Response Adaptive Randomisation within Group Sequential Tests

Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

http://people.bath.ac.uk/mascj

Response Adaptive Randomisation in Clinical Trials Workshop

Cambridge

February 2024

My early interest in Response Adaptive Randomisation

While a PhD student, I worked on the problem of selecting the best out of K treatments with normally distributed responses.

We produced procedures with

Early stopping,

Dropping poorly performing treatment arms,

Response adaptive randomisation.

After a slow start, our paper is regularly cited now in the machine learning literature.

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

Features of Phase III clinical trials

1. Comparing a new treatment against a control (the current standard of care).

2. Testing — as opposed to selection.

If θ represents the "treatment effect", one hopes to reject the null hypothesis H_0: $\theta \leq 0$, showing that the new treatment is superior to the control.

3. Strict control of the Type I error probability — $\alpha = 0.025$ in a one-sided test.

4. Attain high power: $1 - \beta = 0.9$, say, if $\theta = \delta$.

5. Implicit asymmetry between the new treatment and control.

6. A small number of analyses in a group sequential design.

Example of a Phase III trial

For simplicity, consider a trial with responses

$$X_{1,i} \sim N(\mu_1, \sigma^2) \qquad \text{on the new treatment,}$$
$$X_{2,i} \sim N(\mu_2, \sigma^2) \qquad \text{on the control.}$$

The treatment "effect" is $\theta = \mu_1 - \mu_2$. We shall test

$$H_0\text{: } \theta \leq 0 \text{ vs } \theta > 0$$

with type I error probability $\alpha = 0.025$ and power $1 - \beta = 0.9$ when $\theta = \delta = 1$.

In a fixed sample design with equal allocation we require information

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

Suppose σ^2 is such that this requires 100 patients per treatment.

Group sequential design

In a group sequential design, data are analysed on a small number of occasions, K say, during the trial.

Denote by $\widehat{\theta}_k$ the treatment effect estimate at analysis k.

Define

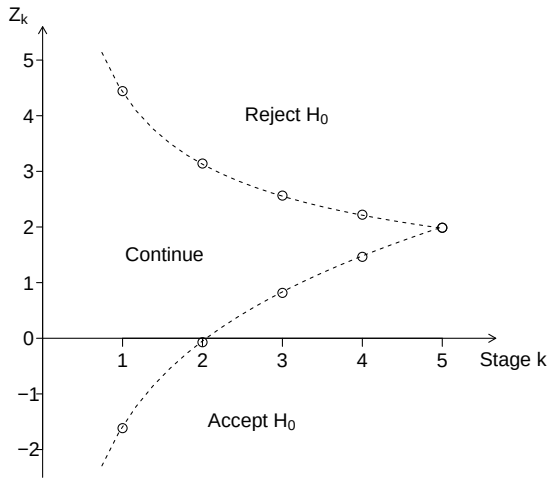
$$\mathcal{I}_{(k)} = \{\mathrm{Var}(\widehat{\theta}_k)\}^{-1},$$

calculated as if the observed sample sizes on each treatment at analysis k had been pre-specified.

A group sequential test monitors

$$Z_k = \widehat{\theta}_k \sqrt{\mathcal{I}_{(k)}}$$

and can stop with an early decision to reject or accept H_0.



Pampallona & Tsiatis (*JSPI*, 1994) design, $\Delta = 0$.

Response Adaptive Randomisation (RAR)

Denote information for μ_1 and μ_2 by $\mathcal{I}_1$ and $\mathcal{I}_2$, respectively.

Then $\mathcal{I}_1$ and $\mathcal{I}_2$ are proportional to the sample sizes on the new treatment and control arms.

Consider a design that aims to minimise a loss function of the form

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

We shall set $a = 4$, which implies a target

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

so we want

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

Group sequential designs and RAR

It is straightforward to include RAR in a group sequential design.

After each analysis, the allocation ratio for the next group can be chosen, based on the current $\widehat{\theta}$.

Group sizes are set so that this group of observations will increase the information for θ by the desired amount.

Jennison & Turnbull (*Seq. Analysis*, 2001) show this adaption does not change the joint distribution of the estimates $\widehat{\theta}_1, \ldots, \widehat{\theta}_K$.

Thus, the original testing boundary can be applied and the design will still have Type I error probability α and power $1 - \beta$ if $\theta = \delta$.

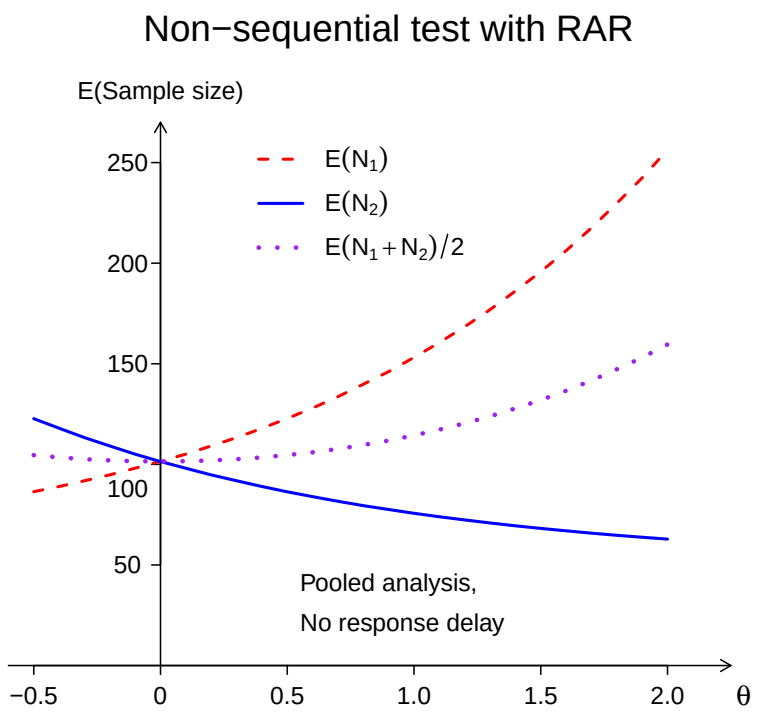
Pooled analysis

One can estimate μ_1, μ_2 and θ from data pooled across groups.

Grouped analysis

Or, to counter the effect of a possible time trends, one can estimate θ within each group, then combine these estimates.

Non-sequential test and Response Adaptive Randomisation

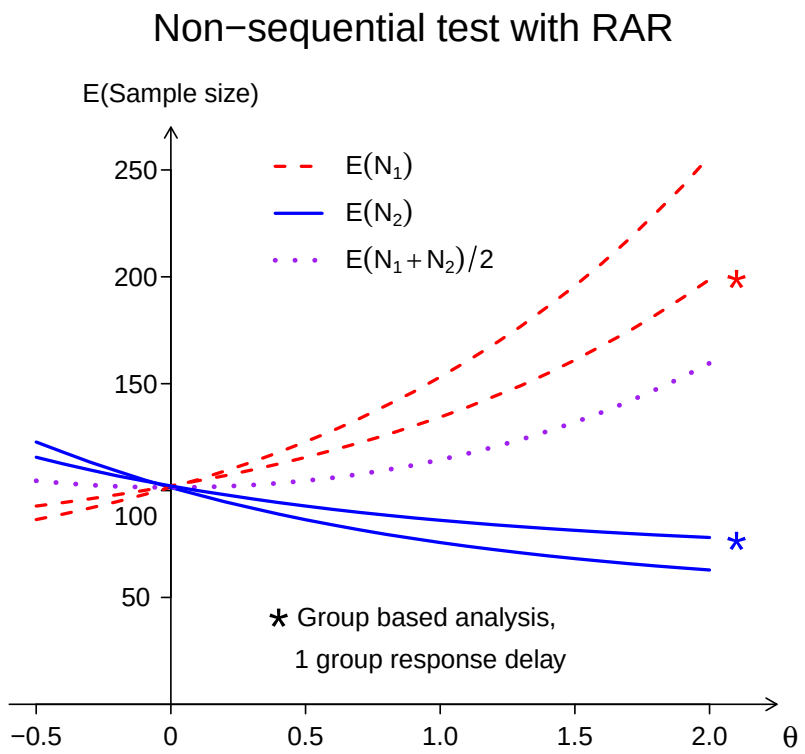


The trial is conducted in 5 stages. From stage 2 onwards, the allocation ratio in each stage is set to achieve $\mathcal{I}_1/\mathcal{I}_2 = 4^{\hat{\theta}/(2\delta)}$.

A non-adaptive design needs sample sizes per treatment arm $N_1 = N_2 = 100$ to achieve the desired Type I error rate and power.

RAR reduces $E(N_2)$ if $\theta > 0$ and reduces $E(N_1)$ if $\theta < 0$.

Non-sequential test with RAR, group based analysis and 1 group delay in observing patient responses

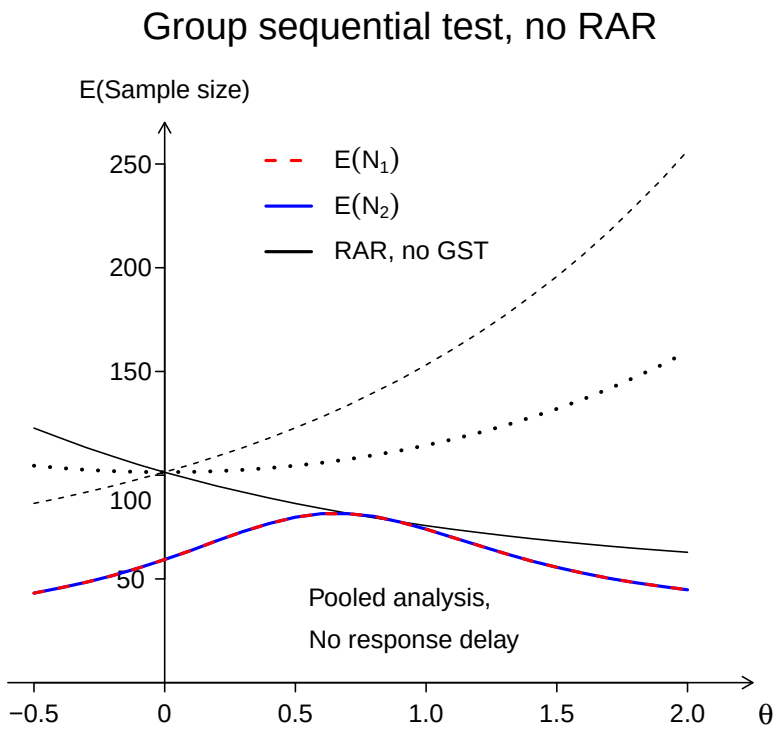


A group based analysis protects against the effects of a time trend.

With delayed response, randomisation is 1 : 1 in the first 2 groups.

RAR is still effective, but its impact is smaller.

A Group Sequential Test (GST) with 1 : 1 allocation ratio (pooled $\widehat{\theta}$ and no delay in observing responses)

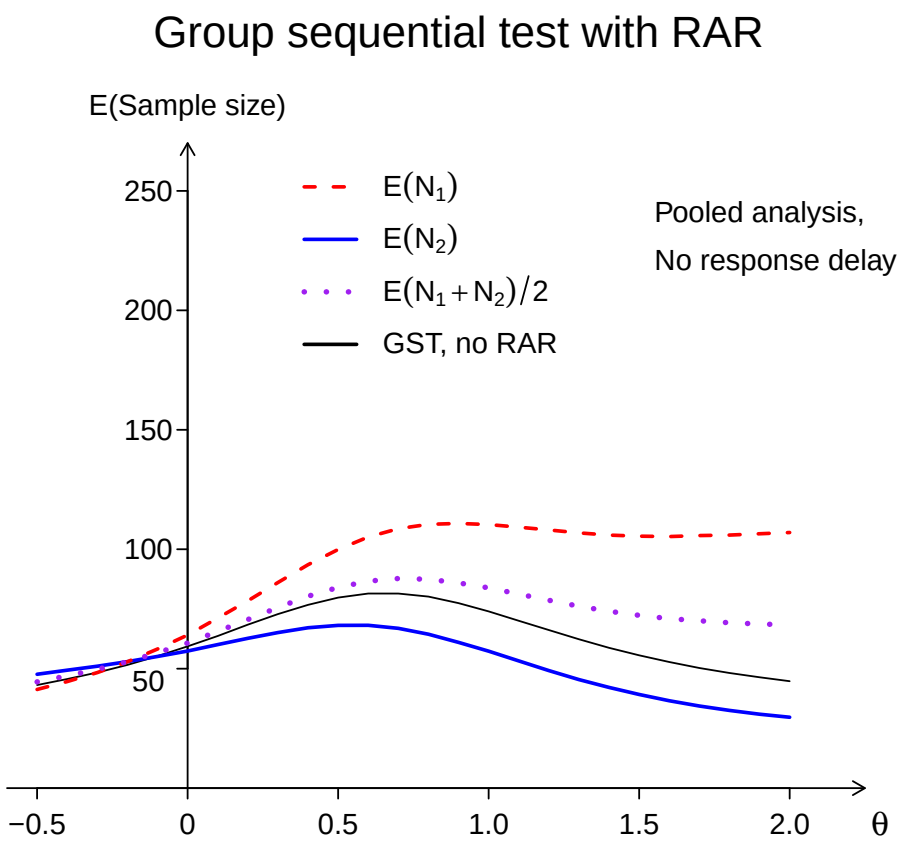


In the GST, $N_1 = N_2$ by design.

The stopping rule reduces both $E(N_1)$ and $E(N_2)$.

Impact is much greater than that of RAR in a non-sequential test.

GST plus RAR (pooled $\widehat{\theta}$, no delay in observing responses)

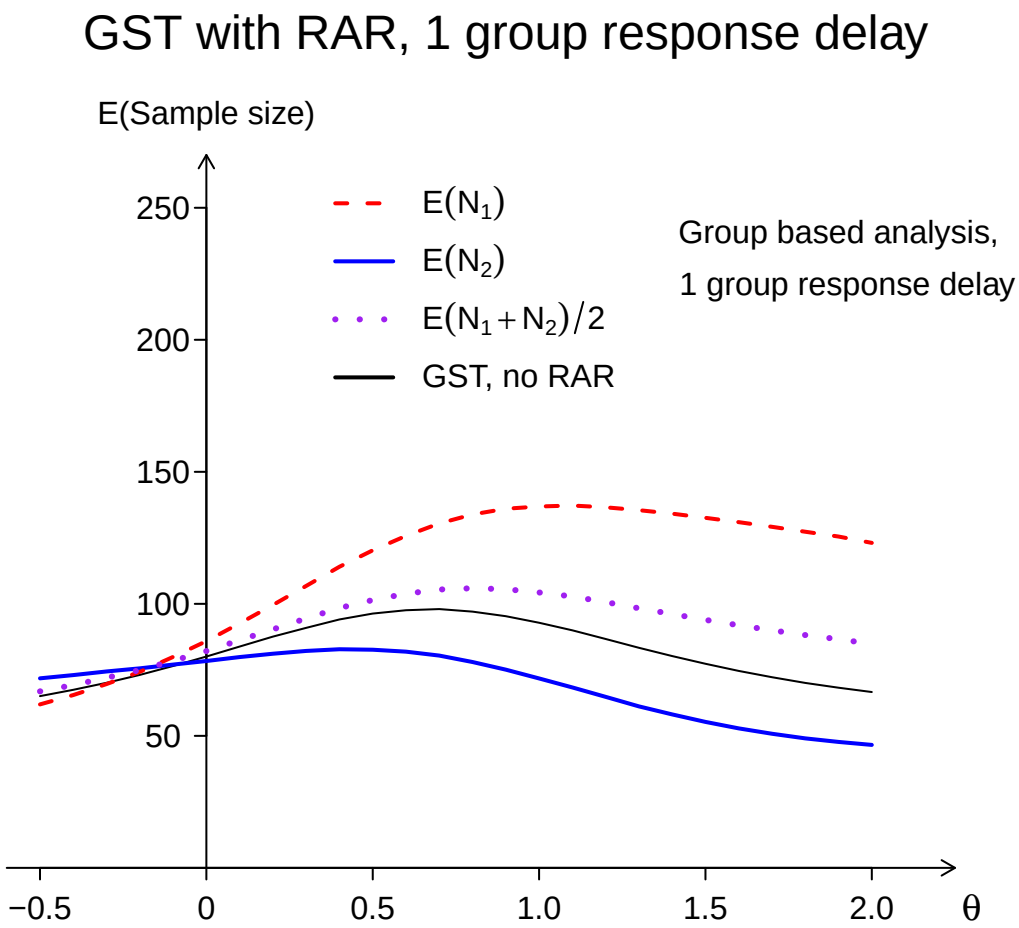


It is straightforward to combine RAR with a group sequential test.

RAR gives further reductions in $E(N_2)$ if $\theta > 0$.

If $\theta < 0$, the GST is likely to stop early to accept H_0: $\theta \leq 0$.

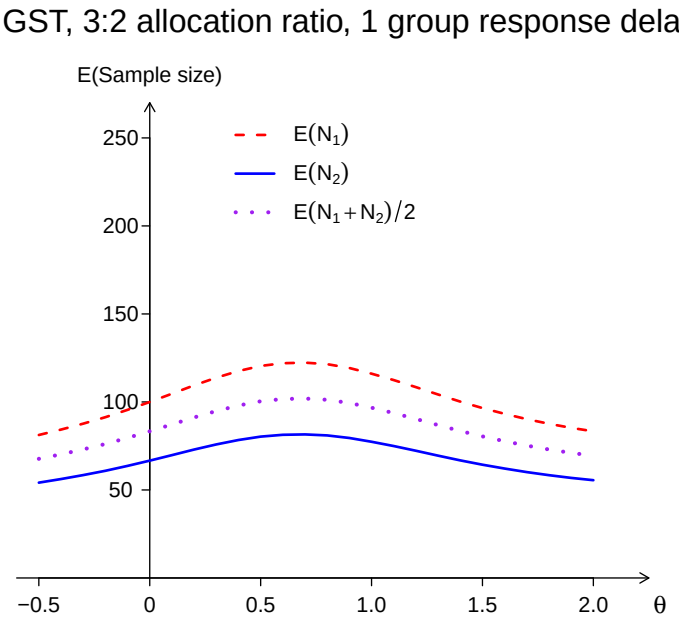
GST plus RAR (group based $\widehat{\theta}$, 1 group delay in responses)



For the GST, $E(N_1)$ and $E(N_2)$ increase due to "pipeline data".

Again, RAR gives further reductions in $E(N_2)$ if $\theta > 0$.

GST with 3 : 2 allocation ratio new treatment : control (group based $\widehat{\theta}$, 1 group delay in responses)

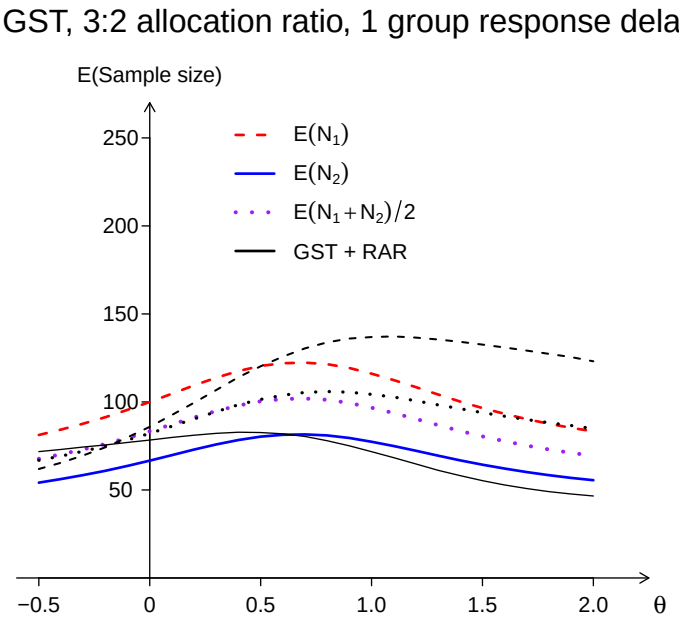


In a Phase III trial, one expects $\theta \geq 0$. And if $\theta < 0$, the GST is likely to stop early.

Interim estimates of θ are noisy, so setting the allocation ratio to give $\mathcal{I}_1/\mathcal{I}_2 = 4^{\hat{\theta}/(2\delta)}$ does not imply $\mathcal{I}_1/\mathcal{I}_2 = 4^{\theta/(2\delta)}$.

A fixed allocation ratio has logistical advantages.

GST with 3 : 2 allocation ratio new treatment : control (group based $\widehat{\theta}$, 1 group delay in responses)



In a Phase III trial, one expects $\theta \geq 0$. And if $\theta < 0$, the GST is likely to stop early.

Interim estimates of θ are noisy, so setting the allocation ratio to give $\mathcal{I}_1/\mathcal{I}_2 = 4^{\hat{\theta}/(2\delta)}$ does not imply $\mathcal{I}_1/\mathcal{I}_2 = 4^{\theta/(2\delta)}$.

A fixed allocation ratio has logistical advantages.

Conclusions

1. Response Adaptive Randomisation can be incorporated in a group sequential design for a Phase III clinical trial.

2. A group sequential test with RAR can reduce the inferior treatment number when compared to a non-adaptive design with a $1:1$ allocation ratio.

3. However, a non-adaptive group sequential test with a fixed $3:2$ allocation ratio gives a similar reduction in the inferior treatment number when $\theta > 0$ — which is all that really matters.

5. RAR may be more suited to multi-arm trials.

6. Also, RAR may be appropriate for trials of treatments for extremely rare diseases, where many of the patients with the disease are in the trial.