Statistica Sinica **9**(1999), 923-937

ON THE BIAS OF ESTIMATION OF A BROWNIAN MOTION DRIFT FOLLOWING GROUP SEQUENTIAL TESTS

Zhengqing Li and David L. DeMets

State University of New York at Albany and University of Wisconsin

Abstract: Group sequential tests have been widely used to control the type I error rate at a prespecified level in comparative clinical trials. It is well known that due to the optional sampling effect, conventional maximum likelihood estimates will exaggerate the treatment difference, and hence a bias is introduced. We consider a group sequentially monitored Brownian motion process. An analytical expression of the bias of the maximum likelihood estimate for the Brownian motion drift is derived based on the alpha spending method of Lan and DeMets (1983). Through this formula, the bias can be evaluated exactly by numerical integration. We study how the Brownian motion drift and various alpha spending functions and interim analysis patterns affect the bias. A bias adjusted estimator is described and its properties are investigated. The behavior of this estimator is studied for differing situations.

Key words and phrases: Alpha spending function, interim analysis, maximum likelihood, robustness, stopping time.

1. Introduction

Group sequential methods for interim monitoring of clinical trials were introduced by Pocock (1977) based on the work on repeated significance tests by Armitage, McPherson and Rowe (1969). The basic idea is to adjust the critical values used at the interim tests of the null hypothesis such that the overall type I error rate is controlled at a prespecified level. Various group sequential strategies have been proposed, including those of Pocock (1977) and O'Brien-Fleming (1979). Two requirements of these methods were equal number of patients between interim analyses, and that the maximum number of interim analyses be specified in advance. In order to avoid these conditions, Lan and DeMets (1983) proposed a more flexible approach referred to as the alpha spending method. The amount of type I error probability spent is a nondecreasing function of information fraction and thus induces a corresponding boundary.

Following the extensive studies of these group sequential tests, there is now a general interest in studying point and interval estimates of treatment difference.

It is well known that when a clinical trial stops early to reject the null hypothesis of treatment equivalence, conventional maximum likelihood estimates will usually overestimate the treatment difference (see for instance Siegmund (1978) and Whitehead (1986)). The reason is that the sampling distributions of these estimates are affected by the planned sequential or group sequential designs although the estimates are not altered. If the observed treatment difference is randomly greater than the true effect, the chances of stopping are higher than if the observed difference is randomly less than the true effect. Bias is the difference between the average observed effect over the sampling space and the true effect. Whitehead (1986) evaluated the magnitude of the bias of maximum likelihood estimation approximately following the sequential probability ratio test and the triangular test, and proposed a bias-adjusted estimate. Hughes and Pocock (1988) did a simulation study based on an estimate of the risk ratio in a typical post-myocardial infarction trial to examine the nature of this bias for various group sequential plans. A Bayesian method, referred to as a shrunken estimate, was proposed to assess the true treatment effect based on interim results (Hughes and Pocock (1988), Pocock and Hughes (1989)). Emerson and Fleming (1990) studied various point and interval estimates for a normal mean. Based on their results, it appeared that the bias-adjusted estimate by Whitehead had the lowest mean squared error among all the estimators considered.

It is well known that the alpha spending method based on the Brownian motion process by Lan and DeMets (1983) provided a very general framework for sequential monitoring of clinical trials (DeMets and Lan (1994), Lan and Zucker (1992)). Asymptotically, many sequentially computed test statistics in survival analyses (Sellke and Siegmund (1983)) as well as in the longitudinal studies (Gange and DeMets (1996)) can be approximated by Brownian motion processes (Lan and Zucker (1992)). It is the goal of this paper to study the bias and the bias-adjusted estimate based on this general framework. Compared to the simulation method used by Hughes and Pocock (1988) and the bias approximation by Whitehead (1986), our bias evaluation and the bias-adjusted estimate are based on exact analytical expressions. Another significance of this study is that the calculation of the bias and the bias-adjusted estimate can be easily incorporated into an existing software (Reboussin, DeMets, Kim and Lan (1995)) which has been used widely in group sequential analyses of clinical trials. Pinheiro and DeMets (1995) also studied this bias for a Gaussian independent increment structure. However, their methods of evaluating bias are different. Their approach was more focused on the simulation method. The variance and the mean squared error of the bias-adjusted estimate were only briefly considered.

In this paper, group sequential methods are described in Section 2, with an emphasis on the alpha spending function approach. Section 3 gives an analytical

expression for the bias of the maximum likelihood estimate of the Brownian motion drift, which can be evaluated through numerical integration. We also compare the bias curves for various alpha spending functions and interim analysis patterns. The properties of a bias-adjusted estimator and its robustness are investigated in Section 4.

2. Group Sequential Monitoring and the Alpha Spending Method

In a clinical trial comparing two treatments, we want to repeatedly test the null hypothesis H_0: there is no treatment difference in the accumulated data, keeping the type I error to a prespecified level. Let $Z(k)$ be the test statistic based on the accumulated data at the kth interim analysis, and let M be the maximum number of interim analyses. If there is a set of probabilities π_k $(k = 1, \ldots, M)$ such that $\sum \pi_k = \alpha$, then the boundary values $\{c_k\}$ can be evaluated through numerical integration (Armitage, McPherson and Rowe (1969)) such that

$$pr\{|Z(1)| < c_1, \ldots, |Z(k-1)| < c_{k-1}, |Z(k)| > c_k | H_0\} = \pi_k. \tag{1}$$

Here the joint distribution (or asymptotic distribution) of $\{Z(1), \ldots, Z(k)\}$ needs to be known, and for many test statistics it is a multivariate normal. At each interim analysis, either the trial is stopped and H_0 is rejected if $|Z(k)| > c_k$ for some k, or otherwise the trial is continued and more data are collected. If $|Z(k)| < c_k$ for all k, then H_0 is not rejected.

The alpha spending function allocates the fixed type I error probability across the course of the clinical trial as a function the information fraction. Let $i(k)$ represent the amount of information available at the kth interim analysis, and let I represent the total information available at the end of the trial. The information fraction at the kth interim analysis is defined as $t_k = i(k)/I$. Lan and DeMets (1983) then specified a nondecreasing function $\alpha^*(t), t \in [0, 1]$ such that $\alpha^*(0) = 0$ and $\alpha^*(1) = \alpha$. The amount of type I error to spend at the kth analysis is $\pi_k = \alpha^*(t_k) - \alpha^*(t_{k-1})$. The boundary values c_k, $k = 1, \ldots, M$ can be determined successively by (1). The alpha spending function of Lan and DeMets (1983) has been applied to various settings such as repeated measurement data (Lee and DeMets (1991)), ordinal categorical data (Gange (1994)) and survival data (Tsiatis, Boucher and Kim (1995)).

Now consider a Brownian motion process $W(t)$, $t \in [0, 1]$, with drift μ and unit variance. We want to test the null hypothesis $H_0 : \mu = 0$. The time scale t for $W(t)$ represents the information fraction after appropriate rescaling (Lan and Zucker (1993)). Suppose that the interim analyses are conducted at times with information fractions $\{t_1, \ldots, t_M\}$, then the two-sided symmetric boundaries c_k, $k = 1, \ldots, M$, for $Z(k) = W(t_k)/(t_k)^{\frac{1}{2}}$ can be calculated using an existing FORTRAN program very rapidly: see the University of Wisconsin-Madison

Biostatistics Technical Report 95 by Reboussin, DeMets, Kim and Lan. The stopping time for the process is

$$T = \min\left\{t_i : \left|W(t_i)/(t_i)^{\frac{1}{2}}\right| > c_i, 1 \le i \le M\right\}, \tag{2}$$

or $T = t_M$ if $\left|W(t_i)/(t_i)^{\frac{1}{2}}\right| \le c_i$ for all $1 \le i \le M$. Note that the stopping time T in (2) depends on the choice of information fractions $\{t_1, \ldots, t_M\}$. To simplify discussions, we assume that the information fractions are fixed in advance. However, in the actual monitoring process, to stop or not depends only on the current and past information fractions.

3. Bias in Estimating the Brownian Motion Drift

Interim analyses in many clinical trials can be approximated by Brownian motion processes with a linear drift. The value of the drift is determined by the treatment effect. In this section, we evaluate the estimation bias of the drift and study its reduction methods for a group sequentially monitored Brownian motion process. In Section 3.1, an exact bias evaluation formula is given, which is applicable to any group sequential procedure which can be approximated by a Brownian motion process. Section 3.2 studies the properties of the bias and compares the bias curves for various group sequential boundaries and interim analysis patterns.

3.1. An analytical expression

Consider the group sequentially monitored Brownian motion process $W(t)$ with drift μ, unit variance and the stopping rule in (2). The maximum likelihood estimate of μ is $\hat{\mu} = W(T)/T$. We shall consider the expectation of $\hat{\mu}$ under the true drift μ

$$E(\hat{\mu}) = E\{W(T)/T\} = \sum_{i=1}^{M} E\{W(T)/T | T = t_i\} pr(T = t_i), \tag{3}$$

where $E\{W(T)/T|T = t_i\}$ denotes the conditional expectation of $W(T)/T$ given that $T = t_i$.

Let $\{c_1, \ldots, c_M\}$ be the symmetric boundaries for $W(T)/(T)^{\frac{1}{2}}$. Define $c_i^* = (t_i)^{\frac{1}{2}} c_i - \mu t_i$ and $b_i^* = -(t_i)^{\frac{1}{2}} c_i - \mu t_i$. Then $\{b_1^*, \ldots, b_M^*\}$ and $\{c_1^*, \ldots, c_M^*\}$ are the upper and lower boundaries for $W(T) - \mu T$. Let $\sigma_i = (t_i - t_{i-1})^{\frac{1}{2}}$ and $\sigma_1 = (t_1)^{\frac{1}{2}}$. Let ϕ denote the standard normal density function. Through recursive integration we define the conditional density functions in a similar manner to Armitage, McPherson and Rowe (1969) as

$$f_1(u) = \phi(u/\sigma_1)\sigma_1^{-1} \tag{4}$$

and

$$f_i(u) = \int_{c^*_{i-1}}^{b^*_{i-1}} \sigma_i^{-1}\phi(u/\sigma_i - x/\sigma_i) f_{i-1}(x)dx. \quad (5)$$

The recursive density functions describe the probability of $(W(T) - \mu T, T)$. Let P_i denote the probability of stopping at or before the ith interim analysis. It can be shown (see Appendix) that the overall bias of $\hat{\mu}$ following a group sequential test can be expressed as

$$\begin{aligned} b(\mu) &= \sum_{i=1}^{M} E\{W(T)/T | T = t_i\}(P_i - P_{i-1}) - \mu \\ &= \sum_{i=1}^{M-1} t_i^{-1}\Big\{\int_{b_i^*}^{\infty} u f_i(u)du + \int_{-\infty}^{c_i^*} u f_i(u)du\Big\} + t_M^{-1}\int_{-\infty}^{\infty} u f_M(u)du. \quad (6) \end{aligned}$$

Let Φ be the cumulative distribution function for a standard normal variable. Define

$$R_i(x) = 1 - \Phi\Big(\frac{b_i^* - x}{\sigma_i}\Big) + \Phi\Big(\frac{c_i^* - x}{\sigma_i}\Big), \text{ and } Q_i(x) = \phi\Big(\frac{b_i^* - x}{\sigma_i}\Big) - \phi\Big(\frac{c_i^* - x}{\sigma_i}\Big).$$

Using Fubini's theorem, the bias can be expressed as

$$b(\mu) = \sum_{i=1}^{M-1} \frac{1}{t_i}\int_{c^*_{i-1}}^{b^*_{i-1}} \{\sigma_i Q_i(x) + x R_i(x)\} f_{i-1}(x)dx + \frac{1}{t_M}\int_{c^*_{M-1}}^{b^*_{M-1}} x f_{M-1}(x)dx, \quad (7)$$

where $c_0^* = b_0^* = 0$ and $f_0(x) = 1$. Computation of $b(\mu)$ at the first analysis involves only the standard normal density, but for the second analysis and beyond, numerical integration is necessary. For example, it can be done using the Newton-Cote methods similar to those described by Armitage, McPherson and Rowe (1969). A FORTRAN program is available from the authors to carry out the computation.

3.2. A numerical comparison

For symmetric boundaries, we have two conclusions immediately. (a) When $\mu = 0$, the bias is zero; if $\mu = 0$, then $c_i^* = -b_i^*$, and $f_i(-u) = f_i(u)$ recursively. Therefore $b(0) = 0$ follows from the fact that the integrands in (6) are odd functions. (b) $b(\mu) = 0$ for a fixed sample size design. If data are reviewed only at t_M, then only the last term in (6) contributes bias, which is zero since $f_M(u) = f_1(u)$. The interpretation of (a) is clear. If the treatment difference is very small, most trials will continue to the last planned analysis and the optional sampling effect is weakened. Thus the group sequential procedure is very close to a fixed sample size design, and the bias is very small.

In comparative clinical trials, the treatment difference is often measured by a sequence of statistics $(Z_1, Z_2, \ldots)$ which are independent normal variables with unknown mean η and unit variance. Let N be the maximum sample size and $S_k = Z_1 + \cdots + Z_k$. We have the following analogy between the partial sum S_n and the Brownian motion process $W(t)$ (Kim and DeMets (1987)):

$$S_{n_k} \sim N(n_k\eta, n_k), \quad W(t_k) \sim N(t_k\mu, t_k),$$

where n_k is the accumulated sample size by the kth interim analysis and t_k is the information fraction at the kth interim analysis. The maximum likelihood estimate of η is $\hat{\eta} = S_{n_k}/n_k$. Suppose that t_k's are proportional to the observed sample size, i.e. $t_k = n_k/N$, then it follows from the arguments of Kim and DeMets (1987) that $\hat{\mu}/\hat{\eta} = (n_k/t_k)^{\frac{1}{2}}$ and $N = n_k/t_k = (\mu/\eta)^2$. Let $\tilde{b}(\eta) = E(\hat{\eta}) - \eta$, then we have

$$\tilde{b}(\eta) = b(\mu)/(N)^{\frac{1}{2}} = b(N^{\frac{1}{2}}\eta)/N^{\frac{1}{2}}, \text{ and } \tilde{b}(\eta)/\eta = b(\mu)/\mu.$$

That is, the proportion of the bias to the true parameter is common no matter which measure, μ or η, is used.

To see how the bias is affected by different boundaries and interim analysis patterns, we compare the bias curves for two-sided symmetric tests at $\alpha = 0.05$ and the following three alpha spending functions:

$$\alpha_1^*(t) = 2\{1-\Phi(z_{\alpha/2}/(t)^{\frac{1}{2}})\}, \quad \alpha_2^*(t) = 2\Big[\frac{1}{2}\alpha\log\{1+(e-1)t\}\Big], \quad \alpha_3^*(t) = 2(\frac{1}{2}\alpha t).$$

Here α_1^* and α_2^* are known to generate boundaries similar to the O'Brien-Fleming and the Pocock boundaries respectively at $\alpha = 0.05$, and α_3^* was previously studied by Lan and DeMets (1983) and Kim and DeMets (1987). The most conservative spending function is α_1^*, followed by α_3^* and α_2^*.

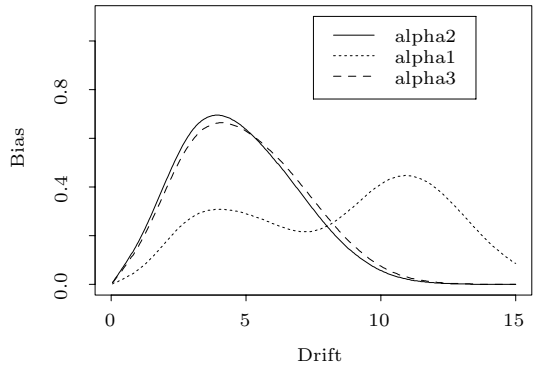

Figure 1. Bias curves for different alpha spending functions at the equal interval analysis $t^{(1)} = \{0.2, 0.4, 0.6, 0.8, 1.0\}$.

Figure 1 plots the bias versus the Brownian motion drift for these three alpha spending functions at equal interval analysis $t^{(1)} = \{0.2, 0.4, 0.6, 0.8, 1.0\}$. One significant feature observed from Figure 1 is that the bias curves are unimodal for α_2^* and α_3^*, while the curve for α_1^* is multi-modal. This occurs because the boundary values for α_1^* (O'Brien-Fleming) are very conservative at the first interim analysis and exhibit a large difference at different analyses. When the drift μ is small the bias is small, since in this situation the design is close to a fixed sample size design. As μ starts to increase, the stopping time exhibits a large variability and the bias increases rapidly. As μ goes to 5 and beyond, the majority of the trials using α_2^* and α_3^* tend to stop at the first interim analysis and the bias starts to decrease. For trials using α_1^*, the bias decreases a little (the first maximum) since most of the trials tend to stop at the second interim analysis. However, the bias starts to increase as μ increases to 8 and beyond. As μ increases to 11, the majority of the trials tend to stop at the first interim analysis. Thus the bias decreases again to produce the second maximum. This result is consistent with the simulation study by Hughes, Freedman and Pocock (1992), as well as that by Pinheiro and DeMets (1995). Another feature observed from the bias curves is that for small or moderate treatment effects, the conservative boundaries based on α_1^* (O'Brien-Fleming) can protect against severe bias while the boundaries based on α_2^* (Pocock) result in the largest bias among the three alpha spending functions. This is consistently true for other interim analysis patterns. Furthermore, an order relation in bias for α^* exists for small or moderate μ. That is, the more convex α^* is, the smaller the bias is. Actually, the more convex α^* is, the more likely a trial tends to stop late. Trials with a small chance of stopping early will produce relatively less bias. These results are also true for the later analysis pattern $t^{(2)} = \{0.3, 0.6, 0.8, 0.9, 1.0\}$ and the early analysis pattern $t^{(3)} = \{0.1, 0.2, 0.3, 0.6, 1.0\}$.

Figure 2 compares the bias for three interim analysis patterns based on α_2^* and α_1^* respectively. Note that for the Pocock-type boundary, the bias for the equal interval analysis is consistently smaller than that for the early analysis, and greater than that for the later analysis. However, for the O'Brien-Fleming-type boundary, such an order relation does not exist and the biases are very close for the three interim analysis patterns when the treatment effects are not large. This is due to the fact that the O'Brien-Fleming-type boundary is extremely large at the early analyses. For a small treatment effect and extremely large boundaries, the probability of stopping early will be very small, and hence the bias produced is small. However, for a large treatment effect, reviewing data more frequently at later stages reduces the bias substantially.

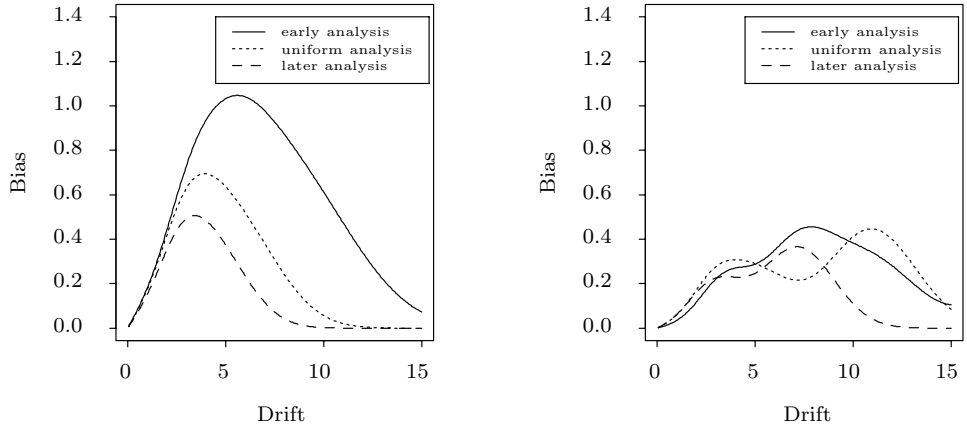

Figure 2. Bias curves for different interim analysis patterns. Figure (a) is based on the Pocock-type boundary and (b) is based on the O'Brien-Fleming-type boundary.

4. Bias Reduction

In this section, we first give a bias reduction method which is independent of group sequential tests. Some analytical results concerning this bias reduction method are given. In Section 4.2, simulation studies show that this method is efficient in reducing the bias while keeping a relatively small variance. The robustness of this method is studied for various group sequential boundaries and interim analysis patterns.

4.1. The bias adjusted estimate and its properties

Let $\hat{\mu}$ be the estimate of μ based on the data. The expected value of $\hat{\mu}$ is given by $E(\hat{\mu}) = \mu + b(\mu)$ where $b(\mu)$ is the bias at μ. The ideal bias-corrected estimate of μ would be $\hat{\mu} - b(\mu)$. In practice, a corrected estimate of μ, say μ^*, can be given by a solution to the following equation (Whitehead (1986))

$$\mu^* = \hat{\mu} - b(\mu^*), \tag{8}$$

where $b(\mu^*)$ is the bias evaluated at the adjusted estimate μ^*.

Let $b^*(\mu) = E(\mu^*) - \mu$ denote the bias for the adjusted estimate μ^*. Thus $b^*(\mu) = b(\mu) - E\{b(\mu^*)\}$, and the amount of bias reduced by μ^* is $E\{b(\mu^*)\}$. If μ^* is very close to μ, $E\{b(\mu^*)\}$ will be close to $b(\mu)$, and hence $b^*(\mu)$ is much smaller than $b(\mu)$. However, in general μ^* is not unbiased although its bias is very small.

Consider the second order Taylor expansion:

$$b(\mu^*) = b(\mu) + (\mu^* - \mu)b'(\mu) + \frac{1}{2}(\mu^* - \mu)^2 b''(\xi),$$

where ξ is between μ^* and μ. Then

$$E\{b(\mu^*)\} = b(\mu) + b^*(\mu)b'(\mu) + \frac{1}{2}E\{b''(\xi)(\mu^* - \mu)^2\}.$$

Since $b^*(\mu) = b(\mu) - E\{b(\mu^*)\}$,

$$b^*(\mu) = -\frac{1}{2}\frac{E\{b''(\xi)(\mu^* - \mu)^2\}}{1 + b'(\mu)}. \tag{9}$$

As μ^* is very close to μ, we have the approximation

$$b^*(\mu) \approx -\frac{1}{2}\frac{b''(\mu)}{1 + b'(\mu)}\,\mathrm{MSE}\,(\mu^*). \tag{10}$$

Therefore, whether the bias of μ^* is positive or negative depends upon the convexity of the bias curve around μ. μ^* underestimates μ if $b(\mu)$ is convex around μ and overestimates μ if $b(\mu)$ is concave.

Whitehead (1986) studied the large-sample variance of the bias-adjusted estimate μ^*, and found

$$\mathrm{Var}\,(\mu^*) \le \frac{\mathrm{Var}\,(\hat{\mu})}{\{1 + b'(\mu)\}^2}. \tag{11}$$

A slight modification can show that

$$\mathrm{MSE}\,(\mu^*) \approx \frac{\mathrm{Var}\,(\hat{\mu})}{\{1 + b'(\mu)\}^2}. \tag{12}$$

Therefore, if $b'(\mu) \ge 0$, $\mathrm{MSE}\,(\mu^*) \le \mathrm{MSE}\,(\hat{\mu})$ and $\mathrm{Var}\,(\mu^*) \le \mathrm{Var}\,(\hat{\mu})$. For the group sequential designs considered in this paper, $b'(\mu)$ is always positive for small or moderate treatment effects, as indicated in Figure 1.

In practice, we can solve equation (8) using the Newton-Raphson method. The initial value $\mu_0^* = \hat{\mu}$ could be used, and usually the first iteration

$$\mu_1^* = \hat{\mu} - \frac{b(\hat{\mu})}{1 + b'(\hat{\mu})} \tag{13}$$

will give a good approximation. Notice that we need to evaluate $b'(\mu)$ in order to solve for μ^*. Define $S_i(x) = b_i^*\phi\{(b_i^* - x)/\sigma_i\} - c_i^*\phi\{(c_i^* - x)/\sigma_i\}$. Using

the relation $b'(\mu) = E\{W^2(T)/T - \mu T\} - 1$ (Whitehead (1986)), one has the expression

$$b'(\mu)+1 = \sum_{i=1}^{M-1}\frac{1}{t_i}\int_{c_{i-1}^*}^{b_{i-1}^*}[(x^2+\sigma_i^2-xt_i)R_i+\sigma_i(x-t_i)Q_i+\sigma_i S_i(x)]f_{i-1}(x)dx + \frac{1}{t_M}\int_{c_{M-1}^*}^{b_{M-1}^*}(x^2+\sigma_M^2-xt_M)f_{M-1}(x)dx.$$

This expression can be evaluated numerically in the same manner as that for $b(\mu)$.

4.2. Robustness of the bias-adjusted estimate

In order to investigate the performance of the adjusted estimate μ^* in (8), a simulation study (simulation size=5000) is conducted based on three alpha spending functions for three interim analysis patterns $t^{(1)}$, $t^{(2)}$ and $t^{(3)}$ respectively. The group sequential boundaries are the two-sided symmetric boundaries at $\alpha = 0.05$, and the alternative is the Brownian motion drift at which the test attains 90% power. The estimated biases, variances and mean squared errors of the two estimators $\hat{\mu}$, μ^* are given in Tables 1, 2 and 3. The estimated $b'(\mu)$ is also given for each value of μ.

Table 1. Comparison of the maximum likelihood estimate and the bias-adjusted estimate based on the boundaries for the uniform analysis $t^{(1)} = \{0.2, 0.4, 0.6, 0.8, 1.0\}$. The value of μ is chosen such that the test attains 90% power.

Boundary function	True μ	$b'(\mu)$	Estimator	Bias	Variance	MSE	Bias percent
α_1^*	3.28	0.048	$\hat{\mu}$	0.289	1.605	1.689	8.82
			μ^*	0.035	1.505	1.506	1.08
α_2^*	3.54	0.056	$\hat{\mu}$	0.677	2.620	3.078	19.11
			μ^*	0.103	2.620	2.631	2.90
α_3^*	3.46	0.078	$\hat{\mu}$	0.644	2.635	3.050	18.62
			μ^*	0.101	2.540	2.550	2.92

The standard error of the estimated bias is less than 0.025.

Note that the bias-adjusted estimate reduces the bias of the maximum likelihood estimate a great deal. For the three alpha spending functions, the bias reduced by the adjusted estimator ranges from 84.3%, 81.4% and 88.6% to 88.4%, 95.3% and 94.0% for $t^{(1)}$, $t^{(2)}$ and $t^{(3)}$, respectively. For the conservative boundaries such as those of the O'Brien-Fleming type, the bias is reduced more effectively than for the Pocock-type boundary. Furthermore, the variance of the

bias-adjusted estimator is generally smaller than that of the maximum likelihood estimate. Consequently, the mean squared error is consistently reduced for all boundaries and interim analysis patterns.

Table 2. Comparison of the maximum likelihood estimate and the bias-adjusted estimate based on the boundaries for the later analysis $t^{(2)} = \{0.3, 0.6, 0.8, 0.9, 1.0\}$. The value of μ is chosen such that the test attains 90% power.

Boundary function	True μ	$b'(\mu)$	Estimator	Bias	Variance	MSE	Bias percent
α_1^*	3.30	0.008	$\hat{\mu}$	0.224	1.490	1.540	6.79
			μ^*	0.011	1.395	1.395	0.32
α_2^*	3.55	0.020	$\hat{\mu}$	0.499	1.945	2.194	14.06
			μ^*	0.083	2.070	2.077	2.35
α_3^*	3.47	0.010	$\hat{\mu}$	0.495	2.000	2.245	14.27
			μ^*	0.092	2.055	2.063	2.65

The standard error of the estimated bias is less than 0.025.

Table 3. Comparison of the maximum likelihood estimate and the bias-adjusted estimate based on the boundaries for the early analysis $t^{(3)} = \{0.1, 0.2, 0.3, 0.6, 1.0\}$. The value of μ is chosen such that the test attains 90% power.

Boundary function	True μ	$b'(\mu)$	Estimator	Bias	Variance	MSE	Bias percent
α_1^*	3.25	0.068	$\hat{\mu}$	0.231	1.480	1.534	7.11
			μ^*	0.014	1.310	1.310	0.43
α_2^*	3.49	0.207	$\hat{\mu}$	0.839	4.090	4.794	24.05
			μ^*	0.096	3.530	3.539	2.74
α_3^*	3.45	0.208	$\hat{\mu}$	0.751	3.850	4.414	21.77
			μ^*	0.076	3.230	3.236	2.21

The standard error of the estimated bias is less than 0.030.

We also studied the bias-adjusted estimator for various values of the Brownian motion drift for the two-sided symmetric O'Brien-Fleming-type boundary for the uniform analysis pattern. The simulation results for 10 values of μ are given in Table 4. Since the proportion of the bias to the true parameter is invariant in the scale of treatment difference, both $|b(\mu)|/\mu$ and $|b^*(\mu)|/\mu$ are listed in Table 4. Note that the bias-adjusted estimate reduces the bias consistently for all values of μ. Two important features can be observed by looking at $b^*(\mu)$. First, $b^*(\mu)$ is negative as $\mu \in A^- = \{0.8, 1.6, 5.6, 6.4, 7.2, 8.0\}$ and positive as $\mu \in A^+ = \{2.4, 3.2, 4.0, 4.8\}$. It can be seen roughly from Figure 1 that $b(\mu)$ is convex at $\mu \in A^-$ and concave at $\mu \in A^+$. Secondly, for $\mu = 1.6$ and 5.6, the

bias $b^*(\mu) = E(\mu^*) - \mu$ is close to zero. This is because $b''(\mu)$ is almost zero at these two points. Therefore the simulation results are consistent with the results given by (9) and (10).

Table 4. Comparison of the maximum likelihood estimate and the bias-adjusted estimate for various values of drift based on the O'Brien-Fleming-type boundary for the uniform analysis $t^{(1)} = \{0.2, 0.4, 0.6, 0.8, 1.0\}$.

μ	$b'(\mu)$	$b(\mu)$	Var $(\hat{\mu})$	$\lvert b(\mu)\rvert/\mu$ (%)	$b^*(\mu)$	Var (μ^*)	$\lvert b^*(\mu)\rvert/\mu$ (%)
0.8	0.073	0.044	1.224	4.97	-0.015	1.055	1.86
1.6	0.118	0.122	1.399	7.14	-0.007	1.192	0.46
2.4	0.113	0.219	1.559	8.79	0.020	1.365	0.83
3.2	0.055	0.288	1.634	8.91	0.042	1.519	1.30
4.0	0.001	0.308	1.730	7.54	0.032	1.705	0.79
4.8	-0.028	0.296	1.857	6.03	0.017	1.887	0.35
5.6	-0.043	0.267	2.035	4.65	-0.005	2.069	0.09
6.4	-0.038	0.233	2.356	3.65	-0.027	2.335	0.42
7.2	0.000	0.216	2.770	2.30	-0.049	2.665	0.68
8.0	0.057	0.238	3.277	3.65	-0.038	3.082	0.47

The standard error of the estimated bias is less than 0.030.

We also observed that the variance of the estimates $\hat{\mu}$ and μ^* increased with the drift μ. This is due to the fact that the probability of stopping early tends to increase as the treatment effect increases. Less information will increase the variation of the estimates. Furthermore, μ^* exhibits smaller variability than $\hat{\mu}$ does when $b'(\mu) > 0$. For the values of μ with $b'(\mu) < 0$ ($\mu = 4.8$, 5.6, 6.4), the variance of μ^* is slightly larger than that of $\hat{\mu}$. This confirmed the results in (11) and (12). Fortunately, it rarely happens in practice that a clinical trial has a drift greater than 5.

5. Discussion

It has been shown that the bias of the maximum likelihood estimate for the Brownian motion drift is a function of the group sequential boundaries, information fractions and the true Brownian motion drift. Therefore, for a given group sequentially monitored clinical trial with the structure of a Brownian motion process, the bias can be estimated based on the estimates of the information fractions and the Brownian motion drift. However, the calculation of the bias-adjusted estimate does not depend on the unknown Brownian motion drift. The bias calculation method described needs numerical integration although alternatively one can calculate it by simulation (Pinheiro and DeMets (1995)). The advantage of using our analytical expression is that it is very easy to incorporate the calculation of the bias and bias-adjusted estimate into an existing Fortran

program of Reboussin, DeMets, Kim and Lan (1995). We only considered the bias evaluation for two-sided symmetric boundaries. For one-sided boundaries (DeMets and Ware (1980)) or two-sided asymmetric boundaries (DeMets and Ware (1982)), the bias can be evaluated similarly. Our study for the Brownian motion process is based on the time scale of the information fraction. If the process is monitored in another scale, such as information, the bias can still be calculated after a simple transformation.

For practical interest, our study was more focused on small or moderate treatment effects. A Brownian motion drift greater than 5 represents a large treatment difference and rarely happens in practice. For example, the designs considered in Section 4.2 attain 90% power at the Brownian motion drift around 3.5. Our study indicated that the conservative boundaries such as the O'Brien-Fleming-type, with a more frequently later analysis pattern, could protect from possible severe bias caused by the maximum likelihood estimate for small or moderate treatment effects. This might provide some useful guidance in the design and monitoring of a clinical trial. Furthermore, our study indicated that the bias-adjusted estimator not only reduced the bias effectively, but also reduced the variance in a realistic range of treatment effects.

After a sequential clinical trial, one may also want an interval estimate in addition to a point estimate. The usual methodology for interval estimates is based on orderings of the sampling space (Tsiatis, Rosner and Mehta (1984), Chang (1989), Emerson and Fleming (1990)). However, as pointed out by Todd, Whitehead and Facey (1996), the ordering approach does not result in an estimator with a simple sampling distribution, it is difficult to remove the bias of point estimates, to combine trial results with those of other trials, and to deduce estimates and confidence intervals for functions of parameters. After modifying the results by Woodroofe (1992), Todd, Whitehead and Facey (1996) gave an approximate confidence interval based on the large sample approximation, and suggested that the bias-adjusted estimate could be presented with Woodroofe's form of confidence interval. However, this confidence interval is based on the maximum likelihood estimate instead of the bias-adjusted estimate. As an issue for future research, it is desirable to construct a confidence interval based on the bias-adjusted estimate.

Appendix

The derivation of the bias formula: It can be shown that f_i describes the probability of values of $W(T) - \mu T$ and the probability of stopping at t_i. Let

$$1 - P_i = \int_{c_i^*}^{b_i^*} f_i(x)dx, \quad Q_i = \int_{b_i^*}^{\infty} f_i(x)dx, \quad \text{and} \quad R_i = \int_{-\infty}^{c_i^*} f_i(x)dx.$$

Then $1 - P_i$ is the probability of not stopping at or before the ith analysis; Q_i is the probability of the event $W(t_i)/(t_i)^{\frac{1}{2}} \in [c_i, \infty]$; R_i is the probability of the event $W(t_i)/(t_i)^{\frac{1}{2}} \in [-\infty, -c_i]$; $Q_i + R_i$ is the probability of stopping at the ith analysis. Notice that

$$P_i - P_{i-1} = Q_i + R_i, \quad \text{and} \quad \sum_{j=1}^{i}(Q_j + R_j) = P_i.$$

For $i \leq M - 1$, it follows from (4) and (5) that the bias, given that the trial has stopped at t_i, is

$$\begin{aligned} E(\hat{\mu}|T = t_i) - \mu &= \frac{E\{W(T)/T; T = t_i\}}{P(T = t_i)} - \mu \\ &= t_i^{-1}(P_i - P_{i-1})^{-1}E\{W(T); T = t_i\} - \mu \\ &= t_i^{-1}(P_i - P_{i-1})^{-1}E\{W(T) - \mu T; T = t_i\} \\ &= t_i^{-1}(P_i - P_{i-1})^{-1}\Big\{\int_{b_i^*}^{\infty} u f_i(u)du + \int_{-\infty}^{c_i^*} u f_i(u)du\Big\}, \end{aligned}$$

where P_i is defined above and $P_0 = 0$. At the last analysis, the trial is also terminated even if $W(T)/(T)^{\frac{1}{2}} \in [-c_M, c_M]$. Thus by similar argument we have

$$E[W(T)/T|T = t_M] - \mu = t_M^{-1}(1 - P_{M-1})^{-1}\int_{-\infty}^{\infty} u f_M(u)du.$$

Consequently, the bias formula (6) follows from (3).

References

Armitage, P., McPherson, C. K. and Rowe, B. C. (1969). Repeated significance tests on accumulating data. *J. Roy. Statist. Soc. Ser. A* **132**, 235-244.

Chang, M. N. (1989). Confidence intervals for a normal mean following a group sequential test. *Biometrics* **45**, 247-254.

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.

Emerson, S. and Fleming, T. R. (1990). Parameter estimation following group sequential hypothesis testing. *Biometrika* **77**, 875-892.

Gange, S. J. (1994). Multivariate ordinal responses in clinical trials. Ph.D. thesis, University of Wisconsin-Madison.

Hughes, M. D., Freedman, L. and Pocock, S. J. (1992). The impact of stopping rules on hetero-geneity of results in overview of clinical trials. *Biometrics* **48**, 41-53.

Hughes, M. D. and Pocock, S. J. (1988). Stopping rules and estimation problems in clinical trials. *Statistics in Medicine* **7**, 1231-1242.

Kim, K. and DeMets, D. L. (1987). Design and analysis of group sequential tests based on the type I error spending rate function. *Biometrika* **74**, 149-154.

Lan, K. K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70**, 659-663.

Lan, K. K. G. and Zucker, D. (1993). Sequential monitoring of clinical trials: the role of information in Brownian motion. *Statistics in Medicine* **12**, 753-765.

Lee, J. W. and DeMets, D. L. (1991). Sequential comparison of change with repeated measurement data. *J. Amer. Statist. Assoc.* **86**, 757-762.

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

Pinheiro, J. C. and DeMets, D. L. (1995). Estimating and reducing bias in group sequential designs with Gaussian independent increment structure. Technical Report 96, Department of Biostatistics, University of Wisconsin-Madison.

Pocock, S. J. (1977). Group sequential methods in the design and analyses of clinical trials. *Biometrika* **64**, 191-199.

Pocock, S. J. and Hughes, M. D. (1989). Practical problems in interim analyses with particular regard to estimation. *Controlled Clinical Trials* **10**, 209s-221s.

Reboussin, D. M., DeMets, D. L., Kim, K and Lan, K. K. G. (1995). Programs for computing group sequential bounds using the Lan-DeMets method. Technical Report 95, Department of Biostatistics, University of Wisconsin-Madison.

Siegmund, D. (1978). Estimation following sequential tests. *Biometrika* **65**, 341-349.

Todd, S., Whitehead, J. and Facey, K. M. (1996). Point and interval estimation following a sequential clinical trial. *Biometrika* **83**, 453-461.

Tsiatis, A. A., Boucher, H. and Kim, K. (1995). Sequential methods for parametric survival models. *Biometrika* **82**, 165-173.

Tsiatis, A. A., Rosner, G. L. and Mehta, C. R. (1984). Exact confidence intervals following a group sequential tests. *Biometrics* **40**, 797-803.

Whitehead, J. (1986). On the bias of maximum likelihood estimation following a sequential test. *Biometrika* **73**, 573-581.

Woodroofe, M. (1992). Estimation after sequential testing: a simple approach for a truncated sequential probability ratio test. *Biometrika* **79**, 347-353.

Department of Biometry and Statistics, One University Place, Rensselaer, NY 12144, U.S.A.

E-mail: ZQL01@health.state.ny.us.

Department of Biostatistics, 600 Highland Avenue K6/446 CSC, University of Wisconsin, Madison, WI 53792, U.S.A.

E-mail: demets@biostat.wisc.edu

(Received April 1997; accepted July 1998)