STATISTICS IN MEDICINE, VOL. 12, 753–765 (1993)

SEQUENTIAL MONITORING OF CLINICAL TRIALS: THE ROLE OF INFORMATION AND BROWNIAN MOTION

K. K. GORDON LAN

Department of Statistics/Computer and Information Systems, The George Washington University, Washington, D.C. 20052, U.S.A.

AND

DAVID M. ZUCKER*

Biostatistics Research Branch, National Heart, Lung, and Blood Institute, Bethesda, Maryland 20892, U.S.A.

SUMMARY

Sequential monitoring has been a topic of major interest in clincal trials methodology over the past two decades. This paper presents a unified conceptual framework for sequential monitoring that covers a wide variety of monitoring procedures in a wide variety of clinical trial settings. The central elements of this framework consist of a suitable concept of statistical information and a scheme for using this concept as a basis for summarizing the accumulating results of a trial in a standardized form, through a stochastic process that can be shown to approximate classical Brownian motion. The ideas are developed in a simple step-by-step fashion and illustrated by several practical examples.

1. INTRODUCTION

In a clinical trial comparing a new treatment with placebo or a standard regimen, ethical considerations usually require periodic examination of the data accruing from the trial for evidence of treatment benefit or harm. It is often appropriate for the trial protocol to include a formal sequential monitoring plan, whereby, upon extreme evidence of treatment benefit, the trial may be terminated early and the investigators may claim that a statistically signficant treatment benefit has been demonstrated.

The mathematical theory of sequential analysis was introduced in the 1940s, motivated by industrial applications, and has continued to develop actively.[1,2] Over the past 20 years, there has been extensive development in the biostatistics literature concerning the sequential monitoring of clinical trials. Seminal contributions were made by Armitage *et al.*,[3] Pocock[4] and O'Brien and Fleming.[5] Lan and DeMets[6] developed a general framework for monitoring trials that emphasizes the role of Brownian motion, thereby making more explicit how monitoring of clinical trials is related to the theory of sequential analysis developed in the mathematical statistics literature. In describing the application of this general framework, Lan and DeMets[7] discuss the role played by the concept of statistical information, but only in informal terms.

The purpose of this expository paper is to describe the role of information and Brownian motion in the sequential monitoring of trials in a manner that is easily accessible to clinical trials

* Present address: Department of Statistics, Hebrew University, Mount Scopus, Jerusalem 91905, Israel.

0277–6715/93/080753–13$11.50

Received November 1991
Revised July 1992

biostatisticians. In Section 2 we introduce the concept of statistical information and illustrate this concept through two one-sample non-sequential examples. In Section 3 we begin the development of a general framework for sequential monitoring by introducing certain important stochastic processes, taking as a starting point the simple problem of inference for a one-sample mean. The development continues in Section 4 with the introduction of the Brownian motion process in the context of inference for a one-sample mean. In Section 5, the general framework for sequential monitoring is described in full form. In Section 6 we describe how the general framework is applied to the following common clinical trials situations: (1) two-sample testing with a single continuous endpoint, (2) two-sample testing with longitudinal (that is, repeated measures) data, and (3) two-sample testing with a survival endpoint. Finally, in Section 7 we give a brief summary.

2. MEASURING STATISTICAL INFORMATION

In sequential monitoring, the statistically appropriate measure of how far a trial has progressed is not the sample size enrolled but rather the amount of statistical information accumulated. The amount of information accumulated is reflected fully by the sample size in certain special cases, but not in general. This section presents a simple introduction to the concept of statistical information through two examples in a one-sample non-sequential framework. Application of the concept to common two-sample clinical trial designs with sequential monitoring is described later in the paper.

A statistical problem typically can be cast as a problem of drawing inference about a parameter θ, such as a population mean. In this section we consider testing the null hypothesis H_0: $\theta = \theta_0$ using a test statistic Z that is based on an estimate $\hat{\theta}$ of θ. The test statistic Z is of the form

$$Z = (\hat{\theta} - \theta_0)/\sqrt{\text{var}(\hat{\theta})}. \tag{1}$$

The statistical information about θ provided by the data is defined to be simply the inverse of $\text{var}(\hat{\theta})$. The idea of defining information as the inverse of the variance of a parameter estimate has obvious intuitive appeal: a higher inverse variance means a lower variance, which means a greater degree of precision in the estimate, which intuitively means a greater level of information about the parameter. The intuitive appeal of this definition should help to explain it to physicians and other investigators.

The two examples below illustrate the definition and show how the amount of information available is related to the power of the test statistic Z to detect departures from H_0. The presentation deals with one-sided testing, but the same arguments apply to two-sided testing.

Example 1: one-sample problem

Suppose that $X_1, \ldots, X_N$ are independent and identically distributed (i.i.d.) observations with mean θ and variance σ^2, and that we wish to draw an inference about the mean θ. The observations could represent measurements of a clinical response variable on each of N patients. Assume for simplicity that σ^2 is known; assume further that the data have been normalized to make σ^2 equal to one. The most commonly used estimator for θ is the sample mean, $\bar{X}_N = (X_1 + \ldots + X_N)/N$. When the Xs are normally distributed, $\bar{X}_N$ is the maximum likelihood estimate (MLE) of θ.

The variance $\text{var}(\bar{X}_N)$ of $\bar{X}_N$ is $1/N$. Therefore, by the definition given above, the information I provided by the data is equal to the sample size N.

Now consider testing the null hypothesis H_0: $\theta = \theta_0$. The test statistic (1) is given by $Z = (\sqrt{N})(\bar{X}_N - \theta_0)$. The expected value of Z for general θ is given by $\mathscr{E}_\theta[Z] = (\sqrt{N})(\theta - \theta_0)$. This equation can be rewritten as

$$\mathscr{E}_\theta[Z] = (\sqrt{I})\delta, \tag{2}$$

where $\delta = \theta - \theta_0$ is the difference between the true value of the parameter θ and the null hypothesis value. By construction, $\text{var}(Z) = 1$. When the Xs have a normal distribution, Z also has a normal distribution. For Xs with an arbitrary distribution, the central limit theorem indicates that when N is large, Z is approximately normally distributed. With Z regarded as normally distributed, the test rule is to reject the null hypothesis when $Z \geqslant z_\alpha$, where z_α is the standard normal critical value corresponding to the desired type I error level (for example, $z_\alpha = 1{\cdot}96$ for a one-sided level of $\alpha = 0{\cdot}025$). In addition, the power of the test statistic Z to reject H_0 is given by

$$\text{power} = \Pr_\theta(Z \geqslant z_\alpha) = \Pr_\theta(Z - \mathscr{E}_\theta[Z] \geqslant z_\alpha - \mathscr{E}_\theta[Z]) = \Phi(\mathscr{E}_\theta[Z] - z_\alpha), \tag{3}$$

where Φ denotes the standard normal distribution function. To achieve a desired power of $1 - \beta$ one needs $\mathscr{E}_\theta[Z] = z_\alpha + z_\beta$, where z_β is the approximate standard normal critical value (for example, $z_\beta = 1{\cdot}28$ for 90 per cent power). Using (2), it follows that the information required to achieve the desired power is

$$I^* = \left(\frac{z_\alpha + z_\beta}{\delta}\right)^2. \tag{4}$$

For this example, the required sample size is $N^* = I^*$.

The next example involves the linear random effects model for longitudinal data. Readers unfamiliar with this setting may wish to skim over this example at a first reading.

Example 2: linear random effects model

Consider a longitudinal study in which a clinical response variable is measured on each of N patients at successive timepoints, with patient i measured at timepoints $\{x_l: l = 1, \ldots, L_i\}$. The number of measurements available may vary from patient to patient because of staggered entry into the study. Let Y_{il} denote the response of patient i at time x_l. Suppose that the Ys satisfy the linear random effects model, defined as

$$Y_{il} = \gamma_i + \theta_i x_l + \varepsilon_{il},$$

where γ_i is a patient-specific intercept, θ_i is a patient-specific slope, and the ε_{il} are i.i.d. error terms, independent of the γ_i and θ_i, with mean zero and variance σ_ε^2 (cf. Laird and Ware[8]). Assume also that the random vectors (γ_i, θ_i) are i.i.d. across patients, with mean (γ, θ) and covariance matrix given by $\text{var}(\gamma_i) = \sigma_\gamma^2$, $\text{var}(\theta_i) = \sigma_\theta^2$ and $\text{cov}(\gamma_i, \theta_i) = \sigma_{\gamma\theta}$. The development below deals with inference for the mean slope θ.

The least squares estimate of the patient-specific slope θ_i is given by

$$\hat{\theta}_i = \frac{\sum_{l=1}^{L_i}(x_l - \bar{x}_i)\,Y_{il}}{\sum_{l=1}^{L_i}(x_l - \bar{x}_i)^2},$$

where $\bar{x}_i = (x_1 + \cdots + x_{L_i})/L_i$. The variance of $\hat{\theta}_i$ is

$$v_i = \sigma_\theta^2 + \left[\sigma_\varepsilon^2 \Big/ \sum_{l=1}^{L_i}(x_l - \bar{x}_i)^2\right].$$

Each $\hat{\theta}_i$ can be regarded as a separate estimate of the mean slope θ. By the general definition of information stated at the beginning of this section, the amount of information associated with $\hat{\theta}_i$ as an estimate of θ is v_i^{-1}. A natural overall estimate of θ is obtained by forming a weighted linear combination of the $\hat{\theta}_i$, with each $\hat{\theta}_i$ weighted according to the amount of information associated with it. The resulting estimate is

$$\hat{\theta} = \left(\sum_{i=1}^{N} v_i^{-1} \hat{\theta}_i \right) \Big/ \left(\sum_{i=1}^{N} v_i^{-1} \right). \tag{5}$$

The estimate $\hat{\theta}$ can be shown to be the MLE of θ under the normal model in which (γ_i, θ_i) is bivariate normally distributed and ε_{il} is normally distributed.

The variance of $\hat{\theta}$ is given by $\text{var}(\hat{\theta}) = [v_1^{-1} + \cdots v_N^{-1}]^{-1}$. Accordingly, the amount of information about θ provided by the data is $I = \text{var}(\hat{\theta})^{-1} = v_1^{-1} + \cdots + v_N^{-1}$. For testing the null hypothesis H_0: $\theta = \theta_0$, the development proceeds as in Example 1.

Notice that the total amount of information provided by the data is equal to the sum over i of the amount of information associated with $\hat{\theta}_i$. The notion that total information may be expressed as the sum of individual contributions of information is reflected in Proposition 2 of Section 3.

The amount of information v_i^{-1} that individual i provides about θ may be re-expressed as

$$v_i^{-1} = \sigma_\theta^{-2} \left(1 + R \Big/ \sum_{l=1}^{L_i} (x_l - \bar{x}_i)^2 \right)^{-1},$$

where $R = \sigma_\varepsilon^2 / \sigma_\theta^2$. It is easy to show that $\sum_{l=1}^{L_i} (x_l - \bar{x}_i)^2$ increases as L_i increases. Assume that the data have been normalized so that $\sigma_\theta^2 = 1$. Then, as $\sum_{l=1}^{L_i} (x_l - \bar{x}_i)^2$ increases to infinity, v_i^{-1} increases to 1. Thus, an individual's information contribution increases gradually from 0 to 1 as more measurements are taken. This situation differs from Example 1, in which the amount of information associated with an individual jumps from 0 to 1 when the individual has the single measurement taken, with no further increase in the individual's information contribution thereafter.

The ideas of the foregoing examples carry over to any statistical problem involving a test of the form (1). Once an expression for the statistical information provided by the data is derived, the general expressions (3) and (4) can be exploited.

In certain settings, such as two sample with survival data (Section 6.3), it is more natural to deal with the test statistic Z, or a related test statistic, than with the parameter estimate $\hat{\theta}$. In this case, an expression for the statistical information can be derived by inverting (2), to obtain

$$I = (\mathscr{E}_\theta[Z]/\delta)^2. \tag{6}$$

In all of the above examples, the statistical information I does not depend upon the value of the parameter θ. This result holds typically in models that are linear in θ but not in models that are non-linear in θ, such as the binomial model or the survival model of Section 6.3. When I depends on θ, it is generally convenient to take the null hypothesis value I_{θ_0} as an approximate working definition of the statistical information. This approximation is usually adequate for practical purposes under moderate deviations from the null hypothesis, such as those represented by the moderate treatment effects that commonly arise in large-scale clinical trials.

3. SEQUENTIAL MONITORING AND STOCHASTIC PROCESSES

The foregoing discussion is now extended to the context of sequential monitoring. A key element in formulating a unified theory is the introduction of various relevant stochastic processes. The initial development here is carried out for the simple one-sample testing problem with a single continuous variable (Example 1 of Section 2). In Section 6 we will describe how the results can be applied to two-sample testing with more complex response variables.

Recall that the simple one-sample problem involves i.i.d. observations $X_1, \ldots, X_N$ with mean θ and variance σ^2. For simplicity, the assumption is made again that $\sigma^2 = 1$. The null hypothesis is H_0: $\theta = 0$.

Define the partial sum S_k by $S_0 = 0$ and $S_k = X_1 + \cdots + X_k$ for $k = 1, \ldots, N$. Elementary calculations establish that

(P1): $$\mathcal{E}[S_k] = k\theta$$

(P2): $$\text{var}(S_k) = k$$

(P3): $$\text{cov}(S_k, S_l) = \min\{k, l\}.$$

Also, by the central limit theorem, $(S_k - k\theta)/\sqrt{k}$ tends to the standard normal distribution N(0, 1) as k tends to infinity.

The variance $\text{var}(S_k) = k$ is equal to the amount of information accrued when there are k observations in the study. Furthermore, the expectation $\mathcal{E}[S_k] = k\theta$ grows linearly with the amount of information. This linear relationship can be used to examine the data 'trend' over the course of the study, as discussed and illustrated graphically by Lan and Wittes.[9] As the study progresses, the amount of information k grows from 0 to N, the final sample size. The quantity $\tau_k = k/N$ represents the fraction of information available at the kth observation, relative to the total information N associated with the planned end of the study. This quantity τ_k will be called the 'information time' associated with the calendar time of the kth observation.[7]

Some important stochastic processes associated with sequential monitoring can now be described. Let t represent elapsed calendar time over the course of the study ($t = 0$ is the start of the study and $t = T$ is the planned end of the study), and let $k(t)$ denote the number of observations accumulated by calendar time t. Also, let $I(t) = k(t)$ denote the amount of information available as of calendar time t, and let t_k denote the calendar time at which the kth observation is taken. Define $S(t) = S_{k(t)}$. Then, if an interim analysis were conducted at calendar time t, the estimate of the mean θ would be $E(t) = S(t)/k(t)$ and the test statistic for testing H_0: $\theta = 0$ would be $Z(t) = S(t)/\sqrt{(k(t))}$.

Properties (P1)–(P3) for the partial sums $\{S_k\}$ lead to the following properties for the processes $\{S(t): t \in [0, T]\}$, $\{E(t): t \in [t_1, T]\}$ and $\{Z(t): t \in [t_1, T]\}$:

(S1): $$\mathcal{E}[S(t)] = I(t)\theta$$

(S2): $$\text{var}(S(t)) = I(t)$$

(S3): $$\text{cov}(S(t), S(u)) = \min\{I(t), I(u)\}.$$

(E1): $$\mathcal{E}[E(t)] = \theta$$

(E2): $$\text{var}(E(t)) = I(t)^{-1}$$

(E3): $$\text{cov}(E(t), E(u)) = \min\{I(t)^{-1}, I(u)^{-1}\}.$$

(Z1): $$\mathscr{E}[Z(t)] = \sqrt{(I(t))}\theta$$

(Z2): $$\text{var}(Z(t)) = 1$$

(Z3): $$\text{cov}(Z(t), Z(u)) = [I(t)/I(u)]^{1/2} \quad \text{for } t < u.$$

In general, a stochastic process satisfying (S1)–(S3), (E1)–(E3) or (Z1)–(Z3), respectively, will be referred to as an *S-process*, an *E-process* or a *Z-process*, with mean θ and information function $I(t)$. For a Z-process, the total information $I(T)$ must be specified separately because it is not uniquely determined by the mean and covariance structure of the process. Given an S-process, an E-process or a Z-process, one may easily obtain processes of the remaining two types through the equations

$$S(t) = I(t)E(t) = \sqrt{(I(t))}Z(t) \tag{7}$$

$$E(t) = S(t)/I(t) = Z(t)/\sqrt{(I(t))} \tag{8}$$

$$Z(t) = S(t)/\sqrt{(I(t))} = \sqrt{(I(t))}E(t). \tag{9}$$

The concepts of an S-process, an E-process and a Z-process are useful because many sequential monitoring situations can be expressed in terms of an (S, E, Z)-process structure, as will be shown in Section 6. In clinical trials practice, study results are most familiarly expressed in terms of a Z-process (representing a sequence of Z-values) or an E-process (representing a sequence of treatment effect estimates), but for the theoretical development in Sections 4 and 5 the most natural starting point is an S-process. The following two simple propositions provide useful tools for formulating sequential schemes in clinical trial settings.

Proposition 1 Let $\{E_0(t)\}$ and $\{E_1(t)\}$ be independent E-processes with respective means θ_0 and θ_1 and respective information functions $I_0(t)$ and $I_1(t)$. Then the process $E(t) = E_1(t) - E_0(t)$ is an E-process with mean $\theta = \theta_1 - \theta_0$ and information function $I(t) = [I_0(t)^{-1} + I_1(t)^{-1}]^{-1}$.

Proposition 2 Let $\{S_1(t)\}$ and $\{S_2(t)\}$ be independent S-processes with common mean θ and respective information functions $I_1(t)$ and $I_2(t)$. Then the process $S(t) = S_1(t) + S_2(t)$ is an S-process with mean θ and information function $I(t) = I_1(t) + I_2(t)$.

The result $I(t) = I_1(t) + I_2(t)$ in Proposition 2 represents a formalized version of the notion that the total amount of information provided by two independent pieces of an S-process, represented by S_1 and S_2, is the sum of the amounts of information associated with each piece.

4. BROWNIAN MOTION

In this section we introduce Brownian motion, which plays a central role in the theory of sequential data monitoring. The presentation here continues with the one-sample problem discussed in the preceding section, and all the notation of that section carries over here. Denote the set of information times by $\mathscr{I} = \{\tau_1, \ldots, \tau_N\}$. Define a new process $B_N(\tau_k)$ for $\tau_k \in \mathscr{I}$ by

$$B_N(\tau_k) = B_N\left(\frac{k}{N}\right) = S_k/\sqrt{N} = S(t_k)/\sqrt{(I(T))}. \tag{10}$$

The quantity $B_N(\tau_k)$ is the B-value defined by Lan and Wittes.[9]

Straightforward calculations based on Properties (S1)–(S3) show the following for two information times τ and τ', with $\Theta = (\sqrt{N})\theta$ (corresponding to $\Theta = (\sqrt{N})\mu$ in Lan and Wittes[9]):

(B1): $$\mathscr{E}[B_N(\tau)] = \Theta\tau$$

(B2): $$\text{var}(B_N(\tau)) = \tau$$

(B3): $$\text{cov}(B_N(\tau), B_N(\tau')) = \min\{\tau, \tau'\}.$$

Note that B_N is defined only at the discrete information times in $\mathscr{I}$. Its definition may be extended to the continuous time scale ($s \in [0, 1]$) by setting $B_N(s) = 0$ for $s < \tau_1 = 1/N$ and $B_N(s) = B_N(\tau_k) = B_N(k/N)$ for s in the interval $[\tau_k, \tau_{k+1})$. In other words, $B_N(s) = B_N([\![Ns]\!]/N)$, with $[\![Ns]\!]$ denoting the greatest integer k such that $k \leqslant Ns$.

The resulting extended version of B_N is an S-process over s in the interval $[0, 1]$ with information function $I^*(s) = \tau_k$ in the interval $[\tau_k, \tau_{k+1})$. In particular, the total information $I^*(1)$ is equal to one. Thus, the process B_N may be viewed simply as a 'standardized' version of the S-process $\{S(t)\}$ obtained by (a) changing the time scale from calendar time t to information time s and (b) rescaling the ordinate so that the total information is normalized to a value of one.

Suppose now that $\theta = 0$ (so that $\Theta = 0$). Then $\mathscr{E}[B_N(s)] = 0$ and

$$\text{cov}(B_N(s_1), B_N(s_2)) = \min\left\{\frac{[\![Ns_1]\!]}{N}, \frac{[\![Ns_2]\!]}{N}\right\} \doteq \min\{s_1, s_2\},$$

where the approximate equality applies for large N. Now for fixed s, $[\![Ns]\!]$ tends to infinity as N tends to infinity. Therefore, by the central limit theorem result for S_k, the random variable $S_{[\![Ns]\!]}/\sqrt{[\![Ns]\!]}$ converges in distribution to N(0, 1) as N tends to infinity. In addition, for fixed s, the quantity $[\![Ns]\!]/N$ converges to s as N tends to infinity. It follows that $B_N(s)$ converges in distribution, as N tends to infinity, to the normal distribution N(0, s).

More generally, it can be shown that for any given set of information times $s_1, \ldots, s_p$ in the interval [0, 1], the random vector $(B_N(s_1), \ldots, B_N(s_p))$ converges in distribution, as N tends to infinity, to a multivariate normal random vector $(B^*(s_1), \ldots, B^*(s_p))$ with mean zero and covariance given by $\text{cov}(B^*(s_q), B^*(s_r)) = \min\{s_q, s_r\}$.

This result implies that the process $\{B_N(t)\}$ may be approximated by a stochastic process known as Brownian motion. The Brownian motion process is defined to be a stochastic process $\{B(s)\colon s \in [0, 1]\}$ such that for any $s_1, \ldots, s_p$ in [0, 1], the random vector $(B(s_1), \ldots, B(s_p))$ has a multivariate normal distribution with mean zero and covariance given by $\text{cov}(B(s_q), B(s_r)) = \min\{s_q, s_r\}$. The Brownian motion process is well known in the stochastic process literature and has been extensively studied.[10]

In the case when θ is not equal to 0, arguments similar to those above show that the random vector $[(B_N(s_1), \ldots, B_N(s_p)) - (\sqrt{N})\theta(s_1, \ldots, s_p)]$ converges in distribution to $(B(s_1), \ldots, B(s_p))$. Roughly speaking, this means that for large N the process B_N is approximately equal in distribution to the process $B_\Theta(s) = B(s) + \Theta s$, which is called Brownian motion with drift Θ. The case when θ is not equal to 0 plays a key role in conditional power calculations and stochastic curtailment in clinial trials, as discussed by Lan and Wittes.[9]

5. GENERAL MONITORING FRAMEWORK BASED ON BROWNIAN MOTION

The groundwork has now been laid for the presentation of a general framework for sequential monitoring. This general framework allows monitoring in a complex clinical trial setting to be carried out using standard methods[4–6] that can be expressed in terms of Brownian motion,

provided that the setting conforms to an (S, E, Z)-process structure. The concept of information time plays a key role in this general framework.

Because most clinical trialists tend to think in terms of Z-values, a natural starting point is the Z-process $\{Z(t)\}$, with mean θ and information function $I(t)$, that comprises the test statistic Z-values corresponding to the interim data at each calendar time t. It is of great importance to note that $\{Z(t)\}$ must have the special covariance structure described by (Z3) to qualify as a Z-process; it does not suffice simply to have $Z(t)$ distributed approximately as N(0, 1) in large samples for each separate t. Let $\tau(t) = I(t)/I(T)$ represents the information time corresponding to calendar time t. Assume that $I(t)$ jumps at K_N calendar times $t_1, \ldots, t_{K_N}$ as increments of information are accrued (for the one-sample example of Sections 3 and 4, a jump occurs each time a new observation is added). Let $\tau_k = \tau(t_k)$ be the information time at the kth of these jumps. Let $\{S(t)\}$ be the S-process corresponding to $\{Z(t)\}$, as given in (7). By analogy with (10), define

$$B_N(\tau_k) = S(t_k)/\sqrt{(I(T))} = (\sqrt{\tau_k})Z(t_k).$$

As was the case in the setting of the previous section, Properties (S1)–(S3) of the S-process imply that $\{B_N(\tau_k)\}$ satisfies (B1)–(B3) with $\Theta = (I(T))\theta$.

In other words, $\{B_N(\tau_k)\}$ has the mean and covariance structure of Brownian motion with drift $\Theta = \sqrt{(I(T))}\theta$. As in the preceding section, the definition of B_N may be extended to general s by setting $B_N(s) = 0$ for $s < \tau_1$ and $B_N(s) = B_N(\tau_k)$ for s in $[\tau_k, \tau_{k+1})$. Furthermore, in typical clinical trial applications, asymptotic arguments analogous to those in the preceding section show that $\{B_N(s): s \in [0, 1]\}$ may be regarded as approximately equal to Brownian motion with drift Θ.

These results allows existing monitoring methods based on Brownian motion to be applied to general situations. For a given calendar time t corresponding to information time s, $B_N(s)$ is related to $Z(t)$ by the equation

$$B_N(s) = (\sqrt{s})Z(t). \tag{11}$$

Thus, in carrying out a monitoring procedure, $Z(t)$ can be converted into $B_N(s)$ directly, without having to form the S-process $\{S(t)\}$ explicitly.

The application of this transformation to Brownian motion in monitoring proceeds as follows. In general, implementing a monitoring scheme involves evaluating probability expressions of the form

$$\Pi = \Pr_{H_0}(Z(t_1^*) < c_1, \ldots, Z(t_{p-1}^*) < c_{p-1}, Z(t_p^*) \geq c_p),$$

where $t_1^*, \ldots, t_p^*$ are the calendar times of the first p looks and $c_1, \ldots, c_p$ are the corresponding critical values. The quantity Π represents the probability of crossing the boundary at the pth look (but not before). Using (11), Π may be re-expressed as

$$\Pi = \Pr_{H_0}(B_N(s_1) < \tilde{c}_1, \ldots, B_N(s_{p-1}) < \tilde{c}_{p-1}, B_N(s_p) \geq \tilde{c}_p), \tag{12}$$

where $\tilde{c}_r = c_r\sqrt{s_r}$. Because B_N behaves asymptotically like an ordinary Brownian motion process $\{B(s)\}$ under H_0, the expression on the right side of (12) is asymptotically equal to the corresponding expression with B_N replaced by B. In turn, the asymptotic expression involving B may be evaluated using numerical techniques described by Lan and DeMets[6] for sequential monitoring of a Brownian motion process.

In situations where the information function $I(t)$ depends on the parameter θ, it is convenient to approximate $I_\theta(t)$ by its null hypothesis value $I_{\theta_0}(t)$, as suggested in Section 2. This approximation

avoids the problems associated with using an estimate of θ in the implementation of the sequential scheme. Again, the approximation typically is quite reasonable under moderate deviations from the null hypothesis.

In the sequential setting, the expression (3) for power no longer strictly applies. When a procedure of the O'Brien–Fleming type is used, though, the effect of sequential monitoring on power is minimal, that is, of the order of 2 percentage points of power.

In short, monitoring the Z-process $\{Z(t)\}$ (or the corresponding S-process or E-process) at $t_1^*, \ldots, t_p^*$ is exactly the same as monitoring the Brownian motion process $\{B(s)\}$ at $s_1, \ldots s_p$. This key idea can be used to adapt any sequential procedure based on Brownian motion to any setting with (S, E, Z)-process structure.

6. MONITORING IN TYPICAL CLINICAL TRIAL SETTINGS

In this section we provide an outline of how the general monitoring framework presented in Section 5 may be applied to some common clinical trial settings.

6.1. A two-sample trial with a single continuous endpoint

Let $X_1, \ldots, X_{N_1}$ denote i.i.d. observations on treatment and let $Y_1, \ldots, Y_{N_0}$ denote i.i.d. observations on control. Let $N = N_0 + N_1$ denote the total planned sample size. Denote the mean of the Xs by $\mu + \delta$ and the mean of the Ys by μ. Assume that the Xs and the Ys have a common variance of 1. The goal is to test the null hypothesis H_0: $\delta = 0$.

Let $\bar{X}_t$ and $\bar{Y}_t$ denote the sample means for treatment and control, respectively, based on the data available as of calendar time t. The standard estimate of δ for this problem based on the data as of time t is the mean difference $\hat{\delta}(t) = \bar{X}_t - \bar{Y}_t$. Because the treatment group data and the control group data can each be viewed separately as independent sets of one-sample data, the development in Section 3 indicates that the processes $\{\bar{X}_t\}$ and $\{\bar{Y}_t\}$ are independent E-processes. The process $\{\bar{X}_t\}$ has mean $\mu + \delta$ and information function $I_1(t) = N_1(t)$, where $N_1(t)$ is the sample size in the treatment group as of calendar time t. The process $\{\bar{Y}_t\}$ has mean μ and information function $I_0(t) = N_0(t)$, where $N_0(t)$ is the sample size in the control group as of calendar time t. Therefore, by Proposition 1 of Section 3, the difference process $\hat{\delta}(t)$ is an E-process with mean δ and information function

$$I(t) = \left(\frac{1}{N_0(t)} + \frac{1}{N_1(t)} \right)^{-1}. \tag{13}$$

Arguments based on the central limit theorem analogous to those in Section 5 show that the process $\{B_N(s)\}$ obtained from the E-process $\{\hat{\delta}(t)\}$ can be approximated by a Brownian motion process. Accordingly, a monitoring scheme for the trial may be formulated as described in Section 5.

The information time $\tau(t) = I(t)/I(T)$ obtained from (13) differs from the sample size ratio $\tilde{\tau}(t) = (N_0(t) + N_1(t))/(N_0 + N_1)$, referred to as the 'process time' by Lan *et al.*[11] However, $\tilde{\tau}(t)$ is nearly equal to $\tau(t)$ in many circumstances. Accordingly, in data monitoring meetings, a rough description in terms of $\tilde{\tau}(t)$ often will be appropriate, though actual calculations should use the rigorous expression for $\tau(t)$ based on (13).

A similar development may be given for the case of a two-sample trial with a binomial endpoint. In this case, though, the development involves the approximation $I \doteq I_{\theta_0}$.

6.2. A two-sample trial with repeated measurements

Consider a longitudinal study in which successive measurements are taken over time on a series of patients (planned total sample size of N), each belonging to one of two independent groups: a control group (group 0) and a treatment group (group 1). Specifically, individual i in group g is measured at timepoints $\{x_l: l = 1, \ldots, L_{gi}\}$. The number of measurements available may vary from patient to patient as in Example 2 of Section 2. Let Y_{gil} denote the response of patient (g, i) at time x_l. The Ys are assumed to follow the model

$$Y_{gil} = \gamma_{gi} + \theta_{gi} x_l + \varepsilon_{gil},$$

where the $(\gamma_{gi}, \theta_{gi})$ are i.i.d. across i, with mean (γ_g, θ_g) and common covariance matrix given by $\mathrm{var}(\gamma_{gi}) = \sigma_\gamma^2$, $\mathrm{var}(\theta_{gi}) = \sigma_\theta^2$ and $\mathrm{cov}(\gamma_{gi}, \theta_{gi}) = \sigma_{\gamma\theta}$, and the ε_{gil} are i.i.d. with mean zero and variance σ_ε^2 and are independent of γ_{gi} and θ_{gi}.

In developing a sequential monitoring scheme for this setting, the first step is to develop results for a specific individual (g, i). For calendar time t, let $L_{gi}(t)$ denote the number of measurements available on individual (g, i) and let $\hat\theta_{gi}(t)$ denote the estimate of the individual-specific slope θ_{gi} based on these measurements, computed in the manner described in Example 2 of Section 2. Further, let $t_1^{(gi)}$ denote the earliest calendar time t for which $L_{gi}(t)$ is equal to or greater than 2, that is, the earliest calendar time t for which $\hat\theta_{gi}(t)$ can be defined. Note that $\hat\theta_{gi}(t) - \theta_{gi}$ may be written as

$$\hat\theta_{gi}(t) - \theta_{gi} = \frac{\sum_{l=1}^{L_{gi}(t)} (x_l - \bar x_{gi}(t))\varepsilon_{gil}}{\sum_{l=1}^{L_{gi}(t)} (x_l - \bar x_{gi}(t))^2},$$

where $\bar x_{gi}(t) = (x_1 + \cdots + x_{L_{gi}(t)})/L_{gi}(t)$. Consider now two calendar times t and t' with $t \geqslant t'$ and $L(t)$, $L(t') \geqslant 2$. Viewing $\hat\theta_{gi}(t)$ and $\hat\theta_{gi}(t')$ as estimates of θ_{gi} and conditioning on the value of θ_{gi}, one obtains

$$\begin{aligned}
\mathrm{cov}(\hat\theta_{gi}(t), \hat\theta_{gi}(t') | \theta_{gi}) &= \mathrm{cov}\left(\frac{\sum_{l=1}^{L_{gi}(t)} (x_l - \bar x_{gi}(t))\varepsilon_{gil}}{\sum_{l=1}^{L_{gi}(t)} (x_l - \bar x_{gi}(t))^2}, \frac{\sum_{l=1}^{L_{gi}(t')} (x_l - \bar x_{gi}(t'))\varepsilon_{gil}}{\sum_{l=1}^{L_{gi}(t')} (x_l - \bar x_{gi}(t'))^2} \right) \\
&= \sigma_\varepsilon^2 \frac{\sum_{l=1}^{L_{gi}(t')} (x_l - \bar x_{gi}(t))(x_l - \bar x_{gi}(t'))}{(\sum_{l=1}^{L_{gi}(t)} (x_l - \bar x_{gi}(t))^2)(\sum_{l'=1}^{L_{gi}(t')} (x_{l'} - \bar x_{gi}(t'))^2)} \\
&= \sigma_\varepsilon^2 \frac{\sum_{l=1}^{L_{gi}(t')} (x_l - \bar x_{gi}(t'))^2}{(\sum_{l=1}^{L_{gi}(t)} (x_l - \bar x_{gi}(t))^2)(\sum_{l'=1}^{L_{gi}(t')} (x_{l'} - \bar x_{gi}(t'))^2)} \\
&= \sigma_\varepsilon^2 \Big/ \sum_{l=1}^{L_{gi}(t)} (x_l - \bar x_{gi}(t))^2.
\end{aligned}$$

Unconditionally, viewed as estimates of θ_g, the random variables $\hat\theta_{gi}(t)$ and $\hat\theta_{gi}(t')$ satisfy

$$\mathrm{cov}(\hat\theta_{gi}(t), \hat\theta_{gi}(t')) = \mathrm{cov}(\mathscr{E}[\hat\theta_{gi}(t)|\theta_{gi}], \mathscr{E}[\hat\theta_{gi}(t')|\theta_{gi}]) + \mathscr{E}[\mathrm{cov}(\hat\theta_{gi}(t), \hat\theta_{gi}(t'))|\theta_{gi})] = v_{gi}(t),$$

where

$$v_{gi}(t) = \mathrm{var}(\hat\theta_{gi}(t)) = \sigma_\theta^2 + \left[\sigma_\varepsilon^2 \Big/ \sum_{l=1}^{L_{gi}(t)} (x_l - \bar x_{gi}(t))^2 \right].$$

Thus, for given g, i, the process $\{\hat{\theta}_{gi}(t)\}$ is an E-process with mean θ_g and information function $I_{gi}(t) = v_{gi}(t)^{-1}$. By (7), the process $\{S_{gi}(t)\}$ defined by $S_{gi}(t) = 0$ for $t < t_1^{(gi)}$ and $S_{gi}(t) = v_{gi}(t)^{-1}\hat{\theta}_{gi}(t)$ for $t \geqslant t_1^{(gi)}$ is an S-process.

The next step is to combine the results across individuals within a group. In particular, because the individuals (g, i) within group g are independent, Proposition 2 of Section 3 implies the process $\{S_g(t)\}$ defined by $S_g(t) = S_{g1}(t) + \cdots + S_{gN_g}(t)$ is also an S-process, with mean θ_g and information function $I_g(t) = v_{g1}(t)^{-1} + \cdots + v_{gN_g}(t)^{-1}$. The E-process that corresponds to $\{S_g(t)\}$ in accordance with (8) is given by

$$\hat{\theta}_g(t) = \left(\sum_{i=1}^{N_g} v_{gi}(t)^{-1}\hat{\theta}_{gi}(t)\right) \Big/ \left(\sum_{i=1}^{N_g} v_{gi}(t)^{-1}\right).$$

The quantity $\hat{\theta}_g(t)$ is the weighted-average overall estimate of θ_g analogous to (5) for the data accumulated up to calendar time t.

The final step is to consider the comparison of the two groups. Because the treatment and control groups are independent, the E-processes $\{\hat{\theta}_0(t)\}$ and $\{\hat{\theta}_1(t)\}$ are independent. Therefore, by Propostion 1 of Section 3, the process $\hat{\delta}(t) = \hat{\theta}_1(t) - \hat{\theta}_0(t)$ is an E-process with mean δ and information function $I(t) = [I_0(t)^{-1} + I_1(t)^{-1}]^{-1}$. As for the case of a single continuous endpoint, arguments analogous to those in Section 4 show that the process $\{B_N(s)\}$ obtained from the E-process $\{\hat{\delta}(t)\}$ can be approximated by a Brownian motion process. Thus, a monitoring scheme for the trial again may be formulated as described in Section 5. A similar development for a more complicated repeated measurements model is given by Wu and Lan.[12]

6.3. A two-sample trial with a survival endpoint

Consider a two-sample survival trial, having a total planned sample size of N, with hazard function $\lambda_0(u)$ in the control group and hazard function $\lambda_1(u)$ in the treatment group, where u is the time from randomization. The most popular test statistic for testing the null hypothesis that the survival patterns in the two groups are identical is the logrank statistic introduced by Mantel[13] and by Peto and Peto.[14] The logrank statistic is well known to have desirable optimality properties under the Cox[15] proportional hazards model, which postulates that $\lambda_1(u) = e^{-\theta}\lambda_0(u)$ for some scalar parameter θ (with the null hypothesis corresponding to $\theta = 0$). In deriving an expression for statistical information, the most natural starting point is the logrank test itself rather than the associated estimate of the parameter θ.

Let $Z(t)$ denote the logrank Z-value based on the data available at calendar time t. Also let $N_g(t)$ denote the sample size in group g as of calendar time t, let $N(t) = N_0(t) + N_1(t)$ denote the total sample size as of calendar time t, and let $d^0(t)$ denote the expected total number of failures among both groups combined as of calendar time t under the null hypothesis H_0: $\theta = 0$.

Assume that the proportional hazards model holds and that the log hazard ratio θ is small. Then, as indicated by Schoenfeld[16] and Freedman,[17] the expected value of $Z(t)$ when $N_0(t) = N_1(t)$ is given approximately by $\mathscr{E}_\theta[Z(t)] = \theta\sqrt{(d^0(t)/4)}$. Accordingly, from (6), the appropriate expression for the statistical information $I(t)$ at calendar time t when $N_0(t) = N_1(t)$ is $I(t) = d^0(t)/4$. For general $N_0(t)$ and $N_1(t)$, similar reasoning shows that the information $I(t)$ is given by

$$I(t) = \left(\frac{1}{N_0(t)} + \frac{1}{N_1(t)}\right)^{-1}\left(\frac{d^0(t)}{N(t)}\right).$$

The process $\{Z(t)\}$ does not satisfy the precise definition of a Z-process, because Properties $(Z1)$ and $(Z3)$ hold exactly only for $\theta = 0$ and approximately only for suitably small θ. Nevertheless, one may formally define $S(t)$ by (7) and $B_N(s)$ as described in Section 4. Asymptotic arguments given by various authors[18-21] imply that $\{B_N(t)\}$ may be approximated for large N by a Brownian motion process under the null hypothesis $\theta = 0$ and by a Brownian motion process with drift for large N and suitably small θ. Thus, the framework of Section 5 may be employed to develop a sequential monitoring scheme for the trial.

7. SUMMARY

Most clinical trials require some form of periodic examination of accruing data for ethical reasons. Often a formal sequential monitoring scheme is appropriate. Accordingly, there has been much attention in the biostatistics literature to sequential monitoring of clinical trials. The literature has focused mainly on the simple case of a two-sample trial with a single continuous endpoint. Many clinical trials, however, involve a more complex data structure.

This paper has presented a general framework for sequential monitoring that may be applied to a wide variety of clinical trial situations. A key role in this framework is played by the concept of statistical information, defined as the inverse of the variance of the parameter estimate of interest, and by the concept of 'information time', defined as the ratio of the amount of information available at a given study calendar time to the amount of information that is to be available at the planned end of the study.

The crucial step is the application of a suitable transformation to the sequence of treatment effect estimates or Z-values observed over the course of the study. In typical clinical trial applications, the sequence of estimates or Z-values at hand describes an E-process or a Z-process, as defined in Section 3, and therefore may be converted to an S-process by (7).

This S-process, in turn, may be converted to a standardized B_N-type form by rescaling the time axis from calendar time to information time and normalizing the process to make the total information equal to one. The B_N-process thereby obtained, in typical clinical trial applications, can be approximated by a Brownian motion process. This result allows clinical trials with complex data structures to be monitored using procedures for monitoring a Brownian motion process.

This approach has been illustrated for three typical clinical trial settings: the two-sample trial with a single continuous endpoint, the two-sample trial with repeated measurements, and the two-sample trial with a survival endpoint. The results given for these situations have direct practical application in clinical trial monitoring. In addition, the same general approach may be used to develop sequential monitoring schemes for other situations.

ACKNOWLEDGEMENTS

Gordon Lan's work was supported by a grant from the National Cancer Institute (CA 55098). The authors thank Mike Proschan and Ed Lakatos for helpful comments.

REFERENCES

1. Wald, A. *Sequential Analysis*, Wiley, New York, 1947 (reprinted by Dover, New York, 1973).
2. Siegmund, D. *Sequential Analysis: Tests and Confidence Intervals*, Springer-Verlag, New York, 1985.
3. Armitage, P., McPherson, C. K. and Rowe, B. C. 'Repeated significance tests on accumulating data', *Journal of the Royal Statistical Society, Series A*, **132**, 235–244 (1969).
4. Pocock, S. J. 'Group sequential methods in the design and analysis of clinical trials', *Biometrika*, **64**, 191–199 (1977).

5. O'Brien, P. C. and Fleming, T. R. 'A multiple testing procedure for clincal trials', *Biometrics*, **35**, 549–556 (1979).
6. Lan, K. K. G. and DeMets, D. L. 'Discrete sequential boundaries for clinical trials', *Biometrika*, **70**, 659–663 (1983).
7. Lan, K. K. G. and DeMets, D. L. 'Group sequential procedures: calendar versus information time', *Statistics in Medicine*, **8**, 1191–1198 (1989).
8. Laird, N. M. and Ware, J. H. 'Random-effects models for longitudinal data', *Biometrics*, **38**, 963–974 (1982).
9. Lan, K. K. G. and Wittes, J. 'The *B*-value: a tool for monitoring data', *Biometrics*, **44**, 579–585 (1988).
10. Karlin, S. and Taylor, H. M. *A First Course in Stochastic Processes*, Academic Press, New York, 1975.
11. Lan, K. K. G., DeMets, D. L. and Halperin, M. 'More flexible sequential and nonsequential designs in long-term clinical trials', *Communications in Statistics*, **A13**, 2339–2353 (1984).
12. Wu, M. C. and Lan, K. K. G. 'Sequential monitoring for comparison of changes in a response variable in clinical studies', *Biometrics*, to appear.
13. Mantel, N. 'Evaluation of survival data and two new rank order statistics arising in its consideration', *Cancer Chemotherapy Reports*, **50**, 163–170 (1966).
14. Peto, R. and Peto, J. 'Asymptotically efficient rank invariant procedures (with discussion)', *Journal of the Royal Statistical Society, Series A*, **135**, 185–206 (1972).
15. Cox, D. R. 'Regression models and life-tables (with discussion)', *Journal of the Royal Statistical Society, Series B*, **34**, 187–220 (1972).
16. Schoenfeld, D. 'The asymptotic properties of nonparametric tests for comparing survival distributions', *Biometrika*, **68**, 316–319 (1981).
17. Freedman, L. S. 'Tables of the number of patients required in clinical trials using the logrank test', *Statistics in Medicine*, **1**, 121–129 (1982).
18. Tsiatis, A. A. 'Repeated significance testing for a general class of statistics used in censored survival analysis', *Journal of the American Statistical Association*, **77**, 855–861 (1982).
19. Sellke, T. and Siegmund, D. 'Sequential analysis of the proportional hazards model', *Biometrika*, **70**, 315–326 (1983).
20. Slud, E. V. 'Sequential linear rank tests for two-sample censored survival data', *Annals of Statistics*, **12**, 551–571 (1984).
21. Gu, M. G. and Lai, T. L. 'Weak convergence of time-sequential censored rank statistics with applications to sequential testing in clinical trials', *Annals of Statistics*, **19**, 1403–1433 (1991).