SORT 44 (2) July-December 2020, 223-264 DOI: 10.2436/20.8080.02.101

Independent increments in group sequential tests: a review

KyungMann Kim[1,*] and Anastasios A. Tsiatis[2]

Abstract

In order to apply group sequential methods for interim analysis for early stopping in clinical trials, the joint distribution of test statistics over time has to be known. Often the distribution is multivariate normal or asymptotically so, and an application of group sequential methods requires multivariate integration to determine the group sequential boundaries. However, if the increments between successive test statistics are independent, the multivariate integration reduces to a univariate integration involving simple recursion based on convolution. This allows application of standard group sequential methods. In this paper we review group sequential methods and the development that established independent increments in test statistics for the primary outcomes of longitudinal or failure time data.

MSC: 62-02, 62H10, 62L10, 62L15.

Keywords: Failure time data, interim analysis, longitudinal data, clinical trials, repeated significance tests, sequential methods.

1 Introduction

In most chronic disease clinical trials, the primary outcome of interest is either longitudinal data taken at successive follow-up visits with possibly missing data or failure time data, i.e. time to an event such as death with possible right censoring. Typically participants enter the study serially in a way known as staggered entry, and the final analysis is conducted either after a pre-specified number of follow-up visits for each participant for longitudinal data or after a pre-specified follow-up period or a pre-specified number of events of interest for failure time data.

For ethical as well as practical reasons, these clinical trials are often monitored sequentially over time during the course of the study, and if a sufficiently large treatment difference is observed at an interim analysis, they may be considered for early stopping to avoid unnecessary experimentation on human subjects. Such an approach is known

[1] Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin, USA. kyungmann.kim@wisc.edu

[2] Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA. tsiatis@ncsu.edu

Received: 15 november 2020
Accepted: 30 november 2020

as a sequential method. When clinical trials are monitored in this way using a sequential method, multiplicity from repeatedly applying statistical testing over time has to be accounted for to control the overall type I error probability at an acceptable significance level. In order to determine the sequential boundaries that preserve the operating characteristic of a statistical test applied repeatedly, the joint distribution of test statistics over time has to be known.

For clinical trials in which the primary outcome of interest is taken only once from each participant, the joint distribution of test statistics over time is simply a product of the distributions of test statistics at each interim analysis as each participant contributes data to the test statistics only once and the increments between successive test statistics are independent. However, for the primary outcome that is either longitudinal or failure time data, it is no longer the case as each participant possibly contributes outcome data to test statistics more than once over interim analyses.

Modern-day clinical trials since the mid 1990s or even earlier have been routinely monitored by data and safety monitoring boards or data monitoring committees to ensure the safety of participants and whether risks versus benefits are acceptable for continuing the study. This is accomplished using standard group sequential methods in interim analyses for possible early stopping if there is clear statistical signal of differences in efficacy of an investigational intervention as compared to a control intervention that may include a placebo or standard of care or if there is major concerns for safety of participants. This review article on independent increments in group sequential tests is an attempt to describe the development of statistical methods for interim analyses leading up to mid 1990s.

For longitudinal data, the joint distribution of test statistics over time has been investigated by many including Armitage, Stratton and Worthington (1985), Geary (1988), Wei, Su and Lachin (1990a), Lee and DeMets (1991, 1992), Reboussin, Lan and DeMets (1992), Su and Lachin (1992), Wu and Lan (1992), Gange and DeMets (1996), and Lee, Kim and Tsiatis (1996). Likewise, for failure time data, the joint distribution of test statistics over time has been investigated by many including Tsiatis (1981, 1982), Gail, DeMets and Slud (1982), Slud and Wei (1982), Sellke and Siegmund (1983), Slud (1994), Tsiatis, Rosner and Tritchler (1985), Gu and Lai (1991), Lin (1992), Gu and Ying (1995), and Tsiatis, Boucher and Kim (1995).

Often the joint distribution turns out to be multivariate normal or at least asymptotically so, and subsequently sequential methods require multivariate numerical integration. The MULNOR program by Schervish (1984) can be used to this end, but it involves very intensive numerical computation. Also the program can handle multivariate integrations of only up to seven dimensions, thus limiting the tests to be applied up to seven times only.

If the increments between successive test statistics are independent, however, the multivariate numerical integration reduces to univariate numerical integration involving simple recursion based on convolution of two independent variables as noted by Armitage, McPherson and Rowe (1969) and McPherson and Armitage (1971). This is ob-

viously the case when the outcomes are measured only once as noted earlier. Moreover, this allows the use of standard group sequential methods such as by Pocock (1977), O'Brien and Fleming (1979), and Lan and DeMets (1983) for design and analysis of group sequential clinical trials.

The joint distributions established by these authors dealt with specific test statistics under selected statistical models for longitudinal data and failure time data. Jennison and Turnbull (1990) and Scharfstein, Tsiatis and Robins (1997), however, provided generalized theory for independent increments in sequential test statistics. The former considered the joint distribution of test statistics for treatment effect in the presence of covariates in regression model setting, while the latter considered the joint distribution of semiparametric-efficient test statistics.

The rest of this paper is organized as follows. In Section 2, we first review the historical development of sequential methods including classical and the so-called group sequential methods specifically for application in clinical trials as a background. We then review repeated significance testing and univariate recursive numerical integration when increments between successive test statistics are independent in contrast to the multivariate numerical integration required for sequential test statistics with correlated increments. In Section 3, we review the historical development for the joint distribution of sequential test statistics and independent increments for group sequential tests of longitudinal data and failure time data. In Section 4, after introducing general notations and formulation of the problem, we review joint distributions of sequentially computed test statistics for general regression models of independent data and various parametric, semiparametric and nonparametric models for longitudinal data and failure time data. In Section 5, we briefly review how the error spending function and information fraction is used for design and analysis of group sequential clinical trials and demonstrate independent increments in sequential test statistics for longitudinal data and failure time data using real clinical trials data and simulated data. We close with concluding remarks and observations in Section 6.

2 Sequential methods

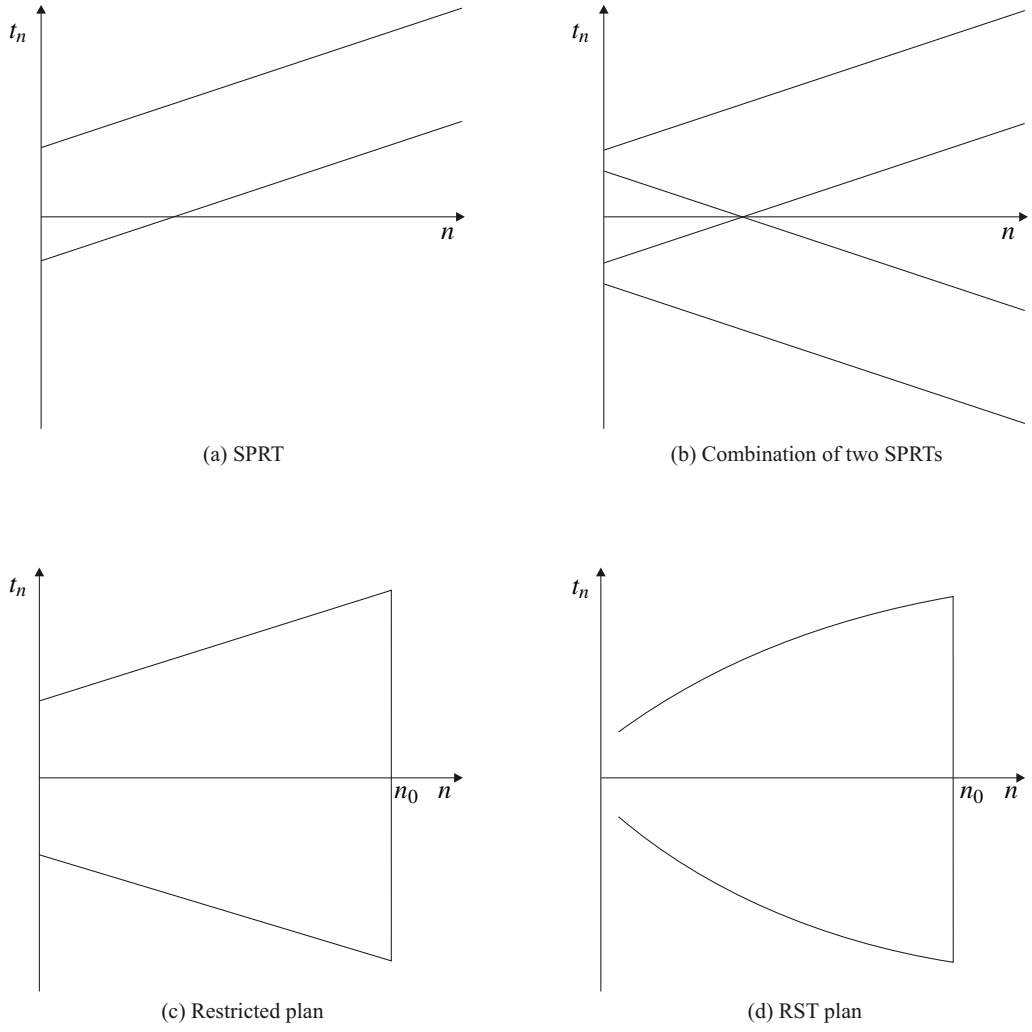
2.1 Early sequential methods

According to Armitage (1990), "[a] scientific investigation is sequential if its conduct at any stage depends on the outcome at previous stages." Probably the earliest application of sequential methods can be found in Dodge and Romig (1929) in which "double sampling schemes" are used in industrial batch sampling for quality monitoring. These two-stage sequential methods were adapted in cancer drug screening trials, e.g. in Gehan (1961), Lee et al. (1979), and Simon (1989). In a theoretical development, Stein (1945) derived a sequential procedure that uses estimated variance from the first-stage sample in choosing the size of the second-stage sample to achieve a desired power of a two-stage t-test.

For a fixed sample test of the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta = \theta_1$, let $f(x;\theta)$ be the probability density or mass function for a random variable X. According to Neyman and Pearson (1933), one rejects H_0 in favor of H_1 if $L_n > c_\alpha$ where

$$L_n = \prod_{i=1}^{n} \frac{f(x_i;\theta_1)}{f(x_i;\theta_0)}$$

is the likelihood ratio. The critical value c_α is determined for the test to be of size α. Then the test is most powerful, that is, the type II error probability β is smallest amongst all tests with size $\leq \alpha$.



(a) SPRT (b) Combination of two SPRTs

(c) Restricted plan (d) RST plan

Figure 1: *Sequential boundaries from Fig. 6.1 in Armitage (1990).*

Following what came to be known as Neyman-Pearson's fundamental lemma above, Wald (1947) developed the sequential probability ratio test (SPRT) to discriminate between two simple hypotheses. Specifically Wald SPRT shows that when the sample size is not fixed in advance, further improvement is possible. The best procedure in a certain sense made precise by Wald and Wolfowitz (1948) is 1) to continue sampling as long as $B < L_n < A$ for some constant $B < 1 < A$ and 2) to stop sampling and decide in favor of H_1 or H_0 as soon as $L_n > A$ or $L_n < B$, respectively, where

$$A \approx \frac{1-\beta}{\alpha} \text{ and } B \approx \frac{\beta}{1-\alpha}.$$

A specific case when $\theta_0 = 0$ and $\theta_1 > 0$ is a one-tailed test as shown in Fig. 1(a). There are two different versions of its generalization for a two-tailed test with $H_1 : \theta \neq 0$. One is a two-tailed test obtained by defining a density function $f_1 = (f_- + f_+)/2$ where f_- and f_+ are the probability density or mass functions corresponding to alternative hypotheses $H_- : \theta < 0$ and $H_+ : \theta > 0$ in two directions as suggested by Wald (1947) (Chapter 9). The other is a combination of two separate one-tailed tests, each with type I error probability $\alpha/2$, by Sobel and Wald (1949), as shown in Fig. 1(b).

One drawback of SPRTs is that sampling may continue indefinitely. A restricted plan by Armitage (1957) is a modification of the two-tailed version of a SPRT by Sobel and Wald (1949) to avoid this possibility by imposing a maximum sample size with the inner wedge removed or pushed out as shown in Fig. 1(c). A similar sequential plan was later developed by Armitage et al. (1969) as a repeated significance test plan as shown in Fig. 1(d) and described in detail in Subsection 2.3 as a means to adjust the critical value to account for multiple testing leading to a constant critical value. Of note, the operating characteristics of these two sequential tests in Figs. 1(c) and 1(d) are very similar.

2.2 "Sampling to reach a foregone conclusion"

Let $X_1, X_2, \ldots$ be independent and identically distributed and drawn from $N(\mu, \sigma^2)$ with known variance σ^2, and consider a statistical test of $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. For a single sampling plan with a fixed sample size n, one would reject H_0 if and only if $|S_n| > 1.96\sigma\sqrt{n}$ at a significance level $\alpha = 0.05$ where $S_n = \sum_{i=1}^{n} X_i$.

A need for adjustment in the critical value for repeated testing is recognized by the law of the iterated logarithm described here. Assume only that X_i, $i = 1, 2, \ldots$ are simply independent and identically distributed with mean μ and finite variance $0 < \sigma^2 < \infty$. In addition assume that n is not fixed in advance, and data become available sequentially one at a time. If S_n is computed for each $n \geq 1$, $|S_n|$ is certain to exceed $1.96\sigma\sqrt{n}$ for some n, even if H_0 is true, for the law of the iterated logarithm asserts that

$$\limsup_{n\to\infty} \frac{S_n - n\mu}{\sigma\sqrt{2n\log\log n}} = 1 \text{ with probability } 1.$$

Thus an unscrupulous experimenter might be tempted to take a sample of size

$$N = \inf\{n \geq 1 : |S_n| > 1.96\sigma\sqrt{n}\},$$

and report as if it were a fixed sample size and claim rejection of H_0 at a significance level 0.05. However, the experimenter may have to spend some time in the process as the expected sample size under this sampling scheme is $\mathrm{E}(N) = \infty$.

That one can reach a nominal significance by testing repeatedly was aptly described as "sampling to reach a foregone conclusion" by Anscombe (1954).

2.3 Repeated significance tests

Controversy regarding control of type I error probability depending on the approach, be it Bayesian, likelihood-based, or frequentist, led Armitage et al. (1969) to evaluate the type I error probability of the sequential testing procedure described above to settle the score, so to speak. The numerical procedure for computing the type I error probability is described below.

Assume as above that $X_1, X_2, \ldots$ are independent and identically distributed normal random variables with mean μ and, without loss of generality, variance 1. To test $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ at a significance level α, sampling is terminated the first time when

$$|S_k| > b_k$$

where $b_1, b_2, \ldots$ are boundary values. With the maximum number of observations K, the boundary values have to satisfy the following:

$$\Pr(|S_k| > b_k \text{ for some } k = 1, \ldots, K) = \alpha$$

or equivalently

$$\Pr(|S_1| \leq b_1, \ldots, |S_K| \leq b_K) = 1 - \alpha.$$

The computation of these probabilities can be simplified by noting that f_k, the probability density function of S_k under H_0 in the sequential procedure, satisfies the following recursive definition base on convolution:

$$f_k(s) = \int_{-b_{k-1}}^{b_{k-1}} f_{k-1}(u)\phi(s-u)du \tag{2.1}$$

where f_1 is the standard normal density function ϕ above. This is so because of the independence between S_{k-1} and $S_k - S_{k-1}$, i.e. independent increments in S_k.

With $k^\star$ denoting the random variable for when $|S_k| > b_k$ for the first time, the probability of stopping at or before k is

$$P_k = \Pr(k^\star \le k) = 1 - \Pr(|S_1| \le b_1, \ldots, |S_k| \le b_k) = 1 - \int_{-b_k}^{b_k} f_k(u)du$$

and the probability of stopping at $k^\star = k$, i.e. the exit probability $\Pr(k^\star = k)$, is simply

$$\begin{aligned} P_k - P_{k-1} &= \Pr(|S_1| \le b_1, \ldots, |S_{k-1}| \le b_{k-1}, |S_k| > b_k) \\ &= \int_{-b_{k-1}}^{b_{k-1}} f_{k-1}(u)\{1 - \Phi(b_k - u) + \Phi(-b_k - u)\}du \end{aligned} \tag{2.2}$$

where Φ is the standard normal distribution function. The overall significance of the sequential procedure is determined by

$$\alpha = 1 - \int_{-b_K}^{b_K} f_K(u)du.$$

The recursive definition of f_k above allows direct computation of these probabilities using standard numerical integration methods, e.g. a Newton-Cotes formula of the second order, i.e. Simpson's rule. This same computational procedure works when $\mu \neq 0$ with X_k replaced by $X_k - \mu$. The above observation led to the notion of repeated significance tests as described in Armitage et al. (1969), which in turn paved the way for development of group sequential methods for clinical trials.

2.4 Group sequential methods for clinical trials

Following the seminal work on sequential analysis by Wald (1947), Bross (1952) and Armitage (1954) appear to have been the first to advocate the use of sequential methods in clinical trials. Different from other settings where savings in sample size was the primary motivation for using sequential methods, it was ethical imperatives in clinical trials in considering early termination to avoid unnecessary experimentation on human subjects in the presence of clear evidence of benefits or harms of interventions.

Suppose that response to treatment is a normal random variable with means μ_A and μ_B for treatments A and B, respectively, and known variance σ^2, a typical two-sample problem. Consider a test of $H_0 : \mu_A = \mu_B$ against $H_1 : \mu_A \neq \mu_B$ or, equivalently, $H_0 : \delta = 0$ against $H_1 : \delta \neq 0$ where $\delta = \mu_A - \mu_B$. A fixed sample size test with a significance level $\alpha = 0.05$ with n participants on each treatment rejects H_0 when

$$Z = \left| \frac{\bar{X}_A - \bar{X}_B}{\sqrt{2\sigma^2/n}} \right| > 1.96$$

where $\bar{X}_A$ and $\bar{X}_B$ denote the sample means.

Group sequential designs call for monitoring of accumulating data over time periodically after groups of observations become available using sequential tests. Wald (1947) (pp 101–103) refers to taking groups of observations and applying SPRTs for binary

outcome. One strategy of a group sequential test is to reject the null hypothesis of no treatment difference if, at any of the interim analyses, the test statistic becomes sufficiently large; otherwise, do not reject (accept) the null hypothesis.

Consider examining the accumulating data after a group of every $2n$ observations, n on each treatment, become available, namely,

$$Y_j = \frac{\bar{X}_{Aj} - \bar{X}_{Bj}}{\sqrt{2\sigma^2/n}} \sim N(\delta^\star, 1)$$

where $\delta^\star = \delta/\sqrt{2\sigma^2/n}$, for up to a maximum of K analyses for a maximum of $2nK$ observations. With the score statistics

$$S_k = \sum_{j=1}^{k} Y_j = \sum_{j=1}^{k} \frac{\bar{X}_{Aj} - \bar{X}_{Bj}}{\sqrt{2\sigma^2/n}} \sim N(\delta^\star k, k) \tag{2.3}$$

or the Wald statistics

$$Z_k = S_k/k^{1/2} \sim N(\delta^* k^{1/2}, 1), \tag{2.4}$$

a group sequential test rejects H_0 for the first time when

$$|S_k| > b_k \text{ or equivalently } |Z_k| > c_k.$$

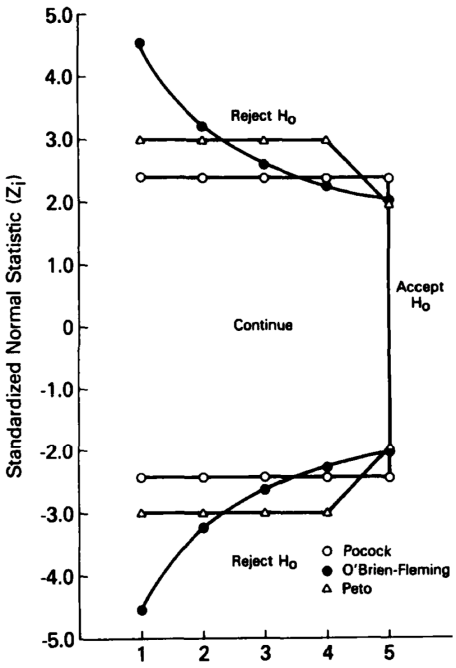


Figure 2: *Group sequential critical values from Fig. 1 in DeMets and Lan (1984).*

Hence, if we want a level α test, we choose the boundary values, $b_1,\ldots,b_K$, or equivalently the critical values, $c_1,\ldots,c_K$, such that, under H_0,

$$\Pr(|S_1|\le b_1,\ldots,|S_K|\le b_K)=\Pr(|Z_1|\le c_1,\ldots,|Z_K|\le c_K)=1-\alpha. \tag{2.5}$$

Note that there is an equal increment of statistical information in sample size, i.e. $2n$, between sequentially computed test statistics and that the increments are independent so that the computational procedure by Armitage et al. (1969) can be used in this type of group sequential tests.

Several group sequential methods are used for determining the boundary or the critical values. These values for Pocock (1977) and O'Brien and Fleming (1979) group sequential methods are obtained by solving (2.5) under the conditions of $c_1=\cdots=c_K$ and $b_1=\cdots=b_K$, respectively (see Fig. 2). Note that Pocock's method is the group sequential version of the repeated significance test method discussed in Subsection 2.3. One practical drawback of these methods is that they depend on the assumption of equal sample size or more generally, equal amount of statistics information, accumulated between two successive analyses. Otherwise the group sequential methods by Pocock (1977) and O'Brien and Fleming (1979) cannot be applied. In order to address this situation, a flexible approach was proposed by Slud and Wei (1982) in which the boundary values, b_k, $k=1,\ldots K$, are determined with prespecified α_k, $k=1,\ldots,K$, such that $\alpha_k=P_k-P_{k-1}$ in (2.2) under the null hypothesis and $\sum_{k=1}^K\alpha_k=\alpha$, the overall significance level. A practical downside to this approach is the arbitrariness in specifying α_ks and the possibility of the group sequential test not meeting the criterion for early stopping at an interim analysis and meeting the criterion at the next interim analysis with the increment in the statistical information between the two interim analyses in the opposite direction, an obvious logical inconsistency.

Generalizing the idea in Slud and Wei (1982), Lan and DeMets (1983) introduced the notion of "alpha spending" instead of arbitrarily specifying α_ks. As a method of allocating the type I error probability α into α_ks as in Slud and Wei (1982), Lan and DeMets (1983) instead proposed allocating the type I error probability α according to an "error spending function," $\alpha^\star(t)$, which is a nondecreasing function of the information time or fraction t, $0\le t\le 1$, defined below with $\alpha^\star(0)=0$ and $\alpha^\star(1)=\alpha$. For $k=1,\ldots,K$, the type I error probability allocated for the k^{th} interim analysis is determined as $\alpha_k=\alpha^\star(t_k)-\alpha^\star(t_{k-1})$ where $t_0=0$ and $t_K=1$ so that $\sum_{k=1}^K\alpha_k=\alpha$. For a one-tailed Pocock (1977) and O'Brien and Fleming (1979) procedures, Lan and DeMets (1983) proposed $\alpha^\star_{\text{P}}(t)=\alpha\log\{1+(e-1)t\}$ and $\alpha^\star_{\text{OF}}(t)=2\{1-\Phi(z_{\alpha/2}/\sqrt{t})\}$, respectively, where z_γ is the upper γ quantile of the standard normal distribution. The information fraction t is the fraction of statistical information corresponding to an interim analysis relative to the maximum information required. For example, $t_k=k/K$ for the group sequential tests with equal samples of size n between two successive analyses as in the score statistics in (2.3). If we consider unequal sample sizes n_k between the $(k-1)^{\text{th}}$ and the k^{th} interim analyses, $t_k=n_k/n_K$ instead.

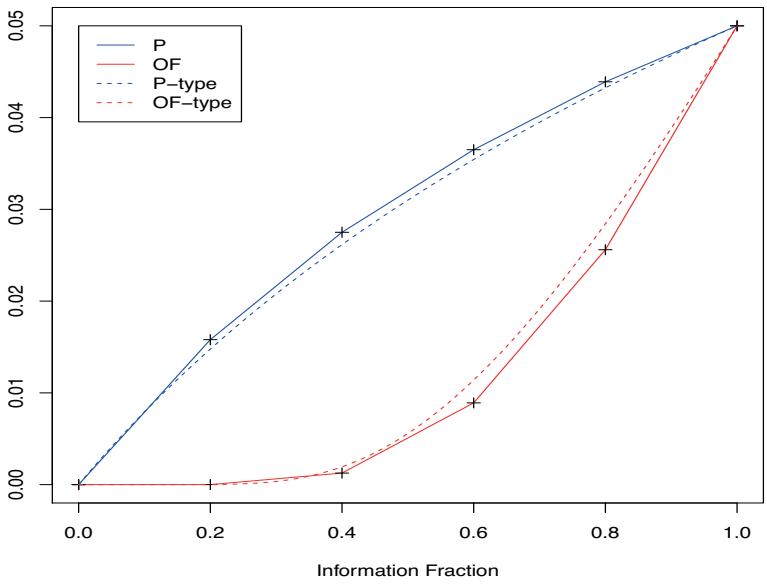


Figure 3: *Cumulative type I error probability for group sequential tests with* $\alpha = 0.05$.

The cumulative type I error probabilities for the Pocock (P) and O'Brien-Fleming (OF) group sequential procedures with $K = 5$ and $\alpha = 0.05$ and the error spending functions $\alpha^{\star}_{\mathrm{P}}(t)$ for Pocock (P-type) and $\alpha^{\star}_{\mathrm{OF}}(t)$ for O'Brien-Fleming (OF-type) from above are plotted in Fig. 3 to indicate similarities between the standard group sequential methods and group sequential methods based on the suitably chosen error spending functions.

From a historical perspective, Pocock (1977), following the repeated significance test of Armitage et al. (1969), popularized the group sequential methods for clinical trials with normal outcome. However, it was Elfring and Schultz (1973) who first coined the term "group sequential designs" for clinical trials with binary outcome. Jennison and Turnbull (1990) present a detailed review of group sequential methods including comparisons of methods by Pocock (1977), O'Brien and Fleming (1979), Slud and Wei (1982), and Lan and DeMets (1983).

2.5 Covariance under independent increments

As noted earlier in Subsection 2.3, in order to apply group sequential methods, one has to solve the following multivariate integral

$$\int_{-b_1}^{b_1} \cdots \int_{-b_K}^{b_K} f(s_1, \ldots, s_K) ds_1 \cdots ds_K = 1 - \alpha$$

where f is the joint density function of the sequential test statistics. However, if the following holds

$$\text{Cov}(S_k, S_l) = \text{Var}(S_k) \text{ or equivalently } \text{Cov}(S_{k-1}, S_k - S_{k-1}) = 0$$

for $1 \le k \le l \le K$ with $S_0 = 0$, i.e. if the sequential test statistics have independent increments, the multivariate integration above becomes univariate integration involving simple recursion based on convolution as indicated in (2.1).

To assess the joint distributions of S_k in (2.3) or Z_k in (2.4), $1 \le k \le K$, consider the fully sequential setting again as in Subsection 2.3. From the standard normal theory and the independent increments structure of S_k, it follows that the joint distribution of the score statistics S_k, $1 \le k \le K$, is multivariate normal with marginals $S_k \sim N(\mu k, k)$ and covariance

$$\text{Cov}(S_k, S_l) = k = \text{Var}(S_k),\ 1 \le k \le l \le K.$$

Since S_k is equivalent to the Wald statistic $Z_k = S_k/k^{1/2} = S_k/\sqrt{\text{Var}(S_k)}$, the corresponding joint distribution of the Wald statistics Z_k, $1 \le k \le K$, are found to be multivariate normal with marginals $Z_k \sim N(\mu k^{1/2}, 1)$ and

$$\text{Cov}(Z_k, Z_l) = \sqrt{k/l} = \sqrt{\text{Var}(S_k)/\text{Var}(S_l)},\ 1 \le k \le l \le K.$$

Hence, any one of the two conditions above gives an independent increments structure of the sequential test statistics. For the two-sample group sequential test as described in Subsection 2.4, replacing μ with δ^*, these results also hold.

More generally, three different test statistics can be considered as in Jennison and Turnbull (1997). For $1 \le k \le l \le K$, the following holds:

$$\hat{\theta}_k \overset{a}{\sim} N(\theta, \mathcal{I}_k^{-1}(\theta)) \text{ and } \text{Cov}(\hat{\theta}_k, \hat{\theta}_l) = \text{Var}(\hat{\theta}_l) = \mathcal{I}_l^{-1}(\theta) \tag{2.6}$$

for the maximum likelihood estimates where $\mathcal{I}_k(\theta)$ is the Fisher information;

$$S_k \overset{a}{\sim} N(\theta \mathcal{I}_k(\theta), \mathcal{I}_k(\theta)) \text{ and } \text{Cov}(S_k, S_l) = \text{Var}(S_k) = \mathcal{I}_k(\theta) \tag{2.7}$$

for the score statistics; and

$$Z_k \overset{a}{\sim} N(\theta \mathcal{I}_k^{1/2}(\theta), 1) \text{ and } \text{Cov}(Z_k, Z_l) = \sqrt{\mathcal{I}_k(\theta)/\mathcal{I}_l(\theta)} \tag{2.8}$$

for the Wald statistics $Z_k = \hat{\theta}_k/\text{SE}(\hat{\theta}_k)$ where SE stands for standard error.

Note that these distributional properties of the sequential test statistics are still true under the general alternatives as well as the null hypothesis, and hence power of the sequential tests can also be evaluated through the univariate integration technique as in McPherson and Armitage (1971). When the underlying distribution is not normal, we consider a class of local alternatives $\{\mu_n\}$, where $\sqrt{n}\mu_n \to \delta \ne 0$. Then normality and an independent increments structure of the sequentially computed test statistics can be established asymptotically under the null and a class of local alternatives so that the standard sequential procedures described in this section are still applicable asymptotically.

2.6 Intuition about independent increments

With normal outcome, it is intuitive that group sequential test statistics would have independent increments, thus allowing application of the classical group sequential methods. With time to event outcome, it is unclear since each participant contributes follow-up data possibly multiple times over group sequential tests. With longitudinal outcome, again it is unclear since each participant contributes follow-up data multiple times longitudinally. Both with longitudinal data and failure time data, a participant contributes data more than once over the course of study in group sequential tests and as a consequence it is not intuitive why sequential tests statistics would have independent increments.

As summarized in Jennison and Turnbull (1990), independent increments structures have been found to hold in many circumstances case by case. Scharfstein et al. (1997) showed with great generality that the efficient score statistics in parametric and semi-parametric models have an independent increments structure. Jennison and Turnbull (1997) also gave a unified explanation based on efficiency of the test statistics for the independent increments structure. For instance, in our fully sequential setting, since the sample mean $\bar{X}_k$ is the maximum likelihood estimator, or least squares estimator of μ, the corresponding sequential score and Wald tests, S_k and Z_k, have an independent increments structure following their theorems. In this review paper, we consider the group sequential score tests with independent increments derived from several estimating methods such as the maximum likelihood and least squares method. For some of them, the independent increments structures are explained by efficiency of the test statistics, while it is not for others.

3 Joint distributions of sequential test statistics

In this section we provide a review of the historical development of independent increments in group sequential tests used in clinical trials with longitudinal data and failure time data as the primary endpoint of interest for evaluation of efficacy of intervention. The emphasis on these types of outcome data is because of the fact that they are widely used in clinical trials in chronic diseases. But more importantly it is not intuitive as to why some group sequential tests for these types of outcome data have an independent increments structure while others do not. This is in contrast to the settings in which outcome data are measured only once from each participant, which intuitively have an independent increment structure.

3.1 Longitudinal data

The joint distribution of sequential test statistics for longitudinal data has been investigated by many authors for application of group sequential methods in clinical trials with such outcome data: Armitage et al. (1985), Geary (1988), Lee and DeMets (1991), Reboussin et al. (1992), and Wu and Lan (1992) based on parametric models; Lee and

DeMets (1992) based on linear rank tests; Su and Lachin (1992) based on a multivariate generalization of the Hodges and Lehmann (1963) estimator of a location shift; Wei et al. (1990a), Gange and DeMets (1996), and Lee et al. (1996) based on semiparametric models in generalized estimation equations; and Spiessens et al. (2002) based on a random-effects model for longitudinal ordinal outcome. Lee (1994) and Spiessens et al. (2000) provide review of some of these sequential tests for longitudinal data.

When the primary outcome is longitudinal data with repeated measurements, each participant can contribute outcome data to test statistics more than once. Thus it is not intuitively obvious that sequential test statistics can have independent increments due to apparent correlation among repeated measurements from the same participant. Indeed the joint distributions of the sequential test statistics by Armitage et al. (1985), Geary (1988), Wei et al. (1990a), Lee and DeMets (1992), and Su and Lachin (1992), all turn out to have correlated increments. But as summarized below, properly formulated test statistics and semiparametric-efficient tests for longitudinal data under various parametric and semiparametric models have independent increments.

Under a linear mixed-effects model of Laird and Ware (1982), Lee and DeMets (1991) show that the asymptotic joint distribution of the sequential test statistics for comparing the rates of change computed over time is multivariate normal under missing at random and includes as special cases those by Armitage et al. (1985) and Geary (1988). Later Reboussin et al. (1992) showed that the test statistics of Lee and DeMets (1991) have an independent increments structure.

In order to account for informative drop-out, Wu and Lan (1992) proposed group sequential tests to compare areas under the response change curves between two treatments based on the two-stage random effects model of Wu and Bailey (1989). It is shown that when the response curve is linear and drop-out non-informative, the test by Wu and Lan (1992) reduces to that by Lee and DeMets (1991) above and that the joint distribution of the test statistics computed over time has independent increments.

Wei et al. (1990a), Gange and DeMets (1996), and Lee et al. (1996) all proposed a group sequential test based on a semiparametric model using the generalized estimating equations approach of Liang and Zeger (1986). Wei et al. (1990a) assume an independence model for the working variance for repeated measures, while Gange and DeMets (1996) and Lee et al. (1996) assume that the covariance matrix for repeated measures is correctly specified or consistently estimated by the working covariance matrix as in Liang et al. (1992).

As indicated by Scharfstein et al. (1997), the joint distribution of the sequentially computed score statistics based on an independence model by Wei et al. (1990a) results in correlated increments as the test is not semiparametric efficient. Gange and DeMets (1996) show that the joint distribution of the regression estimators, i.e. estimators based on the generalized estimating equations, over time is asymptotically multivariate normal with independent increments, while Lee et al. (1996) show that the joint distributions of the sequentially computed score and Wald statistics both are asymptotically multivariate normal with independent increments.

As noted above, standard group sequential methods can be used if one uses an efficient test statistics over time. With random-effects models for ordinal longitudinal data, a Wald-type test can be used with standard group sequential methods. Spiessens et al. (2002) show that, even when the random-effects distribution is misspecified, the joint distribution of the Wald-type test computed over time is asymptotically multivariate normal and showed through simulation studies that a sandwich-type correction to the covariate matrix leads to an approximately independent increments structure.

3.2 Failure time data

Many authors also investigated the joint distribution of sequential test statistics for failure time data under various settings for application of group sequential methods: Tsiatis (1981) and Sellke and Siegmund (1983) under the proportional hazards model; Gail et al. (1982) for two-sample logrank score test; Tsiatis (1982), Slud (1994), and Gu and Lai (1991) for general linear rank tests; Slud and Wei (1982) for the modified Wilcoxon statistics, i.e. a generalized Wilcoxon test by Gehan (1965); Tsiatis et al. (1985) and Gu and Ying (1995) under the proportional hazards model with covariate adjustment; Lin (1992) for logrank tests adjusting for covariates under the accelerated failure time model; and Tsiatis et al. (1995) for general parametric survival models.

When failure time is a primary outcome, each participant can contribute statistical information to group sequential tests more than once before event of interest or random censoring occurs. Hence it seems natural for the increments in successive test statistics to be correlated. Indeed the joint distributions of the test statistics over time by Slud and Wei (1982) for Gehan's test by Gehan (1965) and by Lin (1992) for the logrank test under the accelerated failure time model turn out to have correlated increments. In the case of a general class of linear rank tests, Tsiatis (1982) provides the condition for the weight function under which the joint distribution of the linear rank tests computed over time has independent increments.

Tsiatis (1981) was the first to develop the joint distribution of sequential test statistics and establish independent increments for a sequential test for failure time data. First the asymptotic joint distribution of the sequentially computed score statistics for the proportional hazards model was established and shown to to converge asymptotically to a multivariate Gaussian process with independent increments when participants enter randomly throughout the course of the trial. This allows group sequential methods to be based on the logrank test as a special case of the efficient scores test for the proportional hazards model in clinical trials with failure time data subject to random censoring, thus proving the conjecture made earlier in Armitage (1975) (pp 140–143).

Gail et al. (1982) investigated the operating characteristics of the logrank score test, computed after fixed numbers of events and applied to various group sequential methods, using simulation studies. They show empirically that the joint distribution of the logrank score test computed over time follows a multivariate normal distribution with independent increments reasonably well in a realistic setting in clinical trials.

Tsiatis (1982) generalizes the results in Tsiatis (1981) to a general class of nonparametric linear rank tests statistics and shows that the asymptotic joint distribution of the sequential test statistics within this general class of nonparametric tests is a multivariate normal distribution. This general class of nonparametric tests is characterized by a random function corresponding to the weight functions described by Tarone and Ware (1977) and Prentice and Marek (1979) and as a special case includes a constant weight for the logrank test, a weight function for the modified Wilcoxon test which is the survival function.

Sellke and Siegmund (1983) show that the score process of the partial likelihood and the maximum partial likelihood estimator under the proportional hazards model behave asymptotically like a Brownian motion. This relies on the approximation of the score process by a suitable martingale and a random rescaling of time based on the observed Fisher information. As such, the resulting joint distributions of the score process and the maximum partial likelihood estimator over time both have independent increments.

Slud (1994) shows that under the null hypothesis of no difference in survival distributions the sequentially computed logrank statistics of Mantel (1966) have exactly uncorrelated increments under very general patterns of enrollment, allocation to treatment and lost to follow-up in clinical trials. Gu and Lai (1991) considers the general class of linear rank test statistics investigated in Tsiatis (1982) and develops a general weak convergence theory for the joint distribution of the sequential linear rank test statistics for two sample problems in a realistic clinical trial setting.

Tsiatis et al. (1985) investigates the joint distribution of the sequentially computed efficient scores for the treatment effect derived from a partial likelihood under the proportional hazards model with adjustment for other covariates. They show that the sequential efficient scores test for the treatment effect in the presence of other covariates has asymptotically the same joint distribution as the sequentially computed ordinary logrank test with no covariates. The motivation for this work was the efficiency gain in the test by adjusting for the effects of other covariates. Gu and Ying (1995) show that a general Cox-type partial likelihood score process for staggered entry with covariate adjustment is asymptotically equivalent to a Gaussian process with independent increments, including the case in which the covariates being adjusted for are not independent of the covariates of primary interest, typically a randomized treatment indicator.

Tsiatis et al. (1995) consider the joint distribution of sequentially computed score statistics and the maximum likelihood estimator in parametric models for failure time data in the presence of nuisance parameters. By representing the sequentially computed score test as a stochastic integral of a counting process martingale, they drive the asymptotic joint distribution of the test statistics over time and show that the joint distributions of the score test and the maximum likelihood estimator are multivariate normal with independent increments. This work and the work by Lee et al. (1996) served as a seed for group sequential methods based on semiparametric efficient test statistics by Scharfstein et al. (1997).

Scharfstein et al. (1997) noted that joint distributions of many group sequential statistics used to analyze longitudinal or failure time data are asymptotically multivariate normal with an independent increments structure. This limiting distribution arises naturally when one uses an efficient test statistic to test a single parameter in a semiparametric model. They develop most general results based on semiparametric efficient tests and show that many previously developed cases of independent increments structure are a special case of a semiparametric efficient test.

4 Independent increments

In this section we review most general cases of independent increments for sequential tests for longitudinal and failure time data. First we define some notations and consider the formulation of the problem.

Consider a group sequential study with a maximum number of K interim analyses at calendar times t_k, $k = 1,\ldots,K$. We allow staggered entry of subjects and denote n_k to be the number of subjects who have entered the study at the k^{th} interim analysis. Let Y_{ik} be the outcome of the ith subject. When repeated measures are made as in a longitudinal study, let $Y_{ik} = (Y_{i1k},\ldots,Y_{i,d_{ik},k})^{\top}$ where d_{ik} denote the number of repeated measures of the ith subjects. At each k, Y_{ik}, $i = 1,\ldots,n_k$, are assumed to be independent. Let $X_{ik} = (Z_{ik}, W_{ik})$ denote a $d_{ik} \times p$ dimensional covariate (design) matrix including a treatment indicator Z_{ik} and $p-1$ time-varying covariate vectors W_{ik}, and let $\theta = (\gamma, \beta^{\top})^{\top}$ denote a corresponding parameter vector which consists of a treatment effect parameter γ and covariate effect parameters β. The total number of subjects at the last analysis is set as $n_K = n$, and let T_i be the entry time of the ith subject.

Our primary interest is focused on the group sequential tests with independent increments for the hypotheses of

$$H_0 : \gamma = 0 \text{ vs } H_1 : \gamma \neq 0 \tag{4.1}$$

where the parameters β are regarded as nuisance parameters adjusting for covariates.

A test for the hypotheses in (4.1) is obtained from the "score" vector. At the k^{th} interim analysis, let the p dimensional score vector or, more generally, "estimating equations" to be used to estimate θ, be denoted by

$$S_k(\theta) = \sum_{i=1}^{n_k} S_{ik}(\theta), \tag{4.2}$$

and let $\hat{\theta}_k$ denote the estimator of θ satisfying $S_k(\hat{\theta}_k) = 0$ if it exists. For example, in the fully sequential method described in Subsection 2.3, we can consider a kind of score vector $S_k(\mu) = \sum_{i=1}^{k}(X_i - \mu)$. By solving the estimating equation $S_k(\hat{\mu}_k) = 0$, it produces the estimator $\hat{\mu}_k = \bar{X}_k$ and the Wald test $\hat{\mu}_k/\mathrm{SE}(\hat{\mu}_k) = S_k/\sqrt{k}$, where SE stands for standard error. Note that, under the null hypothesis of $\mu = 0$, the score vector becomes the score test $S_k = S_k(0)$ which is equivalent to the Wald test.

In fact, the score vector given by (4.2) contains several important estimating equation vectors such as the efficient score vector in (4.11) defined by differentiating a log-likelihood with respect to θ and the "least squares" score vector (4.5) obtained from the least squares estimation method. In the sequel, the explicit form of score vectors will be defined case by case.

To construct a sequential score statistics in the presence of nuisance parameters β, we partition, under the null hypothesis of $\gamma = 0$, the score vector (4.2) as

$$S_k(\theta)|_{\gamma=0} = (S_{k,\gamma}(\beta), S_{k,\beta}(\beta)^{\mathsf{T}})^{\mathsf{T}}$$

where $S_{k,\gamma}(\beta)$ denotes a score function with respect to the treatment effects parameter γ and $S_{k,\beta}(\beta)$ denotes a $(p-1)$ dimensional score vector with respect to the nuisance parameters β. Then as test statistics at the k^{th} interim analysis, one can use the score statistics $S_{k,\gamma}(\hat{\beta}_k)$ and Wald statistics $\hat{\gamma}_k/\text{SE}(\hat{\gamma}_k)$ where $\hat{\beta}_k$ is the restricted estimator of β computed under the null hypothesis and $\hat{\gamma}_k$ is the estimator of γ obtained by solving $S_k(\hat{\theta}_k) = 0$. Though both the Wald and the score tests can be used to test the null hypothesis, we will use mainly the score tests for convenience.

The score statistics $S_{k,\gamma}(\hat{\beta}_k)$ are usually expressed, at least approximately, as a linear combination of the scores $S_{k,\gamma}(\beta)$ and $S_{k,\beta}(\beta)$ so that the joint distribution and the independent increment structure of the sequentially computed score statistics can be established by the distributional properties of $S_{k,\gamma}(\beta)$ and $S_{k,\beta}(\beta)$. For example, the Taylor expansions of $S_{k,\gamma}(\hat{\beta}_k)$ and $S_{k,\beta}(\hat{\beta}_k)$ at $\beta = \beta_0$, when applicable, yield

$$S_{k,\gamma}(\hat{\beta}_k) \simeq S_{k,\gamma}(\beta_0) + S'_{k,\gamma}(\beta_0)(\hat{\beta}_k - \beta_0),$$

$$0 = S_{k,\beta}(\hat{\beta}_k) \simeq S_{k,\beta}(\beta_0) + S'_{k,\beta}(\beta_0)(\hat{\beta}_k - \beta_0),$$

where $S'_{k,\gamma}(\beta_0) = \partial S_{k,\gamma}(\beta)/\partial\beta|_{\beta=\beta_0}$ and $S'_{k,\beta}(\beta_0) = \partial S_{k,\beta}(\beta)/\partial\beta|_{\beta=\beta_0}$. They are combined yielding

$$S_{k,\gamma}(\hat{\beta}_k) \simeq S_{k,\gamma}(\beta_0) - S'_{k,\gamma}(\beta_0)\{S'_{k,\beta}(\beta_0)\}^{-1} S_{k,\beta}(\beta_0).$$

Since the score vector (4.2) depends only on observations accumulated up to stage k and it has a form of sum of independent observations, so do $S_{k,\gamma}(\beta)$ and $S_{k,\beta}(\beta)$. Even in the case of repeated measurement ($d_{ik} \geq 2$), we can define $S_{ik}(\theta)$ to accomodate the dependency in Y_{ik} through such a method used in the generalized least squares estimation, and hence the structure of sum of independent observations will still hold. Therefore, applying the central limit theorem, the joint distribution of the sequential score statistics $S_{k,\gamma}(\hat{\beta}_k)$ as well as those of $S_{k,\gamma}(\beta)$ and $S_{k,\beta}(\beta)$ would be a (asymptotic) multivariate normal distribution under some regularity conditions, and also they are expected to have the independent increment structure. If it is the case, the standard group sequential methods described in Subsection 2.4 can be applied to the score statistics $S_{k,\gamma}(\hat{\beta}_k)$ to carry out testing for the null hypothesis.

In the asymptotic approach, to avoid the problem caused when n_k are random, we assume the data structure described in Scharfstein et al. (1997). That is, at the k^{th} interim analysis, consider the accumulated data set $\{Y_{ik}, i = 1,\ldots,n_k\}$ as $\{(Y_{ik}, I(T_i \leq t_k)), i = 1,\ldots,n\}$ where $I(T_i \leq t_k)$ is defined as 1 if the ith patient has entered the study by the time of the k^{th} interim analysis and 0 otherwise. Then the score vector (4.2) can be written as

$$S_k(\theta) = \sum_{i=1}^{n} S_{ik}(\theta) I(T_i \leq t_k), \tag{4.3}$$

and we can establish the asymptotic results based on the total sample size n. With this in mind, we will use the expression of (4.2) rather than that of (4.3).

The more detailed theory of maximum likelihood and generalized least squares estimation can be found, for example, in Cox and Hinkley (1974) and McCullagh and Nelder (1989), respectively.

4.1 Parametric regression models for independent data

We start with the simple model for independent data. Consider a regression model below:

$$\begin{aligned} Y_{ik} &= X_{ik}\theta + \epsilon_{ik} \\ &= Z_{ik}\gamma + W_{ik}\beta + \epsilon_{ik},\ i = 1,\ldots,n_k\ ;\ k = 1,\ldots,K, \end{aligned} \tag{4.4}$$

where the independent error terms ϵ_{ik} have a common distribution function F and a common density function f with mean zero and variance σ^2. Then the usual least squares score vector, at the k^{th} interim analysis, is defined by

$$S_k(\theta) = \sum_{i=1}^{n_k} X_{ik}^{\mathsf{T}}(Y_{ik} - X_{ik}\theta). \tag{4.5}$$

From (4.5), the partitioned scores of $S_k(\theta)$ under the null hypothesis of $\gamma = 0$ in the presence of a nuisance parameter β are given by

$$S_{k,\gamma}(\beta) = \sum_{i=1}^{n_k} Z_{ik}(Y_{ik} - W_{ik}\beta) \tag{4.6}$$

and

$$S_{k,\beta}(\beta) = \sum_{i=1}^{n_k} W_{ik}^{\mathsf{T}}(Y_{ik} - W_{ik}\beta). \tag{4.7}$$

Under the null hypothesis, the restricted estimator of β satisfying $S_{k,\beta}(\hat{\beta}_k) = 0$ in (4.7) is the least squares estimator denoted by

$$\hat{\beta}_k = \left(\sum_{i=1}^{n_k} W_{ik}^{\mathsf{T}} W_{ik} \right)^{-1} \sum_{i=1}^{n_k} W_{ik}^{\mathsf{T}} Y_{ik}.$$

Plugging it into (4.6), the score statistics $S_{k,\gamma}(\hat{\beta}_k)$ are written as a linear combination of observations Y_{ik} as follows:

$$\begin{aligned} S_{k,\gamma}(\hat{\beta}_k) &= \sum_{i=1}^{n_k} Z_{ik}(Y_{ik} - W_{ik}\hat{\beta}_k) \\ &= \sum_{i=1}^{n_k} Z_{ik} Y_{ik} - \left(\sum_{i=1}^{n_k} Z_{ik} W_{ik} \right) \left(\sum_{i=1}^{n_k} W_{ik}^{\mathsf{T}} W_{ik} \right)^{-1} \sum_{i=1}^{n_k} W_{ik}^{\mathsf{T}} Y_{ik}. \end{aligned} \tag{4.8}$$

Note that we can also express the score statistics (4.8) as one having a form of (4.6),

$$S_{k,\gamma}(\hat{\beta}_k) = S_{k,\gamma}(0) - S'_{k,\gamma}(0)\{S'_{k,\beta}(0)\}^{-1} S_{k,\beta}(0)$$

or equivalently,

$$S_{k,\gamma}(\hat{\beta}_k) = S_{k,\gamma}(0) - \Gamma_{k,\gamma\beta}\Gamma_{k,\beta\beta}^{-1} S_{k,\beta}(0) \tag{4.9}$$

where $\Gamma_{k,\gamma\beta}$ and $\Gamma_{k,\beta\beta}$ are submatrices of the partitioned matrix

$$\Gamma_k = \mathrm{Var}\{(S_{k,\gamma}(0), S_{k,\beta}(0)^{\mathsf{T}})^{\mathsf{T}}\} = \begin{bmatrix} \Gamma_{k,\gamma\gamma} & \Gamma_{k,\gamma\beta} \\ \Gamma_{k,\beta\gamma} & \Gamma_{k,\beta\beta} \end{bmatrix}. \tag{4.10}$$

From the equation (4.8), (4.9) and (4.10), we have

$$E\{S_{k,\gamma}(\hat{\beta}_k)\} = \gamma I_k$$

and

$$\mathrm{Var}\{S_{k,\gamma}(\hat{\beta}_k)\} = I_k$$

where

$$I_k = \Gamma_{k,\gamma\gamma} - \Gamma_{k,\gamma\beta}\Gamma_{k,\beta\beta}^{-1}\Gamma_{k,\beta\gamma} = \left\{ \sum_{i=1}^{n_k} Z_{ik}^2 - \left(\sum_{i=1}^{n_k} Z_{ik} W_{ik} \right) \left(\sum_{i=1}^{n_k} W_{ik}^{\mathsf{T}} W_{ik} \right)^{-1} \left(\sum_{i=1}^{n_k} Z_{ik} W_{ik} \right)^{\mathsf{T}} \right\} \sigma^2.$$

To show the independent increments structure of $S_{k,\gamma}(\hat{\beta}_k)$, we first express $S_{l,\gamma}(0)$ and $S_{l,\beta}(0)$ in the equations (4.9) as sums of two independent variables,

$$S_{l,\gamma}(0) = S_{k,\gamma}(0) + \{S_{l,\gamma}(0) - S_{k,\gamma}(0)\}$$

and

$$S_{l,\beta}(0) = S_{k,\beta}(0) + \{S_{l,\beta}(0) - S_{k,\beta}(0)\}$$

for $k \leq l$. Then we can show that

$$\mathrm{Cov}\{S_{k,\gamma}(0), S_{l,\gamma}(0)\} = \mathrm{Var}\{S_{k,\gamma}(0)\} = \Gamma_{k,\gamma\gamma},$$

$$\mathrm{Cov}\{S_{k,\beta}(0), S_{l,\beta}(0)\} = \mathrm{Var}\{S_{k,\beta}(0)\} = \Gamma_{k,\beta\beta}$$

$$\mathrm{Cov}\{S_{k,\gamma}(0), S_{l,\beta}(0)\} = \mathrm{Cov}\{S_{l,\gamma}(0), S_{k,\beta}(0)\} = \mathrm{Cov}\{S_{k,\gamma}(0), S_{k,\beta}(0)\} = \Gamma_{k,\gamma\beta},$$

and

$$\mathrm{Var}\{S_{l,\beta}(0)\} = \mathrm{Var}\{S_{k,\beta}(0)\} = \Gamma_{k,\beta\beta}.$$

These equations produce the independent increments such that

$$\mathrm{Cov}\{S_{k,\gamma}(\hat{\beta}_k), S_{l,\beta}(\hat{\beta}_l)\} = I_k = \mathrm{Var}\{S_{k,\gamma}(\hat{\beta}_k)\}.$$

We established the independent increments structure of sequentially computed score statistics $S_{k,\gamma}(\hat{\beta}_k)$ without normality assumption for the error distribution. Hence, one might construct the exact sequential tests by replacing the normal density function with an underlying density function f in the methods given in Subsection 2.3. If the asymptotic methods are preferred for a non-normal distribution, we can use the asymptotic results established by the multivariate central limit theorem and the Cramér-Wold device. That is, the asymptotic joint distribution of the sequential score statistics $n^{-1/2}S_{k,\gamma}(\hat{\beta}_k)$, $k = 1, \ldots, K\}$, under the null hypothesis, is multivariate normal with mean 0 and covariance matrix

$$\mathrm{Cov}_A\{n^{-1/2}S_{k,\gamma}(\hat{\beta}_k), n^{-1/2}S_{l,\beta}(\hat{\beta}_l)\} = \mathrm{Var}_A\{n^{-1/2}S_{k,\gamma}(\hat{\beta}_k)\} = \bar{I}_k, \;\; 1 \leq k \leq l \leq K,$$

where Cov_A and Var_A denote asymptotic covariance and variance matrices and

$$\bar{I}_k = \lim_{n\to\infty} n^{-1} I_k.$$

Further, under a class of local alternatives $\{\gamma_n\}$, where $\sqrt{n}\gamma_n \to \delta \neq 0$, we can show that the asymptotic distribution of $n^{-1/2}S_{k,\gamma}(\hat{\beta}_k)$ is normal with mean $\delta\bar{I}_k$ and the same variance as under the null hypothesis.

It should be mentioned that the variance σ^2 has been assumed known. In addition to the known variance case, the asymptotic results are still valid if there exists a consistent estimator of the variance when unknown. Although there are some exact tests such as exact t, χ^2 and F tests proposed by Jennison and Turnbull (1991), we restrict our attention to two cases: the known variance case and the unknown variance case where a consistent estimator exists.

For the regression model (4.4), the score vector (4.5) coincides with the efficient score vector based on the likelihood function when the underlying distribution is normal.

As shown in Jennison and Turnbull (1997) and Scharfstein et al. (1997), the asymptotic joint distribution of the sequentially computed efficient score statistics is multivariate normal with independent increments for more general models.

To summarize their results, we consider the model given in Jennison and Turnbull (1997) where Y_{ik} has a density function $f_{ik}(y_{ik};\theta)$ satisfying some regularity conditions necessary to establish the asymptotic results. Then, for observation i, defining the efficient score $S_{ik}(\theta)$ and information matrix I_{ik} as

$$S_{ik}(\theta) = \frac{\partial}{\partial\theta}\log f_{ik}(Y_{ik};\theta) \tag{4.11}$$

and

$$I_{ik}(\theta) = E\left\{-\frac{\partial}{\partial\theta}S_{ik}(\theta)^{\mathsf{T}}\right\},$$

we have, at the k^{th} interim analysis, the efficient score vector $S_k(\theta) = \sum_{i=1}^{n_k} S_{ik}(\theta)$ and information matrix $I_k(\theta) = \sum_{i=1}^{n_k} I_{ik}(\theta)$. Note that $I_k(\theta) = \text{Var}\{S_k(\theta)\}$. Further, taking $\hat{\beta}_k$ as the restricted maximum likelihood estimator of β under the null hypothesis, the efficient score statistics $S_{k,\gamma}(\hat{\beta}_k)$ can be approximated as, for a fixed β_0 of β,

$$S_{k,\gamma}(\hat{\beta}_k) \simeq S_{k,\gamma}(\beta_0) - I_{k,\gamma\beta}I_{k,\beta\beta}^{-1}S_{k,\beta}(\beta_0)$$

where $I_{k,\gamma\beta}$ and $I_{k,\beta\beta}$ are submatrices of the partitioned matrix

$$I_k\{(0,\beta_0')'\} = \text{Var}\{(S_{k,\gamma}(\beta_0), S_{k,\beta}(\beta_0)^{\mathsf{T}})^{\mathsf{T}}\} = \begin{bmatrix} I_{k,\gamma\gamma} & I_{k,\gamma\beta} \\ I_{k,\beta\gamma} & I_{k,\beta\beta} \end{bmatrix}.$$

Therefore, by applying the same arguments as those for the least squares method, it can be shown that the asymptotic joint distribution of the sequential score statistics $n^{-1/2}S_{k,\gamma}(\hat{\beta}_k), k = 1,\ldots,K\}$, is multivariate normal with mean μ and covariance matrix

$$\text{Cov}_A\{n^{-1/2}S_{k,\gamma}(\hat{\beta}_k), n^{-1/2}S_{l,\beta}(\hat{\beta}_l)\} = \bar{I}_k, \;\; 1 \le k \le l \le K,$$

where $\bar{I}_k = \lim_{n\to\infty} n^{-1}(I_{k,\gamma\gamma} - I_{k,\gamma\beta}I_{k,\beta\beta}^{-1}I_{k,\beta\gamma})$ and μ is 0 under the null hypothesis and $\delta\bar{I}_k$ under local alternatives. The variance matrix $\bar{I}_k$ can be replaced by the consistent estimator based on sample information matrices

$$\hat{I}_{ik}\{(0,\hat{\beta}_k^{\mathsf{T}})^{\mathsf{T}}\} = -\frac{\partial}{\partial\theta}S_{ik}(\theta)^{\mathsf{T}}\Big|_{\gamma=0,\beta=\hat{\beta}_k}.$$

4.2 Longitudinal data

In this subsection, we review selected recently developed methods for group sequential tests which, when properly formulated, turn out to have independent increments, starting with parametric models followed by semiparametric models. We still consider the regression model (4.4) and discuss methods based on the generalized least squares estimates and generalized estimating equations rather than the maximum likelihood estimates.

4.2.1 Parametric regression models

For the model (4.4), assume $d_{ik} \geq 1$ and ϵ_{ik} has mean 0 and variance matrix V_{ik}. Then, based on the generalized least squares methods, the score vector $S_k(\theta)$, the generalized least squares estimator $\hat{\beta}_k$ of β under the null hypothesis and score statistics $S_{k,\gamma}(\hat{\beta}_k)$ are given by

$$S_k(\theta) = \sum_{i=1}^{n_k} X_{ik}^{\mathsf{T}} V_{ik}^{-1} (Y_{ik} - X_{ik}\theta), \tag{4.12}$$

$$\hat{\beta}_k = \left(\sum_{i=1}^{n_k} W_{ik}^{\mathsf{T}} V_{ik}^{-1} W_{ik} \right)^{-1} \sum_{i=1}^{n_k} W_{ik}^{\mathsf{T}} V_{ik}^{-1} Y_{ik},$$

and

$$\begin{aligned} S_{k,\gamma}(\hat{\beta}_k) &= \sum_{i=1}^{n_k} Z_{ik}^{\mathsf{T}} V_{ik}^{-1} (Y_{ik} - W_{ik}\hat{\beta}_k) \\ &= \sum_{i=1}^{n_k} Z_{ik}^{\mathsf{T}} V_{ik}^{-1} Y_{ik} - \left(\sum_{i=1}^{n_k} Z_{ik}^{\mathsf{T}} V_{ik}^{-1} W_{ik} \right) \left(\sum_{i=1}^{n_k} W_{ik}^{\mathsf{T}} V_{ik}^{-1} W_{ik} \right)^{-1} \sum_{i=1}^{n_k} W_{ik}^{\mathsf{T}} V_{ik}^{-1} Y_{ik}. \end{aligned}$$

The partitioned scores $S_{k,\gamma}(\beta)$, $S_{k,\beta}(\beta)$ of $S_k(\theta)$ under the null and variance I_k of $S_{k,\gamma}(\hat{\beta}_k)$ are similarly defined as

$$S_{k,\gamma}(\beta) = \sum_{i=1}^{n_k} Z_{ik}^{\mathsf{T}} V_{ik}^{-1} (Y_{ik} - W_{ik}\beta),$$

$$S_{k,\beta}(\beta) = \sum_{i=1}^{n_k} W_{ik}^{\mathsf{T}} V_{ik}^{-1} (Y_{ik} - W_{ik}\beta)$$

and

$$\begin{aligned} I_k &= \Gamma_{k,\gamma\gamma} - \Gamma_{k,\gamma\beta}\Gamma_{k,\beta\beta}^{-1}\Gamma_{k,\beta\gamma} \\ &= \sum_{i=1}^{n_k} Z_{ik}^{\mathsf{T}} V_{ik}^{-1} Z_{ik} - \left(\sum_{i=1}^{n_k} Z_{ik}^{\mathsf{T}} V_{ik}^{-1} W_{ik} \right) \left(\sum_{i=1}^{n_k} W_{ik}^{\mathsf{T}} V_{ik}^{-1} W_{ik} \right)^{-1} \left(\sum_{i=1}^{n_k} Z_{ik}^{\mathsf{T}} V_{ik}^{-1} W_{ik} \right)^{\mathsf{T}} \end{aligned} \tag{4.13}$$

where Γ_k is defined and partitioned the same as (4.10).

Following the arguments developed by Lee, Kim and Tsiatis (1996), we can establish the joint distribution of sequentially computed score statistics $n^{-1/2}S_{k,\gamma}(\hat{\beta}_k)$, $k = 1,\ldots,K$, and can show an independent increments structure. When the underlying distribution is normal, the joint distribution of $S_{k,\gamma}(\hat{\beta}_k)$, $k = 1,\ldots,K$, is multivariate normal with mean μ and

$$\mathrm{Cov}\{S_{k,\gamma}(\hat{\beta}_k), S_{l,\beta}(\hat{\beta}_l)\} = \mathrm{Var}\{S_{k,\gamma}(\hat{\beta}_k)\} = I_k,\ 1 \le k \le l \le K$$

where μ is 0 under H_0 and $\delta\bar{I}_k$ under local alternatives. Furthermore, under suitable regularity conditions for a non-normal underlying distribution, the asymptotic joint distribution of $n^{-1/2}S_{k,\gamma}(\hat{\beta}_k)$, $k = 1,\ldots,K$, is multivariate normal with mean μ and covariance matrix

$$\mathrm{Cov}_A\{n^{-1/2}S_{k,\gamma}(\hat{\beta}_k), n^{-1/2}S_{l,\beta}(\hat{\beta}_l)\} = \mathrm{Var}_A\{n^{-1/2}S_{k,\gamma}(\hat{\beta}_k)\} = \bar{I}_k,\ 1 \le k \le l \le K$$

where $\bar{I}_k = \lim_{n\to\infty} n^{-1}I_k$ and μ is 0 under H_0 and $\delta\bar{I}_k$ under local alternatives. When I_k or $\bar{I}_k$ is unknown, a consistent estimator can be obtained from (4.13) by substituting V_{ik} with $(Y_{ik} - X_{ik}\hat{\theta}_k)(Y_{ik} - X_{ik}\hat{\theta}_k)^\top$ where $\hat{\theta}_k$ is the generalized least squares estimator of θ.

A random effects model can be also applied to construct a sequential procedure to test the null hypothesis $H_0 : \gamma = 0$. Instead of the model (4.4), consider a random effects model

$$Y_{ik} = Z_{ik}\gamma + W_{ik}\beta_i + \epsilon_{ik},\ i = 1,\ldots,n_k \text{ and } k = 1,\ldots,K$$

where γ is a fixed effect parameter, $\epsilon_{ik} \sim N(0,\Sigma_{ik})$ and $\beta_i \sim N(\beta,\Sigma_\beta)$ are all independent. The parameter β_i can be interpreted as participant effects parameter. This model can also be written as $Y_{ik} = Z_{ik}\gamma + W_{ik}\beta + W_{ik}\beta_i^* + \epsilon_{ik}$ where β is also fixed and $\beta_i^* \sim N(0,\Sigma_\beta)$ so that it is included in the model (4.4) with $V_{ik} = W_{ik}\Sigma_\beta W_{ik}^\top + \Sigma_{ik}$.

4.2.2 Semiparametric models

In this subsubsection, we review the results of Lee et al. (1996). Assume that at the k^{th} interim analysis, the marginal mean of Y_{ik} given X_{ik} is

$$E(Y_{ik}|X_{ik}) = \mu_{ik}(\theta) = g(X_{ik},\theta)$$

where g is a known function. Denote a working variance to be used instead of the unknown true variance $V_{ik} = \mathrm{Var}(Y_{ik}|X_{ik})$ by $v_{ik}(\theta,\alpha)$ with additional variance parameters α. Then, the score vector or generalized estimating equations has the form

$$S_k(\theta,\alpha) = \sum_{i=1}^{n_k} S_{ik}(\theta,\alpha) = \sum_{i=1}^{n_k} D_{ik}(\theta)^\top v_{ik}^{-1}(\theta,\alpha)(Y_{ik} - \mu_{ik}(\theta))$$

where $D_{ik}(\theta) = \partial \mu_{ik}(\theta)/\partial\theta$. Note that this score vector is reduced to (4.12) when $\mu_{ik}(\theta) = X_{ik}\theta$ and $v_{ik}(\theta,\alpha) = V_{ik}$, and hence, it can be regarded as a generalization of the least squares methods in Subsection 4.1.

When a consistent estimator $\hat{\alpha}$ of α is available, Liang and Zeger (1986) showed that $S_k(\theta,\hat{\alpha})$ is asymptotically equivalent to $S_k(\theta,\alpha)$, and hence the asymptotic properties regarding the inference on θ remain unchanged when using $S_k(\theta,\hat{\alpha})$ instead of $S_k(\theta,\alpha)$. We assume that α is known or a consistent estimator $\hat{\alpha}$ is available, and denote generalized estimating equations estimators of θ for both cases as the same $\hat{\theta}_k$. Note that $\hat{\theta}_k$ is consistent. For more details about estimation of α, refer to, for example, Crowder (1995) and Lee et al. (1996).

Partition $S_k(\theta,\alpha)$ as $\{S_{k,\gamma}(\gamma,\beta,\alpha), S_{k,\beta}(\gamma,\beta,\alpha)^{\intercal}\}^{\intercal}$ and let $\hat{\beta}_k$ be the restricted generalized estimating equations estimator of β under the null hypothesis. Then, as shown by Rotnitzky and Jewell (1990), it can be shown that the score statistic $S_{k,\gamma}(0,\hat{\beta}_k,\alpha)$ is asymptotically equivalent to $T_k(0,\beta,\alpha)$ where

$$T_k(\gamma,\beta,\alpha) = S_{k,\gamma}(\gamma,\beta,\alpha) - \Gamma_{k,\gamma\beta}\Gamma_{k,\beta\beta}^{-1} S_{k,\beta}(\gamma,\beta,\alpha) \tag{4.14}$$

and $\Gamma_{k,\gamma\beta}$ and $\Gamma_{k,\beta\beta}$ are submatrices of the partitioned matrix of Γ_k,

$$\Gamma_k = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n_k} D_{ik}(\theta)^{\intercal} v_{ik}^{-1}(\theta,\alpha) D_{ik}(\theta) = \begin{bmatrix} \Gamma_{k,\gamma\gamma} & \Gamma_{k,\gamma\beta} \\ \Gamma_{k,\beta\gamma} & \Gamma_{k,\beta\beta} \end{bmatrix}.$$

That is,

$$S_{k,\gamma}(0,\hat{\beta}_k,\alpha) \simeq S_{k,\gamma}(0,\beta,\alpha) - \Gamma_{0k,\gamma\beta}\Gamma_{0k,\beta\beta}^{-1} S_{k,\beta}(0,\beta,\alpha) = T_k(0,\beta,\alpha) \tag{4.15}$$

where the subscript 0 means "evaluated at $\gamma = 0$", or equivalently, "evaluated under the null hypothesis".

Since the score vector $n^{-1/2}S_k(\theta,\alpha)$ has a form of sum of independent variables, the asymptotic distribution of $n^{-1/2}S_k(\theta\alpha)$ is multivariate normal with mean 0 and variance

$$\Omega_k = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n_k} D_{ik}(\theta)^{\intercal} v_{ik}^{-1}(\theta,\alpha) V_{ik}\, v_{ik}^{-1}(\theta,\alpha) D_{ik}(\theta).$$

By asymptotic normality of $n^{-1/2}S_k(\theta,\alpha)$ together with the linear equation (4.14), the asymptotic joint distribution of $\{n^{-1/2}T_k(\gamma,\beta,\alpha), k = 1,\ldots,K\}$ becomes multivariate normal with mean 0. The asymptotic covariance of $n^{-1/2}T_k(\gamma,\beta,\alpha)$ and $n^{-1/2}T_l(\gamma,\beta,\alpha)$, for $k \le l$, is given by

$$M_{kl} = \Omega_{kl,\gamma\gamma} + \Gamma_{k,\gamma\beta}\Gamma_{k,\beta\beta}^{-1}\Omega_{kl,\beta\beta}\Gamma_{l,\beta\beta}^{-1}\Gamma_{l,\beta\gamma} - \Omega_{kl,\gamma\beta}\Gamma_{l,\beta\beta}^{-1}\Gamma_{l,\beta\gamma} - \Gamma_{k,\gamma\beta}\Gamma_{k,\beta\beta}^{-1}\Omega_{kl,\beta\gamma}$$

where $\Omega_{kl,\gamma\gamma}$, $\Omega_{kl,\gamma\beta}$, $\Omega_{kl,\beta\gamma}$ and $\Omega_{kl,\beta\beta}$ are submatrices of the partitioned matrix

$$\Omega_{kl} = \begin{bmatrix} \Omega_{kl,\gamma\gamma} & \Omega_{kl,\gamma\beta} \\ \Omega_{kl,\beta\gamma} & \Omega_{kl,\beta\beta} \end{bmatrix}$$

and

$$\Omega_{kl} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n_k} D_{ik}(\theta)^{\mathsf{T}} v_{ik}^{-1}(\theta,\alpha) V_{ikl}\, v_{il}^{-1}(\theta,\alpha) D_{il}(\theta).$$

Note that V_{ikl} denotes the true covariance matrix of Y_{ik} and Y_{il}. When the true variance functions are correctly specified, as shown in Lee et al. (1996), the asymptotic covariances M_{kl} $(1 \le k \le l \le K)$ are reduced to I_k,

$$I_k = \Gamma_{k,\gamma\gamma} - \Gamma_{k,\gamma\beta}\Gamma_{k,\beta\beta}^{-1}\Gamma_{k,\beta\gamma} = \mathrm{Var}_A\{n^{-1/2}T_k(\gamma,\beta,\alpha)\},$$

indicating an asymptotic independent increments structure.

By applying similar arguments to the equation (4.15), it can be shown that the asymptotic joint distribution of sequential score statistics $n^{-1}S_{k,\gamma}(0,\hat{\beta}_k,\alpha)$, $k = 1,\ldots,K$, is multivariate normal with mean 0 and covariance M_{0kl}, $k,l = 1,\ldots,K$. Furthermore, with a correct specification of the variance functions, we have $M_{0kl} = I_{0k} = \mathrm{Var}_A\{n^{-1/2}S_{k,\gamma}(0,\hat{\beta}_k,\alpha)\}$, which establish an asymptotic independent increments structure of sequentially computed score statistics.

The asymptotic variances Γ_k and Ω_k can be estimated consistently by evaluating $D_{ik}(\theta)$ and $v_{ik}(\theta,\alpha)$ at the consistent estimators $\hat{\alpha}$ and $\hat{\theta}_k$ and by substituting $\{Y_{ik} - \mu_{ik}(\hat{\theta}_k)\}\{Y_{ik} - \mu_{ik}(\hat{\theta}_k)\}^{\mathsf{T}}$ for V_{ik}. Under the null hypothesis, we use $\hat{\theta}_k = (0,\hat{\beta}_k^{\mathsf{T}})^{\mathsf{T}}$. As pointed out by Lee et al. (1996), these consistent estimators also lead to an asymptotic independent increments structure of sequentially computed $n^{-1/2}T_k(\gamma,\beta,\alpha)$ and $n^{-1}S_{k,\gamma}(0,\hat{\beta}_k,\alpha)$ when the variance functions are correctly specified.

4.3 Failure time data

In this subsection, we review the results for a general parametric model, Cox proportional hazards model by Cox (1972), and accelerated failure time model of Lin (1992), in the framework of counting process and martingale integration which can be referred to, for example, Fleming and Harrington (1991) and Anderson et al. (1993)

First, consider the notations for failure time data. Assume that n patients enter the trial at times $e_1,\ldots,e_n$ which are considered as constants. Each patient i has a potential failure time T_i, potential censoring time C_i, treatment indicator Z_i and covariate vector $W_i = (W_{i1},\cdots,W_{ip})$. It is assumed that T_i and C_i are conditionally independent given $\{Z_i,W_i\}$ and $\{T_i,C_i,Z_i,W_i\}$, $i = 1,\ldots,n$, are identically and independently distributed. If the data were analyzed at time t, the observable random variables would be $\{X_i(t),\Delta_i(t),Z_i,W_i\}$ for all $i = 1,\ldots,n$ such that $e_i \le t$. Here $X_i(t) = \min(T_i,C_i,t-e_i)$ is the time to failure or censoring, and $\Delta_i(t) = I\{T_i < \min(C_i,t-e_i)\}$ denotes the failure

indicator. For simplicity, we assume that the covariate vector W_i is time-invariant, but the same results are obtained for a time-varying covariate, as shown in Gu and Ying (1995) for the proportional hazards model.

We assume a hazard function $\lambda(u,Z_i,W_i,\theta)$ with $\theta=(\gamma,\beta^{\mathsf{T}})^{\mathsf{T}}$ where γ is a treatment effect parameter and β is a vector of nuisance parameters denoting covariate effects. As in the previous sections, we are interested in testing the null hypotheses (4.1). For notational simplicity, we set $R_i=(Z_i,W_i^{\mathsf{T}})^{\mathsf{T}}$.

It is convenient to express the failure time data in terms of counting process notation. Define the counting process of observed death for the ith patient at analysis time t by $N_i(u,t)=I\{X_i(t)\le u,\Delta_i(t)=1\}$ for $u\ge 0$. Note that whenever $e_i>t$, $N_i(u,t)=0$ for $u\ge 0$ and when $e_i\le t$, $N_i(u,t)=N_i(t,t)$ for $u\ge t$. Similarly, the at-risk process $Y_i(u,t)=I\{X_i(t)\ge u\}$, which is the indicator of whether the ith patient is at risk u units after entry into the study if the data were analyzed at calendar time t.

With the filtration $\mathcal{F}(u)$, $u\ge 0$, defined by Tsiatis et al. (1995), denote the $\mathcal{F}(u)$ martingale process associated with $N_i(u,t)$ by

$$M_i(u,t)=N_i(u,t)-\int_0^u \lambda(x,R_i,\theta)Y_i(x,t)dx.$$

Also, following the same procedure as in Tsiatis et al. (1995), we define the counting process of death observed between two successive analysis times t_k and t_{k-1} as $DN_i(u,t_1)=N_i(u,t_1)$ and $DN_i(u,t_k)=N_i(u,t_k)-N_i(u,t_{k-1})$, $k=2,\ldots,K$, where $t_1<\cdots<t_K$ denote the analysis times. Let $DY_i(u,t_k)=Y_i(u,t_k)-Y_i(u,t_{k-1})$, then the martingale process associated with $DN_i(u,t)$ can be written by

$$DM_i(u,t)=DN_i(u,t)-\int_0^u \lambda(x,R_i,\theta)DY_i(x,t)dx.$$

Note that since any two processes of $DN_i(u,t_k)$, $k=1,\ldots,K$, will not take jumps at the same time, $DM_i(u,t_k)$ and $DM_i(u,t_l)$ are orthogonal, that is, $\mathrm{Cov}\{DM_i(u,t_k),DM_i(u,t_l)\}=0$ if $k\ne l$. Also, note that $M_i(u,t_k)=\sum_{j=1}^k DM_i(u,t_j)$ for $k=1,\ldots,K$. So far, we defined the data structure, and related counting processes and martingales. We will use this common notations in the next subsubsections.

4.3.1 Parametric regression models

Assume that the hazard function $\lambda(u,R_i,\theta)$ is known, then using standard results for failure time data, the likelihood of the data available at time t is proportional to

$$L(t,\theta)=\Pi_{(i:e_i<t)}[\lambda\{X_i(t),R_i,\theta\}]^{\Delta_i(t)}\exp\left\{-\int_0^{X_i(t)}\lambda(u,R_i,\theta)du\right\}.$$

With the counting process notations, we can express the score vector at analysis time t as

$$\begin{aligned} S(t,\theta) &= \sum_{i=1}^{n} \int_0^{\infty} h(u,R_i,\theta)\{dN_i(u,t) - d\mu_i(u,t,\theta)\} \\ &= \sum_{i=1}^{n} \int_0^{\infty} h(u,R_i,\theta) dM_i(u,t) \end{aligned} \tag{4.16}$$

or equivalently, at analysis time t_k,

$$\begin{aligned} S(t_k,\theta) &= \sum_{j=1}^{k} \sum_{i=1}^{n} \int_0^{\infty} h(u,R_i,\theta)\{dDN_i(u,t_j) - dD\mu_i(u,t,\theta)\} \\ &= \sum_{j=1}^{k} \sum_{i=1}^{n} \int_0^{\infty} h(u,R_i,\theta) dDM_i(u,t_j) \end{aligned} \tag{4.17}$$

where

$$h(u,R_i,\theta) = \partial \log \lambda(u,R_i,\theta)/\partial\theta,$$
$$d\mu_i(u,t,\theta) = \lambda(u,R_i,\theta) Y_i(u,t) du$$

and

$$dD\mu_i(u,t,\theta) = \lambda(u,R_i,\theta) DY_i(u,t) du.$$

Denote the part taken from the second sum in (4.17) by S_j. Then, by the standard arguments for counting processes, e.g. in Fleming and Harrington (1991), the vector of martingale integrals S_j is also martingale, and we have

$$E\{dN_i(u,t)|R_i\} = d\mu_i(u,t,\theta)$$

and

$$E\{dDN_i(u,t)|R_i\} = dD\mu_i(u,t,\theta)$$

so that

$$E\{dM_i(u,t)|R_i\} = E\{dDM_i(u,t)|R_i\} = 0,$$
$$\mathrm{Var}\{dN_i(u,t)|R_i\} = \mathrm{Var}\{dM_i(u,t)|R_i\} = d\mu_i(u,t,\theta),$$

and

$$\mathrm{Var}\{dDN_i(u,t)|R_i\} = \mathrm{Var}\{dDM_i(u,t)|R_i\} = dD\mu_i(u,t,\theta).$$

Hence, S_j has mean 0 and variance

$$\mathrm{Var}(S_j) = \sum_{i=1}^{n} \int_0^{\infty} h(u,R_i,\theta) h(u,R_i,\theta)^{\mathsf{T}} dD\mu_i(u,t_j,\theta).$$

Furthermore, since $\mathrm{Cov}\{DM_i(u,t_k), DM_i(u,t_l)\} = 0$ if $k \neq l$ and observations are independent, S_j, $j = 1,\ldots,k$, are uncorrelated, and hence the score vector $S(t_k,\theta)$ has a similar form to (4.2), sum of uncorrelated variables. Applying the martingale central limit theorem, we can show that the asymptotic joint distribution of $n^{-1/2}S(t_k,\theta)$, $k = 1,\ldots,K$, is multivariate normal with mean 0. It can also be shown that

$$\mathrm{Cov}_A\{n^{-1/2}S(t_k,\theta), n^{-1/2}S(t_l,\theta)\} = \mathrm{Var}_A\{n^{-1/2}S(t_k,\theta)\} = \Gamma(t_k),\ 1 \leq k \leq l \leq K,$$

which indicates an independent increments structure. Here the asymptotic variance

$$\Gamma(t_k) = \lim_{n\to\infty} \int_0^\infty n^{-1} \sum_{i=1}^{n} h(u,R_i,\theta)h(u,R_i,\theta)^{\tau} d\mu_i(u,t_k,\theta). \tag{4.18}$$

Now, partition the score vector $S(t,\theta)$ as $\{S_\gamma(t,\gamma,\beta), S_\beta(t,\gamma,\beta)^{\tau}\}^{\tau}$, where $S_\gamma(t,\gamma,\beta) = \partial \log L(t,\gamma,\beta)/\partial\gamma$ and $S_\beta(t,\gamma,\beta) = \partial \log L(t,\gamma,\beta)/\partial\beta$. Then the score test of the null hypothesis $H_0 : \gamma = 0$ in the presence of nuisance parameters β, evaluated at calendar time t, is given by $S_\gamma(t,0,\hat{\beta}_t)$ where $\hat{\beta}_t$ is the restricted maximum likelihood estimator of β when $\gamma = 0$. Using standard results of likelihood theory, Cox and Hinkley (1974, Sec 9.3), the score test $S_\gamma(t,0,\hat{\beta}_t)$ is asymptotically equivalent to

$$T(t,0,\beta) = S_\gamma(t,0,\beta) - \Gamma_{\gamma\beta}(t)\Gamma_{\beta\beta}^{-1}(t)S_\beta(t,0,\beta)$$

where $\Gamma_{\gamma\beta}$ and $\Gamma_{\beta\beta}$ are submatrices of the partitioned matrix of $\Gamma_0(t)$, which is $\Gamma(t)$ in (4.18) evaluated at $\gamma = 0$,

$$\Gamma_0(t) = \begin{bmatrix} \Gamma_{\gamma\gamma}(t) & \Gamma_{\gamma\beta}(t) \\ \Gamma_{\beta\gamma}(t) & \Gamma_{\beta\beta}(t) \end{bmatrix}.$$

Since $n^{-1/2}T(t,0,\beta)$ is a linear combination of the elements of the score vector $n^{-1/2}$ $S(t,\theta)$, which converges in distribution to a multivariate normal with independent increments, this implies that $n^{-1/2}T(t,0,\beta)$ also converges in distribution to a normal with mean μ and variance $I(t)$, where $I(t) = \Gamma_{\gamma\gamma}(t) - \Gamma_{\gamma\beta}(t)\Gamma_{\beta\beta}^{-1}(t)\Gamma_{\beta\gamma}(t)$, and $\mu = 0$ under the null hypothesis and $\mu = \delta I(t)$ under the local alternatives defined in Subsubsection 4.2.1. Therefore, following the same arguments as in the previous sections, we can show that the asymptotic joint distribution of $\{n^{-1/2}S_\gamma(t_k,0,\hat{\beta}_{t_k}), k = 1,\ldots,K\}$ is multivariate normal with mean μ and covariance $(1 \leq k \leq l \leq K)$

$$\mathrm{Cov}_A\{n^{-1/2}S_\gamma(t_k,0,\hat{\beta}_{t_k}), n^{-1/2}S_\gamma(t_l,0,\hat{\beta}_{t_l})\} = \mathrm{Var}_A\{n^{-1/2}S_\gamma(t_k,0,\hat{\beta}_{t_k})\} = I(t_k), \tag{4.19}$$

which implies an independent increments structure of the asymptotic joint distribution.

Note that $h(u,R_i,\theta)$ does not depend on the calendar time t, and this make it much easier to establish the independent increments structure in (4.18) and (4.19). In fact, it can be shown that when we use a weighted score vector with a weight function

$Q(u,t,\theta)$ which converges in probability to a function $q(u,t,\theta)$, the independent increments structure holds as long as $q(u,t,\theta)$ does not depend on the calendar time t. Therefore, choosing a suitable weight function, we can construct a group sequential test having asymptotic normality and independent increments structure. This may be particularly useful when the efficient test is difficult to be built explicitly, as found in the accelerated failure time model. For weighted tests, the limiting optimal weight function is proportional to the limit of $h(u,R_i,\theta) = \partial \log \lambda(u,R_i,\theta)/\partial\theta$ because the score vectors in (4.16) and (4.17) are efficient scores. Tsiatis (1982) and Lin (1992) also showed, for the proportional hazards model and accelerated failure time model, that the limiting weight functions preserve the independent increments structure.

It is also interesting to note that the score vectors given by (4.16) and (4.17) can be regarded as the score vectors based on the generalized estimating equations accommodating time dependent structure of the failure time data by a stochastic integral. Considering $dN_i(u,t)$ as the ith observation and expressing $h(u,R_i,\theta)$ in (4.16) as

$$h(u,R_i,\theta) = \{\partial d\mu_i(u,t,\theta)/\partial\theta\}/\mathrm{Var}\{dN_i(u,t)|R_i\},$$

we have the generalized estimating equations

$$S(t,\theta) = \sum_{i=1}^{n} \int_0^\infty n^{-1}\{\partial d\mu_i(u,t,\theta)/\partial\theta\}\left[\mathrm{Var}\{dN_i(u,t)|R_i\}\right]^{-1}\{dN_i(u,t) - d\mu_i(u,t,\theta)\}.$$

In this framework, choosing a weight function corresponds to choosing a working variance.

4.3.2 Proportional hazards models

Consider the Cox proportional hazards model where the hazard function $\lambda(u,R_i,\theta)$ is given by

$$\lambda(u,R_i,\theta) = \lambda_0(u)\exp(\theta^{\mathrm{T}}R_i)$$

where λ_0 is an arbitrary baseline hazard fuction. We can express the score vector based on the partial likelihood (Cox, 1975) at the analysis time t as

$$\begin{aligned} U(t,\theta) &= \sum_{i=1}^{n} \int_0^\infty \{R_i - \bar{R}(u,t,\theta)\} dN_i(u,t) \\ &= \sum_{i=1}^{n} \int_0^\infty \{R_i - \bar{R}(u,t,\theta)\} dM_i(u,t) \end{aligned} \tag{4.20}$$

where $\bar{R}(u,t,\theta) = \sum_{i=1}^n R_i Y_i(u,t)\exp(\theta^{\mathrm{T}}R_i)/\sum_{i=1}^n Y_i(u,t)\exp(\theta^{\mathrm{T}}R_i)$.

The partial likelihood score vector (4.20) has the same form as the maximum likelihood score vector (4.16) if $h(u,R_i,\theta)$ in (4.16) is replaced with $R_i - \bar{R}(u,t,\theta)$. Though $R_i - \bar{R}(u,t,\theta)$ may depend on the calendar time t, as shown in Jennison and Turnbull

(1997), the independent increments structure of the score vector $U(t,\theta)$ still holds. Therefore, the arguments in Subsubsection 4.3.1 can be applied to produce the following results:

Denote the score test statistic for the null hypothesis by $U_\gamma(t,0,\hat{\beta})$ where $U_\gamma(t,\gamma,\beta)$ is the first element of the partitioned score vector $U(t,\theta) = \{U_\gamma(t,\gamma,\beta), U_\beta(t,\gamma,\beta)^\top\}^\top$ and $\hat{\beta}$ is the restricted maximum partial likelihood estimator of β when $\gamma = 0$. Then the asymptotic joint distribution of $\{n^{-1/2}U_\gamma(t_k,0,\hat{\beta}_{t_k}), k = 1,\ldots,K\}$ is multivariate normal with mean μ and covariance $(1 \le k \le l \le K)$

$$\mathrm{Cov}_A\{n^{-1/2}U_\gamma(t_k,0,\hat{\beta}_{t_k}), n^{-1/2}U_\gamma(t_l,0,\hat{\beta}_{t_l})\} = \mathrm{Var}_A\{n^{-1/2}U_\gamma(t_k,0,\hat{\beta}_{t_k})\} = I(t_k),$$

which implies an independent increments structure of the asymptotic joint distribution. Here, $I(t) = \Gamma_{\gamma\gamma}(t) - \Gamma_{\gamma\beta}(t)\Gamma^{-1}_{\beta\beta}(t)\Gamma_{\beta\gamma}(t)$ and

$$\Gamma_0(t) = \begin{bmatrix} \Gamma_{\gamma\gamma}(t) & \Gamma_{\gamma\beta}(t) \\ \Gamma_{\beta\gamma}(t) & \Gamma_{\beta\beta}(t) \end{bmatrix},$$

which is obtained by evaluating, at $\gamma = 0$, $\Gamma(t)$,

$$\Gamma(t) = \lim_{n\to\infty} \int_0^\infty n^{-1} \sum_{i=1}^n \{R_i - \bar{R}(u,t,\theta)\}\{R_i - \bar{R}(u,t,\theta)\}^\top Y_i(u,t)\lambda_0(u)\exp(\theta^\top R_i)du.$$

The variance matrix $I(t)$ can be consistently estimated by substituting β and $\lambda_0(u)$ in $\Gamma_0(t)$ with $\hat{\beta}_t$ and the Breslow estimator evaluated under the null hypothesis,

$$\hat{\lambda}_0(u,0,\hat{\beta}_t) = \sum_{i=1}^n dN_i(u,t) \Big/ \sum_{i=1}^n Y_i(u,t)\exp(\hat{\beta}_t^\top W_i).$$

In general, the Breslow estimator is given by

$$\hat{\lambda}_0(u,\hat{\theta}) = \sum_{i=1}^n dN_i(u,t) \Big/ \sum_{i=1}^n Y_i(u,t)\exp(\hat{\theta}^\top R_i)$$

where $\hat{\theta}$ is the maximum partial likelihood estimator of θ, and if there are no covariates, it becomes the Nelson-Aalen estimator by Aalen (1978).

As mentioned in Subsection 4.1, we can consider the weighted score vector $U_Q(t,\theta)$ with a weight function $Q(u,t,\theta)$,

$$U_Q(t,\theta) = \sum_{i=1}^n \int_0^\infty Q(u,t,\theta)\{R_i - \bar{R}(u,t,\theta)\}dN_i(u,t),$$

and we can show that the independent increments structure of sequentially computed score statistics holds when $Q(u,t,\theta)$ converges in probability to a limit $q(u,\theta)$ free of t.

When there are no covariates, the weighted score vector leads to the well known two-sample weighted logrank tests which were studied by Tsiatis (1982).

For a given θ, let $\hat{\lambda}_0(u,\theta)$ denote the Breslow estimator. Then, comparing the partial likelihood score vector (4.20) with the maximum likelihood score vector (4.16), we can show that $U(t,\theta) = S(t,\theta,\hat{\lambda}_0)$ where $S(t,\theta,\hat{\lambda}_0)$ is the score vector obtained by replacing $\lambda_0(u)$ with $\hat{\lambda}_0(u,\theta)$ in (4.16) or (4.17). It seems that the score vector (4.16) can also be expressed in the generalized estimating equations framework as

$$U(t,\theta) = \sum_{i=1}^{n} \int_0^\infty \{\partial d\hat{\mu}_i(u,t,\theta)/\partial\theta\} \left[\widehat{\mathrm{Var}}\{dN_i(u,t)|R_i\}\right]^{-1} \{dN_i(u,t) - d\mu_i(u,t,\theta)\}$$

where $d\hat{\mu}_i(u,t,\theta) = \widehat{\mathrm{Var}}\{dN_i(u,t)|R_i\} = Y_i(u,t)\hat{\lambda}_0(u,\theta)\exp(\theta^{\mathsf{T}}R_i)$. By the consistency of the Breslow estimator $\hat{\lambda}_0(u,\theta)$, this score vector can be regarded as the generalized estimating equations score vector obtained when the true variances are consistently estimated.

4.3.3 Accelerated failure time models

Consider the linear model

$$T_i = \theta^{\mathsf{T}}R_i + \epsilon_i,\ i = 1,\ldots,n$$

where ϵ_i are independent with a common hazard function λ_0. Here, T_is are usually log transformed observation of the original nonnegative failure time data so that they are allowed to have negative values. Further, assume that the treatment indicator Z_i is independent of the covariates W_i as in usual clinical trials.

For a given λ_0, the efficient score vector (4.16) can be written as

$$S(t,\theta) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} R_i\{\lambda_0'(u-\theta^{\mathsf{T}}R_i)/\lambda_0(u-\theta^{\mathsf{T}}R_i)\}\{dN_i(u,t) - Y_i(u,t)\lambda_0(u-\theta^{\mathsf{T}}R_i)du\}.$$

When λ_0 is unknown, replacing $\lambda_0(u)$ with $\hat{\lambda}_0(u,t,\theta)$ and $\lambda_0'(u)/\lambda_0(u)$ with a weight function $Q(u,\theta)$, we have

$$S(t,\theta,\hat{\lambda}_0) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} Q(u,\theta)R_i\{dN_i(u+\theta^{\mathsf{T}}R_i,t) - Y_i(u+\theta^{\mathsf{T}}R_i,t)\hat{\lambda}_0(u,\theta)du\},$$

where $\hat{\lambda}_0(u,t,\theta)du = \sum_{i=1}^n dN_i(u+\theta^{\mathsf{T}}R_i,t)/\sum_{i=1}^n Y_i(u+\theta^{\mathsf{T}}R_i,t)$ is a Nelson-Aalen type estimator of $\lambda_0(u)$ at the analysis time t. Furthermore, let

$$\bar{R}(u,t,\theta) = \sum_{i=1}^{n} R_iY_i(u+\theta^{\mathsf{T}}R_i,t)/\sum_{i=1}^{n} Y_i(u+\theta^{\mathsf{T}}R_i,t).$$

Then $S(t,\theta,\hat{\lambda}_0)$ is equivalent to a rank score vector

$$\begin{aligned} U(t,\theta) &= \sum_{i=1}^{n}\int_{-\infty}^{\infty} Q(u,\theta)\{R_i-\bar{R}(u,t,\theta)\}dN_i(u+\theta^{\mathsf{T}}R_i,t) \\ &= \sum_{i=1}^{n}\int_{-\infty}^{\infty} Q(u,\theta)\{R_i-\bar{R}(u,t,\theta)\}dM_i(u+\theta^{\mathsf{T}}R_i,t) \end{aligned} \tag{4.21}$$

where $M_i(u+\theta^{\mathsf{T}}R_i,t) = N_i(u+\theta^{\mathsf{T}}R_i,t) - \int_{-\infty}^{u} Y_i(x+\theta^{\mathsf{T}}R_i,t)\lambda_0(x)dx$ is a martingale associated with the counting process $N_i(u+\theta^{\mathsf{T}}R_i,t)$ of the residual $X_i(t)-\theta^{\mathsf{T}}R_i$. This class of linear rank tests (4.21) were studied by Tsiatis (1990), Ritov (1990), Wei, Ying and Lin (1990b), and Lin (1992). As shown in Tsiatis (1990), note that the limiting optimal weight function is proportional to $\lambda_0'(u)/\lambda_0(u)$. This rank score vector can also be interpreted in the generalized estimating equations framework, as described in Subsubsection 4.3.2.

At a glance, it seems that the rank score (4.21) has the same form as those of efficient scores for the parametric model and the proportional hazards model, and hence that the same arguments as discussed in the previous sections can be applied. However, as pointed out in several researches such as Tsiatis (1990) and Lin, Wei and Ying (1998), because the rank score is a step function of θ, any exact solution of $U(t,\hat{\theta})=0$ may not exist. Therefore, $\hat{\theta}$ is defined as a value θ for which $U(t,\theta)$ changes sign or as a minimizer of $\| U(t,\theta) \|$ where $\| a \| = (a^{\mathsf{T}}a)^{1/2}$. For more discussions on this minimization problem, refer to Wei et al. (1990b) and Lin et al. (1998).

For simplicity, assume $Q(u,\theta)=1$ temporarily and let $E_i(t,\beta) = X_i(t)-\beta^{\mathsf{T}}W_i$ for $i=1,\ldots,n$. Further, define $N_i^*(u,t) = \Delta_i(t)I\{E_i(t,\beta)\le u\}$ and $Y_i^*(u,t) = I\{E_i(t,\beta)\ge u\}$. Then, under the null hypothesis, $U(t,\theta)$ is partitioned as

$$\begin{aligned} U_\gamma(t,\beta) &= \sum_{i=1}^{n}\int_{-\infty}^{\infty}\{Z_i-\bar{Z}(u,t,\beta)\}dN_i^*(u,t) \\ &= \sum_{i=1}^{n}\int_{-\infty}^{\infty}\{Z_i-\bar{Z}(u,t,\beta)\}dM_i^*(u,t) \end{aligned}$$

and

$$\begin{aligned} U_\beta(t,\beta) &= \sum_{i=1}^{n}\int_{-\infty}^{\infty}\{W_i-\bar{W}(u,t,\beta)\}dN_i^*(u,t) \\ &= \sum_{i=1}^{n}\int_{-\infty}^{\infty}\{W_i-\bar{W}(u,t,\beta)\}dM_i^*(u,t) \end{aligned}$$

where

$$\bar{Z}(u,t,\beta) = \sum_{i=1}^{n} Z_iY_i^*(u,t) \Big/ \sum_{i=1}^{n} Y_i^*(u,t),$$

$$\bar{W}(u,t,\beta) = \sum_{i=1}^{n} W_i Y_i^*(u,t) / \sum_{i=1}^{n} Y_i^*(u,t)$$

and

$$M_i^*(u,t) = N_i^*(u,t) - \int_{-\infty}^{u} Y_i^*(x,t)\lambda_0(x)dx,$$

which is a martingale. Note that the score function $U_\gamma(t,\beta)$ has the same form as the score functions $S_\gamma(t,0,\beta)$ and $U_\gamma(t,0,\beta)$ in Subsubsections 4.3.1 and 4.3.2, respectively, so that we can apply the similar arguments to establish the asymptotic results of $U_\gamma(t,\beta)$. That is, the asymptotic joint distribution of $\{n^{-1/2}U_\gamma(t_k,\beta), k = 1,\ldots,K\}$ is multivariate normal with mean 0 and covariance $(1 \le k \le l \le K)$

$$\mathrm{Cov}_A\{n^{-1/2}U_\gamma(t_k,\beta), n^{-1/2}U_\gamma(t_l,\beta)\} = \mathrm{Var}_A\{n^{-1/2}U_\gamma(t_k,\beta)\} = I(t_k),$$

where

$$I(t) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{Z_i - \bar{Z}(u,t,\beta)\}^2 Y_i^*(u,t)\lambda_0(u)du. \tag{4.22}$$

Since the rank score $U(t,\theta)$ is a step function of θ, we can not apply the usual Taylor expansions to find out a test statistic asymptotically equivalent to $U_\gamma(t,\hat{\beta}_t)$, where the restricted estimator $\hat{\beta}_t$ is the minimizer of $\| U_\beta(t,\beta) \|$. Under the assumption of independence of Z_i and W_i, however, Lin (1992) showed that $U_\gamma(t,\hat{\beta}_t)$ is asymptotically equivalent to $U_\gamma(t,\beta)$. In this case, as shown in Lin (1992), we can simplify $I(t)$ in (4.22). Note that $\bar{Z}(u,t,\beta)$ converges in probability to $\mu_z = E(Z_i)$ and

$$E\left\{\int_{-\infty}^{\infty} Y_i^*(u,t)\lambda_0(u)du\right\} = E\{N_i(\infty,t)\} = \Pr\{\Delta_i(t) = 1\}.$$

Furthermore, Z_i are independent of the other variables so that $I(t) = \sigma_z^2 \Pr\{\Delta_i(t) = 1\}$, where $\sigma_z^2 = E\{(Z_i - \mu_z)^2\}$. Hence, we have that the asymptotic joint distribution of $\{n^{-1/2}U_\gamma(t_k,\hat{\beta}_{t_k}), k = 1,\ldots,K\}$ is multivariate normal with mean 0 and covariance matrix $\{\sigma^2(t_k,t_l); k,l = 1,\ldots,K\}$ where $\sigma^2(t,t') = \sigma^2(t,t) = \sigma_z^2 \Pr\{\Delta_i(t) = 1\}$ for $t \le t'$. Under the null hypothesis, denote $Q(u,\theta)$ and $\hat{\lambda}_0(u,t,\theta)$ by $Q(u,\beta)$ and $\hat{\lambda}_0(u,t,\beta)$. Then, for a given weight function $Q(u,\beta)$, the variance function $\sigma^2(t,t)$ can be consistently estimated by

$$\begin{aligned}
\hat{I}(t) &= n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{\infty} Q^2(u,\hat{\beta}_t)\{Z_i - \bar{Z}(u,t,\hat{\beta}_t)\}^2 Y_i^*(u,t,\hat{\beta}_t)\hat{\lambda}_0(u,t,\hat{\beta}_t)du \\
&= n^{-1} \int_{-\infty}^{\infty} Q^2(u,\hat{\beta}_t) \left\{ \frac{\sum_{i=1}^n Z_j^2 Y_j^*(u,t,\hat{\beta}_t)}{\sum_{i=1}^n Y_j^*(u,t,\hat{\beta}_t)} - \bar{Z}^2(u,t,\hat{\beta}_t) \right\} \sum_{i=1}^{n} dN_i^*(u,t,\hat{\beta}_t)
\end{aligned}$$

where $Y_i^*(u,t,\hat{\beta}_t)$ and $N_i^*(u,t,\hat{\beta}_t)$ are obtained by substituting β with $\hat{\beta}_t$ in $Y_i^*(u,t)$ and $N_i^*(u,t)$, respectively.

5 Examples

5.1 Error spending based on information

As mentioned in Subsection 2.4, standard group sequential methods by Pocock (1977) and O'Brien and Fleming (1979) require equal increments of information at each interim analysis and a pre-specification of the maximum number of analyses. However, these conditions are often not met in practice. The error spending function approach of Lan and DeMets (1983) guarantees an overall type I error probability to a desired significance level without having to fix the number and times of repeated analyses in advance. When designing a study, the number and times of repeated analyses have to be fixed at least tentatively based on the projected duration of enrollment and follow-up and the desired frequency of interim analyses for possible early stopping. This is an issue of particular interest in designing clinical trials with failure time data. A natural approach is to use the notion of statistical information and design the trial as a maximum information trial as in Kim et al. (1995), Lee et al. (1996), and Scharfstein et al. (1997).

At the k^{th} interim analysis, $k = 1, \ldots, K$, denote the standardized score statistics by $S_k = S(\hat{\beta}_k)/\text{SE}\{S(\hat{\beta}_k)\}$ and the standardized Wald test by $W_k = \hat{\gamma}_k/\text{SE}(\hat{\gamma}_k)$ for testing the null hypothesis $\gamma = 0$ in the presence of the nuisance parameters β, where $S(\hat{\beta}_k)$ is the usual score statistics presented in the previous sections with the restricted estimator $\hat{\beta}_k$ of β under the null hypothesis. For the Wald test, $\hat{\gamma}_k$ is obtained from the estimating equations such as the maximum likelihood estimating equations, the least squares estimating equations, the generalized estimating equations and the rank type estimating equations described in the previous sections. Then the information I_k at the k^{th} interim analysis is defined by $I_{k,u} = \text{Var}\{S(\hat{\beta}_k)\}$ for the score test and $I_{k,e} = \{\text{Var}(\hat{\gamma}_k)\}^{-1}$ for Wald test. The information I_k can be estimated by replacing Var with $\widehat{\text{Var}}$. We denote it as $\hat{I}_k$.

For an error spending function $\alpha^\star(t)$ described in Subsection 2.4, we can use the information fraction t_k for the k^{th} interim analysis given by a ratio of the information at the k^{th} interim analysis to the maximum information I_K predetermined by design, i.e. $t_k = I_k/I_K$. At the time of the k^{th} interim analysis, I_k is obtained from the test statistics. The maximum information is defined as

$$I_K = \left(\frac{z_{\alpha/2} + z_\beta}{\gamma_A}\right)^2 \mathfrak{I}_F \tag{5.1}$$

where α and $1-\beta$ are the type I error and the power to be required, respectively; γ_A denotes the treatment effects under the alternative hypothesis; and $\mathfrak{I}_F$ is the so-called inflation factor. The required inflation in statistical information to compensate for the loss of power through multiple testing was discussed by Kim and DeMets (1987). The inflation factor $\mathfrak{I}_F$ is determined as a function of α, β and the number K and timing of repeated testing and depends on the selected error spending function or the group sequential method. Scharfstein et al. (1997) also provide a table of the inflation factors

for methods by Pocock (1977) and O'Brien and Fleming (1979) under various design schemes.

For a given I_K in (5.1), the critical value c_k is calculated by solving the equation

$$\Pr(|Z_1| \le c_1, \ldots, |Z_{k-1}| \le c_{k-1}, |Z_k| > c_k) = \alpha^\star(t_k) - \alpha^\star(t_{k-1}), \tag{5.2}$$

where $(Z_1, \ldots, Z_K)$ is multivariate normal with mean 0 and covariance

$$(I_k/I_l)^{1/2}, \ 1 \le k \le l \le K. \tag{5.3}$$

Reboussin et al. (2000) provided programs for calculating group sequential boundaries using the Lan and DeMets (1983) method. The boundary values $b_{k,u}$ for the score test S_k and $b_{k,e}$ for the Wald test W_k are obtained by replacing t_k in (5.2) with $t_{k,u}$ and $t_{k,e}$, and replacing I_k/I_l in (5.3) with $\hat{I}_{k,u}/\hat{I}_{l,u}$ and $\hat{I}_{k,e}/\hat{I}_{l,e}$, respectively. Here $t_{k,u} = \hat{I}_{k,u}/I_K$ and $t_{k,e} = \hat{I}_{k,e}/I_K$. If $|S_k| > b_{k,u}$ for the score test and $|W_k| > b_{k,e}$ for the Wald test, one stops and rejects the null hypothesis. Note that the covariance (5.3) implies the independent increments structure so that we can use the recursion formula in Subsection 2.3.

5.2 Longitudinal data

To examine the finite sample properties of the "score" test and the Wald test for the semiparametric model for longitudinal data, we use a semiparametric model suggested by the data from the National Cooperative Gallstone Study (NCGS) in Schoenfield et al. (1981). For illustration, we consider only the comparison of cholesterol levels between the placebo (305 patients) group and the high-dose chenodiol (305 patients) group.

The four repeated cholesterol values are modeled as a linear function of the baseline cholesterol value (B_i) and the treatment indicator (T_i) for $i = 1, \ldots, n$ and $j = 1:4$ and the k^{th} interim analysis as

$$E(Y_{ijk}|X_{ik}) = \beta_1 T_i + \beta_{0j} I_{ijk} + \beta_{1j} I_{ijk} B_i.$$

The estimated covariance matrix of the score test statistics and the Wald test statistics over time are, respectively, as follows:

$$\begin{bmatrix} 0.1027 & 0.1044 & 0.1064 & 0.1075 \\ & 0.1490 & 0.1497 & 0.1495 \\ & & 0.1927 & 0.1824 \\ & & & 0.2217 \end{bmatrix}$$

and

$$\begin{bmatrix} 8.8741 & 6.2173 & 5.0184 & 4.3387 \\ & 6.1136 & 4.8675 & 4.1597 \\ & & 4.9627 & 4.0206 \\ & & & 4.1814 \end{bmatrix}.$$

These results confirm empirically the independent increments structure in the sequential test statistics as noted in (2.7) and (2.6) from (2.8), respectively.

5.3 Failure Time data

We describe a simulation study reported in Tsiatis et al. (1995) to illustrate how the group sequential tests for parametric model for failure time data work with moderate sample sizes that are typical in clinical trials. In the simulation, 100 patients were entered uniformly over a 10 year period, and each patient entering the trial in a staggered fashion was randomly allocated with equal probability to one of two treatments indicated by $Z = 0$ or 1. A failure time W_i for patient i was obtained as a function of treatment assignment Z_i and trial entry time E_i by generating an exponentially distributed random variable given by the exponential model with the hazard rate

$$\lambda(u|Z,E,\beta,\theta) = \exp(\theta_1 + \beta Z + \theta_2 E)$$

which is a function of both treatment and entry time. We considered a test of the null hypothesis of no treatment difference, $\beta = 0$, with the nuisance parameters $\theta_1 = 0$ and $\theta_2 = 0.1$.

We analyzed the accumulating data at four times after equal increments in calendar time, i.e. $t = 2.5, 5.0, 7.5$, and 10 years, using all the data available at those times. At each of the four times, we calculated the maximum likelihood estimate $\hat{\beta}(t)$, the score statistic $S_0\{t, \beta = 0, \hat{\theta}(\beta = 0)\}$, and the observed information $\{I^{00}(t)\}^{-1}$. The maximum information was set equal to the average $\{I^{00}(10)\}^{-1}$ obtained from 10,000 repetitions.

To empirically examine the type I error probability, we recorded the proportion of rejections for both the score test and the Wald test, using the Pocock and O'Brien-Fleming type error spending functions at the 0.05 level of significance in another set of 10,000 repetitions. With the group-sequential test based on the Pocock type error spending function $\alpha^{\star}_{\mathrm{P}}(t)$, 598 and 525 of the 10,000 simulations rejected the null hypothesis for the score test and Wald test, respectively. With the group-sequential test based on the O'Brien-Fleming type error spending function $\alpha^{\star}_{\mathrm{OF}}(t) = 2\{1 - \Phi(z_{\alpha/2}/\sqrt{t})\}$, 531 and 519 of the 10,000 simulations rejected the null hypothesis for the score test and the Wald test, respectively. These results seem to suggest that the Wald test produces the type I error probability close to the target significance level as compared to the score test. Also O'Brien-Fleming type group sequential test produces the type I error probability close to the target significance level as compared to the Pocock type group sequential test.

In order to verify the independent increments structure in the sequentially computed test statistics, we computed the empirical correlation matrix of the increments of the score test and the Wald test pre-multiplied by the observed information matrix. They are

$$\begin{bmatrix} 1 & -0.0169 & 0.0011 & 0.0002 \\ & 1 & -0.0059 & 0.0003 \\ & & 1 & -0.0176 \\ & & & 1 \end{bmatrix}$$

for the score test and

$$\begin{bmatrix} 1 & 0.0054 & -0.0028 & -0.0007 \\ & 1 & -0.0032 & -0.0025 \\ & & 1 & -0.0159 \\ & & & 1 \end{bmatrix}$$

for the Wald test. In both the score test and the Wald test, the simulation results appear to confirm the theory, as indicated by the off-diagonal entries being all very close to zero.

As a second example, we consider the Children's Cancer Group study 251 (CCG 251) in which 508 eligible children with untreated acute myeloid leukemia were enrolled between September 1979 and October 1983 in a staggered entry to receive an induction chemotherapy followed by either allogeneic bone marrow transplant or maintenance chemotherapy as reported in Lee and Sather (1995). Post-remission treatment was determined by whether patient had an HLA-matching sibling donor or not without randomization.

A total of 340 children achieved remission and were subsequently allocated to either transplant or chemotherapy. The primary outcome was disease-free survival from the end of induction chemotherapy. As there was apparent cure of disease in a substantial portion of children (30-45%), we analyzed disease-free survival using the mixture model with cure, also known as the cure rate model given by the survival function

$$S(t|Z,\beta,\theta) = \nu_Z + (1-\nu_Z)H_Z(t)$$

where the cure probability ν_Z is parametrized as

$$\nu_Z = \frac{\exp(\alpha+\beta Z)}{1+\exp(\alpha+\beta Z)}$$

and

$$H_Z(t) = \exp\{-\exp(\gamma_o+\gamma_1 Z)t\}^{\delta}.$$

Here β is the parameter of interest and $\theta = (\alpha,\gamma_0,\gamma_1,\delta)'$ is the nuisance parameter.

The study was originally conducted as a fixed sample trial. In order to illustrate the application of group sequential methods using the O'Brien-Fleming type error spending function $\alpha^{\star}_{\mathrm{OF}}(t) = 2\{1-\Phi(z_{\alpha/2}/\sqrt{t})\}$, we applied the score test and the Wald test at yearly intervals starting in October 1982 for three times with 255, 324, and 340 children. Tables ***1*** and ***2*** summarize the results for the score test and the Wald-type test, respectively.

Table 1: *Interim Analyses of CCG 251 with the Score Test.*

k	$\hat{\beta}$	$\widehat{\mathrm{Var}}(\hat{\beta})$	t_k	$\alpha^{\star}_{\mathrm{OF}}(t)$	$\lvert S(t_k)\rvert$	c_k	Reject H_0
1	2.77	4.63	0.307	0.0001	1.29	3.88	No
2	5.97	6.80	0.451	0.0017	2.29	3.14	No
3	9.50	13.38	0.888	0.0348	2.60	2.18	Yes

Table 2: *Interim Analyses of CCG 251 with the Wald Test.*

k	$\hat{\beta}$	$\widehat{\mathrm{Var}}(\hat{\beta})$	t_k	$\alpha^{\star}_{\mathrm{OF}}(t)$	$\lvert W(t_k)\rvert$	c_k	Reject H_0
1	0.515	0.177	0.374	0.0005	1.22	3.48	No
2	0.748	0.125	0.532	0.0043	2.12	2.87	No
3	0.684	0.071	0.931	0.0404	2.56	2.17	Yes

Unlike in Lee and Sather (1995) where the critical values had to be determined using multivariate normal integration as in Schervish (1984), the critical values in Tables ***1*** and ***2*** were determined using univariate normal integration thanks to independent increments. Note that the different test resulted in different estimates of the information fractions at each of the three interim analyses. With both the score test and the Wald test, the trial would have been terminated early in October 1984 with less than full information of 0.888 and 0.931, respectively.

6 Discussion

After the theoretical development in sequential analysis with the seminal work of Wald (1947), ethical imperatives of having to avoid unnecessary experimentation with human subjects in clinical trials motivated early pioneers such as Peter Armitage leading rapid development of sequential methods for clinical trials, e.g. including the first edition of the textbook "Sequential Medical Trials" by Peter Armitage in 1960. Soon there was a recognition, however, that classical sequential methods were not very realistic in most clinical trials and subsequently group sequential methods started to appear in the literature in 1970s.

In order for group sequential methods to be applied correctly, the joint distribution of sequential test statistics computed over time has to be known to determine the group sequential boundaries. In many settings the joint distribution turned out to be a multivariate normal distribution or asymptotically so. This required multivariate normal integration which can be challenging and applicable for up to seven dimensions. However, if the joint distribution has an independent increments structure in its covariance matrix, the multivariate integration reduces to univariate integration involving simple recursion in successive test statistics.

Many authors established the multivariate normality of the joint distributions of sequential test statistics. Many joint distributions turned out to have correlated increments between successive test statistics requiring multivariate normal integration. Examples include tests by Armitage et al. (1985), Geary (1988), Wei et al. (1990a), Lee and DeMets (1992), and Su and Lachin (1992) for longitudinal data and Gehan's test by Slud and Wei (1982) and logrank test under the accelerated failure model by Lin (1992) for failure time data.

Fortunately, joint distributions of many useful test statistics computed over time turn out to have independent increments, thus requiring only univariate integration based on convolution of two independent random variables. Independence of increments in the joint distribution of sequential test statistics was conjectured in Armitage (1975), but the theoretical development started with the initial work in Tsiatis (1981), followed by many noted in Section 3, and culminating with the most general results by Jennison and Turnbull (1997) and Scharfstein et al. (1997).

The limited simulation studies and the real clinical trials data analysis reported here show that the joint distributions of the sequential test statistics investigated have independent increments even for moderate sample sizes. This affirms that standard group sequential methods can be readily applied in interim analysis for possible early stopping of clinical trials in chronic diseases with the very common primary outcome of longitudinal data and failure time data.

Acknowledgements

We thank Professor Pere Puig Casado for inviting us to write this review paper. This is based partly on a presentation given at "A Symposium Honoring Professor Anastasios A. Tsiatis" held on 13 July 2013 in Raleigh, North Carolina, USA, to recognize his contributions in statistical theory and methods on the occasion of his 65th birthday.

Research reported by KMK and AAT was supported in part by grants from the National Institutes of Health at the time of publication.

References

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics*, 6, 701–726.

Anderson, P. K., Borgan, O., Gill, R., and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

Anscombe, F. J. (1954). Fixed-sample-size analysis of sequential observations. *Biometrics*, 10, 89–100.

Armitage, P. (1954). Sequential tests in prophylactic and therapeutic trials. *Quarterly Journal of Medicine*, 23(91), 255–274.

Armitage, P. (1957). Restricted sequential procedures. *Biometrika*, 44, 9–26.

Armitage, P. (1975). *Sequential Medical Trials*, 2nd ed. Wiley, New York.

Armitage, P. (1990). Sequential methods. In Hinkley, D. V., Reid, N., and Snell, E. J. (eds), *Statistical Theory and Modeling: In Honor of Sir David Cox, FRS*. Chapman and Hall, London.

Armitage, P., McPherson, C. K., and Rowe, B. C. (1969). Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society*, Series A, 132, 235–244.

Armitage, P., Stratton, I. M., and Worthington, H. V. (1985). Repeated significance tests for clinical trials with a fixed number of patients and variable follow-up. *Biometrics*, 41, 353–359.

Bross, I. (1952). Sequential medical plans. *Biometrics*, 8, 188–205.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, Series B, 34, 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62, 269–276.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

Crowder, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika*, 82, 407–410.

DeMets, D. L. and Lan, K. K. G. (1994). Interim analysis: the alpha spending function approach. *Statistics in Medicine*, 13, 1341–1352.

Dodge, H. F. and Romig, H. G. (1929). A method of sampling inspection. *Bell System Technical Journal*, 8, 613–631.

Elfring, G. L. and Schultz, J. R. (1973). Group sequential designs for clinical trials. *Biometrics*, 29(3), 471–477.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Gail, M. H., DeMets, D. L., and Slud, E. V. (1982). Simulation studies on increments of the two sample logrank score test for survival time data, with application to group sequential boundaries. In Johnson, R. and Crowley, J. (eds), *Survival Analysis*, IMS Lecture Notes-Monograph Series, 2. Hayward, California: Institute of Mathematical Statistics

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

Geary, D. N. (1988). Sequential testing in clinical trials with repeated measurements. *Biometrika*, 75, 311–318.

Gehan, E. A. (1961). The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *Journal of Chronic Diseases*, 13, 346–353.

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52, 203–223.

Gu, M. G. and Lai, T. L. (1991). Weak convergence of time-sequential censored rank statistics with applications to sequential testing in clinical trials. *Annals of Statistics*, 19, 1403–1433.

Gu, M. and Ying, Z. (1995). Group sequential methods for survival data using partial score processes with covariate adjustment. *Statistica Sinica*, 5, 793–804.

Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *Annals of Mathematical Statistics*, 34, 598–611.

Jennison, C. and Turnbull, B. W. (1990). Statistical approaches to interim monitoring of medical trials: a review and commentary. *Statistical Science*, 5, 299–317.

Jennison, C. and Turnbull, B. W. (1991). Exact calculations for sequential t, χ^2 and F tests. *Biometrika*, 78, 133–141.

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

Kim, K., Boucher, H., and Tsiatis, A. A. (1995). Design and analysis of group sequential logrank tests in maximum duration versus information trials. *Biometrics*, 51, 988–1000.

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

Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics*, 38, 963–974.

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

Lee, J. W. (1994). Group sequential testing in clinical trials with multivariate observations: a review. *Statistics in Medicine*, 13, 101–111.

Lee, J. W. and DeMets, D. L. (1991). Sequential comparison of changes with repeated measurements data. *Journal of the American Statistical Association*, 86, 757–762.

Lee, J. W. and DeMets, D. L. (1992). Sequential rank tests with repeated measurements in clinical trials. *Journal of the American Statistical Association*, 87, 136–142.

Lee, J. W. and Sather, H. N. (1995). Group sequential methods for comparison of cure rates in clinical trials. *Biometrics*, 51, 756–763.

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

Lee, Y. J., Staquet, M., Simon, R., Catane, R., and Muggia, F. (1979). Two-stage plans for patient accrual in phase II cancer clinical trials. *Cancer Treatmetn Reports*, 63, 1721–1726.

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

Liang, K., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analyses for categorial data (with discussion). *Journal of the Royal Statistical Society*, Series B, 54, 3–40.

Lin, D. (1992). Sequential log rank tests adjusting for covariates with the accelerated life model. *Biometrika*, 79, 523–529.

Lin, D. Y., Wei, L. J., and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika*, 85, 605–618.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 50, 163–170.

McCullagh, C. K. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman and Hall, London.

McPherson, C. K. and Armitage, P. (1971). Repeated significance tests on accumulating data when the null hypothesis is not true. *Journal of the Royal Statistical Society*, Series A, 134, 15–26.

Neyman, J. and Pearson, E. S. (1933). IX. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London*, Series A, 231(694–706), 289–337.

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

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

Prentice, R. L. and Marek, P. (1979). A qualitative discrepancy between censored data rank tests. *Biometrics*, 35, 861–867.

Reboussin, D. M., DeMets, D. L., Kim, K., and Lan, K. K. G. (2000). Programs for computing group sequential bounds using the Lan-DeMets method. *Controlled Clinical Trials*, 21, 190–207.

Reboussin, D. M., Lan K. K. G., and DeMets, D. L. (1992). Group sequential testing of longitudinal data. Technical Report 72. Department of Biostatistics, University of Wisconsin-Madison.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Annals of Statistics*, 18, 303–328.

Rotnitzky, A. and Jewell, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, 77, 485–497.

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

Schervish, M. J. (1984). Multivariate normal probabilities with error bound (with corrections in 1985). *Applied Statistics*, 33, 81–94.

Schoenfield, L. J., Lachin, J. M., the Steering Committee, and the NCGS Group (1981). Chenodiol for dissolution of gallstones: The National Cooperative Gallstone Study. *Annals of Internal Medicine*, 95, 257–282.

Sellke, T. and Siegmund, D. (1983). Sequential analysis of the proportional hazards model. *Biometriak*, 70, 315–326.

Simon, R. (1989). Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials*, 10, 1–10.

Slud, E. V. (1984). Sequential linear rank tests for two-sample censored survival data. *Annals of Statistics*, 12(2), 551–571.

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

Sobel, M. and Wald, A. (1949). A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. *Annals of Mathematical Statistics*, 20, 502–522.

Spiessens, B., Lesaffre, E., Verbeke, G., Kim, K., and DeMets, D.L. (2000). An overview of group sequential methods in longitudinal clinical trials. *Statistical Methods in Medical Research*, 9, 497–515.

Spiessens, B., Lesaffre, E., Verbeke, G., and Kim, K. (2002). Group sequential methods for an ordinal logistic random-effects model under misspecification. *Biometrics*, 58, 569–575.

Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Annals of Mathematical Statistics*, 16, 1243–258.

Su, J. Q. and Lachin, J. M. (1992). Group sequential distribution-free methods for the analysis of multivariate observations. *Biometrics*, 48, 1033–1042.

Tarone, R. E. and Ware, J. H. (1977). On distribution-free tests for equality of survival distributions. *Biometrika*, 64, 156–160.

Tsiatis, A. A. (1981). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika*, 68, 311–315.

Tsiatis, A. A. (1982). Repeated significance testing for a general class of statistics used in censored survival analysis. *Journal of the American Statistical Association*. 77, 855–861.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics*, 18, 354–372.

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

Tsiatis, A. A., Rosner, G. L., and Tritchler D. L. (1985). Group sequential tests with censored survival data adjusting for covariates. *Biometrika*, 72, 365–373.

Wald, A. (1947). *Sequential Analysis*. Wiley, New York.

Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 19, 326–339.

Wei, L. J., Su, J. Q., and Lachin, J. M. (1990a). Interim analyses with repeated measurements in sequential clinical trial. *Biometrika*, 77, 359–364.

Wei, L. J., Ying, Z., and Lin, D. Y. (1990b). Linear regression analysis of censored survival data based on rank tests. *Biometrika*, 77, 845–851.

Wu, M. C. and Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*, 45, 939–955.

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