SIMULATION STUDIES ON INCREMENTS OF THE TWO-SAMPLE LOGRANK SCORE TEST FOR SURVIVAL TIME DATA, WITH APPLICATION TO GROUP SEQUENTIAL BOUNDARIES

Mitchell H. Gail

Biometry Branch, National Cancer Institute, Bethesda, Marvland

David L. DeMets

Mathematical and Applied Statistics Branch, National Heart, Lung
and Blood Institute, Bethesda, Maryland

Eric V. Slud

Department of Mathematics, University of Maryland, College Park, Md.

0. SUMMARY

The performance of the logrank statistic, computed after successive fixed numbers of deaths and applied to group sequential boundaries, is evaluated using simulation studies. The group sequential boundaries investigated include those proposed by Haybittle (1971), Pocock (1977), O'Brien and Fleming (1979) and the fixed sample boundary. The data indicate that a simple normal model, based on the assumptions that the increments of the logrank score are uncorrelated and homoscedastic with known variance, leads to reliable predictions of size, power, and average number of groups examined, except when the numbers at risk are very small, as in completely sequential entry. When there is a trend in the lifetime distribution, either in location or dispersion, the size of some group sequential boundaries exceeds nominal levels slightly, whereas the fixed sample logrank test is robust to such trends. The assumptions that the logrank increments are uncorrelated and homoscedastic with known variance are also investigated.

1. Introduction

Pocock (1977) proposed to monitor accumulating clinical trial data with group sequential boundaries which are appropriate for repeated analyses after successive groups of observations. After n groups of d observations each, the standardized statistic $T(nd) = (\sum_{i=1}^{n} \sum_{j=1}^{d} Y_{ij})(\sigma^2 nd)^{-\frac{1}{2}}$ is computed and compared with symmetric, two-sided group sequential boundaries c_n for $n=1,2,\ldots,N$. The null hypothesis is rejected at the smallest $n < N$ for which $|T(nd)| > c_n$. To compute boundaries of appropriate size, it is assumed that the group increments $\sum_{j=1}^{d} Y_{ij}$ are normally distributed, uncorrelated, and homoscedastic with known variance $\sigma^2 d$. Power is computed under the alternative that the group increments have mean δd.

Recently Pocock (1980) suggested that such boundaries could be applied to comparative survival studies by analyzing the standardized logrank statistic at intervals defined by equal numbers of deaths. This idea is closely related to suggestions by Armitage (1975, p. 143) and Jones and Whitehead (1979) for fully sequential analyses. Rigorous asymptotic theory in support of this proposal is available for two special cases, namely, progressive censorship, in which all patients enter simultaneously at the beginning of the experiment, and completely sequential entry, in which the lifetime of one patient is determined before the next enters the study. By referring to the permutational distribution of the linear rank statistic, Chatterjee and Sen (1973) showed that increments of the logrank score (numerator) are uncorrelated under progressive censorship, even for small samples. Asymptotically, their results imply that these increments are normally distributed and homoscedastic with known variance. Sen and Ghosh (1972) obtained these results for sequential entry.

The purpose of our simulations was to cover the intermediate case of staggered entry, which is of practical concern. Even for progressive censorship and sequential entry, simulations were useful to indicate the extent to which asymptotic theory applied. For staggered entry, Tsiatis (1981) has shown the increments to be asymptotically uncorrelated when the intervals are defined by fixed calendar times rather than by fixed numbers of deaths.

We report on the operating characteristics of group sequential boundaries proposed by Haybittle (1971), Pocock (1977) and O'Brien and Fleming (1979). In addition, we provide data on the correlation structure of the increments of the logrank score and test the hypotheses H1 that the increments are uncorrelated, H2 that the increments are uncorrelated and homoscedastic, and H3 that the increments are uncorrelated and homoscedastic with known variance d/4.

Some special studies were undertaken to determine whether group sequential procedures are robust to trends in the life distribution.

2. Methods

2.1 Definition of the Statistics and Boundaries

The computation of the two-sample logrank statistic is particularly simple in the case of continuous survival data (no ties) which we treat. As in Mantel (1966) and Cox (1972), we order the death times $t_1 < t_2 < \cdots < t_d$ to compute the logrank statistic after d deaths. Let p_k denote the proportion of all those patients known to have survived for time t_k or longer who are in group 1, and let $U_k = 1$ or 0 according as the death at t_k is in group 1 or 2. Then the logrank score after d deaths is

$$Z(d) = \sum_{k=1}^{d} (U_k - p_k) .$$

The estimated variance of Z(d) is

$$V(d) = \sum_{k=1}^{d} p_k(1 - p_k) \quad ,$$

which is only slightly less than d/4 in most cases with equal allocation, provided treatment effects are not too large.

After n groups (nd deaths), the statistic

$$T(nd) = Z(nd) \{V(nd)\}^{-\frac{1}{2}}$$

is computed and compared to symmetric two-sided group sequential boundaries c_n for $n=1,2,\ldots,N$, where N is the maximum number of groups to be entered. Tests proceed in the manner of Pocock (1977), with a rejection decision reached for the smallest n such that

$$|T(nd)| > c_n, \quad n=1,2,\ldots,N \quad .$$

We examined four symmetric two-sided size $\alpha = 0.05$ boundaries and studied the case $N=5$ in detail. For $N=5$, the Pocock (1977) boundary (P) is $c_n = 2.413$ for $n=1,2,\ldots,5$. The Haybittle (1971) boundary (H) is $c_n = 3.0$ for $n=1,2,3,4$ and $c_5 = 1.96$. This boundary is conservative and only detects extreme early treatment differences. The O'Brien-Fleming (1979) boundary (O), obtained from their Table 1, is $c_n = (4.149 \times 5/n)^{\frac{1}{2}}$ for $n=1,2,\ldots,5$. For the fixed sample boundary (F), $c_n = 100$ for $n=1,2,3,4$ and $c_5 = 1.96$. The value $c_n = 100$ was chosen for convenience and was never exceeded in our studies.

Suppose Z(nd) were the cumulative sum of nd independent Bernoulli variates corresponding to factors in the partial likelihood of Cox (1972), and that these factors were essentially unaffected by survival information obtained after death nd. Then under the null hypothesis of equality of survival distributions, increments such as $Z(2d) - Z(d)$ would have expectation zero, and correlation zero. With equal allocation, the variance of these increments is approximately d/4. For proportional hazards alternatives with hazard ratio $\exp(\theta)$, the expectation of such an increment would be approximately $d\theta/4$ for small θ. Thus, the power can be expressed in terms of the group non-centrality parameter

$$(1) \qquad \Delta = (d\theta/4)(d/4)^{-\frac{1}{2}} = \theta(d/4)^{\frac{1}{2}} \quad .$$

This quantity Δ is used for tabulations in Pocock (1977). Assuming the normal model holds, the theoretical size, power, and average number of groups $\bar{n}$ may be computed as in Armitage, McPherson and Rowe (1969), McPherson and Armitage (1971), and DeMets and Ware (1980). The theoretical variance of the stopping number n may also be determined from the multinomial distribution of the stopping points.

2.2 Description of the Simulation

The simulated clinical trials had a maximum of Nd = 90 deaths. This number of deaths was determined from equation (1) with N = 1 group so as to yield a power 0.90 for the two-sided 0.05 level logrank test with boundary F against the alternative of a two-fold relative hazard. Lininger, et al, (1979) confirm by simulations that equation (1) indeed yields the correct numbers of deaths required to attain specified power with the boundary F. The simulation proceeded by generating Poisson entry times for each patient up to a maximum of M = 90, 135 or 180 patients. The case M = 90 requires all patients to be followed to death. The case M = 180 allows the trial to stop when 90 patients are either still at risk or yet to be accrued, depending on the rate of entry. Larger values of M were not considered because, in the presence of rapid accrual, the decision would be reached on the basis of early deaths only, and in the case of slow accrual, increasing M beyond 180 has little effect. Each entered patient was then assigned a treatment using an independent Bernoulli variate (usually with equal allocation) and a lifetime (usually exponential). Exponential lifetimes and Poisson entry waiting intervals were generated with the IMSL subroutine GGEXN, and Bernoulli variates were based on the uniform IMSL pseudorandom numbers from GGUBFS. The IBM 370/OSVS was used. At the time when nd deaths occurred, the logrank statistic was calculated by resorting the follow-up times of all patients who had entered the trial to that time. The case of progressive censoring was studied by letting the Poisson entry rate get very large, and the case of sequential entry was studied by letting the entry rate get very small. Each experimental design was studied using 1000 independent simulations of the clinical trial. The proportion of rejections for the empirical estimates of size and power in Tables 2 and 3 are shown as the number of rejections per 1000 trials. Results for each boundary studied are correlated within each experimental design, but all statisitcs were independent across designs.

For Poisson entry experiments with $N=5$, the uncorrelated increments assumption (H1) is tested using the normal theory likelihood ratio statistic computed from formula 7, page 239 in Anderson (1958). The test of H2, the assumptions of uncorrelated homoscedastic increments, is given by equation 7 on page 261 in Anderson. The assumption H3 of independent increments with common variance $d/4$ is tested using equation 7, page 265 in Anderson.

3. Results

3.1 Theoretical Properties of the Group Sequential Boundaries Based on the Normal Model

The results of Table 1 were computed under the normal model using the numerical methods described by DeMets and Ware (1980). The noncentrality parameter Δ was computed from (1) with $d=18$, $N=5$, and relative hazard $\exp(\theta)=2$. The power of H exceeds that of F only because H has size 0.053, slightly in excess of 0.05. The Pocock boundary offers the greatest average savings in n but also has the least power.

TABLE 1. Theoretical properties of four group sequential boundaries with $N=5$

	Haybittle (H)	Pocock (P)	Fixed (F)	O'Brien-Fleming (O)
Null case $\Delta=0$				
size	0.053	0.050	0.050	0.050
$\bar{n}$	4.977	4.876	5.000	4.964
SD(n)	0.268	0.622	0.000	0.241
Alternative $\Delta=1.470$				
power	0.909	0.845	0.907	0.901
$\bar{n}$	3.864	3.083	5.000	3.648
SD(n)	1.313	1.441	0.000	0.989

3.2 Null Case Results with $N = 5$

The null case data in Table 2 are generated using unit exponential lifetimes in both treatment groups. Most, but not all, experimental conditions in Table 2 were studied in two independent simulations. These data give no evidence against the uncorrelated increments assumption H1. Evidence against homoscedasticity (H2) and/or known common variance equal to d/4 (H3) is seen when patients enter sequentially (entry rate 0.001 per year) and when only $M = 90$ patients are admitted. Both these cases require that observation continue until the last patient dies. These are, of course, situations in which p_k may deviate from 0.5 and in which incremental variances $V(nd) - V\{(n-1)d\}$ may deviate markedly from d/4. The smaller variance which results when p_k deviates from 0.5 may account for the slight but consistent elevations in size above nominal levels observed for sequential entry. This holds for all boundaries but is especially pronounced for the Pocock boundary P which has average size 0.069 for sequential entry. The deviation $\delta\bar{n}$ of the average number of groups $\bar{n}$ from predicted is shown for the Pocock boundary. The observed $\bar{n}$ for the Pocock boundary is in good agreement with the predicted value 4.876 except for the case of sequential entry where $\bar{n}$ is slightly smaller than predicted. To summarize, these null experiments are consistent with the uncorrelated increments assumption, and, except for cases when p_k may deviate markedly from 0.5, the homoscedasticity and known variance assumptions are also tenable. The size and average sample number of these experiments are consistent with theory based on the normal model except for minor discrepancies in the case of sequential entry.

3.3 Non-Null Results with $N = 5$

The siutation is different under the alternative hypothesis with exponential lifetime hazards 2.0 and 1.0 in the two treatment groups. The nested hypotheses H1, H2 and H3 are often violated.

The non-null operating characteristics of the group sequential boundaries are detectably different from predictions of the normal model, but the effects are not gross and not of practical importance. The power of the Pocock

TABLE 2. Simulations with N = 5 based on 1000 repetitions for each experiment

NULL CASE

	Progressive Censoring 100,000/Year						Fast Staggered Entry 100/Year					
Total				$H^{\dagger}$	P	Pocock				H	P	Pocock
Patients M	H1	H2	H3	0	F	$\overline{\delta n}$**	H1	H2	H3	0	F	$\overline{\delta n}$
90	12	36*	38*	59 52	53 56	-.003	7	13	13	58 59	53 54	.002
90	16	39*	46*	55 46	52 49	-.023	7	30*	38*	59 57	53 58	.008
135	10	14	14	47 50	49 45	.011	15	18	21	53 46	49 49	.006
135	11	16	19	47 45	45 44	.006	9	15	21	37 37	42 33	.005
180	10	14	17	40 39	48 38	-.008	6	10	17	51 49	55 47	-.028
180	6	6	8	45 40	44 43	.016	8	12	14	54 51	43 52	.012

NULL CASE

	Slow Staggered Entry 10/Year						Sequential Entry 0.001/Year					
Total				H	P	Pocock				H	P	Pocock
Patients M	H1	H2	H3	0	F	$\overline{\delta n}$	H1	H2	H3	0	F	$\overline{\delta n}$
90	4	13	23	51 51	57 47	-.022	14	20	24	67 59	78 59	-.089
90	9	19	32*	57 57	63 52	-.039	10	29*	31*	73 54	74 63	-.094
135	7	22	32*	49 48	48 42	-.011	11	27*	32*	63 55	67 56	-.045
135	2	16	31*	49 51	51 46	-.013	7	36*	44*	67 66	64 56	-.050
180	3	9	19	46 39	60 42	-.034	4	27*	35*	67 62	69 62	-.063
180	6	14	37*	53 56	52 47	-.021	7	18	26*	67 57	63 56	-.059

Table 2 (continued)

RELATIVE HAZARD 2

Total				H	P	Pocock				H	P	Pocock
	Progressive Censoring 100,000/Year						Fast Staggered Entry 100/Year					
Patients	H1	H2	H3	0	F	$\delta\bar{n}$	H1	H2	H3	0	F	$\delta\bar{n}$
90	152*	441*	459*	889 875	811 889	.066	168*	382*	413*	889 891	813 889	.038
90	179*	531*	542*	877 863	794 875	.087	73*	298*	316*	884 869	811 882	.058
135	17	24*	29*	908 896	841 908	.010	9	15	17	896 890	842 895	.088
135	13	25*	28*	909 895	851 906	.012	17	24*	26*	905 901	837 904	.056
180	16	22	32*	900 891	828 900	.162	3	10	11	898 883	819 897	.102
180												

RELATIVE HAZARD 2

Total				H	P	Pocock				H	P	Pocock
	Slow Staggered Entry 10/Year						Sequential Entry 0.001/Year					
Patients	H1	H2	H3	0	F	$\delta\bar{n}$	H1	H2	H3	0	F	$\delta\bar{n}$
90	38*	63*	162*	889 883	811 886	.177	12	22	284*	892 876	800 890	.234
90	35*	52*	144*	884 863	812 880	.082	27*	37*	264*	892 887	825 891	.163
135	49*	50*	123*	915 905	841 913	.016	13	16	240*	891 883	818 888	.209
135	34*	38*	80*	899 877	830 888	.165	13	15	247*	882 875	822 881	.114
180	40*	44*	93*	900 882	816 896	.099	12	21	321*	892 882	835 889	.075
180	53*	57*	168*	899 892	805 899	.207						

* Exceeds the 95th percentile of the corresponding chi-square distribution. The degrees of freedom are 10 for H1, 14 for H2 and 15 for H3.

† Size and power estimates are given for each boundary as the number of rejections per 1000 repetitions.

**The quantity $\delta\bar{n}$ is the deviation of the average number of increments, $\bar{n}$, from expectation. For the null case, $\delta\bar{n} = \bar{n} - 4.876$, and for the relative hazard 2, $\delta\bar{n} = \bar{n} - 3.083$.

boundary is less than the theoretical value 0.845 in every experiment but one and tends to decrease as the entry rate decreases. Nonetheless, the observed power is usually only about 3% less than predicted.

Under the two-fold hazard ratio alternative, the observed values of $\bar{n}$ for the Pocock boundary tend to exceed the predicted value 3.083, especially for entry rate 0.001 per year. These discrepancies range from 0.3 to 7.6% of predicted and are tolerable in practice. For the other boundaries, which have less potential for early stopping, the discrepancies are even smaller.

3.4 Miscellaneous Experiments

A few staggered entry experiments were conducted with $N = 10$, $d = 9$. These experiments conform to H1, H2 and H3 and to the normal model predictions even better than results in Table 2. Some experiments were performed with Weibull lifetimes. For shape parameter 3, null and non-null results were similar to those in Table 2 (entry rate 10/year). With shape parameter 1/3, null and non-null conformance to the normal model was better than indicated in Table 2.

3.5 Robustness Studies When There are Trends in the Life Distribution

The first eight experiments in Table 3 reflect the performance of group sequential boundaries with $N = 5$ and $M = 135$ when there is a time trend in the mean exponential life. The mean lifetime varies linearly with the patient entry index $\gamma = i/135$ for $i = 1,2,\ldots,135$. To obtain a simulated lifetime in the first experiment, a simulated unit exponential lifetime ℓ_0 is transformed to $\ell = (0.1 + 0.9\gamma)\ell_0$. Thus, the average exponential lifetime increases about 10 fold as i ranges from 1 to 135. Other trends are produced from $\ell = (0.5 + 0.5\gamma)\ell_0$, $\ell = (1.0 - 0.5\gamma)\ell_0$, and $\ell = (1.0 - 0.9\gamma)\ell_0$. Under the null hypothesis of no treatment effect, the lifetimes being compared are from the same mixed exponential population. With sequential entry, these null case experiments demonstrate that the increments are negatively correlated for increasing mean life trends and positively correlated for strongly decreasing mean life trends.

TABLE 3. Null case robustness studies with trends in the life distribution for N = 5, M = 135

				Size			
Description of the trend	H1	H2	H3	P	O	H	F
Linear trend in the mean exponential life							
Sequential entry (0.001/yr)							
10 fold increase	143*	370*	896*	82	68	69	55
2 fold increase	51*	135*	192*	71	48	53	40
2 fold decrease	12	33*	112*	67	55	58	51
10 fold decrease	42*	86*	317*	78	57	65	56
Staggered entry (10/yr)							
10 fold increase	122*	291*	690*	86	64	69	56
10 fold decrease	29*	56*	173*	52	48	55	50
Progressive censorship (10^5/yr)							
10 fold increase	6	9	9	50	54	51	48
10 fold decrease	5	9	9	38	47	52	52
Trend in dispersion with constant geometric mean							
Sequential entry (0.001/yr)							
Base 10 increase	69*	219*	363*	69	50	56	45
Base 2 increase	15	42*	44*	65	47	58	49
Base 1	6	16	26*	67	52	63	55
Base 2 decrease	15	20	82*	55	44	50	42
Base 10 decrease	100*	215*	554*	63	44	56	50
Progressive censorship (10^5/yr)							
Base 10 increase	5	15	17	56	63	62	66
Base 10 decrease	10	12	12	51	65	72	70
Trend in mean and dispersion of uniform lifetimes							
Sequential entry (0.001/yr)							
U (12, 12+10γ)	558*	1033*	3062*	99	77	64	49
U (12, 12-10γ)	258*	353*	990*	62	52	56	51

*Exceeds the 95th percentile of the corresponding chi-square distribution.

The size of the Pocock boundary exceeds 0.05 in these cases. These effects are still appreciable at staggered entry rate 10 per year, but they vanish for rapid entry, which is to be anticpated from the theory of progressive censorship applied to a mixed exponential population. Note that the fixed sample test F, and the conservative boundaries O and H, are more robust to such trends.

Seven experiments investigate the effects of a trend in dispersion with constant geometric mean. For the first of these, a unit exponential lifetime ℓ_0 is changed to $\ell = \ell_0 \times 10^{\gamma}$ with probability 1/3, to $\ell = \ell_0 \times 10^{-\gamma}$ with probability 1/3 and to $\ell = \ell_0$ otherwise. In this transformation, 10 is the "base" and $\gamma = i/135$ as before. This yields increasing dispersion. For decreasing dispersion, γ is replaced by $1 - \gamma$. The effects on size and correlation structure are smaller than for a trend in the exponential mean.

Rather dramatic effects on size and correlation structure are seen for trends in the lifetimes which are assumed to be uniform on the support interval indicated. The size of the Pocock boundary is 0.099 in one instance. Again, note the robustness of the fixed sample procedure F.

3.6 Some Null Case Experiments with Two Batches of Patients

We shall briefly mention the results of a number of null case experiments in which a first batch of 100 patients entered at time zero and a second batch of 100 patients entered at the time of death $d = 20$ in the first batch. Group sequential analyses were performed at deaths $d = 20$ and $d = 40$. In some experiments, batches consisted of 50 patients. The principal conclusions of these experiments were:

(1) Loss to follow-up, as might occur if a patient refuses further participation, does not affect size. Loss to follow-up was studied by assuming a constant hazard of withdrawal from the study after the patient is entered. It was found that size is not affected by unequal loss to follow-up in the two treatment groups. Again, size remains near 0.05 even if the loss to follow-up is adaptive in the sense that

the risk of loss to follow-up in the second batch is greater on the treatment which appeared worse at the first analysis. This latter situation could arise in practice if patients who enter learn which treatment appears to be more successful and adhere preferentially to the favored treatment.

(2) Adaptive allocation of 80% of the second batch to that treatment which appreared better (or worse) at the time of the first analysis does not affect size.

(3) Whether the second batch has longer or shorter mean lifetimes than the first batch, size remains near 0.05. If the second batch has a mean lifetime 10 fold greater than the first batch, the logrank score increments are negatively correlated. Interestingly, even if the second batch has a mean lifetime 1000 fold smaller than the first batch, the increments appear uncorrelated. Thus, distortions are more prominent when healthy patients enter later. This asymmetry is also seen in the first six experiments of Table 3. There too, a trend toward healthier patients induces stronger correlations among increments of the logrank score than does a trend toward sicker patients.

4. DISCUSSION

The correlation structure of the logrank score statistic conforms rather well to the normal model (H3) under the null hypothesis except for purely sequential entry. Departures from this simple correlation structure are evident under the alternative of a two-fold hazard ratio and when there are trends in the life distribution. It is not surprising, therefore, that theoretical calculations of operating characteristics work best under the null hypothesis. What is noteworthy is that these calculations are sufficiently accurate to plan experiments under the alternative. For any size α boundary $\{c_n\}$, one can calculate the non-centrality parameter required to obtain a desired power as in DeMets and Ware (1980) or Pocock (1977). Then the required number of deaths per increment is obtained from (1).

Robustness studies demonstrate that the size of the Pocock boundary slightly exceeds 0.05 when there are trends in the lifetime distribution. A typical empirical size is about 0.07. As expected, the boundaries O and H are less affected by such trends, and F is completely robust.

One can adhere to a group sequential plan by analyzing the data when prespecified numbers of deaths have occurred. If one plans to perform repeated analyses at fixed calendar times, instead, predetermined boundaries may turn out to be inappropriate because accrual and death rates are variable and hard to predict.

ACKNOWLEDGEMENTS

We wish to thank James H. Ware for helpful suggestions and Julie Paolella for typing this manuscript.

REFERENCES

Anderson, T.W. (1958). An introduction to multivariate statistical analysis. New York: Wiley.

Armitage, P. (1975). Sequential medical trials. New York; Wiley.

Armitage P., McPherson, C.K., and Rowe, B.C. (1969). Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society A* **132**, 235-244.

Chatterjee, S.K., and Sen, P.K. (1973). Non-parametric testing under progressive censoring. *Calcutta Statistical Association Bulletin* **22**, 13-50.

Cox, D.R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B* **34**, 187-220.

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

Habittle, J.L. (1971). Repeated assessment of results in clinical trials of cancer treatment. *British Journal of Radiology* **44**, 793-797.

Jones, D., and Whitehead, J. (1979). Sequential forms of the logrank and modified Wilcoxon tests for censored data. *Biometrika* **66**, 105-113.

Lininger, L., Gail, M.H., Green, S.B., and Byar, D.P. (1979). Comparison of four tests for equality of survival curves in the presence of stratification and censoring. *Biometrika* **66**, 419-428.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports* **50**, 163-170.

McPherson, C.K., and Armitage, P. (1971). Repeated significance tests on accumulating data when the null hypothesis is not true. *Journal of the Royal Statistical Society A* **134**, 15-25.

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

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

Pocock, S.J. (1980). Group sequential design for clinical trials. *Presented at the American Statistical Association Meetings in Houston*, August 11-14.

Sen, P.K. and Ghosh, M. (1972). On strong convergence of regression rank statistics, *Sankhya A* **34**, 335-348.

Tsiatis, A.A. (1981). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika* **68**, 311-315.