OPTIMAL GROUP SEQUENTIAL DESIGNS FOR THE ANSCOMBE–COLTON MODEL

A. JOHN PETKAU
University of British Columbia

Optimal group sequential designs are obtained for both the original and a truncated version of a selection model formulation of the problem of designing a clinical trial to compare two treatments. These optimal designs can be described as decision boundaries prescribed by nominal significance levels which vary dramatically, becoming less stringent as additional information becomes available. The corresponding Bayes risks indicate the magnitude of the penalty incurred due to both the restriction to group sequential designs and the imposition of varying degrees of truncation, and provide the baseline against which to assess the performance of the different types of group sequential designs that appear to be most commonly used in practice. Increasing the number of groups results in substantial improvements in the performance of the optimal designs, unless the point of truncation is quite small relative to the number of patients whose treatment will be determined by the results of the trial. In the latter circumstances, there is little to be gained through the choice of design; the primary design characteristic is the size of the trial. This result emphasizes the critical importance of planning clinical trials to be of adequate size. Our findings indicate conclusions concerning optimal group sequential designs which differ from those obtained within the usual hypothesis-testing framework and hence have implications for the general problem of designing such comparative clinical trials.

1. Introduction

In long-term clinical trials with sequential patient entry, strict application of fixed sample size designs is unjustified on ethical grounds but it is often argued that fully sequential designs are impractical due to the need for continuous assessment of the accumulating data; this might be particularly difficult to organize in the case of complex, multicenter trials. In such trials it is common practice to assess the accumulating data repeatedly at regularly spaced intervals of time. The planned use of group sequential designs has gained wide acceptance as a convenient approach to the challenge of monitoring such trials.

The problem of designing a clinical trial to compare two treatments usually is addressed from the hypothesis-testing point of view. As the two treatments are considered to be on a more-or-less equal footing, a three-decision formulation of the problem is most common: the null hypothesis of no difference is to be tested against each of two possible alternatives, typically symmetrically located around the null. Control of the error probabilities at the null and specified alternatives is the primary requirement and different designs are compared on the basis of expected sample sizes.

Keywords and phrases: Bayes risk, clinical trials, decision theory, interim monitoring, optimal designs, patient horizon, selection models, truncated designs..

Haybittle (1971) was the first to propose a particular class of group sequential designs in the context of monitoring clinical trials. He proposed carrying out a small number of interim analyses at a very stringent nominal significance level (0.001, say) so that, if no early stopping occured, the interim monitoring could effectively be ignored when carrying out the final analysis. This same general approach was later suggested by Peto et al. (1976). Pocock (1977) proposed a class of designs based on application of repeated significance tests with constant nominal significance levels. Motivated by the intuitive desirability of more stringent tests in the early stages of a trial, O'Brien and Fleming (1979) proposed a third class of designs having the form of a truncated version of the classical Wald sequential probability ratio test. More flexible approaches based on stochastic curtailment, error spending functions and repeated confidence intervals were introduced subsequently by Lan, Simon and Halperin (1982), Lan and DeMets (1983) and Jennison and Turnbull (1984), respectively. An expository discussion of all these approaches appears in Petkau (1996). A large body of literature is now available adapting these and related approaches for a great variety of circumstances; Jennison and Turnbull (2000) provide a comprehensive overview.

Some authors argue that a one-sided alternative is more appropriate for most clinical trials as continuation of a trial to determine whether a new therapy is inferior to a standard therapy may be inappropriate or even unethical; see Whitehead (1997), for example. DeMets and Ware (1980, 1982) and Whitehead and Stratton (1983) were the earliest to consider different types of group sequential designs for one-sided alternatives. The two-decision version of the problem is more closely related to the formulation to be considered in this paper, but the three-decision version is much more customarily employed for designing and monitoring clinical trials. Which version is preferable depends upon the nature and status of the two treatments to be compared in the trial, among other aspects.

Various investigators have considered the relative merits of different types of group sequential designs. Based on numerical evaluations, Pocock (1982) shows that the major reduction in average sample size under the alternative hypothesis for his class of designs is achieved by using a 2-group rather than a fixed sample size (1-group) design, and there is very little extra reduction in going from a 5-group to a 20-group design. He determines optimal 2-group and 5-group designs for the special case of normal responses with known variance and finds that the choice of constant nominal significance levels is near optimal for trials with reasonably large power. Pocock also compares his designs to those proposed by Peto et al. (1976) and O'Brien and Fleming (1979) and concludes that the designs with constant nominal significance levels are to be preferred, provided one is prepared to undertake a large

enough trial to ensure that alternatives of importance can be detected with reasonable power. By considering a more general class of designs containing both the Pocock and the O'Brien and Fleming classes, Wang and Tsiatis (1987) provide further detail supporting these general conclusions. Related work on determining optimal group sequential designs within the hypothesis-testing framework appears in Jennison (1987), Emerson and Fleming (1989) and Eales and Jennison (1992, 1995).

An alternate point of view considers the primary objective of a clinical trial to be optimization of the choice of treatment both for patients in the trial and those whose treatment will be determined by the results of the trial; this point of view has been formalized in various selection models. In this paper, we obtain optimal group sequential designs within the context of both the original and a truncated version of a simple selection model originally investigated by Anscombe (1963) and Colton (1963); in the truncated version of the model, experimentation is allowed to continue to at most a prespecified number of patients rather than to the patient horizon. These optimal designs can be described as decision boundaries prescribed by nominal significance levels which vary dramatically, becoming less stringent as additional information becomes available. The corresponding Bayes risks indicate that, except in particular circumstances, increasing the number of groups results in substantial improvements in performance. The exception occurs when the point of truncation is quite small relative to the number of patients whose treatment will be determined by the results of the trial. The Bayes risks indicate there is then little to be gained through the choice of design; in such circumstances the primary design characteristic is the size of the trial. These results emphasize, perhaps even more so than within the hypothesis-testing framework, the critical importance of planning clinical trials to be of adequate size.

The Anscombe–Colton model can be described as follows. There is a horizon of N patients to be treated with one of two treatments; the horizon should be thought of as the collection of all potential patients. In the initial experimental phase, n pairs of patients are treated sequentially, with different treatments randomly assigned to the patients in each pair. The differences, X_i, in the values of a continuous outcome variable for the ith pair are assumed to be independently and normally distributed with unknown mean μ and known variance σ^2 and to be instantaneously available after treatment. After n is selected by some sequential decision rule, the remaining $N - 2n$ patients are all assigned to the treatment which is inferred to be superior. The loss function measures patient benefit and has two components. The first is $n|\mu|$ which represents the loss in patient benefit incurred during the experimental period where n of the $2n$ patients are assigned to the inferior treatment, and the second is the loss in patient benefit of $(N - 2n)|\mu|$ which

is incurred if the inferior treatment is selected for the final stage. This loss function is intended to capture the essential feature of the problem of designing such a clinical trial: the trade-off between the goals of treating patients well and collecting information on the efficacy of the treatments.

An optimal solution exists for this sequential decision problem within a Bayesian framework when μ is given a prior normal distribution with mean μ_0 and variance σ_0^2. Chernoff and Petkau (1981) provide detailed analytic and numerical descriptions of the Bayes sequential design for a continuous time version of the problem and demonstrate that a simple "continuity correction" yields an excellent approximation to the Bayes sequential design for the discrete time version of the problem even for horizon sizes as small as 100. Related works of Begg and Mehta (1979), Petkau (1980), Lai, Levin, Robbins and Siegmund (1980), Lai, Robbins and Siegmund (1983), and Chernoff and Petkau (1985) concern various aspects of fully sequential designs for this sequential decision problem. In Section 3, we determine Bayes group sequential designs in this untruncated context where the clinical trial could possibly continue to the patient horizon without a decision being made.

Anscombe (1963, p. 371) states very clearly that in coming to this formulation of the problem he has made "every possible simplifying assumption" in order to clarify the issues involved; his hope was that this simple model might provide insight into general principles. In particular, he notes that the time and money cost of experimentation is ignored, but suggests that "a rough allowance for the time and money cost can be made by adding an extra condition to the design problem as already described, that n should not exceed a stated limit." That Anscombe did not pursue this suggestion has led some to criticize the entire formulation. For example, Armitage (1963, p. 386) suggests that in most situations, N would be very large and concludes: "It would in practice be almost certainly impossible for any individual, or existing organization, to carry out trials on the scale required." Such a practical constraint can be easily incorporated; Bayes sequential designs for a truncated version of this model are provided in Petkau (1987). In Section 4 we investigate Bayes group sequential designs in this context.

The results of Sections 3 and 4 facilitate comparison of optimal group sequential designs to optimal fully sequential designs in both the untruncated and truncated versions of the model. The form of these designs provides a qualitative indication of how such experiments should be carried out. The corresponding Bayes risks indicate the magnitude of the penalty incurred due to both the restriction to group sequential designs and the imposition of varying degrees of truncation. Possibly of greater general interest is the comparison of these optimal designs to the different types of group sequential designs that appear to be most commonly used in practice; some results

along these lines are presented in Section 5. The final section discusses the implications of our results for the general problem of designing clinical trials.

2. The reduced problem

This section establishes notation and sketches a reduction of the optimization problem to a more convenient form. For further detail, the reader is referred to Chernoff and Petkau (1981, 1986).

After observing the differences $X_1, \ldots, X_n$ corresponding to n successive pairs of patients, the posterior distribution of μ becomes $N(Y_n^*, s_n^*)$, where

$$Y_n^* = \left(\sigma_0^{-2}\mu_0 + \sigma^{-2}\sum_{i=1}^{n} X_i\right) \Big/ (\sigma_0^{-2} + n\sigma^{-2}), \quad s_n^* = (\sigma_0^{-2} + n\sigma^{-2})^{-1}.$$

As n increases, s_n^* decreases from $s_0^* = \sigma_0^2$ to $s_* = (\sigma_0^{-2} + 0.5N\sigma^{-2})^{-1}$; s_*^{-1} may be regarded as the total potential information for estimating μ. Considering μ as random, it is easily verified that as sampling continues, Y_n^* is a Gaussian process of mean-zero independent increments starting from the point $(y_0^*, s_0^*) = (\mu_0, \sigma_0^2)$.

With the choice $a^2 s_* = 1$, the transformation $Y_n = aY_n^*, s_n = a^2 s_n^*$, transforms the initial point $(y_0^*, s_0^*) = (\mu_0, \sigma_0^2)$ to (y_0, s_0), where

$$y_0 = \mu_0(\sigma_0^{-2} + 0.5N\sigma^{-2})^{1/2}, \quad s_0 = \sigma_0^2(\sigma_0^{-2} + 0.5N\sigma^{-2}),$$

and allows the determination of the Bayes sequential design to be reduced to the following optimal stopping problem: find the stopping time n to minimize the expected risk $E\{d(Y_n, s_n)\}$, where

$$d(y, s) = -(1 - s^{-1})|y|.$$

In this reduced problem, the parameters μ_0, σ_0, σ, and N enter only in the determination of the starting point (y_0, s_0) and the transformation back to the original scale. As n increases,

$$s_n = (\sigma_0^{-2} + 0.5N\sigma^{-2})/(\sigma_0^{-2} + n\sigma^{-2}),$$

decreases from s_0 to 1, while $t_n = s_n^{-1}$, the fraction of the total potential information obtained with n pairs, increases from $t_0 = s_0^{-1}$ to 1. In the limiting case of vague prior information, $s_n = N/2n$ and $t_n = 2n/N$.

The Bayes risk in the original problem corresponding to any stopping time is given by

$$\mathrm{BR} = \sigma^2\sigma_0^{-1}s_0^{1/2}[E\{d(Y_n, s_n)\} + 2(1 - s_0^{-1})s_0^{1/2}\Psi(y_0 s_0^{-1/2})],$$

where $\Psi(u) = \phi(u) + u\{\Phi(u) - 0.5\}$, and ϕ and Φ are the standard normal density and cumulative respectively. In addition, the Bayes expected number of pairs of patients included in the initial experimental phase is given by

$$\text{EP} = \sigma^2\sigma_0^{-2}s_0\{E(s_n^{-1}) - s_0^{-1}\}.$$

Optimal stopping rules can be determined by backward induction. A convenient way of representing these designs is in terms of

$$Z_n = Y_n/s_n^{1/2} = Y_n^*/s_n^{*1/2},$$

the number of standard deviations the current Bayes estimate of μ is away from zero, or in terms of $\beta_n = 1 - \Phi(|Z_n|)$, the nominal significance level for a one-tailed test of the hypothesis $\mu = 0$ based on the prior and the observed data. Symmetric designs can be specified by values $\tilde{z}_n$, or equivalently $\tilde{\beta}_n = 1-\Phi(\tilde{z}_n)$, such that, if n is a permissible stopping time, then the experimental phase is stopped after n pairs of patients have been treated if $|Z_n| \geq \tilde{z}_n$, or equivalently, if $\beta_n \leq \tilde{\beta}_n$; the remaining patients are treated according to the sign of the posterior mean. More explicitly, the decision criterion is

$$\sigma^{-1}\left|\sum_{i=1}^{n} X_i + \mu_0\sigma^2\sigma_0^{-2}\right| \geq (n + \sigma^2\sigma_0^{-2})^{1/2}\tilde{z}_n,$$

which becomes

$$\sigma^{-1}\left|\sum_{i=1}^{n} X_i\right| \geq n^{1/2}\tilde{z}_n,$$

in the limiting case of vague prior information.

3. Untruncated designs

Group sequential designs simply restrict the possible values of n where stopping is allowed. If, for example, the horizon of N patients is split into k groups of m pairs of patients ($2km = N$), then stopping is allowed only at values of $n = im$ ($i = 1, 2, \ldots, k$); of course, stopping is enforced at the kth stage when $n = km = 0.5N$. In the reduced problem this corresponds to stopping being allowed only at values s_{im} given by

$$s_{im} = (\sigma_0^{-2} + 0.5N\sigma^{-2})/(\sigma_0^{-2} + im\sigma^{-2}) = (1 + \gamma)/(1 + i\gamma/k),$$

where $\gamma = 0.5N\sigma_0^2\sigma^{-2} = s_0 - 1$; the parameter γ and the number of groups, k, completely specify the values of s_n at which stopping is permitted. For the sake of simplicity, only equal-sized groups are considered in this paper.

Optimal group sequential designs have been evaluated for the case of k equal-sized groups for various values of $s_0 = 1 + \gamma = 1 + 0.5N\sigma_0^2\sigma^{-2}$. The

Table 1. Optimal 10-group designs.[a]

	$t_0^{-1} = s_0 = 1 + \gamma = 1 + 0.5N\sigma_0^2\sigma^{-2}$					
Stage	10	10^2	10^3	10^4	10^5	10^6
0	0.920	1.135	1.166	1.169	1.170	1.170
1	0.762	0.871	0.884	0.885	0.885	0.885
2	0.647	0.719	0.727	0.728	0.728	0.728
3	0.550	0.601	0.607	0.607	0.607	0.607
4	0.464	0.501	0.505	0.505	0.505	0.505
5	0.383	0.411	0.413	0.414	0.414	0.414
6	0.305	0.324	0.326	0.327	0.327	0.327
7	0.225	0.238	0.239	0.239	0.239	0.239
8	0.137	0.145	0.145	0.145	0.145	0.145
9	0.000	0.000	0.000	0.000	0.000	0.000

[a]Tabulated values are $\tilde{z}$, the number of standard deviations required to stop the experimental phase.

calculations for the last two stages can be carried out explicitly: at stage $k-1$ when only one group (of $2m = N/k$ patients) remains to be treated, $\tilde{z}$ is identically zero (stopping is preferred unless $Z_n \equiv 0$, in which case one is indifferent between stopping and continuing to the k^{th} stage); at stage $k-2$ when two groups remain, $\tilde{z} = c(k-1+k/\gamma)^{-1/2}$, where $c = 0.436327\ldots$ is the positive solution of $\Psi(u) = u$. In tabulating these designs, the value of $\tilde{z}$ at stage 0 (where the experimental phase has not yet been initiated) is also of interest; for values of $|\mu_0/\sigma_0|$ in excess of this value of $\tilde{z}$, it would be optimal not to treat any patients in the experimental phase but to decide between the two treatments strictly on the basis of the prior information.

For a fixed number of groups, as s_0 increases these designs approach the optimal design for the limiting case of vague prior information. In fact, this convergence is quite rapid, as can be seen in Table 1 where the optimal 10-group designs are tabulated for an increasing sequence of values of s_0; the patterns for other numbers of groups were similar.

In most practical situations, $t_0 = s_0^{-1}$, the proportion of the total potential information in the prior, would be quite small. To the accuracy displayed in Table 1, the designs are essentially independent of t_0, provided $t_0 \leq 0.001$; for a fixed number of groups, the same design can be used in all situations provided only that t_0 is small enough. The optimal group sequential designs for k stages ($k = 2, 3, 4, 5$) corresponding to $t_0 = 10^{-6}$ are tabulated in Table 4, together with corresponding designs for the truncated version of the model to be considered in Section 4. Several of these designs are illustrated

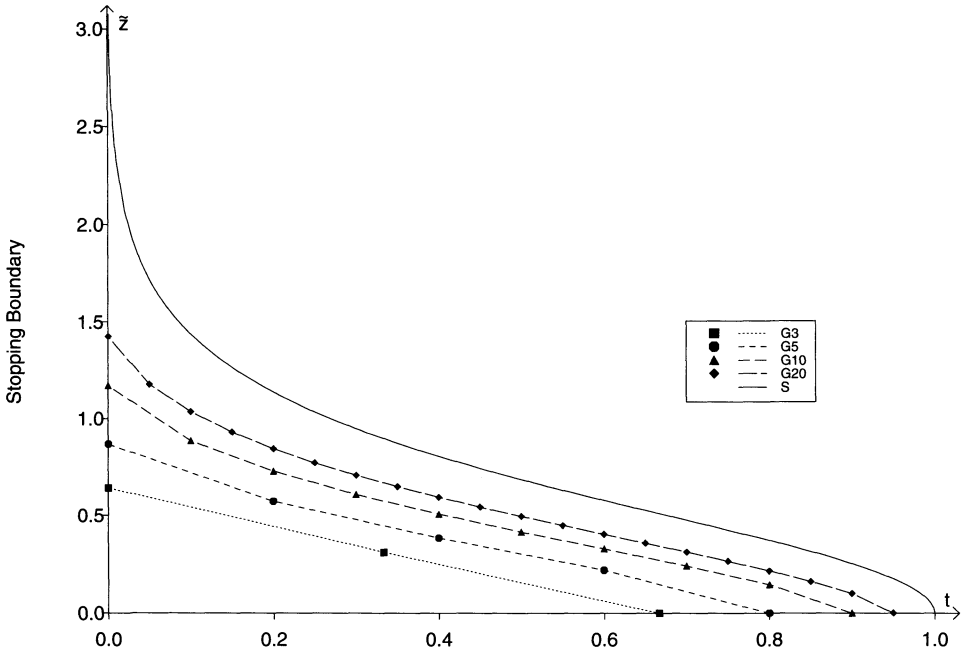

Figure 1. Optimal group sequential designs, G3, G5, G10 and G20, and the optimal fully sequential design, S. Here t is the currently available proportion of the total potential information.

in Figure 1 where, for the sake of comparison, the optimal fully sequential design is also presented.

These optimal designs prescribe nominal significance levels which vary dramatically, becoming less stringent as additional information becomes available. Further, the additional flexibility provided by a greater number of stages allows designs with a greater number of stages to require that the evidence in favour of one of the treatments become more convincing before the experimental phase is stopped and a decision is made.

The Bayes properties of these designs are of even greater interest than the stopping rules. The risk incurred and the expected number of pairs of patients treated in the initial experimental phase of the clinical trial can be evaluated as a function of

$$z_0 = y_0/s_0^{1/2} = y_0^*/s_0^{*1/2} = \mu_0/\sigma_0,$$

at the initial value $t_0^{-1} = s_0 = 1 + \gamma$, at stage 0 of the backward induction which determines the optimal design. We use the normalizations

$$\mathrm{BR} = \sigma^2\sigma_0^{-1}R, \quad \mathrm{EP} = \sigma^2\sigma_0^{-2}P,$$

and tabulate the values of R and P corresponding to $z_0 = 0$ for several of these designs in Tables 2 and 3 respectively. Also included for the sake of comparison are the optimal fully sequential and fixed sample size designs.

Table 2. Normalized Bayes risks for $z_0 = \mu_0/\sigma_0 = 0$.[a]

	$t_0^{-1} = s_0 = 1 + \gamma = 1 + 0.5N\sigma_0^2\sigma^{-2}$					
Design[b]	10	10^2	10^3	10^4	10^5	10^6
G2	3.93	39.89	398.9	3,989.0	39,890	398,900
G3	2.99	27.05	266.4	2,660.0	26,600	266,000
G4	2.56	20.72	200.3	1,996.0	19,950	199,500
G5	2.34	16.98	160.6	1,597.0	15,960	159,600
G10	1.99	9.84	81.8	799.9	7,981	79,790
G20	1.87	6.76	43.0	402.1	3,993	39,900
S	1.78	4.95	10.3	18.6	30	45
F	2.55	10.28	34.7	111.8	356	1,127

[a]Tabulated values are $R = \mathrm{BR}/\sigma^2\sigma_0^{-1}$, where BR is the Bayes risk, rounded to at most the leading four significant digits.

[b]Gk, S and F denote the optimal k-group sequential, fully sequential and fixed sample size designs respectively.

Table 3. Normalized Bayes Expected Sample Sizes for $z_0 = \mu_0/\sigma_0 = 0$.[a]

	$t_0^{-1} = s_0 = 1 + \gamma = 1 + 0.5N\sigma_0^2\sigma^{-2}$					
Design[b]	10	10^2	10^3	10^4	10^5	10^6
G2	4.50	49.50	499.5	5,000.0	50,000	500,000
G3	3.39	34.40	337.5	3,347.0	33,380	333,500
G4	2.85	26.91	256.7	2,521.0	25,070	250,200
G5	2.54	22.48	208.4	2,027.0	20,090	200,300
G10	2.03	14.02	113.1	1,042.0	10,130	100,400
G20	1.87	10.35	67.3	555.0	5,174	50,550
S	1.76	8.11	30.2	103.7	342	1,105
F	1.50	6.33	21.6	70.0	223	706

[a]Tabulated values are $P = \mathrm{EP}/\sigma^2\sigma_0^{-2}$, where EP is the Bayes expected number of pairs of patients in the experimental phase, rounded to at most the leading four significant digits.

[b]Gk, S and F denote the optimal k-group sequential, fully sequential and fixed sample size designs respectively.

A quick glance at Table 2 reveals that the Bayes risks of the optimal k-group sequential designs vary directly with the horizon size N. This contrasts with the behaviour of the optimal fixed sample size and fully sequential designs, for which the Bayes risks are of the order of $N^{1/2}$ and $(\log N)^2$ respectively. Since in the case $z_0 = 0$, the k-group design will include at least one group of $N/2k$ patients in the experimental phase, this behaviour is a direct consequence of the manner in which the group sequential designs

are defined; see the Bayes expected sample sizes tabulated in Table 3. In this untruncated context, these group sequential designs with equal-sized groups cannot possibly perform well for large values of the horizon size.

A more reasonable way to define group sequential designs in this untruncated context would be to choose a fixed total number of pairs of patients and consider sampling this number, rather than the entire horizon, in a group sequential manner with k equal-sized groups. Optimization could then be carried out over both the fixed total number of pairs of patients possibly included in the experimental phase and the strength of evidence required for stopping at each of the k stages. For $k = 1$, this is exactly the best fixed sample size design which, for large horizon sizes, performs substantially better than these group sequential designs although still quite poorly relative to the fully sequential design; see Table 2. Although in many contexts (roughly) equal-sized groups might have administrative advantages, a further generalization would remove this requirement and optimize over the group sizes as well. An elegant and detailed asymptotic investigation of such a formulation is provided by Hald and Keiding (1969, 1972); they establish that, for large horizon sizes, the optimal k-group design involves group sizes increasing in a very precisely specified fashion through successive stages. We do not pursue these more complex formulations here. Instead, in the next section, we incorporate the practical constraint of truncation on the amount of experimentation advocated by Anscombe (1963) and Armitage (1963).

4. Truncated designs

The constraint that n, the number of pairs of patients treated in the initial experimental phase, not exceed a specified number $M < N/2$ is easily incorporated. In the reduced problem, this simply means that s_T, the value of s_n at which stopping is enforced, is given by

$$s_T = (\sigma_0^{-2} + 0.5N\sigma^{-2})/(\sigma_0^{-2} + M\sigma^{-2}) = (1+\gamma)/(1+\eta),$$

where $\eta = M\sigma_0^2\sigma^{-2}$. Also denote $t_T = s_T^{-1}$, the proportion of the total potential information obtained upon reaching s_T, the point of truncation. In the limiting case of vague prior information, $s_T = N/2M$ and $t_T = 2M/N$.

Optimal group sequential designs for this truncated version can be obtained in the same manner as before. Suppose M, the maximum number of pairs of patients allowed in the experimental phase of the trial, is split into k groups of m pairs of patients ($km = M$). Then stopping is allowed only at the values of $n = im$ ($i = 1, 2, \ldots, k$); of course, stopping is enforced after the kth group when $n = km = M$. In this case

$$s_{im} = (\sigma_0^{-2} + 0.5N\sigma^{-2})/(\sigma_0^{-2} + im\sigma^{-2}) = (1+\gamma)/(1+i\eta/k),$$

and the parameters γ and η and the number of groups, k, completely specify the values of s_n at which stopping is permitted. For given values of γ, η and k, the optimal design can be determined by backward induction. As in the untruncated case, only equal-sized groups are considered.

Optimal truncated group sequential designs have been evaluated for the case of k equal-sized groups for various values of $s_T = t_T^{-1}$ and $s_0 = t_0^{-1}$. The calculations for the last stage can be carried out explicitly: at stage $k-1$ when only one group (of $2m = 2M/k$ patients) remains to be treated, $\tilde{z} = c_k(\gamma, \eta)(k + k/\eta)^{-1/2}$, where $c_k(\gamma, \eta)$ is the positive solution of $\Psi(u) = [1 + \eta/k(\gamma - \eta)]u/2$ and can be easily determined numerically.

With a fixed extent of truncation s_T and a fixed number of groups, as s_0 increases these designs approach the optimal design for the limiting case of vague prior information. This convergence is governed by the ratio t_0/t_T, the fraction of the information in the prior relative to that which would be obtained upon reaching the point of truncation. For all practical purposes, the same design can be used in all situations involving a fixed extent of truncation and a fixed number of groups, provided only that t_0/t_T is small enough; computational work indicates that $t_0/t_T \leq 0.001$ suffices. The optimal group sequential designs for k stages ($k = 2, 3, 4, 5$) corresponding to $t_0 = 10^{-6}$ are tabulated in Table 4.

The comparison of designs involving different numbers of groups but a fixed extent of truncation is qualitatively similar to the untruncated case. In particular, these designs also prescribe nominal significance levels that vary dramatically, becoming less stringent as additional information becomes available. A subset of these designs for the case $t_T = 10^{-4}$ are illustrated in Figure 2 in the (z, t') scale, where $t' = t/t_T$ represents the information currently available as a fraction of the information that would be available upon reaching the point of truncation. The optimal truncated fully sequential design is also included for the sake of comparison.

As should be anticipated, as the extent of truncation increases, the optimal k-group designs require that the evidence in favour of one of the treatments become more convincing before the experimental phase is stopped and a decision is made; this can be seen in Table 4 and is illustrated for 10-group designs in Figure 3. An alternate perspective would consider the optimal designs corresponding to increasing amounts of truncation as arising in problems with truncation at a fixed number of pairs of patients M, but increasing horizon sizes N. Table 4 and Figure 3 indicate that factors of ten in the magnitude of the patient horizon have a dramatic effect on the strength of evidence which should prompt a decision at any given stage of the experimental phase.

The risk incurred and the expected numbers of pairs of patients treated in the experimental phase of the clinical trial corresponding to $z_0 = 0$ when

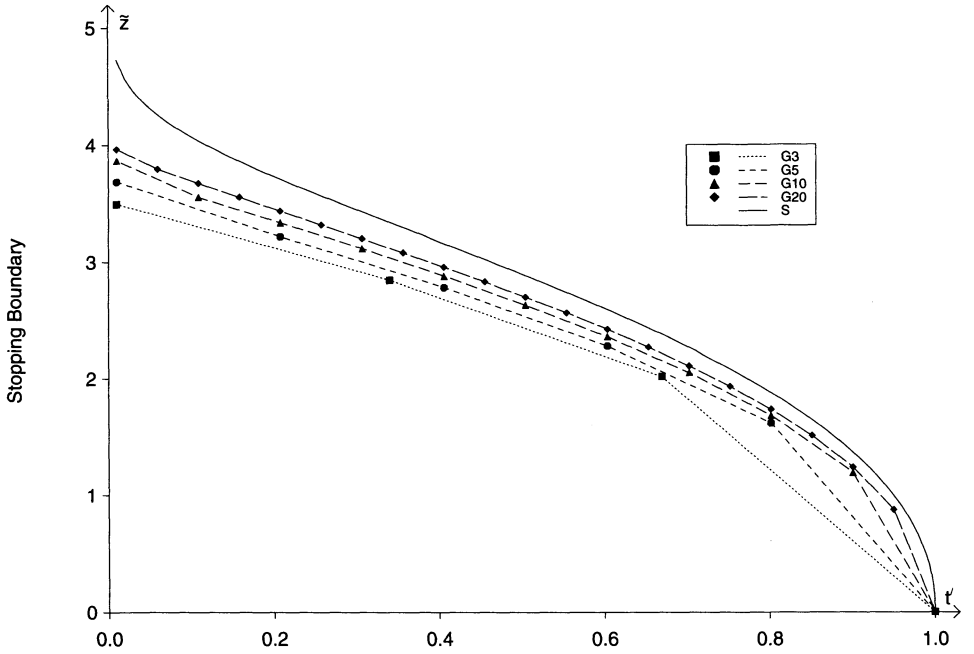

Figure 2. Optimal truncated designs for $t_T = 10^{-4}$. Here $t' = t/t_T$ is the proportion of information currently available relative to the information available upon reaching the point of truncation.

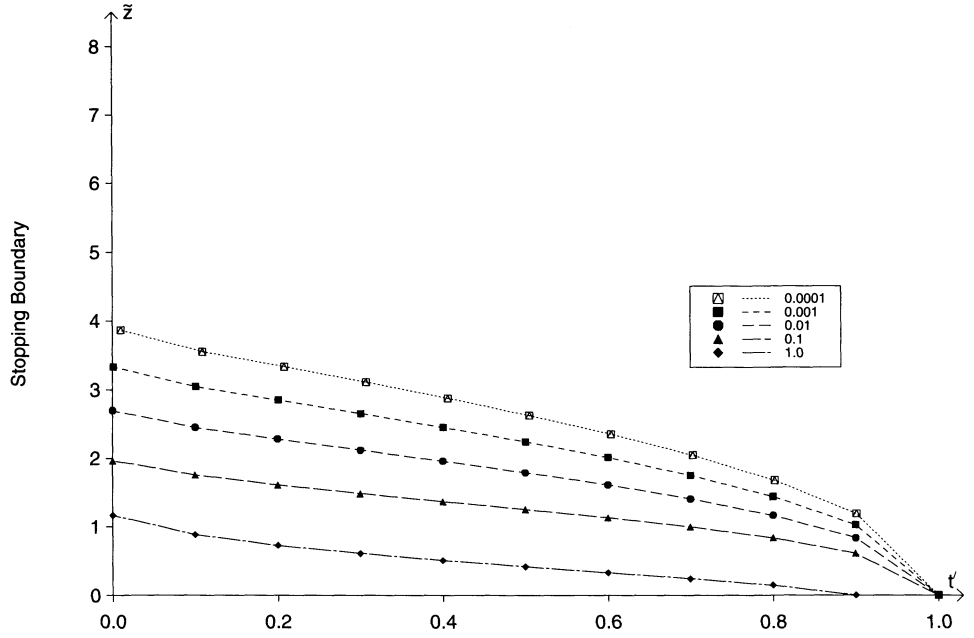

Figure 3. Optimal 10-group designs for differing extents of truncation t_T. Here $t' = t/t_T$ is the proportion of information currently available relative to the information available upon reaching the point of truncation.

Table 4. Optimal truncated group sequential designs.[a]

	$t_T = s_T^{-1} = (1+\eta)/(1+\gamma) = (1+M\sigma_0^2\sigma^{-2})/(1+0.5N\sigma_0^2\sigma^{-2})$				
	0.0001	0.001	0.01	0.10	1.00[b]
Stage			2-Group designs		
0	3.449	2.894	2.224	1.424	0.436
1	2.397	1.995	1.528	0.976	0.000
2	0.000	0.000	0.000	0.000	0.000
Stage			3-Group designs		
0	3.494	2.965	2.351	1.559	0.641
1	2.843	2.385	1.852	1.221	0.309
2	2.012	1.691	1.318	0.877	0.000
3	0.000	0.000	0.000	0.000	0.000
Stage			4-Group designs		
0	3.601	3.007	2.421	1.685	0.767
1	3.071	2.590	2.030	1.367	0.462
2	2.508	2.116	1.659	1.116	0.252
3	1.776	1.502	1.184	0.808	0.000
4	0.000	0.000	0.000	0.000	0.000
Stage			5-Group designs		
0	3.686	3.108	2.464	1.770	0.868
1	3.215	2.723	2.149	1.470	0.572
2	2.781	2.353	1.854	1.260	0.383
3	2.274	1.927	1.523	1.041	0.218
4	1.611	1.369	1.088	0.756	0.000
5	0.000	0.000	0.000	0.000	0.000

[a]Tabulated values are $\tilde{z}$, the number of standard deviations required to stop the experimental phase.

[b]Corresponds to the case of no truncation.

using these optimal designs are tabulated for a few combinations of values of $s_0 = 1+\gamma = 1+0.5N\sigma_0^2\sigma^{-2}$ and $s_0/s_T = 1+\eta = 1+M\sigma_0^2\sigma^{-2}$ in Tables 5 and 6 respectively. Note that if $\sigma_0 \approx \sigma$, for example, the implied values of M span a range of values often encountered in practice. Also included for the sake of comparison are the optimal truncated fully sequential and fixed sample size designs. The optimal fixed sample size design prescribes sampling to the point of truncation for large horizons and early truncation (see Table 3), but in general provides an improvement over a design involving a single group of M pairs of patients.

Table 5. Normalized Bayes risks for $z_0 = \mu_0/\sigma_0 = 0$.[a]

	Design[b]							
	G2	G3	G4	G5	G10	G20	S	F
$1+\eta = 1 + M\sigma_0^2\sigma^{-2}$	$s_0 = 1+\gamma = 1 + 0.5N\sigma_0^2\sigma^{-2} = 10^3$							
100	44.1	31.6	25.6	22.0	15.6	13.1	11.6	34.7
200	81.8	55.9	43.2	35.6	21.1	14.6	10.6	34.7
500	200.1	134.1	101.3	81.7	43.0	24.3	10.4	34.7
1000[c]	398.9	266.4	200.3	160.6	81.8	43.0	10.3	34.7
$1+\eta$	$s_0 = 1+\gamma = 10^4$							
100	80.7	68.6	62.9	59.6	53.8	51.7	50.4	111.8
200	100.4	74.9	62.5	55.2	41.5	35.7	32.3	111.8
500	207.8	142.3	109.8	90.5	52.6	34.9	22.5	111.8
1000	403.1	271.0	205.1	165.8	87.8	49.9	19.8	111.8
$1+\eta$	$s_0 = 1+\gamma = 10^5$							
100	441	430	424	421	416	414	413	479
200	281	256	244	237	224	219	215	358
500	280	215	183	164	127	110	99	356
1000	440	308	242	203	126	89	62	356
$1+\eta$	$s_0 = 1+\gamma = 10^6$							
100	4041	4030	4025	4022	4018	4016	4014	4078
200	2079	2054	2043	2036	2024	2019	2016	2156
500	999	935	903	884	848	832	822	1196
1000	799	668	603	564	488	452	426	1127

[a]Tabulated values are $R = \mathrm{BR}/\sigma^2\sigma_0^{-1}$, where BR is the Bayes risk.

[b]Gk, S and F denote the optimal k-group sequential, fully sequential and fixed sample size designs respectively.

[c]Corresponds to the case of no truncation; entries as in Table 2.

Examination of Table 5 reveals a rather complicated pattern reflecting the fundamental trade-off between knowingly treating half of the patients with the inferior treatment in the experimental phase of the clinical trial and the reliability of the final decision concerning the choice of treatment for the remaining patients in the patient horizon. The rows of Table 5 reveal that substantial improvements in performance can be achieved by increasing the number of groups, unless the point of truncation is quite small relative to the horizon size ($t_T \leq 0.001$ say). In the latter situations, the major contribution to the Bayes risk arises from the uncertainty in the final decision concerning the choice of treatment for the large number of patients remaining to be treated; even the fully sequential designs offer little improvement in

Table 6. Normalized Bayes expected sample sizes corresponding to $z_0 = \mu_0/\sigma_0 = 0$.[a]

	Design[b]							
	G2	G3	G4	G5	G10	G20	S	F
$1+\eta = 1 + M\sigma_0^2\sigma^{-2}$	$s_0 = 1+\gamma = 1 + 0.5N\sigma_0^2\sigma^{-2} = 10^3$							
100	54.9	41.0	34.4	30.6	23.7	21.0	19.3	21.6
200	105.7	75.4	60.7	52.1	35.9	28.9	24.4	21.6
500	255.2	175.0	135.5	112.0	66.3	45.0	29.3	21.6
1000[c]	499.5	337.5	256.7	208.4	113.1	67.3	30.2	21.6
$1+\eta$	$s_0 = 1+\gamma = 10^4$							
100	58.0	45.4	39.6	36.3	30.6	28.4	26.9	70.0
200	110.4	82.3	68.9	61.2	47.1	41.3	37.6	70.0
500	264.0	187.8	150.5	128.6	86.9	68.2	55.3	70.0
1000	516.9	358.7	280.7	234.4	144.6	102.5	70.7	70.0
$1+\eta$	$s_0 = 1+\gamma = 10^5$							
100	60.5	49.0	43.7	40.7	35.7	33.8	32.3	99.0
200	114.2	87.8	75.3	68.2	55.5	50.5	47.0	199.0
500	270.6	197.2	161.7	141.0	102.1	85.2	73.9	222.9
1000	526.7	373.0	297.7	253.4	168.1	129.2	101.0	222.9
$1+\eta$	$s_0 = 1+\gamma = 10^6$							
100	62.7	52.0	47.1	44.4	39.8	38.0	36.5	99.0
200	117.5	92.4	80.7	74.0	62.3	57.7	54.4	199.0
500	276.2	205.1	171.0	151.2	114.3	98.8	88.3	499.0
1000	535.0	384.8	311.6	268.7	186.8	150.1	124.2	706.4

[a]Tabulated values are $P = \mathrm{EP}/\sigma^2\sigma_0^{-2}$, where EP is the Bayes expected number of pairs of patients in the experimental phase.

[b]Gk, S and F denote the optimal k-group sequential, fully sequential and fixed sample size designs respectively.

[c]Corresponds to the case of no truncation; entries as in Table 3.

such situations. As should be anticipated, Table 6 indicates that decreasing Bayes expected sample sizes are associated with the additional flexibility of increasing numbers of groups.

The columns of these tables indicate the effects of varying amounts of truncation. While for a fixed number of groups and a fixed horizon size, the Bayes expected sample sizes increase as the number of pairs of patients at which truncation is enforced increases, the effect on the Bayes risks is not so simple. For the fully sequential design, increasing the point of truncation always provides additional flexibility and therefore results in a reduction in the Bayes risk. Reductions of similar magnitude are evident for the group

sequential designs provided that the point of truncation remains small relative to the horizon size. On the other hand, because the k-group sequential designs will include at least one group of M/k pairs of patients in the case $z_0 = 0$ under consideration, as the point of truncation becomes a reasonable fraction of the horizon size, the Bayes risk begins to be dominated by the contribution due to the experimental phase; in such situations, the Bayes risk will increase as the point of truncation increases.

The columns of these tables also indicate the effects of increasing horizon sizes on designs involving truncation after a fixed number of pairs of patients. While the Bayes expected sample sizes increase very slowly as the horizon size increases by factors of ten, the Bayes risk increases linearly with the horizon size when the point of truncation is quite small relative to the horizon size. In these situations, differences in the Bayes risks among the different truncated designs are minimal and the main point being made by the results in Table 5 is the importance of planning clinical trials to be of adequate size.

5. Comparison with standard group sequential designs

The results of Sections 3 and 4 indicate conclusions about group sequential designs differing from those resulting from a hypothesis-testing formulation of the problem of designing a clinical trial to compare two treatments. In particular, except for situations where the trial is restricted from the outset to involve a very small fraction of the horizon size, substantial improvements in performance result from increasing the number of groups. Further, these optimal designs prescribe nominal significance levels which vary dramatically, becoming less stringent as the amount of information available increases.

The standard group sequential designs considered here correspond to the three original proposals for the three-decision version of the problem of designing a clinical trial to compare two treatments. As mentioned in Section 1, the two-decision version is more closely related to the formulation considered in this paper. But the three-decision version seems to be the usual paradigm for designing and monitoring clinical trials, so the corresponding designs are the natural candidates for comparison.

In terms of $\tilde{z}_i$, the number of standard deviations required for stopping at the i^{th} stage, the three proposals can be described as follows:

HP (Haybittle and Peto et al.) $\tilde{z}_i = z_{\alpha'/2}$ for $i = 1, 2, \ldots, k-1$ and $\tilde{z}_k = z_{\alpha/2}$, where α' is the small nominal level prescribed for each interim analysis and α is the nominal level prescribed for the final analysis.

PK (Pocock) $\tilde{z}_i = c_{\mathrm{PK}}$ for $i = 1, 2, \ldots, k$, where c_{PK} is chosen to achieve the desired overall level α.

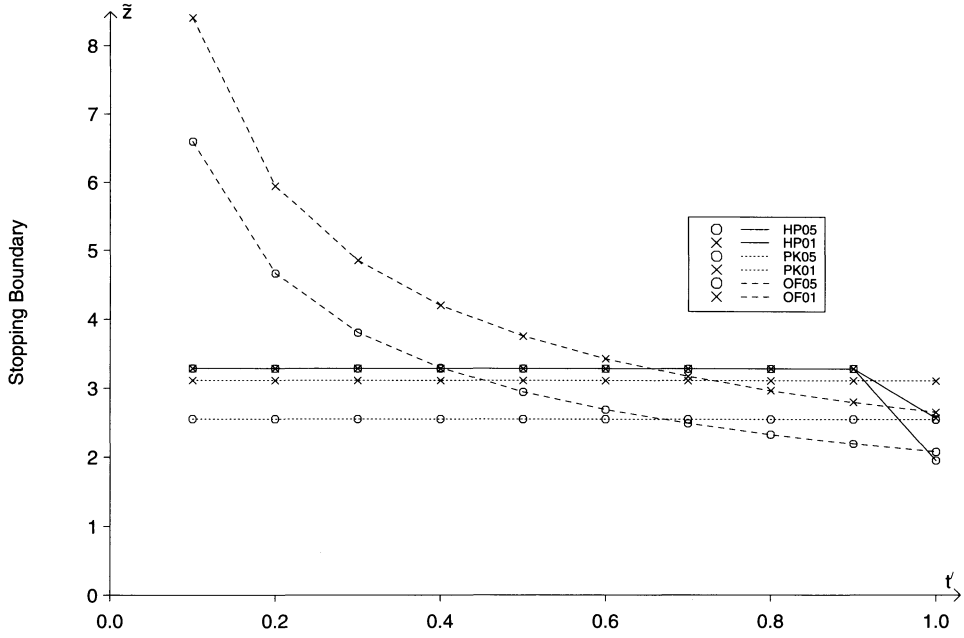

Figure 4. Standard 10-group designs: HP ($\alpha' = 0.001$), PK and OF for $\alpha = 0.05, 0.01$. Here $t' = t/t_T$ is the proportion of information currently available relative to the information available upon reaching the point of truncation.

OF (O'Brien and Fleming) $\tilde{z}_i = c_{OF}\sqrt{k/i}$ for $i = 1, 2, \ldots, k$, where c_{OF} is chosen to achieve the desired overall level α.

To facilitate direct comparison to Figure 3, specific versions corresponding to $\alpha' = 0.001$ and $\alpha = 0.05, 0.01$ for the case of 10 groups are illustrated in Figure 4. In this (z, t') scale, these designs depend upon t_0 and t_T only through the dependence on t_0/t_T of the values of t' at which stopping is allowed ($t' = i/k + (1 - i/k)t_0/t_T$ at the ith stage); the illustrations in Figure 4 are for $t_0/t_T = 10^{-6}$.

Comparison to Figure 3 suggests that except for quite severe truncation, these standard group sequential designs will tend to sample longer than the optimal group sequential designs. The Bayes expected sample sizes of these standard designs are the same for all horizon sizes (more precisely, for all values of $s_0 = 1 + \gamma = 1 + 0.5N\sigma_0^2\sigma^{-2}$) corresponding to a fixed extent of truncation (value of $s_0/s_T = 1 + \eta = 1 + M\sigma_0^2\sigma^{-2}$), and have been evaluated for all cases tabulated in Table 6. Substantial differences are evident among these designs but, for the cases considered, only the Pocock design with level $\alpha = 0.05$ ever yields smaller Bayes expected sample sizes than the optimal group sequential designs.

Of greater relevance is the Bayes risk incurred when using these designs. These standard group sequential designs result from a three-decision formu-

lation of the problem in which a choice between the two treatments is certain to be made only if termination occurs prior to the point of truncation. Yet, even if the evidence is not clear-cut when termination occurs at the point of truncation, within the formulation under consideration a choice must be made: the remaining $N - 2M$ patients must be treated with one of the two treatments. When evaluating the Bayes risks for these designs, the choice of treatment for the remaining patients is made according to the sign of the posterior mean, as this choice minimizes the Bayes risk. It follows that the Bayes risks for these designs depend only on the nominal levels employed for the interim analyses. Bayes risks for the case $s_0/s_T = 1+\eta = 1+M\sigma_0^2\sigma^{-2} = 100$ are tabulated in Table 7.

For the smaller tabulated values of the horizon size, these standard group sequential designs differ considerably in their Bayes risks for all but very small numbers of groups; the Pocock design with level $\alpha = 0.05$ is the clear choice. This design is always the quickest to stop (see Figure 4), but this feature becomes a liability for larger horizon sizes. In the case $s_0 = 10^6$, for example, the Bayes risk for this design actually increases with the number of groups (beyond 4) and it becomes the poorest choice. This reflects the relative uncertainty of an early choice between the treatments with this design; for large horizon sizes, an incorrect choice has serious consequences.

Comparisons among these standard group sequential designs are qualitatively similar for different extents of truncation (values of $s_0/s_T = 1+\eta = 1 + M\sigma_0^2\sigma^{-2}$). These results emphasize the importance of considering the magnitude of the patient horizon when designing clinical trials. All three proposals appear capable of yielding good designs but specific desirable choices of the nominal levels involved are unclear. Clearly these are related to the magnitude of potential patient benefit and the number of patients who might receive this benefit; compare the Pocock designs with levels $\alpha = 0.05$ and 0.01 in Table 7, for example. Unfortunately, it seems these specifications are often made in quite an arbitrary manner with attention restricted to historically justified values, such as 5% and 1%.

Horizons reasonably large relative to the point of truncation typically would be most relevant. There are clear differences among these standard group sequential designs for $s_0 = 10^3$ and 10^4 ($t_T = 0.1$ and 0.01) in Table 7. When the horizon is very large relative to the point of truncation, however, Table 7 indicates that (with the exception of the Pocock design with level $\alpha = 0.05$ which clearly performs worse) there is very little to choose among these standard group sequential designs. Further, in terms of Bayes risk, all are quite comparable to the optimal group sequential design. In such situations, primary attention should be focused on the size of the trial; compared to the substantial reductions in Bayes risk that are possible by reducing the extent

Table 7. Normalized Bayes risks for $z_0 = \mu_0/\sigma_0 = 0$: case $1 + \eta = 1 + M\sigma_0^2\sigma^{-2} = 100$.[a]

	Number of Groups					
	2	3	4	5	10	20
Design[b]	$s_0 = 1 + \gamma = 1 + 0.5N\sigma_0^2\sigma^{-2} = 10^3$					
HP	47.5	36.5	31.2	28.2	22.6	20.3
PK5	45.3	33.5	27.8	24.6	18.8	16.7
PK1	46.4	35.1	29.8	26.8	21.7	19.9
OF5	46.5	36.0	31.8	29.8	27.1	25.5
OF1	48.4	39.3	35.9	34.5	31.6	30.0
Design	$s_0 = 1 + \gamma = 10^4$					
HP	83.5	72.4	67.2	64.2	58.6	56.3
PK5	81.3	69.5	63.9	60.7	55.1	53.2
PK1	82.4	71.1	65.8	62.8	57.7	55.9
OF5	82.5	72.0	67.8	65.8	63.1	61.5
OF1	84.4	75.3	71.9	70.5	67.6	66.0
Design	$s_0 = 1 + \gamma = 10^5$					
HP	443	432	427	424	419	417
PK5	441	430	425	422	418	418
PK1	442	431	426	423	418	416
OF5	442	432	428	426	423	421
OF1	444	435	432	430	428	426
Design	$s_0 = 1 + \gamma = 10^6$					
HP	4043	4032	4027	4024	4019	4019
PK5	4042	4035	4035	4036	4050	4068
PK1	4042	4031	4026	4024	4020	4020
OF5	4042	4031	4027	4025	4023	4021
OF1	4044	4035	4031	4030	4027	4025

[a]Tabulated values are $R = BR/\sigma^2\sigma_0^{-1}$, where R is the Bayes risk.

[b]HP = Haybittle and Peto et al. proposal with $\alpha' = 0.001$, α arbitrary;
PK5, PK1 = Pocock proposal with $\alpha = 0.05, 0.01$;
OF5, OF1 = O'Brien and Fleming proposal with $\alpha = 0.05, 0.01$.

of truncation (see Table 5), differences among these designs are relatively negligible.

6. Discussion

The primary objective of this paper is to provide an evaluation of group sequential designs within an alternative to the usual hypothesis-testing formulation of the problem of designing a clinical trial to compare two treatments.

A very simple alternate formulation is investigated. While the results obtained may not immediately impact upon practice due to various practical aspects which may have been ignored in the model, the general insights provided by the results should stimulate critical thinking and contribute to the ongoing discussion between statisticians and clinical trialists about desirable schemes for the monitoring of clinical trials.

Determination of optimal group sequential designs for this model indicates the general form of desirable stopping rules within this context. Not surprisingly, these designs mimic the form of the optimal fully sequential design but are quicker to stop. Of the standard group sequential designs, only those corresponding to the O'Brien and Fleming proposal are of the same general form, with nominal significance levels varying dramatically, becoming less stringent as additional information becomes available. Evaluation of Bayes properties leads to different conclusions concerning the relative effectiveness of designs involving different numbers of groups than in the usual hypothesis-testing framework; substantial improvements in performance can be achieved by increasing the number of groups (even beyond 5), except in those situations where the point of truncation is very small relative to the horizon size. The standard group sequential designs perform somewhat poorly in this alternate formulation; their relative performance depends critically upon the horizon size. For horizon sizes that are very large relative to the point of truncation, all designs (except perhaps the Pocock design with level $\alpha = 0.05$) perform comparably and the results emphasize the importance of planning clinical trials to be of adequate size.

The results are obtained within the context of a specific model, but these group sequential designs can be employed in a broader setting. Patients need not be paired since the only requirement is that each consecutive group of patients has an equal number on each treatment; this is easily accomplished with treatment assignment by a randomized permuted block design, for example. In fact, provided only that the number of patients on each treatment within each group is reasonable, the performance of the group sequential designs should be relatively insensitive to minor variations in most aspects of the model, including the assumption of normal responses with known variance. On the other hand, the properties of these group sequential designs have been evaluated assuming that the stopping rule can be instantaneously applied. Although the decision to stop the trial need be considered only infrequently with group sequential designs, the same general difficulties arise as with fully sequential designs if response to treatment is not instantaneously available.

The alternate formulation of the problem of designing clinical trials under consideration has often been criticized; see Armitage (1985), for example. The primary criticism concerns this formulation's view of the clinical trial

as a selection procedure; it is argued that the purpose of Phase III trials is more diffuse. While the major purpose may be to assess the relative efficacy of the treatments for the primary response variable upon which the design is based, the recommendation to be made also depends on other factors such as relative efficacy for secondary response variables, side-effects, cost, ease of administration, etc. Further, different clinicians may evaluate these aspects differently, and therefore be led to different recommendations. Consequently, "It will usually be an over-simplification to regard it [the trial] either as merely a trigger for instant decisions or as the sole source of knowledge on which such decisions are based" (Armitage, 1985, p. 20).

But essentially the same criticism can be made of the hypothesis-testing formulation which presumes that the objective of the trial is to decide whether or not there exists an important difference in efficacy for the primary response variable. There is widespread recognition that the probability levels and critical difference involved in this formulation are arbitrary, and no statistician would suggest that rejection or acceptance at a specified level of significance constitutes a complete summary of the information gathered in the course of the trial. No formulation can hope to capture all the features of so complex an undertaking as a clinical trial. By focusing on a particular objective, both formulations provide simple frameworks for the planning of clinical trials.

The main elements of the alternate formulation are the loss function and the patient horizon. In the current problem a loss proportional to some monotonically increasing function of $|\mu|$ for each patient assigned to the inferior treatment seems natural, although asymmetry in the loss might be desirable in certain contexts when an experimental treatment is being compared to a standard. The designs obtained depend upon the particular loss function employed, but the same methodology could be employed to determine optimal designs for other loss functions. One advantage of this general approach is that it forces careful thought about the issues involved and, in particular, how patient benefit should be measured. Of course, the hypothesis-testing formulation also involves loss functions, but the simple 0–1 loss functions involved do not often appear to have been chosen subsequent to careful thought; rather the corresponding expected losses, the probabilities of error, are usually treated as the primary design characteristics.

The patient horizon is admittedly difficult to specify and, indeed, is also conceptually difficult (see Anscombe, 1963, p. 374; Armitage, 1985, p. 20). As already mentioned, this difficulty is closely-related to the difficulty of specifying the levels of Type I and Type II error in the classical approach. Peto (1985, p. 33) emphasizes that even when very clear evidence is published from a clinical trial, medical practice changes only gradually. He argues that models should take into account the impact that publication of particular

results will have. One possibility is to replace the concept of the patient horizon by the number of patients whose treatment will be determined by the results of the trial; this number might vary in some prescribed fashion, depending upon the "strength of evidence" supporting the decision at the time it was made.

The general issue of the appropriateness of a decision-theoretic approach to the problem of designing monitoring schemes for clinical trials has been much discussed; see, for example, the excellent discussion papers by Berry (1993), Whitehead (1993) and Spiegelhalter, Freedman and Parmar (1994). As already noted above, the debate centers around differing views on whether a clinical trial can be viewed as a decision procedure and on whether, in the clinical trials context, it is possible to adequately model a loss function that describes the consequences of all possible actions. As Carlin, Kadane and Gelfand (1998) and Stallard (1998) point out in their discussion of related problems, although it may be difficult to model the loss function accurately, the only alternative is to ignore the consequences of the decisions to be made and use a procedure that is optimal with respect to a loss function that has not been specified.

The fundamental issue concerns the identification of relevant criteria for the design of clinical trials. The hypothesis-testing formulation is useful to the experimenter who is primarily concerned with collecting information about the effects of treatments but ultimately decisions have to made putting this information to use. Schwartz, Flamant and Lellouch (1980) draw a distinction between explanatory and pragmatic types of trials, aimed at understanding and decision respectively. They suggest that significance tests serve a useful purpose in the former, but are less appropriate in the latter. While disagreements may be inevitable in certain situations, the issue merits serious consideration by all statisticians concerned with clinical trials.

Acknowledgements. This research was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada. The assistance of Brian Leroux with some of the computations and Rick White with the figures, as well as the helpful and stimulating comments of many colleagues and the referees are gratefully acknowledged.

REFERENCES

Anscombe, F.J. (1963). Sequential medical trials. *J. Amer. Statist. Assoc.* 58, 365–383.

Armitage, P. (1963). Sequential medical trials. Some comments on F.J. Anscombe's paper. *J. Amer. Statist. Assoc.* 58, 384–387.

Armitage, P. (1985). The search for optimality in clinical trials. *Internat. Statist. Rev.* 53, 15–24.

Begg, C.B. and Mehta, C.R. (1979). Sequential analysis of comparative clinical trials. *Biometrika* 66, 97–103.

Berry, D.A. (1993). A case for Bayesianism in clinical trials (with discussion). *Statist. Med.* 12, 1377–1404.

Carlin, B.P., Kadane, J.B. and Gelfand, A.E. (1998). Approaches for optimal sequential decision analysis in clinical trials. *Biometrics* 54, 964–975.

Chernoff, H. and Petkau, A.J. (1981). Sequential medical trials involving paired data. *Biometrika* 68, 119–132.

Chernoff, H. and Petkau, A.J. (1985). Sequential medical trials with ethical cost. In: *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II (L.M. LeCam and R.A. Olshen, eds.), pp. 521–538. Wadsworth, Belmont.

Chernoff, H. and Petkau, A.J. (1986). Numerical solutions for Bayes sequential decision problems. *SIAM J. Sci. Statist. Comput.* 7, 46–59.

Colton, T. (1963). A model for selecting one of two medical treatments. *J. Amer. Statist. Assoc.* 58, 388–400.

DeMets, D.L. and Ware, J.H. (1980). Group sequential methods for clinical trials with a one-sided hypothesis. *Biometrika* 67, 651–660.

DeMets, D.L. and Ware, J.H. (1982). Asymmetric group sequential boundaries for monitoring clincial trials. *Biometrika* 69, 661–663.

Eales, J.D. and Jennison, C. (1992). An improved method for deriving optimal one-sided group sequential tests. *Biometrika* 79, 13–24.

Eales, J.D. and Jennison, C. (1995). Optimal two-sided group sequential tests. *Sequential Anal.* 14, 273–286.

Emerson, S.S. and Fleming, T.R. (1989). Symmetric group sequential designs. *Biometrics* 45, 905–923.

Hald, A. and Keiding, N. (1969). Asymptotic properties of Bayesian decision rules for two terminal decisions and multiple sampling. I. *J. Roy. Statist. Soc. Ser. B* 31, 455–471.

Hald, A. and Keiding, N. (1972). Asymptotic properties of Bayesian decision rules for two terminal decisions and multiple sampling. II. *J. Roy. Statist. Soc. Ser. B* 34, 55–74.

Haybittle, J.L. (1971). Repeated assessment of results in clinical trials of cancer treatment. *Brit. J. Radiol.* 44, 793–797.

Jennison, C. (1987). Efficient group sequential tests with unpredictable group sizes. *Biometrika* 74, 155–165.

Jennison, C. and Turnbull, B.W. (1984). Repeated confidence intervals for group sequential clinical trials. *Controlled Clin. Trials* 5, 33–45.

Jennison, C. and Turnbull, B.W. (2000). *Group Sequential Methods with Applications to Clinical Trials.* Chapman & Hall/CRC, Boca Raton.

Lai, T.L., Levin, B., Robbins, H. and Siegmund, D. (1980). Sequential medical trials. *Proc. Nat. Acad. Sci. U.S.A.* 77, 3135–3138.

Lai, T.L., Robbins, H. and Siegmund, D. (1983). Sequential design of comparative clinical trials. In: *Recent Advances in Statistics. Papers in Honor of Herman Chernoff on his Sixtieth Birthday* (M.H. Rizvi, J. Rustagi and D. Siegmund, eds.), pp. 51–68. Academic Press, New York.

Lan, K.K.G. and DeMets, D. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* 70, 659–663.

Lan, K.K.G., Simon, R. and Halperin, M. (1982). Stochastically curtailed tests in long-term clinical trials. *Sequential Anal.* 1, 207–219.

O'Brien, P.C. and Fleming, T.R. (1979). A multiple testing procedure for clinical trials. *Biometrics* 35, 549–556.

Petkau, A.J. (1980). Frequentist properties of three stopping rules for comparative clinical trials. *Biometrika* 67, 690–692.

Petkau, A.J. (1987). Truncated sequential medical trials involving paired data. *Canad. J. Statist.* 15, 363–374.

Petkau, A.J. (1996). Statistical and design considerations for multiple sclerosis clinical trials. Chapter 4 in: *Multiple Sclerosis: Advances in Clinical Trial Design, Treatment and Future Perspectives* (D.E. Goodkin and R.A. Rudick, eds.), pp. 63–103. Springer-Verlag, London.

Peto, R. (1985). Discussion of papers by J.A. Bather and P. Armitage. *Internat. Statist. Rev.* 53, 31–34.

Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V., Mantel, N., McPherson, K., Peto, J. and Smith, P.G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. *Brit. J. Cancer* 34, 585–612.

Pocock, S.J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* 64, 191–199.

Pocock, S.J. (1982). Interim analyses for randomized clinical trials: the group sequential approach. *Biometrics* 38, 153-162.

Schwartz, D., Flamant, R. and Lellouch, J. (1980). *Clinical Trials.* Academic Press, London.

Spiegelhalter, D.J., Freedman, L.S. and Parmar, M.K.B. (1994). Bayesian approaches to randomized trials (with discussion). *J. Roy. Statist. Soc. Ser. A* 157, 357–416.

Stallard, N. (1998). Sample size determination for Phase II clinical trials based on Bayesian decision theory. *Biometrics* 54, 279–294.

Wang, S.K. and Tsiatis, A.A. (1987). Approximately optimal one parameter boundaries for group sequential trials. *Biometrics* 43, 193–200.

Whitehead, J. (1993). The case for frequentism in clinical trials (with discussion). *Statist. Med.* 12, 1405–1419.

Whitehead, J. (1997). *The Design and Analysis of Sequential Clincial Trials.* Revised 2nd edition. Wiley, Chichester.

Whitehead, J. and Stratton, I. (1983). Group sequential clincial trials with triangular continuation regions. *Biometrics* 39, 227–236.

A. John Petkau
Department of Statistics
University of British Columbia
333-6356 Agricultural Road
Vancouver, BC V6T 1Z2
Canada
john@stat.ubc.ca