $X'(t')$ has drift $\theta' = (\theta - \bar{\theta})/\Delta$, $= \mp\frac{1}{2}$ at θ_1 and θ_2; corresponding boundary constants are $a_i' = \Delta a_i$ and $b_i' = (b_i - \bar{\theta})/\Delta$. Moreover, such symmetric designs can be adapted to unequal error probability designs (α, β) for testing θ_1 versus θ_2 by choosing θ_1' and θ_2' so that an equal-error-probability (α', α') design for testing θ_1' versus θ_2' has the desired error probabilities α and β at θ_1 and θ_2, respectively; this requires calculation and inversion of the OC function. Whitehead [22] and his PEST software use this technique with $\theta_1' = \theta_1$ and $\alpha' = \alpha$. Thus, the parallel-line and triangular designs of PEST are covered by these two general cases.

In each case, it remains to specify t_s (or r) in (6) or (7). To compute r for a given a, we may first plot $b(\theta)$ and make two initial guesses at suitable values for r. Then determine (numerically) the maximum of the absolute bias of (6) or (7) for each initial guess and iterate until a minimax r, and hence t_s, is determined. We use the sub-density of T on a particular boundary to find the bias of $\hat{\theta}_s$ and to minimax it; see Appendix A. We omit the details.

We carried out this algorithm for symmetric SPRTs and 2-SPRTs, for several commonly used α-values, finding minimax t_s values and the corresponding maximal absolute bias m. We call these *optimal segmented estimates.* Results are in Table 1.

To deal with other α-values, the following approximate formulas were developed empirically from Table 1:

$$t_s = 5.7a - 9.1 \qquad \text{and} \qquad t_s = 3.1a - 4.9$$

for symmetric SPRTs and symmetric 2-SPRTs, respectively. Use of these empirical formulas will yield nearly optimal segmented estimates for reducing bias while avoiding complex computations.

For discrete-time observation (with small increments), we revert to (5)—that is, actual MLEs should be used on the right in (5), rather than values exactly on stopping boundaries as in (6) and (7). An adjustment for overrunning, due to lagged data after reaching a boundary, appears in Section 5.

Table 1: Optimal segmented estimates (t_s) and their maximal absolute bias (m) in symmetric SPRT and 2-SPRT designs. Boundaries for SPRT are $\pm a$; for 2-SPRT, $\pm(a - \frac{1}{4}t)$. Tests have error probabilities α at drift $= \mp\frac{1}{2}^{\star}$.

	Symmetric SPRT			Symmetric 2-SPRT		
α	a	t_s	m	a	t_s	m
0.010	4.595	17.483	0.0012	7.824	19.094	0.0002
0.025	3.664	11.14	0.0015	5.991	13.123	0.0003
0.050	2.944	7.196	0.0019	4.605	8.889	0.0004
0.100	2.197	4.007	0.0026	3.219	5.081	0.0008

* To test θ_1 versus $\theta_2 = \theta_1 + \Delta$, divide t_s by Δ^2 and multiply m by Δ; the SPRT boundaries are $\pm(a/\Delta) + \bar{\theta}t$; the 2-SPRT boundaries are $\pm[(a/\Delta) - (\Delta t/4)] + \bar{\theta}t$.

4. Comparison with other estimates. We shall see that the segmented estimate $\hat{\theta}_s$ not only has a simple explicit expression, but also reduces the bias greatly and has competitive RMS error, compared with alternative estimates.

For $\alpha = 5\%$ and a limited range of θ values, Tables 2 and 3 compare the bias and RMS error of $\hat{\theta}_s$ to those of the MLE $\hat{\theta}_{ML}$, Whitehead's [21] bias-adjusted estimate $\hat{\theta}_a$, and the UMVUE $\tilde{\theta}_u$ of Liu and Hall [16], for the case of symmetric SPRTs and 2-SPRTs. Computations were done using density formulas of Hall [11] in Appendix A and, for the bias-adjusted estimate, approximations given in Appendix B.

Table 2: Bias and RMSE of drift estimates in a symmetric SPRT for testing $\theta = \mp 1/2^\star$ with $\alpha = \beta = 5\%$. Entries, except θ's, have been mutiplied by 1000.

θ	Bias			Root-Mean-Square Error			
	$\hat{\theta}_{ML}$	$\hat{\theta}_a$ **	$\hat{\theta}_s$	$\hat{\theta}_{ML}$	$\hat{\theta}_a$ **	$\tilde{\theta}_u$	$\hat{\theta}_s$
0.0	000	000	0.0	827	452	609	609
0.1	082	055	1.3	824	460	611	611
0.2	154	103	1.9	815	481	618	617
0.3	212	136	1.6	804	515	628	627
0.4	256	148	0.7	795	553	642	641
0.5	286	141	-0.4	791	592	658	658
0.6	306	120	-1.3	792	630	676	677
0.7	319	094	-1.8	798	664	696	697
0.8	327	069	-1.9	809	695	717	718
0.9	332	049	-1.8	822	723	737	739
1.0	335	033	-1.5	837	749	758	760
1.5	339	004	-0.2	926	860	861	861

* To test θ_1 versus $\theta_2 = \theta_1 + \Delta$, multiply bias and RMSE by Δ.
** Based on approximations (10) and (11).

These numerical results show that, in these two cases, $\hat{\theta}_s$ reduces bias more efficiently than does Whitehead's bias-adjusted estimate, and has precision (as measured by the RMS error) almost equal to that of Whitehead's bias-adjusted estimate

Table 3: Bias and RMSE of drift estimates in a symmetric 2-SPRT for testing $\theta = \mp 1/2^{\star}$ with $\alpha = \beta = 5\%$. Entries, except θ's, have been mutiplied by 1000.

θ	Bias			Root-Mean-Square Error			
	$\hat{\theta}_{ML}$	$\tilde{\theta}_a$ **	$\tilde{\theta}_s$	$\hat{\theta}_{ML}$	$\tilde{\theta}_a$ **	$\tilde{\theta}_u$	$\tilde{\theta}_s$
0.0	000	00	0.0	617	402	468	468
0.1	053	28	0.2	615	405	470	470
0.2	101	53	0.3	608	417	477	477
0.3	140	71	0.3	600	434	487	487
0.4	170	79	0.1	594	458	501	501
0.5	189	75	-0.2	592	484	517	517
0.6	202	62	-0.3	596	511	534	534
0.7	209	45	-0.4	604	537	551	552
0.8	213	29	-0.4	615	561	569	570
0.9	215	17	-0.3	629	583	587	588
1.0	216	05	-0.2	645	605	605	606
1.5	217	00	-0.0	722	689	689	689

* To test θ_1 versus $\theta_2 = \theta_1 + \Delta$, multiply bias and RMSE by Δ.

** Based on approximations (10) and (11).

and the UMVUE. The MLE is uniformly the poorest among these, both in bias and in precision; however, the bias-adjusted estimate appears to have uniformly smaller RMS error, especially between the hypothesized values for θ, but its tabulated values are based on approximations. This observation of apparent superiority in RMS error agrees with that of Emerson and Fleming [6], where group sequential designs are the main focus. To choose an estimate, a trade-off between bias and variance may be needed; but if reducing the bias is the main concern, or if simplicity of computation is deemed valuable, then the segmented estimate is surely a strong candidate.

We now consider how close alternative estimates are to the UMVUE. The values of various estimates when stopping on the upper boundary for the symmetric SPRT and the symmetric 2-SPRT designs considered in Tables 2 and 3 are plotted against stopping time in Figure 2. These curves indicate that the optimal segmented estimate is much closer to the UMVUE than is Whitehead's bias-adjusted estimate, consistent with the fact that the segmented estimate has smaller bias than the bias-adjusted estimate.

5. The MADIT example. In this section, we apply various methods of estimation discussed in the previous sections to data collected in the clinical trial MADIT [19]. It was a fully-sequential trial (weekly analyses for two-and-a-half years) with triangular boundaries.

MADIT (Multicenter Automatic Defibrillator Implantation Trial), a multicenter randomized clinical trial, was conducted to evaluate the effectiveness of an implanted automatic defibrillator, compared with conventional drug therapy, to reduce mortality associated with ventricular arrhythmias. Monitoring was based on the

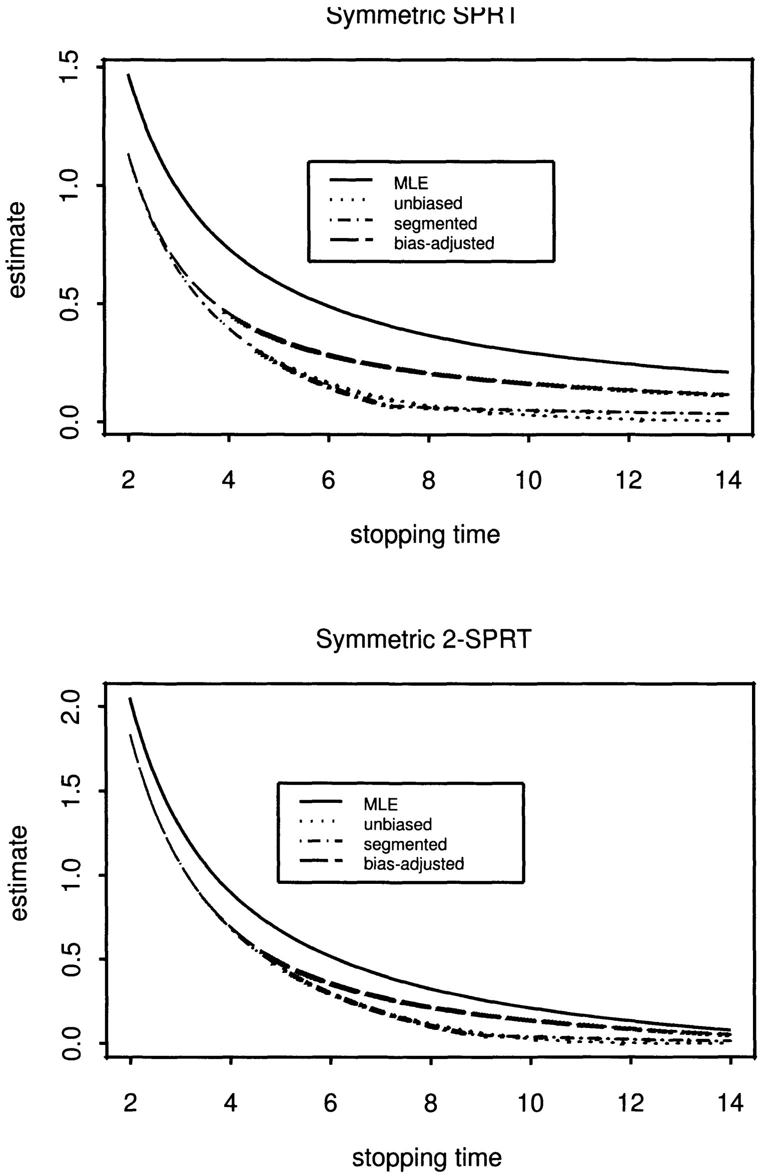

Figure 2. Various estimates in the symmetric SPRT and 2-SPRT designs for testing $\theta = \mp\frac{1}{2}$ with $\alpha = \beta = 5\%$.

log-rank statistic plotted against its estimated variance (with linear interpolation between plotting points). This behaves like a Brownian motion with drift θ equal to the negative of the logarithm of the hazard ratio (HR) for the two treatment groups (see [22, 8]). The trial was designed to have a one-sided significance level at $\theta_1 = 0$ of 2.5% and power at $\theta_2 = 0.622$ (HR $= 0.537$) of 90%. A triangular test was designed for the trial, using the early part of the lower boundary along with the upper boundary for rejection of the null hypothesis, thereby yielding a two-sided significance level of 5% (as described in [22]). Using PEST software, we find that a 2-SPRT of $\theta = 0$ versus $\theta_2' = 0.755$ with $\alpha = \beta = 2.5\%$ has power 90% at θ_2. The stopping boundaries are $x = 7.935 + 0.189t$ and $x = -7.935 + 0.566t$. Correspondingly, the canonical hypotheses are $\mp\frac{1}{2}$ with canonical boundaries $x = \pm(5.991 - \frac{1}{4}t)$.

The upper boundary was reached at $(t, x) = (12.145, 10.230)$ ($= (6.923, 4.259)$ in the canonical design). After some late data came in, a final 'overrunning' position was at $(t^o, x^o) = (13.277,\ 13.167)$. This information is now used to estimate the treatment effect parameter θ.

We first ignore the overrunning. We find the MLE of θ is $x/t = 0.842$, the UMVUE is 0.7163 and Whitehead's bias-adjusted estimate from PEST is 0.743. Using Table 1 and (7), the optimal segmented estimate of drift in the canonical design is $5.991/6.923 - 1/4 - 1/5.991 = 0.448$. Converting this value to obtain the segmented estimate of θ, we find $\tilde{\theta}_s = 0.7157$, close to the UMVUE (and identical if rounding to the nearest thousandth).

Incorporating overrunning, the MLE is found to be $\text{MLE}^o = x^o/t^o = 0.992$. Note that $\text{MLE}^o = (t \cdot \text{MLE} + \Delta x)/t^o$ with Δx being the increment in $X(t)$ after hitting the boundary. We therefore define a segmented estimate by replacing MLE in this formula by the segmented estimate upon reaching the boundary. This yields $(12.145 \times 0.7157 + 2.937)/13.277 = 0.876$ as a segmented estimate incorporating overrunning. Simulations (not detailed here) support our conjecture that this segmented estimate continues to have reduced bias compared with the MLE. No UMVUE is available when there is overrunning (but see [12]). These overrunning estimates are somewhat higher than those at the boundary crossing since the sample path rose slightly more steeply in the overrunning phase; see Figure 1 in [19]

Acknowledgement. The segmented estimate concept originally appeared in the author's PhD dissertation [14], under the supervision of Jack Hall. Jack has been a wonderful academic mentor and dear friend to me ever since I enrolled into the graduate program in statistics (biostatistics) at the University of Rochester in 1993. I am grateful to Jack for his valuable help in preparing this paper, for his precious guidance in my career development and, much more, for his warm friendship that the author values most. Thank you, Jack, and happy birthday!

APPENDIX A: Bias and Precision of the MLE. Consider boundaries in (1). Let

$$\bar{a} = (a_1 + a_2)/2, \quad \bar{b} = (b_1 + b_2)/2, \quad c = a_1 - a_2, \quad b = (b_2 - b_1)/2$$

and

$$r_j = \begin{cases} jc + a_1 \text{ if } j \text{ even} \\ jc - a_2 \text{ if } j \text{ odd}, \end{cases} \qquad s_j = \begin{cases} jc + a_1 \text{ if } j \text{ odd} \\ jc - a_2 \text{ if } j \text{ even}, \end{cases}$$

and let $\phi_t(\cdot)$ be the density of the normal distribution with mean 0 and variance t. Then the density of $(T, X(T))$ upon stopping on the upper boundary at time t is

([11])

$$p_\theta^U(t) = t^{-1}\sum_{j=0}^{\infty}(-1)^j \exp\left\{\frac{b}{c}(r_j^2 - a_1^2) + a_1\tau - \frac{1}{2}\tau^2 t\right\}\cdot r_j\,\phi_t(r_j)\Bigg|_{\tau=\theta-b_1}, \quad t\in(0,t_0).$$

The density $p_\theta^L(t)$ on the lower boundary is given by $p_\theta^U(t)$ with $(a_1, a_2, \theta - b_1, r_j)$ replaced by $(-a_2, -a_1, -\theta+b_2, s_j)$. The density on the vertical boundary is $p_\theta^V(x) = \exp\left\{(\theta-\bar{b})x - \frac{1}{2}(\theta^2 - \bar{b}^2)t_0\right\} p_{\bar{b}}^V(x)$ where

$$\begin{aligned} p_{\bar{b}}^V(x) = \phi_{t_0}(z) + \sum_{j=1}^{\infty}\Big[&\exp\{4bj(jc-\bar{a})\}\,\phi_{t_0}(z-2jc) \\ &- \exp\{2b(2j-1)(jc-a_1)\}\,\phi_{t_0}(z+2jc-2a_1) \\ &- \exp\{2b(2j-1)(jc+a_2)\}\,\phi_{t_0}(z-2jc-2a_2) \\ &+ \exp\{4bj(jc+\bar{a})\}\,\phi_{t_0}(z+2jc)\Big]\Bigg|_{z=x-\bar{b}t_0}. \end{aligned}$$

The first moment (and thence the bias $b(\theta)$) of the MLE $\hat{\theta}_{ML} = X(T)/T$ can be evaluated numerically from

$$\mathbf{E}_\theta\left[\hat{\theta}_{ML}\right] = \int_0^{t_0}(\frac{a_1}{t} + b_1)p_\theta^U(t)dt + \int_0^{t_0}(\frac{a_2}{t} + b_2)p_\theta^L(t)dt + \frac{1}{t_0}\int_{a_2+b_2t_0}^{a_1+b_1t_0} x p_\theta^V(x)dx.$$

The second moment of $\hat{\theta}_{MLE}$, and thence the MSE, may be evaluated similarly. Moments of other estimates may likewise be evaluated numerically.

For symmetric SPRTs, the first two moments can be greatly simplified. After considerable mathematical manipulation [15], we find

$$\mathbf{E}_\theta\left[\hat{\theta}_{ML}\right] = 2a^{-1}\sinh(a\theta)\left(\gamma_1 - \frac{1}{2}\int_0^{a|\theta|} u\,\mathrm{sech}(u)du\right)$$

and

$$\mathbf{E}_\theta\left[\hat{\theta}_{ML}^2\right] = 2\cosh(a\theta)\left(\frac{\theta^2}{4}\int_0^{a|\theta|} u\,\mathrm{sech}(u)du - \frac{\gamma_1}{2}\theta^2 + 3\frac{\gamma_3}{a^2} - \frac{a^2}{4}\int_0^{a|\theta|} u^3\mathrm{sech}(u)du\right)$$

with $\gamma_1 = \sum_{n=0}^{\infty}(-1)^n(2n+1)^{-2} \approx 0.915966$ and $\gamma_3 = \sum_{n=0}^{\infty}(-1)^n(2n+1)^{-4} \approx 0.988945$.

These formulas are used repeatedly for computations of bias and MSE in Sections 3 and 4.

APPENDIX B: Bias and Precision of Whitehead's Estimate. Let $e(\theta) = \mathbf{E}_\theta\left[\hat{\theta}_{ML}\right] = b(\theta) + \theta$ be the expectation of $\hat{\theta}_{ML}$ at θ. The bias-adjusted estimate $\hat{\theta}_a$, defined as the solution to (4), satisfies

$$e(\hat{\theta}_a) = b(\hat{\theta}_a) + \hat{\theta}_a = \hat{\theta}_{ML}.$$

Suppose the mean function $e(\theta)$ of $\hat{\theta}_{ML}$ is nondecreasing and can be differentiated infinitely often (which is satisfied for a broad class of stopping boundaries, including linear boundaries (1) and any group sequential boundaries). Writing $e_{-1}(\cdot)$ for the inverse function of $e(\cdot)$, we have $\hat{\theta}_a = e_{-1}(\hat{\theta}_{ML})$. Expanding $e_{-1}(\hat{\theta}_{ML})$ at $e(\theta)$, we have

$$\hat{\theta}_a - \theta = \sum_{j=1}^{\infty} \frac{e_{-1}^{(j)}(e(\theta))}{j!} (\hat{\theta}_{ML} - e(\theta))^j.$$

Taking expectations of both sides, and assuming orders of integration and summation can be interchanged, we obtain an expression for the bias of $\hat{\theta}_a$:

$$\text{(8)} \qquad \text{Bias}\left[\hat{\theta}_a\right] = \mathbf{E}_\theta\left[\hat{\theta}_a\right] - \theta = \sum_{j=2}^{\infty} \frac{e_{-1}^{(j)}(e(\theta))}{j!} \mathbf{E}_\theta\left[(\hat{\theta}_{ML} - e(\theta))^j\right].$$

Similarly, we have

$$\text{(9)} \qquad \mathbf{E}_\theta\left[\hat{\theta}_a^2\right] = \sum_{j=0}^{\infty} c_j(\theta)\, \mathbf{E}_\theta\left[(\hat{\theta}_{ML} - e(\theta))^j\right]$$

where

$$c_j(\theta) = \sum_{i=0}^{j} \frac{e_{-1}^{(i)}(e(\theta))}{i!} \cdot \frac{e_{-1}^{(j-i)}(e(\theta))}{(j-i)!}.$$

Moments of $\hat{\theta}_{ML}$ can be derived as in Appendix A. Derivatives of $e_{-1}(\theta)$ may by expressed in terms of derivatives of $e(\theta)$ by repeatedly using the identity $e_{-1}(e(\theta)) = \theta$.

The first several leading terms in (8) and (9) are often good enough for approximation when numerically evaluating bias and mean squared error of $\hat{\theta}_a$. Taking terms with $j \leq 2$ and noting that

$$e'_{-1}(e(\theta)) = \frac{1}{e'(\theta)} \qquad \text{and} \qquad e''_{-1}(e(\theta)) = -\frac{e''(\theta)}{(e'(\theta))^3},$$

we find

$$\text{(10)} \qquad \text{Bias}\left[\hat{\theta}_a\right] \approx -\frac{e''(\theta)}{2(e'(\theta))^3} \text{Var}_\theta\left[\hat{\theta}_{ML}\right]$$

and

$$\mathbf{E}_\theta\left[\hat{\theta}_a^2\right] - \theta^2 \approx \frac{1}{(e'(\theta))^3}(e'(\theta) - \theta e''(\theta))\, \text{Var}_\theta\left[\hat{\theta}_{ML}\right].$$

From these we find

$$\text{(11)} \qquad \text{MSE}\left[\hat{\theta}_a\right] = \mathbf{E}\left[\hat{\theta}_a^2\right] - \theta^2 - 2\theta \cdot \text{Bias}\left[\hat{\theta}_a\right] \approx \frac{1}{(e'(\theta))^2} \text{Var}_\theta\left[\hat{\theta}_{ML}\right].$$

REFERENCES

[1] T.W. Anderson. A modification of the sequential probability ratio test to reduce the sample size. *Annals of Mathematical Statistics*, 31:165–197, 1960.

[2] P. Armitage. *Sequential Medical Trials.* Blackwell, Oxford, UK, 2nd edition, 1975.

[3] E. Bellissant, J.F. Duhamel, M. Guillot, A. Pariente-Khayat, G. Olive, and G. Pons. The triangular test to assess the efficacy of metoclopramide in gastroesophageal reflux. *Clinical Pharmacology and Therapeutics*, 61:377–384, 1997.

[4] H. Brunier and J. Whitehead. *PEST: Planning and Evaluation of Sequential Trials Operating Manual.* Medical and Pharmaceutical Statistics Research Unit, University of Reading, Reading, UK, 3.0 edition, 1993.

[5] D.R. Cox. A note on the sequential estimation of means. *Proceedings of the Cambridge Philosophical Society*, 48:447–450, 1952.

[6] S.S. Emerson and T.R. Fleming. Parameter estimation following sequential hypothesis testing. *Biometrika*, 77:875–892, 1990.

[7] B. Ferebee. An unbiased estimator for the drift of a stopped Wiener process. *Journal of Applied Probability*, 20:94–102, 1983.

[8] M. Gu and T.L. Lai. Weak convergence of time-sequential censored rank statistics with application to sequential clinical trials. *Annals of Statistics*, 19:1403–1433, 1991.

[9] W. J. Hall, A. Liu, and K. Ding. Limited-memory approximations for group-sequential tests and estimates. Technical report, Department of Biostatistics, University of Rochester, Rochester, NY, 2001.

[10] W.J. Hall. Sequential minimum probability ratio tests. In I.M. Chakravarti, editor, *Asymptotic theory of statistical tests and estimation : in honor of Wassily Hoeffding*, pages 315–350. Academic Press, New York, 1980.

[11] W.J. Hall. The distribution of Brownian motion on linear stopping boundaries. *Sequential Analysis*, 16:345–352, 1997. Addendum: **17**:123-124.

[12] W.J. Hall and A. Liu. Sequential tests and estimates after overrunning. Technical report, Department of Biostatistics, University of Rochester, Rochester, NY, 2001.

[13] K.K.G. Lan and D.M. Zucker. Sequential monitoring of clinical trials: The role of information and Brownian motion. *Statistics in Medicine*, 12:753–765, 1993.

[14] A. Liu. *Estimation following sequential tests.* PhD thesis, University of Rochester Medical Center, Department of Biostatistics, 1997.

[15] A. Liu. On the maximum likelihood estimate for the drift of Brownian motion following a symmetric sequential probability ratio test. *Communications in Statistics – Theory and Methods*, 26:977–989, 1997.

[16] A. Liu and W.J. Hall. Minimum variance unbiased estimation of the drift of Brownian motion with linear stopping boundaries. *Sequential Analysis*, 17:91–107, 1998.

[17] G. Lorden. 2-sprt's and the modified Kiefer-Weiss problem of minimizing an expected sample size. *Annals of Statistics*, 4:281–291, 1976.

[18] J.S. Montaner, L.M. Lawson, N. Levitt, A. Belzberg, M.T. Schechter, and J. Ruedy. Corticosteroids prevent early deterioration in patients with moderately severe pneumocystis carinii pneumonia and the acquired immunodeficency syndrome (aids). *Annals of Internal Medicine*, 113:14–20, 1990.

[19] A.J. Moss, W.J Hall, D.S. Cannom, J.P. Daubert, S.L. Higgins, H. Klein, J.H. Levine, S. Saksena, A.L. Waldo, D. Wilber, M.W. Brown, and M. Hoe. Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. *New England Journal of Medicine*, 335:1933–1940, 1996.

[20] A. Wald. *Sequential Analysis.* Wiley, New York, 1947.

[21] J. Whitehead. On the bias of maximum likelihood estimation following a sequential test. *Biometrika*, 73:573–581, 1986.

[22] J. Whitehead. *The Design and Analysis of Sequential Clinical Trials.* Wiley, New York, 2nd revised edition, 1997.

BIOMETRY AND MATHEMATICAL STATISTICS BRANCH
NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT
NATIONAL INSTITUTES OF HEALTH
BETHESDA, MD 20892-7510
Liua@mail.nih.gov