Group Sequential Analysis Incorporating Covariate Information

Christopher Jennison
School of Mathematical Sciences,
University of Bath, Bath BA2 7AY, U. K.
and
Bruce W. Turnbull
School of Operations Research and Industrial Engineering
Cornell University, Ithaca, New York 14853-3801, U. S. A.

Abstract

In this paper we survey existing results concerning the joint distribution of the sequence of estimates of the parameter vector when a model is fitted to accumulating data and we provide a unified theory which explains the "independent increments" structure commonly seen in group sequential test statistics. Our theory covers normal linear models, including the case of correlated observations, and asymptotic results extend to generalized linear models and the proportional hazards regression model for survival data. The asymptotic results are derived using standard methods for the non-sequential case and they hold as long as these non-sequential techniques are applicable at each individual analysis. In all cases, the joint distribution of the sequence of parameter estimates has the same form, exactly or asymptotically, as that of the sequence of means of an increasing number of independent, identically distributed normal variables. Thus, our results provide the formal basis for extending the scope of standard group sequential methods to a wide range of problems.

Key words: Censored Survival data; Clinical trials; Counting processes; General linear model; Generalized linear model; Group sequential testing; Longitudinal data; Martingales; Mixed effects model; Proportional hazards regression model; Repeated measurements; Score statistics.

1 Introduction

This research was motivated in part by discussions concerning the design of a long term clinical trial for treatment and prevention of osteoporosis in post-menopausal women comparing a treatment of daily vitamin and mineral supplementation with placebo. During an accrual phase of two years, women were to be recruited and randomly assigned to treatment or placebo and followed for up to five years via a series of physical and radiological examinations at approximately four month intervals. Primary endpoints included "spinal deformity index", a composite score based on the condition of the vertebrae (bone mass, etc.) which was to be evaluated for each subject at each periodic examination. Baseline covariates measured at randomization included age and other demographic characteristics, as well as various health and nutritional status indicators. For ethical, economic and administrative reasons, the accruing results of the trial were to be monitored at approximate six month intervals by a Data and Safety Monitoring Committee (DSMC). An interim analysis of efficacy would be based on testing for a difference of rate of change over time of spinal deformity index between the treatment arms. An analysis of such longitudinal data is often based on a normal linear model with fixed effects for assigned treatment and baseline covariates and random subject effects; the treatment effect then appears as a difference in the slopes of mean spinal deformity index against time for the two treatment arms. Aware of the elevation of Type I and Type II error rates, due to repeated analyses implied by the nature of sequential monitoring, the DSMC requested that monitoring boundaries for this efficacy measure be set in place to act as a guideline to aid in their decision to stop or continue the trial. The establishment of such boundaries requires the computation of the joint distribution of successive estimates of the slope difference parameter based on this complex statistical model.

An alternative efficacy endpoint was also considered. This was the time from randomization to first major fracture of a vertebra. In an interim analysis of such censored failure time data, a semiparametric proportional hazards regression model would be used. Again the challenge was the construction of group sequential boundaries to monitor the treatment effect, which would be based on the joint distribution of successive estimates of the hazard ratio in this second complex statistical model.

In this and other examples, group sequential methods permit a timely response to accumulating information, meeting the ethical and financial imperatives to conclude a trial as soon as the superiority of one treatment becomes apparent or to "abandon a lost cause" when it is clear that the trial results are not going to support further development of an experimental treatment. Following the influential papers of Pocock (1977) and O'Brien and Fleming (1979), group sequential tests have become standard in the monitoring of medical studies. They have been shown to achieve most of the reductions in expected sample size attainable by fully sequential procedures (see Jennison, 1987 and Eales and Jennison 1992) but have the advantage of a lesser administrative burden.

At successive stages of a group sequential procedure, the data that have accumulated so far are evaluated and, based on whether the current value of a summary statistic falls

in some pre-specified set, either (a) the trial is stopped and a terminal decision is made, or (b) the trial continues to the next stage and a new group of observations is gathered; at the final stage only (a) is allowed. The joint distribution of the sequence of summary statistics is needed in order to compute operating characteristics such as size and power of a test, to compute P-values and to construct confidence intervals on termination for parameters of interest. If a Bayes decision formulation is adopted for such a problem, the joint distribution of the sequence of summary statistics is also needed in order to assess future risks and, hence, find Bayes optimal strategies. In this paper we present a unified theory for the joint distribution of a sequence of parameter estimates, including covariate effects, based on accumulating data.

Group sequential procedures that test for a difference between the means of two normal response distributions are readily adapted to other response distributions as long as sequences of test statistics have the same form of joint distribution, at least approximately. The central limit theorem implies the approximate normality of means of groups of independent, identically distributed observations. There are also specialized results for other data types, for example, group sequential logrank tests have been derived for comparing two survival distributions. It is standard practice in clinical trials to stratify treatment allocation with respect to prognostic factors and to fit terms for the effects of these stratification variables and other baseline covariates in a model for patient response. Thus, the sequential monitoring of such trials requires distribution theory for repeated estimates of a treatment effect in a model including regression terms for other explanatory variables. For survival data subject to right censoring, theory has been developed for parametric models (Tsiatis, Boucher and Kim, 1995) and for Cox's proportional hazards regression model (Tsiatis, Rosner and Tritchler, 1985 and Gu and Ying, 1995). Whitehead (1992, Ch. 7.) describes important examples of the sequential analysis of generalized linear models and proportional hazards regression models using score statistics and observed Fisher information; he presents theory for the marginal distribution of each test statistic but does not specifically discuss the joint distribution of a sequence of score statistics.

Further challenges are posed by longitudinal studies such as the osteoporosis trial described above, in which measurements are made repeatedly on each subject over a period of time. For example, if patient entry is staggered, the number of measurements on each subject at a given time will vary, depending on subjects' arrival times. The special issue of Statistics in Medicine, **7**, Number 1/2, 1988, is devoted to longitudinal data analysis and contains a variety of medical examples. The linear mixed effects model (Laird and Ware, 1982) which models correlations between observations on the same subject through random subject effects is a popular choice for normally distributed longitudinal data. Schall (1991) and Breslow and Clayton (1993) describe "generalized linear mixed models" which extend such models to non-normal responses. Treatment effects appear as slopes or intercepts of "linear predictors" and, for the usual ethical reasons, it is desirable to terminate longitudinal studies as soon as a treatment effect is identified. The joint distribution of parameter estimates over a series of analyses is needed to define tests which meet restrictions on error probabilities.

There has been some progress in the group sequential analysis of correlated normal responses. Armitage, Stratton and Worthington (1985) derive group sequential tests for the repeated analysis of counts, motivated by an application in dentistry and Wu and Lan (1992) develop group sequential tests for the effect of treatment on the area under a response curve in an application to the assessment of an anti-smoking programme on the lung function of a group of smokers. Lee and DeMets (1991) obtain group sequential tests for a linear mixed effects model which they apply to a study of calcium supplement effects on repeated bone density measurements and, in Lee and DeMets (1995), the same authors draw comparisons with tests based on a popular heuristic method.

Gange and DeMets (1996) consider the sequential analysis of correlated responses with non-normal distributions using the generalized estimating equations regression model of Liang and Zeger (1986). They derive the asymptotic joint distribution of a sequence of parameter estimates obtained from accumulating data and illustrate the general theory by an application to ordered polytomous outcomes with a proportional odds model.

The purpose of this paper is to provide a comprehensive survey of existing results on the joint distribution of parameter estimates based on accumulating data, especially in the presence of covariates, and to show how these results lie within a unified theory. This general theory explains the commonly found independent increments structure of group sequential test statistics. It also provides new results, adding theoretical support to previously proposed methodology and laying the foundations for further applications of group sequential methods. Our asymptotic theory is a direct extension of well-known theory for the non-sequential case: this has the appealing consequence that the group sequential results do not require any additional regularity conditions and apply whenever it is appropriate to use the standard, non-sequential arguments at each individual analysis.

Additional problems arise if it is necessary to estimate the variance of normally distributed data. We shall not pursue this issue here but we note that a general group sequential theory for the joint distribution of mean and variance parameters generalizing the results of Jennison and Turnbull (1991a) does apply (see Jennison and Turnbull, 1995). Furthermore, group sequential t tests attaining specified Type I and II errors, at least approximately, are available (see Denne, 1996).

In Section 3, we show that the joint distribution of the sequence of maximum likelihood estimates of a parameter vector in a normal linear model retains the familiar form seen in the case of independent, identically distributed normal responses. Equivalently, the related score statistics have independent increments. Thus, if the variance is known, standard group sequential tests can be applied in this general setting. Our theory for normal linear models also applies in the case of correlated errors and provides a simple proof of a number of previous results for longitudinal data and mixed effects models. The pattern emerging from some of these results has started to attain the status of a "folk theorem" — although departures from the standard pattern have also been found. Our result explains why the common pattern arises and when and why exceptions occur. In Section 4 we obtain asymptotic versions of these results for general parametric models, including the important case of generalized linear models. In Section 5 we consider

group sequential analysis of survival data: we discuss existing results for parametric survival models and present a new proof that the sequence of maximum partial likelihood estimates for the parameters of a proportional hazards regression model (Cox, 1972) has, asymptotically, the standard joint distribution. These results provide versions of the "folk theorem" for non-normal data by establishing, asymptotically, the standard covariance structure for sequences of maximum likelihood or maximum partial likelihood parameter estimates and the independent increment property for related sequences of score statistics or Wald statistics.

2 A Standard Joint Distribution for Successive Maximum Likelihood Estimates of θ

In a group sequential study with a maximum of K groups of observations, let $\hat{\theta}^{(k)}$, $k = 1, \ldots, K$, denote the estimates of the parameter vector $\theta = (\theta_1, \ldots, \theta_p)^T$ at successive analyses. Then it is not uncommon to find, at least approximately, that $(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(K)})$ is multivariate normal with

$$\begin{cases} \hat{\theta}^{(k)} \sim N(\theta, \mathcal{I}^{-1}(k, \theta)), & 1 \leq k \leq K, \quad \text{and} \\ Cov(\hat{\theta}^{(k_1)}, \hat{\theta}^{(k_2)}) = Var(\hat{\theta}^{(k_2)}) = \mathcal{I}^{-1}(k_2, \theta), & 1 \leq k_1 \leq k_2 \leq K \end{cases} \tag{1}$$

for certain matrices $\mathcal{I}^{-1}(k, \theta)$, $k = 1, \ldots, K$. This property forms the basis of many group sequential procedures. Suppose, for example, it is required to test the hypothesis H_0: $c^T\theta = \zeta$ for a given p-vector c and scalar constant ζ. Let $\mathcal{I}_k = \{Var(c^T\hat{\theta}^{(k)})\}^{-1} = (c^T\mathcal{I}^{-1}(k, \theta)c)^{-1}$, the Fisher information for $c^T\theta$ at analysis k, $k = 1, \ldots, K$, and $\mu = c^T\theta - \zeta$. A group sequential test of H_0 can be based on the score statistics $S_k = \mathcal{I}_k^{-1}(c^T\hat{\theta}^{(k)} - \zeta)$, $k = 1, \ldots, K$, which follow a multivariate normal distribution with

$$\begin{cases} S_k \sim N(\mathcal{I}_k\mu, \mathcal{I}_k), & 1 \leq k \leq K, \quad \text{and} \\ Cov(S_{k_1}, S_{k_2}) = Var(S_{k_1}) = \mathcal{I}_{k_1}, & 1 \leq k_1 \leq k_2 \leq K. \end{cases} \tag{2}$$

Alternatively, a test can be defined in terms of the standardized statistics

$$Z_k = \mathcal{I}_k^{1/2}(c^T\hat{\theta}^{(k)} - \zeta), \quad k = 1, \ldots, K,$$

known as Wald statistics, which are are jointly multivariate normal with

$$\begin{cases} Z_k \sim N(\mathcal{I}_k^{1/2}\mu, 1), & 1 \leq k \leq K, \quad \text{and} \\ Cov(Z_{k_1}, Z_{k_2}) = \sqrt{(\mathcal{I}_{k_1}/\mathcal{I}_{k_2})}, & 1 \leq k_1 \leq k_2 \leq K. \end{cases} \tag{3}$$

Standard forms of group sequential test give stopping boundaries for the sequences $S_1, \ldots, S_K$ and $Z_1, \ldots, Z_K$. Generally, pre-determined parametric boundaries can be applied if the increments in information between successive analyses, $\mathcal{I}_1, \mathcal{I}_2 - \mathcal{I}_1, \ldots, \mathcal{I}_K -$

$\mathcal{I}_{K-1}$, are equal, but computation is required to adapt to unequal and unpredictable information increments using, for example, the "error spending" methods introduced by Slud and Wei (1982) and Lan and DeMets (1983). It is evident from (2) that the increments S_1, $S_2 - S_1, \ldots, S_K - S_{K-1}$ are independent, hence both sequences $S_1, \ldots, S_K$ and $Z_1, \ldots, Z_K$ are Markov and this property greatly simplifies numerical computations for group sequential tests. See Armitage, McPherson and Rowe (1969), McPherson and Armitage (1971) and Jennison (1994) for details. Without the Markov structure, a general multivariate normal integration routine such as MULNOR (Schervish, 1984) would be needed but this approach is only practical for tests with up to about five groups of observations.

Estimates and test statistics follow the above distributions, (1), (2) and (3) in some simple cases. They do so when observations are independent $N(\theta, \sigma^2)$ variables. Also, in a two treatment comparison of univariate normal observations with variance σ^2, treatment means μ_A and μ_B and cumulative sample sizes at analysis k equal to n_{Ak} and n_{Bk} for treatments A and B respectively, S_k is $\{(n_{Ak}^{-1} + n_{Bk}^{-1})\sigma^2\}^{-1}$ times the difference between mean responses of subjects on treatments A and B at analysis k and $(S_1, \ldots, S_K)$ is multivariate normal and satisfies (2) with $\mu = \mu_A - \mu_B$ and $\mathcal{I}_k = \{(n_{Ak}^{-1} + n_{Bk}^{-1})\sigma^2\}^{-1}$. More generally, when group sizes are large, the central limit theorem implies these properties hold approximately if the $\hat{\theta}^{(k)}$ are sample means of non-normal data.

Other sequences of estimates and test statistics also follow the same form of joint distribution. Tsiatis (1981 and 1982) and Harrington, Fleming and Green (1982) have proved that, asymptotically, sequences of logrank statistics, suitably rescaled, are multivariate normal and satisfy (2). The small-sample accuracy of these approximations has been assessed by Gail, DeMets and Slud (1982) and by DeMets and Gail (1985) and modifications to improve small sample accuracy have been proposed by Jennison (1992). Tsiatis et al. (1995) obtain the same asymptotic property for score statistics in a parametric survival model in the presence of competing risk and end-of-study censoring and Gu and Ying (1995) obtain similar results for semiparametric survival models. Jennison and Turnbull (1985) show that the asymptotic distribution of sequences of Kaplan-Meier estimates of a survival probability obtained from accumulating right censored data is multivariate normal and satisfies (1). However, not all sequences of test statistics follow this standard pattern. Slud and Wei (1982) derive the asymptotic joint distribution of a sequence of modified-Wilcoxon scores, analogous to our $S_1, \ldots, S_K$, for comparing two sets of censored survival data and show that they can have correlated increments if patient entry is staggered. Lin (1991) notes that sequences of sums of linear rank statistics for analyzing multiple survival endpoints are asymptotically multivariate normal but have independent increments *only* if the limiting weight function is independent of survival time. Gu and Ying (1993, Sec. 4) consider the asymptotic joint distribution of successive Buckley-James score statistics (Buckley and James, 1979) for accumulating survival data with covariates, staggered entry and right censoring and show that this also fails to satisfy (2). Jennison and Turnbull (1991b) establish conditions under which successive Mantel-Haenszel estimates of the odds ratio in stratified 2×2 tables are

asymptotically multivariate normal and satisfy (1) but they note that other covariance structures occur if data do not accumulate in a specific manner. The independent increment structure has also been shown not to hold for certain robust score statistics used in semiparametric multiplicative intensity regression models for recurrent event data in complex situations where there may be frailties or other dependencies present — Cook and Lawless (1996, Sec.5), Jiang (1996).

In the group sequential analysis of longitudinal data, sequences of estimates have been shown to follow the standard distribution (1) in some, but not all, cases. Armitage et al. (1985) prove that the differences in mean response of two groups of subjects measured at three successive occasions satisfy (1) under a model assuming independent within-subject variation. However, these authors and Geary (1988) show that this property is lost when a within-subject autoregression term is added. Halperin, Lan, Wright, and Foulkes (1987) prove that estimates of the difference in the mean slopes of two groups in a random effects model follow (1) even when subject entry is staggered and patterns of timing of observations for each subject are irregular. Lee and DeMets (1991) have proposed a group sequential method for comparing rates of change in different forms of linear mixed effects model and, subsequently, Reboussin, Lan and DeMets (1992) have shown that their estimates follow (1). The standard distribution also arises in the asymptotic theory for generalized estimating equation models derived by Gange and DeMets (1996). Wu and Lan (1992) consider a group sequential procedure to compare areas under expected response change curves and show, in important special cases with non-informative censoring, that the sequence of score statistics has, asymptotically, the standard distribution (2) — see, for example, their equations (3.9) and (3.12). However, further counter-examples to (2) are provided by the linear rank statistics of Lee and DeMets (1992), distribution-free multivariate Hodges-Lehmann estimators (Su and Lachin, 1992), and estimates obtained from "independence estimating equations" (Wei, Su and Lachin, 1990).

The following results will explain why the standard pattern occurs so frequently and also indicate when other behaviour is likely to be found.

3 Joint Distribution of $\{\hat{\theta}^{(k)}; k = 1, \ldots, K\}$ for Normal Linear Models

We suppose univariate observations $Y_1, Y_2, \ldots$ are normally distributed with means depending on $\theta = (\theta_1, \ldots, \theta_p)^T$. In a group sequential study with a maximum of K groups of observations, denote the total number of observations in the first k groups by n_k, where $n_1 < \ldots < n_K$, and denote the vector of observations available at the kth analysis by $Y^{(k)} = (Y_1, \ldots, Y_{n_k})^T$, $k = 1, \ldots, K$. We assume the full vector of n_K observations, $Y^{(K)}$, has a multivariate normal distribution with design matrix $X^{(K)}$ and variance matrix $\Sigma^{(K)}\sigma^2$, where $X^{(K)}$ and $\Sigma^{(K)}$ are known. At each analysis, $k = 1, \ldots, K$, we observe $Y^{(k)} \sim N(X^{(k)}\theta, \Sigma^{(k)}\sigma^2)$, where $X^{(k)}$ and $\Sigma^{(k)}$ can be deduced from $X^{(K)}$ and

$\Sigma^{(K)}$ by extracting the elements relating to the first n_k components of $Y^{(K)}$.

Theorem 1 *Suppose* $Y^{(K)} = (Y_1, \ldots, Y_{n_K})^T \sim N(X^{(K)}\theta, \Sigma^{(K)}\sigma^2)$ *with non-singular variance matrix* $\Sigma^{(K)}\sigma^2$ *and the first* n_k *elements of* $Y^{(K)}$ *are available at analyses* $k = 1, \ldots, K$. *Denote the maximum likelihood, or generalized least squares, estimate of* $\theta = (\theta_1, \ldots, \theta_p)^T$ *at analysis* k *by*

$$\hat{\theta}^{(k)} = (X^{(k)T}{\Sigma^{(k)}}^{-1}X^{(k)})^{-1}X^{(k)T}{\Sigma^{(k)}}^{-1}Y^{(k)}, \quad k = 1, \ldots, K. \tag{4}$$

Then, the vectors $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(K)}$ *have a multivariate normal joint distribution with* $\hat{\theta}^{(k)} \sim N(\theta, V_k\sigma^2)$, $k = 1, \ldots, K$, *where* $V_k = ({X^{(k)}}^T{\Sigma^{(k)}}^{-1}X^{(k)})^{-1}$, *and*

$$Cov(\hat{\theta}^{(k_1)}, \hat{\theta}^{(k_2)}) = Var(\hat{\theta}^{(k_2)}), \quad 1 \le k_1 \le k_2 \le K. \tag{5}$$

Proof.
As each $\hat{\theta}^{(k)}$ is a linear function of $Y^{(K)}$, the elements of the vectors $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(K)}$ have a multivariate normal joint distribution. The marginal distribution of each $\hat{\theta}^{(k)}$, $k = 1, \ldots, K$, is well known and it remains to determine the covariances of the estimates $\hat{\theta}^{(k)}$, $k = 1, \ldots, K$. For $k_1 \le k_2$, the data available at analysis k_1 are a subset of the data available at analysis k_2. We can, therefore, write

$$\hat{\theta}^{(k_1)} = A^T Y^{(k_2)}$$

for some $n_{k_2} \times p$ matrix A. As $\hat{\theta}^{(k_1)}$ is an unbiased estimate of θ, $E(A^T Y^{(k_2)}) = A^T X^{(k_2)}\theta = \theta$ for all θ and we can deduce $A^T X^{(k_2)} = I_p$, where I_p denotes the $p \times p$ identity matrix. Hence,

$$\begin{aligned} Cov(\hat{\theta}^{(k_1)}, \hat{\theta}^{(k_2)}) &= Cov(\, A^T Y^{(k_2)},\, ({X^{(k_2)}}^T{\Sigma^{(k_2)}}^{-1}X^{(k_2)})^{-1}{X^{(k_2)}}^T{\Sigma^{(k_2)}}^{-1}Y^{(k_2)}\,) \\ &= A^T\, Var(Y^{(k_2)})\, {\Sigma^{(k_2)}}^{-1}X^{(k_2)}({X^{(k_2)}}^T{\Sigma^{(k_2)}}^{-1}X^{(k_2)})^{-1} \\ &= ({X^{(k_2)}}^T{\Sigma^{(k_2)}}^{-1}X^{(k_2)})^{-1}\sigma^2 = Var(\hat{\theta}^{(k_2)}), \end{aligned}$$

as required. □

Note that Equation (5) can be rewritten as

$$Cov(\hat{\theta}^{(k_2)}, \hat{\theta}^{(k_2)} - \hat{\theta}^{(k_1)}) = 0 \quad \text{for } 1 \le k_1 \le k_2 \le K, \tag{6}$$

showing that $\hat{\theta}^{(k_2)}$ and $\hat{\theta}^{(k_2)} - \hat{\theta}^{(k_1)}$ are independent. It follows that, for $k_1 \le k_2$,

$$Var(\hat{\theta}^{(k_1)}) = Var(\hat{\theta}^{(k_2)}) + Var(\hat{\theta}^{(k_2)} - \hat{\theta}^{(k_1)}). \tag{7}$$

Applying (7) with $\hat{\theta}^{(k_1)}$ replaced by a general unbiased estimate of θ, linear in $Y^{(k_2)}$ gives the familiar Gauss-Markov result that the generalized least squares estimate $\hat{\theta}^{(k_2)}$ is the

minimum variance linear unbiased estimate based on the data available at stage k_2. It also follows from (5) that, for $k \neq k'$,

$$Cov(V_k^{-1}\hat{\theta}^{(k)} - V_{k-1}^{-1}\hat{\theta}^{(k-1)}, V_{k'}^{-1}\hat{\theta}^{(k')} - V_{k'-1}^{-1}\hat{\theta}^{(k'-1)}) = 0.$$

Hence, the process $\{V_k^{-1}\hat{\theta}^{(k)}; k = 1, \ldots, K\}$ has independent increments and we can deduce that $\{\hat{\theta}^{(k)}; k = 1, \ldots, K\}$ is a Markov sequence.

Theorem 1 covers uncorrelated and correlated normal observations, including normal linear mixed effect models as special cases with their particular variance matrices. The covariance properties of estimates derived by Halperin et al. (1987), Reboussin et al. (1992) and Wu and Lan (1992) follow from this result. The restriction in the Theorem to the sequence of maximum likelihood estimates, or equivalently generalized least squares estimates, is critical and explains why the estimates used by Armitage et al. (1985, Sec. 4) and Geary (1988), which are ordinary least squares rather than generalized least squares estimates, do not follow the standard distribution (2).

4 Joint Distribution of $\{\hat{\theta}^{(k)}; k = 1, \ldots, K\}$ for General Parametric Regression Models

Suppose observations Y_i, $i = 1, \ldots, n_K$, are independent and the vector $Y^{(k)} = (Y_1, \ldots, Y_{n_k})^T$ is observed at the kth analysis, $k = 1, \ldots, K$. Denote the density or discrete distribution of Y_i by $f_i(y_i; \theta)$ where θ is a p-dimensional parameter vector. For example, $f_i(y_i; \theta)$ could specify a generalized linear model with linear predictor $x_{i1}\theta_1 + \ldots + x_{ip}\theta_p$. Define the column vector of efficient scores, $U_i(Y_i; \theta)$, and information matrix, $\mathcal{I}_i(\theta)$, for observation i,

$$U_i(Y_i; \theta) = \frac{\partial}{\partial \theta} log f_i(Y_i; \theta) \quad \text{and} \quad \mathcal{I}_i(\theta) = E(-\frac{\partial}{\partial \theta} U_i^T(Y_i; \theta)), \quad i = 1, \ldots, n_K,$$

and the sum of scores and information matrix at analysis k,

$$U(k, \theta) = \sum_{i=1}^{n_k} U_i(Y_i; \theta) \quad \text{and} \quad \mathcal{I}(k, \theta) = \sum_{i=1}^{n_k} \mathcal{I}_i(\theta), \quad k = 1, \ldots, K.$$

We first review standard, fixed sample asymptotic theory applied to a single analysis, k. In the asymptotic setting sample sizes are indexed by a single variable, n, and $n_k \to \infty$ and $\mathcal{I}(k, \theta) \to \infty$ with $\mathcal{I}(k, \theta)/n$ converging to a fixed limit $\overline{\mathcal{I}}(k, \theta)$. Under suitable regularity conditions on the $f_i(y_i; \theta)$, $i = 1, \ldots, n_k$, (see, for example, Cox and Hinkley, 1974, Ch. 9)

$$E(U_i(Y_i; \theta)) = 0 \quad \text{and} \quad \mathcal{I}_i(\theta) = Var(U_i(Y_i; \theta)), \quad i = 1, \ldots, n_k. \tag{8}$$

Further, the maximum likelihood estimate at analyses k satisfies $U(k, \hat{\theta}^{(k)}) = 0$ with probability approaching 1 as $n \to \infty$ and $\hat{\theta}^{(k)}$ is a consistent estimator of θ. Taking a

Taylor series expansion of each element of $U(k,\theta)$ we obtain,

$$n^{-1/2}U^j(k,\theta) = n^{-1}\sum_{i=1}^{n_k} -\left(\frac{\partial}{\partial\theta}U_i^j(Y_i;\theta_{jk}^*)\right)^T n^{1/2}(\hat{\theta}^{(k)}-\theta),$$

$$j = 1,\ldots,p, \quad \text{with probability approaching 1 as } n\to\infty, \tag{9}$$

where $U^j(k,\theta)$ is the jth element of $U(k,\theta)$, $U_i^j(k,\theta)$ is the jth element of $U_i(k,\theta)$, and θ_{jk}^* is a value of θ on the line segment between θ and $\hat{\theta}^{(k)}$. Again with suitable regularity conditions, the weak law of large numbers implies

$$n^{-1}\sum_{i=1}^{n_k} -\frac{\partial}{\partial\theta}U_i^T(Y_i;\theta) \xrightarrow{P} \overline{\mathcal{I}}(k,\theta) \quad \text{as } n\to\infty, \tag{10}$$

where $\xrightarrow{P}$ denotes convergence in probability, and a multivariate central limit theorem gives

$$n^{-1/2}U(k,\theta) \xrightarrow{\mathcal{D}} N(0,\overline{\mathcal{I}}(k,\theta)), \tag{11}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. An important condition for (10) and (11) to apply for a general response distribution is the natural constraint on the experimental design that the fraction of the total information contributed by any individual observation decreases to zero as the sample size increases. The further requirement that $\overline{\mathcal{I}}(k,\theta)$ be uniformly continuous in θ of in a neighbourhood of the true value of θ ensures that the difference between θ_{jk}^* and θ_j is asymptotically negligible and the standard result,

$$n^{1/2}(\hat{\theta}^{(k)}-\theta) \xrightarrow{\mathcal{D}} N(0,\overline{\mathcal{I}}^{-1}(k,\theta)) \quad \text{as } n\to\infty, \tag{12}$$

follows from (9), (10) and (11).

The following theorem establishes the asymptotic joint distribution of $(\hat{\theta}^{(1)},\ldots,\hat{\theta}^{(K)})$ under the regularity conditions used in the above derivation of the marginal distribution of each $\hat{\theta}^{(k)}$. Thus, for any particular group sequential experiment, it suffices to check that standard, non-sequential asymptotic methods can be applied to deduce the asymptotic distributions of the individual $\hat{\theta}^{(k)}$s, in which case the sequential theory follows automatically. In general, the regularity conditions that must be satisfied depend on the type of data being observed and the way they depend on the parameter vector θ. For the important case of generalized linear models, we refer the reader to McCullagh and Nelder (1989) for summaries of asymptotic theory and further references.

Theorem 2 *Suppose observations Y_i are independent with distributions $f_i(y_i;\theta)$, where θ is p-dimensional, and observations $Y_1,\ldots,Y_{n_k}$ are available at analysis k, $k=1,\ldots,K$. Let $\hat{\theta}^{(k)}$ denote the maximum likelihood estimate of θ based on $Y_1,\ldots,Y_{n_k}$ and $\mathcal{I}(k,\theta)$ the Fisher information for θ summed over the first n_k observations. Suppose $n_k\to\infty$ and $\mathcal{I}(k,\theta)\to\infty$ in such a way that $\mathcal{I}(k,\theta)/n\to\overline{\mathcal{I}}(k,\theta)$ for each $k=1,\ldots,K$. Suppose*

the distributions f_i are sufficiently regular that (8), (9), (10) and (11) hold, each $\overline{\mathcal{I}}(k,\theta)$ is uniformly continuous in θ in a neighbourhood of the true value of θ and each $\hat{\theta}^{(k)}$ is consistent for θ as $n \to \infty$.

Then, the joint distribution of $(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(K)})$, is asymptotically multivariate normal,

$$n^{1/2}(\hat{\theta}^{(k)} - \theta) \xrightarrow{\mathcal{D}} N(0, \overline{\mathcal{I}}^{-1}(k,\theta)) \quad \text{as } n \to \infty, \quad k = 1, \ldots, K,$$

and

$$Cov_A(n^{1/2}\hat{\theta}^{(k_1)}, n^{1/2}\hat{\theta}^{(k_2)}) = Var_A(n^{1/2}\hat{\theta}^{(k_2)}) \quad \text{for } 1 \le k_1 \le k_2 \le K,$$

where Var_A and Cov_A are asymptotic variance and covariance matrices.

Proof

The increments in the score statistic, $n^{-1/2}U(1,\theta)$ and $n^{-1/2}(U(k,\theta) - U(k-1,\theta))$, $k = 2, \ldots, K$, depend on distinct sets of independent Y_is and, hence, are independent. With (11), this implies that the increments have independent asymptotic normal distributions with mean 0 and variances $\overline{\mathcal{I}}(1,\theta)$ and $\overline{\mathcal{I}}(k,\theta) - \overline{\mathcal{I}}(k-1,\theta)$, $k = 2, \ldots, K$, respectively. The sequence $\{n^{-1/2}U(1,\theta), \ldots, n^{-1/2}U(K,\theta)\}$ is, thus, asymptotically multivariate normal with mean zero, $Var(n^{-1/2}U(k,\theta)) = \overline{\mathcal{I}}(k,\theta)$, $k = 1, \ldots, K$, and independent increments. This, together with (9) and (10), the consistency of $\hat{\theta}^{(k)}$ and the uniform continuity of each $\overline{\mathcal{I}}(k,\theta)$, implies that $\{n^{1/2}(\hat{\theta}^{(k)} - \theta); k = 1, \ldots, K\}$ has the same limiting joint distribution as $\{\overline{\mathcal{I}}^{-1}(k,\theta)n^{-1/2}U(k,\theta); k = 1, \ldots, K\}$ and the result follows. □

Since the asymptotic distribution of $(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(K)})$ possesses the standard variance structure (1), group sequential tests can be constructed using boundaries computed for the normal case. For example, a group sequential test of H_0: $c^T\theta = \zeta$ can be based on

$$S_k = \mathcal{I}_k n^{1/2}(c^T\hat{\theta}^{(k)} - \zeta), \quad k = 1, \ldots, K,$$

where $\mathcal{I}_k = (c^T\overline{\mathcal{I}}^{-1}(k,\theta)c)^{-1}$, and the vector $(S_1, \ldots, S_K)$ is asymptotically multivariate normal with independent increments. Furthermore, under a series of contiguous alternatives θ^n for which $n^{1/2}(c^T\theta^n - \zeta) \to \mu$, S_k converges in distribution to $N(\mathcal{I}_k\mu, \mathcal{I}_k)$, $k = 1, \ldots, K$, and the standard distribution (2) applies. Alternatively, the standardized statistics

$$Z_k = \mathcal{I}_k^{1/2} n^{1/2}(c^T\hat{\theta}^{(k)} - \zeta), \quad k = 1, \ldots, K,$$

are asymptotically multivariate normal with $Cov_A(Z_{k_1}, Z_{k_2}) = (\mathcal{I}_{k_1}/\mathcal{I}_{k_2})^{1/2}$ for $k_1 \le k_2$, and $Z_k \sim N(\mathcal{I}_k^{1/2}\mu, 1)$ under a series of contiguous alternatives for which $n^{1/2}(c^T\theta^n - \zeta) \to \mu$. This joint distribution agrees with that of the vector $(Z_1, \ldots, Z_K)$ stated in Section 2 and tests designed for normal observations can be applied without modification.

The asymptotic theory for S_k and Z_k is unaffected if $\overline{\mathcal{I}}(k,\theta)$ is replaced by a consistent estimate and, since $\overline{\mathcal{I}}^{-1}(k,\theta)$ is also the asymptotic variance of $n^{1/2}\hat{\theta}^{(k)}$, it is sufficient to have a consistent estimate of this variance. Finite sample approximations are obtained

by considering a particular study as a member of a sequence indexed by a parameter n which tends to ∞. In a small sample analysis we use $(n\overline{\mathcal{I}}_k)^{-1} \simeq Var(c^T\hat{\theta}^{(k)})$ to obtain standardized statistics

$$Z_k = \{Var(c^T\hat{\theta}^{(k)})\}^{-1/2}(c^T\hat{\theta}^{(k)} - \zeta), \quad k = 1, \ldots, K,$$

which are approximately multivariate normal with $Z_k \sim N(\{Var(c^T\hat{\theta}^{(k)})\}^{-1/2}(c^T\theta - \zeta), 1)$ and $Cov(Z_{k_1}, Z_{k_2}) = \{Var(c^T\hat{\theta}^{(k_2)})/Var(c^T\hat{\theta}^{(k_1)})\}^{1/2}$ for $k_1 \leq k_2$. Alternatively, we can define score statistics corresponding to $n^{1/2}S_k$,

$$\tilde{S}_k = \{Var(c^T\hat{\theta}^{(k)})\}^{-1}(c^T\hat{\theta}^{(k)} - \zeta), \quad k = 1, \ldots, K,$$

which are approximately multivariate normal with $\tilde{S}_k \sim N(\{Var(c^T\hat{\theta}^{(k)})\}^{-1}(c^T\theta - \zeta), \{Var(c^T\hat{\theta}^{(k)})\}^{-1})$ and independent increments. Note that all we need here is the maximum likelihood estimate $\hat{\theta}$ and an estimate of its variance at each analysis, both of which are usually provided by computer programmes for statistical model fitting.

It is always advisable to test the adequacy of an asymptotic approximation to small sample behaviour and we would recommend that simulation studies be conducted before using the above results in a new area of application, especially when a sequential procedure affords opportunity for very early stopping. Corrective action can be taken if there are discrepancies in small sample properties. Pocock (1977) suggested implementing group sequential tests for non-normal data by applying repeated significance tests at the nominal levels computed for the normal, known variance case and, even when exact distribution theory is not available for a particular test statistic, one can often employ a good small sample approximation in each repeated significance test rather than use the limiting normal approximation directly. This approach achieves marginal error rates at each analysis close to their intended values while relying on the normal theory to take care of the joint behaviour of decisions at different analyses. It has been shown to work well for group sequential logrank tests (Jennison and Turnbull, 1989 and Jennison, 1992) and for group sequential t and χ^2 tests (Jennison and Turnbull, 1991a). Thus, the asymptotic theory is not necessarily limited by the accuracy of the normal approximation to the marginal distribution of each estimate $\hat{\theta}^{(k)}$, and its description of the joint distribution of $(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(K)})$ remains valuable when small sample approximations are substituted for individual $\hat{\theta}^{(k)}$ distributions.

Our general theory provides a basis for group sequential testing in a range of important applications, including the whole family of generalized linear models. A particular application of the theory is to the sequential analysis of logistic regression models for the decay in germination rates of seeds, for which Whitehead (1989) proposed a sequential monitoring scheme based on the assumption that the sequence of score statistics is approximately multivariate normal with independent increments. Our results provide an asymptotic justification of this assumption to supplement the evidence of Whitehead's small sample simulations.

The general theory can also be extended to studies with correlated observations by replacing the distribution $f_i(Y_i; \theta)$ by the conditional distribution of Y_i given $Y_1, \ldots, Y_{i-1}$

and defining efficient scores $U_i(Y_i;\theta)$ in terms of this conditional distribution (see Cox and Hinkley, 1974, p. 299). The conclusion of Theorem 2 remains valid if it is permissible to apply a weak law of large numbers to sums of conditional information and a central limit theorem to sums of conditional scores to deduce the conditions required in the statement of Theorem 2.

We conclude this Section by noting a remarkably simple proof that all asymptotically efficient estimators have the standard covariance structure (5). Suppose the asymptotic covariance between $\hat{\theta}^{(k_2)}$ and $\hat{\theta}^{(k_2)} - \hat{\theta}^{(k_1)}$ is not zero, then a new estimate of the form $\hat{\theta}^{(k_2)} + \epsilon(\hat{\theta}^{(k_2)} - \hat{\theta}^{(k_1)})$, for ϵ close to zero and of the opposite sign to $Cov(\hat{\theta}^{(k_2)}, \hat{\theta}^{(k_2)} - \hat{\theta}^{(k_1)})$, will have lower variance than $\hat{\theta}^{(k_2)}$, contradicting the asymptotic efficiency of $\hat{\theta}^{(k_2)}$. For example, Lai and Ying (1991) have proved that a modified version of the Buckley-James estimator is asymptotically efficient for the regression analysis of censored normal data and so the above argument establishes the asymptotic covariance structure of successive values of their estimate. In cases such as this, it remains to prove the asymptotic joint normality of the sequence of estimates. We note, conversely, that all the examples described in Section 2 in which the "standard distribution" does not apply involve tests or estimates which are not founded on maximum likelihood methods.

5 Censored Survival Data

5.1 Parametric Survival Models

The sequential analysis of survival data is complicated by the fact that the information accrued between analyses k and $k+1$ depends on the values of variables observed at analysis k, since subjects who have failed at one analysis can contribute no additional information later on. Tsiatis et al. (1995) obtain results that follow our standard pattern for parametric survival models, defined by a parameter vector θ, in the presence of staggered patient entry, competing risk censoring and additional end-of-study censoring at each interim analysis time. They use a martingale central limit theorem to show that, in the notation of Section 4, $(n^{-1/2}U(1,\theta), \ldots, n^{-1/2}U(K,\theta))$ has an asymptotic multivariate normal distribution with independent increments and $n^{-1/2}U(k,\theta) \sim N(0, A(k,\theta))$ for certain matrices $A(k,\theta)$. The proof of independent increments is achieved by writing the martingale associated with the counting process relating to an individual's failure time as a sum of orthogonal martingales, each involving information accruing between a pair of successive analyses. Tsiatis et al. partition θ as $\theta = (\psi, \lambda^T)^T$, where ψ is one-dimensional, and consider the sequence of score statistics, $U_\psi(k, \psi_0, \hat{\lambda}_0^{(k)})$, $k = 1, \ldots, K$, for testing H_0: $\psi = \psi_0$ group sequentially. Here $U(k,\psi,\lambda)$ denotes $U(k,\theta)$ for $\theta = (\psi, \lambda^T)^T$ and $\hat{\lambda}_0^{(k)}$ denotes the maximum likelihood estimate of λ under the restriction $\psi = \psi_0$. They show that these statistics are asymptotically multivariate normal with independent increments and also derive their limiting joint distribution under a sequence of contiguous alternatives.

With a little work, these results can be extended to show that, asymptotically, the

joint distribution of the sequence of maximum likelihood estimates of the full parameter vector, θ, is multivariate normal with the asymptotic covariance structure established in Theorem 2 for general parametric models and uncensored data. Thus, by considering the special case of no censoring, we have an alternative derivation of the result of our Theorem 2 for certain parametric models. However, Tsiatis et al. (1995) limit themselves to the case of continuous distributions for which derivatives of the hazard function satisfy certain conditions. In the absence of censoring, our more direct proof of Theorem 2 applies to both continuous and discrete data; it provides a simpler framework within which to assess regularity conditions and, as noted in Section 4, can also be extended to accommodate dependent observations.

Tests of H_0: $\psi = \psi_0$ based on the score statistics, $U_\psi(k, \psi_0, \hat{\lambda}_0^{(k)})$, and the Wald statistics derived from the components $\hat{\psi}^{(k)}$ of the unrestricted maximum likelihood estimates $\hat{\theta}^{(k)}$, $k = 1, \ldots, K$, are asymptotically equivalent to the first order. It may be that one or the other has better second order properties and it will then be preferable to base a group sequential test on that statistic. Whitehead (1992, p. 67) notes the use of $U_\psi(k, \psi_0, \hat{\lambda}_0^{(k)})$ rather than $U_\psi(k, \psi_0, \hat{\lambda}^{(k)})$ in sequential testing was first proposed by Bartlett (1946). The diffusion approximation to a sequence of likelihood ratios described in Bartlett's paper is of particular interest as it anticipates the independent increment property of score statistics which is so crucial to the general theory.

5.2 The Proportional Hazards Regression Model

By far the most popular model for survival data is the semiparametric proportional hazards regression model (Cox, 1972) and this is most commonly fitted by maximum partial likelihood (Cox, 1975). Gu and Ying (1995) have found the joint distribution of the sequence of score statistics for testing a null hypothesis in a proportional hazards model with a general form of relative risk and time-varying covariates. These score statistics are related to maximum partial likelihood estimates of the full parameter vector, the asymptotic joint distribution of which we derive in the following theorem. The estimates conform to the standard pattern (1) and this implies the independent increments structure of the score statistics. Tsiatis et al. (1985) have previously considered group sequential analysis of the proportional hazards regression model and derived the joint distribution of a sequence of estimates of one treatment parameter under the assumption of independence between the treatment variable and other covariates. Since it is not uncommon, even in quite sizeable randomized trials, for some covariates to be poorly balanced between treatment groups, our result provides important reassurance that the joint distribution of the sequence of estimates of the treatment hazard ratio, adjusted for estimated covariate effects, can still be approximated by the standard multivariate normal distribution (1). A similar assumption of independence between treatment and covariates is made by Lin (1992) in developing a group sequential theory for the accelerated life model and it appears plausible that this assumption might also be weakened.

We follow the development presented for the fixed sample case by Andersen, Borgan,

Gill and Keiding (1993, Sec. VII) with changes of notation to restrict attention to a single type of failure and to include multiple analyses. Although we only discuss the proportional hazards regression model and covariates which are fixed over time, our methods can also be used to obtain group sequential results for the full class of semiparametric multiplicative hazard models considered by Andersen et al. (1993) with time-varying covariates.

Suppose independent subjects $i = 1, \ldots, n$ are observed, each having a covariate vector $Z_i = (Z_{i1}, \ldots, Z_{ip})$ and potential survival time X_i. Entry of subjects may be staggered and survival is subject to competing risk censoring and end-of-study censoring at interim analyses. Thus, each subject has a non-decreasing sequence of potential censoring times U_{ik}, $k = 1, \ldots, K$. At the kth analysis, we observe $\tilde{X}_{ik} = min(X_i, U_{ik})$ and the indicator of an observed failure, $D_{ik} = I(\tilde{X}_{ik} = X_i)$, where $I(A)$ takes the value 1 if A occurs and 0 otherwise. We assume a proportional hazards regression model in which the survival distribution for subject i is continuous with hazard rate

$$\lambda_i(t) = \lambda_0(t) e^{\theta^T Z_i}, \quad t > 0, \tag{13}$$

where $\lambda_0(t)$ is an unknown baseline hazard function and θ the parameter vector to be estimated. We assume censoring to be non-informative for θ (see Andersen et al., 1993, Sec. III.2.3).

To describe the group sequential results, we first need to describe the standard martingale approach to survival data. Define $Y_{ik}(t) = I(\tilde{X}_{ik} \geq t)$ for $t > 0$, the indicator that individual i is observed at analysis k to be at risk at survival time t, measured from time of entry to the study, and define the counting process $N_{ik}(t) = I(\tilde{X}_{ik} \leq t, D_{ik} = 1)$, $t > 0$. For a given value of n, the processes $N_{ik}(t)$ and $Y_{ik}(t)$ and the vector Z_i are defined on a sample space $(\Omega, \mathcal{F}, \mathcal{P})$ adapted to a filtration $(\mathcal{F}_t)$, the dependence on n of the sample space and filtration being suppressed in this notation. For each $t > 0$ take $\mathcal{F}_t$ to be the σ-field generated by the union over $k \in \{1, \ldots, K\}$ of $\{(N_{ik}(u), Y_{ik}(u), Z_i); 0 < u \leq t, i = 1, \ldots, n\}$ and $\mathcal{F}$ the σ-field generated by the union of $\mathcal{F}_t$ over $t > 0$. The compensated version of $N_{ik}(t)$,

$$M_{ik}(t) = N_{ik}(t) - \int_0^t \lambda_0(u) e^{\theta^T Z_i} Y_{ik}(u)\, du, \quad t > 0,$$

is a local square integrable martingale with respect to $(\mathcal{F}_t)$ with predictable variation process

$$\langle M_{ik} \rangle(t) = \int_0^t \lambda_0(u) e^{\theta^T Z_i} Y_{ik}(u)\, du, \quad t > 0.$$

For $t > 0$ and $k = 1, \ldots, K$ define

$$S_k^{(0)}(\theta, t) = \sum_{i=1}^n Y_{ik}(t) e^{\theta^T Z_i}, \quad S_k^{(1)}(\theta, t) = \sum_{i=1}^n Z_i Y_{ik}(t) e^{\theta^T Z_i} \quad \text{and} \quad S_k^{(2)}(\theta, t) = \sum_{i=1}^n Z_i^{\otimes 2} Y_{ik}(t) e^{\theta^T Z_i},$$

where $A^{\otimes 2}$ denotes AA^T, and also define

$$E_k(\theta, t) = \frac{S_k^{(1)}(\theta, t)}{S_k^{(0)}(\theta, t)} \quad \text{and} \quad V_k(\theta, t) = \frac{S_k^{(2)}(\theta, t)}{S_k^{(0)}(\theta, t)} - E_k(\theta, t)^{\otimes 2}.$$

For fixed t and k, $S_k^{(0)}(\theta,t)$ is a scalar, $S_k^{(1)}(\theta,t)$ and $E_k(\theta,t)$ are p-vectors and $S_k^{(2)}(\theta,t)$ and $V_k(\theta,t)$ are $p\times p$ matrices. The log partial likelihood for θ based on data available at analysis k on each patient's survival up to time $t>0$ from entry to the study is

$$C_t(k,\theta)=\sum_{i=1}^{n}\int_0^t \log\left\{\frac{e^{\theta^T Z_i}}{\sum_{l=1}^n Y_{lk}(u)e^{\theta^T Z_l}}\right\}dN_{ik}(u),\quad k=1,\ldots,K.$$

Denote $\partial/\partial\theta\{C_t(k,\theta)\}$ by $U_t(k,\theta)$, the jth element of which can be written as

$$U_t^j(k,\theta)=\sum_{i=1}^{n}\int_0^t (Z_{ij}-E_{kj}(\theta,u))\,dN_{ik}(u)=\sum_{i=1}^{n}\int_0^t (Z_{ij}-E_{kj}(\theta,u))\,dM_{ik}(u), \qquad (14)$$

where $E_{kj}(\theta,u)$ is the jth element of $E_k(\theta,u)$, and write $\partial^2/\partial\theta^2\{C_t(k,\theta)\}$ as $-\mathcal{I}_t(k,\theta)$, where

$$\mathcal{I}_t(k,\theta)=\sum_{i=1}^{n}\int_0^t V_k(\theta,u)\,dN_{ik}(u).$$

Our distributional results are obtained in an asymptotic setting as $n\to\infty$ and concern estimates of θ based on survival up to time τ, the largest survival time at which a non-zero proportion of subjects can be expected to be at risk. We require $\int_0^\tau \lambda_0(t)\,dt<\infty$ and introduce the following group sequential analogues of conditions VII.2.1 and VII.2.2 of Andersen et al. (1993). Note that our subscript k indexes interim analyses but Andersen et al's subscript h, indexing failure type, is not needed as we consider only one form of failure.

Condition 1. Let θ_0 be the true value of θ. Then there is a neighbourhood $\mathcal{H}$ of θ_0 and, for each $k=1,\ldots,K$, there are scalar, p-vector and $p\times p$ matrix functions $s_k^{(0)}$, $s_k^{(1)}$ and $s_k^{(2)}$, respectively, defined on $\mathcal{H}\times[0,\tau]$ such that

$$sup\{\|n^{-1}S_k^{(m)}(\theta,t)-s_k^{(m)}(\theta,t)\|;\theta\in\mathcal{H},t\in[0,\tau]\}\xrightarrow{P}0\quad\text{as } n\to\infty,\quad m=0,1,2,$$

and each $s_k^{(m)}(\theta,t)$ is a continuous function of $\theta\in\mathcal{H}$ uniformly in $t\in[0,\tau]$ and bounded on $\mathcal{H}\times[0,\tau]$.

Further, for each $k=1,\ldots,K$, $s_k^{(0)}(\theta_0,t)$ is bounded away from zero on $t\in[0,\tau]$, $s_k^{(1)}(\theta,t)=\partial/\partial\theta\{s_k^{(0)}(\theta,t)\}$ and $s_k^{(2)}(\theta,t)=\partial^2/\partial\theta^2\{s_k^{(0)}(\theta,t)\}$ for $\theta\in\mathcal{H}$ and $t\in[0,\tau]$, and $\Sigma_{k\tau}=\int_0^\tau v_k(\theta_0,t)s_k^{(0)}(\theta_0,t)\lambda_0(t)\,dt$ is positive definite, where $v_k=s_k^{(2)}/s_k^{(0)}-(s_k^{(1)}/s_k^{(0)})^{\otimes 2}$.

Condition 2. There exists a $\delta>0$ such that

$$n^{-1/2}sup\{|Z_i|I(\theta_0^T Z_i>-\delta|Z_i|);i=1,\ldots,n\}\xrightarrow{P}0\quad\text{as } n\to\infty.$$

Andersen et al. (1993) remark that Condition 2 is satisfied if the distribution of covariates is bounded by that of a random variable with finite rth moment for some $r>2$. We are now ready to prove a group sequential version of their Theorem VII.2.2.

Theorem 3 *Suppose $\hat{\theta}^{(k)}$, $k = 1, \ldots, K$, are maximum partial likelihood parameter estimates for the proportional hazards model (13) based on accumulating data. Suppose also that $\int_0^\tau \lambda_0(t)\, dt < \infty$ and Conditions 1 and 2 above are satisfied. Then, for each $k = 1, \ldots, K$, the probability that the equation $U_\tau(k, \theta) = 0$ has a unique solution, $\hat{\theta}^{(k)}$, tends to 1 and $\hat{\theta}^{(k)} \xrightarrow{P} \theta_0$ as $n \to \infty$.*

The joint distribution of $(\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(K)})$, is asymptotically multivariate normal,

$$n^{1/2}(\hat{\theta}^{(k)} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \Sigma_{k\tau}^{-1}),$$

and

$$Cov_A(n^{1/2}\hat{\theta}^{(k_1)}, n^{1/2}\hat{\theta}^{(k_2)}) = Var_A(n^{1/2}\hat{\theta}^{(k_2)}) \quad \text{for } 1 \le k_1 \le k_2 \le K,$$

where Var_A and Cov_A are asymptotic variance and covariance matrices. Also,

$$n^{-1}\mathcal{I}_\tau(k, \hat{\theta}^{(k)}) \xrightarrow{P} \Sigma_{k\tau} \quad \text{as } n \to \infty.$$

Proof.
The consistency of each $\hat{\theta}^{(k)}$, $k = 1, \ldots, K$, follows from Theorem VII.2.1 of Andersen et al. (1993) and we obtain distributional results by adapting the proof of their Theorem VII.2.2.

We use the same device as Tsiatis et al. (1995) and subdivide the counting processes $N_{ik}(t)$, $i = 1, \ldots, n$, $k = 1, \ldots, K$, into elements associated with observation of subject i between successive pairs of analyses. Let $DN_{i1}(t) = N_{i1}(t)$ and $DN_{ik}(t) = N_{ik}(t) - N_{i(k-1)}(t)$, $k = 2, \ldots, K$, so $DN_{ik}(t)$ is a counting process with intensity $DY_{ik}(t)\lambda_0(t)e^{\theta^T Z_i}$, where $DY_{i1}(t) = Y_{i1}(t)$ and $DY_{ik}(t) = Y_{ik}(t) - Y_{i(k-1)}(t)$, $k = 2, \ldots, K$. Let $DM_{ik}(t)$ be the compensated version of $DN_{ik}(t)$,

$$DM_{ik}(t) = DN_{ik}(t) - \int_0^t \lambda_0(u)e^{\theta^T Z_i} DY_{ik}(u)\, du, \quad t > 0, \;\; k = 1, \ldots, K,$$

then DM_{ik} is a local square integrable martingale with respect to $(\mathcal{F}_t)$. As subjects are independent, $DM_{i_1k_1}(t)$ and $DM_{i_2k_2}(t)$ are orthogonal if $i_1 \neq i_2$ and, since it is not possible for a jump to occur in both $DN_{ik_1}(t)$ and $DN_{ik_2}(t)$, the martingales DM_{ik_1} and DM_{ik_2} are orthogonal if $k_1 \neq k_2$. This orthogonality provides the key to establishing that the sequence $n^{-1/2}U_\tau(1, \theta_0), \ldots, n^{-1/2}U_\tau(K, \theta_0)$ has asymptotically independent increments.

From (14), we can now write

$$n^{-1/2}U_t^j(k, \theta_0) = \sum_{l=1}^{k}\sum_{i=1}^{n} \int_0^t H_{ij}^{(n)}(k, u)\, dDM_{il}(u) \quad j = 1, \ldots, p, \quad k = 1, \ldots, K, \qquad (15)$$

where $H_{ij}^{(n)}(k, t) = n^{-1/2}(Z_{ij} - E_{kj}(\theta_0, t))$. The right hand side of (15) is a linear combination of stochastic integrals of predictable and locally bounded processes with respect to the local square integrable martingales DM_{ik}, hence, the $n^{-1/2}U_t^j(k, \theta_0)$,

considered as processes in t, are also local square integrable martingales. For $k_1 \leq k_2$, the predictable covariance process between $n^{-1/2}U_t^{j_1}(k_1,\theta_0)$ and $n^{-1/2}U_t^{j_2}(k_2,\theta_0)$ is

$$\langle n^{-1/2}U_t^{j_1}(k_1,\theta_0), n^{-1/2}U_t^{j_2}(k_2,\theta_0)\rangle(t) = \sum_{l=1}^{k_1}\sum_{i=1}^{n}\int_0^t H_{ij_1}^{(n)}(k_1,u)H_{ij_2}^{(n)}(k_2,u)\lambda_0(u)e^{\theta^T Z_i}DY_{il}(u)\,du,$$

$$= \sum_{i=1}^{n}\int_0^t H_{ij_1}^{(n)}(k_1,u)H_{ij_2}^{(n)}(k_2,u)\lambda_0(u)e^{\theta^T Z_i}Y_{ik_1}(u)\,du$$

$$= \sum_{i=1}^{n}\int_0^t H_{ij_1}^{(n)}(k_1,u)H_{ij_2}^{(n)}(k_1,u)\lambda_0(u)e^{\theta^T Z_i}Y_{ik_1}(u)\,du, \quad t>0,$$

since $H_{ij_2}^{(n)}(k_2,u) - H_{ij_2}^{(n)}(k_1,u)$ is constant over i and $\Sigma_{i=1}^n H_{ij_1}^{(n)}(k_1,u)e^{\theta^T Z_i}Y_{ik_1}(u) = 0$ for all u. We can now rejoin Andersen et al's (1993) proof of their Theorem VII.2.2 and obtain

$$\langle n^{-1/2}U_t^{j_1}(k_1,\theta_0), n^{-1/2}U_t^{j_2}(k_2,\theta_0)\rangle(t) \xrightarrow{P}$$

$$\int_0^t v_{k_1 j_1 j_2}(\theta_0,u)s_{k_1}^{(0)}(\theta_0,u)\lambda_0(u)\,du, \quad j_1,j_2 = 1,\ldots,p, \quad k_1,k_2 = 1,\ldots,K,$$

where $v_{kj_1j_2}(\theta,u)$ denotes the element in row j_1 and column j_2 of $v_k(\theta,u)$. Hence, we deduce the asymptotic multivariate normal distribution of $U_\tau(1,\theta_0),\ldots,U_\tau(K,\theta_0)$ by Rebolledo's martingale central limit theorem, and the asymptotic multivariate normal distribution of $\hat\theta^{(1)},\ldots,\hat\theta^{(K)}$ from the usual Taylor series expansion of the form (9). □

As in Theorem 2, the transition between fixed sample and group sequential cases is relatively straightforward since we are able to appeal to Andersen et al. (1993) for the proofs of various technical points. The regularity conditions for the group sequential results are exactly those needed to derive the asymptotic distribution of each estimate $\hat\theta^{(k)}$, $k = 1,\ldots,K$, individually. Thus the discussion by Andersen et al. (1993, Ch. VII) of situations in which their regularity conditions apply and when the condition $\int_0^\tau \lambda_0(t)\,dt < \infty$ can be relaxed to allow use of $\tau = \infty$ applies also to the group sequential case.

Construction of group sequential tests is straightforward and one can follow the procedure exactly as described in Section 4, using the asymptotic results to approximate the small sample distribution of the sequence $\hat\theta^{(1)},\ldots,\hat\theta^{(K)}$. In doing this, we approximate $Var(\theta^{(k)})$ by the observable quantity $\{\mathcal{I}_\tau(k,\hat\theta^{(k)})\}^{-1}$ since $n^{-1}\mathcal{I}_\tau(k,\hat\theta^{(k)})$ is a consistent estimate for $\Sigma_{k\tau}$. Again, all that is needed is the maximum likelihood estimate $\hat\theta$ and an estimate of its variance at each analysis, both of which are usually provided by computer programmes for statistical model fitting.

The results of Sellke and Siegmund (1983) and Slud (1984) are related to our Theorem 3 but concern the process obtained by following an estimate of θ continuously in calendar time as data accumulate. These results apply only to a scalar parameter, although Sellke and Siegmund (1983) do discuss the case of a vector θ and it is possible that their methods could produce results for a vector θ if applied in our asymptotic setting, rather than in their somewhat different asymptotic framework.

6 Discussion

We have presented a general distribution theory for sequences of maximum likelihood estimates of a parameter vector. Examples include normal observations, both uncorrelated and correlated, general parametric regression models, and censored survival data. This theory forms the basis of a unified approach to group sequential analysis and supports the use of existing methods in a wide range of applications.

In fact, this general theory can be extended further. Scharfstein, Tsiatis and Robins (1997) show in great generality that efficient score statistics in parametric and semiparametric models have the independent increments structure (2); their theory extends to all semiparametric efficient test statistics and includes the result that Wald statistics based on semiparametric efficient estimators have the asymptotic distribution (3). It is informative to investigate how this very general theory can be used to produce the results we have obtained in parametric and semiparametric cases. The asymptotic normality and efficiency of score statistics at individual analyses are required as conditions for Scharfstein et al's (1997) key Theorem 1 and, in general, these properties must be verified on a case by case basis to confirm that the Theorem can be applied. For parametric models, standard techniques establish the asymptotic normality and efficiency of maximum likelihood estimates and score statistics, subject to the usual regularity conditions; in our Theorem 2 we pursued these standard maximum likelihood methods a little further to give a simple, direct proof of the asymptotic joint distribution of a sequence of parameter estimates and, hence, of score statistics. The general theory for semiparametric models applies to the proportional hazards regression model, where the baseline hazard function is an infinite-dimensional nuisance parameter. Again, some work is needed to establish asymptotic normality and efficiency of the partial likelihood score statistics before Scharfstein et al. (1997) can apply their Theorem 1 and these authors cite the results of Andersen et al. (1993), obtained using martingale tools; we continued to work with these martingale methods and needed only a little additional effort to prove our Theorem 3.

Our remark in the last paragraph of Section 4 can be recast in terms of score statistics to give a heuristic explanation of Scharfstein et al's (1997) Theorem 1. Suppose statistics $S_1, \ldots, S_K$ are to be used for testing H_0: $c^T\theta = \zeta$ against contiguous alternatives θ^n for which $n^{1/2}(c^T\theta^n - \zeta) \to \mu$, as described in Section 4, and asymptotically $S_k \sim N(\mu\mathcal{I}_k, \mathcal{I}_k)$, $k = 1, \ldots, K$. To be efficient, each S_k must maximize the value of $\mathcal{I}_k$ in this distribution over all other statistics based on the data available at analysis k. Assume for the moment that $S_1, \ldots, S_K$ are efficient but do not have asymptotically independent increments, in particular, $Cov(S_{k_2} - S_{k_1}, S_{k_1}) = \delta \neq 0$. Some algebra shows that the mean and variance of

$$\left(\frac{\mathcal{I}_{k_2} + \epsilon\,\mathcal{I}_{k_1}}{\mathcal{I}_{k_2} + 2\epsilon\,\mathcal{I}_{k_1} + 2\epsilon\,\delta + \epsilon^2\mathcal{I}_{k_1}}\right)(S_{k_2} + \epsilon S_{k_1})$$

are

$$\mu \mathcal{I}_{k_2} \left(\frac{\mathcal{I}_{k_2} + 2\epsilon \mathcal{I}_{k_1} + \epsilon^2 \mathcal{I}_{k_1}^2 / \mathcal{I}_{k_2}}{\mathcal{I}_{k_2} + 2\epsilon \mathcal{I}_{k_1} + 2\epsilon\, \delta + \epsilon^2 \mathcal{I}_{k_1}} \right) \quad \text{and} \quad \mathcal{I}_{k_2} \left(\frac{\mathcal{I}_{k_2} + 2\epsilon \mathcal{I}_{k_1} + \epsilon^2 \mathcal{I}_{k_1}^2 / \mathcal{I}_{k_2}}{\mathcal{I}_{k_2} + 2\epsilon \mathcal{I}_{k_1} + 2\epsilon\, \delta + \epsilon^2 \mathcal{I}_{k_1}} \right)$$

respectively, and taking ϵ small with opposite sign to δ gives a more efficient statistic than S_{k_2}. This contradicts the assumption of asymptotic efficiency and it follows that efficient score statistics must have asymptotically independent increments. It remains to show that asymptotically the joint distribution of $(S_1, \ldots, S_K)$ is multivariate normal; however, in Scharfstein et al's (1997) Theorem 1, this property is easily obtained as a consequence of the assumed "regularity and asymptotic linearity" of $S_1, \ldots, S_K$.

Generalizing in another direction to normal linear models with unknown variance, Jennison and Turnbull (1991a, 1995) have derived joint distribution theory for the sequence of pairs of estimates for the parameter vector determining observation means and the variance factor σ^2 that arises in a group sequential study. There is still scope for further research: in a longitudinal study it will often be necessary to estimate the correlation structure of responses from a given subject during the study but group sequential methods are currently available only for known correlations; new theory for the joint distribution of a sequence of pairs of statistics, for example the hazard ratio between two survival curves and an odds ratio for a separate binary outcome, is also needed to apply the group sequential tests for a general bivariate response, proposed for the case of normal data by Jennison and Turnbull (1993a).

Various difficulties must be surmounted in the practical implementation of group sequential tests. Scharfstein et al. (1997) describe a strategy for modifying the length of accrual or the duration of follow-up as observed information levels become known during the course of a clinical trial, in order to achieve target values of Type I error and power. Their example of monitoring CD4+ cell counts in an AIDS clinical trial provides an excellent illustration of the use of "error spending" methods and innovative design strategies to cope with unequal and unpredictable increments in information between analyses.

One problem that usually disappears in asymptotic theory when a sequence of contiguous alternatives $\theta^n \rightarrow \theta_0$ is considered is the dependence of an observation's variance on its mean. This can be a troublesome factor in small samples as the information for θ and, hence, the variances of $\hat{\theta}^{(k)}$, $k = 1, \ldots, K$, in Equation (1) depend on the true value of θ. An adaptive approach to this problem which works well for binary responses is described by Jennison and Turnbull (1993b and c).

We note finally that fitting regression models to accumulating data raises issues of "model uncertainty" if the covariates to be included at each interim analysis are selected in a data-dependent manner. It is difficult enough to calibrate tests following such data-driven model choice in the fixed sample case and the sequential analogue of this problem has many more degrees of freedom.

ACKNOWLEDGEMENT

This research was supported in part by Grant PHS CA 66218 from the U. S. National Institutes of Health.

REFERENCES

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

Armitage, P., McPherson, C.K. and Rowe, B.C. (1969) Repeated Significance Tests on Accumulating Data," *J. R. Statist.Soc.* 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.

Bartlett, M. S. (1946), "The Large-Sample Theory of Sequential Tests," *Proc. Camb. Phil. Soc.*, **42**, 239–244.

Breslow, N. E. and Clayton. D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *J. Am. Statist. Assoc.*, **88**, 9–25.

Buckley, J. and James, I. (1979), "Linear Regression with Censored Data," *Biometrika*, **66**, 429–436.

Cook, R.J. and Lawless, J.F. (1996), "Interim Monitoring of Longitudinal Comparative Studies with Recurrent Event Responses," *Biometrics* **52**, 1311-1323.

Cox, D. R. (1972), "Regression Models and Life Tables (with Discussion)," *J. R. Statist. Soc.* B, **34**, 187–220.

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

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

DeMets, D. L. and Gail, M. H. (1985), "Use of Logrank Tests and Group Sequential Methods at Fixed Calendar Times," *Biometrics*, **41**, 1039–1044.

Denne, J. S. (1996), *Sequential Procedures for Sample Size Estimation.* PhD Thesis, University of Bath, U.K.

Eales, J. D. and Jennison, C. (1992), "An Improved Method for Deriving Optimal One-sided Group Sequential Tests," *Biometrika*, **79**, 13–24.

Gail, M. H., DeMets, D. L. and Slud, E. V. (1982), "Simulation Studies on Increments of the Two-sample Logrank Score Test for Survival Data, with Application to Group Sequential Boundaries," In *Survival Analysis*, Monograph Series 2, Ed. J. Crowley and R. Johnson, Pp. 287–301. Hayward, California: IMS Lecture Notes.

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.

Gu, M. and Ying, Z. (1993), "Sequential Analysis for Censored Regression Data," *J. Am. Statist. Assoc.*, **88**, 890–898.

Gu, M. and Ying, Z. (1995), "Group Sequential Methods for Survival Data Using Partial Likelihood Score Processes with Covariate Adjustment," *Statistica Sinica*, **5**, 793–804.

Halperin, M. Lan, K. K. G., Wright, E. C., and Foulkes, M. A. (1987), "Stochastic Curtailing for Comparison of Slopes in Longitudinal Studies," *Controlled Clinical Trials*, **8**, 315–326.

Harrington, D. P., Fleming, T. R. and Green, S. J. (1982), "Procedures for Serial Testing in Censored Survival Data," In *Survival Analysis*, Monograph Series 2, Ed. J. Crowley and R. Johnson, Pp. 269–286. Hayward, California: IMS Lecture Notes.

Jennison, C. (1987), "Efficient Group Sequential Tests with Unpredictable Group Sizes," *Biometrika*, **74**, 155–165.

Jennison, C. (1992), "Bootstrap Tests and Confidence Intervals for a Hazard Ratio when the Number of Observed Failures is Small, with Applications to Group Sequential Survival Studies," In *Computing Science and Statistics, Volume 22*, (Eds C. Page and R. LePage), Pp. 89–97. New York: Springer-Verlag.

Jennison, C. (1994), "Numerical Computations for Group Sequential Tests," In *Computing Science and Statistics, Volume 25*, (Eds M. Tarter and M. D. Lock), Pp. 263–272. Interface Foundation of North America.

Jennison, C. and Turnbull, B. W. (1985), "Repeated Confidence Intervals for the Median Survival Time," *Biometrika*, **72**, 619–625.

Jennison, C. and Turnbull, B. W. (1989), " Interim Analyses: The Repeated Confidence Interval Approach (with Discussion)," *J. R. Statist. Soc. B*, **51**, 305–361.

Jennison, C. and Turnbull, B. W. (1991a), "Exact Calculations for Sequential t, χ^2 and F Tests," *Biometrika*, **78**, 133–141.

Jennison, C. and Turnbull, B. W. (1991b), "A Note on the Asymptotic Joint Distribution of Successive Mantel-Haenszel Estimates of the Odds Ratio Based On Accumulating Data," *Sequential Analysis*, **10**, 201–209.

Jennison, C. and Turnbull, B. W. (1993a), "Group Sequential Tests for Bivariate Response: Interim Analyses of Clinical Trials with Both Efficacy and Safety Endpoints," *Biometrics*, **49**, 741-752.

Jennison, C. and Turnbull, B. W. (1993b), "Sequential Equivalence Testing and Repeated Confidence Intervals, with Applications to Normal and Binary Responses," *Biometrics*, **49**, 31–43.

Jennison, C. and Turnbull, B. W. (1993c), "One-sided Sequential Tests to Establish Equivalence Between Treatments with Special Reference to Normal and Binary Responses," In *Multiple Comparisons in Biostatistics: Current Research In The Topics of C. W. Dunnett*, Ed. F. M. Hoppe, Pp. 315–330. New York: Dekker.

Jennison, C. and Turnbull, B. W. (1995), "Distribution Theory for Group Sequential Analysis of General Linear Models," *Bath University Statistics Report 95:01*

Jiang, W. (1996), "Interim Analyses with Non-standard Increment Structures: Applications to Recurrent Events with Frailty," Submitted for Publication.

Lai, T. L. and Ying Z. (1991), "Large Sample Theory for a Modified Buckley-James Estimator for Regression Analysis with Censored Data," *Annals of Statistics*, **19**, 1370–1402.

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. and DeMets, D. L. (1991), "Sequential Comparison of Changes with Repeated Measurements Data," *J. Am. Statist. Assoc.*, **86**, 757–762.

Lee, J. W. and DeMets, D. L. (1992), "Sequential Rank Tests with Repeated Measurements in Clinical Trials," *J. Am. Statist. Assoc.*, **87**, 136–142.

Lee, J. W. and DeMets, D. L. (1995), "Group Sequential Comparison of Changes: Ad-hoc Versus More Exact Method," *Biometrics*, **51**, 21–30.

Liang, K.-Y, and Zeger, S. L. (1986), "Longitudinal Data Analysis using Generalized Linear Models," *Biometrika*, **73**, 13–22.

Lin, D. Y. (1991), "Nonparametric Sequential Testing in Clinical Trials with Incomplete Multivariate Observations," *Biometrika*, **78**, 123–131.

Lin, D. Y. (1992), "Sequential Log Rank Tests Adjusted for Covariates with the Accelerated Life Model," *Biometrika*, **79**, 523–529.

McCullagh, P. and Nelder, J. A. (1989), "*Generalized Linear Models*, 2nd Edition. London: Chapman and Hall.

McPherson, C. K. and Armitage, P. (1971), "Repeated Significance Tests On Accumulating Data when the Null Hypothesis is Not True," *J. R. Statist. Soc.* A, **134**, 15–25.

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.

Reboussin, D. M., Lan, K. K. G. and DeMets, D. L. (1992), "Group Sequential Testing of Longitudinal Data," Univ. of Wisconsin, Dept. of Biostatistics, Tech. Report 72.

Schall, R. (1991), "Estimation in Generalized Linear Models with Random Effects," *Biometrika*, **78**, 719–727.

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," *J. Am. Statist. Assoc.*, **92**,

Schervish, M. J. (1984), "Algorithm AS 195. Multivariate Normal Probabilities with Error Bound," *Appl. Statist.*, **33**, 81–94.

Sellke, T. and Siegmund, D. (1983), "Sequential Analysis of the Proportional Hazards Model," *Biometrika*, **70**, 315–326.

Slud, E. V. (1984), "Sequential Linear Rank Tests for Two-sample Censored Survival Data," *Ann. Statist.*, **12**, 551–571.

Slud, E. V. and Wei, L. J. (1982), "Two-sample Repeated Significance Tests Based on the Modified Wilcoxon Statistic," *J. Am. Statist. Assoc.*, **77**, 862–868.

Su, J. Q. and Lachin, J. M. (1992), "Group Sequential Distribution-free Methods for the Analysis of Multivariate Observations," *Biometrics*, **48**, 1033–1042.

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), "Group Sequential Methods for Survival Analysis with Staggered Entry," In *Survival Analysis*, Monograph Series 2, Ed. J. Crowley and R. Johnson, Pp. 257–268. Hayward, California: IMS Lecture Notes.

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.

Wei, L. J., Su, J. Q., and Lachin, M. J. (1990), "Interim Analyses with Repeated Measurements in a Sequential Clinical Trial," *Biometrika*, **77**, 359–364.

Whitehead, J. (1989), "Sequential Methods for Monitoring Declining Quality, with Applications to Long-term Storage of Seeds," *Biometrics*, **45**, 13–22.

Whitehead, J. (1992) *the Design and Analysis of Sequential Clinical Trials*, 2nd Edition. Chichester: Ellis Horwood.

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.