
UNIVERSITY OF WISCONSIN
DEPARTMENT OF BIostatISTICS

**Technical
Report #123**

September 1997

**BIAS CORRECTION IN GROUP SEQUENTIAL ANALYSIS
WITH CORRELATED DATA**

**Roger P. Qu, PhD
David L. DeMets, Ph.D.
Department of Biostatistics**

MADISON, WISCONSIN

Bias Correction in Group Sequential Analysis With Correlated Data

Roger P. Qu and David L. DeMets

Hoechst Marion Roussel, Inc. and University of Wisconsin

Abstract: This paper focuses on bias of group sequential estimate of treatment effect for correlated data using generalized estimating equation (GEE) method and Lan and DeMets alpha-spending function. Linear and logistic regressions are used to examine (a) the magnitude of the bias of sequential estimate with correlated data; (b) the influence of the true correlation structure on bias; and (c) bias under different monitoring patterns. Finally, a bias-corrected sequential estimate is proposed using a Brownian motion approximation and numerical simulation. Logistic regression is used to illustrate and to assess the performance of the proposed method.

Key words and phrases: Alpha-spending function; Brownian motion; Correlation structure; Interim analysis; GEE; Linear regression; Logistic regression; Sequential boundaries; Simulation; Working correlation.

1. Introduction

Introduced by Pocock in 1977, group sequential analysis has increasingly become a standard method in clinical trials. Driven by the philosophy that a clinical trial should be terminated once a clinically relevant treatment effect becomes convincing, group sequential analysis has been applied to interim monitoring for a variety of statistical endpoints including continuous, binary, and survival data (DeMets and Lan, 1994). The basic idea in group sequential analysis is to adjust the critical values for the interim analyses such that the overall significance level is controlled at a pre-assigned level. Different interim monitoring schemes have been proposed. Pocock (1977) and O'Brien and Fleming (1979) methods require a pre-assigned

maximum number of analyses and an equal number of observations between analyses. In 1983, Lan and DeMets introduced the alpha-spending function method which allocates the total pre-assigned overall significance level α through a pre-specified monotonic function that does not require fixed number of pre-specified interim analyses or equal increments.

Most applications of group sequential methods in the past have been for independent observations. With recent development of repeated measurements analysis, group sequential methodology has been applied to correlated data. Lee and DeMets (1991) and Wu and Lan (1992) considered sequential comparison of a treatment effect for linear mixed effect models (Laird and Ware, 1982) using a multivariate integration technique; Wei, Su and Lachin (1990) applied group sequential analysis to correlated data using the generalized estimating equation (GEE) method (Liang and Zeger, 1986) with Slud and Wei's (1982) Type I error spending; and Gange and DeMets (1996) applied the Lan and DeMets' alpha-spending function to GEE models. For a more detailed review of group sequential analysis for correlated data, see Lee (1994) and Wu and Lan (1992).

All these results provide useful tools in sequential monitoring for a variety of spectrums. However, it has been known that sequential analysis is prone to exaggeration of treatment effect due to early stopping of a trial (Emerson and Fleming, 1990, Huges and Pocock, 1988, Huges, Freedman and Pocock, 1992, Pinheiro and DeMets 1995). The boundaries in group sequential methods are determined for the protection of the overall significance level only. Nevertheless, traditional estimates of treatment effect are not adjusted accordingly for sequential monitoring. Whitehead (1986) proposed an adjustment to the maximum likelihood estimate following a sequential probability ratio or triangular test. Emerson and Fleming (1990) studied estimate of normal means and pointed out that Whitehead's adjustment has the lowest mean squared error among all estimators over a wide range of alternatives. The essence of Whitehead's method relies on the feasibility of an assessment of the sequential

bias when the treatment effect is given. For independently collected observations, the distribution of the sequential maximum likelihood estimate can be approximated by Brownian motion, and thus its bias can be assessed. Using Brownian motion and recursive integration, Li (1996) evaluated the bias of the sequential maximum likelihood estimate. For correlated data, when the true correlation structure is known, the sequential estimate possesses a Gaussian independent increment structure (GIIS) (Gange and DeMets, 1996), which allows Brownian motion approximation to the distribution of the estimate of the treatment effect (Lan and Zucker, 1993). Pinherio and DeMets (1995) considered bias correction of the sequential estimate with GIIS using Whitehead's adjustment.

Both Li (1996) and Pinherio and DeMets' (1995) results utilized a Brownian motion approximation and showed satisfactory bias reduction using Whitehead's adjustment. However, when the true correlation structure of observations is unknown, the performance of bias reduction using Brownian motion is unclear. Since the GEE method does not require knowledge of true correlation structure given the alpha-spending function, the boundaries can be determined using the asymptotic normality of the GEE estimate and Mulnor's subroutine (Schervish, 1984) for calculation of multivariate normal boundaries given probability. But the lack of the known correlation structure prevents one from assessing the bias of the sequential estimate, and thus Whitehead's adjustment is not feasible. Perhaps the most forbidding factor in using Whitehead's adjustment to the GEE estimate is the extremely low speed of Mulnor's subroutine when the dimension of the multivariate normal distribution is higher than 3.

In the following, we first briefly review the GEE method for correlated data in Section 2.1 and the Lan and DeMets alpha-spending function for sequential analysis in Section 2.2. In Section 3, bias of the group sequential estimate using the independence working correlation in GEE estimation and the alpha-spending function in conjunction with the

multivariate integration method is examined for linear and logistic regressions. The influence of the true correlation structure on the sequential bias is studied using one-dependence, AR-1, exchangeable and independence structures (defined later). A bias-corrected sequential estimate using the independence working correlation structure in the GEE estimation and Brownian motion approximation is proposed in Section 4. The procedure is illustrated using logistic regression. Finally, Section 5 concludes our paper with some discussion.

2. Group sequential analysis for correlated data

2.1 GEE method for correlated data

Let K be the total number of interim analyses, and $N(k)$ be the number of subjects available for the k th interim analysis, $k = 1, 2, \dots, K$. Let $\mathbf{y}_{ik} = (y_{ik1}, y_{ik2}, \dots, y_{ikd_1}, \dots, y_{ik(d_{ik}-1)}, \dots, y_{ikd_{ik}})'$ be the d_{ik} responses from the i th subject at the k th interim analysis, $i = 1, 2, \dots, N(k)$. Thus, the additional responses obtained during the $(k-1)$ th to k th analysis are $y_{ikd_{ik-1}+1}, y_{ikd_{ik-1}+2}, \dots, y_{ikd_{ik}}$ for the i th subject. For $k < l$, it is possible that $d_{ik} = d_{il}$ which indicates that no additional response from the i th subject was collected during the $(k-1)$ th to k th analysis. Let \mathbf{x}_{ikj} be the covariate vector corresponding to y_{ikj} for $j = 1, 2, \dots, d_{ik}$, $i = 1, 2, \dots, N(k)$ and $k = 1, 2, \dots, K$.

The GEE method specifies the marginal distribution of y_{ikj} as

$$f(y_{ikj}) = \exp \{ [y_{ikj}\theta_{ikj} - a(\theta_{ikj}) + b(y_{ikj})] \phi \},$$

where $\theta_{ikj} = h(\eta_{ikj})$ for some monotonic link function h , $\eta_{ikj} = \mathbf{x}'_{ikj}\boldsymbol{\beta}$, and ϕ is a nuisance parameter. Then

$$E(y_{ikj}) = a'(\theta_{ikj}), \quad \text{var}(y_{ikj}) = a''(\theta_{ikj})/\phi.$$

Let $\mathbf{a}'(\boldsymbol{\theta}_{ik}) = (a'(\theta_{ik1}), \dots, a'(\theta_{ikd_{ik}}))'$, $\mathbf{A}_{ik}(\boldsymbol{\beta}) = \text{diag}\{a''(\theta_{ikj})\}$, $\mathbf{M}_{ik}(\boldsymbol{\beta}) = \partial \mathbf{a}'(\boldsymbol{\theta}_{ik})/\partial \boldsymbol{\beta}$ and

$\mathbf{r}_{ik}(\boldsymbol{\beta}) = \mathbf{y}_{ik} - E(\mathbf{y}_{ik})$. Let $\mathbf{R}_{ik}(\boldsymbol{\rho})$ be a working correlation matrix and

$$V_{ik}(\boldsymbol{\beta}, \boldsymbol{\rho}, \phi) = \mathbf{A}_{ik}^{1/2}(\boldsymbol{\beta})\mathbf{R}_{ik}(\boldsymbol{\rho})\mathbf{A}_{ik}^{1/2}(\boldsymbol{\beta})/\phi$$

be the working covariance matrix of \mathbf{y}_{ik} . The estimate of $\boldsymbol{\beta}$ at the k -th interim analysis can be obtained by solving the generalized estimating equation defined in Liang and Zeger (1986)

$$U_k(\boldsymbol{\beta}) = \sum_{i=1}^{N(k)} \mathbf{M}_{ik}(\boldsymbol{\beta})^t V_{ik}^{-1}(\boldsymbol{\beta}, \boldsymbol{\rho}, \phi) \mathbf{r}_{ik}(\boldsymbol{\beta}) = 0, \quad (1)$$

where the superscript “t” indicates transpose of a matrix. In solving (1), once an updated estimate of $\boldsymbol{\beta}$ is obtained, the parameters $\boldsymbol{\rho}$ and ϕ can be estimated using Pearson residuals and method of moments or other methods. Thus (1) can be viewed as a function of $\boldsymbol{\beta}$ alone by replacing $\boldsymbol{\rho}$ and ϕ by their consistent estimates $\boldsymbol{\rho}(\hat{\boldsymbol{\beta}}, \phi(\hat{\boldsymbol{\beta}}))$ and $\phi(\hat{\boldsymbol{\beta}})$. When $\mathbf{R}_{ik}(\boldsymbol{\rho})$ is the true correlation matrix of \mathbf{y}_{ik} , $\hat{\boldsymbol{\beta}}^{(k)}$ is most efficient in the family of working correlation matrices. Under standard regularity conditions, the joint distribution of $(\hat{\boldsymbol{\beta}}^{(1)}, \dots, \hat{\boldsymbol{\beta}}^{(k)})$ is asymptotically normal with mean $(\boldsymbol{\beta}, \dots, \boldsymbol{\beta})$. Let γ be the treatment effect under consideration, $\boldsymbol{\xi}$ be the rest of parameters in $\boldsymbol{\beta}$, i.e., $\boldsymbol{\beta} = (\boldsymbol{\xi}', \gamma)'$, and $\gamma^{(k)}$ be the estimate of γ corresponding to $\hat{\boldsymbol{\beta}}^{(k)}$. Then the joint distribution of $(\hat{\gamma}^{(1)}, \dots, \hat{\gamma}^{(k)})$ is asymptotically normal with mean (γ, \dots, γ) . For a more detailed description of GEE estimation for group sequential method, see Ganga and DeMets (1996).

In the following, the sequential bias is examined for four true correlation structures: one-dependence \mathbf{R}_1 , exchangeable \mathbf{R}_2 , AR-1 \mathbf{R}_3 and independence $\mathbf{R}_4 = \mathbf{I}$ (Liang and Zeger, 1986). Given a proper value of the pair-wise correlation ρ , for the one-dependence correlation structure, $\text{corr}(y_i, y_j) = \rho$ if $|i - j| = 1$ and 0 otherwise; for the exchangeable, $\text{corr}(y_i, y_j) = \rho$ for $i \neq j$; for the AR-1, $\text{corr}(y_i, y_j) = \rho^{|i-j|}$ for $i \neq j$; and observations from the independence structure are independent of each other.

2.2 The alpha-spending function

Unlike the Pocock (1977), O'Brien and Fleming (1979) and Slud and Wei (1982) methods which pre-determine a fixed number of sequential monitorings in a clinical trial, the alpha-spending function technique introduced by Lan and DeMets (1983) allocates the total pre-assigned significance level α through a given monotonic function $\alpha(t^*)$ on $[0,1]$ with $\alpha(0) = 0$ and $\alpha(1) = \alpha$, where the argument t^* represents the amount of information fraction available at the time of monitoring. The alpha-spending function method relaxes the requirement of fixed number of monitoring and thus allows for a flexible schedule of interim analyses.

Any monotonic function $\alpha(\cdot)$ on $[0, 1]$ with $\alpha(0) = 0$ and $\alpha(1) = \alpha$ can serve as an alpha-spending function for sequential monitorings. Let Φ be the cumulative distribution function and z_ν be the upper ν -percentile of the standard normal distribution $N(0,1)$. Here are four alpha-spending functions: a. $\alpha_1(t) = 2[1 - \Phi(z_{\alpha/2}/\sqrt{t})]$; b. $\alpha_2(t) = \alpha t$; c. $\alpha_3(t) = \alpha t^2$; and d. $\alpha_4(t) = \alpha \log[1 + (e - 1)t]$.

The boundaries derived by α_1 and α_4 are very similar to those given by Pocock (1977) and O'Brien and Fleming (1979), respectively. The O'Brien-Fleming type boundaries are large at early stages, and thus conservative, and close to the fixed sample cut-point at the end of the trials. The Pocock type boundaries are almost constant cross the entire trial. The level of possibly early stopping of a trial increases from α_1 to α_4 .

Let $\psi_k = \text{var}^{-1}(\gamma^{(k)})$, $k = 1, 2, \dots, K$. Then the amount of information accumulated by the k -th interim analysis is defined as ψ_k , where ψ_K measures the total information available by the end of the trial. The information fraction at the k -th analysis is thus $t_k^* = \psi_k/\psi_K = \text{var}(\gamma^{(K)})/\text{var}(\gamma^{(k)})$. The information defined in this way reflects the amount of statistical information about the treatment effect γ and thus has more statistical appeal. In the simplest one-sample problem with identically and independently distributed observations with variance one, $\psi_k = N(k)$, the number of subjects observed by the k -th interim analysis.

and the information fraction $t_k^* = N(k)/N(K)$ which is a function of $N(k)$ and the total sample size $N(K)$. In general, specification of t^* needs estimates of total sample size $N(K)$ and variation of observations. All these quantities are needed in general for sample size calculation in a planned clinical trial, or can be estimated using information collected at the time of interim analyses.

The information fraction t^* is needed for determination of boundaries for sequential monitorings. Once t_k^* is determined, one can use $\alpha(\cdot)$ to calculate the amount of significance level spent at the k th analysis. When the information fraction can not be evaluated exactly, there are a number of ways to estimate it, including using calendar time or number of responses observed at the time of the analysis. It has been shown that determination of t^* is fairly robust in terms of Type I error and power in general, see Lan and Zucker (1993) and Lan, Reboussin and DeMets (1994) and our examples in Section 5. At the time being, let t_k^* be the information fraction estimated by one of the possibilities. Then the sequential boundaries c_1, c_2, \dots, c_K are calculated via

$$P \left(\left| \frac{\hat{\gamma}^{(1)}}{\sqrt{\psi_1}} \right| \leq c_1, \dots, \left| \frac{\hat{\gamma}^{(k-1)}}{\sqrt{\psi_{k-1}}} \right| \leq c_{k-1}, \left| \frac{\hat{\gamma}^{(k)}}{\sqrt{\psi_k}} \right| > c_k \right) = \alpha(t_k^*) - \alpha(t_{k-1}^*),$$

for $k = 1, 2, \dots, K$.

When the working correlation matrix $\mathbf{R}_{ik}(\boldsymbol{\rho})$ is the true correlation matrix, then for $k < l$, the covariance of \mathbf{y}_{ik} and \mathbf{y}_{il} equals the variance of \mathbf{y}_{il} (Gange and DeMets, 1996). Thus, the sequence $\{\hat{\gamma}^{(j)}\}$ has an independent increment structure, i.e., for $j < k < l$, $\text{cov}(\hat{\gamma}^{(k)} - \hat{\gamma}^{(j)}, \hat{\gamma}^{(l)} - \hat{\gamma}^{(k)}) = 0$. Hence the Brownian motion approximation and existing software (Reboussin, DeMets, Lan and Kim, 1992) can be used in conjunction with the α -spending function technique of Lan and DeMets (1983) to calculate the sequential boundaries. Even if $\mathbf{R}_{ik}(\boldsymbol{\rho})$ is not the true correlation matrix, the joint distribution of $(\hat{\gamma}^{(1)}, \dots, \hat{\gamma}^{(k)})$ is still asymptotically normal. Using the asymptotic normality of $(\hat{\gamma}^{(1)}, \dots, \hat{\gamma}^{(k)})$ and Mul-

nor's subroutine, one can calculate the asymptotic sequential boundaries according to the allocation of significance level.

3. Bias of group sequential estimate

We now examine the bias associated with group sequential analysis. Since the estimate following group sequential analysis does not take into account that the trial is stopped early, it inevitably exaggerates the treatment effect. Let $\hat{\gamma}$ denote the estimate of γ at stopping time τ , where τ is defined as $\inf\{k \mid |\hat{\gamma}^{(k)}| > c_k, k = 1, 2, \dots, K-1\}$, or $\tau = K$ otherwise. The expectation of $\hat{\gamma}$ equals

$$\begin{aligned} E(\hat{\gamma}) &= \sum_{k=1}^K E \left\{ \hat{\gamma}^{(k)} I_{[|\hat{\gamma}^{(j)}/\sqrt{\psi_j}| \leq c_j, j=1,2,\dots,k-1, |\hat{\gamma}^{(k)}/\sqrt{\psi_k}| > c_k]} \right\} \\ &+ E \left\{ \hat{\gamma}^{(K)} I_{[|\hat{\gamma}^{(j)}/\sqrt{\psi_j}| \leq c_j, j=1,2,\dots,K]} \right\} \end{aligned}$$

where $I_{[A]}$ is the indicator function of an event A . Let $b(\gamma)$ denote the bias of $\hat{\gamma}$, i.e., $b(\gamma) = E(\hat{\gamma}) - \gamma$. One can see that if $\gamma = 0$, i.e., there is no treatment effect, then $b(0) = 0$ since the joint distribution of $\hat{\gamma}_{(1)}$ and $\hat{\gamma}_{(2)}$ in this situation is symmetric about zero. If γ is positive, then for $K = 2$, the simplest case with only one interim analysis,

$$\begin{aligned} b(\gamma) &= E(\hat{\gamma}^{(1)}) + E \left\{ (\hat{\gamma}^{(2)} - \hat{\gamma}^{(1)}) I_{[|\hat{\gamma}^{(1)}/\sqrt{\psi_1}| \leq c_1]} \right\} - \gamma \\ &\approx E \left\{ (\hat{\gamma}^{(2)} - \hat{\gamma}^{(1)}) I_{[\hat{\gamma}^{(1)}/\sqrt{\psi_1} \leq c_1]} \right\}. \end{aligned}$$

Note that only when $\hat{\gamma}^{(1)}/\sqrt{\psi_1}$ is small, is the difference $\hat{\gamma}^{(2)} - \hat{\gamma}^{(1)}$ taken into account in the bias calculation. Thus $\hat{\gamma}$ tends to be an over-estimate for positive γ . Also, the later the first interim analysis, the closer $\hat{\gamma}^{(2)}$ and $\hat{\gamma}^{(1)}$ are, and thus the smaller the bias in general.

The two marginal distributions we examine here are a Gaussian distribution with identical link and a binomial distribution with logit link. The former corresponds to the usual linear regression with normal deviates and the latter to the logistic regression. Assume that the observations are from a randomized clinical trial with treatment effect γ , that is $\eta_{ikj} = \xi + \gamma x_i$.

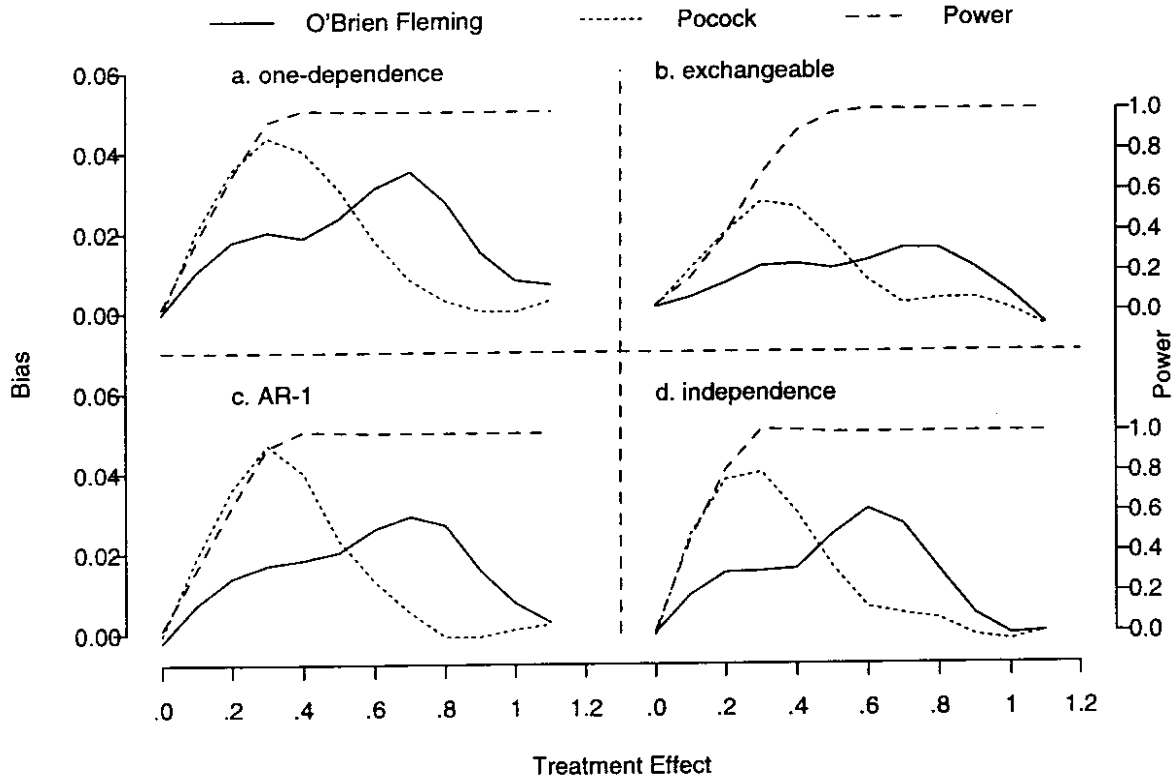


Figure 1: Bias and average power for linear regression with $\rho = 0.3$.

where x_i equals one if the i -th subject receives treatment and zero otherwise. The pair-wise correlation ρ is 0.3 or 0.7, reflecting different levels of correlation. The sample size n is 100 with 50 subjects in each group. The true correlation structures used in the simulation are the one-dependence \mathbf{R}_1 , exchangeable \mathbf{R}_2 , AR-1 \mathbf{R}_3 and independence \mathbf{R}_4 , while the working correlation is the independence structure. For simplicity, assume that the observations from each subject are evenly spaced throughout the course of the trial. Thus the design of the trial is completely balanced in terms of data collection. As a result, one can show that the information fraction defined using ψ_k equals the proportion of observations collected and the

fraction of time elapsed at each interim analysis. The simulation was performed with 2000 replicates.

3.1 Linear regression

For linear regression, $y_{ikj} = \xi + \gamma x_i + \epsilon_{ikj}$, $\epsilon_{ikj} \sim N(0, \sigma^2)$. Thus $h(\eta) = \eta = \theta = \xi + \gamma x$. $\phi = \sigma^{-2}$, $a(\theta) = 2^{-1}\theta^2$, $b(y, \phi) = -2^{-1}[y^2/\sigma^2 + \log(2\pi\sigma^2)]$. Assume $\sigma^2 = 1$ and $\xi = 1$. Suppose that there were ten observations from each subject and that the trial was monitored at the times when 20%, 50%, 70% and 100% of observations have been collected,

Figure 1 presents plots of bias for $\rho = 0.3$. Overall, the O'Brien-Fleming boundaries have smaller bias than Pocock's boundaries. The power is almost the same for the O'Brien-Fleming and Pocock boundaries, and their average power is also shown in Figure 1. Pocock's boundaries produce a large bias for small values of γ , while O'Brien-Fleming's boundaries have their peaks when Pocock's boundaries are almost unbiased. However, this occurs for extremely large, almost unlikely treatment effect. The Pocock boundaries are smaller than the O'Brien-Fleming boundaries for early monitoring and thus produce larger bias than O'Brien-Fleming boundaries for a small treatment effect. For the O'Brien-Fleming boundaries, since it is very hard to stop early for a small treatment effect, the trial will be carried almost to its designed end and thus the bias is much smaller than for the Pocock boundaries. After reaching their respective peaks, the bias of both types of boundaries then decreases with the increase of treatment effect. All these properties are consistent with those observed by others such as Pinheiro and DeMets (1995) and Li (1996).

The influence of true correlation structure is clearly demonstrated here by the similar bias pattern of plots a, c, and d of Figure 1. For the one-dependence structure, only adjacent observations are correlated, for the AR-1 structure, correlation of observations more than one-lag apart is almost negligible. Thus bias from these correlation structures behave almost

the same. Interestingly, the exchangeable structure has much smaller bias compared to other structures, including even the independence correlation. In fact, because of the correlation among all observations from single subject, $(\hat{\gamma}^{(1)}, \dots, \hat{\gamma}^{(k)})$ are more correlated with each other than for other three correlation structures, and hence produce less bias. The value of pair-wise correlation ρ has little influence on the overall level of correlation for the three correlation structures \mathbf{R}_1 , \mathbf{R}_2 , and \mathbf{R}_4 and thus their bias pattern and magnitude for $\rho = 0.7$ are almost the same as for $\rho = 0.3$. Figure 2 shows the bias for the exchangeable structure with $\rho = 0.7$. For the exchangeable structure, the magnitude of bias for $\rho = 0.7$ should be less than for $\rho = 0.3$ since interim estimates are more correlated for larger values of ρ than smaller values.

3.2 Logistic regression

Suppose the response y is binary with possible values 0 and 1. Let p_{ikj} be the probability that y_{ikj} is equal to 1 and assume that $p_{ikj} = \frac{e^{\xi + \gamma x_i}}{1 + e^{\xi + \gamma x_i}}$ and $\theta_{ikj} = \xi + \gamma x_i$. Then

$$f(y_{ikj}) = p_{ikj}^{y_{ikj}} (1 - p_{ikj})^{1 - y_{ikj}} = \exp\{y_{ikj} \cdot \theta_{ikj} - \log(1 + e^{\theta_{ikj}})\}$$

Thus $h(\eta) = \eta = \theta$, $\phi = 1$ and $a(\theta) = \log(1 + e^\theta)$. Assume that $\xi = .1$. Suppose that there were eight observations from each subject and that the trial is monitored at the times when 25%, 50%, 75% and 100% of observations are collected. Figure 3 is the counterpart of Figure 1 for logistic regression. Again, bias for one-dependence, AR-1 and independence structures is almost the same, and the exchangeable structure has smaller bias than the other three. For Pocock's boundaries, the pattern of bias is similar to the linear regression case; for O'Brien-Fleming's boundaries, the bias keeps increasing for the range of γ examined. The probability of 1 is 0.5 and 0.82 for γ equal to 0 and 1.5 respectively. Thus the range of increase of response rate from control to treatment group is from 0% to 64% for $\gamma \in (0, 1.5)$. The average power reaches 100% when $\gamma = 1.0$ for all four situations. Thus O'Brien-Fleming's boundaries

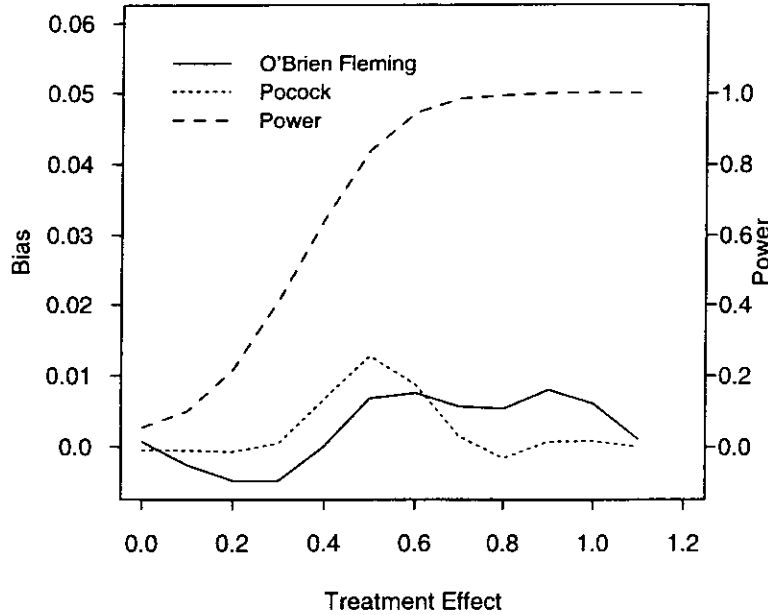


Figure 2: Bias and average power for linear regression for true exchangeable correlation structure with $\rho = 0.7$.

produce less bias than Pocock's for most practical situations. Similar to the linear regression case, the bias for $\rho = 0.7$ is almost the same as for $\rho = 0.3$ for the one-dependence, AR-1 and independence structures and smaller than for $\rho = 0.3$ for the exchangeable structure (not shown here).

3.3 Influence of different monitoring patterns

The previous simulations assumed uniform monitoring of the clinical trials, i.e., interim analyses were performed at times with equally spaced information fractions. To examine

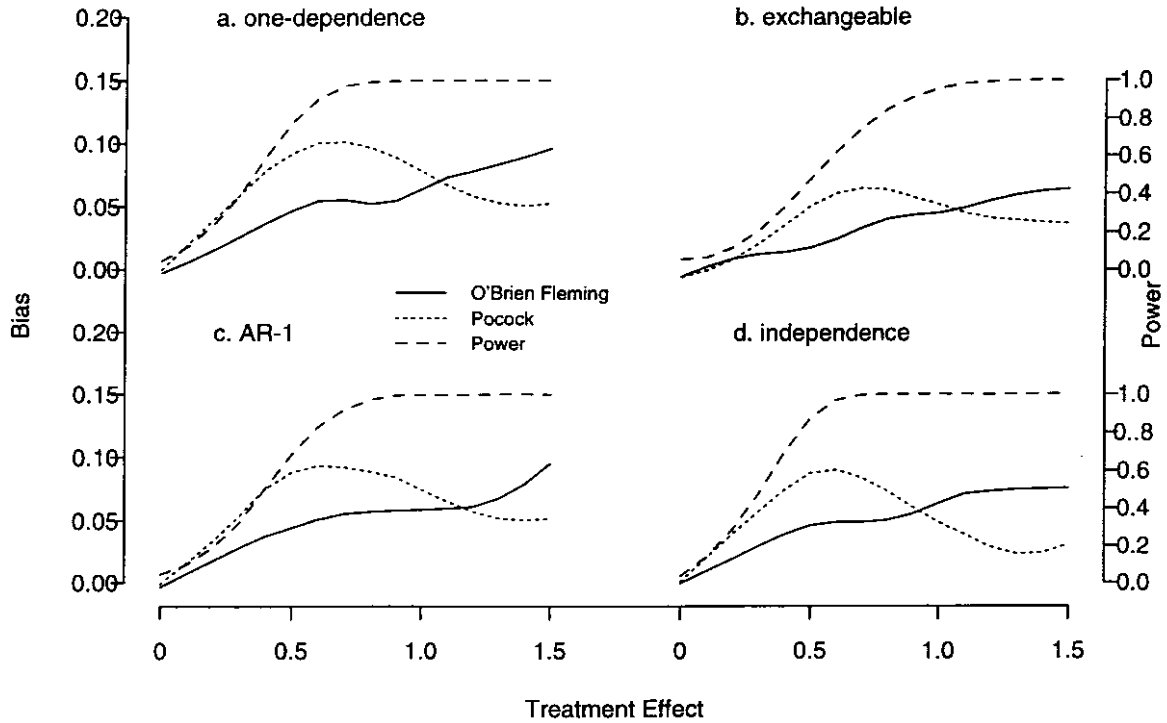


Figure 3: Bias and average power for logistic regression with $\rho = 0.3$.

the sequential bias for difference monitoring schemes, assume that the interim analyses were performed when 2, 3, 5, and 8 observations were collected from each subject for an early monitoring scheme and 4, 6, 7, and 8 observations for a late monitoring scheme. Figure 4 is the bias for logistic regression with $\rho = 0.3$ and Pocock's boundaries for different monitoring schemes. It can be seen that the late monitoring has much smaller bias than the other two schemes. Using the number of observations collected, the information fractions for the early and late monitoring are 0.25, 0.375, 0.625 and 1, and 0.50, 0.75, 0.875 and 1, respectively. Since information fractions of the early and uniform monitoring schedule considered here

do not disagree too much, the bias of these two monitoring schemes almost can not be distinguished from each other. Simulations for other true correlation structures and the linear regression, as well as for O'Brien-Fleming's boundaries, reveal similar results, and thus will not be shown here. Similar results were observed by others also, e.g., Li (1996) and Pinherio and DeMets (1995).

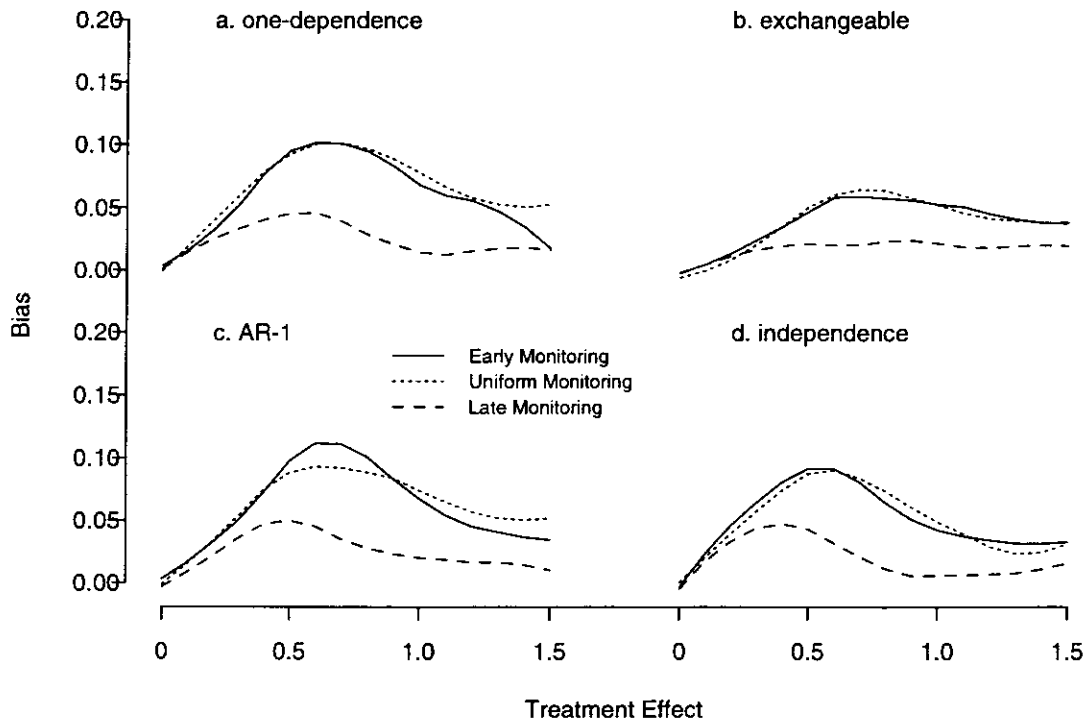


Figure 4: Bias using Pocock's boundaries for logistic regression with $\rho = 0.3$.

4. Bias reduction of group sequential estimate

4.1 Bias estimation

We now consider the Whitehead bias correction. The Whitehead bias-corrected estimate γ^* is defined as the solution of

$$\gamma^* = \hat{\gamma} - b(\gamma^*). \quad (2)$$

The bias of γ^* is $b^*(\gamma^*) = E(\gamma^*) - \gamma = [E(\hat{\gamma}) - \gamma] - E[b(\gamma^*)] = b(\gamma) - E[b(\gamma^*)]$. So the bias of $\hat{\gamma}$ is reduced by the amount of $E[b(\gamma^*)]$. See Li (1996) for more discussion of the properties of γ^* .

In practice, (2) can be solved numerically using Newton-Raphson iteration

$$\gamma_i^* = \gamma_{i-1}^* + \frac{(\hat{\gamma} - \gamma_{i-1}^*) - b(\gamma_{i-1}^*)}{1 + b'(\gamma_i^*)}, i = 1, 2, \dots,$$

where $b'(\gamma)$ denotes the derivative of the bias function $b(\gamma)$, and $\gamma_0^* = \hat{\gamma}$. The iteration requires evaluation of $b'(\gamma)$ which can be done numerically. The key in using the Whitehead correction is the estimate of the sequential bias given the treatment effect. In the following, we are going to use Brownian motion to approximate the sequential bias $b(\gamma)$.

Recall the information fraction $t_k^* = \psi_k/\psi_K$. If the working correlation is the true correlation, then for $k < l$, $\text{cov}(\hat{\gamma}^k, \hat{\gamma}^l) = \text{var}(\hat{\gamma}^l)$ (Gange and DeMets, 1996). As a result

$$\begin{pmatrix} \hat{\gamma}^{(1)} \\ \dots \\ \hat{\gamma}^{(K)} \end{pmatrix} \sim N \left[\begin{pmatrix} \gamma \\ \dots \\ \gamma \end{pmatrix}, \psi_K \left(\frac{\min(t_i^*, t_j^*)}{t_i^* \cdot t_j^*} \right)_{i,j} \right]. \quad (3)$$

Define

$$B_N(t_k^*) = \sqrt{t_k^*} \frac{\hat{\gamma}^k}{\psi_k}.$$

Then, we have (a) $E[B_N(t_k^*)] = t_k^* \cdot (\sqrt{\psi_K} \cdot \gamma)$; (b) $\text{var}[B_N(t_k^*)] = t_k^*$; and (c) $\text{cov}[B_N(t_k^*), B_N(t_l^*)] = \min\{t_k^*, t_l^*\}$. Hence, $B_N(t^*)$ resembles a Brownian motion on $[0,1]$. Thus, instead of Mulnor's

subroutine, one can use the existing subroutine of Reboussin et al. (1992) for the calculation of sequential boundaries.

The use of the Reboussin et al. subroutine is not the only advantage of using the true correlation in the GEE estimation. Since (3) is the true joint distribution of the interim estimate, one can use it to estimate the bias of the sequential estimate of γ . Given the sequential boundaries $\{c_k\}$ calculated under null hypothesis, and a value of γ , one proceeds as follows:

1. Calculate ψ_1, \dots, ψ_K and t_1^*, \dots, t_K^* ;
2. Generate $\hat{\gamma}^{(1)}, \dots, \hat{\gamma}^{(K)}$ from distribution (3) ;
3. Compare $\{|\hat{\gamma}^{(k)} \sqrt{\psi_k}|\}$ with $\{c_k\}$ and calculate $\hat{\gamma} = \hat{\gamma}^{(\tau)}$, where τ is the stopping time;
4. repeat steps 2 and 3 M times.

Let $\{\hat{\gamma}_i\}_{i=1}^M$ be the M estimates of γ . The sequential bias of γ is then estimated as

$$b(\gamma) = \frac{\sum_{i=1}^M \hat{\gamma}_i}{M} - \gamma.$$

4.2 Numerical simulations

The above procedure is based on the assumption that the working correlation \mathbf{R} is the true correlation structure. In practice, the true correlation structure is unknown, and thus the independence structure is often a natural alternative. If the independence structure is close to the true correlation, then this procedure should give satisfactory results. If the independence structure is far wrong from correct, the performance of using the independence structure is unknown. For the previous logistic regression example with $\rho = 0.3$, Figure 5 shows the original and bias-corrected sequential estimate using the O'Brien-Fleming boundaries. The bias correction is surprisingly good even for the true exchangeable structure, and

the amount of bias reduction is almost the same for all the four structures. Since the bias for the exchangeable model is relatively smaller than for the other three structures, the procedure over-corrects the sequential estimate slightly. But the magnitude of over-correction is almost negligible compared to the original bias. In fact, no matter what the true correlation structure is, the mean of (3) is always correct. Whether the working correlation is the true structure or not only influences the variance structure of (3).

The first half of Table 1 summarizes the operating characteristics of the original and bias-corrected sequential estimates for the true one-dependence correlation structure for $\gamma = 0, 1.4(0.2)$. Although the bias of corrected estimate is much smaller than the original estimate, its standard deviation is almost the same as the original. This is consistent with the fact that the bias-corrected estimate actually shrinks the original sequential estimate. The ratio of bias to the true value of γ is significantly smaller for the corrected bias than for the original bias. For the true exchangeable structure, the standard deviation of the bias-corrected estimate is slightly smaller than that of the original. Since the true one-dependence structure is almost the same as the independence working correlation, its standard deviation is smaller than for the exchangeable case. The comparison of operating characteristics of the two estimates for the true AR-1 and independence structures is very similar to the one-dependence and thus not shown here.

Sometimes, the nature of the observations may be useful in determining the form of the true correlation structure, as in a family cohort study where the correlation of any two family members is often the same for the whole family and leads to an exchangeable structure. For a fixed sample size, use of true correlation in the GEE estimation enjoys several desirable statistical properties, such as small variances and higher power. Unfortunately, the relation between power and bias is not monotonic at the micro level in the group sequential method. Consider the logistic regression model again and assume that the true exchangeable correla-

Table 1: Operating characteristics of sequential and bias-corrected estimates using O'Brien-Fleming's boundaries for logistic regression with true one-dependence or exchangeable correlation structure and $\rho = 0.3$.

	γ	$b(\gamma)$	$std(\hat{\gamma})$	$b(\gamma)/\gamma$	$b^*(\gamma)$	$std(\gamma^*)$	$b^*(\gamma)/\gamma$
One-dependence	0.0	-0.015	0.209	*	-0.013	0.196	*
	0.2	0.011	0.228	0.059	-0.002	0.214	-0.009
	0.4	0.024	0.247	0.062	-0.001	0.235	-0.004
	0.6	0.047	0.262	0.079	0.011	0.255	0.019
	0.8	0.039	0.275	0.049	-0.001	0.268	-0.001
	1.0	0.064	0.288	0.064	0.018	0.280	0.018
	1.2	0.073	0.353	0.061	0.022	0.347	0.018
	1.4	0.057	0.360	0.041	0.002	0.358	0.001
Exchangeable	0.0	-0.005	0.283	*	-0.005	0.264	*
	0.2	-0.005	0.296	-0.025	-0.017	0.277	-0.089
	0.4	0.016	0.319	0.042	-0.008	0.304	-0.019
	0.6	0.003	0.334	0.005	-0.029	0.322	-0.048
	0.8	0.043	0.347	0.054	0.003	0.337	0.004
	1.0	0.034	0.353	0.034	-0.012	0.344	-0.012
	1.2	0.061	0.420	0.051	0.010	0.413	0.008
	1.4	0.053	0.422	0.038	-0.000	0.419	-0.000

tion is used in the GEE estimation. The information fractions for the uniform monitoring using the independence correlation in the GEE estimation are 0.25, 0.50, 0.75 and 1, and the corresponding O'Brien-Fleming boundaries are 4.332, 2.963, 2.359 and 2.014.

The information fractions using the true exchangeable structure are 0.596, 0.816, 0.930 and 1. Using the Brownian motion approximation, the O'Brien-Fleming boundaries are 2.679, 2.259, 2.143 and 2.091. Since observations from single subject are all correlated, the two observations obtained at the first interim analysis contain information about observations not collected, and the amount of statistical information reaches to almost 60% of the totally available. Thus the O'Brien-Fleming boundary for the first monitoring is only 2.679.

Figure 6 shows the original and bias-corrected sequential estimate using the true correla-

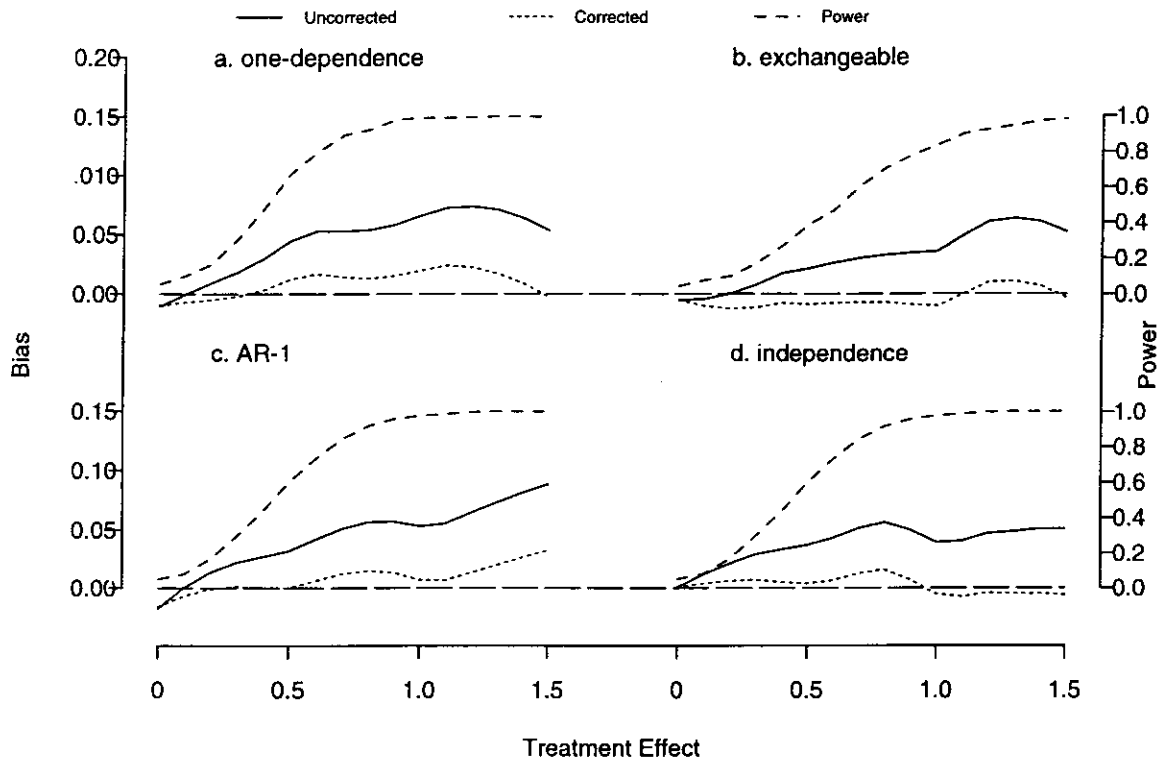


Figure 5: Bias reduction using O’Brien-Fleming’s boundaries and the independence working correlation structure for logistic regression with $\rho = 0.3$.

tion structure in both the GEE estimation and the Brownian motion approximation. Both types of bias are much larger than their counterpart in Figure 5. Table 2 shows power and average stopping time (AST) using the true exchangeable and the independence working structures. Obviously, the independence structure has a larger average stopping time (AST) because of its larger boundaries than the exchangeable. But the independence structure almost gives the same power as the true correlation structure. Thus using the independence structure only delays the stopping time of the trial.

Table 2: Power and average stopping time (AST) of sequential analysis with O'Brien-Fleming's boundaries calculated using the independence working correlation or the true exchangeable correlation structure for logistic regression with $\rho = 0.3$.

	γ	0.0	0.2	0.40	0.60	0.80	1.0	1.2	1.4
Independence	Power	0.042	0.097	0.269	0.466	0.709	0.838	0.926	0.971
	AST	3.977	3.951	3.805	3.578	3.217	2.897	2.582	2.372
Exchangeable	Power	0.053	0.093	0.248	0.480	0.690	0.865	0.933	0.977
	AST	3.925	3.845	3.580	3.179	2.677	2.126	1.836	1.506

5. Discussion

This paper examines the bias of the group sequential estimate of treatment effect with correlated data using GEE estimation. Because the true correlation structure is unknown in most practical cases, our results focus on the use of the independence working correlation structure. Our simulation results are very close to what reported for independent observations by others. With the independence working correlation model, the influence of the true correlation structure is determined by the overall level of correlation of observations from individual subjects. The higher the overall level of correlation is, the smaller the bias of the sequential estimate. For the four correlation structures we examined, only the true exchangeable correlation structure distinguishes itself from other structures by having smaller bias.

One of the difficulties associated with the bias reduction is the choice of working correlation in the sequential analysis. For the traditional sequential analysis, the use of the independence working correlation still preserves the overall significance level. For the bias reduction, the performance of using the independence working correlation is unclear because of the requirement of bias evaluation for the Whitehead method. When the working correlation is the true correlation, we have shown how to correct the bias of sequential estimate

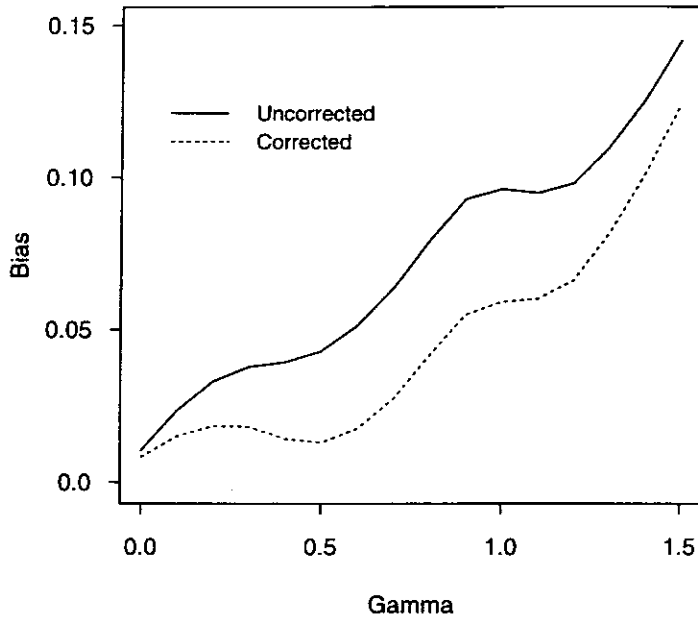


Figure 6: Bias reduction using O’Brien-Fleming’s boundaries and the true exchangeable correlation structure for logistic regression with $\rho = 0.3$.

using Monte Carlo simulation.

With the more realistic independence working correlation model, we have shown that the bias reduction is also substantial. Although over-corrected for the true exchangeable structure, the bias-corrected estimate is still highly preferred than the original estimate.

The previous results focused on the O’Brien-Fleming type boundaries. For the previous example of logistic regression with $\rho = 0.3$ and uniform monitoring, Table 3 shows the original and corrected biases for $\gamma = 0$, and 1.1 for the Pocock boundaries with the independence

working correlation. One can see that both the original bias and the level of bias reduction decrease with sample size. For fixed sample size, the amount of reduction is almost constant, i.e., independent of correlation structure, because of the use of independence working correlation in the bias evaluation. The Type I errors are well within acceptable range of the nominal level 0.05 except for the 0.081 for the true independence correlation structure with sample size $n = 50$. Aside from random error, this may also be due to the poor normal approximation to the joint distribution of GEE estimate of γ for small sample size. Simulation for O'Brien-Fleming boundaries for different sample sizes showed similar results to Table 3 and thus is not reported here.

We showed the magnitude of bias reduction for the logistic regression with equally spaced observations. However, with a fixed set of boundaries $\{c_k\}$ and for a given set of information fractions $\{t_i^*\}$ and ψ_K , the level of reduction is irrelevant to the model assumption since distribution (3) is determined completely by $\{t_i^*\}$ and ψ_K . Thus, the results we observed reflect the performance of this method in general. For most realistic situations, the O'Brien-Fleming type boundaries give relatively smaller bias compared to the Pocock type boundaries. Also, given the information fractions, the amount of reduction using (3) is mainly determined by and positively related to the magnitude of the original bias.

Acknowledgement

Qu's work was supported by grant CA-09565 from the National Cancer Institute.

Table 3: Operating characteristics of sequential and bias-corrected estimates using Pocock's boundaries for logistic regression using the independence working correlation structure with $\rho = 0.3$ and sample size n .

	Type I error	$\gamma = 0$				power	$\gamma = 1.1$			
		$b(\gamma)$	$\text{std}(\hat{\gamma})$	$b^*(\gamma)$	$\text{std}(\gamma^*)$		$b(\gamma)$	$\text{std}(\hat{\gamma})$	$b^*(\gamma)$	$\text{std}(\gamma^*)$
<i>n=50</i>										
1-dep.	0.048	-0.002	0.304	-0.001	0.266	0.980	0.177	0.424	0.080	0.449
Exch.	0.047	-0.000	0.405	0.000	0.353	0.740	0.090	0.470	-0.003	0.483
AR-1	0.049	-0.008	0.313	-0.007	0.271	0.945	0.129	0.399	0.032	0.419
Indep.	0.081	0.003	0.272	0.002	0.234	0.997	0.131	0.354	0.032	0.377
<i>n=100</i>										
1-dep.	0.052	0.009	0.216	0.008	0.189	1.000	0.088	0.315	0.053	0.340
Exch.	0.041	-0.009	0.269	-0.008	0.234	0.971	0.049	0.322	0.012	0.347
AR-1	0.044	-0.006	0.219	-0.006	0.192	1.000	0.053	0.306	0.016	0.331
Indep.	0.047	0.004	0.185	0.003	0.162	1.000	0.063	0.293	0.027	0.317
<i>n=150</i>										
1-dep.	0.056	-0.009	0.177	-0.008	0.155	1.000	0.015	0.263	0.000	0.277
Exch.	0.049	-0.005	0.229	-0.004	0.200	0.999	0.009	0.278	-0.006	0.294
AR-1	0.045	-0.003	0.176	-0.003	0.154	1.000	0.039	0.269	0.026	0.283
Indep.	0.048	0.000	0.148	0.000	0.130	1.000	0.006	0.245	-0.007	0.259
<i>n=200</i>										
1-dep.	0.044	-0.003	0.144	-0.002	0.126	1.000	0.006	0.250	0.001	0.257
Exch.	0.037	0.007	0.192	0.006	0.167	1.000	0.013	0.249	0.007	0.257
AR-1	0.052	-0.005	0.163	-0.005	0.143	1.000	0.004	0.242	-0.000	0.249
Indep.	0.066	-0.000	0.129	-0.000	0.112	1.000	0.003	0.216	-0.000	0.222

References

- DeMets, D. L. & Lan, K. K. G. (1994). Interim analysis: The alpha-spending function approach. *Statistics in Medicine* **13**, 1341–1352.
- Emerson, S. & Fleming, T. R. (1990). Parameter estimation following group sequential hypothesis testing. *Biometrika* **4**, 875–892.
- Gange, S. J. and DeMets, D. L. (1996). Sequential monitoring of clinical trials with correlated responses. *Biometrika* **83**, 157–167
- Huges, M. D. and Pocock, S. J. (1988). Stopping rules and estimation problems in clinical trials, *Statistical in Medicine* **7**, 1231–1242.
- Huges, M. D., Freedman, L. and Pocock, S. J. (1992). The impact of stopping rules on heterogeneity of results in overview of clinical trials, *Biometrics* **48**, 41–53.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **70**, 659–663.
- Lan, K. K. G. & DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70**, 659–663.
- Lan, K. K. G. & Zucker, D. (1993). Sequential monitoring of clinical trials: The role of information in Brownian motion. *Statistics in Medicine* **12**, 753–765.
- Lan, K. K. G., Reboussin, D. M. and DeMets, D. L. (1994). Information and information fractions for design and sequential monitoring of clinical trials. *Communications in Statistics, Part A—Theory and Methods* **23**, 403–420.
- Lee, J. W. (1994). Group sequential testing in clinical trials with multivariate observations: A review. *Statistics in Medicine* **13**, 101–111.
- Lee, J. W. & DeMets, D. L. (1991). Sequential comparison of change with repeated measurement data. *J. Am. Statist. Assoc.* **86**, 757–62.
- Liang, K. Y. and Zeger S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.
- Li. Z. Q. (1996). Bias issue and sequential checking of the proportional hazards assumption following group sequential tests. *Ph.D. Thesis*, Department of Statistics, University of Wisconsin-Madison.
- O'Brien, P. C. & Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics* **35**, 549–556.
- Pinheiro, J. C. & 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**, 191–99.

Reboussin, D. M., DeMets, D. L., Kim K. and Lan, K. K. G. (1992). Programs for computing group sequential bounds using the Lan-Demets method. *Technical Report 60*, Department of Biostatistics, University of Wisconsin-Madison.

Schervish, M. J. (1984). Algorithm as 195. Multivariate normal probability with error bound. *Applied Statistics* **33**, 81–94.

Slud, E. and Wei, L. J. (1982). Two sample repeated significance tests based on the modified Wilcoxon statistic. *Journal of the American Statistical Association* **77**, 862–868.

Wei, L. J., Su, J. Q. and Lachin, J. M. (1990). Interim analyses with repeated measurements in a sequential clinical trial. *Biometrika* **77**, 359–264.

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

Wu, M. C. and Lan, K. K. Gordon (1992). Sequential monitoring for comparison of changes in a response variable in clinical studies. *Biometrics* **48**, 765–779.

Department of Biometrics, H4-M2316, Hoechst Marion Roussel, Inc., 10236 Marion Drive, Kansas City, MO 64137-1405, U.S.A. Tel: (816) 966-4257.

Department of Biostatistics, University of Wisconsin-Madison, 600 Highland Avenue, Madison, Wisconsin 53792, U.S.A.