Published in final edited form as:

J Biopharm Stat. 2017 ; 27(1): 135–147. doi:10.1080/10543406.2016.1167071.

More data, less information? Potential for nonmonotonic information growth using GEE

Abigail B. Shoben, PhD,

Division of Biostatistics, The Ohio State University, Columbus, OH

Kyle D. Rudser, PhD, and

Division of Biostatistics, University of Minnesota, Minneapolis, MN

Scott S. Emerson, MD, PhD

Department of Biostatistics, University of Washington, Seattle, WA

Abstract

Statistical intuition suggests that increasing the total number of observations available for analysis should increase the precision with which parameters can be estimated. Such monotonic growth of statistical information is of particular importance when data are analyzed sequentially, such as in confirmatory clinical trials. However, monotonic information growth is not always guaranteed, even when using a valid, but inefficient estimator. In this paper we demonstrate the theoretical possibility of nonmonotonic information growth when using generalized estimating equations (GEE) to estimate a slope and provide intuition for why this possibility exists. We use theoretical and simulation-based results to characterize situations that may result in nonmonotonic information growth. Nonmonotonic information growth is most likely to occur when (1) accrual is fast relative to follow-up on each individual, (2) correlation among measurements from the same individual is high, and (3) measurements are becoming more variable further from randomization. In situations that may lead to nonmonotonic information growth, study designers should plan interim analyses to avoid situations most likely to result in nonmonotonic information growth.

Keywords

Group sequential trials; Information growth; Longitudinal data

1 Introduction

In many clinical trials, interim analyses are desirable for ethical and financial reasons, however, specialized sequential methods are needed to handle the additional statistical issues that arise from multiple analysis times (Whitehead, 1997; Jennison and Turnbull, 2000). In particular, several methods are commonly used to maintain the nominal type I error rate, including the group sequential stopping rules proposed by Pocock (1977), O'Brien and Fleming (1979), Whitehead and Stratton (1983), and Lan and DeMets (1983). In their

implementation, all of these methods require an estimate of the proportion of statistical information at each analysis time relative to the final expected information, and the implicit assumption is that the amount of statistical information grows monotonically during the study as more data are acquired.

We consider here the sequential analysis of longitudinal clinical trials that have repeated measurements of the clinical outcomes over the time of follow-up. Such trial designs are common when studying outcomes of tumor growth in cancer (Stein et al., 2011), CD4+ cell counts in HIV / AIDS (Pantazis et al., 2008), or cognitive decline in Alzheimer's Disease (Aisen et al., 2008). For example, a recently completed trial of tolvaptan in autosomal dominant polycystic kidney disease considered the effect of treatment over time on both a primary endpoint of kidney size by MRI and a secondary endpoint of estimated glomerular filtration rate as estimated from serum creatinine (Torres et al., 2012). Patients were randomized in a 2:1 ratio to receive tolvaptan or placebo, with protocol specified MRI imaging of kidney size annually and assessment of serum creatinine every four months over a three year period of treatment. In such a trial, several methods of summarizing the treatment effect over time are possible, including the magnitude of the endpoint at the final measurement, the time averaged measurement over time, and the rate of change (slope) over time. All of these are special cases of a weighted area under the curve approach (Kittelson et al., 2005). In this paper we consider measures such as those chosen in the tolvaptan trial and focus specifically on the rate of change over time computed using all available measurements.

Compared to clinical trials with only a single measurement at a single observation time, estimating the information growth in longitudinal clinical trials is more complicated, even if the data are independent and homoscedastic. The simplistic approach of estimating the information growth as the ratio of the number of available measurements relative to the final expected number of measurements overestimates the current statistical information in such a setting (Wu and Lan, 1992). The expected correlation among observations from the same individual and the potential for heteroscedastic data add additional difficulty when estimating information growth in the typical longitudinal study. Potential correlation and heteroscedasticity lead to the possibility of nonmonotonic information growth where the variance of the estimated treatment effect could be larger at some later interim analyses than it was at a prior analysis.

The possibility of nonmonotonic information growth has been previously considered. In group sequential trials with a single measurement made on each individual, several authors have conjectured that the imprecision inherent in estimating within group variances or baseline event rates at the earliest of interim analyses might lead to boundaries that do not reflect the true known proportionate information available at each analysis, boundaries that do not necessarily adhere to the desired shape function, and even a spurious appearance of nonmonotonic information growth during the monitoring of a study (Scharfstein et al., 1997; Burington and Emerson, 2003). However, those authors noted that situations truly leading to nonmonotonic information growth are likely rare in practice due to the relatively large increments of information typically accrued between successive analyses: the monotonic

increase in available data usually overwhelms the potential non-monotonicity due to estimates of the nuisance parameters across the analyses.

Furthermore, existing work has shown that using the efficient estimator leads to an independent increment structure (and thus monotonic information growth) in a group sequential trial (Scharfstein et al., 1997; Jennison and Turnbull, 1997). Therefore longitudinal trials that use the efficient estimator must have monotonic information growth. However, using the efficient estimator for longitudinal trials may be impossible in practice due to uncertainty in the true probability distribution for the endpoint measurements. Regulatory agencies require that analysis plans for phase 3 trials be fully prespecified, so many designs opt for analysis methods that are robust to many possible circumstances, including nonlinearities of treatment effect over time and various potential correlation structures. Consequences of using such inefficient, but useful estimators, have not been fully explored. Even in the longitudinal case, others have speculated that such nonmonotonicities would be rare and arise only in practice due to estimation of the information growth with nuisance parameters (Scharfstein et al., 1997). In this paper, we expand on that work by documenting situations in which the information growth is truly nonmonotonic.

We restrict our attention to the case where the statistic of interest is a linear contrast over time (slope) estimated using a generalized estimating equations (GEE) approach (Liang and Zeger, 1986). We illustrate with theory and examples the possibility of nonmonotonic information growth when GEE is used. We then describe the implications of nonmonotonic information growth and provide suggestions for avoiding it in practice.

2 Theoretical Possibility

2.1 Notation and Terminology

In a clinical trial where a difference in slopes between treatment groups is the primary outcome of interest, the corresponding linear model is : $E(Y | Z, D) = \beta_0 + \beta_1 Z + \beta_2 D + \beta_3 ZD$. In this model, Z indicates measurement time from baseline and D is the binary indicator of treatment status. Instead of this interaction model, we equivalently consider the slope parameter from the stratified model, $E(Y | Z) = \beta_0 + \beta_1 Z$, where the parameter of interest is simply β_1 . Our estimate of the slope parameter of interest is $\hat{\beta}_1$, the second element in the vector $\hat{\beta}$.

In sequential analysis, we consider two analysis times, t_j and t_k with $j < k$, such that the number of measurements available at time t_k is always greater than the number of measurements available at time t_j . We use M_j to indicate the total number of measurements available at analysis time t_j and let z_m denote the value of Z for the m -th observation ($m = 1, \dots, M_j$).

For each individual i , let the vector of measurement times be z_i . In single analysis designs, the data are considered to be “balanced” if this vector is the same for all individuals observed, that is $z_i = z \forall i$. This commonly occurs in laboratory or other planned experiments in which all individuals are measured at the same observation times after the start of the study.

In clinical trials with interim analyses, the concept of “balance” is more complicated because more observations on the same individual may be made after an interim analysis. We reserve the term “balance” to refer to the scenario where the vector of observation times z is constant at all interim analysis times. If an interim analysis takes place at a point such that all individuals who are included in an interim analysis have the same observation times and no further measurements will be observed on these individuals, then it is a point of balance. Such balance may occur, for example in laboratory studies where follow up is short. If an interim analysis takes place after the study is completed on half the animals, that analysis takes place at point of balance because all individuals in the interim analysis have the same observation times and these observation times are the same as what would be true at the final analysis.

In contrast, we say that the data are not balanced but “level” if at an interim analysis time the vector of observation times is the same for all individuals $z(t_j)$, but not all observation times that would be observed at the end of the study have been measured (i.e., $z(t_j) \neq z(t_j)$). As an example, consider a situation in which animals were to be measured first at randomization and then at week 5, week 10, and week 15. If an interim analysis were planned after week 5 (after half the planned measurements), all individuals would have the same vector of observation times at the interim analysis, but this vector is not the same as what would be true at the end of the study. These data are not balanced because the interim and final observation time vector is different. However, these data are “level” because at each observation time observed at the time of the interim analysis (baseline and week 5), there are an equal number of measurements.

2.2 GEE Background

The method of generalized estimating equations includes the choice of a working covariance matrix, W . For a continuous outcome (with identity link and assumption of no mean-variance relationship), the estimates produced from this method are easily written in closed form:

$$\hat{\beta} = (X^T W^{-1} X)^{-1} X^T W^{-1} Y, \quad (1)$$

although iteration may be necessary to estimate $\hat{W}^{-1}$ from the data.

For any choice of working covariance matrix, W , the true variance of the estimated parameters is:

$$\text{Var}(\hat{\beta}) = (X^T W^{-1} X)^{-1} X^T W^{-1} V W^{-1} X (X^T W^{-1} X)^{-1}, \quad (2)$$

where V denotes the true covariance matrix, $\text{Var}(Y) = V$.

If $W = V$, then equation 2 simplifies to $\text{Var}_V(\hat{\beta}) = (X^T W^{-1} X)^{-1}$ and we use $\text{Var}_V(\hat{\beta})$ to denote that this variance matches the variance of the efficient estimator (using weighted least squares (WLS) with known V) when $\text{Var}(Y) = V$. However, in many settings, the form of the

working covariance matrix (W) does not match the true covariance matrix (V). Although this difference may result in inefficient estimates (Mancl and Leroux, 1996), using the robust, Huber-White sandwich standard errors will lead to appropriate inference (Liang and Zeger, 1986). However, the use of the inefficient estimator leads to the possibility of nonmonotonic information growth.

Using the identity matrix as the form of the working covariance matrix ($W=I$), is often a particularly appealing choice. Using working independence means that the estimate of the parameters of interest will match exactly the estimates that would be obtained through standard least squares regression methods (Pan et al., 2000), and using working independence ensures that the estimates will be unbiased even in scenarios with time-varying covariates that can lead to bias using other working covariance structures (Pepe and Anderson, 1994).

2.3 Correlated, Homoscedastic Data

We first restrict our attention to the case of homoscedastic, correlated data, such that only the correlation structure of the working covariance matrix may be misspecified. For simplicity and without loss of generality, assume that the common, constant variance, is equal to 1. Let v_{rc} denote the elements of the true covariance matrix V at analysis time t_j ($r = 1, \dots, M_j$, $c = 1, \dots, M_j$). Note that most elements of this covariance matrix are equal to 0; only elements corresponding to measurements made on the same individual will be nonzero. In a similar manner, let w_{rc} denote the elements of the inverse of the working covariance matrix, W^{-1} . In the special case of using GEE with working independence, we have $w_{rc} = 1$ if $r = c$ and $w_{rc} = 0$ otherwise.

Using matrix algebra and equation 2, we obtain the following fraction comparing the variance of our estimate of interest at two analysis times, t_j and t_k :

$$\frac{Var(\hat{\beta}_{1k})}{Var(\hat{\beta}_{1j})} = \left(\frac{M_j Var_j(z)}{M_k Var_k(z)} \right)^2 \frac{\left(\sum_c^{M_k} \sum_r^{M_k} v_{rc} (z_r - \bar{z}_k)(z_c - \bar{z}_k) \right)}{\left(\sum_c^{M_j} \sum_r^{M_j} v_{rc} (z_r - \bar{z}_j)(z_c - \bar{z}_j) \right)} \quad (3)$$

where $\bar{z}_j$ indicates the mean of the Z measurements available at analysis time

t_j ($\bar{z}_j = (\sum_m^{M_j} z_m) / M_j$) and $Var_j(z)$ indicates the variance of these same measurements.

From equation 3, it is clear that the first term involving M_j , M_k and the variances of the predictor space at each analysis time must be less than 1 because that is the ratio of the variances of the efficient estimator with independent data. However, there is no such guarantee with the term on the right side. Adding additional measurements such that highly correlated measurements are now at points of high leverage could lead the top term to be larger than the bottom and thus could result in nonmonotonic information growth; in other words, the counterintuitive scenario of lower statistical statistical precision with more total measurements at time t_k .

Intuition might suggest that the reason for this possible absolute inefficiency is due to the weighting on the measurements when determining the estimate of $\hat{\beta}_1$. Using working independence weights observations taken at the same measurement time z equally, regardless of how many other measurements are made on each individual. Such weighting does not generally result in a great loss of efficiency with level data, where all individuals have the same number of measurements and the same measurement times. In fact, when data are truly exchangeable, using working independence is equivalent to the efficient estimator if the data are level. However, this equal weighting is not efficient when a few individuals have more measurements than the others and measurements on the same individual are correlated and may lead to an unexpected loss of statistical information despite greater total measurements.

2.4 Independent, Heteroscedastic data

Similarly, we consider independent, but heteroscedastic data. Although such data are unlikely in a longitudinal setting, focusing on the impact of heteroscedasticity separately from correlation provides some useful intuition. As was the case with homoscedastic, correlated data, using GEE with working independence in this setting will lead to valid, but inefficient estimates, and the possibility of non-monotonic information growth.

Let σ_m^2 denote the variance of the m -th observation for simplicity in the notation. Then,

$$\frac{Var(\hat{\beta}_{1k})}{Var(\hat{\beta}_{1j})} = \left(\frac{M_j Var_j(z)}{M_k Var_k(z)} \right)^2 \left(\frac{\sum_m^{M_k} \sigma_m^2 (z_m - \bar{z}_k)^2}{\sum_m^{M_j} \sigma_m^2 (z_m - \bar{z}_j)^2} \right) \quad (4)$$

This formula illustrates that the variance of the slope using working independence will be larger when the data at the extremes of the predictor space (i.e. with large values of $z_m - \bar{z}$) are more variable. As in the case of correlated, homoscedastic data, the first term involving ratios of M_j , M_k , $Var_j(z)$, and $Var_k(z)$ must always be less than 1 in a group sequential design. However, the second term in equation 4 will can be greater than 1, particular when data are increasing in variability. As was true with correlated, homoscedastic data, nonmonotonic information growth can occur if the second term is sufficiently greater than 1 to overwhelm the first term. Such a scenario is possible if the additional points at analysis time t_k result in points of high leverage having large variances.

2.5 Correlated, Heteroscedastic Data

With correlated, heteroscedastic data, or when the working covariance matrix takes a form other than independence, the theoretical results do not provide as much intuition. Nevertheless, we show in the supplementary materials (Web Appendix) that the variance of our statistic of interest, $\hat{\beta}_1$, can be written as the sum of two terms:

$$Var(\hat{\beta}_1) = Var_w(\hat{\beta}_1) + \Delta \quad (5)$$

where $\text{Var}_W(\hat{\beta}_1)$ is the variance of the statistic if the true covariance matrix matched the working covariance matrix used (i.e., if $\text{Var}(Y) = W$).

By separating the variance of the estimate in this way, we can see that the second term in equation 5, $\frac{1}{n} \sum_{i=1}^n \frac{1}{w_i}$, is the one that may lead to nonmonotonic information growth; if the second term were zero, then the first term would equal to the variance of an efficient estimator and therefore the sequential statistics must have independent increments and grow monotonically. This second term is due to the inefficient weights (such that the elements of the working covariance matrix are not equal to the true covariance matrix) and, as expected, is larger in magnitude when discrepancies are large.

Nonmonotonic information growth can occur in such circumstances when the magnitude of $\frac{1}{n} \sum_{i=1}^n \frac{1}{w_i}$ changes over the course of the study – if w_i were constant, then $\text{Var}(\hat{\beta}_1)$ would be consistently decreasing (because $\text{Var}_W(\hat{\beta}_1)$ must be decreasing). As shown previously with correlated, homoscedastic data (in section 2.3) and with independent, heteroscedastic data (section 2.4), using an “incorrect” form of the working covariance matrix is nearly efficient (small values of $\frac{1}{n} \sum_{i=1}^n \frac{1}{w_i}$) when data are level (or nearly level) but is far less efficient when highly correlated (or highly variable) measurements are added at points of high leverage (larger values of $\frac{1}{n} \sum_{i=1}^n \frac{1}{w_i}$). Thus, transitions from available data that are level (or nearly level) to data that are less level create differences in the magnitude of $\frac{1}{n} \sum_{i=1}^n \frac{1}{w_i}$ that may cause nonmonotonic information growth if the increase in $\frac{1}{n} \sum_{i=1}^n \frac{1}{w_i}$ overwhelms the decrease in $\text{Var}_W(\hat{\beta}_1)$. Intuition is similar in the general case – nonmonotonic information growth can result when going from a place of near levelness (and nearly efficient weighting) to a point where some observations have high leverage (and the weighting is very different from what would be efficient). A more detailed derivation in the general case is not possible due to the lack of closed form solution to the iteratively determined elements of working covariance matrices; they are not independent and not correct for the true covariance matrix.

3 Simulated Examples of Nonmonotonic Growth

To provide intuition for which factors most influence the possibility of nonmonotonic information growth and provide illustrations of when it might occur, we considered simulations that varied factors in three categories. We considered factors related to (1) the correlation of measurements from the same participant, including possible heteroscedasticity, (2) factors related to the trial design, and (3) factors related to statistical analysis choices. Factors related to the correlation of measurements were: the structure of the true correlation matrix (AR(1) or exchangeable) and the amount of heteroscedasticity present (none, some, or substantial). Factors related to trial design were the number of measurements planned for each participant (10 or 4 measurements/participant) and the accrual rate of participants relative to the planned timing of measurements (fast, medium, or slow). Factors related to statistical analysis were limited to the choice of working correlation matrix in GEE (independent, exchangeable, AR(1), or unstructured).

We choose to include the “unstructured” working covariance matrix because some authors (Gange and DeMets, 1996) 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. Others have suggested that using a working covariance matrix with consistently estimated parameters (even if misspecified) will lead to nearly independent increments (Lee et al., 1996), suggesting that any choice of working covariance will lead to monotonic information growth. All simulations were done using the geepack package (Yan and Fine, 2004) in R (R Development Core Team, 2008).

3.1 Correlated, Homoscedastic Data

We first demonstrate the possibility of nonmonotonic information growth when data are correlated, but homoscedastic. In this example, 10 measurements are made on each individual, one at baseline, and one at each of nine follow-up times. The accrual occurs over the first two months. When the true correlation structure is AR(1), we chose $\rho = 0.95$ as an example of high within individual correlation. In an attempt to make the exchangeable correlation structure as equivalent as possible, we chose $\rho = 0.8338$ when the true correlation structure was exchangeable. With this choice, the average correlation among all pairs of measurements on an individual at the end of the study would be equal for both the true exchangeable and true AR(1) scenarios. To ensure comparability between these two scenarios, 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). In each scenario, four working covariance matrices were considered: independence, exchangeable, AR(1), and unstructured. For each of the two scenarios, there are two plots of the information growth. The scaled plots show the relative loss of efficiency compared to using the correctly specified form of the working covariance matrix, while the unscaled plots show true information growth curves (information at time t_j is scaled by the information at the final analysis).

Additional scenarios considered only exchangeable, homoscedastic data and varied the correlation (low: $\rho = 0.3$, medium: $\rho = 0.6$, and high: $\rho = 0.9$) or the accrual rate (fast: 1 month, medium: 5 months, and slow: 10 months).

In the first scenario, using working independence does lead to nonmonotonic information growth (figure 1). This nonmonotonic growth occurs exactly when highly correlated observations are at points of high leverage using the independence working covariance matrix. When only a portion of subjects have additional measurements further along in the study, there is a greater loss of efficiency using working independence (compared to working exchangeable) and nonmonotonic information growth is observed. As might be expected, the amount of this non-monotonicity depends on several factors. The degree of correlation among measurements on the same individual affects the true information growth when using an independence working covariance matrix. When the true data are exchangeable with low correlation ($\rho = 0.3$), the information growth is nearly the same between the exchangeable and independence working covariance matrices (figure 2A). As the correlation increases, using working independence becomes less efficient at interim points in the trial and can lead to nonmonotonic information growth (figure 2: A-C).

Another factor influencing possible nonmonotonicities in the information growth is the rate of accrual relative to the time between consecutive measurements on an individual. When accrual is fast relative to the time between measurements the design becomes nearly level at

certain points during calendar time. In the limit of this scenario, levelness may be achieved if the accrual time is shorter than the time between consecutive measurements.

Nonmonotonicity is most pronounced when the accrual period is short relative to the follow up on each individual (figure 2D). In this case, working independence is equivalent to working exchangeable at all places of levelness during the study (information growth curves touch). However, working independence is not efficient at other places during the study where there is not levelness and this inefficiency results in nonmonotonic information growth when using working independence (equation 3). The short accrual relative to the time between measurements leads to dramatic nonmonotonicity between places of levelness and these nonmonotonicities become more extreme as the study progresses. For example, when all individuals have four measurements but before anyone has a fifth (time=4.0), using working exchangeable and working independence is equivalent due to levelness (s.e.=0.0081, IG=0.061). However, when 20% of individuals have the fifth measurement (time=4.2), working independence is far less efficient than working exchangeable (s.e.=0.0097 vs. 0.0074). This change in statistical efficiency leads to nonmonotonic information growth using working independence (IG=0.043). This drop in information is more extreme when more measurements are available. For example, when all individuals have nine of ten measurements but no one yet has a tenth (time=9.0) working independence and working exchangeable are again equivalent (s.e.=0.0024, IG=0.73). When 20% of individuals have a tenth measurement, the statistical information has increased for working exchangeable as expected (s.e.=0.0023, IG=0.78) while it has decreased for the working independence case (s.e.=0.0026, IG=0.61).

A long, slow accrual relative to the length of follow up leads to relative inefficiency using working independence, but does not tend to lead to noticeable nonmonotonic information growth (figure 2F). The relative inefficiency is due to the lack of levelness at any point in the trial except the end of the study.

Using choices other than independence for the working covariance matrix may also lead to nonmonotonic information growth. In particular, the “unstructured” working covariance matrix performs poorly when the design is far from level, yielding results similar to those using an independence working covariance matrix (figure 1). When the design is far from level, some estimated parameters in the working covariance structure are quite variable because there are 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. 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 do not reflect another practical difficulty of using an unstructured working covariance matrix in the setting of a group sequential clinical trial.

In contrast, both working exchangeable and working AR(1) did not result in nonmonotonic information growth, however there is still concern that using the misspecified form of correlation matrix could lead to a relative loss of efficiency. The scaled plots in figure 1 show the relative loss of efficiency compared to using the correctly specified form of the working covariance matrix. In this example, using an exchangeable working covariance

matrix appears to be most desirable; it does lose some efficiency relative to using AR(1) when the truth is AR(1) (relative efficiency = 89%), but never did the information growth become nonmonotonic. In contrast, when the true data are exchangeable, using a working AR(1) structure led to a dramatic drop in efficiency (relative efficiency = 50%).

3.2 Independent, Heteroscedastic Data

Briefly, we demonstrate the possible nonmonotonicity of the information growth when the data are independent, but heteroscedastic. We consider predictor-variance heteroscedasticity, such that only the predictor value (z) impacts the amount of variability (rather than mean-variance heteroscedasticity). We model such heteroscedasticity as:

$$Y_{iz} \sim (\mu_z, \sigma^2(a+bz)^\gamma), \quad (6)$$

where μ_z denotes that the mean depends on the value of the predictor z and σ , a , b , and γ are constants.

For the sake of illustration, we consider the same accrual pattern and measurement schedule used previously (2 month uniform accrual, measurements at baseline and months 1-9 thereafter). The heteroscedasticity was generated as in equation 6, with $\sigma^2 = 1$, $a = 1$, $b = 1$, and $\gamma = 2$, indicating substantial heteroscedasticity. The true information growth was obtained by simulation and is shown in figure 3. In this setting, particularly later in calendar time (as measurements are becoming increasingly more variable), the true information decreases as the first few more variable measurements are obtained, as expected from equation 4. Eventually, as more measurements are accrued, the information again increases, as would be expected with increasing the number of observations and reducing the leverage of the points with the highest variability.

The estimated information growth shown in figure 3 is the (incorrect) information growth from using ordinary least squares estimation with independent data and the assumption of constant variance. If these “model-based” standard errors are used, the estimated information growth will differ from the true information growth. The model-based estimates of the standard error 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 adding a few measurements with more variability. This estimate increases as more measurements with increased variability are added, so it is typically highest only after all additional measurements have been taken. This increase in estimated variance is usually offset by gains in the number of measurements, leading to more typical behavior of the information growth curve. However, under extreme heteroscedasticity (γ very large), assuming a constant variance could lead to dramatic cases in which the estimated information is lowest after all of the more variable measurements have been accrued.

3.3 Correlated, Heteroscedastic Data

To explore the relationship between correlated and heteroscedastic data and their joint impact on information growth, we consider a scenario in which every individual will have

four measurements at equally spaced study times 0-3. We assume that an interim analysis will be conducted before a fourth measurement is obtained on any individual (at time t_j which is fixed), and calculate the relative information of the statistic from this analysis with the statistic from an interim analysis after we have the fourth measurement on a fraction of individuals (at time t_k , which we will vary). We transform this ratio to a log scale for ease of

illustration and plot $\log\left(\frac{\text{Var}(\hat{\beta}_{1j})}{\text{Var}(\hat{\beta}_{1k})}\right)$, so negative values indicate less statistical precision at the later analysis time. We consider three different accrual patterns: Fast, Medium, and Slow. Under Fast accrual, all individuals in the study have three measurements before any individual has a fourth measurement made (so the design is completely level at time t_j). Under Medium accrual, two-thirds of individuals in the study have three measurements before the first person has a fourth measurement made. Under Slow accrual, only one-third of individuals in the study have three measurements before the first person has a fourth measurement made. We also varied the amount of heteroscedasticity from none ($\gamma = 0$ in equation 6), to some ($\gamma = 1$), to substantial ($\gamma = 2$). The true data were exchangeable and analysis was done using GEE with working independence.

Figure 4 shows places in which an interim analysis occurring later in calendar time leads to a more variable estimate of the slope. Such nonmonotonic information occurs when the correlation is at least moderate, when accrual is fast, and when interim analyses are spaced close together.

When accrual is medium or slow relative to follow-up, there is little problem with nonmonotonicity, even with very strong correlation and large amounts of heteroscedasticity. Nonmonotonicity occurs only with high correlations (ρ at least > 0.5) and only when a small fraction of individuals have the fourth measurement. In order for nonmonotonicity to be a concern in these scenarios, interim analyses would have to be spaced very close together, even with heteroscedastic and highly correlated data.

However, if accrual is fast relative to follow-up (such that nearly all individuals are accrued before a second measurement is made on any individual), then nonmonotonic information is of more concern. With high correlations ($\rho > 0.5$), nonmonotonic information growth is likely unless care is taken to schedule analyses such that the amount of levelness is similar between the interim analysis times. Even with low to moderate correlations ($\rho < 0.5$), nonmonotonic information is possible with heteroscedasticity, but only when an interim analysis occurs shortly after one that was at a point of levelness. As long as such a situation is avoided when planning a trial, nonmonotonic information growth is unlikely with low to moderate correlations and heteroscedastic data.

The primary source of possible nonmonotonic information growth in fast accruing trials is the correlation of the data; even with substantial heteroscedasticity, nonmonotonic information growth is not a concern when the correlation of measurements is low ($\rho < 0.3$). Results when the data were truly AR(1) were similar and are available in the Web Appendix.

4 Practical Implications

First, it should be noted that there are important scientific reasons to include measurements from times further from randomization, even if doing so detracts from the precision with which we can estimate the slope. From a statistical perspective only, we could obtain a more precise estimate of the slope, by ignoring the newly acquired data when nonmonotonic information growth is present. However, this approach places complete faith in the underlying linear model being true; the additional measurements further from randomization provide evidence (or lack thereof) of the linear trend further away from randomization. This evidence of a linear trend should be included, despite any statistical concerns about increased variability of the estimate.

The possibility of nonmonotonic information growth presents a problem for standard group sequential methods. If a study encounters nonmonotonic information growth in truth, there are three options available: (1) Do not conduct the planned interim analysis, or delay it until a time at which the information has increased again such that the later analysis is more statistically informative than prior analyses. (2) Construct boundaries that correctly spend some amount of small, prespecified type I error at this analysis, despite the decreased information. This second approach will lead to nonmonotonic boundaries on the sample mean scale, in which interim estimates that would have signaled stopping at the earlier (but less variable) analysis time would not stop at this later analysis time. (3) To avoid nonmonotonic boundaries on the sample mean scale, use the sample mean scale boundaries from the previous analysis time; remaining analyses could be completed by accounting for the constrained boundaries. However, if there is extreme nonmonotonic information growth between two analyses, it is possible that this approach would spend all of the type I error for the trial (or even more than the nominal amount) at this point, which would preclude conducting future analyses.

The best recommendation for future studies would be to avoid planning interim analyses in scenarios that might lead to nonmonotonic information growth. Specifically, in longitudinal studies with rapid accrual, researchers should avoid planning successive interim analyses such that the first is at a place of levelness or near levelness and the second is at a point with only a handful of measurements at the last observed measurement time. If such planning is impossible, researchers should instead plan for the possibility of nonmonotonic information growth using one of the three options detailed above. Researchers using GEE methods should also strongly consider an exchangeable working correlation matrix, as the possibility of nonmonotonic information growth is lessened using exchangeable rather than independence working correlation.

Although beyond the scope of this paper, we note briefly that missing data can also disrupt the levelness of a longitudinal design and could potentially result in nonmonotonic information. For example, if individuals accrued later in the trial were not observed at a planned final observation time, the data would be missing completely at random (MCAR) but including all measurements in the final analysis using working independence may be less efficient than using data from all individuals at all observations but the final one. This situation is analogous to interim analyses where data are “missing” on some individuals due

to different accrual patterns. We would expect that in general, missing data under MCAR would not lead to nonmonotonic information growth because it would likely decrease the amount of levelness over the duration of the study. This situation is more similar to the case of long accrual relative to follow up, which lead to less efficient estimation, but not nonmonotonic information growth.

Finally, although nonmonotonic information growth is a particular problem for clinical trials utilizing sequential analysis methods, the possibility of losing information by adding additional measurements is of interest to a broader group as well. The possibility of “losing” statistical precision provides further incentive to balance designs as much as possible (although scientific concerns may preclude such designs in many real-life settings). As with clinical trial investigators, we would suggest utilizing an exchangeable working correlation matrix with GEE whenever possible to reduce the impact on the standard errors from an unbalanced design. Following such practices would limit potential problems associated with nonmonotonic information growth.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Aisen PS, Schneider LS, Sano M, Diaz-Arrastia R, van Dyck CH, Weiner MF, Bottiglieri T, Jin S, Stokes KT, Thomas RG, Thal LJ. High-dose B vitamin supplementation and cognitive decline in Alzheimer disease: A randomized controlled trial. *Journal of the American Medical Association*. 2008; 300:1774–1783. [PubMed: 18854539]
- Burington BE, Emerson SS. Flexible implementations of group sequential stopping rules using constrained boundaries. *Biometrics*. 2003; 59:770–777. [PubMed: 14969454]
- Gange SJ, DeMets DL. Sequential monitoring of clinical trials with correlated responses. *Biometrika*. 1996; 83:157–167.
- Jennison C, Turnbull BW. Group sequential analysis incorporating covariate information. *Journal of the American Statistical Association*. 1997; 92:1330–1341.
- Jennison, C., Turnbull, BW. *Group Sequential Methods With Applications to Clinical Trials*. CRC Press; 2000.
- Kittelson JM, Sharples K, Emerson SS. Group sequential clinical trials for longitudinal data with analyses using summary statistics. *Statistics in Medicine*. 2005; 24:2457–2475. [PubMed: 15977295]
- Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika*. 1983; 70:659–663.
- Lee SJ, Kim K, Tsiatis AA. Repeated significance testing in longitudinal clinical trials. *Biometrika*. 1996; 83:779–789.
- Liang K, Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
- Mancl LA, Leroux BG. Efficiency of regression estimates for clustered data. *Biometrics*. 1996; 52:500–511. [PubMed: 10766502]
- O’Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979; 35:549–556. [PubMed: 497341]
- Pan W, Louis TA, Connett JE. A note on marginal linear regression with correlated response data. *The American Statistician*. 2000; 54:191–195.

- Pantazis N, Touloumi G, Vanhems P, Gill J, Buchere HC, Porter K. CASCADE Collaboration. The effect of antiretroviral treatment of different durations in primary HIV infection. *AIDS*. 2008; 22:2441–2450. [PubMed: 19005267]
- Pepe M, Anderson G. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics Simulation and Computation*. 1994; 23:939–951.
- Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika*. 1977; 64:191–200.
- R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2008.
- Scharfstein DO, Tsiatis AA, Robins JM. Semiparametric efficiency and its implications on the design and analysis of group sequential studies. *Journal of the American Statistical Association*. 1997; 92:1342–1350.
- Stein WD, Gulley JL, Schlom J, Madan RA, Dahut W, Figg WD, Ning Y-m, Arlen PM, Price D, Bates SE, Fojo T. Tumor regression and growth rates determined in five intramural nci prostate cancer trials: The growth rate constant as an indicator of therapeutic efficacy. *Clinical Cancer Research*. 2011; 17:907–917. [PubMed: 21106727]
- Torres VE, Chapman AB, Devuyst O, Gansevoort RT, Grantham JJ, Higashihara E, Perrone RD, Krasa HB, Ouyang J, Czerwiec FS. TEMPO 3:4 Trial Investigators. Tolvaptan in Patients with Autosomal Dominant Polycystic Kidney Disease. *New England Journal of Medicine*. 2012; 367:2407–2418. [PubMed: 23121377]
- Whitehead, J. The design and analysis of sequential clinical trials. John Wiley & Sons; 1997.
- Whitehead J, Stratton I. Group sequential clinical trials with triangular continuation regions. *Biometrics*. 1983; 39:227–236. [PubMed: 6871351]
- Wu MC, Lan KG. Sequential monitoring for comparison of changes in a response variable in clinical studies. *Biometrics*. 1992; 48:765–779. [PubMed: 1420840]
- Yan J, Fine J. Estimating equations for association structures. *Statistics in Medicine*. 2004; 23:859–880. [PubMed: 15027075]

Figure 1.

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.

Figure 2.

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 $\rho = 0.8338$.

Figure 3.

Plot illustrating information growth when the data are very heteroscedastic. The solid line represents the true, nonmonotonic information growth, simulated empirically. The dashed line represents the model-based estimates of the information growth.

Figure 4.

Places of possible nonmonotonic “information” using GEE with working independence when the data are truly exchangeable. These plots consider the relative amount of information between one interim analysis before any individuals have a 4th (final) measurement and one interim analysis at points later in calendar time when various fractions of 4th measurements have been obtained. Plots vary by the amount of heteroscedasticity present from none (top) to large (bottom).