Biometrika (1987), **74**, 1, *pp.* 149–54
Printed in Great Britain

Design and analysis of group sequential tests based on the type I error spending rate function

BY KYUNGMANN KIM AND DAVID L. DEMETS
Biostatistics Center, University of Wisconsin, Madison, Wisconsin 53706, U.S.A.

SUMMARY

Lan & DeMets (1983) devised a method of constructing discrete group sequential boundaries by using the type I error spending rate function. It is extended so as to generate asymmetric as well as symmetric two-sided boundaries for clinical trials. The design aspect of this procedure is explored in terms of the maximum sample size needed to achieve a desired level of power and the expected stopping times under both null and alternative hypotheses. Finally, these properties are employed in search of appropriate type I error spending rate functions for differing situations.

Some key words: Brownian motion process; Expected stopping time; First exit time; Maximum sample size; Survival study.

1. INTRODUCTION

Many large-scale clinical trials comparing two treatments, *A* and *B*, accumulate data during the course of the study and typically these data are analysed on a periodic basis. These trials may be characterized as one of two types. The first type measures the patient response to treatment relatively soon after entry. Thus, data are available on groups of patients where a group is defined as patients entered since the latest interim analysis. In all K such groups will be available. The second type of trial enters a cohort of patients over a relatively short period and follows them to the time of death or some nonfatal event. In this case, events are added between analyses and correspond to groups.

In order to preserve the type I error for repeated testing of accumulating data, Pocock (1977) modified earlier work of Armitage, McPherson & Rowe (1969) for trials of the first type. Basically, the jth group of size $2n$ is compared using some normalized statistic Z_j, where n is the number of patients in a group on each treatment. For example, $Z_j = \surd\{n/(2\sigma)\}(\bar{X}_{Aj} - \bar{X}_{Bj})$, where $\bar{X}$ represents the sample mean and σ the common standard deviation. Other examples are provided by Pocock (1977). Then $\bar{Z}_k = (Z_1 + \cdots + Z_k)/k$ is computed for the kth interim analysis for $k = 1, \ldots, K$ and it is compared to some critical value z_P^* which assures a type I error of size α; $\bar{Z}_K$ would represent a total sample of $2nK$ with no early termination. O'Brien & Fleming (1979) introduced a procedure which compared $\bar{Z}_k$ with $\surd(K/k)z_{\text{OBF}}^*$. Further work by Tsiatis (1982) applied these types of group sequential procedures for survival type trials.

While the above procedures are very useful, one constraint is that the number of groups K must be specified in advance. Another is that the groups must be of equal size, $2n$. A procedure by Lan & DeMets (1983) removes these requirements by selecting a 'type I error spending rate function' α^* in advance. Functions were provided which correspond roughly to versions of the Pocock and the O'Brien–Fleming boundaries and details were given for one-sided group sequential boundaries. They suggested using these for two-sided symmetric boundaries with appropriate adjustment for the overall significance level. In this paper, four areas are investigated: (i) further evaluation of three type I error spending rate functions suggested by Lan & DeMets (1983) in an asymmetric as well as symmetric setting plus two new functions: (ii) behaviour of these functions under three patterns of interim analyses, uniform, late and early; (iii) design aspects of the Lan-DeMets procedure; and, finally, (iv) appropriate choice of α^* for differing situations using some selected criteria.

2. EXTENSION OF THE LAN-DEMETS PROCEDURE

The procedure by Lan & DeMets (1983) can be easily extended for generating two-sided asymmetric as well as symmetric boundaries as follows: define α^* as before, i.e.

$$\alpha^*(t) = \text{pr}\,(\tau \leqslant t),$$

but with a different first exit time

$$\tau = \inf\{0 \leqslant t \leqslant 1\colon\ W_t > b_t \text{ or } W_t < c_t\}$$

for some boundary points $c_t < 0 < b_t$. Here W_t is a standard Brownian motion process. Then the upper and the lower boundary satisfy, respectively,

$$\text{pr}\,(W_1 > b_1) = \tfrac{1}{2}\alpha_U^*(t_1), \quad \text{pr}(W_1 < c_1) = \tfrac{1}{2}\alpha_L^*(t_1),$$

and

$$\text{pr}\,(c_1 \leqslant W_1 \leqslant b_1, \ldots, c_{k-1} \leqslant W_{k-1} \leqslant b_{k-1},\ W_k > b_k) = \tfrac{1}{2}\{\alpha_U^*(t_k) - \alpha_U^*(t_{k-1})\},$$
$$\text{pr}\,(c_1 \leqslant W_1 \leqslant b_1, \ldots, c_{k-1} \leqslant W_{k-1} \leqslant b_{k-1},\ W_k < c_k) = \tfrac{1}{2}\{\alpha_L^*(t_k) - \alpha_L^*(t_{k-1})\}.$$

In order to generate symmetric boundaries, simply choose $\alpha_L^* = \alpha_U^*$ and $c_k = -b_k$. A FORTRAN program is available from the authors. Also note that the t_k can be fixed and the corresponding b_k and c_k calculated as and when desired.

In this paper, the following α^*'s are studied:

$$\alpha_1^*(t) = 2\{1 - \Phi(z_{\alpha/2}/\surd t)\}, \quad \alpha_2^*(t) = 2[\tfrac{1}{2}\alpha \log\{1 + (e-1)t\}],$$
$$\alpha_3^*(t) = 2(\tfrac{1}{2}\alpha t), \quad \alpha_4^*(t) = 2(\tfrac{1}{2}\alpha t^{3/2}), \quad \alpha_5^*(t) = 2(\tfrac{1}{2}\alpha t^2).$$

Figure 1 shows how each α^* spends the type I error α. These α_1^*, α_2^* and α_3^* were previously studied by Lan & DeMets (1983). Here α_1^* and α_2^* are known to generate boundaries similar to the O'Brien-Fleming and the Pocock boundaries, respectively; α_3^*, α_4^* and α_5^* are intermediate to α_1^* and α_2^*. Further rationale for considering α_3^*, α_4^* and α_5^* will be discussed in § 5.

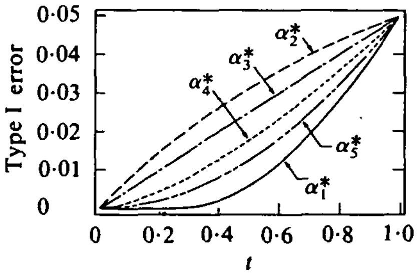

Fig. 1. Mode of spending type I error for α_i^* at $\alpha = 0{\cdot}05$ for $i = 1, 2, 3, 4, 5$.

Table 1. *Symmetric boundaries for $W_{t_k}/\surd t_k$ at $\alpha = 0{\cdot}05$*

(a) *Equal interval analyses*

$t^{(1)}$	0·2	0·4	0·6	0·8	1·0
α_1^*	4·89	3·36	2·68	2·29	2·03
α_2^*	2·44	2·43	2·41	2·40	2·39
α_3^*	2·58	2·49	2·41	2·34	2·28
α_4^*	2·85	2·59	2·43	2·29	2·18
α_5^*	3·09	2·72	2·47	2·28	2·11

(b) *Late analyses*

$t^{(2)}$	0·3	0·6	0·8	0·9	1·0
α_1^*	3·93	2·67	2·29	2·19	2·08
α_2^*	2·31	2·32	2·38	2·44	2·44
α_3^*	2·43	2·34	2·32	2·35	2·33
α_4^*	2·64	2·37	2·28	2·28	2·23
α_5^*	2·84	2·43	2·27	2·24	2·16

(c) *Early analyses*

$t^{(3)}$	0·1	0·2	0·3	0·6	1·0
α_1^*	7·02	4·89	3·93	2·67	1·98
α_2^*	2·66	2·63	2·59	2·35	2·28
α_3^*	2·81	2·74	2·67	2·36	2·18
α_4^*	3·17	2·94	2·80	2·38	2·10
α_5^*	3·49	3·16	2·95	2·43	2·05

Table1 gives typical symmetric boundaries for the α^*'s considered above for five interim analyses at the equal interval analyses, $t^{(1)} = \{0{\cdot}2, 0{\cdot}4, 0{\cdot}6, 0{\cdot}8, 1{\cdot}0\}$, the more frequent late analyses, $t^{(2)} = \{0{\cdot}3, 0{\cdot}6, 0{\cdot}8, 0{\cdot}9, 1{\cdot}0\}$, and the more frequent early analyses, $t^{(3)} = \{0{\cdot}1, 0{\cdot}2, 0{\cdot}3, 0{\cdot}6, 1{\cdot}0\}$. Note that the boundary values for α_1^*, α_2^* and α_3^* at $t^{(1)}$ are almost the same as the one-sided boundary values in Table 1 of Lan & DeMets (1983). The time patterns do substantially influence the boundaries for all the α^*'s. Note that, for $t^{(2)}$, α_2^* actually generates increasing boundary values. The first few boundaries for α_3^*, α_4^* and α_5^* are smaller but then become larger in the last interim analysis when compared with the boundaries for $t^{(1)}$.

In contrast, $t^{(3)}$ causes all the initial boundary values to be much larger than those for $t^{(1)}$. However, the boundary values for the last interim analysis are uniformly smaller over the α^*'s when compared with $t^{(1)}$. These results suggest that the more frequent early analysis pattern is quite conservative early on, but the last boundary value is small, even smaller than that of the equal interval analysis pattern. The frequent late interim analysis pattern requires a larger adjustment in the last boundary value than the others.

If one needs to specify asymmetric boundaries to account for an adverse side effect of a treatment (DeMets & Ware, 1982), one can use a pair of α^*'s; one for specifying the upper boundary and the other for the lower boundary. For example, if the side effect of a treatment is intolerable and if one wants to be conservative about the positive effect, one can use α_1^* for the upper boundary and α_2^* for the lower boundary.

3. Design aspects

Consider a sequence of independent normal variables $Z_1, Z_2, \ldots$ with unknown mean ζ and unit variance and let $S_k = Z_1 + \cdots + Z_k$. Now consider a group sequential test of $H_0: \zeta = 0$ based on some discrete group sequential boundaries and let N be the maximum sample size. In comparative clinical trials, each Z_j can be thought of as a difference between treatments. Note the following analogy between the partial sum S_n and the Brownian motion process W_t;

$$S_{n_k} \sim N(n_k\zeta, n_k), \quad W_{t_k} \sim N(t_k\xi, t_k),$$

Table 2. *Values of ξ attaining a power of* 0·90

	α_1^*	α_2^*	α_3^*	α_4^*	α_5^*
$t^{(1)}$	3·28	3·54	3·46	3·38	3·34
$t^{(2)}$	3·30	3·55	3·47	3·39	3·35
$t^{(3)}$	3·25	3·49	3·41	3·34	3·30

Table 3. *Expected stopping times when* $\alpha = 0{\cdot}05$ *and* $\beta = 0{\cdot}10$

	(a) *Under* H_0					(b) *Under* H_1				
	α_1^*	α_2^*	α_3^*	α_4^*	α_5^*	α_1^*	α_2^*	α_3^*	α_4^*	α_5^*
$t^{(1)}$	0·993	0·976	0·980	0·985	0·988	0·745	0·587	0·613	0·649	0·675
$t^{(2)}$	0·992	0·978	0·981	0·985	0·988	0·742	0·574	0·602	0·637	0·666
$t^{(3)}$	0·997	0·977	0·982	0·988	0·991	0·819	0·628	0·659	0·698	0·730

where n_k is the accumulated sample size by the kth interim analysis and t_k is the scaled process time of the kth interim analysis. If the test score is z_c, then the maximum likelihood estimators of the mean and the drift parameter are $\hat{\zeta} = z_c/\surd n_k$ and $\hat{\xi} = z_c/\surd t_k$ respectively. Therefore, $\hat{\xi}/\hat{\zeta} = \surd(n_k/t_k)$.

Suppose now that $n_k/t_k \equiv N$ for all k; that is, t_k's are proportional to the sample sizes. Given a boundary, one can determine numerically ξ which would attain a desired power, say, $1-\beta$, by solving for ξ the equation

$$1-\beta = \mathrm{pr}_\xi\,(\tau \leqslant 1) = \mathrm{pr}_\xi\,(W_t > b_t \text{ or } W_t < c_t \text{ for } t \in [0, 1]).$$

The values of ξ satisfying the above equation for the boundaries in Table 1 are given in Table 2. Therefore, the maximum sample size for each treatment in a clinical trial is determined by

$$N \equiv n_k/t_k = (\xi/\zeta)^2,$$

where ξ is obtained from Table 2 and ζ is the treatment difference to be detected.

To compute the expected stopping time, one can use the argument of DeMets & Ware (1980) as follows. Define the stopping time as $\tau^* = \tau$ if the boundaries are crossed and $\tau^* = 1$ otherwise. Then

$$E(\tau^*) = \sum_{k=1}^{K-1} t_k \,\mathrm{pr}\,(\tau = t_k) + t_K\{1 - \mathrm{pr}\,(\tau \leqslant t_{K-1})\} = 1 - \sum_{k=1}^{K-1} (1-t_k)\,\mathrm{pr}\,(\tau = t_k).$$

Table 3 gives the expected stopping times under the null hypothesis, $\xi = 0$, and under the alternative hypotheses, ξ's in Table 2. For all three interim analysis patterns, there is an order relation in the expected stopping times for α^*'s; that is, the more convex is α^*, as in Fig. 1, the larger is the expected stopping time. By using Tables 2 and 3, one can compute the expected sample size by $N^* = NE(\tau^*)$.

4. WHICH TYPE I ERROR SPENDING RATE FUNCTION TO USE?

Slud & Wei (1982) recommended that repeated significance tests be used only with a steadily increasing sequence of $\alpha^*(t_k) - \alpha^*(t_{k-1})$ so that the resulting boundary values $b_k/\surd t_k$ decrease. By using steadily decreasing boundary values, an embarrassing possibility of a nonsignificant result at t_{k+1} when, for some reason, a decision to stop the trial with a significant result at t_k was deferred, becomes less likely, though may not be completely avoided. Note that, for the pattern $t^{(2)}$, α_2^* and α_3^*, which are not strictly convex, generate nondecreasing boundaries. Hence, we recommend using strictly convex α^*. Both the Pocock and the O'Brien–Fleming-type boundaries are extreme. One alternative is to use convex combinations of them or to use α^*'s fitting in between them such as α_3^*, α_4^* or α_5^*; see Fig. 1.

Table 2 conveys information for appropriate choice of α^*. Note that the extreme α^* such as α_1^* can detect a smaller difference with the same power as compared to other α^*'s. This implies that a difference is more likely to be detected with a high level of power when boundaries are generated by an extreme α^*. This is one of the justifications for recommending the O'Brien–Fleming-type boundary.

One can also use the expected stopping time as a criterion for choosing appropriate α^*; see Table 3. If the short-term effect is of interest, then early stopping would be desirable. In such cases, one can choose α^* with a smaller expected stopping time or α^* that would generate boundaries similar to the Pocock boundaries so that there is an ample chance of an early rejection of no difference if indeed there is a real difference. If early stopping is unacceptable for some reason or if one wants to be conservative early on, one could use α^* that would generate boundaries similar to the O'Brien–Fleming boundaries.

The choice of α^* is complicated further because of the randomness of t_k's. Therefore, it is inevitably ad hoc in nature. One precaution in practice is that α^* should be chosen prior to the collection of data to avoid being criticized as 'data-dependent'.

5. Discussion

If the number and times of interim analyses are known, then α^* does not necesssarily have to be specified. As long as one knows how much of the overall significance level is to be spent between each interim analysis, one can generate boundaries by the same numerical procedure. In this regard, the Slud–Wei procedure is a special case of the Lan–DeMets procedure. The major distinction is that the Slud–Wei procedure deals directly with the actual process time and requires the number and the times of interim analyses predetermined.

In a survival analysis, the times of interim analyses are based on the number of deaths so that $t_k = d_k/D$, where d_k is the number of deaths observed by the kth interim analysis and D is the maximum number of deaths to be observed throughout the course of the study. Note that D should be chosen so as to achieve a certain level of power at the design stage of a study.

From any continuous sequential boundary b_t in $\tau = \inf\{0 \leqslant t \leqslant 1\colon W_t > b_t\}$, one can always find the corresponding discrete version of it via $\alpha^*(t) = \text{pr}(\tau \leqslant t)$. It would be interesting to characterize α^* corresponding to an arbitrary discrete group sequential boundary. We hope to pursue this issue later.

Acknowledgements

Comments from a referee were most helpful for the revision of this paper. All the computations were carried out on the Statistics Research Computer of the Department of Statistics, University of Wisconsin-Madison.

REFERENCES

ARMITAGE, P., MCPHERSON, C. K. & ROWE, B. C. (1969). Repeated significance tests on accumulating data. *J.R. Statist. Soc.* A **132**, 235-44.

DEMETS, D. L. & WARE, J. H. (1980). Group sequential methods for clinical trials with a one-sided hypothesis. *Biometrika* **67**, 651-60.

DEMETS, D. L. & WARE, J. H. (1982). Asymmetric group sequential boundaries for monitoring clinical trials. *Biometrika* **69**, 661-3.

LAN, K. K. G. & DEMETS, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70**, 659-63.

O'BRIEN, P. C. & FLEMING, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics* **35**, 549-56.

POCOCK, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64**, 191-9.

SLUD, E. V. & WEI, L. J. (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic. *J. Am. Statist. Assoc.* **77**, 862-8.

TSIATIS, A. (1982). Repeated significance testing for a general class of statistics used in censored survival analysis. *J. Am. Statist. Assoc.* **77**, 855-61.

[*Received April* 1985. *Revised April* 1986]