The Annals of Statistics
1996, Vol. 24, No. 4, 1765–1791

AN O'BRIEN–FLEMING SEQUENTIAL TRIAL FOR COMPARING THREE TREATMENTS

BY REBECCA A. BETENSKY

Harvard School of Public Health

We consider a sequential procedure for comparing three treatments with the goal of ultimately selecting the best treatment. This procedure starts with a sequential test to detect an overall treatment difference and eliminates the apparently inferior treatment if this test rejects the equality of the treatments. It then proceeds with a sequential test of the remaining two treatments. We base these sequential tests on the stopping boundaries popularized by O'Brien and Fleming. Our procedure is similar in structure to that used by Siegmund in conjunction with modified repeated significance tests. We compare the performances of the two procedures via a simulation experiment. We derive analytic approximations for an error probability, the power and the expected sample size of our procedure, which we compare to simulated values. Furthermore, we propose a modification of the procedure for the comparison of a standard treatment with experimental treatments.

1. Introduction. Large phase III clinical trials frequently compare three or more treatment options. These trials are extremely costly, and thus a design that identifies the best treatment with the fewest subjects and with high accuracy is desirable. Furthermore, if there are two equivalent treatments that are better than the third, this design should eliminate the third, but neither of the better two. And if all three treatments are equivalent, it should not eliminate any one. In this paper we investigate a sequential procedure with these features.

For simplicity, we assume that the treatments produce instantaneous normally distributed responses with a common known variance. Our goal is to select the treatment with the largest mean response. Initially, we assume that all three treatments are on equal footing and should be handled symmetrically. The procedure starts with a sequential test of H_0: $\mu_1 = \mu_2 = \mu_3$, where μ_i, $i = 1, 2, 3$, are the treatment means. If the stopping boundary of this test is crossed, there is evidence that the treatments are not equal and so the apparently inferior treatment is eliminated. Once this occurs, a second sequential test of $H_0^{i,j}$: $\mu_i = \mu_j$ begins, where i and j are the two remaining treatments. If the stopping boundary for this test is crossed, a best treatment is identified. These tests are *closed*, meaning they are truncated at a prespecified maximum sample size. Siegmund (1993) used this same structure for his procedure.

There has been much work on experimental designs for ranking treatments and selecting the best of several treatments [see Bechhofer, Santner and Goldsman (1995)]. Proposals in the literature can be classified according to

Received March 1994; revised January 1996.
AMS 1991 *subject classifications*. Primary 62L10; secondary 62N10.
Key words and phrases. Repeated significance test, ranking procedure, boundary crossing.

whether they are single staged or multistaged, whether they are closed or open and whether or not they eliminate inferior treatments during the trial. Our proposal can be classified as multistaged, closed and eliminating.

The earliest proposal for a multistaged, closed, eliminating procedure appears to be that of Paulson (1964). He took the *indifference-zone* approach, in which he ensured with high probability the correct selection of the best treatment, given that the mean of the best treatment is separated from that of the second best by some known constant. Bechhofer, Santner and Goldsman (1995) report that Paulson has recently greatly improved upon his original procedure. This new procedure compares each remaining treatment with all other treatments through their sample totals after adjusting for the fact that some may have been eliminated.

A drawback of the indifference-zone approach is that the distance between the best and second best treatments is usually not known. Another approach, initiated by Kao and Lai (1980), ensures with high probability that the mean of the selected treatment is within a fixed constant of the mean of the best treatment. Hsu and Edwards (1983) propose a procedure that improves upon Kao and Lai's (1980) procedure in that it ensures with high probability that the eliminated treatments are not the best and that the treatments remaining at the end of the trial are within certain small distances of the best. A drawback of this procedure in some applications is that it is not considered an error to eliminate a treatment equivalent to the best. Other recent similar proposals are found in Hughes (1993) and Follmann, Proschan and Geller (1994).

The motivation of our design is broader in scope than simply ensuring correct selection with high probability. Rather, we attempt to simultaneously minimize the important errors that could be made in the selection process as well as the expected sample size. However, if an investigator had some prior notion of the configuration of the true treatment means, this could be used in choosing the design parameters of our procedure to ensure correct selection with high probability. Presumably, we could assess the efficiency of our procedure relative to other procedures that are based on this probability requirement through this means. Jennison, Johnstone and Turnbull (1982) derive a lower bound for a measure of efficiency for procedures based on the indifference-zone probability statement, as well as the probability statement used by Kao and Lai (1980).

Additionally, our design differs from those described above in using a global test for treatment differences when more than two treatments are under consideration, rather than pairwise tests. In fact, our procedure can be viewed as a sequential version of Fisher's least significant difference procedure for multiple comparisons. That is, first a global test is carried out, and only if the null hypothesis is rejected is a pairwise test done. This is in contrast to the procedure of Follmann, Proschan and Geller (1994), for example, which reduces to Tukey's procedure for multiple comparisons in the fixed sample case.

We base our procedure on the widely used, so-called "O'Brien–Fleming" stopping boundaries. Siegmund (1985), page 260, points out that these bound-

aries were proposed by other authors prior to O'Brien and Fleming (1979). Tests based on them can be viewed as truncated versions of the sequential probability ratio tests proposed by Wald (1947). They were proposed also by Rao (1950) and Miller (1970) as sequential versions of locally most powerful tests. Samuel-Cahn (1974) used them as well. O'Brien and Fleming (1979) may have contributed to the popularization of these boundaries by extending them to grouped data. Siegmund (1993) bases his procedure on modified repeated significance tests, which are the main competitors of O'Brien–Fleming tests.

To define these tests, let $Y_{i,j}$ denote the immediate response of patient j to treatment i, $i = 1, \ldots, r$, $j = 1, 2, \ldots$. Assume that the $Y_{i,j}$ are independent and normally distributed with means μ_i and variance 1. Then the log-likelihood ratio statistic for testing the null hypothesis of equality of all treatments, H_0: $\mu_1 = \cdots = \mu_r$, is given by $T_n^2/2n$, where $T_n^2 = \sum_{i=1}^{r}(Y_{i\cdot} - Y_{\cdot\cdot}/r)^2$, $Y_{i\cdot} = \sum_{j=1}^{n} Y_{i,j}$ and $Y_{\cdot\cdot} = \sum_{i=1}^{r}\sum_{j=1}^{n} Y_{i,j}$. The O'Brien–Fleming test is based on a constant boundary for T_n^2, the "unnormalized" statistic, and the repeated significance test is based on a constant boundary for T_n^2/n. A clear advantage of the O'Brien–Fleming procedure is that it is based on fewer design parameters than the modified repeated significance procedure.

We begin in Section 2 by describing our procedure and defining some notation. Using Monte Carlo simulations, we compare our procedure to the three-treatment procedure based on repeated significance tests in Section 3. In Section 4 we develop analytic approximations for an error probability and the power of the test, and in Section 5 we derive an approximation for the expected sample size associated with the three-treatment procedure. In these two sections, we assess the accuracy of our analytic approximations through comparisons with values from simulations. In Section 6 we discuss how the procedure might be modified for the comparison of treatments that are not on equal footing, and in Section 7 we conclude.

2. Definition of the procedure. Let $Y_{i,j}$ denote the immediate response of patient j to treatment i, $i = 1, 2, 3$, $j = 1, 2, \ldots$. We assume that the $Y_{i,j}$ are independent and normally distributed with means μ_i and variance 1, and that a large mean response corresponds to a good treatment. In the first stage of sampling, the observations consist of the triples $Y_j = (Y_{1,j}, Y_{2,j}, Y_{3,j})'$. For the mathematical formulation of the procedure, it is useful to change coordinates via an orthogonal transformation, $X_j = CY_j$, where

$$C = \begin{pmatrix} 6^{-1/2} & 6^{-1/2} & -2 \cdot 6^{-1/2} \\ 2^{-1/2} & -2^{-1/2} & 0 \\ 3^{-1/2} & 3^{-1/2} & 3^{-1/2} \end{pmatrix}.$$

Likewise, let $\theta_1 = 6^{-1/2}(\mu_1 + \mu_2 - 2\mu_3)$ and $\theta_2 = 2^{-1/2}(\mu_1 - \mu_2)$ be the two orthogonal contrasts. The null hypothesis of no treatment effect in the original coordinates, that is, $\mu_1 = \mu_2 = \mu_3$, translates into $\theta_1 = \theta_2 = 0$ in the new coordinates. Let $S_{i,n} = \sum_{j=1}^{n} X_{i,j}$ and $S_n = (S_{1,n}, S_{2,n})'$. In order to

make pairwise comparisons between the treatments, it is useful to define the quantities $X_j^{i_1,i_2} = 2^{-1/2}(Y_{i_1,j} - Y_{i_2,j})$ and $S_n^{i_1,i_2} = \sum_{j=1}^n X_j^{i_1,i_2}$. In terms of $X_{1,j}$ and $X_{2,j}$, $X_j^{1,2} = X_{2,j}$, $X_j^{1,3} = (3^{-1/2}X_{1,j} + X_{2,j})/2$ and $X_j^{2,3} = (3^{-1/2}X_{1,j} - X_{2,j})/2$. It is useful for future calculations to note that there is a one-to-one correspondence between events of the type $S_n^{i_1,i_2} = \max_{k,l}|S_n^{k,l}|$ and the particular sector of the circle in which ω_n, the angle between S_n and the positive x_1-axis, is found. The six possible sectors are determined by the x_1-axis and the lines rotated $\pi/3$ and $-\pi/3$ from the positive x_1-axis.

The log-likelihood ratio statistic for testing the hypothesis of no treatment difference is $\|S_n\|^2/2n$ and the O'Brien–Fleming sequential test stops as soon as the "unnormalized" version of this statistic, $\|S_n\|^2$, crosses a certain constant level. Namely, the first stage of our procedure samples $X_1, X_2, \ldots$ until $T_1 \wedge m = \min(T_1, m)$, where

$$T_1 = \inf\{n\colon \|S_n\| > b_1\}. \tag{1}$$

There are three possible outcomes to consider:

1. If $T_1 < m$ and $S_{T_1}^{i_1,i_3} = \max_{k,l}|S_{T_1}^{k,l}|$, treatment i_3 is eliminated and the second stage of this procedure samples $X_{T_1+1}^{i_1,i_2}, X_{T_1+2}^{i_1,i_2}, \ldots$ until $T_2 \wedge m$, where

$$T_2 = \inf\{n\colon n \geq T_1,\ \|S_n^{i_1,i_2}\| > b_2\}.$$

   If $T_2 \leq m$ and $S_{T_2}^{i_1,i_2} > b_2$, treatment i_1 is selected as the best. If $T_2 \leq m$ and $S_{T_2}^{i_1,i_2} < -b_2$, treatment i_2 is selected as the best. If $T_2 > m$, then neither i_1 nor i_2 is selected.
2. If $T_1 \geq m$, $\|S_m\| > b_1$ and $S_m^{i_1,i_3} = \max_{k,l}|S_m^{k,l}|$, treatment i_3 is eliminated. If $S_m^{i_1,i_2} > b_2$, treatment i_1 is selected as best, whereas if $S_m^{i_1,i_2} < -b_2$, treatment i_2 is selected. If $-b_2 \leq S_m^{i_1,i_2} \leq b_2$, then neither treatment is selected.
3. If $T_1 \geq m$ and $\|S_m\| \leq b_1$, then the null hypothesis of no treatment effect is not rejected.

The parameters b_1 and b_2 are chosen to meet several requirements. First, b_1 is chosen so that the significance level for the test of H_0: $\theta_1 = \theta_2 = 0$, given by $P_0\{T_1 \leq m\}$, equals the desired α. Likewise, b_2 is chosen to protect against the error of eliminating one of two equivalent treatments that are far superior to the third. In this case, the first stage will terminate early and the error of eliminating either of the remaining treatments can be approximated by the significance level of a two-treatment O'Brien–Fleming procedure with parameters m and b_2, which is set to α as well.

Finally, b_1 and b_2 are chosen to satisfy $b_2 < b_1\, 3^{1/2}/2$, which ensures that with high probability, two treatments that are nearly equivalent can be eliminated at the same time. To see this, suppose that treatments 2 and 3 are equivalent and inferior to treatment 1. Suppose that at time T_1, the angle

between $\|S_{T_1}\|$ and the positive x_1-axis, ω_{T_1}, is slightly less than $\pi/3$. Then treatment 3 is eliminated and treatments 1 and 2 remain under consideration. Since $\|S_{T_1}\| > b_1$ and ω_{T_1} is close to $\pi/3$, the x_2-component of S_{T_1}, $S_{T_1}^{1,2}$, is likely to exceed $b_1\ 3^{1/2}/2$ and thus b_2 since $b_2 < b_1\ 3^{1/2}/2$. This leads to the elimination of treatment 2 at T_1 as well.

Siegmund's (1993) modified repeated significance version of this procedure is based on the stopping rules

$$T_1 = \inf\{n\colon n \geq m_0,\ \|S_n\| > b_1' n^{1/2}\}$$

and

$$T_2 = \inf\{n\colon n \geq T_1,\ \|S_n^{i_1, i_2}\| > b_2' n^{1/2}\},$$

where m_0 serves to prevent unreasonably early stopping. Additional parameters c_1 and c_2 replace b_1' and b_2' at time m in order to increase the power of the procedure. The O'Brien–Fleming procedure has three fewer design parameters than the modified repeated significance procedure, which is a clear practical advantage.

3. A comparison of procedures. Table 1 contains the results of a Monte Carlo experiment designed to compare the performance of the three-treatment procedure based on both modified repeated significance tests and O'Brien–Fleming sequential tests. The entries of the table are based on 9999 repetitions when $\theta_2 = 0$ and 2500 repetitions when $\theta_2 > 0$. The columns labeled "rs" contain the results from Siegmund's (1993) repeated significance test simulation found in his Table 1 (the column headed "p_1," which we define below, is not from his table because he does not simulate this value). They are included for comparison with the columns labeled "OF" which contain the results of the simulation based on O'Brien–Fleming tests. The parameters for the repeated significance tests are $m_0 = 10$, $m = 50$, $b_1 = 3.5$, $c_1 = 2.5$, $b_2 = 2.92$, and $c_2 = 2.05$. The parameters for the O'Brien–Fleming tests are $m = 50$, $b_1 = 18.52$ and $b_2 = 15.31$. They are chosen analogously to those used by Siegmund (1993) so that the probabilities of the errors of declaring a treatment difference when there is none and of eliminating treatment 1 or 2 when they are equivalent (assuming that treatment 3 is eliminated immediately) are equal to 0.05. The values of b_1 and b_2 used for the O'Brien–Fleming test are taken from Table 2(b) of Jennison and Turnbull (1991). The values of θ_1 and θ_2 used in the simulations are taken from Siegmund's (1993) table and represent various relationships between the three treatments, always subject to the ordering $\mu_1 \geq \mu_2 \geq \mu_3$. Each group of four (θ_1, θ_2) pairs in the table headed by the $(\theta_1, 0)$ pair has the same value of $\|\theta\|$, up to rounding errors.

The notation introduced by Siegmund (1993) for the important operating characteristics of this procedure is the following: $p_1 = P_{\|\theta\|}\{T_1 \leq m\}$ is the probability of rejecting the initial null hypothesis of no treatment effect; p_2 is the probability of mistakenly eliminating treatment 1 or 2 when they are equivalent, that is, when $\theta_2 = 0$, and of correctly selecting treatment 1 as best when it is, that is, when $\theta_2 > 0$; $E_1 = E_{\|\theta\|}\{T_1 \wedge m\}$ is the expected number

TABLE 1
Operating characteristics of the three-treatment procedure

		p_1		p_2		E_1		E_2		Total	
θ_1	θ_2	rs	OF	rs	OF	rs	OF	rs	OF	rs	OF
0.00	0.00	0.051	0.047			49.7	49.6	49.9	49.9	149.5	149.5
0.70	0.00	0.996	0.995	0.050	0.050	23.3	26.6	49.2	49.5	121.7	125.6
0.35	0.61	0.998	0.996	0.982	0.978	23.3	26.5	28.7	29.9	80.7	86.3
0.50	0.50	0.996	0.995	0.942	0.933	22.9	26.2	32.2	32.2	87.3	90.6
0.61	0.35	0.996	0.994	0.670	0.674	22.8	26.4	41.2	40.1	105.2	106.6
0.60	0.00	0.972	0.972	0.048	0.047	29.4	30.5	49.4	49.5	128.2	129.5
0.30	0.52	0.975	0.970	0.906	0.912	28.9	30.5	35.2	34.5	99.3	99.5
0.42	0.42	0.967	0.972	0.822	0.813	29.4	30.8	38.2	37.1	105.8	105.0
0.52	0.30	0.971	0.973	0.552	0.537	29.5	30.4	44.0	42.7	117.5	115.8
0.50	0.00	0.890	0.888	0.047	0.049	35.8	35.3	49.6	49.5	135.0	134.3
0.25	0.43	0.878	0.872	0.756	0.734	36.3	35.7	41.8	39.9	119.9	115.5
0.35	0.35	0.877	0.880	0.629	0.637	35.9	35.6	43.1	41.5	122.1	118.6
0.43	0.25	0.882	0.872	0.400	0.393	35.8	35.9	46.2	45.5	128.2	126.9
0.40	0.00	0.708	0.697	0.046	0.048	41.6	40.6	49.6	49.5	140.8	139.6
0.20	0.35	0.711	0.714	0.503	0.507	42.0	40.3	45.8	44.1	133.6	128.5
0.28	0.28	0.712	0.682	0.398	0.423	42.3	40.8	46.9	45.1	136.1	131.0
0.35	0.20	0.699	0.715	0.260	0.263	41.5	40.2	47.9	47.2	137.3	134.6
0.30	0.00	0.452	0.442	0.041	0.039	46.3	44.9	49.6	49.6	145.5	144.1
0.15	0.26	0.435	0.439	0.264	0.253	46.3	45.1	48.5	47.6	143.3	140.3
0.21	0.21	0.449	0.442	0.214	0.234	46.2	45.0	48.6	47.9	143.3	140.8
0.26	0.15	0.430	0.448	0.141	0.134	46.1	44.9	49.2	48.7	144.5	142.3

of triples observed in the first stage of the test, and $E_2 = E_{(\theta_1, \theta_2)}\{T_2 \wedge m\}$ is the total number of triples from the first stage of the test and pairs from the second stage of the test.

The different shapes of the stopping boundaries of the two tests suggest that the first stage of the three-treatment test should terminate earlier for the repeated significance test for larger values of $\|\theta\|$ and earlier for the O'Brien–Fleming test for smaller values of $\|\theta\|$. This is the case for the two-treatment versions of these tests. As indicated in Table 1 in the columns headed "E_1," this is exactly what occurs in this example. Table 1 indicates also that the two tests perform comparably with respect to error probabilities and power.

It is more difficult to anticipate the effects of the different boundaries after the second stage of sampling. Consider the case as in the example when $\mu_1 \geq \mu_2 \geq \mu_3$. It is plausible that the final expected sample sizes will be comparable for large values of $\|\theta\|$ and small to moderate values of $\theta_2 = 2^{-1/2}(\mu_1 - \mu_2)$ and for small values of $\|\theta\|$. This will happen if the large potential sample size for the second stage of sampling that arises from the early stopping of the repeated significance test for large $\|\theta\|$ is almost entirely used up by the large sample size needed to test $\theta_2 = 0$ versus small alternatives (where θ_2 is most likely to be the drift of the treatment difference that will be observed in

the second stage of the test when $\mu_1 \geq \mu_2 \geq \mu_3$). Likewise, it will happen that the larger initial sample size needed by the O'Brien–Fleming test for large $\|\theta\|$ is followed by a smaller sample size required by the test for a smaller θ_2. For small $\|\theta\|$, both tests are likely to use most of the available observations in the first stage of sampling and so the final expected sample sizes for both tests will be very close to m.

In the example in Table 1 the simulated final expected sample sizes for the two tests are comparable for all of the parameter configurations studied; for each set of θ_1 and θ_2 the values of E_2 are never separated by more than two units. For all values of $\|\theta\|$, when $\theta_2 = 0$, the final expected sample sizes for the two tests are most similar. It is also the case that when $\theta_2 > 0$ the simulated values of E_2 for the two tests are closest together for $\|\theta\| = 0.3$ as compared with other values of $\|\theta\|$. More of the predicted patterns of behavior under the different parameter configurations would likely be evident had the table included larger values of $\|\theta\|$.

It is also plausible that the repeated significance–based test will maintain its smaller sample sizes for the case of large values of $\|\theta\|$ and large values of θ_2. In fact, in this example there are only two cases in which E_2 for the repeated significance test is less than or equal to E_2 for the O'Brien–Fleming test: $\theta_1 = 0.35$, $\theta_2 = 0.61$ and $\theta_1 = 0.5$, $\theta_2 = 0.5$. As predicted, these are the two configurations with both the largest $\|\theta\|$ and the largest θ_2. Again, had the table included larger values of $\|\theta\|$, it is likely that there would have been more such cases of the repeated significance test stopping earlier than the O'Brien–Fleming test.

The other comparison of interest to make between the two procedures is of the expected total observations, $3E_1 + 2(E_2 - E_1)$, they each take. This value leads to a more informative comparison than that based on E_2 alone because it is expressed in interpretable units, that is, the number of observations, whereas E_2 is composed of some triples and some pairs of observations. Table 1 indicates that the two procedures are quite similar in this measure as well; they are never separated by more than five observations.

In conclusion, this example suggests that the error probabilities and power of the three-treatment procedure based on repeated significance tests and O'Brien–Fleming tests are comparable. The basis for a decision as to which type of stopping rule to use should then be expected sample size, which depends on the specific alternative hypothesis. If the treatment effect is likely to be large, that is, $\|\theta\| \geq 0.7$, then it appears that there will be a slight advantage to using repeated significance tests. If it is likely to be moderate, that is, $\|\theta\| = 0.5$ or $\|\theta\| = 0.4$, there appears to be a slight advantage to using the O'Brien–Fleming tests. For other values of $\|\theta\|$, the two tests appear to be so similar that it will not matter which is used. Thus, the differences in expected sample sizes that are found in the two-treatment versions of these tests emerge in the three-treatment versions as well, although for three treatments they seem to be of lesser magnitude. Although we have considered only a single simulation study, the behavior of the two-treatment versions of these tests implies that our results should apply to other parameter configurations.

4. Analytic approximations: significance level and power. In this section we derive analytic approximations for two probabilities of interest. For notational convenience, we now assume that treatment 1 is superior or equivalent to treatment 2, which is superior or equivalent to treatment 3. The first probability is the significance level attached to the second stage of the procedure, namely the error probability of eliminating treatment 1 or 2 when they are equivalent and superior to treatment 3. The second probability is the power of the procedure to correctly select treatment 1 as the best when it is the best. An approximation for the significance level attached to the first stage of sampling, $P_0\{T_1 \le m\}$, is found in Betensky (1992).

The basic ingredient for the approximations of both probabilities is an approximation for the joint probability

$$P_\theta\{T_1 = n,\ \|S_n\| \in d\xi,\ \omega_n \in d\omega\}, \tag{2}$$

where $\xi > b_1$ and $\theta = (\theta_1, \theta_2)$.

Letting $\xi = b_1 + x$, the probability in (2) can be approximated informally by

$$\begin{aligned}&\{1 - \Phi(x - b_1/n) - \nu(2b_1/n)\exp(2b_1x/n)[1 - \Phi(x + b_1/n)]\}\\ &\quad\times \exp(\theta_1\xi\cos\omega + \theta_2\xi\sin\omega - \xi^2/2n - n\|\theta\|^2/2)\xi\,d\xi\,d\omega/2\pi n,\end{aligned} \tag{3}$$

where

$$\nu(\mu) = 2\mu^{-2}\exp\left[-2\sum_{n=1}^{\infty} n^{-1}\Phi(-|\mu|n^{1/2}/2)\right]$$

accounts for the excess over the boundary of the discrete-time process.

The heuristic justification for this approximation follows the same approach used by Siegmund (1985), Section 9.5, for a similar one-dimensional result for the square root boundary. By conditioning on $\|S_n\|$ and ω_n, (2) can be written as

$$P_\theta\{T_1 = n|\,\|S_n\| = \xi,\ \omega_n = \omega\}P_\theta\{\|S_n\| \in d\xi,\ \omega_n \in d\omega\},$$

which can be written equivalently as

$$P_0\{T_1 > n - 1|S_n = (\xi, 0)\}P_\theta\{\|S_n\| \in d\xi,\ \omega_n \in d\omega\} \tag{4}$$

by the rotational invariance of T_1 and the sufficiency of S_n for θ. The conditional probability in (4) can be shown to converge to $P_{b_1/n}\{S_{1i} > x \text{ for all } i \ge 1\}$ as a consequence of Lemma 2 of Woodroofe (1978). Furthermore,

$$\begin{aligned}&P_{b_1/n}\{S_{1i} > x \text{ for all } i \ge 1\}\\ &\quad= \int_x^\infty P_{b_1/n}\{S_{1i} > x \text{ for all } i \ge 2|S_{11} = y\}P_{b_1/n}\{S_{11} \in dy\}\\ &\quad= \int_x^\infty P_{b_1/n}\{S_{1i} > x - y \text{ for all } i \ge 1\}\phi(y - b_1/n)\,dy\\ &\quad= \int_x^\infty P_{b_1/n}\{\tilde\tau(x - y) = \infty\}\phi(y - b_1/n)\,dy,\end{aligned} \tag{5}$$

where $\tilde{\tau}(-x) = \inf\{n\colon S_{1n} \le -x\}$. In order to evaluate this integral further, it is necessary to use the large x approximation for $P_\mu\{\tilde{\tau}(-x) < \infty\}$ given in (9.65) in Siegmund (1985). That is, as $-(x-y) \to \infty$,

$$P_{b_1/n}\{\tilde{\tau}(x-y) = \infty\} \sim 1 - \nu(2b_1/n)e^{2(x-y)b_1/n}.$$

Substituting this approximation in (5), even though not all values of $-(x-y)$ are large, enables us to evaluate the integral to be

$$1 - \Phi(x - b_1/n) - \nu(2b_1/n)\exp(2b_1x/n)[1 - \Phi(x + b_1/n)],$$

which may be approximated further by

$$1 - \Phi(x - b_1/n) - \exp(2b_1(x-\rho)/n)[1 - \Phi(x + b_1/n)] \tag{6}$$

using $\nu(\mu) \sim \exp(-\rho\mu)$ and $\rho \cong 0.583$ [see Siegmund (1985), page 82]. The joint probability in (4) is calculated by a simple change of variables from Cartesian to polar coordinates.

Clearly, the use of this approximation will not lead to mathematically correct results because of the substitutions made in the above derivation. Nevertheless, it may lead to sufficiently accurate results that are usable in practice. One way to assess the accuracy of this approximation for the joint probability is to integrate it numerically over ω and ξ and sum it over n to get an approximation for $P_{\|\theta\|}\{T_1 \le m\}$. This can then be compared with the approximation given in Betensky (1992), Theorem 4.6, which is reasonably accurate in many examples. As indicated in Table 2, the approximation given here slightly underestimates the joint probability (insofar as this can be concluded based on its numerical integral) and does so slightly more for larger $\|\theta\|$ than for smaller $\|\theta\|$.

In Section 4.1 we outline the approximation for the error, $P_{(\theta_1, 0)}\{$treatment 1 or 2 is eliminated$\}$, and in Section 4.2 we outline the approximation for the power, $P_{(\theta_1, \theta_2)}\{$treatment 1 is selected$\}$. Both of these approximations involve conditioning on T_1, $\|S_{T_1}\|$ and ω_{T_1}, for which the approximation for the joint probability given in (3) is necessary. In Section 4.3 we compare these approximations with values from the simulation described in Section 3.

TABLE 2
Evaluation of $P_\theta\{T_1 = n,\ \|S_n\| \in d\xi,\ \omega_n \in d\omega\}$ *via* $P_\theta\{T_1 \le m\}$

$\|\theta\|$	**Approximation for density summed**	**Approximation for** $P_\theta\{T_1 \le m\}$
0.7	0.9513	0.9948
0.6	0.9333	0.9696
0.5	0.8562	0.8833
0.4	0.6802	0.6966
0.3	0.4330	0.4397

4.1. $P_{(\theta_1,0)}\{$*treatment* 1 *or* 2 *is eliminated*$\}$. It is convenient to express the event {treatment 1 or 2 is eliminated} as the union of the disjoint events {treatment 1 or 2 is eliminated after stage 1 of sampling} and {treatment 3 is eliminated after stage 1 of sampling and treatment 1 or 2 is eliminated after stage 2 of sampling}. Therefore,

$$P_{(\theta_1,0)}\{\text{treatment 1 or 2 is eliminated}\}$$

$$= P_{(\theta_1,0)}\{T_1 \le m,\ \omega_{T_1} \in [\pi/3, 5\pi/3]\} \tag{7}$$

$$+ P_{(\theta_1,0)}\{T_1 \le m,\ \omega_{T_1} \in [-\pi/3, \pi/3],\ T_2 \le m\}. \tag{8}$$

This follows from the fact that $\omega_{T_1} \in [\pi/3, 5\pi/3]$ if and only if $\max_{k,l} S_{T_1}^{k,l} = S_{T_1}^{3,1}$, $S_{T_1}^{2,1}$, $S_{T_1}^{12}$ or $S_{T_1}^{3,2}$ and that $\omega_{T_1} \in [-\pi/3, \pi/3]$ if and only if $\max_{k,l} S_{T_1}^{k,l} = S_{T_1}^{1,3}$ or $S_{T_1}^{2,3}$. The probability in (7) can be written as the joint probability (2) integrated over ω and ξ and summed over n:

$$\int_{\pi/3}^{5\pi/3} \int_{b_1}^{\infty} \sum_{n=1}^{m} P_{(\theta_1,0)}\{T_1 = n,\ \|S_n\| \in d\xi,\ \omega_n \in d\omega\}. \tag{9}$$

The approximation for (8) is more complicated than that for (7) because it involves T_2 as well as T_1. By conditioning on T_1, $\|S_{T_1}\|$ and ω_{T_1}, it can be written as

$$\int_{-\pi/3}^{\pi/3} \int_{b_1}^{\infty} \sum_{n=1}^{m} P_{(\theta_1,0)}\{T_2 \le m | T_1 = n,\ \|S_n\| = \xi,\ \omega_n = \omega\} \tag{10}$$

$$\times P_{(\theta_1,0)}\{T_1 = n,\ \|S_n\| \in d\xi,\ \omega_n \in d\omega\}.$$

The conditional probability in (10) is easily seen to equal

$$\begin{cases} 1, & \text{if } |\xi \sin\omega| \ge b_2, \\ P_0\{\tau \le m - n\}, & \text{if } |\xi \sin\omega| < b_2, \end{cases} \tag{11}$$

where

$$\tau = \inf\{k\colon S_k \ge b_2 - \xi \sin\omega \text{ or } S_k \le -b_2 - \xi \sin\omega\}. \tag{12}$$

Therefore, in the region $|\xi \sin\omega| < b_2$, the conditional probability in (10) is approximately equal to

$$P_0\{\tau_{b_2 - \xi\sin\omega} \le m - n\} + P_0\{\tau_{b_2 + \xi\sin\omega} \le m - n\}$$

$$\cong 1 - \Phi((m-n)^{-1/2}(b_2 - \xi\sin\omega))$$

$$+ \Phi(-(m-n)^{-1/2}(b_2 - \xi\sin\omega + 2\rho))$$

$$+ 1 - \Phi((m-n)^{-1/2}(b_2 + \xi\sin\omega))$$

$$+ \Phi(-(m-n)^{-1/2}(b_2 + \xi\sin\omega + 2\rho)),$$

where $\tau_a = \inf\{k\colon S_k \geq a\}$ and $\rho \cong 0.583$. The presence of the term ρ is to account for the excess over the boundary from the discrete process and is suggested by Siegmund (1985), Section 3.5. This approximation can be substituted into (10) along with (3) and it can be evaluated numerically to give (8).

A similar approach based on an approximation for the joint probability (2) can be used for approximating this error probability in the three-treatment procedure based on the repeated significance tests studied by Siegmund (1993). This approach as applied to Siegmund's problem is discussed in Betensky (1992) where it is shown to perform well and thus to be a good alternative to the ad hoc continuity correction he applies to a Brownian motion approximation.

4.2. $P_{(\theta_1,\theta_2)}\{\textit{treatment } 1 \textit{ is selected}\}$. It is convenient to express the event {treatment 1 is selected} as the union of the disjoint events {treatment 3 is eliminated at T_1, treatment 2 is eliminated at T_2} and {treatment 2 is eliminated at T_1, treatment 3 is eliminated at T_2}. Therefore,

$$P_{(\theta_1,\theta_2)}\{\text{treatment 1 is selected}\}$$

$$= P_{(\theta_1,\theta_2)}\{T_1 \leq m,\ \omega_{T_1} \in [-\pi/3, \pi/3],\ T_2 \leq m,\ S^{1,2}_{T_2} \geq b_2\} \tag{13}$$

$$+ P_{(\theta_1,\theta_2)}\{T_1 \leq m,\ \omega_{T_1} \in [\pi/3, \pi],\ T_2 \leq m,\ S^{1,3}_{T_2} \geq b_2\}, \tag{14}$$

where (13) can be written as

$$\int_{-\pi/3}^{\pi/3}\int_{b_1}^{\infty}\sum_{n=1}^{m} P_\theta\{T_2 \leq m,\ S^{1,2}_{T_2} \geq b_2 | T_1 = n,\ \|S_n\| = \xi,\ \omega_n = \omega\} \tag{15}$$
$$\times P_{(\theta_1,0)}\{T_1 = n,\ \|S_n\| \in d\xi,\ \omega_n \in d\omega\}$$

and (14) can be written as

$$\int_{\pi/3}^{\pi}\int_{b_1}^{\infty}\sum_{n=1}^{m} P_\theta\{T_2 \leq m,\ S^{1,3}_{T_2} \geq b_2 | T_1 = n,\ \|S_n\| = \xi,\ \omega_n = \omega\} \tag{16}$$
$$\times P_{(\theta_1,0)}\{T_1 = n,\ \|S_n\| \in d\xi,\ \omega_n \in d\omega\}.$$

The conditional probability in (15) is equal to

$$\begin{cases} 1, & \text{if } \xi \sin\omega \geq b_2, \\ 0, & \text{if } \xi \sin\omega \leq -b_2, \\ P_{\theta_2}\{\tau \leq m-n,\ S_\tau \geq b_2 - \xi\sin\omega\}, & \text{if } |\xi\sin\omega| < b_2, \end{cases} \tag{17}$$

where τ is defined in (12). Likewise, the conditional probability in (16) can be written analogously with $(3^{1/2}\xi\cos\omega + \xi\sin\omega)/2$ replacing $\xi\sin\omega$, $(3^{1/2}\theta_1 + \theta_2)/2$ replacing θ_2 and τ', replacing τ, where

$$\tau' = \inf\{k\colon S_k \geq b_2 - (3^{1/2}\xi\cos\omega + \xi\sin\omega)/2$$
$$\text{or } S_k \leq -b_2 - (3^{1/2}\xi\cos\omega + \xi\sin\omega)/2\}.$$

The probability in (17) may be approximated by

$$
\begin{aligned}
&P_{\theta_2}\{\tau_{b_2-\xi\sin\omega} \le m-n\} \\
&\quad \cong 1-\Phi((m-n)^{-1/2}(b_2-\xi\sin\omega)-(m-n)^{1/2}\theta_2) \\
&\qquad + e^{2(b_2-\xi\sin\omega+\rho)\theta_2} \\
&\qquad\quad \times \Phi(-(m-n)^{-1/2}(b_2-\xi\sin\omega+2\rho)-(m-n)^{1/2}\theta_2)
\end{aligned} \tag{18}
$$

and the analogous probability based on τ' is dealt with similarly. These approximations together with (3) are substituted in (15) and (16) to approximate the power.

4.3. *Numerical results.* Table 3 contains the approximations for the error probability (i.e., $P_{(\theta_1,0)}\{$eliminate treatment 1 or 2$\}$) and the power (i.e., $P_{(\theta_1,\theta_2)}\{$select treatment 1$\}$) as well as values from the simulation of the O'Brien–Fleming three-treatment trial described in Section 3. For the parameter values in this example ($m = 50$, $b_1 = 18.52$, $b_2 = 15.31$), the numerical approximation for the error consisting of the sum of (9) and (10) appears to be quite accurate. The approximation for the power consisting of the sum of (15) and (16) performs reasonably well for this example. The values from the approximation are almost always less than those from the simulation, sug-

TABLE 3
Error ($\theta_2 = 0$) *and significance level* ($\theta_2 > 0$) *from stage* 2 *of sampling*

θ_1	θ_2	Monte Carlo	Approximation
0.70	0.00	0.050	0.048
0.35	0.61	0.978	0.934
0.50	0.50	0.933	0.889
0.61	0.35	0.674	0.642
0.60	0.00	0.047	0.048
0.30	0.52	0.912	0.873
0.42	0.42	0.813	0.783
0.52	0.30	0.537	0.517
0.50	0.00	0.049	0.047
0.25	0.43	0.734	0.718
0.35	0.35	0.637	0.619
0.43	0.25	0.393	0.381
0.40	0.00	0.048	0.046
0.20	0.35	0.507	0.496
0.28	0.28	0.423	0.407
0.35	0.20	0.263	0.249
0.30	0.00	0.039	0.042
0.15	0.26	0.253	0.249
0.21	0.21	0.234	0.209
0.26	0.15	0.134	0.135

gesting that the downward bias in the approximation of the joint probability (2) may be playing a role here.

5. Expected sample size. Two approximations for the expected sample size after the second stage of sampling are derived in this section. The first is an analytic approximation valid up to terms of constant order as $m \to \infty$ and is discussed in Section 5.1. It follows the same approach used by Siegmund (1993) to approximate the expected sample size for the three-treatment trial based on repeated significance tests. However, our approximation is accurate up to terms of constant order, whereas Siegmund's (1993) approximation does not include terms smaller than order $m^{1/2}$. Consequently, our proof is slightly more complicated than that given in Siegmund (1993). The second approximation, described in Section 5.2, is based on summing probabilities and is similar in approach to the approximations developed in Section 4.1. Like those approximations, this one is not mathematically correct due to a variety of substitutions made. Nevertheless, it is more accurate than the analytic approximation for the numerical example considered. In Section 5.3 these approximations are compared with each other and with values from the simulation of the O'Brien–Fleming three-treatment trial described in Section 3.

5.1. *Analytic approximation of* $E_\theta(T_2 \wedge m)$. The basis for this approximation for the expected sample size is the following identity given in Siegmund (1993), Section 4:

$$\begin{aligned} E_\theta(T_2 \wedge m) = {} & E_\theta(T_1 \wedge m) + E_\theta(T_2 - T_1;\ T_1 \le m) \\ & - E_\theta(T_2 - m;\ T_1 \le m < T_2). \end{aligned} \tag{19}$$

In analyzing this it is most interesting to assume that $\mu_1 \gg \mu_2 \ge \mu_3$ because when μ_1 is close to μ_2 the second stage of sampling will use up most of the available patients so that $E_\theta(T_2 \wedge m)$ will be very close to m. Also, as in Siegmund (1993), we assume that

$$\|\theta\| = b_1/m + \Delta_1/m^{1/2} \tag{20}$$

and that

$$\theta_2 = b_2/m + \Delta_2/m^{1/2} \tag{21}$$

which give the relation

$$\theta_2 = b_2\|\theta\|/b_1 + \Delta_3/m^{1/2}, \tag{22}$$

where $\Delta_3 = \Delta_2 - \Delta_1 b_2/b_1$.

Every expectation in (19) may be taken over the set $\{|T_1 - b_1/\|\theta\|| < m^{5/8}\}$ without changing $E_\theta(T_2 \wedge m)$ by more than an exponentially small amount. This follows easily from bounding this expectation taken over the complementary set by

$$m(P_\theta\{T_1 > b_1/\|\theta\| + m^{5/8}\} + P_\theta\{T_1 < b_1/\|\theta\| - m^{5/8}\})$$

and bounding these probabilities using approximations in Betensky (1992).

Proposition 1 contains an approximation for $E_\theta(T_1 \wedge m)$, Proposition 2 contains an approximation for $E_\theta(T_2 - T_1;\ T_1 \le m)$ and Proposition 3 contains an approximation for $E_\theta(T_2 - m;\ T_1 \le m < T_2)$. These three results are combined via (19) to give an approximation for $E_\theta(T_2 \wedge m)$.

PROPOSITION 1. *Assume that* (20) *holds. Then, as* $m \to \infty$,

$$E_{\|\theta\|}(T_1 \wedge m) = \|\theta\|^{-1}\big(b_1 - m^{1/2}[\phi(\Delta_1) - \Delta_1\Phi(-\Delta_1)] - \|\theta\|^{-1}/2 + \rho_{\|\theta\|}\Phi(\Delta_1)\big) + o(1), \tag{23}$$

where $\rho_{\|\theta\|} = \lim_{m\to\infty} E_\theta(\|S_{T_1}\| - b_1)$ *counts for excess of the discrete process* $\|S_{T_1}\|$ *over the boundary* b_1 *and is approximately equal to* $0.583 + \|\theta\|/4$ *for small* $\|\theta\|$ [*see Siegmund* (1985), *page* 228].

PROOF. The proof proceeds by writing $E_{\|\theta\|}(T_1 \wedge m)$ as the sum

$$E_{\|\theta\|}(T_1) - E_{\|\theta\|}(T_1 - m;\ T_1 > m)$$

and approximating each component individually. The approximation for the first term follows from an application of Siegmund's (1985) Theorem 9.28. The conditions of this theorem are straightforward to verify. The approximation for the second term follows from an analysis similar to that used in the proof of Siegmund's (1985) Proposition 11.13. Other results that are used are approximations for the distribution of $\|S_m\|$ given in Betensky (1992), Theorem 4.5, and for $\rho_{\|\theta\|}$, the correction for discrete time. □

In the following results, where applicable, we assume that (20), (21) and (22) hold and we let $\gamma = (\|\theta\| b_1/m)^{1/2}\Delta_3/\theta_1$.

PROPOSITION 2. *As* $m \to \infty$,

$$E_\theta(T_2 - T_1;\ T_1 \le m) = C\Big\{ \frac{b_1^{1/2}\theta_1}{\theta_2\|\theta\|^{3/2}}[\phi(\gamma) - \gamma\Phi(-\gamma)]P_{\|\theta\|}\{T_1 \le m\} + \Phi(\Delta_1)\big[\Phi(-\gamma)(\rho_{\theta_2}\theta_2^{-1} - \rho_{\|\theta\|}\|\theta\|^{-1}) + (\gamma\phi(\gamma) + \Phi(-\gamma))\|\theta\|^{-2}/2\big]\Big\} + o(1),$$

where $C = 1$ *if* $\mu_2 \gg \mu_3$, $C = 2$ *if* $\mu_2 = \mu_3$ *and* $\rho_\mu \cong 0.583 + \mu/4$.

REMARK. For use in practice, we approximate $P_{\|\theta\|}\{T_1 \le m\}$ by

$$\Phi(\Delta_1) + \frac{\phi(\Delta_1)}{2m^{1/2}\|\theta\|} + \frac{\phi(\Delta_1)b_1^{1/2}e^{-2\rho b_1/m}}{m\|\theta\|^{1/2}(\|\theta\| + b_1/m)}$$

as in Theorem 4.6 in Betensky (1992). Although this is not a *bona fide* asymptotic approximation, it works well in practice.

PROOF. Up to an exponentially small error, $E_\theta(T_2 - T_1;\ T_1 \le m)$ can be written as

$$E_\theta\Big(T_2 - T_1;\ T_1 \le m,\ S_{T_1}^{1,3} = \max_{k,l} S_{T_1}^{k,l}\Big) + E_\theta\Big(T_2 - T_1;\ T_1 \le m,\ S_{T_1}^{1,2} = \max_{k,l} S_{T_1}^{k,l}\Big). \tag{24}$$

When μ_2 is much larger than μ_3, the second term is negligible, whereas when μ_2 is roughly equal to μ_3, the two terms are approximately equal. Therefore, it suffices also to consider the first term in (24). As in Siegmund's analysis, this first term is equal to

$$\sum_{n=1}^{m} \int_0^\infty P_\theta\{T_1 = n,\ S_{2,n} \in b_2 - dx\} E_\theta \tau(x) \tag{25}$$

plus an exponentially small error term, where $\tau(x) = \inf\{k\colon S_{2,k} \ge x\}$. Because it suffices to consider the expectation $E_\theta(T_2 \wedge m)$ over the set $\{|T_1 - b_1/\|\theta\|| < m^{5/8}\}$, it suffices also to consider this component of the expectation, $E_\theta(T_2 - T_1;\ T_1 \le m,\ S_{T_1}^{1,3} = \max_{k,l} S_{T_1}^{k,l})$, over that same set. Therefore, the effective range of summation in (25) is $b_1/\|\theta\| - m^{5/8} < n < b_1/\|\theta\| + m^{5/8}$. For n in this range the effective range of integration for x is $m^{1/8} < x < m^{5/8}$. This follows from arguments in Proposition 4.27 of Siegmund (1985). The expectation $E_\theta\tau(x)$ is approximately equal to $(x + \rho_{\theta_2})/\theta_2$, where $\rho_{\theta_2} \cong \rho + \theta_2/4$. The probability in (25) can be written as

$$\int_0^\infty P_\theta\{S_{2,T_1} \in b_2 - dx | T_1 = n,\ \|S_{T_1}\| = b_1 + r\} P_\theta\{T_1 = n,\ \|S_{T_1}\| \in b_1 + dr\}$$

and so the first step in this proof is to approximate the conditional probability

$$P_\theta\{S_{2,T_1} \in b_2 - dx | T_1, \|S_{T_1}\|\}. \tag{26}$$

LEMMA 1. *Assume that $x = O(m^{1/2})$. Then, as $m \to \infty$,*

$$\begin{aligned} &P_\theta\{S_{2,T_1} \in b_2 - dx | T_1, \|S_{T_1}\|\} \\ &\quad = \frac{\|\theta\|^{3/2}}{b_1^{1/2}\theta_1} \phi\Big(\frac{\|\theta\|^{3/2}}{b_1^{1/2}\theta_1}\Big(x + \frac{b_1\Delta_3}{m^{1/2}\|\theta\|}\Big)\Big) \\ &\qquad \times \Big[1 - \frac{\Delta_3\theta_2}{m^{1/2}\theta_1^2} - \frac{b_2 R\Delta_3\|\theta\|^2}{m^{1/2}b_1\theta_1^2} + \frac{b_1\Delta_3^3\theta_2\|\theta\|}{2m^{3/2}\theta_1^4} \\ &\qquad\quad + x\Big(\frac{3\Delta_3^2\theta_2\|\theta\|^2}{2m\theta_1^4} - \frac{b_2\|\theta\|^2}{b_1^2\theta_1^2} - \frac{b_2 R\|\theta\|^3}{b_1^2\theta_1^2}\Big) \\ &\qquad\quad + x^2\frac{3\Delta_3\theta_2\|\theta\|^3}{2m^{1/2}b_1\theta_1^4} + x^3\Big(\frac{\theta_2\|\theta\|^4}{2b_1^2\theta_1^4}\Big) + O_p(m^{-1})\Big]\,dx, \end{aligned} \tag{27}$$

where $R = \|S_{T_1}\| - b_1$.

REMARK. Although the effective range of integration for x extends to $m^{5/8}$, for the purpose of evaluating (25) up to terms of constant order, it suffices to assume that $x = O(m^{1/2})$ and to expand (26) up to terms of order $m^{-1/2}$.

PROOF. The starting point for this approximation is the likelihood ratio calculation given in Siegmund (1985), page 114, which shows that

$$P_\theta\{S_{2,T_1} \in b_2 - dx | T_1, \|S_{T_1}\|\} = \frac{\exp(\|\theta\|\,\|S_{T_1}\|\cos[\sin^{-1}((b_2-x)/\|S_{T_1}\|) - \sin^{-1}(\theta_2/\|\theta\|)])}{2\pi I_0(\|\theta\|\,\|S_{T_1}\|)[\|S_{T_1}\|^2 - (b_2-x)^2]^{1/2}}, \tag{28}$$

where I_0 denotes the modified Bessel function and $I_0(\nu) = (2\pi\nu)^{-1/2}\exp(\nu)(1+O(\nu^{-1}))$. The rest of the proof consists of extensive Taylor series expansions using (20) and (21) and the fact that $\|S_{T_1}\| = b_1 + R$, where R is stochastically bounded. □

The proof of the proposition concludes with the evaluation of

$$\sum_{n=1}^{m}\int_0^\infty P_\theta\{T_1 = n,\ \|S_{T_1}\| \in b_1 + dr\} \times \int_{m^{1/8}}^{m^{5/8}} b_1^{-1/2}\theta_1^{-1}\|\theta\|^{3/2}\phi(b_1^{-1/2}\theta_1^{-1}\|\theta\|^{3/2}x + \gamma) \times (k_0 + k_1x + k_2x^2 + k_3x^3 + O_p(m^{-1}))(x + \rho_{\theta_2})\,dx/\theta_2, \tag{29}$$

where the constants k_0, k_1, k_2, k_3 are obtained from (27). It is convenient to change the variable of integration from x to

$$y = b_1^{-1/2}\theta_1^{-1}\|\theta\|^{3/2}x + \gamma.$$

The inner integral in (29) is equal to

$$\frac{b_1^{1/2}\theta_1}{\theta_2\|\theta\|^{3/2}}[\phi(\gamma) - \gamma\Phi(-\gamma)] + \frac{\rho_{\theta_2}}{\theta_2}\Phi(-\gamma) - \frac{b_2R}{b_1\theta_2}\Phi(-\gamma) + \frac{1}{2\|\theta\|^2}[\gamma\phi(\gamma) + \Phi(-\gamma)] + o_p(1)$$

and thus (29) is equal to

$$\frac{b_1^{1/2}\theta_1}{\theta_2\|\theta\|^{3/2}}[\phi(\gamma) - \gamma\Phi(-\gamma)]P_{\|\theta\|}\{T_1 \le m\} + \Phi(\Delta_1)[\theta_2^{-1}\rho_{\theta_2}\Phi(-\gamma) + \|\theta\|^{-2}(\gamma\phi(\gamma) + \Phi(-\gamma))/2] - \theta_2^{-1}(b_2/b_1)\Phi(-\gamma)E_\theta(\|S_{T_1}\| - b_1;\ T_1 \le m) + o(1).$$

Finally, since the excess over the boundary, $\|S_{T_1}\| - b_1$, and T_1 are asymptotically independent [Betensky (1992), Theorem 4.2],

$$E_\theta(\|S_{T_1}\| - b_1;\ T_1 \le m) \to \rho_{\|\theta\|}\Phi(\Delta_1),$$

where $\rho_{\|\theta\|} = \lim_{m\to\infty} E_\theta(\|S_{T_1}\| - b_1)$. This and another application of (22) give the result of the proposition. □

PROPOSITION 3. *As* $m \to \infty$,

$$\begin{aligned} &E_\theta(T_2 - m;\ T_1 \le m < T_2) \\ &\quad = \int_0^\infty (m^{1/2}y + \rho_{\theta_2})\phi(\Delta_2 + y) \\ &\qquad \times \Bigg[1 - \Phi\Bigg(-\frac{\Delta_1\|\theta\|}{\theta_1} + \frac{\Delta_2\theta_2}{\theta_1} - \frac{\Delta_3^2\|\theta\|}{2m^{1/2}\theta_1^2} \\ &\qquad\qquad + y\Bigg(\frac{\theta_2}{\theta_1} - \frac{\Delta_3\|\theta\|^2}{m^{1/2}\theta_1^3}\Bigg) - y^2\frac{\|\theta\|^2}{2m^{1/2}\theta_1^2}\Bigg)\Bigg]\,dy/\theta_2 \\ &\qquad + \frac{b_1\theta_1}{m\theta_2\|\theta\|^2}[\phi(\gamma) - \gamma\Phi(-\gamma)]m^{1/2}P_{\|\theta\|}\{T_1 < m,\ \|S_m\| < b_1\}o(1). \end{aligned}$$

REMARK. For use in practice, we approximate $P_{\|\theta\|}\{T_1 < m,\ \|S_m\| < b_1\}$ by $m^{-1/2}\phi(\Delta_1)e^{-2\rho b_1/m}/(\|\theta\| + b_1/m)$ as in Theorem 4.4 of Betensky (1992).

PROOF. As in Siegmund (1993), this expectation can be written as

$$\begin{aligned} &E_\theta\Big(T_2 - m;\ T_1 \le m < T_2,\ S_{T_1}^{1,3} = \max_{k,l} S_{T_1}^{k,l}\Big) \\ &\quad + E_\theta\Big(T_2 - m;\ T_1 \le m < T_2,\ S_{T_1}^{1,2} = \max_{k,l} S_{T_1}^{k,l}\Big) \end{aligned} \tag{30}$$

up to an exponentially small error. The first term in (30) is equal to

$$\int_0^\infty P_\theta\{T_1 \le m,\ T_2 > m,\ S_{T_1}^{1,3} = \max_{k,l} S_{T_1}^{k,l},\ S_{2,m} \in b_2 - dx\}E_\theta\tau(x), \tag{31}$$

where $\tau(x) = \inf\{k\colon S_{2,k} \ge x\}$ and $E_\theta(\tau(x)) \sim (x + \rho_{\theta_2})/\theta_2$. It is easy to see that the important range for x in this integral is $m^{1/8} < x < m^{5/8}$ as well. It can be shown that the constraints $S_{T_1}^{1,3} = \max_{k,l} S_{T_1}^{k,l}$ and $T_2 > m$ can be eliminated from the probability in (31) without changing (31) by more than an exponentially small amount. Therefore, the integral in (31) is equal to

$$\begin{aligned} &\int_{m^{1/8}}^{m^{5/8}} P_\theta\{\|S_m\| \ge b_1,\ S_{2,m} \in b_2 - dx\}(x + \rho_{\theta_2})\,dx/\theta_2 \\ &\quad + \int_{m^{1/8}}^{m^{5/8}} P_\theta\{T_1 < m,\ \|S_m\| < b_1,\ S_{2,m} \in b_2 - dx\}(x + \rho_{\theta_2})\,dx/\theta_2 \end{aligned} \tag{32}$$

plus an exponentially small error. Lemma 2 contains an approximation for the probability in the first integral in (32) and Lemma 3 contains an approximation for the second integral in (32). □

LEMMA 2. *Assume that* $x = O(m^{1/2})$. *Then, as* $m \to \infty$,

$$P_\theta\{\|S_m\| \geq b_1,\ S_{2,m} \in b_2 - dx\}$$
$$\sim \phi(\Delta_2 + m^{-1/2}x)\Bigg[1 - \Phi\Bigg(-\frac{\Delta_1\|\theta\|}{\theta_1} + \frac{\Delta_2\theta_2}{\theta_1} - \frac{\Delta_3^2\|\theta\|}{2m^{1/2}\theta_1^2} + \frac{x}{m^{1/2}}\left(\frac{\theta_2}{\theta_1} - \frac{\Delta_3\|\theta\|^2}{m^{1/2}\theta_1^3}\right) - \frac{x^2\|\theta\|^2}{2m^{3/2}\theta_1^2} + O(m^{-1})\Bigg)\Bigg]\frac{dx}{m^{1/2}}.$$

REMARK. As in Lemma 1, it suffices to expand this probability for $x = O(m^{1/2})$ even though x can be as large as $m^{5/8}$.

PROOF. Up to an exponentially small error, this probability can be written as

$$P_\theta\{\|S_m\| \geq b_1,\ S_{2,m} \in b_2 - dx\}$$
$$= \phi(\Delta_2 + m^{-1/2}x)[1 - \Phi(m^{-1/2}[b_1^2 - (b_2 - x)^2]^{1/2} - m^{1/2}\theta_1)]m^{-1/2}\,dx.$$

The rest of the proof consists of a sequence of Taylor series expansions, each of which makes repeated use of the assumptions. □

LEMMA 3. *Assume that* (20), (21) *and* (22) *hold. Then, as* $m \to \infty$,

$$\int_{m^{1/8}}^{m^{5/8}} P_\theta\{T_1 < m,\ \|S_m\| < b_1,\ S_{2,m} \in b_2 - dx\}((x + \rho_{\theta_2})/\theta_2)\,dx$$
$$= \frac{b_1\theta_1}{m\theta_2\|\theta\|^2}[\phi(\gamma) - \gamma\Phi(-\gamma)]m^{1/2}P_{\|\theta\|}\{T_1 < m,\ \|S_m\| < b_1\} + o(1).$$

PROOF. First note that

$$\int_{m^{1/8}}^{m^{5/8}} P_\theta\{T_1 < m,\ \|S_m\| < b_1,\ S_{2,m} \in b_2 - dx\}(x + \rho_{\theta_2})\,dx/\theta_2$$
$$\sim \int_{m^{1/8}}^{m^{5/8}} P_\theta\{T_1 < m,\ b_1 - y < \|S_m\| < b_1,\ S_{2,m} \in b_2 - dx\}(x + \rho_{\theta_2})\,dx/\theta_2$$

because, as shown in the proof of Theorem 4.4 in Betensky (1992), under the assumption (20), the major contribution to the joint probability $P_\theta\{T_1 < m,\ \|S_m\| < b_1\}$ occurs for $\|S_m\|$ close to b_1, and for $\|S_m\|$ outside of this region, the joint probability is exponentially small. Therefore, this integral is equal to

$$\int_0^y \int_{m^{1/8}}^{m^{5/8}} P_\theta\{T_1 < m,\ \|S_m\| \in b_1 - dr\}P_\theta\{S_{2,m} \in b_2 - dx \mid \|S_m\| = b_1 - r\}$$
$$\times (x + \rho_{\theta_2})\,dx/\theta_2$$

for fixed $y > 0$. Because r is small relative to b_1, the result of Proposition 2 may be used to evaluate the integral with respect to x to obtain

$$\int_0^y P_\theta\{T_1 < m,\ \|S_m\| \in b_1 - dr\}$$
$$\times \left[\frac{b_1^{1/2}\theta_1}{\|\theta\|^{3/2}\theta_2}[\phi(\gamma) - \gamma\Phi(-\gamma)] + \theta_2^{-1}\rho_{\theta_2}\Phi(-\gamma)\right.$$
$$\left. + \|\theta\|^{-2}[\gamma\phi(\gamma) + \Phi(-\gamma)]/2 + r\theta_2^{-1}(b_2/b_1)\Phi(-\gamma) + o_p(1)\right].$$

The integral with respect to r is

$$\frac{b_1\theta_1}{m\theta_2\|\theta\|^2}[\phi(\gamma) - \gamma\Phi(-\gamma)]m^{1/2}P_{\|\theta\|}\{T_1 < m,\ \|S_m\| < b_1\}$$
$$+ \theta_2^{-1}(b_2/b_1)\Phi(-\gamma)E_\theta(b_1 - \|S_m\|;\ T_1 < m,\ b_1 - y < \|S_m\| < b_1) + o(1).$$

The result of the lemma now follows from the fact that the last term in this expression is negligible because $b_1 - \|S_m\| < y$, y is taken to be fixed and the joint probability of T_1 and $\|S_m\|$ is $O(m^{-1/2})$. □

The results of Propositions 1, 2 and 3, combined by means of (19), yield an approximation for $E_\theta(T_2 \wedge m)$. Because of the asymptotic relations that are assumed for this theorem, there are many mathematically equivalent ways that this approximation could be expressed. It is likely that certain forms of the approximation perform better on certain numerical examples than others. Because this result is accurate up to constant order terms, the proof is more complicated than the proof of Siegmund's (1993) analogous result for the square root boundary up to terms of order $m^{1/2}$. However, the use of the straight line boundary does simplify some aspects of this proof because the value of the process $\|S_n\|$ at the stopping time T_1 is $b_1 + R$ rather than $b_1T_1^{1/2} + R$ as in the square root boundary case.

5.2. *$E_\theta(T_2 \wedge m)$ via summing probabilities.* An alternative approach to approximating $E_\theta(T_2 \wedge m)$ is by summing probabilities. The expected sample size $E_\theta(T_2 \wedge m)$ can be written as

$$m - \sum_{n=1}^{m-1}\int_0^{2\pi}\int_{b_1}^{\infty}\sum_{j=n}^{m-1} P_\theta\{T_2 \le j | T_1 = n,\ \|S_n\| = \xi,\ \omega_n = \omega\}$$
$$\times P_\theta\{T_1 = n,\ \|S_n\| \in d\xi,\ \omega_n \in d\omega\}. \tag{33}$$

The joint probability in (33) is approximated as in (3) and the conditional probability is approximated as in (11) by

$$\begin{cases} 1, & \text{if } |c| \ge b_2, \\ P_\mu\{\tau_{b_2-c} \le j - n\} + P_{-\mu}\{\tau_{b_2+c} \le j - n\}, & \text{if } |c| < b_2, \end{cases}$$

where

$$\mu = \begin{cases} \theta_2, & \text{if } \omega_n \in [-\pi/3, \pi/3], \\ (3^{1/2}\theta_1 + \theta_2)/2, & \text{if } \omega_n \in [\pi/3, \pi], \\ (3^{1/2}\theta_1 - \theta_2)/2, & \text{if } \omega_n \in [\pi, 5\pi/3], \end{cases}$$

$$c = \begin{cases} \xi \sin \omega, & \text{if } \omega_n \in [-\pi/3, \pi/3], \\ (3^{1/2}\xi \cos \omega + \xi \sin \omega)/2, & \text{if } \omega_n \in [\pi/3, \pi], \\ (3^{1/2}\xi \cos \omega - \xi \sin \omega)/2, & \text{if } \omega_n \in [\pi, 5\pi/3], \end{cases}$$

and the probabilities of the form $P_\mu\{\tau_{b_2-c} \le j - n\}$ are approximated as in (18). The approximation (33) is then computed using multiple numerical integration.

5.3. *Numerical results.* Table 4 contains the approximations for the expected sample size of the three-treatment trial given in Sections 5.1 and 5.2 as well as values from the simulation of the three-treatment O'Brien–Fleming trial described in Section 3. The analytic approximation is reasonably accurate for this example, for $\|\theta\| \ge 0.4$. Given the many appearances of $\|\theta\|^{-1}$ and θ_2^{-1} in this approximation, it cannot be expected to do as well in the region of small $\|\theta\|$ and θ_2 values. In this table the constant C from Proposition 2 is taken to be 2 when $\mu_2 = \mu_3$, that is, when $\theta_2 = 3^{1/2}\theta_1$, and is taken to be

TABLE 4
Expected sample size of stage 2 of sampling: $E_\theta\{T_2 \wedge m\}$

θ_1	θ_2	Monte Carlo	Theorem 1	Numerical integration
0.70	0.00	49.47		49.50
0.35	0.61	29.94	29.49	30.63
0.50	0.50	32.20	31.77	33.18
0.61	0.35	40.06	39.06	40.55
0.60	0.00	49.51		49.50
0.30	0.52	34.50	34.15	35.05
0.42	0.42	37.11	36.44	37.71
0.52	0.30	42.66	41.94	43.15
0.50	0.00	49.50		49.52
0.25	0.43	39.94	39.10	40.02
0.35	0.35	41.54	40.63	41.83
0.43	0.25	45.45	44.96	45.48
0.40	0.00	49.53		49.56
0.20	0.35	44.07	42.36	44.23
0.28	0.28	45.11	43.78	45.44
0.35	0.20	47.24	47.04	47.34
0.30	0.00	49.63		49.63
0.15	0.26	47.58	41.87	47.54
0.21	0.21	47.86	43.44	47.96
0.26	0.15	48.70	44.17	48.70

1 otherwise. This should imply that the approximation is most accurate for those (θ_1, θ_2) pairs with $\theta_2 = 3^{1/2}\theta_1$. In fact, although the approximation is accurate for all (θ_1, θ_2) pairs, this does appear to be the case for large $\|\theta\|$. However, for smaller values of $\|\theta\|$, the approximation seems to be slightly more accurate for the other (θ_1, θ_2) pairs. This may just be an artifact of $\|\theta\|$ and θ_2 values that are too small. Actually, with some more effort, the constant C could probably be defined more precisely as a function of θ_1 and θ_2. Also, it is possible that alternative, equivalent expressions of the approximation could lead to improved numerical results.

Although the approximation given in Section 5.2 is relatively slow computationally because of the multiple numerical integrations, Table 4 indicates that it is quite accurate. It has the added advantage over the analytic approximation of being applicable for the $\theta_2 = 0$ entries of the table. However, these entries are the least interesting in the table as they are well approximated by m.

6. Standard versus experimental treatments. It is often the case that investigators conduct experiments in which new treatments are compared with an established standard treatment. In such experimental situations, it is appropriate to use the one-sided alternative hypothesis that the new treatments are better than the standard. That is, the burden should be on the new treatments to prove themselves superior to the standard, which is already well established. It is appropriate also that the experiment should be designed to stop as soon as there is sufficient evidence that the new treatments are not superior to the standard. There is nothing to be gained by letting the experiment use its maximum sample size in this situation because there is usually already a good understanding of the positive and adverse responses to the standard treatment. Various authors have adapted two-treatment sequential trials to one-sided alternatives and to stop early to accept the null hypothesis [e.g., DeMets and Ware (1980, 1982) and Siegmund (1986)]. Here we extend Siegmund's (1986) approach to the three-treatment setting for modified repeated significance tests.

Assume that treatment 3 is the standard treatment. The goal of the first stage of sampling should be to determine whether either of the new treatments is superior to the standard. Therefore, the first stage of sampling is defined by the alternative hypothesis of H_1: $\mu_1 > \mu_3$ or $\mu_2 > \mu_3$. It could be argued that the first stage should also be used to compare treatments 1 and 2. However, it seems more appropriate for this experimental situation to postpone this comparison until it is determined that either one of these new treatments is better than the standard. Define

$$T_1 = \inf\{n\colon n \geq m_0,\ S_n^{1,3} \geq b_1 n^{1/2} \text{ or } S_n^{1,3} \leq -b_2 n^{1/2} + \delta n \text{ or } S_n^{2,3} \geq b_1 n^{1/2} \text{ or } S_n^{2,3} \leq -b_2 n^{1/2} + \delta n\}, \tag{34}$$

and stop the first stage of sampling at $\min(T_1, m)$. If $T_1 \leq m$ and $S_{T_1}^{i,3} \geq b_1 T_1^{1/2}$ or if $T_1 > m$ and $S_m^{i,3} \geq c_1 m^{1/2}$, then conclude that treatment i is superior

to treatment 3, where $i = 1$ or 2. If $T_1 \le m$ and $S_{T_1}^{i,3} \le -b_2 T_1^{1/2} + \delta T_1$, then conclude that treatment i is not superior to treatment 3.

If either new treatment is determined to be superior to treatment 3, eliminate treatment 3 and continue sampling until $\min(T_2, m)$ where

$$T_2 = \inf\{n\colon n \ge T_1,\ |S_n^{1,2}| \ge b_3 n^{1/2}\},$$

and declare treatment i to be the best if $T_2 \le m$ and $S_{T_2}^{i,j} \ge b_3 T_2^{1/2}$ or if $T_2 > m$ and $S_m^{i,j} \ge c_2 m^{1/2}$, where $(i, j) = (1, 2)$ or $(2, 1)$.

Alternatively, if treatment i is found to be no better than treatment 3 after the first stage of sampling, eliminate treatment i and continue sampling until $\min(T_{2,j}, m)$ where

$$T_{2,j} = \inf\{n\colon n \ge T_1, S_n^{j,3} \ge b_1 n^{1/2} \text{ or } S_n^{j,3} \le -b_2 n^{1/2} + \delta n\},$$

where $j \ne i$, $j \ne 3$, and declare treatment j to be the best if $T_{2,j} \le m$ and $S_{T_{2,j}}^{j,3} \ge b_1 T_{2,j}^{1/2}$ or if $T_{2,j} > m$ and $S_m^{j,3} \ge c_1 m^{1/2}$, and declare no treatment difference if $T_{2,j} \le m$ and $S_{T_{2,j}}^{j,3} \le -b_2 T_{2,j}^{1/2} + \delta T_{2,j}$.

If it happens that two of the possible boundary crossing events defining T_1 occur at the same time, such as $S_{T_1}^{1,3} \ge b_1 T_1^{1/2}$ and $S_{T_1}^{2,3} \le -b_2 T_1^{1/2} + \delta T_1$, then we eliminate treatment 3 because treatment 1 has proven itself better than treatment 3, and we continue comparing treatments 1 and 2 using T_2. Presumably, treatment 1 will prove itself better than treatment 2 because treatment 2 does not appear to be any better than treatment 3.

With an appropriate choice of parameters, the proposed procedure should stop significantly early relative to the original procedure under the null hypothesis. A reasonable goal is for the expected sample size after the second stage to be roughly $3m/2$. At best, with a good choice of parameters, the proposed procedure will not display a significant loss of power or increase in sample size under the alternative hypothesis of $\mu_1 \ge \mu_2 \ge \mu_3$.

To make this procedure comparable to Siegmund's (1993) simulation of the original procedure described in Section 3, the Type I error of declaring a treatment difference when there is none must be set to 0.05/3. This is because the first stage of sampling for the original procedure stops and rejects H_0 when any one of the three random walks representing the three pairwise comparisons crosses an upper or lower boundary. The modified procedure stops and rejects H_0 after either stage of sampling as soon as either one of only two random walks crosses an upper boundary. Therefore, there are two boundaries that could be crossed for the modified procedure, versus six for the original procedure. A good approximation to an overall significance level of 0.05/3 is achieved by choosing b_1 and c_1 to be the boundary values that give a significance level of 0.05/3 for the two-sided two-treatment repeated significance test.

The other error of interest in the original procedure, the error of eliminating treatment 1 or 2 when they are equivalent, must be redefined for the modified

procedure. For this procedure it should be considered an error to eliminate treatment 1 or 2 only when they are being compared with each other. It is to be expected that treatment 1 or 2 might be eliminated by treatment 3 when the differences between the treatments are not that great, and so this is an error of lesser magnitude than if treatment 1 or 2 eliminates the other. Therefore, b_3 and c_2 are chosen to be the 0.05 level boundary values for the two-sided repeated significance test with maximum sample size of m.

In order to compare this modified procedure to Siegmund's (1993) simulation of the repeated significance three-treatment procedure, the parameters can be chosen according to the guidelines explained above to be $b_1 = 3.45$, $c_1 = 2.45$, $b_3 = 2.92$, $c_2 = 2.05$ and $\delta = 0.75$. For simplicity, b_2 is chosen to be equal to b_1. The error probability and power, p_1, is defined exactly as in Section 3, except that it may involve both stages of sampling. If treatment 3 eliminates treatment 1 after the first stage of sampling, then no conclusion regarding the presence of a treatment effect can be made until after the second stage of sampling. The error, p_2, of eliminating treatment 1 or 2 when they are equivalent is counted only when they are compared against each other. The power, p_2, and the expected sample sizes E_1 and E_2 are defined as for the original procedure. Table 5 lists these operating characteristics, along with the expected total sample size, for the modified procedure based on a

TABLE 5
Modified procedure for comparing standard vs. new treatments

θ_1	θ_2	p_1	p_2	E_1	E_2	Total
0.00	0.00	0.017		17.97	26.02	70.01
0.70	0.00	0.993	0.044	24.27	49.37	123.01
0.35	0.61	0.965	0.961	16.94	32.07	81.08
0.50	0.50	0.990	0.948	19.26	32.84	84.94
0.61	0.35	0.993	0.714	21.60	40.87	103.34
0.60	0.00	0.961	0.041	29.29	49.07	127.43
0.30	0.52	0.890	0.879	18.23	37.08	92.39
0.42	0.42	0.942	0.857	21.91	38.84	99.59
0.52	0.30	0.961	0.625	25.00	43.42	111.84
0.50	0.00	0.863	0.032	32.63	48.36	129.35
0.25	0.43	0.720	0.698	19.50	40.79	101.08
0.35	0.35	0.820	0.712	23.38	42.13	107.64
0.43	0.25	0.859	0.527	27.84	45.00	117.84
0.40	0.00	0.660	0.021	33.24	46.54	126.32
0.20	0.35	0.506	0.480	20.20	40.73	101.66
0.28	0.28	0.599	0.500	24.20	42.65	109.50
0.35	0.20	0.668	0.413	28.33	44.75	117.83
0.30	0.00	0.416	0.012	30.52	42.85	116.22
0.15	0.26	0.276	0.248	20.48	37.80	96.08
0.21	0.21	0.350	0.276	23.79	39.77	103.33
0.26	0.15	0.397	0.242	27.08	41.72	110.52

simulation. These values should be compared with the values in the columns headed "rs" in Table 1.

As desired, when there is no treatment difference, that is, $\theta_1 = \theta_2 = 0$, the value of p_1 is 0.017, approximately 0.05/3, and the value of the expected total sample size is 70.01, approximately $3m/2 = 75$. There is no value for p_2 listed in the table because it has no meaning for these parameter values. The values of p_1 are comparable to those from the original procedure for large $\|\theta\|$, except that they are slightly lower when $\mu_1 > \mu_2 = \mu_3$. This is probably because the treatment 2 versus treatment 3 comparison crosses its lower boundary before the treatment 1 versus treatment 3 comparison, crosses its upper boundary, and the treatment 1 versus treatment 3 comparison, which continues in the second stage, does not have enough time to detect treatment 1 as better than treatment 3. Also, the values of p_1 drop with $\|\theta\|$ because as the treatment differences decrease in magnitude, the procedure is more likely to stop early in favor of no treatment difference than to detect very small differences.

The values of p_2 when it is an error probability are slightly lower for the modified procedure for large $\|\theta\|$ and get progressively lower for smaller $\|\theta\|$. This is due to the different definitions of p_2 for the two procedures. The values of this error probability are smaller for the modified procedure because it is counting fewer events as errors than is the original procedure. For large $\|\theta\|$, for the original procedure, it is likely that treatment 3 is eliminated after the first stage of sampling and so the event of treatment 1 or 2 eliminating the other constitutes the major contribution to p_2. However, for smaller $\|\theta\|$, this event constitutes less of the contribution to p_2 for the original procedure and so p_2 from the modified procedure will represent proportionately less of p_2 from the original procedure. As the power, p_2 for the modified procedure exceeds that for the original procedure throughout the table, except when $\mu_1 > \mu_2 = \mu_3$. This may be happening for the same reason that the values of p_1 for these entries are smaller than for the original procedure.

There is quite a bit of freedom in the choice of parameters for this procedure. Even though the performance of this procedure could surely be improved for this example, it already is reasonably good. It achieves the goal of stopping early as soon as there is evidence that the new treatments are not superior to the standard at no great loss of power.

Alternatively, an adaptation of the O'Brien–Fleming two-sample test could be extended to the three-treatment procedure by defining the stopping rules

$$\begin{aligned} T_1 &= \inf\{n\colon S_n^{1,3} \geq b_1 \text{ or } S_n^{1,3} \leq -b_2 + \delta n \\ &\qquad \text{or } S_n^{2,3} \geq b_1 \text{ or } S_n^{2,3} \leq -b_2 + \delta n\}, \\ T_2 &= \inf\{n\colon n \geq T_1,\ |S_n^{1,2}| \geq b_3\} \end{aligned}$$

and

$$T_{2,j} = \inf\{n\colon n \geq T_1,\ S_n^{j,3} \geq b_1 \text{ or } S_n^{j,3} \leq -b_2 + \delta n\}.$$

Another possibility is to base a procedure on sequential probability ratio tests so that

$$T_1 = \inf\{n\colon S_n^{1,3} \geq b_1 + \delta n \text{ or } S_n^{1,3} \leq -b_2 + \delta n \text{ or } S_n^{2,3} \geq b_1 + \delta n \text{ or } S_n^{2,3} \leq -b_2 + \delta n\},$$

$$T_2 = \inf\{n\colon n \geq T_1,\ |S_n^{1,2}| \geq b_3\}$$

and

$$T_{2,j} = \inf\{n\colon n \geq T_1,\ S_n^{j,3} \geq b_1 + \delta n \text{ or } S_n^{j,3} \leq -b_2 + \delta n\}.$$

Paulson (1962) takes an indifference-zone approach to this problem. A different approach is to apply what has been proposed for comparing several new treatments with a standard in the fixed sample case. Dunnett (1955) studies this problem and proposes a procedure that results in narrower confidence limits for all of the comparisons between the standard treatment and each new treatment. He recommends that $n(d-1)^{1/2}$ observations be taken on the standard treatment, where there are $d-1$ experimental treatments, and n observations taken on each of the experimental treatments. An example in which this approach was taken is the Coronary Drug Project [Coronary Drug Project Research Group (1973)]. This study allocated 2.5 times as many patients to the placebo treatment as it did to any of the five active treatments. This ratio was chosen similarly to Dunnett's to minimize the variance for the five placebo versus experimental comparisons to be made of five-year mortality. This approach could be adapted to the three-treatment sequential procedure and it is likely to lead to better estimates of treatment differences. It is not so obvious, however, what the effects would be on total sample size.

7. Discussion. We have examined a procedure for sequentially selecting the best of three treatments that is truncated at a maximum sample size and that eliminates an apparently inferior treatment when there is sufficient evidence to do so. Our procedure is based on the popular O'Brien–Fleming stopping boundaries. We have compared it to that investigated by Siegmund (1993) based on repeated significance boundaries. We have derived analytic approximations for various operating characteristics of interest. These provide insight into the roles of the various parameters that define the procedure. In so doing, they are useful for both the design and analysis of such procedures. In addition, we have proposed several modifications to our procedure for the comparison of a standard treatment to experimental treatments, when early stopping in favor of the null hypothesis may be desirable.

It is straightforward to apply this procedure and the approximations to grouped responses, which is a more realistic scenario than continuous monitoring, via an appropriate change of scale. Betensky (1995) has extended this procedure to censored survival data which are more common than instantaneous normal responses.

Woodroofe and Coad (1995) have recently used "very weak expansions" introduced by Woodroofe (1986, 1989) to define a confidence region for the treat-

ment effect vector at the end of the trial. This greatly increases the usefulness of the procedure in practice. Woodroofe and Coad (1995) find that their approximations are slightly better for our O'Brien–Fleming-based procedure, than for the repeated significance–based procedure of Siegmund (1993).

To make this procedure even more useful in practice, it is necessary to develop it for the case of unknown and perhaps unequal variances among the treatment groups. This extension would make unequal sampling from the treatment groups feasible, and raises the possibility of an adaptive sampling plan. It seems likely that the methods of Siegmund (1985), Section 5.4, could be applied to this end, although the analytic approximations will certainly be more complicated.

Another extension of interest is the comparison of more than three treatments. An analogous procedure could be defined, based on global sequential tests at each stage. Beyond the first stage, analytic approximations seem impractical, although some crude results may be possible.

Acknowledgment. I thank David Siegmund for advising this research for my dissertation at Stanford University.

REFERENCES

BECHHOFER, R. E., SANTNER, T. J. and GOLDSMAN, D. M. (1995). *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons*. Wiley, New York.

BETENSKY, R. A. (1992). A study of sequential procedures for comparing three treatments. Ph.D. dissertation, Stanford Univ.

BETENSKY, R. A. (1995). Sequential analysis of censored survival data from three treatment groups. Unpublished manuscript.

Coronary Drug Project Research Group (1973). The Coronary Drug Project: design, methods, and baseline results. *Circulation* **47** 1–50.

DEMETS, D. L. and WARE, J. H. (1980). Group sequential methods for clinical trials with a one-sided hypothesis. *Biometrika* **67** 651–660.

DEMETS, D. L. and WARE, J. H. (1982). Asymmetric group sequential boundaries for monitoring clinical trials. *Biometrika* **69** 661–663.

DUNNETT, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50** 1096–1121.

FOLLMANN, D. A., PROSCHAN, M. A. and GELLER, N. L. (1994). Monitoring pairwise comparisons in multi-armed clinical trials. *Biometrics* **50** 325–336.

HSU, J. C. and EDWARDS, D. G. (1983). Sequential multiple comparisons with the best. *J. Amer. Statist. Assoc.* **78** 958–964.

HUGHES, M. D. (1993). Stopping guidelines for clinical trials with multiple treatments. *Statistics in Medicine* **12** 901–915.

JENNISON, C., JOHNSTONE, I. M. and TURNBULL, B. W. (1982). Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) **2** 55–86. Academic Press, New York.

JENNISON, C. and TURNBULL, B. W. (1991). Exact calculations for sequential t, χ^2 and F tests. *Biometrika* **78** 133–141.

KAO, S. C. and LAI, T. L. (1980). Sequential selection procedures based on confidence sequences for normal populations. *Comm. Statist. A—Theory Methods* **9** 1657–1676.

MILLER, R. G., JR. (1970). Sequential rank tests—one sample case. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 97–108. Univ. California Press, Berkeley.

O'BRIEN, P. C. and FLEMING, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics* **35** 549–556.

PAULSON, E. (1962). A sequential procedure for comparing several experimental categories with a standard or control. *Ann. Math. Statist.* **33** 438–443.

PAULSON, E. (1964). A sequential procedure for selecting the population with the largest mean from k normal populations. *Ann. Math. Statist.* **35** 174–180.

RAO, C. R. (1950). Sequential tests of null hypotheses. *Sankyā Ser. A* **10** 361–370.

SAMUEL-CAHN, E. (1974). Repeated significance tests II, for hypotheses about the normal distribution. *Comm. Statist. A—Theory Methods* **3** 711–733.

SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York.

SIEGMUND, D. (1986). Boundary crossing probabilities and statistical applications. *Ann. Statist.* **14** 361–404.

SIEGMUND, D. (1993). A sequential clinical trial for comparing three treatments. *Ann. Statist.* **21** 464–483.

WALD, A. (1947). *Sequential Analysis*. Wiley, New York.

WOODROOFE, M. (1978). Large deviations of the likelihood ratio statistic with applications to sequential testing. *Ann. Statist.* **6** 72–84.

WOODROOFE, M. (1986). Very weak expansions for sequential confidence levels. *Ann. Statist.* **14** 1049–1067.

WOODROOFE, M. (1989). Very weak expansions for sequentially designed experiments: linear models. *Ann. Statist.* **17** 1087–1102.

WOODROOFE, M. and COAD, D. S. (1995). Corrected confidence sets for sequentially designed experiments. Unpublished manuscript.

DEPARTMENT OF BIOSTATISTICS
HARVARD SCHOOL OF PUBLIC HEALTH
677 HUNTINGTON AVENUE
BOSTON, MASSACHUSETTS 02115