

Semiparametric Efficiency and Its Implication on the Design and Analysis of Group-Sequential Studies



Daniel O. Scharfstein; Anastasios A. Tsiatis; James M. Robins

Journal of the American Statistical Association, Vol. 92, No. 440 (Dec., 1997),
1342-1350.

Stable URL:

<http://links.jstor.org/sici?sici=0162-1459%28199712%2992%3A440%3C1342%3ASEAIIO%3E2.0.CO%3B2-5>

Journal of the American Statistical Association is currently published by American Statistical Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Semiparametric Efficiency and Its Implication on the Design and Analysis of Group-Sequential Studies

Daniel O. SCHARFSTEIN, Anastasios A. TSIATIS, and James M. ROBINS

Authors have shown that the time-sequential joint distributions of many statistics used to analyze data arising from group-sequential time-to-event and longitudinal studies are multivariate normal with an independent increments covariance structure. In Theorem 1 of this article, we demonstrate that this limiting distribution arises naturally when one uses an efficient test statistic to test a single parameter in a semiparametric or parametric model. Because we are able to think of many of the statistics in the literature in this fashion, the limiting distribution under investigation is just a special case of Theorem 1. Using this general structure, we then develop an information-based design and monitoring procedure that can be applied to any type of model for any type of group-sequential study provided that there is a unique parameter of interest that can be efficiently tested.

KEY WORDS: Independent increment; Information-based design and monitoring; Longitudinal study; Maximum information trial; Time-to-event study.

1. INTRODUCTION

In many experiments, investigators collect data over time in an effort to test a particular research hypothesis. In a time-to-event study, for example, individuals are enrolled during some accrual period and are followed until they experience some event of interest, they withdraw from the study, or the study is terminated. During this process, covariates are collected on each individual. Researchers are often interested in understanding the relationship between the time to event and a single covariate, possibly controlling for potential confounding factors. In a typical longitudinal study, individuals may be enrolled over time, and repeated outcome measurements are collected on each subject along with the times of measurements and other relevant covariates. In this setting, researchers may be interested in assessing whether the trajectory of the outcome variable changes according to a single covariate.

For ethical as well as practical reasons, researchers may wish to monitor a study that unfolds over time at interim analysis times to assess whether there is "enough evidence" in support of the research hypothesis to warrant early termination of the study. In general, we refer to such an experiment as a group-sequential study. In this setting, a test statistic is typically computed at each analysis time and compared to a stopping boundary. Due to repeated looks at the data, this boundary is adjusted to maintain some predetermined overall significance level. To design and monitor this type of study, the joint distribution of the sequentially computed statistics must be derived.

Often, the research hypothesis under investigation can be rephrased as a hypothesis testing question involving a single, scalar parameter in a given statistical model. In Section 3 of this article we derive the asymptotic joint distribution

of the sequentially computed semiparametric efficient test statistics for testing a single parameter in a general parametric or semiparametric model. A Wald statistic based on a semiparametric efficient estimator of the parameter of interest is one example of an efficient test statistic. Under mild regularity conditions, we demonstrate that the limiting distribution is multivariate normal with an independent increments covariance structure. Furthermore, the joint distribution depends only on the statistical information available at the interim analysis times. For ease of exposition, we focus on the case of a single, scalar parameter of interest; however, no additional effort is needed to establish this same distributional structure in the multiparameter setting.

Many statistics have been developed to analyze data from group-sequential studies, and their time-sequential distributional properties have been derived on a case-by-case basis. Because we are able to view many of these statistics as efficient test statistics for testing a single parameter in an appropriate parametric or semiparametric model, we know that their limiting distribution is a special case of our main result. For time-to-event studies, the special cases include the work of Tsiatis, Boucher, and Kim (1995) who derived the joint distribution of the sequentially computed score tests and maximum likelihood estimates for a single parameter in a general parametric model for the survival distribution. Gu and Ying (1995), generalizing the work of Sellke and Siegmund (1983), Tsiatis (1981), and Tsiatis, Rosner, and Tritchler (1985), assumed that the time-to-event distribution followed a proportional hazards model and showed that the time-sequential partial likelihood score statistics have the outlined limiting distribution. This result holds because the partial likelihood score statistic can be shown to be a semiparametric efficient test statistic. Gu and Lai (1991), Slud (1984), and Tsiatis (1982) characterized the time sequential distribution of a general class of weighted logrank statistics where the weights can depend on unknown functions of calendar and study time. When these weight functions do not depend on the calendar time of analysis, the independent increments structure emerges.

Daniel D. Scharfstein is Assistant Professor of Biostatistics, Johns Hopkins School of Hygiene and Public Health, Baltimore, MD 21205. Anastasios A. Tsiatis is Professor of Statistics, North Carolina State University, Raleigh, NC 27695. James M. Robins is Professor of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA 02115. This research was sponsored in part by National Institutes of Health grants R01-AI-31789 and R01-AI-32475 from NIAID and R01-CA-51962 from NCI. The authors are grateful to Joseph Hogan and Andrea Rotnitzky for their help.

Harrington and Fleming's (1982) G^p test statistics are members of this latter class. Andersen, Borgan, Gill, and Kieding (ABGK; 1993, sec. V.2) presented the semiparametric model for which the G^p statistics are semiparametric efficient. For longitudinal studies, Reboussin, Lan, and DeMets (1992), building on the work of Lee and DeMets (1991), assumed that the repeated measures on an individual followed a random-effects model (Laird and Ware 1982) and showed that the time-sequential distribution of the maximum likelihood estimates of the fixed effects is multivariate normal with an independent increments covariance structure. Wu and Lan (1992) established a similar result for the linear random-effects model where one is interested in comparing the rates of change in repeated responses between two treatment groups. In the situation where repeated-measures data are assumed to follow a marginal mean model, Gange and DeMets (1996) and Lee, Kim, and Tsiatis (1996) derived the limiting distribution of the time-sequential parameter estimates obtained by the generalized estimating equation methodology of Liang and Zeger (1986). The limiting distribution has uncorrelated increments only if the "working variances" are equal to or consistently estimate the true variances. This is the same condition required by Chamberlain (1987) to show efficiency of these parameter estimates. Jennison and Turnbull (1997) summarized many of these individual results and also established the outlined distributional structure for efficient estimates of the parametric component in normal linear models, general parametric regression models, and the proportional hazards regression model. They also utilized the efficiency of the estimates to provide a simple, heuristic proof of why the independent increments property must hold.

Using our main distributional result, we show in Section 4 how to exploit this structure to develop a general information-based design and monitoring procedure. We provide detailed examples in Section 5. In Section 6 we present a simulation study that assesses the degree of conformance between our design and monitoring procedure and the desired operating characteristics (i.e., power and type I error) of the study. Finally, we devote the last section to a summary of our results as well as future research directions.

2. DATA AND MODEL

Suppose that the full data collected on the i th ($i = 1, \dots, n$) individual in the study are described by a continuous-time random-data process $X_i = \{E_i, (D_{u,i}, u \geq 0)\}$, where E_i is the entry time into the study and $D_{u,i}$ represents all of the additional data collected during the u time units on study. We assume that the data are collected over a fixed time interval. Therefore, letting n approach infinity means that an increasing number of individuals arrive in that fixed time frame. Furthermore, we assume that these data processes are iid. The restriction of the full data process to any calendar time t is represented by the process $X_{t,i} = \{\Delta_i(t), \Delta_i(t)E_i, (D_{u,i}, 0 \leq u \leq t - E_i)\}$, where $\Delta_i(t) = 1(E_i \leq t)$ and $1(A)$ denotes the indicator function of the event A . Because the full data processes are iid, we know that $X_{t,1}, \dots, X_{t,n}$ are also iid.

Let X denote the full data collected on an individual. We assume that the set of all probability measures that may have generated X satisfies a semiparametric model of the form $\mathcal{S} = \{P_{\beta,\theta}: \beta \in \beta, \theta \in \theta\}$, where each probability measure in \mathcal{S} is indexed by a parameter of interest $\beta \in \beta \subset R^1$ and a nuisance parameter θ taking values in an infinite dimensional set θ . We also assume that all of the measures in \mathcal{S} are dominated by a common σ -finite measure ν and define the corresponding density by $f(\cdot, \beta, \theta)$. We let (β_0, θ_0) index the true probability measure.

For $t > 0$, let X_t represent the data available at time t . Then X_t is a measurable transformation of X . This transformation induces a set of restricted probability measures at time t , $\{P_{\beta,\theta}^t: \beta \in \beta, \theta \in \theta\}$, which are dominated by a common σ -finite measure ν_t . We denote the corresponding density by $f_t(\cdot, \beta, \theta)$. We assume that β is identifiable at all t , but allow for the fact that θ may not be. That is, there may be multiple values of θ that yield the same density for X_t . To avoid identifiability problems, we assume that there exists a many-to-one, continuously differential function $h_t: \theta \rightarrow \theta_t$, where $\theta_t = h_t(\theta)$ and $P_{\beta,\theta_t}^t = P_{\beta',\theta'_t}^t$ implies that $\beta = \beta'$ and $\theta_t = \theta'_t$. We consider $\mathcal{S}^t = \{P_{\beta,\theta_t}^t: \beta \in \beta, \theta_t \in \theta_t\}$ to be the corresponding semiparametric model at time t , and $\mathcal{L}_t(x_t, \beta, \theta_t)$ to be the likelihood function. With $\theta_{t0} = h_t(\theta_0)$, we let (β_0, θ_{t0}) index the true probability measure.

3. TIME-SEQUENTIAL ASYMPTOTIC JOINT DISTRIBUTION OF SEMIPARAMETRIC EFFICIENT TEST STATISTICS

At interim analysis time t , we have n independent copies of X_t . Let $\mathbf{X}_{t,n}$ denote a vector containing these copies. Using these data, we are interested in testing the null hypothesis $H_0: \beta = \beta_0$ against local alternatives. One way of performing this test at time t is to use a semiparametric efficient test statistic, $Z_t^*(\mathbf{X}_{t,n}, \beta_0)$. A special type of efficient test statistic is the Wald statistic based on a semiparametric efficient estimate of β_0 . If we let $\hat{\beta}_t^*(\mathbf{X}_{t,n})$ denote a semiparametric efficient estimator of β_0 at time t , then the efficient Wald statistic is of the form

$$W_t^*(\mathbf{X}_{t,n}, \beta_0) = \frac{\hat{\beta}_t^*(\mathbf{X}_{t,n}) - \beta_0}{\text{SE}[\hat{\beta}_t^*(\mathbf{X}_{t,n})]},$$

where $\text{SE}[\hat{\beta}_t^*(\mathbf{X}_{t,n})]$ is the standard error of $\hat{\beta}_t^*(\mathbf{X}_{t,n})$. In this section we present the asymptotic distribution of the sequentially computed semiparametric efficient test statistics. The proof of this result is given in the Appendix.

The main result is built on the theory of semiparametric efficient estimation as presented by Bickel, Klaassen, Ritov, and Wellner (BKRW; 1993) and Newey (1990) and the corresponding theory of semiparametric efficient hypothesis testing as described by ABGK (1993), Choi, Hall, and Schick (1996), and Robins and Rotnitzky (1992). Before establishing this result, we formally define a semiparametric efficient test statistic. For this discussion, we define a parametric submodel at time t , $\mathcal{P}^t = \{P_{\beta,\eta_t}^t: \beta \in \beta, \eta_t \in \eta_t\}$, to be the set of probability measures indexed by β and a nuisance parameter η_t taking on values in a finite dimen-

sional set $\eta_t \subset \theta_t$, and where there exists (β_0, η_{t0}) such that $P_{\beta_0, \theta_{t0}}^t = P_{\beta_0, \eta_{t0}}^t$. We let $\mathcal{L}_t(x_t, \beta, \eta_t)$ denote the likelihood function associated with \mathcal{P}^t . The likelihood or parametric submodel is said to be regular if it follows the smoothness conditions outlined in definition A.1 of Newey (1990) and the information matrix is nonsingular on $\beta \times \eta_t$. We make reference to θ_t and η_t when referring to the properties of the semiparametric model and the parametric submodel.

To perform a hypothesis test of $H_0: \beta = \beta_0$ against the class of local alternatives $\{\beta_n\}$, where $\sqrt{n}(\beta_n - \beta_0) \rightarrow \tau \neq 0$, we need a test statistic whose asymptotic distribution is known both under the null hypothesis and under local alternatives. We restrict attention to the class of test statistics that are regular and asymptotically linear (RAL). A test statistic $Z_t(\mathbf{X}_{t,n}, \beta_0)$ is regular if its limiting distribution is the same for all local data-generating processes $P_{\beta_0, \eta_{t,n}}^t$, where $\sqrt{n}(\eta_{t,n} - \eta_{t0})$ is bounded. A test statistic is asymptotically linear if it admits the expansion $\sum_{i=1}^n \phi_t(X_{t,i}, \beta_0, \theta_{t0})/\sqrt{n} + o_p(1)$, where $E_{\beta_0, \theta_t}[\phi_t(X_t, \beta_0, \theta_t)] = 0$ and $\text{var}_{\beta_0, \theta_t}[\phi_t(X_t, \beta_0, \theta_t)] = 1$, whatever be θ_t . We refer to $\phi_t(X_t, \beta_0, \theta_{t0})$ as a standardized "testing" influence function. We impose the additional condition that $E_{\beta, \eta_t}[\phi_t(X_t, \beta_0, \theta_{t0})^2]$ exists and is continuous at (β_0, η_{t0}) for each regular parametric submodel. By convention, we assume that an expectation without a subscript is taken with respect to the true probability measure, whereas expectations subscripted by (β, θ_t) and (β, η_t) are taken with respect to P_{β, θ_t}^t and P_{β, η_t}^t .

To establish the distribution of $Z_t(\mathbf{X}_{t,n}, \beta_0)$ under local data-generating processes, we define the nuisance tangent space at time t , $\Lambda_t(X_t)$, to be the space spanned by all linear combinations of the nuisance scores at time t from all regular parametric submodels. More formally, we define $\Lambda_t(X_t)$ to be the mean squared closure of the set of all random variables $a'S_{t, \eta_t}(X_t)$, where $S_{t, \eta_t}(X_t)$ is the score for η_t in some regular parametric submodel (i.e., $S_{t, \eta_t}(X_t) = \partial \log \mathcal{L}_t(X_t, \beta_0, \eta_{t0})/\partial \eta_t$) and a is a conformable constant vector. We consider $\Lambda_t(X_t)$ to be a subset of a Hilbert space of random variables H , with inner product $E[H_1 H_2]$ and $E[H^2] < \infty$. The projection of a random variable H on $\Lambda_t(X_t)$, $\Pi[H|\Lambda_t(X_t)]$, exists and is the element of $\Lambda_t(X_t)$ that satisfies $E[(H - \Pi[H|\Lambda_t(X_t)])l] = 0$ for all $l \in \Lambda_t(X_t)$. We define the efficient score for β at time t as $S_{t, \text{eff}}(X_t) = S_{t, \beta}(X_t) - \Pi[S_{t, \beta}(X_t)|\Lambda_t(X_t)]$, where $S_{t, \beta}(X_t)$ is the score for β evaluated at the truth (i.e., $S_{t, \beta}(X_t) = \partial \log \mathcal{L}_t(X_t, \beta_0, \theta_{t0})/\partial \beta$). The information bound \mathcal{I}_t is equal to the variance of $S_{t, \text{eff}}(X_t)$. The following lemma is borrowed from appendix 1 of Robins and Rotnitzky (1992).

Lemma 1. A RAL test statistic $Z_t(\mathbf{X}_{t,n}, \beta_0)$ with influence function $\phi_t(X_t, \beta_0, \theta_{t0})$ is asymptotically normal with mean $\tau E[\phi_t(X_t, \beta_0, \theta_{t0})S_{t, \text{eff}}(X_t)]$ and variance 1 under a local data-generating process $P_{\beta_n, \eta_{t,n}}^t$ in any regular parametric submodel where $\sqrt{n}(\beta_n - \beta_0) \rightarrow \tau$ and $\sqrt{n}(\eta_{t,n} - \eta_{t0}) \rightarrow \rho$.

Using the Cauchy-Schwarz inequality and the fact that

the variance of $\phi_t(X_t, \beta_0, \theta_{t0})$ is 1, we know that the mean under local alternatives, $\tau E[\phi_t(X_t, \beta_0, \theta_{t0})S_{t, \text{eff}}(X_t)]$, is bounded from above by $\tau \mathcal{I}_t^{1/2}$. A RAL test statistic $Z_t^*(\mathbf{X}_{t,n}, \beta_0)$ with influence function $\phi_t^*(X_t, \beta_0, \theta_{t0}) = \mathcal{I}_t^{-1/2} S_{t, \text{eff}}(X_t)$ will attain this upper bound. This statistic is referred to as a RAL semiparametric efficient test statistic. In fact, it can be shown that this test statistic is the basis of an optimal asymptotic size α test within the class of tests that are asymptotically unbiased and based on test statistics that are RAL (ABGK 1993; Choi et al. 1996).

To show how an efficient Wald statistic fits into the foregoing testing framework, we rely on the theory of semiparametric efficient estimation as discussed by BKRW (1993) and Newey (1990). In these works, a semiparametric efficient estimator of β_0 at time t is an estimator that is regular and is such that, at the truth,

$$\sqrt{n}(\hat{\beta}_t^*(\mathbf{X}_{t,n}) - \beta_0) = \sum_{i=1}^n \mathcal{I}_t^{-1} S_{t, \text{eff}}(X_{t,i})/\sqrt{n} + o_p(1).$$

Because we assume that $\text{SE}[\hat{\beta}_t^*(\mathbf{X}_{t,n})]^2/(\mathcal{I}_t^{-1})$ converges in probability to 1, we know that the efficient Wald statistic $W_t^*(\mathbf{X}_{t,n}, \beta_0)$ is a RAL, semiparametric efficient test statistic.

Theorem 1. Within a given semiparametric model with parameter of interest β , suppose that we are interested in testing the null hypothesis $H_0: \beta = \beta_0$ against local alternatives β_n , where $\sqrt{n}(\beta_n - \beta_0) \rightarrow \tau \neq 0$. Furthermore, suppose that $Z_{t_j}^*(\mathbf{X}_{t_j,n}, \beta_0)$ ($j = 1, \dots, J, t_1 < t_2 < \dots < t_J$) are RAL semiparametric efficient test statistics, with respective influence functions $\phi_{t_j}^*(X_{t_j}, \beta_0, \theta_{t_j,0}) = \mathcal{I}_{t_j}^{-1/2} S_{t_j, \text{eff}}(X_{t_j})$. Then,

$$(\hat{\mathcal{I}}_{t_1}(\mathbf{X}_{t_1,n})^{1/2} Z_{t_1}^*(\mathbf{X}_{t_1,n}, \beta_0), \dots, \hat{\mathcal{I}}_{t_J}(\mathbf{X}_{t_J,n})^{1/2} Z_{t_J}^*(\mathbf{X}_{t_J,n}, \beta_0))' \xrightarrow{\mathcal{D}} \mathcal{N}(\delta, \Sigma).$$

Here δ a $J \times 1$ vector, contains 0s under the null hypothesis and elements $\tau \mathcal{I}_{t_j}$ under local alternatives, Σ is a $J \times J$ matrix with an independent increments covariance structure (i.e., the (j, j') th element of Σ , where $j \leq j'$, is equal to \mathcal{I}_{t_j}), and $\hat{\mathcal{I}}_{t_j}(\mathbf{X}_{t_j,n})$ is a consistent estimate for \mathcal{I}_{t_j} .

Remark. As a special case, this theorem holds for all parametric models where $Z_t^*(\mathbf{X}_{t,n}, \beta_0)$ could be either the standardized score statistic for β evaluated at β_0 and the restricted maximum likelihood estimator of the nuisance parameters or a Wald statistic based on the maximum likelihood estimator of β_0 . In this setting, \mathcal{I}_t is the Fisher information based on the data available at time t .

In this theorem efficiency is a sufficient condition for the emergence of the independent increments structure. However, it is not necessary. There are sequentially computed RAL inefficient test statistics and sequentially computed Wald statistics based on RAL inefficient estimators that also have this asymptotic distributional structure (e.g., the Mantel-Haenszel estimator of the log hazard ratio in a two-sample proportional hazards survival model).

Finally, note that Theorem 1 can be generalized to the situation where the parameter of interest is multidimensional. In this setting testing would be carried out via quadratic forms that would have an asymptotic chi-squared distribution. The independent increments property can then be exploited to greatly simplify the construction of group-sequential boundaries (see Jennison and Turnbull 1991).

4. INFORMATION-BASED DESIGN AND MONITORING PROCEDURE

This framework affords us the opportunity to generalize the theory of group-sequential methods. Classical designs have focused on determining the maximum sample size required to ensure the proper operating characteristics of the study. In this framework, however, we see that the distributions of the test statistics are more generally related to the information available at the time of analysis. Thus we focus our design considerations on the maximum information required to attain the desired characteristics. As a result, we use what we call an *information-based design and monitoring procedure*.

4.1 Design

The first step in the design of a group-sequential study is to select a model for the full-data process and characterize the research hypothesis through a single parameter, β . For example, in a time-to-event study designed to compare two treatment groups, we may be willing to assume that the log hazard ratio is constant over time. In this case, we could use a proportional hazards model and analyze the data using either the logrank test statistic or a Wald statistic based on the maximum partial likelihood estimator of β . For this general procedure, we use the efficient Wald statistic, $W_t^*(\mathbf{X}_{t,n}, \beta_0)$, but this restriction is not critical, because of the asymptotic equivalence between semiparametric efficient test statistics. Note that many efficient test statistics along with estimates of the information bound are readily available from standard statistical software packages.

We are interested in testing the null hypothesis $H_0: \beta = \beta_0$ versus the alternative $H_A: \beta \neq \beta_0$. In the second step, we need to specify the desired operating characteristics of the study. Thus we need to identify the alternative of interest, β_A , along with the power, $1 - \gamma$, with which we want to be able to detect this alternative. In this step we also need to decide on the level of significance, α , and the maximum number of interim analyses to be performed, J .

The next step is to determine the method of calculating the stopping boundary. We propose using the α -spending approach of Lan and DeMets (1983), which is a method of determining the amount of the overall type I error to utilize at each interim analysis based on some estimate of the proportion of information collected. In particular, we need to decide what kind of α -spending function to use. In general, we denote the spending function by $\alpha(\pi)$, where $0 \leq \pi \leq 1$. It is important to remember that $\alpha(\pi)$ is a nondecreasing function with $\alpha(0) = 0$ and $\alpha(1) = \alpha$. Two common use functions have received considerable attention in the literature. These functions correspond to what are referred to as

the O'Brien–Fleming boundary and the Pocock boundary (Lan and DeMets 1983). The former function tends to be very conservative at the early stages of the study, whereas the latter is more liberal. For an O'Brien–Fleming boundary, we take $\alpha(\pi) = 4 - 4\Phi(z_{\alpha/4}/\sqrt{\pi})$, and for a Pocock boundary, we take $\alpha(\pi) = \alpha \log(1 + (\exp\{1\} - 1)\pi)$. In these formulas, $\Phi(\cdot)$ and z_x are the cumulative density function and $1 - x$ quantile of a standard normal random variable.

The fourth step in the design stage is to compute the maximum information required to preserve the operating characteristics of the study. To see how this maximum information is determined, suppose that we are analyzing the data at only one analysis time, t^* , and we ask the question “how much information is required to have power $1 - \gamma$ and type I error α ?” For sufficiently large n , we know from Theorem 1 that under $H_0: \beta = \beta_0$,

$$W_{t^*}^*(\mathbf{X}_{t^*,n}, \beta_0) \approx N(0, 1),$$

and under local alternatives of the form $\beta_n = \beta_0 + \tau/\sqrt{n}$,

$$W_{t^*}^*(\mathbf{X}_{t^*,n}, \beta_0) \approx N(\tau\sqrt{\mathcal{I}_{t^*}}, 1).$$

The test $H_0: \beta = \beta_0$ versus local alternatives will have power $1 - \gamma$ and size α if $\tau = (z_{\alpha/2} + z_\gamma)/\sqrt{\mathcal{I}_{t^*}}$. To handle a fixed alternative, we suppose that $n = \lceil \tau/(\beta_A - \beta_0) \rceil^2$. This implies that $\beta_n = \beta_A$. Assuming that n is sufficiently large, we then have the total information required at t^* is

$$n\mathcal{I}_{t^*} = \left(\frac{z_{\alpha/2} + z_\gamma}{\beta_A - \beta_0} \right)^2.$$

But when the data are monitored at some interim analysis times, there is a loss of power. Hence the maximum information must be inflated. So we can define maximum information as

$$\text{MI} = \left(\frac{z_{\alpha/2} + z_\gamma}{\beta_A - \beta_0} \right)^2 \text{IF}, \quad (1)$$

where IF denotes the inflation factor. Assuming that the interim analyses are performed after accumulating equal amounts of information, and using the fact that the test statistics are proportional to the partial sum of independent normal random variables, we can use the results of Wang and Tsiatis (1987) to determine the inflation factor as a function of J, α, γ , and the type of spending function. Table 1 presents a series of representative inflation factors. The important fact about these inflation factors is that they are highly insensitive to even gross deviations in the equal information assumption.

To assess whether this maximum information is attainable within the resources of the study, we must conduct some theoretical calculations or simulations to compute the number of subjects and other relevant design parameters (e.g., duration of the trial, accrual rate, number of repeated measurements, etc.) necessary to achieve this amount of information. These calculations require initial guesses for the nuisance parameters. At this point, serious miscalculation can occur. We recommend that the interim analysis times be used not only to assess the relationship of interest, but

Table 1. Inflation Factors as a Function of J , α , γ , and the Type of Spending Function

J	Spending function	$\alpha = .05$			$\alpha = .01$		
		Power = $1 - \gamma$			Power = $1 - \gamma$		
		.80	.90	.95	.80	.90	.95
2	Pocock	1.11	1.10	1.09	1.09	1.08	1.08
	O-F	1.01	1.01	1.01	1.00	1.00	1.00
3	Pocock	1.17	1.15	1.14	1.14	1.12	1.12
	O-F	1.02	1.02	1.02	1.01	1.01	1.01
4	Pocock	1.20	1.18	1.17	1.17	1.15	1.14
	O-F	1.02	1.02	1.02	1.01	1.01	1.01
5	Pocock	1.23	1.21	1.19	1.19	1.17	1.16
	O-F	1.03	1.03	1.02	1.02	1.01	1.01
6	Pocock	1.25	1.22	1.21	1.20	1.19	1.17
	O-F	1.03	1.03	1.03	1.02	1.02	1.02
7	Pocock	1.26	1.24	1.22	1.22	1.20	1.18
	O-F	1.03	1.03	1.03	1.02	1.02	1.02

also to update the initial guesses and potentially revise the study design to help attain the maximum information (see Scharfstein and Tsiatis 1997).

4.2 Monitoring

The monitoring phase of the procedure is related to the work of Lan and Zucker (1993). In a group-sequential study, we monitor the study at interim times t_1, \dots, t_J . At each analysis time, say t_j , we compute the Wald test statistic, $W_{t_j}^*(\mathbf{X}_{t_j,n}, \beta_0)$, which incorporates all of the information up until the analysis time. If this statistic exceeds the stopping boundary value b_j , that is,

$$|W_{t_j}^*(\mathbf{X}_{t_j,n}, \beta_0)| \geq b_j,$$

then we may terminate the study and reject the null hypothesis. To calculate boundary values that preserve the operating characteristics of the study, we utilize the α -spending approach of Lan and DeMets (1983) along with the special distributional structure of the efficient test statistics. That is, we take advantage of the fact that the time-sequential limiting distribution of the efficient test statistics is normal with an independent increments covariance structure.

At the first analysis time t_1 , we can estimate the information time by $\pi_1 = \{SE[\hat{\beta}_{t_1}(\mathbf{X}_{t_1,n})]^{-2}\}/MI$, because $SE[\hat{\beta}_{t_1}(\mathbf{X}_{t_1,n})]^{-2}$ is an estimate for the total information available at time t_1 . Then the first boundary cutoff is the value of b_1 where $\Pr_{\beta=\beta_0}(|W_{t_1}^*(\mathbf{X}_{t_1,n}, \beta_0)| \geq b_1) = \alpha(\pi_1)$. Because $W_{t_1}^*(\mathbf{X}_{t_1,n}, \beta_0)$ is normal with mean 0 and variance 1, we know that $b_1 = z_{\alpha(\pi_1)/2}$. After this cutoff is determined, we compute the observed value of the test statistic. If it exceeds b_1 , then we stop and reject H_0 ; otherwise, we continue to the next monitoring time.

Consider the j th ($j = 2, \dots, J - 1$) analysis time, and suppose that the boundary values b_1, \dots, b_{j-1} have been computed. At this analysis time, we estimate the information time by $\pi_j = \{SE[\hat{\beta}_{t_j}(\mathbf{X}_{t_j,n})]^{-2}\}/MI$, and we find the boundary value b_j that solves the following equation:

$$\begin{aligned} &\Pr_{\beta=\beta_0} (|W_{t_1}^*(\mathbf{X}_{t_1,n}, \beta_0)| \\ &\leq b_1, \dots, |W_{t_{j-1}}^*(\mathbf{X}_{t_{j-1},n}, \beta_0)| \leq b_{j-1}, \\ &|W_{t_j}^*(\mathbf{X}_{t_j,n}, \beta_0)| \geq b_j) = \alpha(\pi_j) - \alpha(\pi_{j-1}). \end{aligned}$$

This solution is easily computed by using the independent increments property of the test statistics and the recursive numerical integration algorithm of Armitage, McPherson, and Rowe (1969). Again, the observed value of test statistic is determined and compared to its cutoff. If it exceeds b_j , then we stop and reject H_0 ; otherwise, we continue to the next monitoring time.

At the final analysis time, we use up the remaining significance level. That is, we compute b_J , where

$$\begin{aligned} &\Pr_{\beta=\beta_0} (|W_{t_1}^*(\mathbf{X}_{t_1,n}, \beta_0)| \\ &\leq b_1, \dots, |W_{t_{J-1}}^*(\mathbf{X}_{t_{J-1},n}, \beta_0)| \leq b_{J-1}, \\ &|W_{t_J}^*(\mathbf{X}_{t_J,n}, \beta_0)| \geq b_J) = \alpha - \alpha(\pi_{J-1}). \end{aligned}$$

There are two main ways of implementing this monitoring scheme. First, we can conduct the study for a fixed duration—a *maximum duration trial*. In this type of monitoring scheme several situations require special treatment. There may be a monitoring time, t_j , where the observed information is greater than or equal to MI ($\pi_j \geq 1$). In this event, we treat t_j as the final analysis time. That is, we suggest spending the remaining type I error and terminating the study. At this point, the significance level is maintained at α , and the power of the study is greater than $1 - \gamma$. The second situation arises when the final analysis time is reached but the observed information is less than MI ($\pi_J < 1$). Because this scheme deals with a fixed trial duration, we spend the remaining type I error, terminate the study, and report that it is underpowered. The discrepancy in power from the original design can be quite significant. This situation can seriously detract from the credibility of the study and provides increased motivation for using the information at earlier interim analysis times to help restructure the elements of the study that will help us attain the maximum level of information. Another problem that can arise is that the estimated information times are not monotonic. That is, the information time at an earlier analysis is greater than that of a later analysis. Although theoretically this cannot happen because the expected information must increase with time, it potentially can occur in practice because we are estimating this quantity. If nonmonotonicity is encountered, then we recommend that the interim analysis be bypassed until a point where the information time is greater than all previously calculated information times. In practice this latter problem very rarely arises. In a simulation study that we conducted (see Sec. 6), we never encountered any nonmonotonicities.

An alternative to the maximum duration trial is a *maximum information trial*. Here the study is monitored and continued either until the null hypothesis is rejected or until a time t where $SE[\hat{\beta}_t(\mathbf{X}_{t,n})]^{-2} = MI$. In this type of trial, we can also encounter the nonmonotonicity issue and recommend the same approach as earlier. Because the power and type I error are likely to be maintained under this scheme, we recommend implementing a maximum information trial. In practice we realize that it may be difficult to implement these schemes and adhere to the original monitoring plan. That is, it may require fewer or more interim analyses than originally specified. If fewer looks are required, then the

study will be overpowered. That is, suppose that we designed a study to be monitored five times with 90% power. The maximum information required in this study design would be larger than that of the study with fewer looks and the same power. Therefore, the study with fewer looks that collects the larger amount of information will have power greater than 90%. By similar reasoning, if more looks are needed than originally planned, then the study will be underpowered. However, it is possible to argue analytically that in practice the underpowering will be minimal.

5. EXAMPLES

5.1 Monitoring: The ACTG 019 Study

AIDS Clinical Trial Group Protocol 019 (ACTG 019) was a phase III clinical trial in which 1,338 adults with asymptomatic human immunodeficiency virus (HIV) infection who had CD4+ cell counts of fewer than 500/mm³ were randomized to receive one of three treatment protocols: placebo (428 subjects); zidovudine, 500 mg/day (453); or zidovudine, 1500 mg/day (457). After entry into the study, CD4+ cell counts were collected on each patient at baseline, week 8, week 16, and every 16 weeks thereafter. The primary goal of the study was to compare the relative efficacy of the three treatment arms with respect to the time from randomization until the onset of AIDS, ARC (AIDS-related complex), or death. Enrollment in the study began on July 14, 1987, and continued until May 10, 1989, at which time the data were frozen in preparation for the first full interim analysis time to be presented in August 1989. At the data safety monitoring board meeting, the placebo arm was terminated in favor of the low-dose zidovudine arm (Volberding et al. 1990).

In retrospect, suppose that we had designed the trial to compare the relative efficacy of the placebo versus low-dose zidovudine arms based on changes in CD4+ counts over time. A plausible model for these data is to assume that the log CD4+ cell counts for an individual follow a mixed-effects linear model where the measurements vary around a straight line whose intercept varies around a general population intercept and whose slope varies around a treatment-specific slope. Specifically, suppose that for the i th subject, K_i measurements are collected at times from entry ($U_{i,1}, \dots, U_{i,K_i}$). Note that $U_{i,j}$ takes values in the set $\{0, 8, 16, 32, 48, \dots\}$. Let $\mathbf{Y}_i = (Y_{i,U_{i,1}}, \dots, Y_{i,U_{i,K_i}})$ denote the vector of repeated measurements observed on the i th individual. Let Z_i be the treatment indicator ($Z_i = 1$ for zidovudine, 0 otherwise). For notational convenience, define $\mathbf{R}_i = (E_i, K_i, U_{i,1}, \dots, U_{i,K_i})$. Then the model assumes that for each individual i ($i = 1, \dots, n$),

$$\mathbf{Y}_i | Z_i, \mathbf{R}_i = \mathbf{W}_i \boldsymbol{\alpha} + \mathbf{V}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad (2)$$

where \mathbf{W}_i is a $K_i \times 3$ matrix with the j th row equal to $(1, U_{i,j}, Z_i U_{i,j})$, $\boldsymbol{\alpha} = (\xi, \zeta, \beta)'$, \mathbf{V}_i is a $K_i \times 2$ matrix with the j th row equal to $(1, U_{i,j})$, $\boldsymbol{\varepsilon}_i$ conditional on \mathbf{R}_i is an independently distributed $N(0, \sigma^2 \mathbf{I}_{\mathbf{R}_i})$ random vector, and $\mathbf{b}_i = (a_i, b_i)'$ is distributed $N(0, \Sigma_D)$ independent of each other and of $\boldsymbol{\varepsilon}_i$. Here $\mathbf{I}_{\mathbf{R}_i}$ is a $K_i \times K_i$ identity matrix and Σ_D is an $r \times r$ positive definite covariance matrix.

We are interested in testing the null hypothesis $H_0: \beta = 0$ versus the alternative hypothesis $H_A: \beta \neq 0$. To perform this test, we can use a Wald statistic based on the maximum likelihood estimator of β . Suppose that the researchers were interested in detecting the alternative $e^{52\beta} = 1.15$ with 80% power. Furthermore, suppose that the study was planned for four interim analyses to be conducted every 6 months with a stopping rule based on an O'Brien-Fleming boundary and a two-sided significance level of 5%. Using Equation (1) and Table 1, we see that the maximum information required to achieve the desired operating characteristics is

$$MI = \left(\frac{1.96 + .84}{\log(1.15)/52} \right)^2 * 1.02 = 1,106,993.50.$$

We reconstructed the data as it would have appeared at the following interim analysis times: 1/15/88, 7/15/88, 1/15/89, and 7/15/89. Figure 1 shows the predicted log CD4+ cell counts over time for each of the two treatment groups based on fitting Model (2) at each of the four interim analyses. Table 2 presents a summary of the interim analyses. Based on the data collected by the first interim analysis, the model predicts that both treatment groups have an almost identical increase in log CD4+ cell counts over time. At this point, the number of repeated measures per individual is quite small; hence we do not rely too heavily on this analysis. As more data are collected, the difference between the two groups unfolds. We see that the trial should be stopped at the fourth look. Hence we can conclude that low zidovudine is a more effective treatment than placebo in terms of slowing the loss of CD4+ cell counts over an 80-week period.

5.2 Design

Suppose that we were asked to design the foregoing trial from scratch. If we are willing to assume that Model (2) holds, then how can we convert the targeted maximum information number into a trial design (e.g., number of patients to enroll, number of repeated measurements, study length, etc.)? To make this conversion, we need to make some initial guesses for the nuisance parameters, and then either perform some theoretical calculations or conduct a simulation. Because theoretical computations are not always tractable, we have adopted the latter approach. In this example we use the results from fitting Model (2) at the aforementioned final analysis to help define some of the nuisance parameters. That is, we assume that $\xi = 5.756$, $\zeta = -.0021$, $\sigma^2 = .182$, and

$$\Sigma_D = \begin{bmatrix} .238960 & .003688 \\ .003688 & .000057 \end{bmatrix}.$$

Furthermore, we assume that subjects enroll in the study according to a Poisson process, have their CD4+ cell counts measured at fixed intervals, and are independently lost to follow-up according to an exponential distribution with mean 4.48 years (i.e., 20% drop out within the first year of the study). For simplicity, we accrue patients for the duration of the study provided that they have the potential for more than one CD4+ measurement. A more realistic scenario would consider the length of accrual as a design

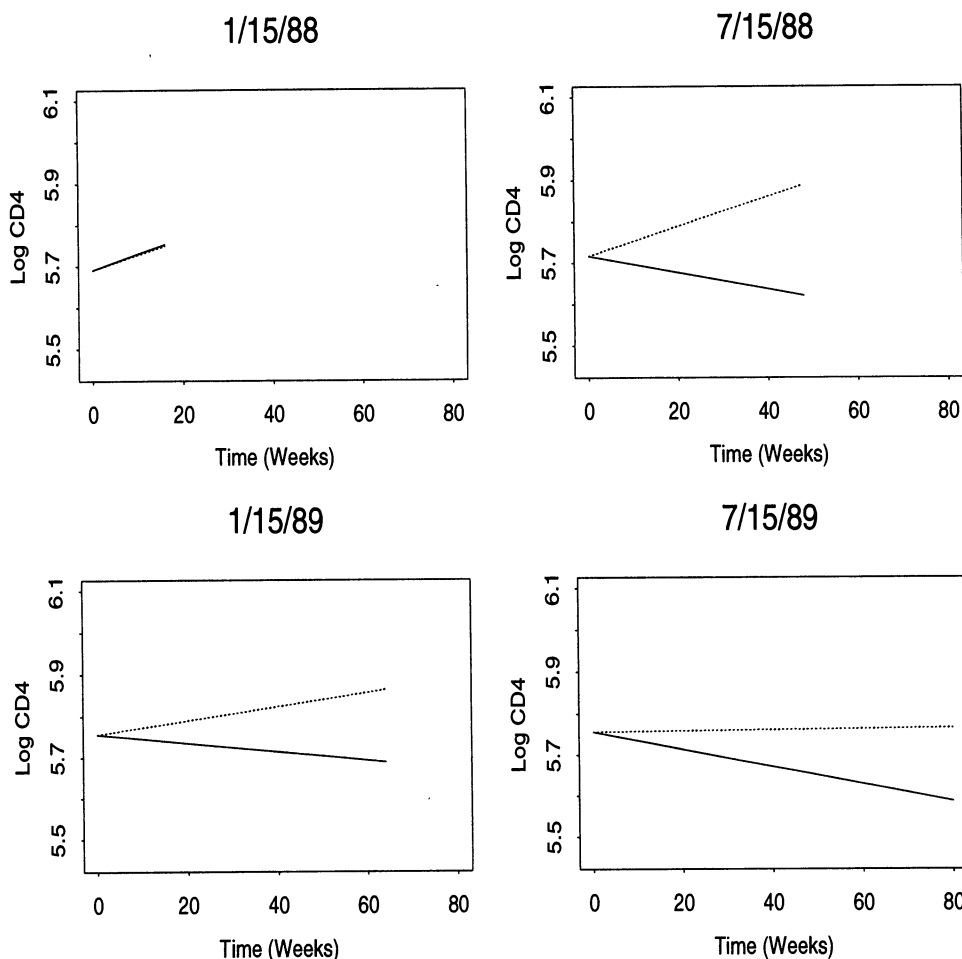


Figure 1. Predicted log CD4+ Cell Counts Over Time for Each of the Two Treatment Groups Based on Fitting Model (2) at Each of the Four Interim Analysis Times. The placebo and low-dose zidovudine groups are represented by the solid and dotted lines.

parameter. To figure out the trial design, we simulate data under the alternative hypothesis ($\beta = \log(1.5)/52$) and the foregoing assumptions for a variety of entry rates and interval lengths. For each combination of these parameters, we compute the average study length at which an estimate for the information at the end of the study is equal to the targeted maximum information. The information is estimated by one over the square of the standard error of the parameter estimate. Each of these averages is based on 25 simulations. We considered 10 different entry rates ranging from 5 to 50 subjects per week and six different interval lengths ranging from 4 to 24 weeks. The results are presented in Table 3. The designs in this table are all equivalent in terms of the outlined operating characteristics. Study planners may prefer certain designs for practical reasons, such as cost, number of patients available, or targeted study length.

6. SIMULATION STUDY

Consider the longitudinal study discussed in the previous section. Suppose that study planners anticipate that there are roughly 1,000 patients available to participate in the experiment. Using Table 3, they may decide to design a 77-week study in which 15 subjects are enrolled per week for the first 69 weeks and CD4+ cell measurements are taken at time of enrollment and every 8 weeks thereafter. Using this design and the initial guesses for the nuisance parameters, we simulate data under the null and alternative hypotheses and see how well our method performs in terms of the desired operating characteristics. We add the constraint that the data are monitored every 19 weeks provided that the information accrued is greater than 10% and the trial will be terminated at 100% information if it has not been stopped already. Running 5,000 simulations under each hypothesis,

Table 2. Summary of Interim Analyses for the ACTG019 Trial

j	Date of analysis	No. enrolled	Mean no. of counts collected	$\hat{\beta}_j^*(\mathbf{X}_{t_j, n})$	$ W_j^*(\mathbf{X}_{t_j, n}, \beta_0) $	π_j	b_j	Decision
1	1/15/88	378	1.94	-.00022	.0270	.0136	8.0000	Continue
2	7/15/88	513	3.22	.00559	2.0708	.1240	6.2579	Continue
3	1/15/89	773	3.60	.00273	2.2632	.6194	2.6229	Continue
4	7/15/89	881	3.93	.00226	2.2682	.9065	1.9799	Stop

Table 3. Average Study Length (Weeks),^a Average Number of Measurements per Subject,^b and Average Number of Subjects Enrolled^c for 60 Combinations of Entry Rates and Interval Lengths

Entry rate (per week)	Interval length (weeks)					
	4	8	12	16	20	24
5	99.70 ^a	117.61	130.19	137.56	147.62	150.23
	11.73 ^b	7.16	5.47	4.49	3.94	3.46
	478.96 ^c	549.60	598.08	609.40	626.24	631.04
10	75.99	89.70	99.73	103.67	109.96	113.87
	9.50	5.92	4.53	3.74	3.32	2.97
	721.92	821.00	869.88	868.92	892.88	907.08
15	65.77	76.47	84.35	89.33	92.36	96.99
	8.48	5.27	4.07	3.44	3.00	2.71
	936.00	1,032.24	1,077.48	1,091.32	1,079.52	1,099.40
20	59.63	70.30	75.76	80.34	83.53	86.08
	7.80	4.93	3.83	3.21	2.81	2.57
	1,117.12	1,252.76	1,268.80	1,285.20	1,273.68	1,237.24
25	55.28	64.56	70.06	74.03	77.42	79.63
	7.36	4.64	3.61	3.05	2.70	2.47
	1,281.76	1,415.80	1,449.12	1,463.04	1,432.56	1,387.44
30	52.19	60.30	65.56	68.86	72.32	74.83
	7.02	4.46	3.46	2.93	2.61	2.37
	1,461.44	1,567.24	1,608.00	1,590.28	1,581.92	1,542.04
35	49.56	57.62	62.37	65.75	68.49	71.85
	6.78	4.29	3.35	2.85	2.55	2.31
	1,593.96	1,739.04	1,760.08	1,734.36	1,702.64	1,673.68
40	47.69	55.51	59.50	63.29	65.17	68.21
	6.55	4.18	3.25	2.78	2.48	2.27
	1,751.08	1,893.92	1,901.92	1,887.64	1,808.80	1,754.56
45	46.07	52.73	57.22	60.30	62.88	64.48
	6.36	4.03	3.17	2.72	2.42	2.23
	1,885.32	2,025.60	2,044.80	1,990.72	1,930.76	1,821.00
50	44.27	50.92	55.31	57.87	61.14	62.83
	6.20	3.95	3.12	2.67	2.37	2.21
	2,026.04	2,142.44	2,169.84	2,092.36	2,069.36	1,933.32

we found that the type I error was .0512 and the power was .8076. These results are quite close to our targeted levels of .05 and .80.

The results of this simulation study are encouraging but are limited in the sense that in practice, the nuisance parameters are not known at the study design phase. However, as mentioned at the end of Section 4.1, it is possible to adaptively revise the study design as better estimates of the nuisance parameters become available at the successive interim analysis times. As future research, we intend to investigate the effect of this adaptive procedure on the operating characteristics of the study.

7. SUMMARY

The purpose of this article was to unify many of the results presented in the literature on group-sequential studies. Authors have shown that the distributions of many of the statistics used to analyze data arising from group-sequential time-to-event and longitudinal studies have a special structure. Namely, the asymptotic joint distribution of the time-sequential statistics is multivariate normal with an independent increments covariance structure. In Theorem 1 we demonstrated that this limiting distribution arises naturally when one efficiently estimates or tests a single parameter in a semiparametric or parametric model. In particular, the

joint distribution depends only on the statistical information at the interim analysis times.

Given this general structure, we then developed an information-based design and monitoring procedure in Section 4, followed by examples in Section 5. As the simulation study in Section 6 illustrated, the design procedure applied to a particular longitudinal study worked quite well in terms of preserving the desired operating characteristics.

Without any additional effort, the results of Theorem 1 can be generalized to the efficient testing of a multidimensional parameter in a semiparametric or parametric model. As future research, we intend to develop an information-based design and monitoring procedure for studies where the hypothesis under investigation involves multiple parameters. In addition, we recognize that our results hinge on asymptotic theory, and we would like to see how well our design and analysis procedure performs in smaller samples. Finally, all of the results in this article assume that we have correctly specified the model for the full-data process. We plan to investigate the feasibility of a procedure that sequentially checks model accuracy and allows the possibility of adaptively changing the model.

APPENDIX: PROOF OF THEOREM 1

First, note that $\hat{I}_{t_j}(\mathbf{X}_{t_j,n})^{1/2} Z_{t_j}^*(\mathbf{X}_{t_j,n}, \beta_0)$ is equal to

$\sum_{i=1}^n S_{t_j, \text{eff}}(X_{t_j, i})/\sqrt{n} + o_p(1)$ ($j = 1, \dots, J$). The asymptotic normality follows from a straightforward application of the multivariate central limit theorem, Slutsky's theorem, and contiguity. To establish the covariance structure, we consider two arbitrary time points, s and t , where $0 < s < t$. It is sufficient to establish that $E[S_{s, \text{eff}}(X_s)(S_{t, \text{eff}}(X_t) - S_{s, \text{eff}}(X_s))] = 0$. To do this, we write

$$\begin{aligned} & E[S_{s, \text{eff}}(X_s)(S_{t, \text{eff}}(X_t) - S_{s, \text{eff}}(X_s))] \\ &= E[S_{s, \text{eff}}(X_s)(S_{t, \beta}(X_t) - S_{s, \beta}(X_s))] \\ & \quad + E[S_{s, \text{eff}}(X_s)\Pi[S_{s, \beta}(X_s)|\Lambda_s(X_s)]] \\ & \quad - E[S_{s, \text{eff}}(X_s)\Pi[S_{t, \beta}(X_t)|\Lambda_t(X_t)]], \end{aligned}$$

and we consider each term of the right side. To show that the first term is 0, it suffices to show that $E[S_{t, \beta}(X_t) - S_{s, \beta}(X_s)|X_s] = 0$ or, equivalently, that $E[S_{t, \beta}(X_t)|X_s] = S_{s, \beta}(X_s)$. To see why this term is 0, it is helpful to think of the data at time s as "missing" data relative to the "full" data at time t . With this in mind, we can invoke proposition A.5.5 of BKRW (1993), which states that the score for a parameter in a missing-data model is the conditional expectation of the score for the parameter in a full-data model given the observed data. The second term is 0 because the efficient score for β at time s is orthogonal to every element of $\Lambda_s(X_s)$, which includes $\Pi[S_{s, \beta}(X_s)|\Lambda_s(X_s)]$. Let $\mathcal{H}_t(X_t)$ be a Hilbert space of random variables $H(X_t)$ with inner product $E[H_1(X_t)H_2(X_t)]$ and $E[H(X_t)^2] < \infty$. By lemma A.1 of Robins, Rotnitzky, and Zhao (1994), we know that $\Lambda_s(X_s)$ (i.e., the missing-data nuisance tangent space) is the closure of the set $\{\mathbf{g}(\mathbf{p}_t(H(X_t))): H(X_t) \in \mathcal{H}_t(X_t)\}$, where $\mathbf{p}_t(H(X_t)) = \Pi[H(X_t)|\Lambda(X_t)]$ is the projection of $H(X_t)$ onto $\Lambda_t(X_t)$ (i.e., the full-data nuisance tangent space) and $\mathbf{g}(\cdot) = E[\cdot|X_s]$. Now we wish to establish the relationship between $\Lambda_s(X_s)^\perp$ and $\Lambda_t(X_t)$. By theorem 1 of section 6.6 of Luenberger (1969), we know that $\Lambda_s(X_s)$ is also equal to the orthogonal complement of the null space of $\mathbf{p}_t^* \mathbf{g}^*$, where \mathbf{p}_t^* and \mathbf{g}^* are the adjoints of \mathbf{p}_t and \mathbf{g} . Because any null space is closed, we know that $\Lambda_s(X_s)^\perp$ is equal to the null space itself. Furthermore, \mathbf{p}_t is self-adjoint, because $E[(H(X_t) - \mathbf{p}_t(H(X_t)))\mathbf{p}_t(L(X_t))] = E[(L(X_t) - \mathbf{p}_t(L(X_t)))\mathbf{p}_t(H(X_t))] = 0$, which implies that $E[H(X_t)\mathbf{p}_t(L(X_t))] = E[\mathbf{p}_t(H(X_t))L(X_t)]$. In addition, \mathbf{g}^* is the identity operator, because $E[E[H(X_t)|X_s]A(X_s)] = E[E[H(X_t)A(X_s)|X_s]] = E[H(X_t)A(X_s)]$. Therefore, we can write $\Lambda_s(X_s)^\perp$ as the set $\{A(X_s): \mathbf{p}_t(A(X_s)) = 0\}$. This implies that $\Lambda_s(X_s)^\perp \subset \Lambda_t(X_t)^\perp$. Because $S_{s, \text{eff}}(X_s) \in \Lambda_s(X_s)^\perp$, it is also an element of $\Lambda_t(X_t)^\perp$ and we know that the third term is 0.

[Received February 1996. Revised February 1997.]

REFERENCES

- Andersen, P. K., Borgan, O., Gill, R. D., 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," *Journal of the Royal Statistical Society, Ser. A*, 132, 235-244.
- Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, Baltimore: Johns Hopkins University Press.
- Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305-324.
- Choi, S., Hall, W. J., and Schick, A. (1996), "Asymptotically Uniformly Most Powerful Tests in Parametric and Semiparametric Models," *The Annals of Statistics*, 24, 841-861.
- Gange, S. J., and DeMets, D. L. (1996), "Sequential Monitoring of Clinical Trials With Correlated Responses," *Biometrika*, 83, 157-167.
- Gu, M., and Lai, T. (1991), "Weak Convergence of Time-Sequential Censored Rank Statistics With Applications to Sequential Testing in Clinical Trials," *The Annals of Statistics*, 19, 1403-1433.
- Gu, M., and Ying, Z. (1995), "Group Sequential Methods for Survival Data Using Partial Score Processes With Covariate Adjustment," *Statistica Sinica*, 5, 793-804.
- Harrington, D. P., and Fleming, T. R. (1982), "A Class of Rank Test Procedures for Censored Survival Data," *Biometrika*, 69, 553-566.
- Jennison, C., and Turnbull, B. W. (1991), "Exact Calculations for Sequential t , χ^2 , and F Tests," *Biometrika*, 78, 133-141.
- (1997), "Group Sequential Analysis Incorporating Covariate Information," *Journal of the American Statistical Association*, 92, 1330-1341.
- Laird, N. M., and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963-974.
- Lan, G. K. K., and DeMets, D. L. (1983), "Discrete Sequential Boundaries for Clinical Trials," *Biometrika*, 70, 659-663.
- Lan, G. K. K., and Zucker, D. M. (1993), "Sequential Monitoring of Clinical Trials: The Role of Information and Brownian Motion," *Statistics in Medicine*, 12, 753-765.
- Lee, J. W., and DeMets, D. L. (1991), "Sequential Comparison of Changes With Repeated Measurements Data," *Journal of the American Statistical Association*, 86, 757-762.
- Lee, S. J., Kim, K., and Tsiatis, A. A. (1996), "Repeated Significance Testing in Longitudinal Clinical Trials," *Biometrika*, 83, 779-789.
- Liang, K., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13-22.
- Luenberger, D. G. (1969), *Optimization by Vector Space Methods*, New York: Wiley.
- Newey, W. K. (1990), "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99-135.
- Reboussin, D. M., Lan, K. K. G., and DeMets, D. L. (1992), "Group Sequential Testing of Longitudinal Data," Technical Report 72, University of Wisconsin.
- Robins, J. M., and Rotnitzky, A. (1992), "Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers," in *AIDS Epidemiology—Methodological Issues*, eds. N. Jewell, K. Dietz, and V. Farewell, Boston: Birkhauser, pp. 297-331.
- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), "Estimation of Regression Coefficients When Some Regressors are not Always Observed," *Journal of the American Statistical Association*, 89, 846-866.
- Scharfstein, D. O., and Tsiatis, A. A. (in press), "The Use of Simulation and Bootstrap in Information-Based Group Sequential Studies," *Statistics in Medicine*, 16.
- Sellke, T., and Siegmund, D. (1983), "Sequential Analysis of the Proportional Hazards Model," *Biometrika*, 70, 315-326.
- Slud, E. (1984), "Sequential Linear Rank Statistics for Two Sample Censored Survival Data," *The Annals of Statistics*, 12, 551-571.
- 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.
- (1982), "Repeated Significance Testing for a General Class of Statistics Used in Censored Survival Analysis," *Journal of the American Statistical Association*, 77, 855-861.
- Tsiatis, A. A., 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.
- Volberding, P. A., Lagakos, S. A., Koch, M. A., Pettinelli, C., Myers, M. W., Booth, D. K., Balfour, H. H., Reichman, R. C., Bartlett, J. A., Hirsch, M. S., Murphy, R. L., Hardy, W. D., Soeiro, R., Fischl, M. A., Bartlett, J. G., Merigan, T. C., Hyslop, N. E., Richman, D. D., Valentine, F. T., and Corey, L. (1990), "Zidovudine in Asymptomatic Human Immunodeficiency Virus Infection: A Controlled Trial in Persons With Fewer Than 500 CD4-Positive Cells per Cubic Millimeter," *New England Journal of Medicine*, 322, 941-949.
- Wang, S. K., and Tsiatis, A. A. (1987), "Approximately Optimal One-Parameter Boundaries for Group Sequential Trials," *Biometrics*, 43, 193-199.
- Wu, M. C., and Lan, G. K. K. (1992), "Sequential Monitoring for Comparison of Changes in a Response Variable in Clinical Studies," *Biometrics*, 48, 765-779.