Submitted to Statistical Science

Group sequential designs with response-adaptive randomisation

Christopher Jennison

Abstract. Group sequential Phase III trial designs enable early stopping for positive or negative study outcomes. Response-adaptive randomisation can be included in such designs with the sampling ratio in each group of subjects determined by the current treatment effect estimate. We demonstrate the potential of adaptive randomisation to reduce the number of patients receiving the inferior treatment, even when there is a delay in observing each patient's response. We also observe that using a fixed but unequal sampling ratio may offer a simpler way to achieve the same objectives.

Key words and phrases: Response-adaptive randomisation, Group sequential test, Inferior treatment number.

1. DESIGNING A PHASE III CLINICAL TRIAL

I would like to thank the authors for their thought provoking paper. I shall focus my comments on the design of a Phase III clinical trial in which a new treatment is to be compared to a standard. A typical Phase III trial investigates the responses of a large number of patients with half randomised to the new treatment and half to the control. For patients in the trial, it is advantageous if a higher proportion receive the better of the two treatments, reducing the "inferior treatment number" (ITN). Looking beyond the trial, it is desirable to reach an early decision that will make a successful new treatment available to the general patient population as soon as possible.

For simplicity, consider a trial with normally distributed responses $X_{1,i} \sim N(\mu_1, \sigma^2)$, $i = 1, 2, \ldots$, on the new treatment and $\mathrm{X}_{2,i} \sim N(\mu_2, \sigma^2)$, $i = 1, 2, \ldots$, on the control, where σ^2 is known. Denote the treatment effect by $\theta = \mu_1 - \mu_2$ and suppose a positive value of θ indicates that the new treatment is superior. A hypothesis test of H_0: $\theta \le 0$ vs $\theta > 0$ is to be conducted with type I error probability $\alpha = 0.025$ and power $1 - \beta = 0.9$ if $\theta = \delta$, where δ denotes the anticipated treatment effect or a minimal clinically significant treatment effect.

We shall quantify sample sizes through Fisher information for μ_1, μ_2 and θ. If the numbers of subjects on the new treatment and control are n_1 and n_2, respectively, then the information for μ_1 is

$$\mathcal{I}_1 \,=\, 1/\mathrm{Var}(\hat{\mu}_1) \,=\, n_1/\sigma^2,$$

the information for μ_2 is

$$\mathcal{I}_2 \,=\, 1/\mathrm{Var}(\hat{\mu}_2) \,=\, n_2/\sigma^2,$$

and the information for θ is

$$\mathcal{I} = 1/\mathrm{Var}(\hat{\theta}) = 1/\mathrm{Var}(\hat{\mu}_1 - \hat{\mu}_2) \,=\, \{1/\mathcal{I}_1 + 1/\mathcal{I}_2\}^{-1}.$$

A fixed sample size design requires information

$$\mathcal{I}_{\theta,fix} \,=\, \frac{\{\Phi^{-1}(1-\alpha) + \Phi^{-1}(1-\beta)\}^2}{\delta^2}$$

for θ, where Φ denotes the standard normal cumulative distribution function.

With equal allocation, information $\mathcal{I}_{\theta,fix}$ is achieved by setting $n_1 = n_2 = 2\,\sigma^2\mathcal{I}_{\theta,fix}$, so $\mathcal{I}^{(1)} = \mathcal{I}^{(2)} = 2\,\mathcal{I}_{\theta,fix}$. Response-adaptive randomisation (RAR) can be used to allocate a higher proportion of subjects to the superior treatment, i.e., to the new treatment if $\theta > 0$ or to the control if $\theta < 0$. Note that in our example, with the same response variance for each treatment, the total sample size is minimised when $n_1 = n_2 = 2\,\sigma^2\mathcal{I}_{\theta,fix}$, so an RAR scheme that allocates subjects unequally will inevitably require a larger total sample size.

Chapter 17 of [3] describes trial designs that aim to minimise a loss function of the form

$$L(\theta) \,=\, \begin{cases} \mathcal{I}_1 + a^{\theta/\delta}\,\mathcal{I}_2 & \text{if } \theta \ge 0, \\ a^{-\theta/\delta}\,\mathcal{I}_1 + \mathcal{I}_2 & \text{if } \theta \le 0, \end{cases}$$

where a is a chosen constant. The ratio that minimises $L(\theta)$ for a fixed value of $\mathcal{I} = \{1/\mathcal{I}_1 + 1/\mathcal{I}_2\}^{-1}$ is

$$\mathcal{I}_1/\mathcal{I}_2 \,=\, a^{\theta/(2\delta)}. \tag{1}$$

Thus, if we set $a = 4$, the goal is to achieve

$$\begin{aligned} \mathcal{I}_1 \,&=\, \mathcal{I}_2 && \text{when } \theta = 0, \\ \mathcal{I}_1 \,&=\, \sqrt{2}\,\mathcal{I}_2 && \text{when } \theta = \delta/2, \\ \mathcal{I}_1 \,&=\, 2\,\mathcal{I}_2 && \text{when } \theta = \delta. \end{aligned}$$

We shall restrict attention to the case $a = 4$ in this short report but other choices are, of course, possible,

Department of Mathematical Sciences, University of Bath, Bath, UK (e-mail: c.jennison@bath.ac.uk).

2. AN RAR DESIGN WITH θ ESTIMATED FROM DATA POOLED OVER GROUPS

An adaptive sampling rule can be used to achieve the ratio (1). Suppose patients are observed in groups. Let n_{1j} and n_{2j} denote the number of patients randomised to treatments 1 and 2, respectively, in group j and let

$$n_1^k = \sum_{j=1}^{k} n_{1j} \text{ and } n_2^k = \sum_{j=1}^{k} n_{2j}$$

be the cumulative sample sizes on treatments 1 and 2 after k groups. We denote by

$$\hat{\mu}_{1,k} = \frac{1}{n_1^k} \sum_{i=1}^{n_1^k} X_{1,i} \tag{2}$$

and

$$\hat{\mu}_{2,k} = \frac{1}{n_2^k} \sum_{i=1}^{n_2^k} X_{2,i} \tag{3}$$

the estimates of μ_1 and μ_2 obtained from pooled analyses of the responses in the first k groups. Define the associated information levels

$$\mathcal{I}_1^k = n_1^k/\sigma^2 \text{ and } \mathcal{I}_2^k = n_2^k/\sigma^2.$$

Let $\hat{\theta}_{p,k} = \hat{\mu}_{1,k} - \hat{\mu}_{2,k}$ and note that the information for θ after group k is

$$\mathcal{I}_{p,k} = \{1/\mathcal{I}_1^k + 1/\mathcal{I}_2^k\}^{-1}.$$

We have simulated a design in which data are gathered in 5 groups. The first group has equal sampling and yields information levels $\mathcal{I}_1^1 = \mathcal{I}_2^1 = (1/5)\,2\mathcal{I}_{\theta,fix}$. Then for $k = 2, \ldots, 5$, the sampling ratio in group k is chosen so that after group k, the cumulative information levels $\mathcal{I}_1^k$ and $\mathcal{I}_2^k$ for μ_1 and μ_2 satisfy

$$\{1/\mathcal{I}_1^k + 1/\mathcal{I}_2^k\}^{-1} = (k/5)\,\mathcal{I}_{\theta,fix} \tag{4}$$

and

$$\mathcal{I}_1^k/\mathcal{I}_2^k = a^{\hat{\theta}_{p,k-1}/(2\delta)}. \tag{5}$$

If the change in the target ratio is too high, it may not be possible to reach the ratio (5) after the next group, in which case all subjects are taken from the appropriate treatment arm with a sample size satisfying (4).

Theory presented in Chapter 17 of [3] and in [4] shows that the final estimate of θ is distributed as

$$\hat{\theta}_{p,5} \sim N(\theta, \mathcal{I}_{\theta,fix}^{-1})$$

as long as the sampling rule for each stage k is a function of $\hat{\theta}_{p,1}, \ldots, \hat{\theta}_{p,k-1}$ (and does not, for example, involve estimates of the individual means μ_1 and μ_2). Thus, the final hypothesis test can be carried out exactly as in a fixed sample design with no adaptive sampling.

TABLE 1
Designs with no early stopping

	Target		RAR using pooled $\hat{\theta}_{p,k}$		RAR using group based $\hat{\theta}_{g,k}$	
θ	$E(N_1)$	$E(N_2)$	$E(N_1)$	$E(N_2)$	$E(N_1)$	$E(N_2)$
$-\delta/2$	85.4	120.7	86.4	122.8	90.0	120.1
0	100.0	100.0	101.4	101.5	102.5	102.5
$\delta/2$	120.7	85.4	122.8	86.4	120.1	90.0
δ	150.0	75.0	153.0	75.7	145.0	81.2
$3\delta/2$	191.4	67.7	195.9	68.2	180.2	75.0
2δ	250.0	62.5	256.4	62.8	229.8	70.6

Expected sample size for each treatment arm is expressed as a percentage of $n = 2\sigma^2\mathcal{I}_{\theta,fix}$, the sample size per treatment in a non-adaptive design. Results are based on 1 million simulations and are accurate to the number of digits shown.

The results of applying the above approach are presented in Table 1. Here $E(N_1)$ and $E(N_2)$ can be regarded as the expected sample sizes on treatments 1 and 2 when σ^2 and δ are such that a non-adaptive design requires 100 observations per treatment. Alternatively, for general σ^2 and δ, these numbers are the expected sample sizes on each treatment expressed as a percentage of that required in a non-adaptive design. Columns 2 and 3 show the target values given by the ratio (1) and we see these numbers are closely matched by the values for the RAR design in columns 4 and 5. Thus, the desired reductions in the ITN are achieved.

3. AN RAR DESIGN COMBINING WITHIN GROUP ESTIMATES OF θ

In the method of Section 2, pooling data from the five groups can be problematic if responses are affected by a time trend, since differences in the distribution of observations over time for the two treatments can lead to bias in the final estimate of θ. Suppose, for example, the time trend increases the mean of all observations in group j by Δ_j where $\Delta_1, \Delta_2, \ldots$ is an increasing sequence. The contribution of these trend terms to (2) is

$$\sum_{j=1}^{k} \frac{n_{1j}}{n_1^k}\,\Delta_j \tag{6}$$

while the contribution to (3) is

$$\sum_{j=1}^{k} \frac{n_{2j}}{n_2^k}\,\Delta_j. \tag{7}$$

Suppose n_{1j} increases with j and n_{2j} decreases with j. Then, since Δ_j increases with j, term (6) is greater than term (7) and thus the time trend produces an increase in the expected value of $\hat{\theta}_{p,k} = \hat{\mu}_{1,k} - \hat{\mu}_{2,k}$.

In order to protect against the effects of a time trend we can proceed as follows. In group 1, sample the treatments

equally. Choose sample sizes per treatment in this group such that the information levels for μ_1 and μ_2 satisfy $\mathcal{I}_1^{(1)} = \mathcal{I}_2^{(1)} = (1/5)\,2\mathcal{I}_{\theta,fix}$ and, thus, the information for θ is $\mathcal{I}_{g,1} = (1/5)\,\mathcal{I}_{\theta,fix}$. Denote the within group estimates of μ_1 and μ_2 by $\hat{\mu}_1^{(1)}$ and $\hat{\mu}_2^{(1)}$ and define the treatment effect estimate $\hat{\theta}_{g,1} = \hat{\mu}_1^{(1)} - \hat{\mu}_2^{(1)}$.

For $k = 2, \ldots, 5$, set sample sizes in group k such that information levels $\mathcal{I}_1^{(k)}$ and $\mathcal{I}_2^{(k)}$ for μ_1 and μ_2 based on group k data alone satisfy

$$\{1/\mathcal{I}_1^{(k)} + 1/\mathcal{I}_2^{(k)}\}^{-1} \;=\; (1/5)\,\mathcal{I}_{\theta,fix}. \tag{8}$$

and

$$\mathcal{I}_1^{(k)}/\mathcal{I}_2^{(k)} \;=\; a^{\hat{\theta}_{g,k-1}/(2\delta)}, \tag{9}$$

where $\hat{\theta}_{g,k-1}$ is as defined below. Denoting the estimates of μ_1 and μ_2 from group j data by $\hat{\mu}_1^{(j)}$ and $\hat{\mu}_2^{(j)}$, we define the "group based" estimate of θ after group k by

$$\hat{\theta}_{g,k} \;=\; \frac{1}{k}\sum_{j=1}^{k}(\hat{\mu}_1^{(j)} - \hat{\mu}_2^{(j)}). \tag{10}$$

It is straightforward to check that, under our original model, the final estimate of θ is distributed as

$$\hat{\theta}_{g,5} \sim N(\theta, \mathcal{I}_{\theta,fix}^{-1}). \tag{11}$$

Furthermore, if the effect of a time trend is to increase the mean response in group j on both treatments by a constant Δ_j, the distributions of $\hat{\theta}_{g,k}$, $k = 1, \ldots, 5$, are unaffected and (11) still holds.

Results for this design in columns 6 and 7 of Table 1 show that the attained sampling ratio is quite close to its target value for each value of θ. Although the sampling ratio is closer to one and for some θ values there is a small increase in total sample size compared to the pooled data method of Section 2, this may be regarded as an acceptable price to pay for the assurance that the method remains valid in the presence of a time trend.

4. DELAYED RESPONSES

So far, we have supposed all the information from subjects in group $k-1$ is available when setting the allocation ratio for group k. In practice, there is often a delay in observing a patient's response and decisions must be made using the data available at the start of each group. Suppose the time taken to observe a response is equal to the time taken to recruit each group of patients. Then, the randomisation ratio for group k will be based on the estimate of θ from the first $k-2$ groups.

Applying the pooled data approach of Section 2, we simulated an RAR design in which subjects are randomised equally to the two treatments in groups 1 and 2 while, for groups 3, 4 and 5, we solve equations (4) and (5) but with $\hat{\theta}_{p,k-1}$ replaced by $\hat{\theta}_{p,k-2}$ in (5). We see from the results in columns 4 and 5 of Table 2 that the delayed response has little effect when estimates of θ are based on pooled data.

TABLE 2
Designs with no early stopping and one group delay

	Target		RAR using pooled $\hat{\theta}_{p,k}$		RAR using group based $\hat{\theta}_{g,k}$	
θ	$E(N_1)$	$E(N_2)$	$E(N_1)$	$E(N_2)$	$E(N_1)$	$E(N_2)$
$-\delta/2$	85.4	120.7	86.7	123.7	92.8	115.5
0	100.0	100.0	102.0	102.0	102.2	102.2
$\delta/2$	120.7	85.4	123.7	86.7	115.5	92.8
δ	150.0	75.0	154.4	75.9	134.4	86.1
$3\delta/2$	191.4	67.7	198.0	68.3	161.1	81.4
2δ	250.0	62.5	259.5	62.9	198.7	78.1

Expected sample size for each treatment arm is expressed as a percentage of the sample size per treatment in a non-adaptive design. Results are based on 1 million simulations.

In simulating an RAR design with the grouped data approach of Section 3, we randomised subjects equally to the two treatments in groups 1 and 2 then set allocation ratios for groups 3, 4 and 5 by solving (8) and (9) with $\hat{\theta}_{g,k-1}$ replaced by $\hat{\theta}_{g,k-2}$. In this case, it is not possible to compensate later for the equal allocation in groups 1 and 2 and, thus, the overall sampling ratio is closer to one. In keeping with the less extreme sampling ratios, the total sample sizes under $\theta = 3\delta/2$ and $\theta = 2\delta$ are now smaller than for the pooled data method of Section 2.

5. GROUP SEQUENTIAL TESTS WITH RAR

When interim analyses are conducted during a trial, it is natural to consider a group sequential test which allows early stopping if results support this. We created a group sequential stopping boundary for our example based on cumulative Z-statistics $\{Z_1, \ldots, Z_5\}$. We employed a Pampallona and Tsiatis [5] design with shape parameter $\Delta = 0$ (see also Sec. 4.2 of [3]). This has a maximum information level for θ of $\mathcal{I}_{\theta,max} = 1.09\,\mathcal{I}_{\theta,fix}$ and analyses take place with information levels

$$\mathcal{I}_{\theta,k} \;=\; (k/5)\,\mathcal{I}_{\theta,max}, \quad k = 1, \ldots, 5. \tag{12}$$

At analysis k, the trial is stopped to accept H_0: $\theta \leq 0$ if $Z_k < a_k$ and to reject H_0 in favour of $\theta > 0$ if $Z_k > b_k$, where

$$(a_1, \ldots, a_5) = (-1.615, -0.071, 0.816, 1.464, 1.987)$$

and

$$(b_1, \ldots, b_5) = (4.442, 3.141, 2.565, 2.221, 1.987).$$

The stopping boundary is illustrated in Figure 1.

Values of $E(N_1)$ and $E(N_2)$ for selected treatment effects θ are shown in columns 2 and 3 of Table 3. Values of the ITN are clearly lower than those in Table 1 when

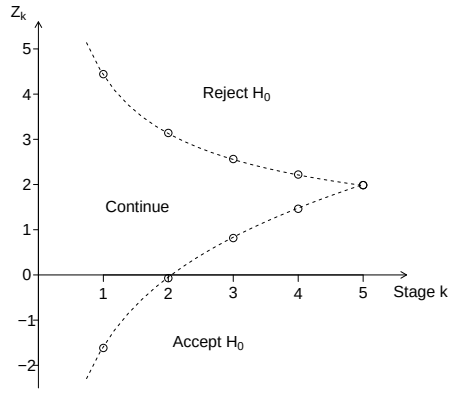

FIG 1. *Group sequential stopping boundary*

RAR is used in a non-sequential test, and reductions in expected sample size on the superior treatment are even greater. These results demonstrate that application of a group sequential stopping rule can be more effective in reducing the number of subjects on the inferior treatment than using RAR with a non-sequential trial design.

TABLE 3
Designs with early stopping

	Non-adaptive		RAR using pooled $\hat{\theta}_{p,k}$		RAR using group based $\hat{\theta}_{g,k}$	
θ	$E(N_1)$	$E(N_2)$	$E(N_1)$	$E(N_2)$	$E(N_1)$	$E(N_2)$
$-\delta/2$	43.2	43.2	41.3	47.7	42.5	45.4
0	59.3	59.3	64.2	57.5	63.8	57.6
$\delta/2$	79.8	79.8	99.9	68.2	96.0	71.4
δ	74.0	74.0	110.3	57.3	99.2	62.8
$3\delta/2$	55.6	55.6	105.6	39.2	84.5	46.0
2δ	44.8	44.8	106.8	29.7	76.9	36.8

Expected sample size for each treatment arm is expressed as a percentage of the sample size per treatment in a non-adaptive design. Results are based on 1 million simulations.

Ideally, one should take advantage of both group sequential stopping and adaptive randomisation. Theory in Chapter 17 of [3] and in [4] justifies the use of RAR in a group sequential design with Z-statistics $Z_k = \hat{\theta}_k\sqrt{\mathcal{I}_{\theta,k}}$, where $\hat{\theta}_k = \hat{\theta}_{p,k}$ in the pooled data approach or $\hat{\theta}_k = \hat{\theta}_{g,k}$ in the grouped data approach. We simulated group sequential designs incorporating RAR with the same rules for determining sampling ratios as before. The resulting expected sample sizes when using estimates $\hat{\theta}_{p,k}$ based on pooled data are shown in columns 4 and 5 of Table 3, and results using group based estimates $\hat{\theta}_{g,k}$ are given in columns 6 and 7.

These results demonstrate that using RAR together with a group sequential test can be highly effective in reducing the ITN. As in the case of a non-sequential test, RAR increases the total sample size. It is notable that $E(N_1)$ is only a little smaller that $E(N_2)$ when $\theta = -\delta/2$ and this can be attributed to the asymmetry of the stopping boundary: from analysis 2 onwards, negative values of $\hat{\theta}_{p,k}$ or $\hat{\theta}_{g,k}$ lead to termination of the trial with acceptance of H_0 rather than continuation with a higher proportion of subjects on the control treatment.

We now consider delayed patient responses. For simplicity, we restrict attention to group sequential designs with information at analyses 1 to 5 given by (12) and the same stopping rule as before, and which do not make use of responses observed after the trial is stopped. (For designs that make proper use of data obtained after stopping recruitment, see the delayed response group sequential tests of Hampson and Jennison [1].) In the results shown in Table 4, $E(N_1)$ and $E(N_2)$ include the subjects whose responses are observed after termination and so do not affect the decision to reject or accept H_0. While the delayed response has an impact on $E(N_1)$ and $E(N_2)$ for all three types of design, the application of RAR still helps reduce the number of subjects on the inferior treatment.

TABLE 4
Designs with early stopping and one group delay

	Non-adaptive		RAR using pooled $\hat{\theta}_{p,k}$		RAR using group based $\hat{\theta}_{g,k}$	
θ	$E(N_1)$	$E(N_2)$	$E(N_1)$	$E(N_2)$	$E(N_1)$	$E(N_2)$
$-\delta/2$	65.0	65.0	62.0	71.9	64.3	67.2
0	80.0	80.0	85.9	78.4	84.2	78.5
$\delta/2$	96.3	96.3	120.2	82.7	111.1	88.8
δ	92.9	92.9	137.0	71.8	117.4	82.3
$3\delta/2$	77.2	77.2	132.7	55.3	106.1	67.7
2δ	66.6	66.6	123.2	46.6	98.7	58.7

Expected sample size for each treatment arm is expressed as a percentage of the sample size per treatment in a non-adaptive design. Results are based on 1 million simulations.

6. DISCUSSION

We have shown that RAR can be incorporated into a group sequential design and this approach can be effective in reducing the expected number of subjects receiving the inferior treatment at the price of an increase in total sample size. Designs can be made robust to a time trend, although then the reduction in the number of subjects on the inferior treatment is smaller. The benefits of RAR are also reduced if there is a delay in observing subjects' responses since fewer subjects are affected by each change in the sampling ratio.

A further requirement in a Phase III trial may be to observe sufficiently many patients on the new treatment to evaluate safety. In a group sequential design, one can adapt the sampling ratio from group 2 onwards but only

allow early stopping after some later point in order to ensure a sufficiently high number of subjects have been allocated to the new treatment. In an RAR scheme, the number of subjects per treatment is random and depends on the sampling ratios up to the point at which the trial terminates, so the analysis at which the number of subjects on the new treatment reaches a specified value may be data dependent. Another option, which removes this uncertainty, is to randomise a fixed, higher proportion of patients to the new treatment throughout the trial. We have implemented such a scheme in our simulations and Table 5 presents results when a 3:2 sampling ratio is used throughout the study.

When a constant 3:2 sampling ratio is applied with the higher number of patients on the new treatment, we have $E(N_2) = (2/3)E(N_1)$ at all values of θ, and the total expected sample size is 4.2% more than for a design with equal allocation. That $E(N_1) > E(N_2)$ for negative values of θ is not ideal but since the trial is motivated by a belief that $\theta > 0$, this is not a major drawback. Comparisons of columns 4 and 5 of Table 5 with columns 4 to 7 of Table 3 and columns 6 and 7 of Table 5 with columns 4 to 7 of Table 4 show this simple design is competitive with RAR methods in reducing ITN, especially if it is deemed necessary to use group based estimates of $\hat{\theta}_{g,k}$ in the RAR design. These results indicate that the simpler option of using a fixed, unequal sampling ratio in a group sequential design should be considered as an alternative to more complex RAR designs.

TABLE 5
Designs with fixed 3:2 sampling ratio

	No early stopping		Early stopping with no delay		Early stopping with delay	
θ	$E(N_1)$	$E(N_2)$	$E(N_1)$	$E(N_2)$	$E(N_1)$	$E(N_2)$
$-\delta/2$	125.0	83.3	54.0	36.0	81.3	54.2
0	125.0	83.3	74.1	49.4	100.0	66.7
$\delta/2$	125.0	83.3	99.8	66.5	120.4	80.2
δ	125.0	83.3	92.5	61.7	116.1	77.4
$3\delta/2$	125.0	83.3	69.5	46.3	96.6	64.4
2δ	125.0	83.3	56.0	37.3	83.2	55.5

Expected sample size for each treatment arm is expressed as a percentage of the sample size per treatment in a non-adaptive design. Results are based on 1 million simulations.

In the above discussion, we have assumed that the new treatment is expected to be at least as effective as the control, so expected sample sizes under negative values of θ are not a cause for concern. This may be regarded as a reasonable assumption in many Phase III trials as such studies are only conducted when previous trials have produced promising results for the new treatment. We would welcome the authors' comments on when they believe a more symmetric attitude to the two treatments would be appropriate, for example, in a comparison of two competing treatments for a rare disease.

Finally, we note that this discussion has been confined to the case of a two-treatment comparison and quite different features may arise in studies with multiple treatment arms. My own interest in multi-arm adaptive trials started many years ago in a collaboration with Iain Johnstone and Bruce Turnbull [2] and it is pleasing to see a number of recent citations to this work which suggest our proposals had some merit.

REFERENCES

[1] HAMPSON, L. V. and JENNISON, C. (2013). Group sequential tests for delayed responses (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75** 3–54.

[2] JENNISON, C., JOHNSTONE, I. M. and TURNBULL, B. W. (1982). Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means. In *Statistical Decision Theory and Related Topics III* 55–86. Elsevier.

[3] JENNISON, C. and TURNBULL, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC.

[4] JENNISON, C. and TURNBULL, B. W. (2001). Group sequential tests with outcome-dependent treatment assignment. *Sequential Analysis* **20** 209–234.

[5] PAMPALLONA, S. and TSIATIS, A. A. (1994). Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the null hypothesis. *Journal of Statistical Planning and Inference* **42** 19–35.