BIOMETRICS 58, 89–97
March 2002

Monitoring the Rates of Composite Events with Censored Data in Phase II Clinical Trials

Ying Kuen Cheung

Department of Biostatistics, Mailman School of Public Health, Columbia University, 630 West 168th Street, New York, New York 10032, U.S.A.
email: cheung@biostat.columbia.edu

and

Peter F. Thall

Department of Biostatistics, Box 447, M. D. Anderson Cancer Center, University of Texas, 1515 Holcombe Boulevard, Houston, Texas 77030, U.S.A.

SUMMARY. In many phase II clinical trials, interim monitoring is based on the probability of a binary event, response, defined in terms of one or more time-to-event variables within a time period of fixed length. Such outcome-adaptive methods may require repeated interim suspension of accrual in order to follow each patient for the time period required to evaluate response. This may increase trial duration, and eligible patients arriving during such delays either must wait for accrual to reopen or be treated outside the trial. Alternatively, monitoring may be done continuously by ignoring censored data each time the stopping rule is applied, which wastes information. We propose an adaptive Bayesian method that eliminates these problems. At each patient's accrual time, an approximate posterior for the response probability based on all of the event-time data is used to compute an early stopping criterion. Application to a leukemia trial with a composite event shows that the method can reduce trial duration substantially while maintaining the reliability of interim decisions.

KEY WORDS: Approximate posterior; Competing risks; Dependent censoring; Historical data; Interim analyses; Mixture of beta distributions.

1. Introduction

For both ethical and practical reasons, most phase II clinical trials have provisions for stopping early if results are disappointing. Aside from maximum sample size, the method for outcome-adaptive interim decision making is the essential component of a phase II trial design. Various approaches to interim monitoring in phase II have been proposed. For binary outcomes, these include frequentist group-sequential test-based designs (Fleming, 1982; Simon, 1989) and Bayesian designs (Thall and Simon, 1994; Heitjan, 1997). Similarly, both frequentist methods (Bryant and Day, 1995; Conaway and Petroni, 1995) and Bayesian methods (Thall, Simon, and Estey, 1995; Thall and Sung, 1998) have been proposed for monitoring multiple events.

In this article, we focus on phase II trials where patient outcome is characterized as the binary indicator of a composite event defined in terms of one or more time-to-event variables. For example, suppose that the goal is to monitor the probability of surviving at least 6 months in a trial involving a rapidly fatal disease and that the trial will be conducted by accruing and treating successive cohorts of patients. To properly apply an outcome-adaptive interim decision rule that decides after each cohort whether to stop the trial or treat another cohort, apparently accrual must be suspended until all patients in the current cohort have been followed long enough to score their binary outcomes. A cohort size of one is safest but imposes an unrealistic trial duration, while larger cohort sizes yield a trial that is more feasible but less safe. A seemingly reasonable alternative is to conduct interim analyses based on data only from patients who have been followed completely. This is inefficient because it ignores essential information such as, e.g., the fact that a patient currently has survived 5 months. It also may reduce the trial's safety, especially if the accrual rate is high relative to the length of the evaluation period. For example, given an accrual rate of three patients per month, if accrual is not suspended, then 18 patients may be treated before even the first patient's 6-month outcome can be evaluated.

We propose a Bayesian sequential adaptive procedure for continuous monitoring that utilizes all current *information at* each interim decision time, including both censored and uncensored observations. An approximate posterior for the response probability is computed by replacing nuisance parameters in the likelihood with consistent estimates. Continuous

monitoring based on the approximate posterior (CMAP) is carried out by applying the stopping criterion of Thall and Simon (1994) at each new patient's accrual time. The main practical advantage of the method is that it maintains the reliability of the sequential decision procedure while eliminating the need to either suspend accrual or ignore censored observations.

We formulate the probability model and describe three basic cases in Section 2. The approximate posterior is derived in Section 3, and we define CMAP in Section 4. We apply the method to design an acute leukemia trial in Section 5, with extension to a randomized phase II trial in Section 6. Section 7 concludes with a discussion.

2. A Working Likelihood

Let B denote the event of interest, with $\theta = \text{pr}(B)$, and denote the event indicator by $\mathbf{1}(\cdot)$. Let $N(t)$ denote the current number of patients who have been accrued and treated at a given calendar time t during the trial, with $t = 0$ the trial starting time. If all the $\mathbf{1}(B_i)$'s were observed, then the current likelihood would be

$$\prod_{i=1}^{N(t)} \theta^{\mathbf{1}(B_i)}(1-\theta)^{1-\mathbf{1}(B_i)}. \tag{2.1}$$

However, among the $N(t)$ patients treated up to the current calendar time, if patient i has not been followed for his/her full observation period, then $\mathbf{1}(B_i)$ may not be known. Our strategy for developing a method to monitor θ is to first write down the likelihood of the data that are actually observed and construct a working likelihood by replacing nuisance parameters in the likelihood by estimates. We then apply an early stopping rule using the approximate posterior of θ based on the working likelihood.

We first introduce notation to reconcile calendar time, t, and an individual patient's time, s, where $s = 0$ is the start of the patient's treatment. Let U be the calendar time when the patient enters the trial. We will suppress i for simplicity when considering one patient. For any calendar time t, the patient's follow-up time is

$$C(t) = \begin{cases} 0 & \text{if } t < U \\ t-U & \text{if } U \le t < U+T \\ T & \text{if } t \ge U+T, \end{cases}$$

where T is the fixed length of time required to observe B. Let $A(s)$ be an event defined for $0 \le s \le T$ that is observed only if $s \le C(t)$, with $A(T) = B$. We define the patient's observable process in calendar time as

$$A^o(t) = A\{C(t)\}, \qquad t \ge 0$$

and denote $Y(t) = \mathbf{1}\{A^o(t)\}$. Thus, if the patient is followed for a period of length at least T, then it is known whether the outcome B has occurred; otherwise, only the partial information $Y(t)$ is available.

Case 1: Simple event. First consider the simplest case involving only one event-time variable. Let Z denote the patient's survival time and suppose that it is desired to monitor the probability of $B = \{Z \le T\}$, the event that the patient dies before T. In this case, $A(s) = \{Z \le s\}$ is the event that the patient does not survive a period of at least s for $0 \le s \le T$, and the patient's observable process is

$$A^o(t) = \begin{cases} A(0) = \{Z < 0\} & \text{if } t < U \\ A(t-U) = \{Z < t-U\} & \text{if } U \le t < U+T \\ A(T) = \{Z < T\} & \text{if } t \ge U+T. \end{cases}$$

In this case, $Y(t) = \mathbf{1}(Z < T) = \mathbf{1}(B)$ for $t \ge U+T$.

In general, the likelihood function of the data that are actually observed at t is

$$\prod_{i=1}^{N(t)} \left[\text{pr}\left\{A_i^o(t)\right\}\right]^{Y_i(t)} \left[1 - \text{pr}\left\{A_i^o(t)\right\}\right]^{1-Y_i(t)}.$$

Our approach is based on the probability decomposition

$$\text{pr}\left\{A_i^o(t)\right\} = \text{pr}\left\{A_i^o(t) \mid B_i\right\}\theta + \text{pr}\left\{A_i^o(t) \mid \bar{B}_i\right\}(1-\theta), \tag{2.2}$$

where $\bar{B}$ denotes the complement of B. Note that $\omega_{i1}(t) = \text{pr}\{A_i^o(t) \mid B_i\}$ is the probability of observing the observable event at t given that B_i would have occurred had follow-up of patient i been complete. Similarly, $\omega_{i2}(t) = \text{pr}\{A_i^o(t) \mid \bar{B}_i\}$ is the analogous probability given that B_i would not have occurred. Since the scientific focus is θ while $\omega_{i1}(t)$ and $\omega_{i2}(t)$ are nuisance parameters arising because some $\mathbf{1}(B_i)$'s are not available at t, we replace $\omega_{i1}(t)$ and $\omega_{i2}(t)$ with patient-specific estimates, $\hat{\omega}_{i1}(t)$ and $\hat{\omega}_2(t)$. This gives the working likelihood at time t,

$$L(\theta,t) = \prod_{i=1}^{N(t)} \{\hat{\omega}_{i1}(t)\theta + \hat{\omega}_{i2}(t)(1-\theta)\}^{Y_i(t)} \times \{1 - \hat{\omega}_{i1}(t)\theta - \hat{\omega}_{i2}(t)(1-\theta)\}^{1-Y_i(t)}. \tag{2.3}$$

While the estimators $\hat{\omega}_{i1}$ and $\hat{\omega}_{i2}$ may be constructed in several ways, in any case, we require that they respect the constraints

$$\hat{\omega}_{i1}(0) = 0, \qquad \hat{\omega}_{i1}(t) = 1; \qquad \hat{\omega}_{i2}(0) = 0, \qquad \hat{\omega}_{i2}(t) = 0 \tag{2.4}$$

for all i and $t \ge U_i + T$. These constraints ensure that $L(\theta,t)$ reduces to the binomial likelihood (2.1) when all patients are followed completely.

Returning to case 1, we set $\hat{\omega}_{i2}(t) \equiv 0$ because death by time $C_i(t)$ is impossible for $C_i(t) \le T$ given survival beyond T, i.e., $\omega_{i2}(t) = 0$. Consequently, in this case,

$$L(\theta,t) = \prod_{i=1}^{N(t)} \{\hat{\omega}_{i1}(t)\theta\}^{Y_i(t)} \{1 - \hat{\omega}_{i1}(t)\theta\}^{1-Y_i(t)},$$

where $1-\theta$ is the probability of response, which is survival beyond T. This is the weighted likelihood used by Cheung and Chappell (2000) in the context of dose finding using the continual reassessment method (O'Quigley, Pepe, and Fisher, 1990), with toxicity in place of death. The following two cases include more complicated situations where B is a composite event involving two or more time-to-event time variables.

Case 2: Composite event. Suppose there is a desirable response whose occurrence is subject to censoring by a terminal failure but not the other way around. An important example is a cancer chemotherapy trial where response is disease remission and terminal failure is death. Denote the times to response and terminal failure by X and Z, respectively. We are interested in the probability of response without failure by

time T, which is the composite event $B = \{X \leq T < Z\}$. In this case, the observable process is $A^o(t) = \{X \leq C(t) < Z\}$ and the conditional probabilities in (2.2) are

$$\begin{aligned}\omega_{i1}(t) &= \frac{\text{pr}\{X_i \leq C_i(t), Z_i > T\}}{\text{pr}(B_i)} \\ &= \text{pr}\{X_i \leq C_i(t) \mid X_i \leq T < Z_i\}\end{aligned}$$

and

$$\begin{aligned}\omega_{i2}(t) &= \frac{\text{pr}\{X_i \leq C_i(t) < Z_i \leq T\}}{[1 - \text{pr}(B_i)]} \\ &= \rho_i \text{pr}\{X_i \leq C_i(t) \mid X_i \leq T\} \\ &\quad \times [1 - \text{pr}\{Z_i \leq C_i(t) \mid Z_i \leq T\}],\end{aligned}$$

where

$$\rho_i = \frac{\text{pr}\{X_i \leq C_i(t) < Z_i < T\}\,\text{pr}\{X_i \leq T\}\,\text{pr}\{Z_i \leq T\}}{\text{pr}\{X_i \leq C_i(t)\}\text{pr}\{C_i(t) < Z_i < T\}\,[1 - \text{pr}(B_i)]}.$$

When X_i and Z_i are independent, ρ_i reduces to $\text{pr}(X_i \leq T)\text{pr}(Z_i \leq T \mid \bar{B}_i)$. We will use these decompositions to develop estimators of $\omega_{i1}(t)$ and $\omega_{i2}(t)$.

Case 2 accommodates settings with multiple terminal failure types by defining Z to be the minimum of the failure times. This arises, e.g., in cancer chemotherapy trials where disease remission, toxicity, and death are all monitored and B is the event that the patient is alive with disease in remission and without severe toxicity at time T. In this case, Z is the minimum of the times to toxicity and death.

Case 3: Composite event with competing risks. An additional complication arises when the desirable response and a nonfatal failure are competing risks in that the occurrence of one censors the other. This case includes our illustrative application, a leukemia chemotherapy trial where a patient either achieves complete remission (CR) or his/her leukemia is declared resistant to the treatment. The desired event is that the patient achieves CR before T and before being declared resistant and that the patient is still alive at T. In general, denoting the nonterminal failure time by $\tilde{X}$, the composite event of interest is $B = \{X \leq T \wedge \tilde{X}, Z > T\}$. The observable process is $A^o(t) = \{X \leq C(t) \wedge \tilde{X}, Z > C(t)\}$, and the decompositions of the two conditional probabilities are

$$\begin{aligned}\omega_{i1}(t) &= \frac{\text{pr}(X_i \leq C_i(t) \wedge \tilde{X}_i, Z_i > T)}{\text{pr}(B_i)} \\ &= \text{pr}(X_i \leq C_i(t) \wedge \tilde{X}_i \mid X_i \leq T \wedge \tilde{X}_i, Z_i > T)\end{aligned}$$

and

$$\begin{aligned}\omega_{i2}(t) &= \frac{\text{pr}(X_i \leq C_i(t) \wedge \tilde{X}_i, C_i(t) < Z_i \leq T)}{[1 - \text{pr}(B_i)]} \\ &= \rho_i \text{pr}(X_i \leq C_i(t) \wedge \tilde{X}_i \mid X_i \leq T \wedge \tilde{X}_i) \\ &\quad \times [1 - \text{pr}(Z_i \leq C_i(t) \mid Z_i \leq T)],\end{aligned}$$

where

$$\begin{aligned}\rho_i = &\Big[\text{pr}(X_i \leq C_i(t) \wedge \tilde{X}_i, C_i(t) < Z_i < T) \\ &\times \text{pr}(X_i \leq T \wedge \tilde{X}_i)\text{pr}(Z_i \leq T)\Big] \\ \div &\Big[\text{pr}\{X_i \leq C_i(t) \wedge \tilde{X}_i\} \\ &\times \text{pr}\{C_i(t) < Z_i < T\}\,[1 - \text{pr}(B_i)]\Big].\end{aligned}$$

Similar to case 2, ρ_i reduces to $\text{pr}(X_i \leq T \wedge \tilde{X}_i)\text{pr}(Z_i \leq T \mid \bar{B}_i)$ if Z_i and $(X_i, \tilde{X}_i)$ are independent.

THEOREM 1: *Suppose the true event probability θ_0 is bounded away from zero and one. If $\hat{\omega}_{ij}(t)$ converges almost surely to $\omega_{ij}(t)$ for $j = 1, 2$ and all i as $N(t) \to \infty$, then $\hat{\theta} = \arg\max L(\theta, t)$ is strongly consistent for θ_0.*

A proof is given in the Appendix. Theorem 1, coupled with log concavity of $L(\theta, t)$ in θ for given t, suggests that the likelihood, and thus the posterior density, will tend to concentrate near the true value θ_0.

To compute the working likelihood, we will use the following family of estimators for $\omega_{i1}(t)$ and $\omega_{i2}(t)$. Let $m_{c1}(t)$ and $m_{c2}(t)$ denote the respective numbers of patients for whom B_i and $\bar{B}_i$ have been observed by calendar time t. We define each estimator to be a weighted average of an empirical and a prior component as follows:

$$\begin{aligned}\hat{\omega}_{i1}(t) = &\frac{m_{c1}(t)}{m_{c1}(t) + m_0} \frac{\sum_j \mathbf{1}\{C_j(t) \geq T, A_j(C_i(t)) \cap B_j\}}{m_{c1}(t)} \\ &+ \frac{m_0}{m_{c1}(t) + m_0} f_{i1}(t) \qquad (2.5)\end{aligned}$$

and

$$\begin{aligned}\hat{\omega}_{i2}(t) = &\frac{m_{c2}(t)}{m_{c2}(t) + m_0} \frac{\sum_j \mathbf{1}\{C_j(t) \geq T, A_j(C_i(t)) \cap \bar{B}_j\}}{m_{c2}(t)} \\ &+ \frac{m_0}{m_{c2}(t) + m_0} f_{i2}(t), \qquad (2.6)\end{aligned}$$

with the convention that $0/0 \equiv 1$. For assumed prior parameters $\gamma > 0$ and $0 < \rho \leq 1$, the prior components of $\hat{\omega}_{i1}(t)$ and $\hat{\omega}_{i2}(t)$ in (2.5) and (2.6) are

$$f_{i1}(t) = \left\{\frac{C_i(t)}{T}\right\}^\gamma$$

and

$$f_{i2}(t) = \rho\left\{\frac{C_i(t)}{T}\right\}^\gamma \left[1 - \left\{\frac{C_i(t)}{T}\right\}^\gamma\right], \qquad (2.7)$$

with m_0 playing the role of an assumed prior sample size. We will give $f_{i1}(t)$ and $f_{i2}(t)$ small weights by setting $m_0 = 1$. For $\gamma = 1$, these correspond to uniform priors on the event times in the observation period $[0, T]$, whereas $\gamma < 1$ and $\gamma > 1$ correspond to the prior beliefs that the events are more likely earlier or later, respectively, in the interval.

The prior components f_{i1} and f_{i2} play a crucial role early in the trial by stabilizing the estimates since the empirical components are highly variable when there are few observations. As the sample grows, the empirical components take over in that, under independent censoring, $\hat{\omega}_{i1}(t)$ and $\hat{\omega}_{i2}(t)$ are consistent for $\omega_{i1}(t)$ and $\omega_{i2}(t)$. Numerically, ρ_i can be greater than one without violating the constraints (2.4). On the other hand, $f_{i2} \leq f_{i1}$ uniformly when $\rho \leq 1$. This inequality is a reasonable constraint because of how A_i and B_i are defined. Other forms for the functions f_{i1} and f_{i2} may be used, provided they respect the constraints (2.4). If prior knowledge is available about the event-time distributions, then informative choices for f_{i1} and f_{i2} may be more appropriate than (2.7). Similarly, values of $m_0 > 1$ may be used. However, as shown by the simulations reported below, the above family

Table 1
Likelihood contributions of an individual patient, given $(Y_i, \hat{\omega}_{i1}, \hat{\omega}_{i2})$

Weight functions	$Y_i = 0$	$Y_i = 1$
$\hat{\omega}_{i2} = \hat{\omega}_{i1} = 0$	1	Impossible
$\hat{\omega}_{i2} = 0, \hat{\omega}_{i1} = 1$	$1-\theta$	θ
$0 = \hat{\omega}_{i2} < \hat{\omega}_{i1} < 1$	$1-\hat{\omega}_{i1}\theta$	θ
$0 = \hat{\omega}_{i1} < \hat{\omega}_{i2} < 1$	$\theta + \frac{1-\hat{\omega}_{i2}}{\hat{\omega}_{i2}}$	$1-\theta$
$0 < \hat{\omega}_{i1} < \hat{\omega}_{i2} < 1$	$\theta + \frac{1-\hat{\omega}_{i2}}{\hat{\omega}_{i2}-\hat{\omega}_{i1}}$	$1-\theta\frac{\hat{\omega}_{i2}-\hat{\omega}_{i1}}{\hat{\omega}_{i2}}$
$0 < \hat{\omega}_{i2} = \hat{\omega}_{i1} < 1$	1	1
$0 < \hat{\omega}_{i2} < \hat{\omega}_{i1} < 1$	$1-\theta\frac{\hat{\omega}_{i1}-\hat{\omega}_{i2}}{1-\hat{\omega}_{i2}}$	$\theta + \frac{\hat{\omega}_{i2}}{\hat{\omega}_{i1}-\hat{\omega}_{i2}}$
$0 < \hat{\omega}_{i2} < \hat{\omega}_{i1} = 1$	$1-\theta$	$\theta + \frac{\hat{\omega}_{i2}}{1-\hat{\omega}_{i2}}$

with $m_0 = 1$ and f_{ij}'s given by (2.7) appears to be appropriate for most cases.

3. The Approximate Posterior of θ

In this section, we derive the approximate posterior of θ based on the working likelihood. Because this utilizes all of the data available at any calendar time, an adaptive safety monitoring rule based on θ may be applied at the time each new patient is accrued, thus obviating the need to ever suspend accrual.

The following computations show that, if θ follows a beta prior, i.e., the p.d.f. with parameters a and b is

$$\text{beta}(\theta; a, b) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}, \qquad 0 < \theta < 1,$$

where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ and $\Gamma(\cdot)$ is the gamma function, then its approximate posterior is a weighted average of beta distributions, with the weights explicit functions of $\hat{\omega}_{i1}(t)$ and $\hat{\omega}_{i2}(t)$. Table 1 gives the possible forms of the contribution of the ith patient to the working likelihood (2.3). Each entry of Table 1 may be expressed as one of the two general forms (i) $\theta + \alpha_i$ for some $\alpha_i \geq 0$ or (ii) $1 - \beta_i\theta$ for some $0 < \beta_i \leq 1$, where α_i and β_i are determined by $\hat{\omega}_{i1}(t)$ and $\hat{\omega}_{i2}(t)$.

Let n_1, n_2, n_3, n_4 be the numbers of patients whose contributions to the working likelihood take the respective forms $\theta + \alpha_i$ for $\alpha_i > 0$, $1 - \beta_i\theta$ for $0 < \beta_i < 1$, θ, and $1 - \theta$. Thus, $\Sigma_{l=1}^{4} n_l = N(t)$. For brevity, we suppress the dependence of the α_i's, β_i's, and n_l's on t. After appropriately reindexing patients, the working likelihood can be written as

$$L(\theta, t) = \prod_{i=1}^{n_1}(\theta + \alpha_i) \prod_{i=n_1+1}^{n_1+n_2} (1 - \beta_i\theta)\theta^{n_3}(1-\theta)^{n_4}. \quad (3.1)$$

The first product in (3.1) may be written as the sum

$$\prod_{i=1}^{n_1}(\theta + \alpha_i) = \sum_{j=0}^{n_1} \Omega_j \theta^{n_1 - j},$$

where

$$\Omega_0 \equiv 1 \quad \text{and} \quad \Omega_j = \sum_{\mathcal{A}_j} \alpha_{i_1} \cdots \alpha_{i_j}, \qquad j > 0,$$

and $\mathcal{A}_j$ is the set of ${}_{n_1}C_j$ choices of j distinct integers from $\{1, \ldots, n_1\}$. Similarly,

$$\prod_{i=n_1+1}^{n_1+n_2} (1 - \beta_i\theta) = \prod_{i=n_1+1}^{n_1+n_2} \{(1-\theta) + \theta(1-\beta_i)\}$$
$$= \sum_{k=0}^{n_2} \Psi_k \theta^k (1-\theta)^{n_2-k},$$

where

$$\Psi = 1 \quad \text{and} \quad \Psi_k = \sum_{\mathcal{B}_k} (1-\beta_{i_1}) \cdots (1-\beta_{i_k})$$

and $\mathcal{B}_k$ is the set of ${}_{n_2}C_k$ choices of k distinct integers from $\{n_1 + 1, \ldots, n_1 + n_2\}$. Thus, the working likelihood may be written

$$L(\theta, t) = \sum_{j=0}^{n_1} \sum_{k=0}^{n_2} \Omega_j \Psi_k \theta^{n_1+n_3+k-j}(1-\theta)^{n_2+n_4-k}.$$

Straightforward computations now yield the following approximate posterior of θ.

THEOREM 2: *If θ follows a beta(a,b) prior, then the approximate posterior probability density of θ given the data observed at time t is*

$$p(\theta \mid \mathbf{Y}, \hat{\omega}) = \sum_{j=0}^{n_1} \sum_{k=0}^{n_2} \left(\frac{\Phi_{jk}}{\sum_{r,s} \Phi_{rs}} \right) \text{beta}(\theta; a_{jk}, b_k), \quad (3.2)$$

where $a_{jk} = a + n_1 + n_3 + k - j$, $b_k = b + n_2 + n_4 - k$, and $\Phi_{jk} = \Omega_j \Psi_k B(a_{jk}, b_k)$.

4. Continuous Monitoring Based on the Approximate Posterior

In this section, we apply the above formulation to develop a practical methodology for interim monitoring based on a composite binary event. Consider a single-arm phase II clinical trial of an experimental treatment in which the primary question is whether it improves upon a standard therapy in terms of the probability θ of the event B. We assume without loss of generality that B is desirable since an adverse event, such as toxicity or death, is accommodated by dealing with its complement.

We define our proposed method in the context of the phase II monitoring method given by Thall and Simon (1994, hereafter TS). To reflect the typical phase II setting, the TS method uses an informative beta(a_S, b_S) prior on the event rate of the standard therapy θ_S based on historical data or clinical experience and a relatively noninformative beta(a_E, b_E) prior on the experimental treatment's rate θ_E. In a typical application, the amount of information $a_S + b_S$ in the prior on θ_S reflects the number of historical patients treated with the standard therapy, whereas $a_E + b_E$ is usually set equal to one or two. The TS method terminates the trial at an interim analysis if, for some targeted improvement $\delta > 0$ and fixed lower probability cutoff p_L,

$$\text{pr}(\theta_E > \theta_S + \delta \mid \text{data}) \leq p_L. \quad (4.1)$$

This rule is applied after each successive cohort of a predetermined size c, up to a maximum of N patients. In practice, the

parameters p_L, N, and c may be calibrated to obtain a design with good operating characteristics (OCs), whereas the minimal improvement δ is elicited from the physicians. The TS method formalizes the notions that there is uncertainty about both θ_E and θ_S and that safety monitoring in single-arm phase II trials is inherently comparative.

In trials where it is appropriate to stop early if there is evidence that the experimental treatment is superior, TS use the additional stopping criterion

$$\text{pr}(\theta_E > \theta_S + \delta \mid \text{data}) \geq p_U, \tag{4.2}$$

where p_U is a fixed upper probability cutoff. This criterion may not be used when it is clinically desirable to continue using a promising treatment, although time, cost, drug availability, and the need to publish new medical results expeditiously are also relevant. For simplicity, we focus on trials that use only the safety stopping criterion (4.1). Extension to trials using (4.2) is straightforward.

Application of the TS design is limited by the logistical problems described in the Introduction. This is the case with any clinical trial design having interim decisions based on the probability of a binary event defined in terms of one or more time-to-event variables in a time period of nontrivial length T. The TS method requires all current patients to be followed completely so that the posterior distribution of θ_E can be calculated and the stopping criterion (4.1) evaluated before new patients are accrued. For values of T that are large relative to the patient accrual rate, this may be unrealistic. Practical compromises are to apply (4.1) after successive patient cohorts of size $c > 1$, which we denote by TS(c), or to monitor continuously using only the complete data from patients who have been followed the entire time period T while ignoring censored observations, which we denote by TSCD.

We define continuous monitoring based on the approximate posterior (CMAP) of θ to be the TS method with the stopping criterion (4.1) applied each time a new patient is accrued, using (3.2) computed from the working likelihood (2.3). Assuming a beta(a_E, b_E) prior on θ_E, by Theorem 2, the approximate posterior of θ_E is determined by a_E, b_E and the estimates $\{\hat{\omega}_{i1}(t), \hat{\omega}_{i2}(t), i = 1, \ldots, N(t)\}$. In turn, the $\hat{\omega}_{ij}(t)$'s are determined by $\{f_{i1}(t), f_{i2}(t), i = 1, \ldots, N(t)\}$, which are defined in terms of ρ and γ. Thus, the CMAP design parameters are $\boldsymbol{\Delta} = (a_S, b_S, a_E, b_E, \delta, p_L, N, \rho, \gamma)$. Given N and δ, a value of γ reflecting prior knowledge about the event-time distributions on $[0, T]$ and given the beta prior parameters (a_S, b_S, a_E, b_E), the values of ρ and p_L may be chosen to obtain a design with desirable OCs. This may be done by first assuming $c = 1$ and obtaining $(a_S, b_S, a_E, b_E, \delta, p_L, N)$ according to the guidelines in TS (1994). For randomized phase II trials, the number of experimental treatments also plays a role at this stage and can be studied along with p_L and N, as in Thall and Sung (1998). Given these parameters, exploring ρ = 0.50, 1.00, or 2.00 should be adequate. If prior knowledge about γ is not available, then studying γ = 1.0 or 1.5 should suffice. To ensure that this process is reliable, the simulation scenarios should be formulated to provide a reasonable representation of what may be observed in the trial.

5. Application to an Acute Leukemia Trial

5.1 *Historical Data*

We now illustrate CMAP and study its properties in the context of a phase II trial of an experimental treatment for patients with newly diagnosed acute myelogenous leukemia (AML) or myelodysplastic syndromes (MDS). The entry criteria include the prognostically unfavorable $-5/-7$ cytogenetic abnormality characterized by loss of portions of the 5th or 7th chromosomes. Patient outcome corresponds to case 3 as described in Section 2. Denoting the respective times to complete remission (CR), being declared resistant, and death by X, $\tilde{X}$, and Z as before, response is the composite event that the patient is alive and in CR at day 90, $B = \{X < 90 \wedge \tilde{X}, 90 < Z\}$.

Among 335 AML/MDS patients with $-5/-7$ treated at the M. D. Anderson Cancer Center between 1990 and 1999, 154 (46%) achieved CR. Among these 154, 10 died before day 90, so the 90-day response rate was 144/335 (43%). As the first step in developing a design, we analyze the historical data by fitting the parametric competing risks model proposed by Shen and Thall (1998). Let Z_0 denote the time to death without either antecedent nonfatal event, R the residual survival time subsequent to CR, and $\tilde{R}$ the residual survival time subsequent to the patient's disease being declared resistant. Thus, $X, \tilde{X}, R, \tilde{R}$, and Z_0 are latent variables and the overall survival time Z takes one of the following three possible forms:

$$Z = \begin{cases} Z_0 & \text{if } Z_0 < X \wedge \tilde{X} \\ X + R & \text{if } X < Z_0 \wedge \tilde{X} \\ \tilde{X} + \tilde{R} & \text{if } \tilde{X} < Z_0 \wedge X. \end{cases}$$

Because only one of the three variables (X, R), $(\tilde{X}, \tilde{R})$, and Z_0 is observed on each patient, we assume that they are independent since any probability model including parameters characterizing their association is not identifiable (Tsiatis, 1975; Prentice et al., 1978). Marginally, each of the five variables $X, \tilde{X}, R, \tilde{R}$, and Z_0 is assumed to follow a three-parameter generalized odds rate model (Dabrowska and Doksum, 1988), characterized by the survival function

$$\bar{F}(t; \lambda, \phi, \zeta) = \{1 + \zeta(t/\lambda)^{\phi}\}^{-1/\zeta}, \qquad \lambda > 0,\ \phi > 0,\ \zeta > 0,$$

and $F = 1 - \bar{F}$ is the cumulative distribution function (c.d.f.). Denote the marginal c.d.f.'s of X and R by F_X and F_R. The association between X and R is modeled by the bivariate c.d.f.,

$$F_{X,R}(x, r) = F_X(x)F_R(r)\{1 + \alpha \bar{F}_X(x)\bar{F}_R(r)\}$$

for $x, r > 0$ and $-1 < \alpha < 1$, with the same functional form for the bivariate c.d.f. of the pair $(\tilde{X}, \tilde{R})$. Because this 17-parameter model accounts for the manner in which B depends on the event times, θ is a complex function of the model parameters. The maximum likelihood estimates (MLEs) of the model parameters based on the historical data are summarized in Table 2. The model-based MLE of θ_S is 0.44, which is very close to the simple binomial estimate 0.43.

5.2 *Simulation Study*

To compare CMAP to the TS method, we simulated the leukemia trial 2000 times using each design under a range

Table 2
Maximum likelihood estimates (with standard error) of the competing risks model for the historical data

	Death without CR or resistance, Z_0	Complete remission		Resistance	
		X	R	$\tilde{X}$	$\tilde{R}$
$\hat{\alpha}$		−0.200 (0.412)		0.185 (0.291)	
$\hat{\phi}$	1.222 (0.139)	19.08 (4.008)	2.183 (0.386)	3.681 (1.005)	1.768 (0.309)
$\hat{\lambda}$	100.9 (16.55)	27.18 (0.855)	226.5 (34.44)	78.82 (13.60)	83.85 (11.78)
$\hat{\zeta}$	0.416 (0.687)	20.75 (4.901)	1.557 (0.505)	1.216 (1.362)	1.475 (0.524)

of clinical scenarios. To conserve space, we summarize the results under the four scenarios described in Table 3. In all cases, we assumed *a priori* that $\theta_S \sim$ beta(145, 192) and $\theta_E \sim$ beta(0.86, 1.14) and applied the early stopping rule (4.1) with $\delta = 0.15$, $p_L = 0.05$, minimum sample size 10, and maximum sample size 60. Patient accrual was simulated as a Poisson process with rate five per month. Patients' event times were generated under the competing risks model described above. For the first two scenarios, we constrained the model parameters to give fixed response probability 0.44, the MLE of θ based on the historical data. The event-time parameters were calibrated so that events occurred within the 90-day observation interval on average either the same as seen historically (scenario 1) or later in the interval (scenario 2). Scenarios 3 and 4 have fixed response probability 0.59, reflecting the targeted 0.15 increase in θ, and parameters were calibrated so that the experimental treatment improves all three aspects of response, with shorter median time to CR and longer median times to death and resistance (scenario 3) or a modest improvement in CR but a large improvement in survival (scenario 4).

Figure 1 displays the probability of rejecting the experimental treatment using TS(1), TS(5), TSCD, or CMAP with six values of ρ and $\gamma = 1.0$. Each plotted value is the mean from the 2000 simulated trials. Overall, the rejection probabilities of all methods studied do not differ substantively, although TS(5) and TSCD have slightly lower rejection probabilities in all cases.

Figure 2 summarizes the trial duration distributions of the TS and CMAP designs. The reduction in trial duration provided by CMAP compared with TS(1) and TS(5) is quite striking under all four scenarios. By avoiding suspension of accrual, CMAP reduces median trial duration by 30–40% compared with TS(5). The differences are so large that a trial considered not feasible using TS(1) or TS(5) might be quite feasible using CMAP. The graphs also reveal that the variation in trial duration is much smaller using CMAP, which would allow investigators to budget time more precisely when planning the trial. The durations under TSCD are only slightly larger than with CMAP, as expected since both methods monitor continuously.

Figure 3 illustrates the achieved sample-size distributions of the designs under the two null scenarios. On average, CMAP requires slightly more patients than TS(1) or TS(5) in the null case, while TSCD requires substantially more patients than CMAP. This seems to be the price that is paid by the fact that TSCD ignores censored data. We also ran simulations for CMAP woth $\gamma = 1.5$, which gives similar OCs to CMAP with $\gamma = 1.0$, and hence the results are not displayed here. CMAP with $\gamma = 1.5$ requires slightly more patients than with $\gamma = 1.0$, which may be due to the fact that larger γ produces smaller prior values of $\hat{\omega}_{i1}(t)$, and thus more patients are required to stop early. For this reason, it may be preferable to apply CMAP with $\gamma = 1.0$. With fixed $\theta_E = 0.59$ under scenarios 3 or 4, TS(1) and TS(5) turn away on average 37 and 51 patients, respectively, whereas CMAP, by definition, does not turn away any patients. Since the sample size is very likely to be 60 under all designs when $\theta_E = 0.59$, in practice, the TS(1) and TS(5) designs would require on average 97–111 eligible patients in order to obtain 60 for the trial.

In summary, compared with TS(1) and TS(5), CMAP has the advantages of a much shorter trial duration and the fact that no eligible patients are turned away from the trial; hence, the actual number of patients required is much smaller. These advantages should greatly outweigh the slight increase in sample size with CMAP in the null cases. Compared with TSCD,

Table 3
Fixed values of the parameters and the medians (in days) of the competing risks model used in the simulations. Only parameters that differ from the MLEs obtained from the historical data are given.

		Z_0		X		$\tilde{X}$	
Scenario	θ_E	λ	med	λ	med	λ	med
1. Historical	0.44	100.9	84	27.18	49	78.82	81
2. Later events	0.44	121.0	101	32.00	58	97.00	99
3. Overall improvement	0.59	121.0	101	19.13	35	97.00	99
4. Improved survival	0.59	161.0	135	21.83	40	97.00	99

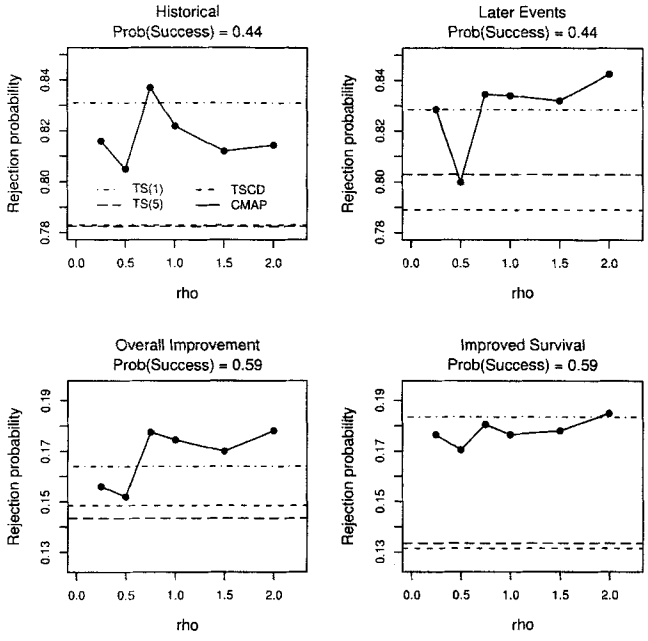

Figure 1. Probabilities of rejecting the experimental treatment using CMAP with $\rho = 0.25, 0.5, 0.75, 1.0, 1.5, 2.0$ and $\gamma = 1.0$ versus the TS designs.

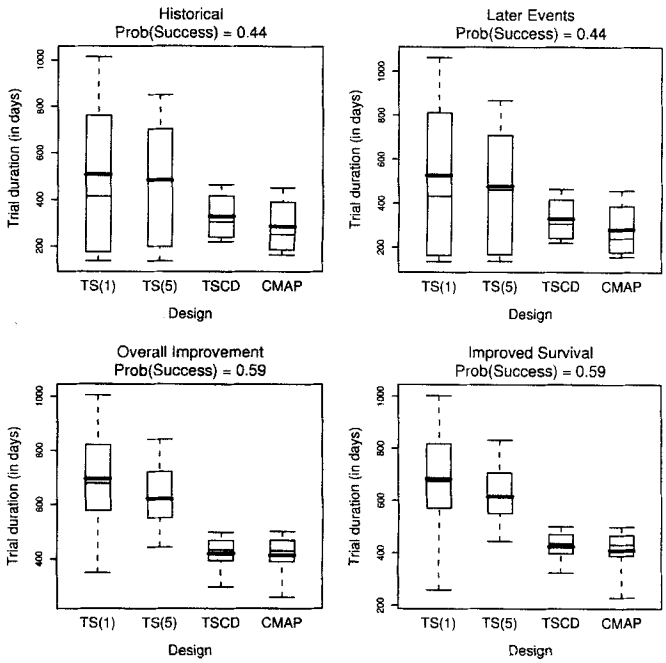

Figure 2. Trial duration of the TS designs and CMAP with $\rho = 0.5$ and $\gamma = 1.0$. Each box spans the interquartile range, the darker line inside is the mean, and the lighter is the median. The whiskers extend to the 10th and 90th percentiles.

CMAP has a slightly shorter duration in any case but a substantively smaller null sample size. The following section shows that these differences may be more substantial in a randomized selection trial.

6. Randomized Phase II Trials

We next consider randomized phase II trials where the goal is to select one or more of m experimental treatments, $\mathrm{E}_1, \ldots, \mathrm{E}_m$, for future evaluation. In such settings, it is still appropriate to monitor the safety of each E_k by comparison to historical data, precisely as if it were the only arm of a conventional phase II trial. To apply CMAP in this more general context, there are now m pairs of estimators, $\{\hat{\omega}_{i1k}(t), \hat{\omega}_{i2k}(t), k = 1, \ldots, m\}$, since the m treatments may have different time-to-event variable distributions. The working likelihood is now the product of m components, each taking the form (2.3).

As an illustration, we extend the leukemia trial example to include $m = 3$ experimental treatments, with response probabilities $\{\theta_{\mathrm{E}_1}, \theta_{\mathrm{E}_2}, \theta_{\mathrm{E}_3}\}$. The same priors on θ_{S} and the θ_{E_j}'s as before are assumed. In addition to applying the early stopping criterion (4.1) to each arm, any E_k is dropped if it is inferior to the others:

$$\Pr\left(\theta_{\mathrm{E}_k} < \max_{k' \neq k} \theta_{\mathrm{E}_{k'}} \mid \text{data}\right) > 0.90.$$

If one or more arms are dropped, then the remaining patients, up to a maximum of $N = 90$, are randomized fairly to the remaining arms. The entire trial is terminated if all three arms are dropped. We evaluate the designs in the two situations where the treatments' event-time distributions correspond to the three-treatment scenarios $(1, 1, 2)$ and $(1, 2, 3)$ in Table 3. Under scenario $(1, 1, 2)$, it is best to terminate all three arms since each has null success rate 0.44. Under scenario $(1, 2, 3)$, the success rates are $(0.44, 0.44, 0.59)$, so it is most desirable to select E_3 and drop both E_1 and E_2.

Table 4 summarizes OCs of this three-arm randomized trial under the TS designs and CMAP with $\rho = 0.5$ and $\gamma = 1.0$; all designs use $p_{\mathrm{L}} = 0.10$. As in the single-arm case, compared with CMAP, the TS(1) and TC(5) designs greatly inflate trial duration and TSCD inflates both duration and sample size. Recall that, in general, TSCD is less likely to reject the treatment in the single-arm case. This deficiency is magnified under scenario $(1, 1, 2)$ of the randomized trial, where TSCD greatly inflates the probability of incorrectly selecting one of the treatments. Another undesirable effect is seen under sce-

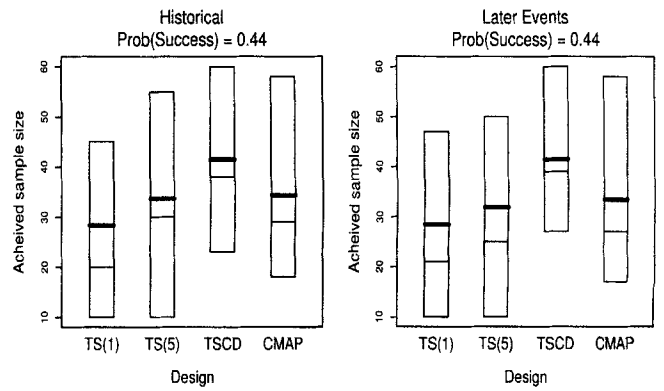

Figure 3. Achieved sample sizes of the TS designs and CMAP with $\rho = 0.5$ and $\gamma = 1.0$ under the two null scenarios. Each box spans the interquartile range, the darker line inside is the mean, and the lighter is the median.

Table 4
Selection probabilities, median achieved sample size, and median trial duration (in days) for the three-arm randomized trial under the TS designs and under CMAP with $\rho = 0.5$ and $\gamma = 1.0$. Correct selection probabilities are given in boldface.

		Selection probabilities				Sample size			
Scenario	Design	E_1	E_2	E_3	None	E_1	E_2	E_3	Duration
(1, 1, 2)	TS(1)	0.09	0.09	0.09	**0.73**	13	13	13	879
	TS(5)	0.13	0.13	0.13	**0.62**	15	20	15	749
	TSCD	0.15	0.15	0.16	**0.54**	23	23	23	606
	CMAP	0.08	0.10	0.08	**0.74**	14	15	13	481
(1, 2, 3)	TS(1)	0.05	0.06	**0.65**	0.24	11	11	51	950
	TS(5)	0.06	0.06	**0.71**	0.17	15	15	50	821
	TSCD	0.07	0.08	**0.69**	0.15	20	21	40	632
	CMAP	0.04	0.05	**0.63**	0.28	13	12	47	597

nario (1, 2, 3), where, compared with CMAP, TSCD shifts patients away from the more desirable treatment E_3 to the two inferior arms. It thus appears that the advantage of CMAP over TSCD is more pronounced in this type of randomized trial.

7. Discussion

We have proposed an adaptive Bayesian method for monitoring the probabilities of composite events in phase II trials where a fixed time period of nontrivial length is required to evaluate each patient's response. A key component of the procedure is the method for estimating the conditional probabilities $\omega_{i1}(t)$ and $\omega_{i2}(t)$. Our estimators $\hat{\omega}_{i1}(t)$ and $\hat{\omega}_{i2}(t)$ are weighted averages of a prior and an empirical component. Choices of the empirical component other than those given by (2.5) and (2.6) can be contemplated. Under case 1, a consistent empirical component for $\hat{\omega}_{i1}(t)$ is $\{1 - KM(C_i(t))\}/\{1 - KM(T)\}$, where $KM(\cdot)$ is the product limit estimate of the survivor function (Kaplan and Meier, 1958). Another alternative is

$$\frac{\kappa}{\nu+1} + \frac{1}{\nu+1}\left(\frac{C_i(t) - z_{(\kappa)}}{z_{(\kappa+1)} - z_{(\kappa)}}\right), \tag{7.1}$$

where ν is the number of events (deaths), $0 \equiv z_{(0)} < z_{(1)} \leq \cdots \leq z_{(\nu)} < z_{(\nu+1)} \equiv T$ are the ordered event times, and $\kappa = \max_{0 \leq j \leq \nu}\{j : u \geq z_{(j)}\}$. This function, ignoring the censored nonevents, assigns equal mass on all $\nu + 1$ intervals $(z_{(j)}, z_{(j+1)}], j = 0, \ldots, \nu$ and assumes a piecewise distribution that is uniform within each interval. This was proposed by Cheung and Chappell (2000), in the context of a dose-finding trial, as an estimate of the conditional distribution of the time to toxicity given that toxicity occurred within the observation period. If the event-time distributions are continuous, then the function (7.1) has the advantage that it also is continuous. However, generalization to more complicated situations with dependent censoring, such as cases 2 and 3, is not straightforward.

A more direct approach would be to model the event-time distributions, as we did in fitting the historical data in Section 5.1, and derive the posterior of θ from this fully Bayesian model. Instead, our approach focuses directly on θ, the parameter of interest for monitoring. This provides a simpler and more natural way to specify a prior on θ. The price for this is the need to estimate the ω_{ij}'s, and consequently the posterior of θ is approximate. We feel that the benefits of our method, including both its practicality and desirable OCs compared with other methods, outweigh these limitations.

In case 1, the mixture distribution that is the approximate posterior of θ under the working likelihood has the same components as the posterior of a nonparametric Bayes estimator with a Dirichlet process prior (Susarla and Van Ryzin 1976, Theorem 5), although the mixing probabilities may differ. Elicitation of a Dirichlet process prior is not straightforward since it requires a finite measure on $\mathcal{R}^+$ *a priori*. In contrast, our procedure requires only prior beliefs regarding the event rate at time T, while the nuisance parameters, which are of infinite dimension, are replaced by the estimates $\hat{\omega}_1$ and $\hat{\omega}_2$, which may easily incorporate prior knowledge of the time distributions.

Our proposed method is related to a Bayesian approach proposed by Follmann and Albert (1999) that addresses the case of a binary outcome defined in terms of one censored time-to-event variable, our case 1. They assume a Dirichlet prior on discrete event-time probabilities to produce a posterior distribution that is a mixture of Dirichlets. However, their method does not accommodate composite events defined in terms of two or more censored time-to-event variables, such as our cases 2 and 3.

An *R*-function that implements both the CMAP and TS methods is available from the first author on request.

Acknowledgements

We thank Yu Shen for providing code to analyze the historical data with the competing risks model. We are also grateful to a referee for constructive comments on an earlier draft. Dr Thall's work was partially supported by NIH grant CA-83932.

Résumé

Dans beaucoup d'essais cliniques de Phase II, on mesure de façon séquentielle la probabilité d'un événement binaire (la 'réponse'), critère éventuellement composite, défini à partir d'une ou plusieurs variables de temps évaluant la survenue d'événements dans une période donnée. Ces méthodes, qui requièrent de s'adapter aux résultats observés, peuvent nécessiter de suspendre à plusieurs reprises le recrutement des patients afin de pouvoir suivre chaque patient le temps nécessaire

à l'évaluation de sa réponse. La durée de l'essai s'en voit augmentée et les patients éligibles qui se présentent dans ces périodes d'interruption doivent attendre la reprise des inclusions ou bien être traités en dehors de l'essai. Une autre approche, qui n'utilise pas toute l'information disponible, consiste à ignorer les données censurées au moment où on applique les règles d'arrêt de la procédure séquentielle. Nous proposons ici une méthode adaptative bayésienne qui élimine ces problèmes. A chaque inclusion d'un patient, on calcule un critère d'arrêt en utilisant une approximation de la probabilité a posteriori de la réponse, approximation qui tient compte de la totalité des données disponibles (y compris les données des patients dont la réponse finale est encore indéterminée). Un exemple d'application à un essai dans la leucémie avec événement composite montre que cette méthode peut réduire la durée de l'essai de manière importante tout en préservant la fiabilité des décisions prises en cours d'essai au vu de la procédure séquentielle.

References

Bryant, J. and Day, R. (1995). Incorporating toxicity considerations into the design of two-stage phase II clinical trials. *Biometrics* **51,** 1372–1383.

Cheung, Y. K. and Chappell, R. (2000). Sequential designs for phase I clinical trials with late-onset toxicities. *Biometrics* **56,** 1177–1182.

Conaway, M. R. and Petroni, G. R. (1995). Bivariate sequential designs for phase II trials. *Biometrics* **51,** 656–664.

Dabrowska, D. M. and Doksum, K. A. (1988). Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association* **83,** 744–749.

Fleming, T. R. (1982). One-sample multiple testing procedure for phase II clinical trials. *Biometrics* **38,** 143–151.

Follmann, D. A. and Albert, P. S. (1999). Bayesian monitoring of event rates with censored data. *Biometrics* **55,** 603–607.

Heitjan, D. F. (1997). Bayesian interim analysis of phase II cancer clinical trials. *Statistics in Medicine* **16,** 1791–1802.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53,** 457–481.

O'Quigley, J., Pepe, M., and Fisher, L. (1990). Continual reassessment method: A practical design for phase I clinical trials in cancer *Biometrics* **46,** 33–48.

Prentice, R. L., Kalbfleish, J. D., Peterson, A. V., Jr., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34,** 541–554.

Shen, Y. and Thall, P. F. (1998). Parametric likelihoods for multiple nonfatal competing risks and death. *Statistics in Medicine* **17,** 999–1015.

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

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *Journal of the American Statistical Association* **71,** 897–902.

Thall, P. F. and Simon, R. (1994). Practical Bayesian guidelines for phase IIb clinical trials. *Biometrics* **50,** 337–349.

Thall, P. F. and Sung, H. G. (1998). Some extensions and applications of a Bayesian strategy for monitoring multiple outcomes in clinical trials. *Statistics in Medicine* **17,** 1563–1580.

Thall, P. F., Simon, R., and Estey, E. H. (1995). Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. *Statistics in Medicine* **14,** 357–379.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Science* **72,** 20–22.

Received February 2001. *Revised October* 2001. *Accepted November* 2001.

Appendix

Proof of Theorem 1

For clarity, the calendar time t is omitted in the notation when no ambiguity arises. The maximum likelihood estimate $\hat{\theta} = \arg\max L(\theta)$ solves

$$J(\theta) = \frac{1}{N}\sum_{i=1}^{N}\{\hat{\omega}_{i1} - \hat{\omega}_{i2}\}\left[\frac{Y_i}{\hat{\xi}_i(\theta)} - \frac{1-Y_i}{1-\hat{\xi}_i(\theta)}\right] = 0,$$

where $\hat{\xi}_i(\theta) = \hat{\omega}_{i1}\theta + \hat{\omega}_{i2}(1-\theta)$. Now let

$$J_0(\theta) = \frac{1}{N}\sum_{i=1}^{N}\{\omega_{i1} - \omega_{i2}\}\left[\frac{\pi_i}{\xi_i(\theta)} - \frac{1-\pi_i}{1-\xi_i(\theta)}\right],$$

where $\pi_i = \mathrm{pr}(A_i^o)$ and $\xi_i(\theta) = \omega_{i1}\theta + \omega_{i2}(1-\theta)$.

Then $|J(\theta) - J_0(\theta)|$ is bounded above by the sum of three terms,

$$\Delta_1(\theta) = \left|\frac{1}{N}\sum_{i=1}^{N}\{\hat{\omega}_{i1} - \omega_{i1} - \hat{\omega}_{i2} + \omega_{i2}\}\left[\frac{Y_i}{\hat{\xi}_i(\theta)} - \frac{1-Y_i}{1-\hat{\xi}_i(\theta)}\right]\right|$$

$$\Delta_2(\theta) = \left|\frac{1}{N}\sum_{i=1}^{N}\{\omega_{i1} - \omega_{i2}\} \times \left[\frac{Y_i}{\hat{\xi}_i(\theta)} - \frac{Y_i}{\xi_i(\theta)} - \frac{1-Y_i}{1-\hat{\xi}_i(\theta)} + \frac{1-Y_i}{1-\xi_i(\theta)}\right]\right|$$

$$\Delta_3(\theta) = \left|\frac{1}{N}\sum_{i=1}^{N}\{\omega_{i1} - \omega_{i2}\}\left[\frac{Y_i - \pi_i}{\xi_i(\theta)\{1-\xi_i(\theta)\}}\right]\right|.$$

$\Delta_1(\theta)$ and $\Delta_2(\theta)$ converge to zero uniformly in θ by assumption. By Kolmogorov's Strong Law of Large Numbers, $\Delta_3(\theta)$ converges to zero almost surely for all $\theta \in [\epsilon, 1-\epsilon]$ for some $\epsilon > 0$. Consequently, we have

$$\sup_{\theta\in[\epsilon,1-\epsilon]} |J(\theta) - J_0(\theta)| \to 0 \qquad \text{(A.1)}$$

with probability one as $N \to 0$. Next, it is easy to verify that $J_0(\theta_0) = 0$, and since J_0 is strictly decreasing by checking its first derivative on $(\epsilon, 1-\epsilon)$, the root is unique. Similarly, $\hat{\theta}$ is the unique solution of $J(\theta) = 0$ if it exists. Therefore, (A.1) implies $|\hat{\theta} - \theta_0| \to 0$ almost surely by continuity of J and J_0.