ESTIMATING AND REDUCING BIAS IN GROUP SEQUENTIAL DESIGNS WITH GAUSSIAN INDEPENDENT INCREMENT STRUCTURE

José C. Pinheiro, PhD
David L. DeMets, PhD
Department of Biostatistics

MADISON, WISCONSIN
Estimating and Reducing Bias in Group Sequential Designs with Gaussian Independent Increment Structure

José C. Pinheiro and David L. DeMets

Department of Biostatistics, University of Wisconsin – Madison

Group sequential methods are frequently used in data monitoring of clinical trials to detect early therapeutic benefit or unexpected toxicity that might lead to early termination of the study. A question that has raised some concern in the use of sequential testing procedures is the bias associated with the estimates of treatment differences. It is known that clinical trials that stop early due to evidence of therapeutic benefit are prone to exaggerate the magnitude of the treatment effect.

We consider methods for estimating and reducing the bias of treatment differences estimators in group sequential designs with Gaussian independent increment structure. We derive an analytical expression for the bias and give an easy-to-calculate approximate bound for its variation. A simulation estimate of the bias, based on a Gaussian independent increment structure, is also described and a related bias reduced estimator is considered.

1 Introduction

Group sequential methods are frequently used in data monitoring of clinical trials to detect early therapeutic benefit or unexpected toxicity that might lead to early termination of the study (DeMets, 1987). Different methods have been proposed to take into account the effect of repeatedly testing the data on the overall significance level of the hypothesis test for treatment effect (Pocock, 1977; O'Brien and Fleming, 1979; Lan and DeMets, 1983). All of these methods are based on the idea of
adjusting the significance levels (or, equivalently, the critical values) of the individual analyses, so that the overall probability of detecting a significant effect, under the null hypothesis of no treatment difference, is kept at a pre-specified level $\alpha$. The most flexible of these methods, which includes the other methods as particular cases, is the *alpha spending function*, proposed by Lan and DeMets (1983).

Even though the alpha spending function is of general applicability, its most common use (for which computer programs for calculating the sequential boundaries are available) is when the sequence of test statistics for the interim analyses has a *Gaussian independent increment structure (GIIS)* (DeMets and Lan, 1994), characterized by

$$\hat{\theta}(t_i) \overset{D}{=} \frac{\sigma B(t_i)}{t_i}, \ i = 1, \ldots, K$$

(1)

where $\hat{\theta}(t_i)$ is the test statistic corresponding to the $i$th analysis, $B(t)$ is a Brownian motion with $\text{Var}[B(1)] = 1$, $\sigma^2$ is the variance of $\hat{\theta}(t_K)$, $t_i$ is the fraction of the total information available at the $i$th interim analysis (i.e. $t_i = \sigma^2 / \text{Var}[\hat{\theta}(t_i)]$) also called the *information fraction* (Lan, Reboussin and DeMets, 1994), and $K$ is the total number of analyses planned for the study. The symbol $\overset{D}{=} \text{ denotes equality in distribution. It has been shown that the GIIS holds, at least approximately, for most test statistics used in applications of group sequential methods. Among others, the GIIS occurs in comparison of means, proportions, logrank statistitics, and slopes of linear mixed effects models (Tsiatis, 1982; Lan et al., 1994; Tsiatis, Boucher and Kim, 1995).

A question that has raised some concern in the use of sequential testing procedures is the bias associated with the estimators of treatment differences (Hughes and Pocock, 1988 and Hughes, Freeman, and Pocock 1992). It is known that clinical trials that stop early due to evidence of therapeutic benefit are prone to exaggerate the magnitude of the treatment difference (Hughes, Freedman and Pocock, 1992). In this paper we consider methods for estimating and reducing the bias of estimators of treatment differences in group sequential designs for which the GIIS assumption is valid. Section 2 includes methods for estimating the bias, both analytically and by simulation. The relationship between the bias and the frequency pattern of the sequential analysis is studied in Section 3. In section 4 we consider a method for reducing the bias in group sequential designs estimation, based on the results presented in Whitehead (1986). Our conclusions are included in section 5.
2 Estimating the bias

In this section we consider the estimation of the bias associated with estimators of treatment differences in group sequential designs. Two approaches are described: the first one provides an analytical formula for the bias and an easy-to-calculate approximate bound for its variation, based on a second-order Taylor expansion, and the second one uses simulation, based on the GIIS assumption.

2.1 Analytical expression for the bias

Let \( B(t) \) represent a Brownian motion with drift parameter \( \mu \) and consider the \( K \) interim analyses information fractions \( t = \{t_1, t_2, \ldots, t_{K-1}, t_K \} \). Let \( c_j = \{c_j(t_1), c_j(t_2), \ldots, c_j(t_{K-1}), c_j(t_K) \} \), \( j = 1, 2 \) denote the lower and upper critical values for the interim analyses, determined by the alpha spending function, and define \( \ell = \inf \{i \in \{1, 2, \ldots, K\} \mid B(t_i) \notin [c_1(t_i), c_2(t_i)] \} \), with the convention that \( \ell = K \), if \( B(t_i) \in [c_1(t_i), c_2(t_i)], i = 1, \ldots, K \). Now define \( \tau = t_\ell \), the information fraction at the time the trial is stopped. Since \( \tau \) is a bounded stopping time, it follows from Wald’s equations (Siegmund, 1985) that \( E[B(\tau)] = \mu E(\tau) \).

We want to evaluate \( E_\theta [\hat{\theta}(\tau)] \), the expected value of the test statistic at the time the sequential testing is stopped, when the true treatment difference is equal to \( \theta \). We assume here that \( \hat{\theta}(t_i) \) is an unbiased estimator of \( \theta \), which is equivalent to assuming that \( \theta = \sigma \mu \).

In the case of symmetric boundaries, we have that \( \hat{\theta}(\tau) \) is unbiased under the null hypothesis of no treatment effect. That is, if \( c_1 = -c_2 \),

\[
E_0 \left[ \hat{\theta}(\tau) \right] = \sigma E_0 \left[ \frac{B(\tau)}{\tau} \right] = \sigma E_0 \left\{ E_0 \left[ B(\tau) \mid \ell \right] / \tau \right\} \quad \text{and}
\]

\[
E_0 \left[ B(\tau) \mid \ell = i \right] = E_0 \left[ B(t_i) \mid \ell = i \right] = E_0 \left[ B(t_i) ; \ell = i \right] / P_0(\ell = i) = 0
\]

where the last equality follows from the fact that the event \( \{\ell = i\} \) defines a symmetric region in \( R^1 \), under symmetric boundaries and GIIS.

We now consider the general case of asymmetric boundaries, and/or significant treatment differences. We use a modified version of a general result presented in Whitehead (1986). Let \( g \) be a real function defined over \( R^2 \), such that \( E_\mu |g[B(\tau), \tau]| < \infty, \quad \forall \mu \in R \), then under the GIIS
assumption it follows that

\[
\frac{\partial}{\partial \mu} E_\mu \{ g \left[ B(\tau), \tau \right] \} = E_\mu \{ [B(\tau) - \mu \tau] g \left[ B(\tau), \tau \right] \}
\]

(3)

The proof is included in the Appendix.

It follows from \( \theta = \sigma \mu \) and (3) that

\[
\frac{\partial}{\partial \theta} E_\theta \{ g \left[ B(\tau), \tau \right] \} = \left(1/\sigma\right) E_\theta \{ [B(\tau) - (\theta \tau) / \sigma] g \left[ B(\tau), \tau \right] \}
\]

(4)

By setting \( g(x, y) = y^{-1} \) and noting that \( E_\theta (\tau^{-1}) \leq t_i^{-1} < \infty, \forall \theta \in R \), it follows as a corollary of (4) that, under GIIS,

\[
\text{Bias}_\theta \left[ \hat{\theta}(\tau) \right] = E_\theta \left[ \hat{\theta}(\tau) \right] - \theta = \sigma E_\theta \{ [B(\tau) - (\theta \tau) / \sigma] / \tau \}
\]

\[
= \sigma^2 \frac{\partial}{\partial \theta} E_\theta (\tau^{-1}) = \sigma^2 \sum_{i=1}^{K-1} \left( t_i^{-1} - 1 \right) \frac{\partial}{\partial \theta} \pi_i (\theta)
\]

where \( \pi_i (\theta) = P_\theta (\tau = t_i) \) denotes the \( i \)th exit probability for the sequential tests.

In order to use (5), one only needs estimates for the information fractions \( (t_1, \ldots, t_{K-1}) \) and the variance of the test statistic at the final analysis \( \sigma^2 \). These are usually available at the planning stage of a clinical trial, so that (5) can be used to evaluate potential bias problems prior to the beginning of the study. Updated estimates of \( t_1, \ldots, t_{K-1} \) and \( \sigma^2 \) can be obtained at the time the trial is stopped and used in (5) to produce a bias reduced estimate, as described in Section 4. Computer packages available for calculating the group sequential boundaries (Reboussin, DeMets, Lan and Kim, 1992) also provide the exit probabilities for given values of \( \theta \) as a by product. These programs can be used in conjunction with numerical derivatives and/or interpolation splines to calculate the bias according to (5). Alternatively, simulation methods based on the GIIS assumption may be employed to estimate the bias, as described in Section 2.2.

Formula (5) gives interesting insights on the relationship between the bias and the true treatment difference \( \theta \). For any reasonable two-sided sequential boundaries, \( E_\theta (\tau^{-1}) \) is a monotonically increasing function of \( |\theta| \), as the trial should end earlier when treatment effects, either good or bad, are present. Therefore, we have that \( \text{sign} \left\{ \text{Bias}_\theta \left[ \hat{\theta}(\tau) \right] \right\} = \text{sign} (\theta) \) and the sequential testing
is prone to exaggerate treatment effects, as noticed before by other authors (Hughes et al., 1992). Since \( E_\theta (\tau^{-1}) \) attains a minimum at \( \theta = 0 \), its derivative is zero at this point and \( \text{Bias}_0 \{ \hat{\theta} (\tau) \} = 0 \), as shown in (2). Typically, for small values of \(|\theta|\), the trial will proceed until the last analysis, as \(|\theta|\) increases the trial will tend to stop at earlier analyses, and \( P_\theta (\tau = t_1) \uparrow 1 \) when \(|\theta| \uparrow \infty \). The bias will be larger for those values of \( \theta \) such that the exit probabilities change more rapidly, but it will converge to zero as \(|\theta|\) becomes large. In the case of one-sided boundaries, \( E_\theta (\tau^{-1}) \) will increase monotonically with \(|\theta|\) for \( \theta \) in the alternative hypothesis and \( \text{sign} \{ \text{Bias}_\theta \{ \hat{\theta} (\tau) \} \} = \text{sign} (\theta) \) in that region of the parameter space.

Using a second order Taylor expansion of \( E \{ \hat{\theta} (\tau) \} = E \{ \sigma B(\tau)/\tau \} \) around \( E \{ B(\tau) \} \) and \( E(\tau) \) it follows

\[
\text{Bias}_\theta \{ \hat{\theta} (\tau) \} = E_\theta \{ \hat{\theta} (\tau) \} - \theta \approx \frac{\theta \text{Var} (\tau) - \sigma \text{Cov} [B(\tau), \tau]}{E^2(\tau)}
\]

This approximate expression for the bias is hard to use in practice, since \( \text{Cov} [B(\tau), \tau] \) cannot be easily estimated from the data. Using the facts that \([B(t) - (\theta/\sigma)t]^2 - t\) is a martingale (Durrett, 1990, p. 358) and \( \tau \) is a bounded stopping time, it follows from the Cauchy-Schwarz inequality that

\[
|\text{Cov} [B(\tau) - (\theta/\sigma)\tau, \tau]| \leq \{ \text{Var} [B(\tau) - (\theta/\sigma)\tau] \text{Var}(\tau) \}^{1/2} = [E(\tau) \text{Var}(\tau)]^{1/2}
\]

and an easy-to-calculate approximate upper bound for the bias variation can be obtained as follows

\[
|\text{Bias}_\theta \{ \hat{\theta} (\tau) \}| \approx \left| \frac{\theta \text{Var} (\tau) - \sigma \text{Cov} [B(\tau), \tau]}{E^2(\tau)} \right| = \left| \frac{\sigma \text{Cov} [B(\tau) - (\theta/\sigma)\tau, \tau]}{E^2(\tau)} \right| \leq \left[ \frac{\sigma^2 \text{Var}(\tau)}{E^3(\tau)} \right]^{1/2}
\]

Since preliminary estimates of \( \sigma^2 \) and the information fractions \( t_1, \ldots, t_{K-1} \) are usually available at the time the clinical trial is being designed, the distribution of \( \tau \) can be estimated (under GIIS) and the bound in (7) calculated.

To illustrate the use of the exact formula for the bias in (5) and the upper bound in (7), we consider an artificial example of a clinical trial, with four planned analyses, in which the treatment effect is estimated by the mean difference between treatment and placebo patients for a normally
distributed response variable, with one observation per patient. We assume that twenty-five patients on each group enter the study between interim analyses and that the variance of the response variable is equal to two. We estimated the true bias and the approximate upper bound for the bias through Monte Carlo simulation with 100,000 replications. The true bias was obtained by fitting an interpolation B-spline (DeBoor, 1978) to the Monte Carlo estimates of $E_{\theta}(\tau^{-1})$ and calculating its first-order derivative. We used the sequential boundaries proposed by O'Brien and Fleming (1979) and Pocock (1977), with overall significance level of 0.05. Figure 1 presents the simulation results, including the Monte Carlo bias (i.e. $\sum_{i=1}^{100,000} \hat{\theta}_{i} (\tau) / 100,000 - \theta$). We see that the upper bound (7)

![Graph](image)

**Figure 1:** Exact bias and approximate upper bound for the bias, calculated by Monte Carlo simulation using O’Brien-Fleming’s and Pocock’s boundaries. Circles represent the Monte Carlo bias.

approximates the variation in the bias reasonably well, especially for the O’Brien-Fleming boundaries. The bias corresponding to the Pocock boundaries is larger than the one corresponding to the O’Brien-Fleming boundaries for smaller treatment differences ($\theta \leq 1.4$), but the situation is reversed for very large treatment differences. This behavior is associated with the level of difficulty for early trial termination, which is easier when Pocock’s boundaries are used, and will be discussed in more detail in the next section. The bias curves in Figure 1 are similar to those presented in Hughes et al. (1992, Figure 1).
2.2 Estimating the bias through simulation

The bias of $\hat{\theta}(\tau)$ can also be assessed through simulation, if estimates of $\sigma^2$ and of the information fractions $t_1, \ldots, t_{K-1}$ are available. This is usually true in group sequential designs, as discussed in Section 2.1.

The motivation behind the simulation approach for the bias estimation is that, under the GIIS assumption, the covariance structure of $\tilde{\Theta} = [\tilde{\theta}(t_1), \ldots, \tilde{\theta}(t_K)]^T$ is fairly simple and depends only upon $\sigma^2$ and the information fractions. Letting $\Sigma$ represent the variance-covariance matrix of $\tilde{\Theta}$, we have

$$\Sigma_{ij} = \text{Cov} [\tilde{\theta}(t_i), \tilde{\theta}(t_j)] = \sigma^2 \frac{\min(t_i, t_j)}{t_it_j} \quad (8)$$

The simulation assessment of the bias is done as follows.

1. For a given value of $\theta$, generate $\tilde{\theta}_1^*, \ldots, \tilde{\theta}_N^*$ independently and according to a $N_K(\theta, 1, \Sigma)$ distribution, for sufficiently large number of simulations $N$ (in the example described below, $N = 10,000$ was sufficient).

2. Obtain $\tilde{\theta}_1^*(\tau), \ldots, \tilde{\theta}_N^*(\tau)$ by applying the sequential testing procedure to the random samples generated in the first step.

3. Estimate the bias of $\hat{\theta}(\tau)$ at $\theta$ as

$$\hat{\text{Bias}}_{\theta}[\tilde{\theta}(\tau)] = \frac{\sum_{i=1}^N \tilde{\theta}_i^*(\tau)}{N} - \theta \quad (9)$$

$\hat{\text{Bias}}_{\theta}[\tilde{\theta}(\tau)]$ can be regarded as an approximate parametric bootstrap estimator of the bias (Efron and Tibshirani, 1993). The numerical algorithm used to implement the simulation bias estimation is simple and can be programmed to run very efficiently: a version of the algorithm we wrote in C takes no more than 2 seconds to obtain $\hat{\text{Bias}}_{\theta}[\tilde{\theta}(\tau)]$ for $K = 5$ and $N = 10,000$, when executed on a Sun SPARC 20 workstation. Other interesting features of the algorithm is that it can be used for any group sequential design in which the underlying test statistic can be assumed to have, at least approximately, GIIS. During the simulation, we can also store additional information about the estimation procedure, such as the confidence levels for the usual confidence intervals and the power of the underlying test.
We illustrate the use of this methodology with a Monte Carlo simulation study based on an example described in Smith, Sempo, Smith and Gilligan (1989) of a study of calcium supplement effects on bone density. In this study, 37 postmenopausal women received calcium and 37 other were assigned a placebo. Ten or eleven observations were taken on each woman, during a period of 5 years. The primary question of the study was to test whether the calcium supplement decreased the rate of decay of bone density. Lee and DeMets (1992) used this example to illustrate the use of sequential rank tests with repeated measures data.

We simplify the problem here and assume that eleven observations were made on each woman, and that these were taken at times 0.3, 0.5, 0.7, 1, 1.3, 1.7, 2, 2.8, 3.2, 4, and 4.5 years for all women. Following Lee and DeMets (1992), we assume that five analyses were planned at times 1, 2, 3, 4, and 5 years. We use the following linear mixed effects model (Laird and Ware, 1982) to represent the jth measurement on the ith patient.

\[ y_{ij} = \beta_0 + [\beta_{1C} I_i + \beta_{1T} (1 - I_i) + b_i] \text{time}_j + \epsilon_{ij} \]  

(10)

where \( \beta_0 \) represent the (common) intercept, \( \beta_{1C} \) represents the control group slope, \( \beta_{1T} \) represents the treatment group slope, \( I_i \) is an indicator variable taking value one when the ith patient belongs to the control group and zero otherwise, \( b_i \) represents the random effect associated with the slope, \( \text{time}_j \) represents the time of the jth measurement, and \( \epsilon_{ij} \) represents the within-patient error term. The \( b_i \) are assumed to be independent and identically distributed with \( \mathcal{N}(0, \sigma_b^2) \) distribution, while the \( \epsilon_{ij} \) are assumed to be independent and identically distributed with \( \mathcal{N}(0, \sigma_r^2) \) distribution and independent of the \( b_i \). In this model, the treatment difference is defined as \( \theta = \beta_{1T} - \beta_{1C} \).

In the Monte Carlo simulation study we use \( \beta_0 = 68, \beta_{1C} = -3, \sigma_b^2 = 0.09, \sigma_r^2 = 4 \), and vary \( \theta \) between 0 and 3.5, using steps of size 0.05. For each \( \theta \), 2,000 samples were generated according to model (10) and the maximum likelihood estimates (Lindstrom and Bates, 1988) of the parameters in the model were obtained. The Monte Carlo estimate of the bias is simply the average of the 2,000 estimated treatment differences minus the true treatment difference.

We use the alpha spending function method of Lan and DeMets (1983) to derive the critical boundaries for the sequential tests. Four different alpha spending functions are used:
1. \( \alpha_1^*(t) = 2 \left[ 1 - \Phi \left( \frac{z_{\alpha/2}}{\sqrt{t}} \right) \right] \), where \( \Phi \) denotes the c.d.f. of a \( N(0, 1) \) random variable, \( z_{\alpha/2} \) denotes the \( (1 - \alpha/2) \) percentile of the \( N(0, 1) \) distribution, and \( \alpha \) denotes the overall significance level. This spending function gives boundaries similar to those proposed by O'Brien and Fleming (1979):

2. \( \alpha_2^*(t) = \alpha t^4 \);

3. \( \alpha_3^*(t) = \alpha t^2 \);

4. \( \alpha_4^*(t) = \alpha \log [1 + (e - 1) t] \). This spending function gives boundaries similar to those proposed by Pocock (1977).

Figure 2 presents the symmetric two-sided critical boundaries corresponding to each of the \( \alpha^* \) functions above, for \( \alpha = 0.05 \). There is a decreasing level of difficulty for early trial termination when we go from \( \alpha_1^* \) to \( \alpha_4^* \). The alpha spending functions considered here cover most practical applications of this method to group sequential designs.

![Critical boundaries for alpha spending functions used in the Monte Carlo simulation.](image)

Figure 2: Critical boundaries for alpha spending functions used in the Monte Carlo simulation.

We also estimated the bias of the treatment difference estimator using simulation based on the GIIS, as described previously. The information fractions corresponding to (10) are equal to 0.067, 0.320, 0.490, 0.843, and 1, and \( \sigma^2 = \text{Var}[\hat{\theta}(t_5)] = 0.0065 \).
Figure 3 presents the Monte Carlo estimates of the bias and the GIIS simulation estimates of the bias, for the spending functions $\alpha^*_1$ through $\alpha^*_4$. The bias curve corresponding to the GIIS simulation shows the same overall pattern as the Monte Carlo bias curve, but the two tend to be out of phase for larger values of $\theta$, especially for $\alpha^*_1$. We note that the standard deviation of $\hat{\theta}(t_i)$ varies between 0.08 ($i = 5$) and 0.31 ($i = 1$), so that values of $\theta$ greater than 1 correspond to considerably large treatment differences (relative to the variability in the estimators), which are less likely to be encountered in practice. For $\theta \leq 1$ there is a strong agreement between the Monte Carlo and the GIIS simulation results.

![Graph showing Monte Carlo and GIIS simulation bias curves.]

Figure 3: Monte Carlo and GIIS simulation estimated bias curves.

The discrepancy between the Monte Carlo and the GIIS simulation bias estimates becomes more pronounced for values of $\theta$ such that the sequential tests have increasing probability of stopping at the first interim analysis. Figure 4 illustrates this fact for the $\alpha^*_2$ spending function. When the sequential procedure is stopped at the first interim analysis, only four observations are available on each patient. The assumption of GIIS structure for the treatment differences estimator in the linear mixed model holds only asymptotically (Lan et al., 1994) and therefore the smaller the sample
size, the worse the GIIS approximation is. Fortunately, when the trial is stopped at the first interim analysis, it is likely that the treatment effect, or toxicity, is very large, so that the bias becomes a less important concern.

![Figure 4: Relationship between Monte Carlo and GIIS simulation estimates of the bias and the average stopping time in the Monte Carlo simulation for $\alpha_2^*$ spending function.](image)

Another interesting issue that can be observed in Figure 3 is how the different spending functions compare in terms of bias. We see that, as we go from the hardest ($\alpha_1^*$) to the easiest ($\alpha_4^*$) level of early trial termination, the tallest bias peak moves from right to left. As mentioned before, this tallest peak is related to the change in termination from the second to the first interim analysis. This change will occur for smaller values of $\theta$ as we decrease the difficulty of early termination. Therefore, if bias is a major concern in the analysis, more conservative spending functions such as $\alpha_1^*$ or $\alpha_2^*$ should be preferred.

3 Bias under different analysis patterns

An important issue related to the bias in group sequential designs is how it is affected by the frequency pattern of the interim analyses. Frequently, a uniform pattern is used, meaning that the successive analyses occur at equally spaced information fractions. In some instances, one may look more frequently at earlier times, especially if there is early indication of treatment benefit or toxicity. Other times, one may look more frequently later into the study, as when there is no early
indication of treatment effect or toxicity.

In this section we investigate the bias resulting from these early, uniform, and late analysis patterns, using the results of Section 2 and two examples of group sequential designs.

First we consider an artificial example of a clinical trial in which five analyses are planned and the treatment effect is estimated as the mean difference between placebo and treatment arms, for a normally distributed response variable, with one observation per patient. The projected patient accrual is equal to 100 patients per treatment arm. The information fractions assumed for the early, uniform, and late analysis patterns are given in Table 1.

Table 1: Information fractions under early, uniform, and late analysis patterns.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Information Fractions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>0.1 0.2 0.3 0.6</td>
</tr>
<tr>
<td>Uniform</td>
<td>0.2 0.4 0.6 0.8</td>
</tr>
<tr>
<td>Late</td>
<td>0.4 0.7 0.8 0.9</td>
</tr>
</tbody>
</table>

The bias of the treatment difference estimator for each analysis pattern was estimated through Monte Carlo simulation with 100,000 replicates, using the results of Section 2. We used the O'Brien–Fleming and the Pocock sequential boundaries, with overall significance level of 0.05. Figure 5 presents the resulting bias curves.

Figure 5: Bias curves for early, uniform, and late analysis patterns.
The late analysis pattern is the one that offers the greatest protection against bias. Its bias curve peaks at smaller values of the true treatment difference $\theta$, but its maximum bias is considerably smaller than the maximum bias of the other two strategies. The early analysis pattern has the worst performance in terms of bias, especially for the Pocock boundaries. The uniform analysis pattern offers an intermediate protection against bias. As noted before in Section 2.1, the Pocock boundaries give less protection against bias than the O'Brien-Fleming boundaries for smaller values of $\theta$. Also, the maximum bias under the Pocock boundaries is larger than the maximum bias under the O'Brien-Fleming boundaries, for all three analysis patterns.

We consider now the calcium supplement study example of Section 2.2, under different analysis patterns. We will restrict ourselves here to the alpha spending functions $\alpha^*_1$ and $\alpha^*_3$, which give boundaries similar to the O'Brien-Fleming and the Pocock boundaries, respectively.

When planning the interim analyses for a repeated measures study, like the calcium supplement study, one usually uses the fraction of the elapsed calendar time (with respect to duration of the study) or the fraction of measurements observed as a proxy for the information fraction (Lan et al., 1994). We use the calendar time approach here when defining the different analysis patterns, but also calculate the corresponding true information fractions. It can be shown that the information fraction corresponding to time $t_j$ in model (10) is given by

$$t_j = \left[ \frac{\sum_{k=1}^{j} \text{time}_k^2}{1 + (\sigma_1^2/\sigma_2^2) \sum_{k=1}^{j} \text{time}_k^2} \right] / \left[ \frac{\sum_{k=1}^{11} \text{time}_k^2}{1 + (\sigma_1^2/\sigma_2^2) \sum_{k=1}^{11} \text{time}_k^2} \right] \quad (11)$$

Table 2 presents the elapsed calendar time fractions and the information fractions assumed for the early, uniform, and late analysis patterns in the calcium supplement example. We note that the information fraction is in general smaller than the elapsed calendar time fraction.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Time Fractions</th>
<th>Information Fractions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>0.1 0.2 0.3 0.4</td>
<td>0.013 0.067 0.124 0.320</td>
</tr>
<tr>
<td>Uniform</td>
<td>0.2 0.4 0.6 0.8</td>
<td>0.067 0.320 0.491 0.843</td>
</tr>
<tr>
<td>Late</td>
<td>0.5 0.6 0.7 0.8</td>
<td>0.320 0.491 0.659 0.843</td>
</tr>
</tbody>
</table>

The bias of the treatment difference estimator was calculated through Monte Carlo simulation.
with 2,000 replicates, as described in Section 2.2. Figure 6 presents the corresponding bias curves.

![Bias curves for early, uniform, and late analysis patterns in the calcium supplement example.](image)

Figure 6: Bias curves for early, uniform, and late analysis patterns in the calcium supplement example.

Again we conclude that the late analysis pattern is the one that provides greater protection against bias. The maximum bias under the late analysis pattern is about 40% of the maximum bias under the uniform analysis pattern and 16% of the maximum bias under the early analysis pattern. The O'Brien–Fleming boundaries offer greater protection against bias for smaller values of \( \theta \), under all three analysis patterns.

### 4 Reducing the bias

The simulation approach to the bias estimation described in section 2.2 can also be used to reduce the bias in group sequential designs with GIIS. Let \( g(\theta) = E_\theta[\hat{\theta}(\tau)] \), then, provided we have estimates for \( \sigma^2 \) and the information fractions \( t_1, \ldots, t_{K-1} \), we can use the GIIS simulation to estimate \( g \) (or equivalently, \( g(\theta) - \theta \), the bias). Let \( \hat{g} \) represent the corresponding estimate of \( g \). We consider the following bias reduced (BR) estimator for \( \theta \).

\[
\hat{\theta}_{BR}(\tau) = \hat{g}^{-1}\left[\hat{\theta}(\tau)\right]
\]  

(12)
The bias reduction formula in (12) assumes that \( \tilde{g} \) is invertible. As discussed in Section 2.1, the bias is a monotonically increasing function of \( \theta \) for any reasonable two-sided boundaries and for any one-sided boundaries. Hence, \( \tilde{g} \) will be invertible in all cases of practical interest.

The bias reduced estimator (12) is equivalent to the \( \tilde{\theta} \) estimator proposed by Whitehead (1986, Section 5), which uses the fact that \( \theta = g(\theta) - \text{Bias}_\theta \left( \tilde{\theta}(\tau) \right) \).

The rationale behind the bias reduced estimator \( \tilde{\theta}_{BR}(\tau) \) is that

\[
\tilde{\theta}_{BR}(\tau) = \tilde{\theta}(\tau) - \left\{ \tilde{g} \left[ \tilde{\theta}_{BR}(\tau) \right] - \tilde{\theta}_{BR}(\tau) \right\}
\]

which is equal to the initial estimate \( \tilde{\theta}(\tau) \) minus the estimated bias associated with \( \tilde{\theta}_{BR}(\tau) \).

The expression in (13) also provides an algorithm for estimating \( \tilde{\theta}_{BR}(\tau) \). Letting \( \tilde{\theta}_0 = \tilde{\theta}(\tau) \), we use the recursive relation

\[
\tilde{\theta}_{i+1} = \tilde{\theta}(\tau) - \left[ \tilde{g} \left( \tilde{\theta}_i \right) - \tilde{\theta}_i \right] \quad \text{(14)}
\]

and iterate on \( i \) until convergence. In the examples that we have analyzed convergence rarely occurred in more than seven iterations. Alternatively, the Newton-Raphson algorithm may be used, giving the iterative relation

\[
\tilde{\theta}_{i+1} = \tilde{\theta}_i + \frac{\tilde{\theta}(\tau) - \tilde{g} \left( \tilde{\theta}_i \right)}{\tilde{g}' \left( \tilde{\theta}_i \right)} \quad \text{(15)}
\]

This approach will generally converge faster than (14), but it has the disadvantage of requiring the estimation of \( \tilde{g}' \left( \tilde{\theta}_i \right) \), which will generally have to be done numerically.

To compare the computational performance of the two numerical methods for obtaining \( \tilde{\theta}_{BR}(\tau) \), we calculated their user’s times until convergence for 100 simulated values of \( \tilde{\theta}(\tau) \), using a Sun SPARC 20 workstation. A group sequential design with five planned analyses, uniform analysis pattern, and \( \sigma^2 = 0.2 \) was used in the simulation. A total of 10,000 Monte Carlo replicates were used in each calculation of \( \tilde{g} \). Both the O’Brien–Fleming and the Pocock boundaries were considered in the simulation. The simpler recursive method in general converged faster to the desired solution than the Newton-Raphson method. The median user’s times until convergence for the recursive algorithm were 1.70 and 1.71 seconds respectively for the O’Brien–Fleming and the Pocock boundaries, while for the Newton-Raphson algorithm these were respectively 3.42 and 3.43 seconds. Nevertheless, the differences in performance are of the order of a few seconds and should
not be noticed in practical applications, unless one is planning to use the bias reduced estimator in a simulation study.

We applied the bias reduction technique described in (12) to the simulated data from the calcium supplement study of section 2.2. Figure 7 presents the Monte Carlo estimates of the bias curves for $\hat{\theta}(\tau)$ and $\hat{\theta}_{BR}(\tau)$.

![Graph showing bias curves for $\hat{\theta}(\tau)$ and $\hat{\theta}_{BR}(\tau)$]

Figure 7: Monte Carlo simulation estimated bias curves for $\hat{\theta}(\tau)$ and $\hat{\theta}_{BR}(\tau)$.

There is a substantial reduction in the bias of $\hat{\theta}_{BR}(\tau)$, especially for $\theta \leq 1$. This coincides with the region where the GIIIS approximation for the variance-covariance structure of $\hat{\theta}(\tau)$ is more accurate (c.f. Section 2.2). This is also the region where the bias is more of a concern, since for larger treatment differences the relative magnitude of the bias becomes smaller.

Confidence intervals for $\theta$ based on $\hat{\theta}_{BR}(\tau)$ can be derived from the usual confidence intervals based on $\hat{\theta}(\tau)$, by applying $\hat{g}^{-1}$ to the endpoints of the interval. Alternatively, an approximation to the variance of $\hat{\theta}_{BR}(\tau)$ can be derived using the delta method (Efron and Tibshirani, 1993, p. 314),
as below,

\[
\text{Var} \left[ \hat{\theta}_{BR}(\tau) \right] \approx \text{Var} \left[ \hat{\theta}(\tau) \right] \left[ \frac{d g^{-1}(\theta)}{d \theta} \right]_{g(\theta)}^2 = \text{Var} \left[ \hat{\theta}(\tau) \right] \left[ g'(\theta) \right]^{-2} \tag{16}
\]

and confidence intervals based on the normal distribution can be obtained. An estimate of (16) is given by \((\hat{\sigma}^2 / \tau) \left[ \hat{g}' \left( \hat{\theta}_{BR}(\tau) \right) \right]^{-2} \).

5 Conclusions

Estimated treatment differences following a group sequential clinical trial are prone to exaggerate the treatment benefit, or toxicity. The bias tends to increase, in absolute value, with the variability in the trial stopping time, being more pronounced at early termination trials. Since the stopping pattern of a group sequential design is heavily dependent upon the choice of boundaries, so will be the bias. The easier the early termination of a trial is, the smaller the treatment differences at which the maximum bias will occur. Therefore, if bias is a major concern in a clinical trial, more conservative sequential boundaries should be chosen.

When the sequence of estimators used in a clinical trial has, at least approximately, a Gaussian independent increment structure, estimates and bounds for the bias can be derived before hand. In this case, bias reduced estimates of the treatment difference can be obtained and will give more conservative estimates of treatment benefit and toxicity.

The frequency pattern of the interim analysis in a clinical trial has a major impact on the bias of the treatment difference estimator. Having the interim analyses later into the trial gives the best protection against bias, while looking at the data more frequently at the early stages of the trial may lead to severe bias.

An issue that needs more investigation in the area of group sequential methods is the impact of the stopping rule on variance estimation. This aspect of the problem was not considered here, but certainly deserves more investigation. Another interesting topic is the estimation and reduction of bias when the GIIS assumption is not valid. This would be particularly useful for early termination trials, when the GIIS may no longer provide a good approximation to the variance-covariance structure of the estimators.
References


Appendix

We include here the proof of result (3) of Section 2.1. First we note that, by the GIIS assumption,

\[ E_\mu \{ g [B (\tau) , \tau] \} = \sum_{i=1}^{K} E_\mu \{ g [B (\tau) , \tau] ; \ell = i \} = \sum_{i=1}^{K} \int_{\{t=i\}} g (x, t_i) f_i (x) \, dx \]

where \( f_i (x) = \exp \left[ - \frac{(x - \mu t_i)^2}{2 t_i} \right] / (2\pi t_i)^{1/2} \). Thus \( \partial f_i / \partial \mu (x) = (x - \mu t_i) f_i (x) \). It then follows from an application of the dominated convergence theorem to exponential families of distribution (Lehmann, 1986, p. 59) that

\[
\frac{\partial}{\partial \mu} \sum_{i=1}^{K} \int_{\{t=i\}} g (x, t_i) f_i (x) \, dx = \sum_{i=1}^{K} \int_{\{t=i\}} g (x, t_i) \frac{\partial f_i (x)}{\partial \mu} \, dx \\
= \sum_{i=1}^{K} \int_{\{t=i\}} (x - \mu t_i) g (x, t_i) f_i (x) \, dx = \sum_{i=1}^{K} E_\mu \{ [B (\tau) - \mu \tau] g [B (\tau) , \tau] ; \ell = i \} \\
= E_\mu \{ [B (\tau) - \mu \tau] g [B (\tau) , \tau] \}
\]
as we wanted to show.