Practical Bayesian Guidelines for Phase IIB Clinical Trials

Peter F. Thall; Richard Simon

Biometrics, Vol. 50, No. 2 (Jun., 1994), 337-349.

Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28199406%2950%3A2%3C337%3APBGFPI%3E2.0.CO%3B2-V

Biometrics is currently published by International Biometric Society.

BIOMETRICS 50, 337–349
June 1994

Practical Bayesian Guidelines for Phase IIB Clinical Trials

Peter F. Thall

Department of Biomathematics, Box 237, M. D. Anderson Cancer Center,
University of Texas,
1515 Holcombe Boulevard, Houston, Texas 77030, U.S.A.

and

Richard Simon

Biometric Research Branch, CTEP, Division of Cancer Treatment,
National Cancer Institute,
6130 Executive Boulevard, Rockville, Maryland 20892, U.S.A.

SUMMARY

A Phase IIB clinical trial typically is a single-arm study aimed at deciding whether a new treatment E is sufficiently promising, relative to a standard therapy, S, to include in a large-scale randomized trial. Thus, Phase IIB trials are inherently comparative even though a standard therapy arm usually is not included. Uncertainty regarding the response rate Θ_S of S is rarely made explicit, either in planning the trial or interpreting its results. We propose practical Bayesian guidelines for deciding whether E is promising relative to S in settings where patient response is binary and the data are monitored continuously. The design requires specification of an informative prior for Θ_S, a targeted improvement for E, and bounds on the allowed sample size. No explicit specification of a loss function is required. Sampling continues until E is shown to be either promising or not promising relative to S with high posterior probability, or the maximum sample size is reached. The design provides decision boundaries, a probability distribution for the sample size at termination, and operating characteristics under fixed response probabilities with E.

1. Introduction

Phase II clinical trials are usually single-arm studies aimed at estimating the activity of a new therapy. They are especially prominent in cancer therapeutics, where new treatments frequently arise as combinations of chemotherapeutic agents and growth factors or as varying dose or radiation fractionation schedules. Early Phase II trials, known as Phase IIA trials, are conducted to determine whether a drug has any antidisease activity. Phase IIB trials are conducted to better determine the degree of activity of drugs or combinations known to be active. Such trials are of great importance because they are used to identify treatments that are sufficiently promising, relative to a standard therapy S, to include in a randomized comparative trial with S. Thus, Phase II trials are inherently comparative, regardless of whether they include a control arm of patients treated with S. They provide an essential bridge between small Phase I trials, which determine the maximum tolerated doses of drugs, and large-scale randomized Phase III trials.

To implement most of the commonly used designs for Phase II trials, such as those proposed by Gehan (1961), Fleming (1982), and Simon (1989), the clinician must specify a single value of the patient response rate Θ_S to S. In many cases, however, there is uncertainty regarding Θ_S. For example, it is common for a clinician to give a range of values when asked to provide the response rate of S. Under such circumstances it may be inappropriate to insist on a single value of Θ_S for planning purposes, since the resulting design and analysis would treat an inherently variable quantity as if it were a constant. A more realistic approach should explicitly account for the clinician's uncertainty regarding Θ_S, both in planning the Phase II trial and in interpreting its results.

When reliable historical data on S are available, they may be incorporated formally into the trial design and possibly into the subsequent statistical inferences as well. Thall and Simon (1990) use

Key words: Bayesian inference; Design; Phase II trial; Prior distribution; Sample size.

empirical Bayes methods to incorporate data from historical pilot studies of S into both the trial design and a final test of efficacy. In the absence of such data on S, clinical experience and current belief regarding the efficacy of S may be represented by a probability distribution on Θ_S, and in this case a Bayesian approach is appropriate. Although the two approaches differ philosophically, both provide formal models for randomness in the standard therapy response rate.

In this paper we present a Bayesian approach to the design and analysis of Phase II clinical trials. We consider settings where patient response is binary (success or failure) and the data are monitored continuously. The aim is to provide simple, practical guidelines for deciding whether a new therapy E is promising relative to S while accounting explicitly for uncertainty regarding the response rates of S and E. The design requires the clinician to provide a prior for Θ_S, a targeted improvement for E, and bounds $n_{\min}$ and $n_{\max}$ on the allowable sample size. A flat or weakly informative prior for Θ_E is used. No explicit specification of a loss function is required. The trial continues until E is shown with high posterior probability to be either promising or not promising or until $n_{\max}$ is reached. In application, the design simply consists of a sequence of upper and lower decision cutoffs for continuously monitored data and frequentist operating characteristics, including the probability distribution of the sample size N.

Mehta and Cain (1984) take a similar approach to fixed-sample-size pilot studies, using posterior probabilities that E is active compared to a fixed standard for S to derive charts for early termination and for posterior probability intervals. Ho (1991) examines some frequentist properties of a group sequential Bayesian rule for comparing two Gaussian samples. Analogously to the approach taken by Ho, we use a Bayesian framework to obtain decision rules and then evaluate the behavior of the design so derived under fixed values of the experimental treatment success probability.

Specific criteria for generating the decision boundaries are given in Section 2. Methods for eliciting and quantifying prior information are discussed in Section 3. Section 4 presents numerical operating characteristics of the design for a range of priors and decision parameters. Guidelines for selecting design parameters based on these results are provided. In particular, the formulation appears useful for determining whether a nonrandomized Phase II trial is appropriate, since highly disperse priors on Θ_S produce designs that are unlikely to yield conclusive results without very large sample sizes. An application is described in Section 5, followed by a general discussion in Section 6.

2. Decision Criteria

The sequence of patient responses to E will be denoted $Y_1, Y_2, \ldots$, with each $Y_i = 0$ or 1 as the treatment is a failure or a success. The total number of successes out of the first n patients is thus $X_n = Y_1 + \cdots + Y_n$. We assume that conditional on Θ_E, the Y_i's are exchangeable with $\Pr[Y_1 = 1] = \Theta_E = 1 - \Pr[Y_1 = 0]$. In particular, this implies that the response rate does not "drift" over the course of the trial. We address settings in which the data are monitored continuously, i.e., all X_n are observed up to some predetermined practical limit $n_{\max}$, and the clinicians wish to declare E promising, not promising, or to terminate the trial at any time based on the most recent data. Such circumstances reflect actual clinical practice in decision-making with experimental therapies in many Phase II trials, as compared to what is assumed implicitly by fixed-sample-size one-stage or group sequential protocols. We shall also assume that an informative prior π_S for Θ_S may be elicited from the clinicians, but require that the prior π_E of Θ_E be at most slightly informative. Reasons for this latter requirement will be discussed in Section 3. We model π_S and π_E as beta distributions for simplicity and convenience. This will be denoted $\pi_t = \text{beta}(a_t, b_t)$ for $t = S$ or E. In particular, a beta(a, b) distribution has mean $\mu = a/(a + b)$ and variance $\mu(1 - \mu)/(1 + a + b)$, so in general it may be characterized equivalently in terms of its mean and the concentration parameter $c = a + b$.

Denote the probability density function (pdf) and cumulative distribution function (cdf) of a beta(a, b) distribution by $f(\cdot; a, b)$ and $F(\cdot; a, b)$, respectively. Our criterion function for determining the trial decision cutoffs is the posterior probability

$$\lambda(x, n; \pi_S, \pi_E, \delta_0) = \Pr(\Theta_S + \delta_0 < \Theta_E | X_n = x \text{ out of } n)$$

$$= \int_0^{1-\delta_0} \{1 - F(p + \delta_0; a_E + x, b_E + n - x)\} f(p; a_S, b_S)\, dp,$$

for $n = 1, 2, \ldots, n_{\max}$, using the fact that the posterior density of Θ_E given X_n is beta($a_E + X_n$, $b_E + n - X_n$). This is easily evaluated via numerical integration. Let p_L and p_U denote predetermined probabilities, with p_L a small value such as .01–.05, and p_U a large value such as .95–.99. The upper and lower decision cutoffs are

$$U_n = \text{the smallest integer } x \text{ such that } \lambda(x, n; \pi_S, \pi_E, 0) \geq p_U,$$

$$L_n = \text{the largest integer } x < U_n \text{ such that } \lambda(x, n; \pi_S, \pi_E, \delta_0) \leq p_L.$$

The decision rule at stage n is as follows:

If $X_n \geq U_n$, stop and declare E promising.
If $X_n \leq L_n$, stop and declare E not promising.
If $L_n < X_n < U_n$ and $n < n_{\max}$, then continue, i.e., treat another patient.

The trial is declared inconclusive if X_n has not hit a stopping boundary by $n = n_{\max}$.

The rationale for the lower cutoff criterion is that a treatment that is very unlikely to provide an improvement of at least δ_0 over Θ_S does not warrant further consideration. The upper cutoff criterion simply says that any treatment that is highly likely to offer an improvement over S is considered promising. Although the probabilities p_L and p_U bear a superficial resemblance to Type I error rate and power in classical hypothesis testing, they are in fact quite different criteria.

3. The Priors

For eliciting a prior on Θ_S, we find it convenient to describe the beta(a, b) distribution equivalently in terms of its mean $\mu = a/(a + b)$ and W_{90} = the width of the 90% probability interval running from the 5th to the 95th percentiles. The clinician is first asked to specify the mean of Θ_S and the value of W_{90}. A computer plot of the corresponding beta(a, b) pdf may be used to enhance the clinician's conceptualization of the prior, with this process repeated until appropriate values of a and b are obtained. A nice discussion of the process of eliciting and quantifying beta(a, b) priors is given by Lindley and Phillips (1976).

Since both π_S and π_E play fundamental roles in determining the decision boundaries of the design, it is essential that π_E be formulated so that it realistically reflects limited knowledge about Θ_E while also generating a practically useful design. In terms of the concentration parameter $c_E = a_E + b_E$ of π_E, we thus require that $2 \leq c_E \leq 10$: the dispersion of π_E is roughly no more than that of the uniform distribution on [0, 1] but no smaller than that of the posterior corresponding to a small pilot study of E. Given a targeted improvement δ_0 for E over S, we set the mean π_E equal to $\mu_S + \delta_0/2$. This centers π_E at the average of the means of the priors corresponding to the most pessimistic view that E is on average identical to S and the optimistic view that E provides the targeted improvement δ_0. For the prior used here, $a_E = c_E(\mu_S + \delta_0/2)$ and $b_E = c_E[1 - (\mu_S + \delta_0/2)]$. Thus π_E is determined by μ_S, c_E, and δ_0, and in particular the mean of π_E is an explicit function of the mean of π_S and the targeted improvement.

The use of a flat prior on Θ_E is motivated by several considerations. If we allow π_E to be highly concentrated around $\mu_S + \delta_0$ or even $\mu_S + \delta_0/2$, i.e., if π_E is informative and optimistic, then $\Pr[\Theta_S < \Theta_E]$ will be large a priori and the upper criterion may be satisfied without treating any patients at all. While such prior optimism regarding the efficacy of E is encouraging, we shall insist on empirical clinical information as a basis for deciding whether E will be administered to a large number of patients in a Phase III clinical trial. Simply put, we require that the Phase II trial provide real clinical experience with E and we do not allow the prior to dominate the data. By the same token, a prior for Θ_E that is highly concentrated around μ_S reflects a rather pessimistic view, and subsequent clinical results that per se show a highly favorable performance by E may not produce a posterior on Θ_E that satisfies the upper decision criterion. Moreover, if clinicians really feel such prior pessimism regarding Θ_E, then a Phase II trial of E probably is not warranted in the first place.

Our prior π_E is similar to but not the same as "reference prior" in the sense that we wish the posterior to be dominated by the data and not the prior so that the trial results may be used by others who have their own priors. Thus we have chosen it by first varying π_E and studying the corresponding posterior probability distributions, in terms of the empirical behavior of the resulting decision rules, then restricting π_E to a set of priors yielding designs with desirable properties. Bernardo (1979, p. 115) provides a formal basis for choosing a reference prior by defining it to be that which maximizes the amount of missing information about the parameter that would be expected a posteriori based on an infinitely large sample. To our knowledge no reference prior in the sense of Bernardo has been developed for Bernoulli sampling with a sequential stopping rule.

The numerical value $W_{90} = .20$ corresponds to reasonably informative priors on Θ_S. For example, for $\mu = .20$ this converts to $(a, b) = (8.15, 32.6)$, corresponding roughly to the beta posterior obtained from a study of 39 patients with seven successes, starting from a uniform prior. From another viewpoint $\Pr[\Theta_S < .40] = .997$ for this prior, so it would be appropriate in the case where the clinician believes that the mean success rate of S is .20 and moreover is nearly certain that it is

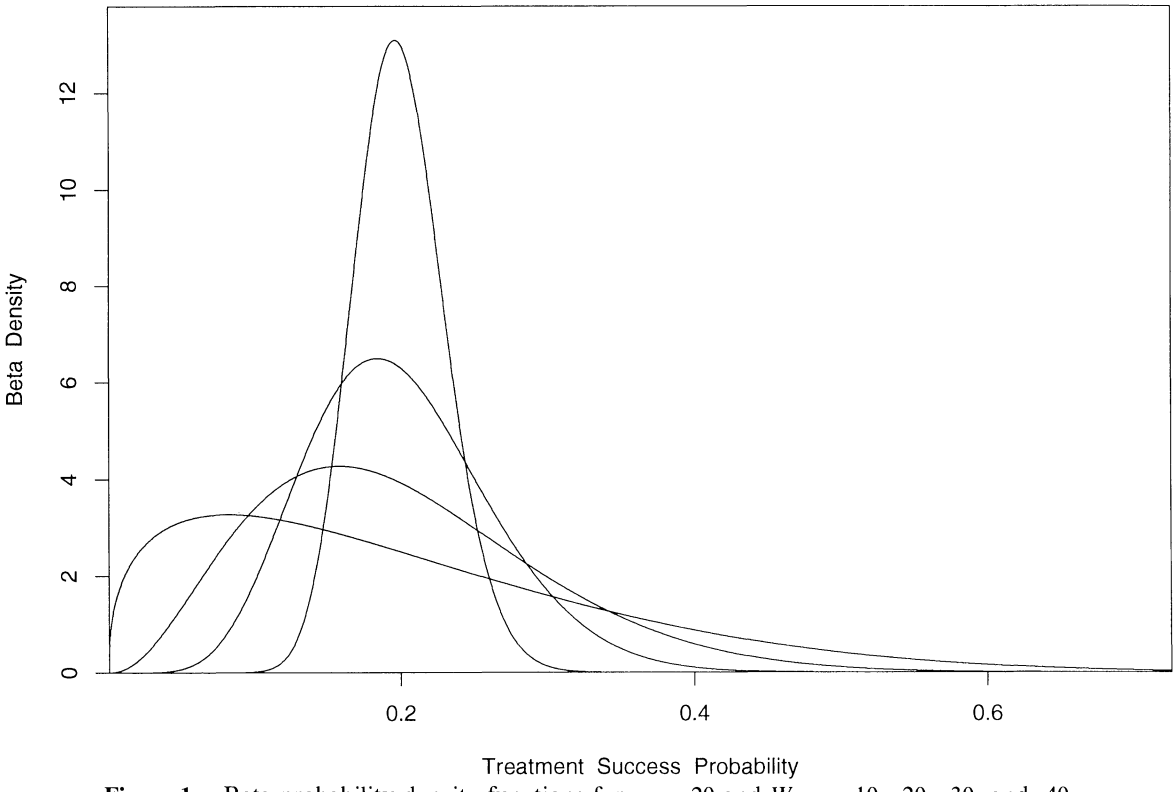

Figure 1. Beta probability density functions for $\mu = .20$ and $W_{90} = .10, .20, .30$, and .40.

no larger than .40. For $\mu = .50$ and $W_{90} = .20$, the values of (a, b) increase to (33.4, 33.4), corresponding analogously to the posterior from a study of roughly 65 patients. This illustrates the fact that more information is required to obtain the same posterior probability interval width when the mean is .50 rather than .20. For $W_{90} = .30$, corresponding to a more disperse prior on Θ_S, the respective beta parameters decrease to (3.28, 13.10) for $\mu = .20$ and (14.6, 14.6) for $\mu = .50$. These correspond roughly to the beta posteriors obtained from studies of the standard treatment based on 14 and 27 patients, respectively. In contrast, $W_{90} = .10$ corresponds to a highly informative prior on Θ_S, since for $\mu_S = .20$ this would be the beta posterior obtained from a study of 169 patients. To compare the dispersion of π_E to that of π_S, we note that for $\mu_E = .30$, which is determined by $\mu_S = .20$ and $\delta_0 = .20$, the values $c_E = 2$ and 10 correspond to $W_{90} = .716$ and .442, respectively, so our recommended priors on Θ_E are quite disperse compared to those used for Θ_S.

Beta densities with mean .20 and W_{90} ranging from .10 to .40 are shown in Figure 1. This illustrates the facts that a prior with $W_{90} = .10$ is highly concentrated about its mean, whereas $W_{90} = .40$ corresponds to a prior so disperse that it reflects almost no real knowledge about the success rate.

The design parameters thus consist of (μ_S, W_{90}), c_E and δ_0 to determine the priors, and (p_L, p_U) for the decision boundaries. We allow δ_0 to play a dual role, since it determines μ_E and also parameterizes λ for comparison to p_L when determining the lower boundary. Once these few parameters are specified, the design is obtained by first computing the decision boundaries $\{L_n, U_n, n = 1, 2, \ldots\}$, with these used in turn to evaluate the design's operating characteristics.

4. Frequentist Operating Characteristics

We next evaluate the design's behavior under a variety of circumstances corresponding to what might realistically be anticipated in the clinical setting described earlier. While the design's decision boundaries are obtained based on a Bayesian framework, which regards the success probability of E as a random quantity in order to reflect actual uncertainty, we evaluate the design's behavior under fixed values of this success probability, which we denote by p_E. Each computation requires the design parameters μ_S, W_{90}, c_E, and δ_0, the decision boundaries, and the assumed true probability p_E of a successful treatment with E. We study the designs for values of p_E ranging from μ_S to $\mu_S + .20$. The value $p_E = \mu_S$ corresponds to an experimental treatment having true success probability equal to the mean of the clinician's prior on the standard treatment success probability, and

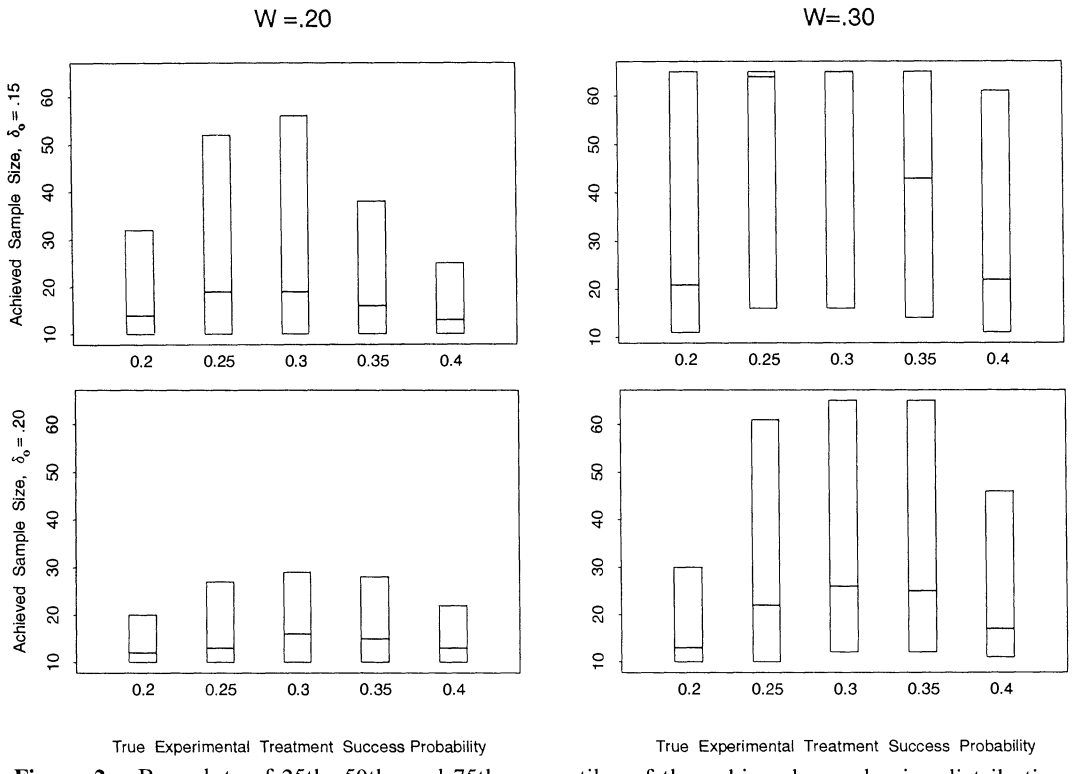

Figure 2. Box plots of 25th, 50th, and 75th percentiles of the achieved sample size distribution for $\mu_S = .20$ and $c_E = 2$.

in this case E does not provide a treatment advance over S. The values $p_E = \mu_S + .15$ and $\mu_S + .20$ correspond to cases in which E is clinically superior to S. To avoid possible confusion, we note the important distinction between the prior mean $\mu_E = \mu_S + \delta_0/2$ of π_E, which plays a fundamental role in the computation of λ and hence in determining the decision boundaries, and the assumed fixed values of p_E used to evaluate the design.

Each design's operating characteristics are summarized in terms of $p_+ = \Pr[\text{E is declared promising}]$, $p_- = \Pr[\text{E is declared not promising}]$ and the empirical 25th, 50th, and 75th percentiles of the achieved sample size N. The values of p_+ and p_- and of the probability distribution of N were computed analytically using the recursion given in the Appendix. The sample size boundaries $n_{\min} = 10$ and $n_{\max} = 65$ and cutoff criteria probabilities $(p_L, p_U) = (.05, .95)$ were used throughout, unless otherwise indicated. The value $n_{\max} = 65$ was chosen for illustration because Phase II trials of larger size often are impractical.

General patterns of variation in the sample size are illustrated in Figure 2, which presents box plots of the distribution of N for standard prior mean $\mu_S = .20$ and experimental treatment prior concentration parameter $c_E = a_E + b_E = 2$. Figure 3 presents corresponding graphs of p_+, p_-, and the probability $1 - p_+ - p_-$ of an inconclusive trial. In both Figures 2 and 3, the true experimental treatment success probability p_E is varied from $\mu_S = .20$ to $\mu_S + .20 = .40$ for each combination of $W_{90} = .20, .30$ and $\delta_0 = .15, .20$. The patterns in which N, p_+, and p_- each vary with W_{90}, δ_0, and p_E are similar for other values of μ_S and c_E.

A general message of Figure 2 is that both the median and variability of the achieved sample size increase as the prior on Θ_S becomes less informative or as the targeted improvement δ_0 is decreased. For $W_{90} = .20$ in this case the median of N varies from 12 to 19; when W_{90} is increased to .30, N becomes moderately higher at $\delta_0 = .20$ and substantially higher at $\delta_0 = .15$. Values of the true success rate p_E midway between μ_S and $\mu_S + \delta_0$ produce the largest values of N.

The plots in Figure 3 show that the Type I and Type II error rates and the probability of an inconclusive trial all increase as the prior on Θ_S becomes more disperse or as the targeted improvement is decreased. In terms of both sample size distribution and decision probabilities, the design based on $W_{90} = .20$ and $\delta_0 = .20$ appears to be the most attractive. If prior knowledge about S is characterized by $W_{90} = .30$ then N is very likely to take on substantially larger values compared to the case where $W_{90} = .20$, and the trial has a nontrivial probability of being inconclusive. The

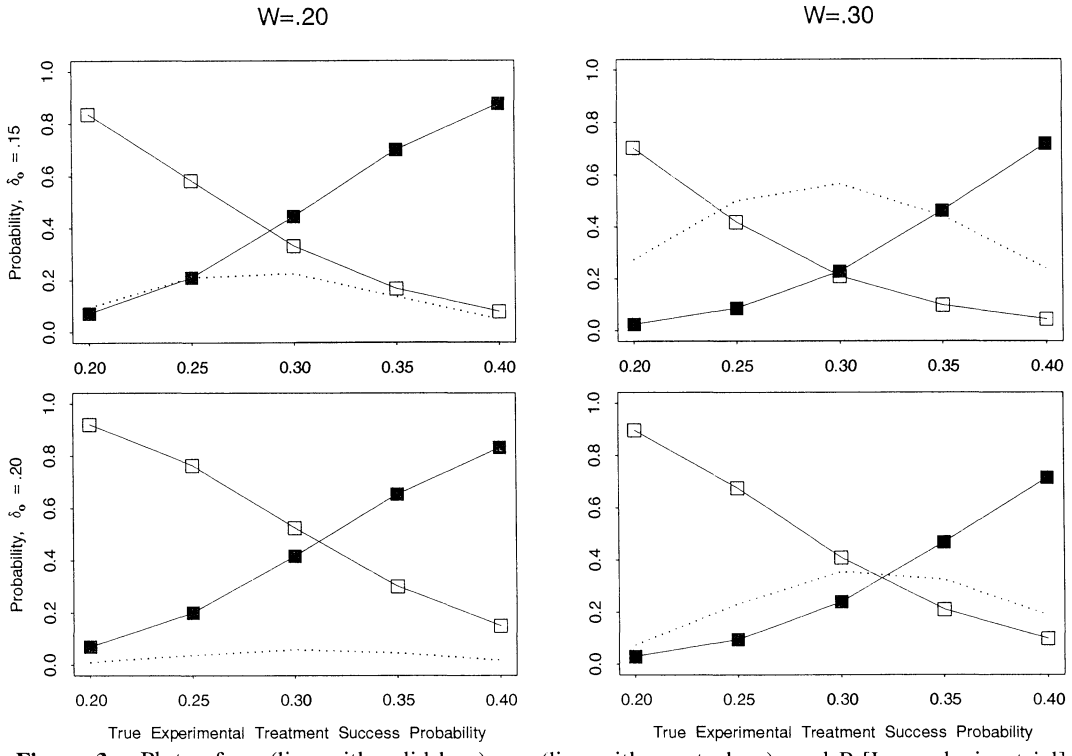

Figure 3. Plots of p_+ (line with solid box), p_- (line with empty box), and Pr[Inconclusive trial] $= 1 - p_+ - p_-$ (dashed line), for $\mu_S = .20$ and $c_E = 2$.

design based on a target improvement of $\delta_0 = .20$ may still provide acceptable operating characteristics in this case, but use of $\delta_0 = .15$ is probably ill advised.

The ways in which N, p_+, and p_- each vary with W_{90} and δ_0 in Figures 2 and 3 also hold true for other combinations of μ_S and c_E. The corresponding designs for $c_E = 2$ and $\mu_S = .50$ are about equally attractive compared to those shown in Figures 2 and 3, so it appears that varying the standard mean response rates from .20 to .50 does not have a large effect on the design's operating characteristics. When the concentration parameter c_E of π_E is increased from 2 to 10, however, the problems noted above for $W_{90} = .30$ become even more extreme. The situation $c_E = 10$ and $W_{90} = .30$ would arise, e.g., in the case $\mu_S = .20$, if the priors were based, respectively, on two earlier studies in which 14 patients were treated with S and 8 patients were treated with E. This is very different from the setting that we address with our design, namely one in which there is considerable experience with S but little is known about E, so it is not surprising in the former case that a single-arm study of E has poor operating characteristics.

Numerical results corresponding to standard treatment prior means $\mu_S = .20$ and .50 are presented in Tables 1 and 2, respectively. These tables illustrate the manner in which the design's operating characteristics vary with c_E and μ_S. The other design parameters are varied through each combination of $W_{90} = .20, .30$ and $\delta_0 = .15, .20$. Both p_+ and p_- decrease or remain constant as c_E is increased from 2 to 10, with a few small exceptions. This is because more data are required to reach a decision when the prior on Θ_E is more concentrated around its mean $\mu_S + \delta_0/2$. Several interesting patterns emerge as μ_S is increased from .20 to .50: in all cases p_+ increases and p_- decreases, while achieved sample size increases for $p_E = \mu_S$. These results may be attributed to the fact that the binomial variance is largest for values of p close to .50.

An important issue is how the design behaves for highly informative priors, since this corresponds to circumstances in which the clinicians have had a great deal of experience with the standard therapy. The first two rows of Table 3 present the operating characteristics of a design based on a prior with $W_{90} = .10$, followed by the corresponding values for $W_{90} = .20$ repeated from Table 1 to facilitate comparison. Since the values of p_- for $p_E = \mu_S + .20$ are rather high, specifically $p_- = .173$ for $W_{90} = .10$ and $p_- = .150$ for $W_{90} = .20$, we hypothesized that more stringent decision criteria might produce a more attractive overall design. We therefore examined behavior of the corresponding designs with $p_L = .02$ and $p_U = .98$, the results of which are also given in Table 3. The more extreme cutoffs produce a design with very attractive operating characteristics when $W_{90} = .10$.

Table 1

Operating characteristics for $\mu_S = .20$, $(p_L, p_U) = (.05, .95)$, $(n_{min}, n_{max}) = (10, 65)$, *and true* Pr[Success] = p_E.

		$c_E = 2$		$c_E = 10$	
		$W_{90} = .20$	$W_{90} = .30$	$W_{90} = .20$	$W_{90} = .30$
$\delta_0 = .15$	**N**	10 14 32	11 21 65	12 20 45	19 43 65
$p_E = \mu_S$	p_+	.071	.024	.046	.008
	p_-	.835	.703	.813	.592
$\delta_0 = .15$	**N**	10 16 38	14 43 65	12 26 55	30 65 65
$p_E = \mu_S + .15$	p_+	.700	.461	.685	.333
	p_-	.165	.097	.109	.045
$\delta_0 = .15$	**N**	10 13 25	11 22 61	11 18 34	17 41 65
$p_E = \mu_S + .20$	p_+	.875	.717	.883	.609
	p_-	.077	.043	.042	.016
$\delta_0 = .20$	**N**	10 12 20	10 13 30	10 13 24	11 19 39
$p_E = \mu_S$	p_+	.070	.029	.058	.013
	p_-	.920	.896	.929	.866
$\delta_0 = .20$	**N**	10 13 22	11 17 46	11 15 25	14 31 65
$p_E = \mu_S + .20$	p_+	.832	.714	.844	.643
	p_-	.150	.098	.131	.063

*N = (25th, 50th, 75th) percentiles of achieved sample size, p_+ = Pr[E declared promising], and p_- = Pr[E declared not promising].

Although this is certainly a matter of opinion, we recommend calibrating the values of p_L and p_U in this manner to obtain desirable designs when π_S is highly informative.

The use of more stringent decision criteria reveals what may be a subtle effect of the fixed upper bound on N. Note that as (p_L, p_U) change from (.05, .95) to (.02, .98) in Table 3, the value of p_+ for $p_E = \mu_S + .20$ increases from .827 to .872 when $W_{90} = .10$, but decreases from .832 to .766 when $W_{90} = .20$. This may be attributed to the fact that N is more disperse for $W_{90} = .20$ and thus hits the upper bound of 65 more often when the termination criteria are made more stringent, resulting in a a loss of power.

The case of a disperse prior on Θ_S is more problematic. Tables 1 and 2 indicate that when $W_{90} = .30$ it is advisable to use a targeted improvement of $\delta_0 = .20$ rather than .15, since the latter produces undesirable designs in most of the cases studied. To examine the effects of an even more disperse

Table 2

Operating characteristics for $\mu_S = .50$, $(p_L, p_U) = (.05, .95)$, $(n_{min}, n_{max}) = (10, 65)$, *and true* Pr[Success] = p_E.

		$c_E = 2$		$c_E = 10$	
		$W_{90} = .20$	$W_{90} = .30$	$W_{90} = .20$	$W_{90} = .30$
$\delta_0 = .15$	**N**	11 18 42	12 28 65	12 25 55	22 65 65
$p_E = \mu_S$	p_+	.134	.085	.094	.034
	p_-	.721	.567	.703	.464
$\delta_0 = .15$	**N**	10 18 39	10 28 65	14 24 53	22 62 65
$p_E = \mu_S + .15$	p_+	.754	.601	.729	.487
	p_-	.110	.063	.081	.028
$\delta_0 = .15$	**N**	10 13 26	10 17 40	11 19 32	16 30 62
$p_E = \mu_S + .20$	p_+	.917	.832	.914	.764
	p_-	.040	.022	.025	.008
$\delta_0 = .20$	**N**	10 12 23	11 16 35	11 16 26	14 25 52
$p_E = \mu_S$	p_+	.139	.097	.108	.045
	p_-	.842	.778	.865	.759
$\delta_0 = .20$	**N**	10 12 21	10 13 31	11 16 26	12 24 50
$p_E = \mu_S + .20$	p_+	.881	.835	.889	.787
	p_-	.101	.052	.087	.032

*N = (25th, 50th, 75th) percentiles of achieved sample size, p_+ = Pr[E declared promising], and p_- = Pr[E declared not promising].

Table 3
Operating characteristics for highly informative priors on Θ_S. *All designs correspond to* $\mu_S = .20$, $\delta_0 = .20$, *and* $c_E = 2$.

		p_E	N	p_+	p_-
$(p_L, p_U) = (.05, .95)$	$W_{90} = .10$	.20	10 12 16	.107	.893
		.40	10 11 18	.827	.173
	$W_{90} = .20$	.20	10 12 20	.070	.920
		.40	10 13 22	.832	.150
$(p_L, p_U) = (.02, .98)$	$W_{90} = .10$	.20	10 14 29	.052	.932
		.40	11 17 28	.872	.107
	$W_{90} = .20$	.20	11 19 38	.022	.887
		.40	11 22 50	.766	.065

prior, we considered a design based on π_S with $W_{90} = .40$; the results are presented in Table 4 along with those of the corresponding design with $W_{90} = .30$ repeated from Table 1 for comparison. The highly disperse prior produces a design requiring a very large number of patients and having an unacceptably low probability of detecting a large true improvement. Although relaxing the decision criteria to $(p_L, p_U) = (.10, .90)$ produces a design with more acceptable properties, we believe that when knowledge about the standard therapy is so limited a randomized control arm of patients treated with S should be included, since this deals with the problem more directly by obtaining empirical information on Θ_S where it is lacking. In this case an unbalanced randomized two-arm trial may be appropriate. This situation is analogous to the case of limited or highly variable historical data on S in the empirical Bayes setting treated by Thall and Simon (1990). Since it is not entirely obvious how such a trial would be formulated using the present approach, we leave the relevant details for future consideration.

An alternative approach to evaluating operating characteristics is to incorporate a priori uncertainty regarding Θ_E. We did this by first sampling a value of p_E according to π_E, then simulating a sequence of independent, identically distributed (iid) binary responses $Y_1, Y_2, \ldots$ according to p_E and comparing the successive sums $X_n = Y_1 + \cdots + Y_n$ to the decision boundaries. This was

Table 4
Operating characteristics for a highly disperse prior on Θ_S. *Both designs correspond to* $\mu_S = .20$, $\delta_0 = .20$, $c_E = 2$, *and* $(p_L, p_U) = (.05, .95)$.

W_{90}	p_E	N	p_+	p_-
.30	.20	10 13 30	.029	.896
	.40	11 17 46	.714	.098
.40	.20	10 20 57	.008	.762
	.40	14 65 65	.343	.064

Table 5
Unconditional probabilities of declaring E *promising or not promising for* $\mu_S = .20$, $(p_L, p_U) = (.05, .95)$, $(n_{min}, n_{max}) = (10, 65)$ *based on the prior distribution of* Θ_E

		$c_E = 2$		$c_E = 10$	
		$W_{90} = .20$	$W_{90} = .30$	$W_{90} = .20$	$W_{90} = .30$
$\delta_0 = .15$	N	10 10 13	10 11 19	12 17 37	14 32 65
	p_+	.388	.328	.357	.238
	p_-	.574	.546	.498	.393
	$1 - p_+ - p_-$	.038	.126	.145	.369
$\delta_0 = .20$	N	10 10 12	10 10 14	10 13 22	11 19 54
	p_+	.404	.354	.431	.301
	p_-	.587	.573	.542	.480
	$1 - p_+ - p_-$	.009	.073	.027	.219

*N = (25th, 50th, 75th) percentiles of achieved sample size, p_+ = Pr[E declared promising], and p_- = Pr[E declared not promising].

repeated 2,000 times for each design and the proportions of runs in which E was declared promising, not promising, or the simulated trial was inconclusive were recorded. The results for $\mu_S = .20$ are given in Table 5. The values of p_+ and p_- in Table 5 have very different interpretations from the operating characteristics given in Tables 1–4, because randomizing p_E according to π_E reflects what may be anticipated a priori averaging over a flat prior on Θ_E, rather than for a single fixed value of p_E. For example, with $\delta_0 = .15$, $c_E = 2$, and $W_{90} = .20$, the unconditional prior expectation is .388 that there will be a positive result, and .038 that the trial will be inconclusive.

5. An Application

The approach proposed here has been used to design a clinical trial of fludarabine + ara-C + granulocyte colony stimulating factor (G-CSF) to treat poor-prognosis patients suffering from acute myelogenous leukemia (AML) at The University of Texas M. D. Anderson Cancer Center. This AML patient subgroup is defined in terms of unfavorable cytogenetic abnormalities and the presence of an antecedent hematologic disorder. Using established treatments, to date there are virtually no survivors beyond 2 years among such patients. The clinical endpoint for the trial is complete remission (CR) of the leukemia. The innovative aspect of the treatment is the use of the growth factor G-CSF, since the standard therapy currently is fludarabine + ara-C alone.

The Bayesian approach was adopted in part due to the desire to carry out the study quickly and efficiently, since the number of patients accrued in this group is limited to about 4 per month while their survival is so poor with current treatment. Moreover, it is both feasible and desirable to continuously monitor AML patient responses at M. D. Anderson. The clinician's prior on the standard success rate had a mean of .50 with $W_{90} = .20$, reflecting overall clinical experience and a recent study in which 22 of 45 such patients treated with fludarabine + ara-C achieved CR. The design based on these parameters with $c_E = 2$ and $\delta_0 = .20$ was implemented. As can be seen from Table 2, this design has an 88% probability of correctly identifying a 20% improvement in CR rate and a 10% probability of missing such an improvement. If the true E success rate is only .50, there is an 84% probability of correctly declaring the treatment not promising and a 14% probability of incorrectly declaring it promising. For true success rates (.50, .55, .60, .65, .70) the corresponding sample size distributions have medians (12, 15, 15, 15, 12) and 75th percentiles (23, 28, 29, 26, 21). Based on the current accrual rate, the median trial duration thus should be 3–5 months, and even if $p_E = .60$ there is a 75% chance that the trial will terminate after about 7 months.

6. Discussion

Many authors have proposed the use of Bayesian methods in clinical trials and medical statistics. Nearly three decades ago Novick and Grizzle (1965) and Cornfield (1966a, 1966b), among others, argued cogently for such an approach, although their suggestions seem to have been largely ignored in clinical practice. In recent years the biostatistical literature has shown a renewed interest in Bayesian methods, including general articles by Berry (1985, 1987), Racine et al. (1986), and Spiegelhalter and Freedman (1988).

Sylvester and Staquet (1977, 1980) and more recently Sylvester (1988) have proposed decision-theoretic Bayesian methods for Phase II clinical trials. They take the classical approach in which a decision rule is obtained by minimizing the Bayes risk, using a two-point prior on Θ_E. They address the problem of optimizing the sample size and decision cutoff of a single-stage design where n is fixed, with the objective simply to determine whether a new drug is active. In contrast, we intentionally avoid the difficulties associated with specification of a loss function, although a loss function certainly is defined implicitly in our formulation.

If a trial is inconclusive using our approach, the disposition of E will depend on what other regimens are available, the results that have been obtained for them, and additional considerations such as toxicity and cost. Even in such a case, the current trial will have been successful in producing a useful estimate of the efficacy of E. Moreover, it is unlikely that investigators or sponsors would be interested in Phase II trials that are likely to be as large as some Phase III trials.

Several generalizations and modifications of our approach are motivated by purely practical considerations. The first is the idea noted earlier that a randomized two-arm trial may be more appropriate in situations where limited knowledge about S is reflected in a highly disperse prior. This was discussed by Meier (1975). Thall and Simon (1990) dealt with the analogous problem in an empirical Bayes setting where a single-stage design is desired and historical data on S are limited. They determined the proportions of patients randomized to E and S arms by minimizing the variance of the estimated mean treatment effect difference.

Another design modification would be to impose an additional early termination rule based on the predictive probability of obtaining a conclusive result, i.e., the conditional probability based on

current data that E will be declared either promising or not promising after accrual of $n_{\max}$ patients. The use of predictive probabilities for decision making in clinical trials has been discussed by several authors, including Herson (1979), Spiegelhalter and Freedman (1986), Grieve (1988), and Choi and Pepple (1989). In the present context, a desirable goal might be to terminate the trial early if it is highly unlikely that the trial will ever produce a conclusive result. Formally, for $m < n_{\max}$ and $x \leq z \leq x + n_{\max} - m$, the conditional distribution of $X_{n_{\max}}$ given X_m is the beta–binomial pdf

$$\Pr[X_{n_{\max}} = z | X_m = x] = \binom{n_{\max} - m}{z - x} \frac{B(z + a_E,\ n_{\max} - z + b_E)}{B(x + a_E,\ m - x + b_E)},$$

where $B(u, v)$ is the beta function. In general, this could be used to compute $\Pr[X_{n_{\max}} = U_{n_{\max}}$ or $L_{n_{\max}} | X_m = x]$ at any interim point and the trial terminated if it is less than a specified cutoff p_t. Based on our numerical results, such a rule would be of most value when the true success rate of E is midway between μ_S and $\mu_S + \delta_0$.

A slightly different objective in some Phase II trials is to determine whether E is not unacceptably worse than S. This might be appropriate if negative side effects associated with E were substantially less severe than those of S. Also, Phase II studies sometimes have the objective of ensuring that a candidate experimental treatment for a randomized trial appears competitive to the standard. The approach proposed here could be easily modified to accommodate this objective by replacing the event $[\Theta_S < \Theta_E]$ used for the upper probability comparison with, say, $[\Theta_S - .05 < \Theta_E]$.

In a cooperative group setting it may be impractical to monitor data continuously. For example, interim monitoring for common diseases such as non–small-cell lung cancer may slow down the trial since registration must be suspended while the continuation criteria are checked. In such a setting, especially where further confirmation of an informative prior on Θ_E is desired, interim monitoring and preliminary stopping rules may be undesirable. The interpretation of results at the end of the trial could consist of informal examination of the posterior distribution of $\Theta_E - \Theta_S$ for a set of prior distributions, including an informative prior if that is of interest. This type of application is rather different from what we have in mind here; in addition to the ability to monitor the data continuously, we find the decision structure of the proposed Phase II design very useful both for planning purposes and for interpretation of results. However, a fixed-sample-size version of our design could be constructed based on the width of the posterior distribution of Θ_E or $\Theta_E - \Theta_S$.

Phase IIB trials are inherently comparative, except in the case of a new disease or where there is no effective treatment, hence a numerical standard for Θ_S is required for designing the trial. Since this must be obtained from some combination of empirical data and subjective clinical experience, there will always be uncertainty regarding Θ_S, whether one regards it as a random parameter in a Bayesian setting or as a statistical estimate from a frequentist viewpoint. Using an empirical Bayes approach, which is in fact frequentist and not Bayesian, Thall and Simon (1990) showed that ignoring the variability inherent in a statistical estimate $\hat{\theta}_S$ leads to an underestimate of the required trial sample size or to inflated Type I and Type II error rates. In the Bayesian setting, this is analogous to declaring $\Pr[\Theta_S + \delta_0 < \Theta_E | X_n = x$ out of $n]$ either very large or very small because Θ_S is incorrectly treated as a constant and the new data on E are thus given unrealistically heavy weight. Under either approach, ignoring the randomness in Θ_S or in $\hat{\theta}_S$ thus may lead to an overstatement of the Phase II trial results, which may lead in turn to large-scale randomized trials of inferior new treatments or to the discarding of superior new treatments. We thus argue that, regardless of statistical philosophy, a proper scientific approach should begin with an honest account of the information available on S, whether it is objective data or subjective clinical experience.

Résumé

Un essai thérapeutique de Phase IIB est spécifiquement un essai à une branche visant à décider si un nouveau traitement E est suffisamment prometteur, comparé à une thérapeutique standard S, pour entreprendre un essai randomisé à grande échelle. De la sorte, les essais de Phase IIB sont comparatifs de façon enhérente, même si une branche correspondant à la thérapeutique standard n'est habituellement pas incluse. L'incertitude concernant la réponse Θ_S de S est rarement rendue explicite, que ce soit en planifiant l'essai ou en interprétant ses résultats. Nous proposons des lignes directrices pratiques, bayésiennes, pour décider si E est prometteur par rapport à S dans des contextes où la réponse du patient est binaire, et où les données sont suivies continûment. Le plan nécessite la spécification d'une information à priori pour Θ_S, d'une amélioration souhaitée par E, et de limites pour la taille d'échantillon possible. Aucune spécification explicite de fonction de perte n'est requise. L'échantillonnage se poursuit jusqu'à ce que E soit considéré soit prometteur soit non prometteur relativement à S, avec une probabilité a postériori élevée, ou bien que la taille maximum

d'échantillon soit atteinte. Le plan permet d'obtenir les limites de décision, une distribution de probabilité correspondant à la taille d'échantillon effectuée en fin d'essai, et les caractéristiques de performance sous des probabilités de réponse à E fixées.

References

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions*. New York: Dover.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (with Discussion). *Journal of the Royal Statistical Society, Series B* **41,** 113–147.

Berry, D. A. (1985). Interim analyses in clinical trials: Classical vs. Bayesian approaches. *Statistics in Medicine* **4,** 521–526.

Berry, D. A. (1987). Interim analysis in clinical trials: The role of the likelihood principle. *The American Statistician* **41,** 117–122.

Choi, S. C. and Pepple, P. A. (1989). Monitoring clinical trials based on predictive probability of significance. *Biometrics* **45,** 317–323.

Cornfield, J. (1966a). Sequential trials, sequential analysis and the likelihood principle. *The American Statistician* **20,** 18–23.

Cornfield, J. (1966b). A Bayesian test of some classical hypotheses—with application to sequential clinical trials. *Journal of the American Statistical Association* **61,** 577–594.

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

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

Grieve, A. P. (1988). Some uses of predictive distributions in pharmaceutical research. In *Biometry—Clinical Trials and Related Topics*, T. Okuno (ed.). Amsterdam: Elsevier.

Herson, J. (1979). Predictive probability early termination plans for Phase II clinical trials. *Biometrics* **35,** 775–783.

Ho, C. H. (1991). Some frequentist properties of a Bayesian method in clinical trials. *Biometrical Journal* **33,** 735–740.

Lindley, D. V. and Phillips, L. D. (1976). Inference for a Bernoulli process (a Bayesian view). *The American Statistician* **30,** 112–119.

Mehta, C. R. and Cain, K. C. (1984). Charts for the early stopping of pilot studies. *Journal of Clinical Oncology* **2,** 676–682.

Meier, P. (1975). Statistics and medical experimentation. *Biometrics* **31,** 511–529.

Novick, M. R. and Grizzle, J. E. (1965). A Bayesian approach to the analysis of data from clinical trials. *Journal of the American Statistical Association* **60,** 81–96.

Racine, A., Grieve, A. P., Fluhler, H., and Smith, A. F. M. (1986). Bayesian methods in practice: Experiences in the pharmaceutical industry (with Discussion). *Applied Statistics* **36,** 93–150.

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

Spiegelhalter, D. J. and Freedman, L. S. (1986). A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. *Statistics in Medicine* **5,** 1–13.

Spiegelhalter, D. J. and Freedman, L. S. (1988). Bayesian approaches to clinical trials (with Discussion). In *Bayesian Statistics 3*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (eds), 453–477. Oxford: Clarendon Press.

Sylvester, R. J. (1988). A Bayesian approach to the design of Phase II clinical trials. *Biometrics* **44,** 823–836.

Sylvester, R. J. and Staquet, M. J. (1977). Decision theory and Phase II clinical trials in cancer. *Cancer Treatment Reports* **64,** 519–524.

Sylvester, R. J. and Staquet, M. J. (1980). Design of Phase II clinical trials in cancer using decision theory. *Cancer Treatment Reports* **64,** 519–524.

Thall, P. F. and Simon, R. (1990). Incorporating historical control data in planning Phase II clinical trials. *Statistics in Medicine* **9,** 215–228.

Received March 1992; *revised August* 1992, *and January and March* 1993; *accepted March* 1993.

Appendix

The probability distribution of N, $p_+ = \Pr[\text{E is declared promising}]$, and $p_- = \Pr[\text{E is declared not promising}]$ may be computed under the assumption of beta priors as follows. At stage n define $L_n = -1$ if it is impossible to declare E not promising, and $U_n = n + 1$ if it is impossible to declare E promising, e.g., if $n < n_{\min}$. Recall that the beta pdf and cdf are denoted $f(p;\, a,\, b) = p^{a-1}(1-p)^{b-1}/B(a,\, b)$ and $F(p;\, a,\, b) = \int_0^p f(u;\, a,\, b)\, du$, for $0 \leq p \leq 1$, $a > 0$, and $b > 0$. Each of the following results holds for these values of p, a, and b.

Lemma. (a) $F(p; a, b) \geq F(p; a + 1, b)$, and (b) $F(p; a, b) \leq F(p; a, b + 1)$.

Proof. (a) This is an immediate consequence of the decomposition

$$F(p; a, b) = \frac{\Gamma(a + b)}{\Gamma(a + 1)\Gamma(b)} p^a(1 - p)^b + F(p; a + 1, b),$$

given as 26.5.16 by Abramowitz and Stegun (1965), since the first term is greater than or equal to 0. For part (b), $F(p; a, b) = 1 - F(1 - p; b, a)$ by 26.5.2 of Abramowitz and Stegun (1965), while $F(1 - p; b, a) \geq F(1 - p; b + 1, a)$ by part (a). Thus $F(p; a, b) \leq 1 - F(1 - p; b + 1, a) = 1 - \{1 - F(p; a, b + 1)\} = F(p; a, b + 1)$.

Denote $\xi(x, n) = 1 - \lambda(x, n; \pi_S, \pi_E, 0)$ for convenience, and note that $\lambda(x, n; \pi_S, \pi_E, 0) \geq p_U$ if and only if $\xi(x, n) \leq 1 - p_U = q_U$.

Theorem 1. If $\xi(x, n) \leq q_U$ then $\xi(x + 1, n + 1) \leq q_U$.

Proof.

$$\xi(x + 1, n + 1) = \int_0^1 f(p; a_S; b_S)F(p; a_E + x + 1, b_E + n - x)\, dp$$

$$\leq \int_0^1 f(p; a_S, b_S)F(p; a_E + x, b_E + n - x)\, dp \qquad (A.1)$$

by part (a) of the Lemma. Expression (A.1) equals $\xi(x, n)$, which implies the desired result.

Theorem 2. If $\xi(x, n + 1) \leq q_U$ then $\xi(x, n) \leq q_U$.

Remark 1. If $U_n \leq n$, then $\xi(U_n, n) \leq q_U$. Theorem 1 implies that $\xi(U_n + 1, n + 1) \leq q_U$. Since U_{n+1} is the smallest integer y such that $\xi(y, n + 1) \leq q_U$, it follows that $U_{n+1} \leq U_n + 1$, hence as $n \to n + 1$ the upper boundary increases by at most 1. If $U_n = n + 1$, then since $U_{n+1} \leq n + 2$, it follows that $U_{n+1} \leq U_n + 1$. Similarly, Theorem 2 implies that the upper boundary is monotone nondecreasing.

The proof of Theorem 2 is similar to that of Theorem 1, but uses part (b) of the lemma. For the lower boundary, denote $\eta(x, n) = 1 - \lambda(x, n; \pi_S, \pi_E, \delta_0)$ and $q_L = 1 - p_L$, so that $\lambda(x, n; \pi_S, \pi_E, \delta_0) \leq p_L$ if and only if $\eta(x, n) \geq q_L$. Theorem 3 is proved similarly to Theorems 1 and 2.

Theorem 3. If $\eta(x, n) \geq q_L$ then $\eta(x, n + 1) \geq q_L$ and $\eta(x - 1, n - 1) \geq q_L$.

Remark 2. Theorem 3 implies that the lower boundary is monotone nondecreasing and increases by at most 1 as $n \to n + 1$. Theorems 1–3 together establish the facts that $L_n = L_{n-1}$ or $L_{n-1} + 1$ and $U_n = U_{n-1}$ or $U_{n-1} + 1$ for $n > n_{min}$, i.e., each boundary either stays the same or increases by 1 as n increases by 1.

Denote $C_n = \{L_n + 1, \ldots, U_n - 1\}$. In particular, $C_n = \{0, \ldots, n\}$ for $n \leq n_{min}$. Denote the event that the trial has continued through the nth stage by $\mathcal{A}_n = \{X_j \in C_j, 1 \leq j \leq n\}$, the indicator of the set S by $I[S]$, and let $p = \Pr[Y_1 = 1] = 1 - q$. The respective probabilities that E is declared promising or not promising at stage n are $p_+(n) = \Pr[X_n \geq U_n \text{ and } \mathcal{A}_{n-1}]$ and $p_-(n) = \Pr[X_n \leq L_n \text{ and } \mathcal{A}_{n-1}]$, for $n_{min} \leq n \leq n_{max}$. Denote $\tau_n(x) = \Pr[X_n = x \text{ and } \mathcal{A}_{n-1}]$ for $x = 0, \ldots, n$ and $n \geq 2$, with $\tau_1(x) = p^x q^{1-x}$, $x = 0, 1$. The following recursion allows $\{\tau_n(x), L_n \leq x \leq U_n\}$ to be computed in sequence for each $n \geq 2$ using only p, L_n, U_n, and the values of $\{\tau_{n-1}(x), L_{n-1} \leq x \leq U_{n-1}\}$.

Theorem 4. $\tau_n(x) = p\tau_{n-1}(x - 1)I[x - 1 \in C_{n-1}] + q\tau_{n-1}(x)I[x \in C_{n-1}]$.

Proof of Theorem 4.

$$\tau_n(x) = \sum_{j=0}^{n-1} \Pr(X_n = x, X_{n-1} = j \text{ and } \mathcal{A}_{n-1})$$

$$= \Pr(X_n = x, X_{n-1} = x - 1 \text{ and } \mathcal{A}_{n-1}) + \Pr(X_n = x, X_{n-1} = x \text{ and } \mathcal{A}_{n-1}). \qquad (A.2)$$

The first probability in (A.2) equals

$$\Pr(X_n = x,\ X_{n-1} = x-1, \text{ and } \mathcal{A}_{n-2})I[x-1 \in C_{n-1}]$$

$$= \Pr(X_n = x | X_{n-1} = x-1 \text{ and } \mathcal{A}_{n-2})\tau_{n-1}(x-1)I[x-1 \in C_{n-1}]. \tag{A.3}$$

By the Markov property of the random walk $\{X_n,\ n \geq 1\}$, expression (A.3) equals

$$\Pr(X_n = x | X_{n-1} = x-1)\tau_{n-1}(x-1)I[x-1 \in C_{n-1}] = p\tau_{n-1}(x-1)I[x-1 \in C_{n-1}].$$

Likewise, the second probability in (A.2) equals $q\tau_{n-1}(x)I[x \in C_{n-1}]$, which proves the theorem.

Theorem 5. The probability of declaring E promising at stage n is

$$p_+(n) = \begin{cases} \Pr[X_n \geq U_n], & n = n_{\min}, \\ p\Pr[X_{n-1} = U_{n-1} - 1 \text{ and } \mathcal{A}_{n-2}]I[U_n = U_{n-1}], & n > n_{\min}. \end{cases}$$

The probability of declaring E not promising at stage n is

$$p_-(n) = \begin{cases} \Pr[X_n \leq L_n], & n = n_{\min}, \\ q\Pr[X_{n-1} = L_{n-1} + 1 \text{ and } \mathcal{A}_{n-2}]I[L_n = L_{n-1} + 1], & n > n_{\min}. \end{cases}$$

Proof of Theorem 5. The equalities are immediate for $n = n_{\min}$, since the trial cannot terminate prior to $n_{\min}$. Consider $n > n_{\min}$. By Theorem 4,

$$\Pr(X_n = L_n \text{ and } \mathcal{A}_{n-1}) = p\Pr(X_{n-1} = L_n - 1 \text{ and } \mathcal{A}_{n-2})I[L_n - 1 \in C_{n-1}]$$

$$+ q\Pr(X_{n-1} = L_n \text{ and } \mathcal{A}_{n-2})I[L_n \in C_{n-1}]. \tag{A.4}$$

By Remark 2, $L_{n-1} \geq L_n - 1$, and $L_n \in C_{n-1}$ if and only if $L_{n-1} = L_n - 1$. Thus $L_n - 1 \notin C_{n-1}$, and (A.4) reduces to

$$\Pr(X_n = L_n \text{ and } \mathcal{A}_{n-1}) = q\Pr(X_{n-1} = L_n \text{ and } \mathcal{A}_{n-2})I[L_n \in C_{n-1}]$$

$$= q\Pr(X_{n-1} = L_{n-1} + 1 \text{ and } \mathcal{A}_{n-2})I[L_n = L_{n-1} + 1]. \tag{A.5}$$

For $k \geq 1$,

$$\Pr(X_n = L_n - k \text{ and } \mathcal{A}_{n-1}) = p\Pr(X_{n-1} = L_n - k - 1 \text{ and } \mathcal{A}_{n-2})I[L_n - k - 1 \in C_{n-1}]$$

$$+ q\Pr(X_{n-1} = L_n - k \text{ and } \mathcal{A}_{n-2})I[L_n - k \in C_{n-1}]. \tag{A.6}$$

Again, since the lower boundary is monotone nondecreasing with jumps of at most 1 as $n - 1 \to n$, $L_{n-1} \geq L_n - 1$. Thus $L_n - k \notin C_{n-1}$ for any $k \geq 1$, and with (A.6) this implies

$$\Pr(X_n = L_n - k \text{ and } \mathcal{A}_{n-1}) = 0, \quad k \geq 1. \tag{A.7}$$

Since $p_-(n) = \Pr[X_n \leq L_n \text{ and } \mathcal{A}_{n-1}]$, it follows from (A.5) and (A.7) that

$$p_-(n) = q\Pr(X_{n-1} = L_{n-1} + 1 \text{ and } \mathcal{A}_{n-2})I[L_n = L_{n-1} + 1].$$

A similar argument holds for $p_+(n)$.

The operating characteristics are then simply

$$p_+ = \sum_{n=n_{\min}}^{n_{\max}} p_+(n) \quad \text{and} \quad p_- = \sum_{n=n_{\min}}^{n_{\max}} p_-(n),$$

and the distribution of N is obtained from the fact that

$$\Pr[N = n] = \begin{cases} p_+(n) + p_-(n), & n < n_{\max}, \\ \Pr[\mathcal{A}_{n-1}], & n = n_{\max}. \end{cases}$$