View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Collection Of Biostatistics Research Archive

UW Biostatistics Working Paper Series

2-5-2010

Estimates of Information Growth in Longitudinal Clinical Trials

Abigail Shoben
University of Washington, ashoben@uw.edu

Kyle Rudser
University of Minnesota, rudser@umn.edu

Scott S. Emerson
University of Washington, semerson@u.washington.edu

Suggested Citation

Shoben, Abigail; Rudser, Kyle; and Emerson, Scott S., "Estimates of Information Growth in Longitudinal Clinical Trials" (February 2010). *UW Biostatistics Working Paper Series.* Working Paper 358.
http://biostats.bepress.com/uwbiostat/paper358

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.
Copyright © 2011 by the authors

Estimates of Information Growth in Longitudinal Clinical Trials

Abigail B. Shoben [1,*], Kyle D. Rudser [2], Scott S. Emerson [1]

[1] Department of Biostatistics, Box 357232, University of Washington, Seattle, Washington, 98195, U.S.A.

[2] Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota, 55414, U.S.A.

[*] *email: ashoben@u.washington.edu*

Summary

In group sequential clinical trials, it is necessary to estimate the amount of information present at interim analysis times relative to the amount of information that would be present at the final analysis. If only one measurement is made per individual, this is often the ratio of sample sizes available at the interim and final analyses. However, as discussed by Wu and Lan (1992), when the statistic of interest is a change over time, as with longitudinal data, such an approach overstates the information. In this paper, we discuss other problems that can result in overestimating the information, such as heteroscedasticity and correlated observations. We demonstrate that when using an inefficient estimator on unbalanced data, the true information growth can be nonmonotonic across interim analyses.

Key Words: Unbalanced data; Inefficient estimators.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

1 Introduction

In many group sequential clinical trials, repeated measurements are made on continuous outcomes for each individual over a specified follow up time period. Such longitudinal measures may be used to evaluate possible treatment effects, such as time averaged measurement (area under the curve), rate of change (slope), or difference between initial and final values, any of which can be taken as specific cases of a weighted area under the curve approach (Kittelson et al., 2005). For illustrative purposes, we consider when the outcome of interest is the rate of change over time, such as when monitoring tumor growth, CD4+ cell counts, or cognitive decline in Alzheimer's Disease.

For ethical and financial reasons, it is usually necessary to perform interim analyses of data from clinical trials before all study data have been collected. In order to maintain the type I error rate, several methods are commonly used, including the stopping rules proposed by Pocock (1977), O'Brien and Fleming (1979), Whitehead and Stratton (1983), and the error spending function approach of Lan and DeMets (1983). However, in order to be implemented in a flexible manner, all methods require an estimate of the amount of statistical information present at each analysis time. This estimated information can then be used to generate appropriate stopping boundaries at any particular analysis time.

In order to make correct statistical inference in a study, it is important that the true information growth be accurately modeled. Failure to do so may lead to grossly incorrect type I and type II errors. In settings where only one outcome measurement is obtained on each individual, the proportionate information at a particular analysis time is often calculated as a ratio of the number of measurements at the analysis time to the total expected

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

number of measurements; in the case of survival data, the current information is the number of events observed relative to the total expected. However, even when the proportionate information can be correctly computed, problems can arise due to the need to estimate nuisance parameters affecting the variability of measures of treatment effect. Burington and Emerson (2003) noted that imprecision of the estimated nuisance parameters can lead to error spending boundaries that do not reflect the true known proportionate information available at each analysis, while boundaries constrained on other scales will not necessarily adhere to the desired boundary shape function. Several authors have further conjectured that the imprecision inherent in estimating within group variances or baseline event rates at the earliest of interim analyses might lead to a spurious appearance of nonmonotonic information growth during the monitoring of a study (Scharfstein et al., 1997; Burington and Emerson, 2003) Such situations are likely rare in practice due to the relatively large increments of information typically accrued between successive analyses: the monotonic increase in available data overwhelms the potential nonmonotonicity in the estimates of the nuisance parameters across the analyses.

Further complications arise when treatment effects are measured by a contrast across study time in a longitudinal clinical trial. In that setting, the number of available measurements relative to the final expected number of measurements overestimates the current statistical information, even when the true effect is linear in time and the data are homoscedastic (Wu and Lan, 1992). This argues that a naive approach to estimating information based solely on sample size is problematic. However, a larger concern would be settings in which the true information growth might be nonmonotonic in that the variability of the estimated treatment effect was higher at a later interim analysis than it was at one conducted earlier. Were that to occur, standard stopping boundaries for group sequential trials would not be

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

appropriate because the assumption of independent increments is violated.

Previous authors (Scharfstein et al., 1997; Jennison and Turnbull, 1997) showed that using the efficient estimator leads to an independent increment structure in a group sequential trial and that the information growth therefore must be monotonic. They note that using an inefficient estimator does not preclude independent increments, but conjecture that it may lead to nonmonotonicity in some circumstances. However, Scharfstein et al. (1997) speculate that such nonmonotonicities are rare and only arise in practice due to estimation of the information growth as discussed above. In this manuscript, we expand on their work by presenting situations in which the information growth is nonmonotonic in truth.

For this paper, we restrict attention to the case where the statistic of interest is of a linear contrast over time. First, we consider the case of independent longitudinal data and investigate the consequences of inaccurately estimating the information growth in a clinical trial. We then consider the case of correlated longitudinal data and discuss issues regarding information growth when using generalized estimating equations (GEE) in this setting including scenarios leading to nonmonotonic information growth.

2 Notation and Group Sequential Methods

In a group sequential trial testing a null hypothesis of $H_0 : \theta = \theta_0$ against a one-sided alternative, a stopping rule is defined over a schedule of analyses, occurring at times t_1, ..., t_J, where J is the maximal number of analyses. These stopping rules are typically defined in terms of a continuation set, $C_j = (a_j, b_j] \cup [c_j, d_j)$, with $-\infty \leq a_j \leq b_j \leq c_j \leq d_j \leq \infty$. If the test statistic, S_j is contained in the continuation set, C_j, the trial continues to the next

COBRA
Collection of Biostatistics
Research Archive

analysis time, t_{j+1}. By defining C_J as the empty set the trial is assured of having no more than J analyses. These continuation sets are determined in part by the amount of statistical information at each analysis time, relative to the amount of information expected at the end of the trial. This will be referred to as the information growth over the course of the study, which we compute as $\frac{1/Var(\hat{\theta}_j)}{1/Var(\hat{\theta}_J)}$ and denote by π_j for analysis time t_j.

Using the unified family approach (Kittelson and Emerson, 1999), the boundaries are determined by the following formula, with parameters A, P, and R specified to determine the behavior of the boundaries at possible early termination points (the $*$ denotes a, b, c, or d).

$$\nu_*(\pi; A, P, R, G) = (A_* + \pi_j^{-P_*}(1 - \pi_j)^{R_*})G_* \tag{1}$$

The parameter G is found by iterative search to obtain desired operating characteristics. These operating characteristics usually focus on the type I and type II statistical errors as well as some measure of "conservatism" at the early analyses. Specifically this conservatism is often considered as testing for the treatment effect of interest with a smaller type I error rate, such as 0.005 before continuing the study. We note that this conservatism could also be thought of as increasing the size of the effect for which efficacy would be declared, i.e. this early test has a boundary that would lead to a 95% confidence interval suggestive of a larger effect than would the 95% confidence interval at the end of the study. In any case, it is necessary to weigh the scientific and clinical implications of the stopping rules when determining boundaries. If a trial were to stop early, additional information that might be of interest, such as being able to characterize long-term effects or gaining increased knowledge of potential safety issues, would not be obtained.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

In the unified family approach above, $A_* = 0$, $P_* \geq 0$, $R_* = 0$ corresponds to the Wang and Tsiatis (1987) one-parameter boundaries. In this parameterization, increasing values of P_* lead to decreased early conservatism. In the special case $P_* = 0.5$ these boundaries correspond to the Pocock (1977) boundaries, if $P_* = 0$, the boundaries correspond to those of O'Brien and Fleming (1979), and if $P_* = \infty$, there is no early stopping.

In most trials where one outcome measurement is made on each individual accrued, $\pi_j = \frac{N_j}{N_J}$, the ratio of the number of measurements at the time t_j to the number that would be available at the end of the trial. If the primary outcome is survival, then usually $\pi_j = \frac{D_j}{D_J}$, the ratio of the number of events observed by time t_j to the number expected at the end of the study. However, neither of these holds in the longitudinal case, which is the setting for this manuscript.

3 Information Growth with a Linear Model

We consider the case where the treatment effect of interest is the change over time, for example the difference in rates of tumor growth between a placebo and a treatment group. For now, we focus on the case where there is no correlation in the data. One example in which this might occur is if the population being studied over time was a group of rats with the same treatment and same starting point for the disease and then at each time point a single rat was selected for analysis. Independent data are obtained when the analysis at each time point cannot be performed on the same rat over time. In the case of the tumor growth example, measuring the tumor at a particular time point requires an autopsy and thus each rat is only measured once. In the models below, we use x to denote the time from

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

randomization and we reserve reference to time to denote analysis times (t_j) in the study. For simplicity we restrict attention to a one-sample model; in a randomized controlled trial both groups would be analyzed similarly.

Let the outcome measurement Y_{ik} be the observation for cluster i at some time x_k from randomization. We assume measurements are known to be independent across clusters. The total number of measurements expected to be observed over the course of the study is thus $I * K$ where I denotes the number of clusters observed and K denotes the number of observations that will be made on each cluster over the duration of the study. In the example of tumor growth in rats, a cluster would be a single rat. Each observation Y_{ik} would be made when one of the rats was sacrificed to measure the tumor growth. In other studies, a cluster could simply be one participant measured at K time points from the time that participant was randomized.

In this model the regression formula is:

$$E(Y|X = x) = \beta_0 + \beta_1 * x. \tag{2}$$

The parameter of interest β_1 is the change in the outcome (Y) over time from randomization (x). Consider initially the case when the true treatment effect corresponds exactly to the analysis method used; data are homoscedastic and the treatment effect (slope) is constant over the length of the follow up time. In this setting the standard error of $\hat{\beta}_1$ can be calculated exactly. For a particular analysis time t_j, the standard error depends on the number of measurements observed by that point in the study, the variance of the outcome at

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

any fixed point in time, and the variance of the predictor variable (time from randomization):

$$Var_j(\hat{\beta}_1) = \frac{\sigma^2_{y|x}}{n_j * Var_j(x)}. \tag{3}$$

Thus, the estimate of the information growth $(\frac{1/Var(\hat{\theta}_j)}{1/Var(\hat{\theta}_J)})$ that relies only on the fraction $(\frac{N_j}{N_J})$ of the total measurements made does not account for the fact that the variance of the time from randomization $(Var(x))$ is also increasing with study time $(Var_j(x) \leq Var_J(x))$. Therefore, simply using the proportion of expected total measurements as a surrogate for the information growth will overestimate the true information growth.

Figure 1 shows the fraction of information present as a function of study time for different accrual periods. Here there are 10 observed time points (i.e., k=1,...,10) with measurements taken at baseline and at each month thereafter, (i.e., $x_1 = 0$, $x_2 = 1$, ..., $x_{10} = 9$). The accrual periods vary from nearly instantaneous (0.1 month accrual) to long relative to the length of follow up (30 months). In all cases, using the proportion of expected measurements as a surrogate for the amount of information (dashed line) overestimates the actual amount of information present. Qualitatively, this overestimation is most problematic when the accrual time is short relative to the length of follow up. If the accrual period is long relative to the length of follow up on a cluster, the true information curve is close to the information curve approximated by the relative number of measurements. Intuitively, this is reasonable. In an extreme case where all measurements for the study are made on the last recruited cluster prior to accrual of a new cluster, the increase in information would almost entirely be due to the next measurement obtained, rather than increasing variability of the x measurements.

In this situation estimating the information growth using the model-based standard errors yields correct estimates of the information at various analysis times; however simply using

COBRA
Collection of Biostatistics
Research Archive

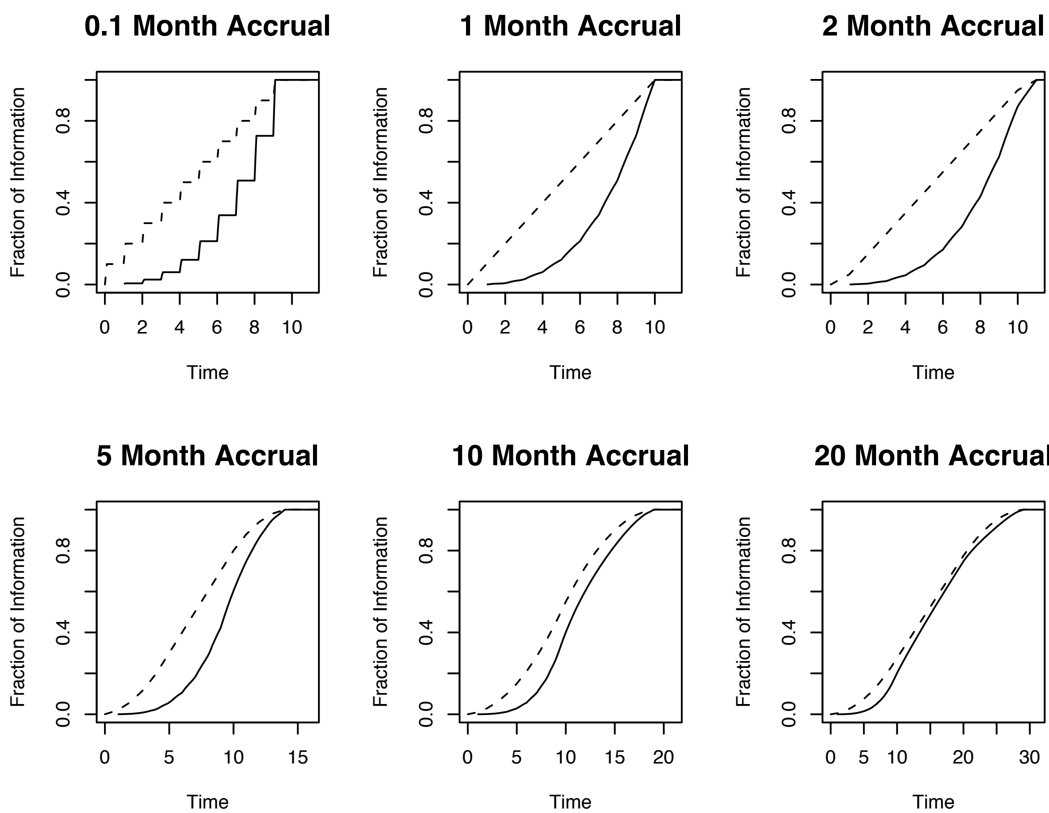

Figure 1: Plots showing the true information growth (solid line) relative to the information growth that would be estimated from the fraction of the total number of measurements (dashed line). In all cases, estimating the IG by the number of measurements overestimates the true information.

the number of measurements made does not. Using the incorrect estimates of the information to form boundaries can result in incorrect type I errors. This will be discussed in greater detail in the next section.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

4 Impact of Incorrect Information Growth Estimates

4.1 Incorrect Type I Error

If the relative amount of information is overestimated at an interim analysis time t_j, the type I error will be inflated. This is a result of overstating the precision with which our parameter of interest is known, thus setting the stopping boundaries too narrow for the analysis at time t_j. Therefore, under the null, the estimate at this time point is outside of these boundaries more often than expected and the type I error is inflated.

To illustrate, consider 10 measurements taken at times 0-9, with an accrual period of 2 months (i.e., the calendar time for all subjects to complete the study is 11 months). Assume that four analyses are planned to be evenly spaced in calendar time (i.e., every 2.75 months), and analyses are to be conducted using either the stopping boundary shape described by Pocock (1977) or O'Brien and Fleming (1979). We consider single boundary stopping rules that allow for early stopping only under the alternative, and two boundary stopping rules that allow for early stopping for futility as well. The designs are constructed to maintain a fixed two-sided 0.05 type I error rate and to have 97.5% power for a specific alternative when using the true information growth.

Table 1 shows the dramatic increase in type I error when using the naive information growth estimates in this setting. The boundaries are constructed to maintain a fixed 0.05 error rate, yet the single boundary type I error is 0.34 using the Pocock boundary and 0.21 using an O'Brien-Fleming boundary. For two boundary designs, the type I errors are 0.31 and 0.20, respectively. In this case, the estimated information is greater than the

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Table 1: Stopping probabilities (SP) under the null and the alternative with 97.5% power using Pocock and O'Brien-Fleming (OBF) stopping boundaries.

Naive IG									
		Single Boundary				Two Boundary			
Calendar		Pocock		OBF		Pocock		OBF	
Time	Naive IG	SP_{null}	SP_{alt}	SP_{null}	SP_{alt}	SP_{null}	SP_{alt}	SP_{null}	SP_{alt}
0.25	0.225	0.274	0.452	0.138	0.270	0.277	0.499	0.141	0.274
0.50	0.50	0.056	0.274	0.046	0.308	0.028	0.163	0.040	0.271
0.75	0.775	0.011	0.181	0.018	0.298	0.003	0.042	0.012	0.206
1.00	1.00	0.002	0.072	0.008	0.107	<0.001	0.003	0.003	0.047
Overall power		0.34	0.978	0.21	0.983	0.31	0.71	0.20	0.80
True IG									
		Single Boundary				Two Boundary			
Calendar		Pocock		OBF		Pocock		OBF	
Time	True IG	SP_{null}	SP_{alt}	SP_{null}	SP_{alt}	SP_{null}	SP_{alt}	SP_{null}	SP_{alt}
0.25	0.015	0.007	0.027	<0.001	<0.001	0.007	0.032	<0.001	<0.001
0.50	0.14	0.007	0.187	<0.001	<0.001	0.007	0.244	<0.001	<0.001
0.75	0.48	0.006	0.521	0.002	0.456	0.006	0.557	0.002	0.462
1.00	1.00	0.005	0.266	0.023	0.519	0.004	0.142	0.023	0.513
Overall power		0.025	0.975	0.025	0.975	0.025	0.975	0.025	0.975

true information, which causes the boundaries at interim analyses to be too narrow. This results in some null trials being declared effective when they would not have been if stopping boundaries constructed with the correct information had been used. This can be seen by noting the difference in early stopping probabilities at each analysis under the naive estimates of the information and under the true information when the null hypothesis is true. The inflation of the type I errors is slightly less for designs with with both efficacy and futility boundaries, as some trials are stopped prematurely early for futility, thus preventing these trials from contributing to the type I error.

COBRA
Collection of Biostatistics
Research Archive

4.2 Loss of Power

When using a single efficacy stopping boundary, there is a very slight increase in power under the alternative because the interim boundaries are closer to the null and these make it more likely for trials to be declared effective. When using both efficacy and futility boundaries, in addition to the inflation of the type I error, there is also a loss of power due to the overestimated information. Table 1 also shows that for the alternative with 97.5% power under the true information growth, the power is only 71% using Pocock boundaries and 80% using O'Brien-Fleming. This is again due to the overestimated information causing the interim boundaries to be too narrow; in this case the boundary for futility causes some trials that would eventually reject the null to be stopped early for futility.

4.3 Nonmonotonic Boundaries

If the estimated information growth is nonmonotonic, the stopping boundaries can be nonmonotonic as well. In extreme cases, an estimated or true nonmonotonicity in the information growth can lead to boundaries that preclude stopping at the interim analysis entirely. This will be further explored in the next two sections that deal with heteroscedasticity and correlated observations, respectively.

5 Information Growth and Heteroscedastic Data

We next examine the case in which the measurements are heteroscedastic, such as if the measurements were becoming increasingly variable over time, but that the analysis model

COBRA
Collection of Biostatistics
Research Archive

(erroneously) presumes homoscedasticity. This might occur if there were rigid entry criteria for the study at baseline, such as a systolic blood pressure measurement between 130 and 140 and we were interested in measuring it over time. In one such case with increasing variability, the data might be generated from a model such as the one below where x again is the time from randomization.

$$Y_{ix} \sim (\mu_x, \sigma^2(x+1)^\gamma) \tag{4}$$

As a consequence of this increasing variability, both the true information growth and the analysis model-based estimate of the information growth can be nonmonotonic (figure 2). If the effect is truly linear but later measurements are much more variable than early ones, the later measurements will detract from the precision with which we can estimate the slope. However, even if this extreme case is true, there may be more pressing reasons to continue with the later measurements, such as providing evidence that the linear trend exists (or does not exist) over a longer range of time.

In figure 2 the accrual pattern and measurement schedule match those from the previous section (2 month accrual, measurements at baseline and months 1-9 thereafter). The heteroscedasticity was generated as in equation (4), with $\sigma^2 = 1$ and $\gamma = 2$. The analysis model-based estimates of the information growth from ordinary least squares regression assuming constant variance are different from the true information growth obtained by simulation. For the true information growth, when the first measurement that is more variable occurs, the information may in fact drop. Then, as more measurements that are more variable are accrued during this interval, the information increases, as would be expected with increasing the number of observations. The initial drop in information is in part attributable

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

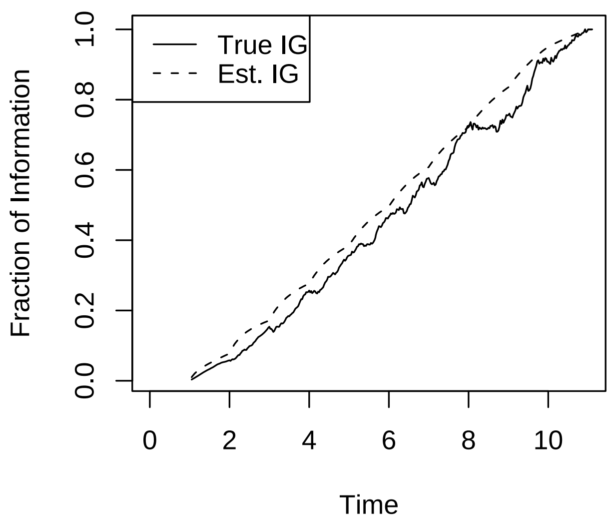

Figure 2: Plot illustrating information growth when the data are truly heteroscedastic. The solid line represents the true, nonmonotonic information growth, simulated empirically. The dashed line represents the model-based estimates of the information growth. The above was simulated assuming $\sigma = 1$ and $\gamma = 2$ in a model like that of equation (4).

to the amount of influence the first points can have on the estimated slope. As the first more variable measurements are of high influence, the actual variability of the $\hat{\beta}_1$ parameter is increased when these first points are added. When balance is achieved, such as at the end of a study with no dropout and no missing measurements, heteroscedascitiy is less of a concern, because no point is overly influential.

In contrast, the analysis model-based estimates of the information growth are constrained by the assumption of constant variance. Thus, when adding points that are more variable, the estimate of the constant variance is only slightly altered by a few measurements with more variability. The estimate of the variance increases as more measurements with increased variability are added, so the estimate of the constant variance is highest only after all the

COBRA
Collection of Biostatistics
Research Archive

additional measurements have been taken. This is usually offset by gains in the number of measurements leading to more typical behavior of the information growth curve. However, it can lead to dramatic cases in which the estimated information is lowest after all of the more variable measurements have been accrued if the later measurements are much more variable than the earlier ones (e.g., if γ is extremely large in the above data generation model).

5.1 Consequences of Incorrect Information Growth Estimates

As before, overestimating the information growth at a particular analysis point (such as using the model-based estimate of the information growth in this setting) can lead to incorrect type I error. In the extreme cases where the true information growth is nonmonotonic, and the estimated information at a planned analysis time is less than the information at a previous analysis time (i.e., $\pi_j > \pi_k; j < k$), one obvious solution is to not do an analysis at this time point. Certainly, the standard software available for conducting group sequential analyses cannot be used in such a circumstance, and in the case of possible nonmonotonicity due to heteroscedasticity, it is expected that the information will improve by the end of the trial (when balance is again obtained). Therefore, if extreme heteroscedascity is a concern, it would make sense to plan for interim analyses at times where as much balance as possible will be attained.

6 Information Growth with Correlated Observations

We now turn attention to the case where multiple longitudinal measurements are correlated, such as when measuring tumor growth in an individual over time. One approach to such

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

data is to use generalized estimating equations (GEE) to estimate the treatment effect over time (Liang and Zeger, 1986).

When GEE is used, a "working" covariance matrix structure is specified to use as weights in the estimation of the parameters. Common choices for the specification of the working covariance matrix $W(\rho)$ include independence, exchangeable ($W_{ij} = \rho, i \neq j$), and auto regressive with order one (AR(1); $W_{ij} = \rho^{|i-j|}$). A completely unstructured working covariance matrix can also be used; in this case all off diagonal elements of the matrix are estimated separately. If the working covariance matrix is not independence, an iterative process is used to solve for $W(\hat{\rho})$ and $\hat{\beta}$. In most cases, the estimates of $\hat{\beta}$ will be unbiased regardless of the choice of working covariance, however in certain circumstances with time-varying covariates, the estimates may be biased if working independence is not used (Pepe and Anderson, 1994). When the working covariance matrix is close to the truth, the estimates will be most efficient (Wang and Carey, 2003). However, using simple forms of the working covariance matrix may be justified in many situations due to a lack of knowledge of the true correlation structure, a desire to be robust to possible misspecification of the correlation or the linear model, or the convenience in estimation when a simple structure of the working covariance matrix is used.

In GEE, the regression parameter estimates are given by:

$$\hat{\beta} = (X^T W^{-1} X)^{-1} X^T W^{-1} Y. \tag{5}$$

The estimated standard errors then account for the correlation between observations made on the same cluster, where V is the true correlation between clusters (usually estimated with

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

the empirical version in GEE):

$$Var(\hat{\beta}) = (X^T W^{-1} X)^{-1} X^T W^{-1} \hat{V} W^{-1} X (X^T W^{-1} X)^{-1} \quad (6)$$

If the correlation structure is known, then letting $W = V$ makes the GEE and GLS estimates equivalent, and this estimator is the best linear unbiased estimator by the Gauss-Markov theorem. If the true correlation structure is unknown but the form of the working covariance matrix is correctly specified such that $W \rightarrow_p V$, then this estimator will be asymptotically efficient. If the working covariance matrix is independence, then the point estimates for $\hat{\beta}$ match those that would be obtained using ordinary least squares regression (OLS) exactly. If not, point estimates for $\hat{\beta}$ will be like those that would have been obtained using GLS with a similar correlation structure.

As noted previously, the amount of additional information obtained between interim analysis times during a longitudinal clinical trial depends on several factors, including the number of additional measurements and the increased variability in the predictor variable (time since randomization). With correlated measurements the amount of additional information will also depend on the degree of correlation between the new measurements and the previous measurements. In the extreme case of nearly perfectly correlated data and no new individuals (clusters), there is almost no new statistical information obtained, even as additional measurements are collected, because the new measurements are almost completely determined by knowledge of the high correlation and the assumed linear model. If the correlation between the new measurements and the old ones is not correctly accounted for in the estimate of $\hat{\beta}$, problems can arise, specifically nonmonotonic information growth when the addition of measurements results in increased variability of the point estimate.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Using GEE to estimate the longitudinal treatment effect can lead to nonmonotonic information growth curves. This can occur even when everything is correctly specified: the linear contrast is exactly correct, the data are homoscedastic, the clusters are correctly identified, and the "robust" standard error estimates are derived using the sandwich estimator. One situation in which this can occur is when the design is unbalanced. For instance, at an interim analysis during the conduct of a clinical trial, some individuals may have had three total measurements while other individuals have only had two. Previous authors have noted that using an independence working covariance matrix can lead to relative inefficiency in this setting compared to using a working covariance matrix that matches the form of the true data (Wang and Carey, 2003) . We note that using an incorrectly specified working covariance matrix (and thus inefficient weights) can lead to absolute inefficiencies as well. Such absolute inefficiency can lead to nonmonotonic information growth curves: In certain situations there is more statistical information when everyone has just two measurements than when everyone has two measurements and a handful of individuals have three measurements.

Intuition might suggest that the reason for this absolute inefficiency is due to the weighting on the measurements when determining the estimate of $\hat{\beta}$. For example, we noted previously that if an independence working covariance matrix is used, the estimate of $\hat{\beta}_1$ will match exactly the estimate that would have been obtained through OLS regression ignoring the correlation within an individual. When all individuals have the same number of measurements at the same time points from randomization, all subjects are weighted equally and this does not generally result in a great loss of efficiency. However, this is not true when a few individuals have more measurements than the others. Compared to the case of all independent measurements, the line fit with just two observations on each subject when those observations are highly correlated is much less variable (there is a gain in information

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

due to the positive correlation within an individual). If only a handful of these highly correlated subjects have measurements at a more extreme time point, these subjects have greater influence on the slope (as if they were new, independent measurements), and the variability of the slope increases due to chance selection of different measurements over hypothetical repeated experiments. This can actually increase the true variability of the slope, unless the correlation with other measurements is properly accounted for by downweighting the additional observations (using the working covariance matrix) relative to the weights that would be used in OLS regression.

Figure 3 shows the true information growth curves under situations where the true effect is linear, the data are homoscedastic, and the correlation within individuals is high. To be consistent with our prior setting, 10 measurements were made on each individual, one at baseline, and one at each of nine follow up times. For an example of high within individual correlation, we chose an AR(1) structure with $\rho = 0.95$. In an attempt to make the exchangeable correlation structure as equivalent as possible, ρ for the exchangeable case was chosen such that the average correlation between all pairs of measurements on an individual at the end of the study would be equal to that in the AR(1) structure ($\rho = 0.8338$). Finally, to ensure comparability, the number of individuals in the AR(1) case was increased such that the final amount of statistical information was equivalent between the AR(1) and exchangeable cases (2170 and 500 individuals, respectively). Four working covariance matrices were used in each simulation: independence, exchangeable, AR(1), and unstructured. The plots demonstrate the nonmonotonic behavior using the independence working covariance matrix in this setting. The scaled plots show the relative loss of efficiency compared to using the correctly specified form of the working covariance matrix. In this setting, using an exchangeable working covariance matrix appears to be most desirable; it does lose some

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

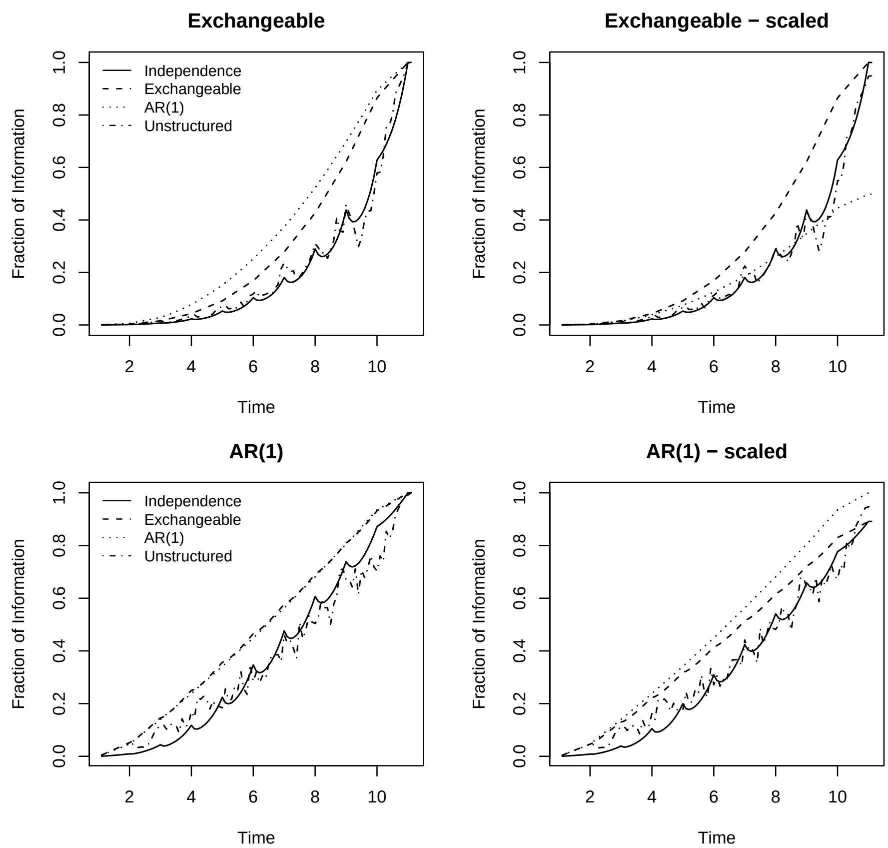

Figure 3: Plot illustrating information growth over time using GEE when the data are truly linear. The true correlation structure is either exchangeable or AR(1) and the plots show the information growth using each of four working covariance matrices. The scaled graphs show the true information growth relative to the amount of information when the working covariance matrix is exactly specified.

efficiency relative to using AR(1) when the truth is AR(1) (relative efficiency = 89%), but does not become nonmonotonic. In contrast, when the true data are exchangeable, using a working AR(1) structure leads to a dramatic drop in efficiency (relative efficiency = 50%).

Some authors have suggested the use of an "unstructured" working covariance matrix with GEE to provide nearly efficient estimation without pre-specifying the form of the working covariance (Gange and DeMets, 1996). Others have suggested that using a working

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

covariance matrix with consistently estimated parameters (even if misspecified) will lead to nearly independent increments (Lee et al., 1996). Using an "unstructured" working covariance matrix does lead to nearly efficient estimation when the design is balanced (when all subjects have equal numbers of measurements), however in preliminary investigations, it performs poorly when the design is markedly unbalanced, yielding results similar to using an independence working covariance matrix (figure 3). These simulations were done using the `geepack` package in R (Yan and Fine, 2004). When the design is unbalanced, some estimated parameters in the working covariance structure appear to be quite variable due to only a few observations contributing to those estimates. For this reason, using the unstructured working covariance can lead to many of the same problems as using the independence working covariance, which suggests that in most circumstances the use of the exchangeable working covariance matrix would be preferred. In addition, in a small number of simulations (approximately 3%) using an unstructured working correlation matrix meant that the GEE estimates did not converge. These cases were excluded from our estimates, and hence the graphs underestimate the true magnitude of the problem.

As might be expected, the degree of correlation within measurements on the same individual affects the true information growth when using an independence working covariance matrix (figure 4). When the true data are exchangeable with low correlation ($\rho = 0.3$) and the same study design as before (2 month accrual, 10 measurements per individual), the information growth is nearly the same between the exchangeable and independence working covariance matrices (figure 4A). As the correlation increases, using working independence becomes less efficient at interim points in the trial and can lead to nonmonotonic information growth (figure 4: A-C).

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

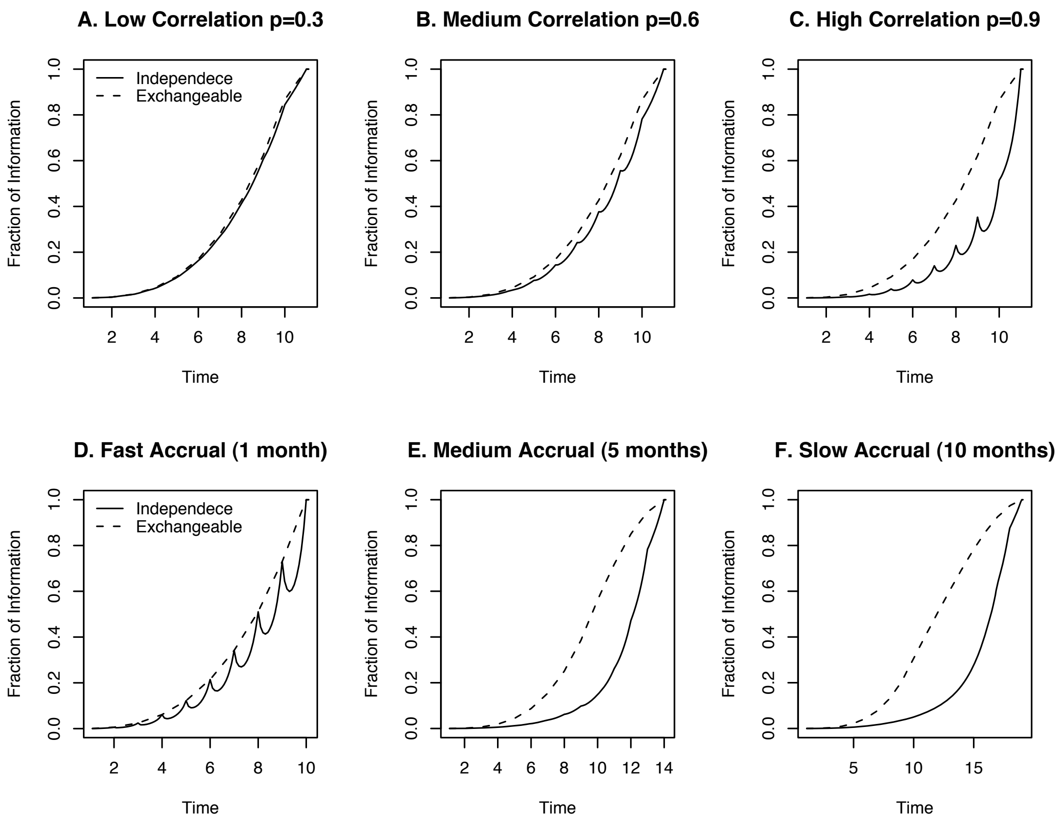

Figure 4: Plots illustrating the effect of the within individual correlation and the accrual pattern on the information growth over time using GEE. In all cases, the data are truly linear the covariance within individuals has an exchangeable structure, and 10 measurements are made on each individual (at baseline and months 1-9). For plots A-C, accrual was fixed at 2 months and for plots D-F the correlation was fixed at $\rho = 0.8338$.

When the design is completely balanced (as might occur at the end of a study with no dropout), the estimates using independence and exchangeable working covariance matrices are the same. Such balance may also be achieved during a study if the accrual period is shorter than the time between consecutive measurements on an individual. In our example where individuals are measured every month, this would occur if everyone were accrued within one month (e.g., every individual has a first measurement before anyone has a second as in figure 4D). However, when the design is far from balanced (as might occur during a long

COBRA
Collection of Biostatistics
Research Archive

accrual period), working independence will be noticeably less efficient than exchangeable at interim points in the study when the true data are exchangeable (e.g. figure 4F).

A long accrual period when using an independence working covariance matrix leads to relative inefficiency, but does not tend to lead to noticeable nonmonotonic information growth. Nonmonotonicity is most pronounced when the accrual period is short relative to the follow up on each individual and if the correlation within an individual is high (figure 4D). Consider a case of high within subject correlation ($\rho = 0.8338$) and short accrual (so that all individuals have two measurements before anyone has a third). In this situation, the amount of statistical information decreases when the first individual gets at third measurement, and continues to decrease until slightly more than 10% of the study population has a third measurement. The amount of information present when everyone had two measurements but no one had a third is not surpassed until more than 50% of the new third measurements are obtained. This becomes even more striking as the study continues. When everyone has nine measurements but no one yet has ten, the amount of statistical information decreases when the first person gets a tenth measurement and continues to decrease until approximately 30% have a tenth measurement. The amount of information is not greater than the amount when no one had a tenth until 70% have a tenth measurement.

6.1 Consequences of Nonmonotonic Information Growth

GEE can have nonmonotonic information growth in truth, but the information growth is well estimated in these cases by the so-called "robust" standard errors computed using the sandwich estimator. However, some of the problems of nonmonotonic information growth discussed in the previous section remain. Specifically, analyses should be planned to take

COBRA
Collection of Biostatistics
Research Archive

place at points in the study where the design will have as much balance as possible to avoid situations where nonmonotonic information growth may potentially occur. Using an exchangeable working covariance will also be helpful in avoiding potentially nonmonotonic information growth while preserving most of the efficiency even if the true correlation structure is AR(1).

7 Conclusions/Discussion

There are different considerations for the planning of analyses in a group sequential clinical trial than for planning a trial with only one analysis time. In particular, for a group sequential clinical trial the rules determining the schedule of analyses must be completely pre-specified and the behavior and estimation of the information growth over the course of the study must be considered. We have demonstrated that poorly estimated information growth can lead to substantial inflation of type I error and loss of power, and we have also shown that when using GEE in certain circumstances the true information growth may be nonmonotonic.

If nonmonotonicity were to occur in a trial, it could be ignored, by refusing to proceed with the scheduled interim analysis. It could also be avoided by moving the interim analysis time to the point in the trial where the maximum information thus far in the trial was achieved. However, in both of these scenarios the planned interim analysis is not being performed. In many cases, we may not truly believe that the data are exactly linear, and may be interested in a linear contrast over time. If the planned analysis is not done due to nonmonotonic information growth, the linear contrast intended to be estimated at that time point is not assessed, and any ethical and efficiency concerns that motivated the use of

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

a stopping rule are not being addressed. Furthermore, the nonmontonic information growth is clear evidence of a violation of the independent increments assumption, and application of stopping boundaries derived under that assumption may be problematic.

It should be noted that the pathological behavior of the information growth was observed in extreme cases with unusually high correlation between observations. However, it is nonetheless important to maintain the correct type I error by using the correct information growth. In the case of GEE, using the exchangeable working covariance will tend to avoid the possibility of nonmonotonic information growth and would seem to preserve most of the efficiency, even when the true correlation structure is not exchangeable. This paper has focused on the effects of poorly estimated covariance when the mean model is exactly correct; future work will investigate the effects of model misspecification on the estimated information growth in these longitudinal settings.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

References

Burington, B. E. and Emerson, S. S. (2003). Flexible implementations of group sequential stopping rules using constrained boundaries. *Biometrics* **59,** 770–777.

Gange, S. J. and DeMets, D. L. (1996). Sequential monitoring of clinical trials with correlated responses. *Biometrika* **83,** 157–167.

Jennison, C. and Turnbull, B. W. (1997). Group sequential analysis incorporating covariate information. *Journal of the American Statistical Association* **92,** 1330–1341.

Kittelson, J. M. and Emerson, S. S. (1999). A unifying family of group sequential test designs. *Biometrics* **55,** 874–882.

Kittelson, J. M., Sharples, K., and Emerson, S. S. (2005). Group sequential clinical trials for longitudinal data with analyses using summary statistics. *Statistics in Medicine* **24,** 2457–2475.

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

Lee, S. J., Kim, K., and Tsiatis, A. A. (1996). Repeated significance testing in longitudinal clinical trials. *Biometrika* **83,** 779–789.

Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalizeed linear models. *Biometrika* **73,** 13–22.

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

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/uwbiostat/paper358

Pepe, M. and Anderson, G. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics Simulation and Computation* **23,** 939–951.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64,** 191–200.

Scharfstein, D. O., Tsiatis, A. A., and Robins, J. M. (1997). Semiparametric efficiency and its implications on the design and analysis of group sequential studies. *Journal of the American Statistical Association* **92,** 1342–1350.

Wang, S. K. and Tsiatis, A. A. (1987). Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics* **43,** 193–199.

Wang, Y. and Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations. *Biometrika* **99,** 29–41.

Whitehead, J. and Stratton, I. (1983). Group sequential clinical trials with triangular continuation regions (corr: V39 p1137). *Biometrics* **39,** 227–236.

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

Yan, J. and Fine, J. (2004). Estimating equations for association structures. *Statistics in Medicine* **23,** 859–880.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press