Adaptive Methods in Clinical Research

Lecture 2: Group Sequential Designs

David Robertson

MRC Biostatistics Unit

17th September 2024

Overview

1. Introduction to group sequential designs
2. Type I error rate, power and expected sample size
3. Stopping boundaries

Planning an RCT

- A randomised controlled trial (RCT) is carried out to compare the effectiveness of a new experimental treatment versus a control (placebo or standard treatment).
- Generally, frequentist operating characteristics are controlled: the type I error rate α, and power $(1 - \beta)$
- Let θ denote some measure of the difference between the effectiveness of the new treatment and the control.
- Testing the null hypothesis $H_0 : \theta \leq 0$ against the alternative $H_1 : \theta > 0$, using a suitable test statistic Z
- *Fixed sample test*: reject H_0 if $Z > c$, where $P(Z > c) \leq \alpha$ under H_0. Choose sample size n so that $P(Z > c) \geq 1 - \beta$ when $\theta = \delta$

- Consider a group sequential design with a total of J analyses.
- At jth analysis, test statistic Z_j is calculated using patients assessed so far.
- A general one-sided group sequential test is defined by constants (l_j, u_j) with $l_j < u_j$ for $j = 1, \ldots, J$ and $l_J = u_J$

Group sequential designs

- Consider a group sequential design with a total of J analyses.
- At jth analysis, test statistic Z_j is calculated using patients assessed so far.
- A general one-sided group sequential test is defined by constants (l_j, u_j) with $l_j < u_j$ for $j = 1, \ldots, J$ and $l_J = u_J$

After group $j = 1, \ldots, J - 1$

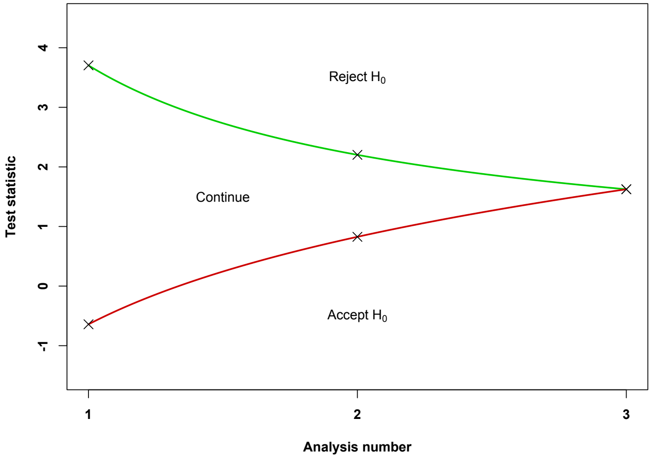
if $Z_j \geq u_j$	stop, reject H_0	(*early stopping for efficacy*)
if $Z_j \leq l_j$	stop, do not reject H_0	(*early stopping for lack of benefit*)
otherwise	continue to group $j + 1$	

after group J

if $Z_J \geq u_J$	stop, reject H_0
if $Z_J < l_J$	stop, do not reject H_0

MRC Biostatistics Unit · UNIVERSITY OF CAMBRIDGE

Group sequential design schematic



- A naive way to apply group sequential designs: test H_0 at significance level α several times throughout the trial.
- E.g., $u_1 = \cdots = u_J = 1.64$, $l_1 = \cdots, l_{J-1} = -\infty$, $l_J = 1.64$
- This inflates the probability of making a type I error:

- A naive way to apply group sequential designs: test H_0 at significance level α several times throughout the trial.
- E.g., $u_1 = \cdots = u_J = 1.64$, $l_1 = \cdots , l_{J-1} = -\infty$, $l_J = 1.64$
- This inflates the probability of making a type I error:

Number of analyses	Type I error rate
1	0.050
2	
3	
5	
10	

Repeated testing of H_0

- A naive way to apply group sequential designs: test H_0 at significance level α several times throughout the trial.
- E.g., $u_1 = \cdots = u_J = 1.64$, $l_1 = \cdots, l_{J-1} = -\infty$, $l_J = 1.64$
- This inflates the probability of making a type I error:

Number of analyses	Type I error rate
1	0.050
2	0.080
3	0.101
5	0.130
10	0.172

- Need to adjust the critical value used to ensure the overall type I error rate is controlled.

History

- A long history: theory dates back to Wald in the 1940s, with early medical applications by Armitage in the 1950s/60s
- Now one of the most commonly used type of adaptive designs

Calculating error probabilities

- Let $\boldsymbol{l} = (l_1, \ldots, l_J)$ be the *lack of benefit boundaries* and $\boldsymbol{u} = (u_1, \ldots, u_J)$ the *efficacy boundaries*.
- Let n_j denote the number of patients who have been assessed on each arm by the jth analysis, and let $\boldsymbol{n} = (n_1, \ldots, n_J)$

MRC Biostatistics Unit UNIVERSITY OF CAMBRIDGE

- Let $\boldsymbol{l} = (l_1, \ldots, l_J)$ be the *lack of benefit boundaries* and $\boldsymbol{u} = (u_1, \ldots, u_J)$ the *efficacy boundaries*.
- Let n_j denote the number of patients who have been assessed on each arm by the jth analysis, and let $\boldsymbol{n} = (n_1, \ldots, n_J)$
- Problem: how to choose $\boldsymbol{l}$, $\boldsymbol{u}$ and $\boldsymbol{n}$ so that design has overall type I error rate α and power $(1-\beta)$
- To solve this we can solve a simpler problem: for a trial design with parameters $\boldsymbol{l}$, $\boldsymbol{u}$ and $\boldsymbol{n}$, what is the type I error rate and power?
- A useful theorem can be used.

Calculating error probabilities

- Let $\hat{\theta}_j$ be the maximum likelihood estimate (MLE) of θ at analysis j
- Let $\mathcal{I}_j$ denote the (Fisher) *information* at analysis j
  - Asymptotic variance of MLE given by $\text{var}(\hat{\theta}_j) = \frac{1}{\mathcal{I}_j}$
- At analysis j, the Wald statistic, Z_j is calculated:

$$Z_j = \frac{\hat{\theta}_j}{\sqrt{\text{var}(\hat{\theta}_j)}} = \hat{\theta}_j \sqrt{\mathcal{I}_j} \tag{1}$$

MRC Biostatistics Unit
UNIVERSITY OF CAMBRIDGE

- **Theorem 1**: the asymptotic joint distribution of $(Z_1, \ldots, Z_J)$ given $(\mathcal{I}_1, \ldots, \mathcal{I}_J)$ has the following properties:
  - $(Z_1, \ldots, Z_J)$ is multivariate normal
  - $E(Z_j) = \theta\sqrt{\mathcal{I}_j}$
  - $\mathrm{Cov}(Z_{j_1}, Z_{j_2}) = \sqrt{\mathcal{I}_{j_1}/\mathcal{I}_{j_2}}$ for $1 \leq j_1 \leq j_2 \leq J$

  For proof (and required regularity conditions), see Jennison and Turnbull (JASA 1997)

- **Key message**: Group sequential design theory is applicable in a very wide variety of design scenarios
  - E.g. Binary outcomes, time-to-event outcomes, GLMs, covariate adjustment, . . .

Example: normally distributed endpoint

- Responses $Y_{0i} \sim N(\mu_0, \sigma^2)$ for patients on control treatment, and $Y_{1i} \sim N(\mu_1, \sigma^2)$ for patients on experimental treatment. Assume σ^2 known.
- Parameter $\theta = \mu_1 - \mu_0$ is of interest.

Example: normally distributed endpoint

- Responses $Y_{0i} \sim N(\mu_0, \sigma^2)$ for patients on control treatment, and $Y_{1i} \sim N(\mu_1, \sigma^2)$ for patients on experimental treatment. Assume σ^2 known.
- Parameter $\theta = \mu_1 - \mu_0$ is of interest.
- At analysis j, MLE $\hat{\theta}_j$ is

$$\bar{Y}_1^{(j)} - \bar{Y}_0^{(j)} = \left(\frac{1}{n_j} \sum_{i=1}^{n_j} Y_{1i} - \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{0i} \right) \sim N(\theta \, , \, \frac{2\sigma^2}{n_j})$$

- Wald test statistic is $Z_j = (\bar{Y}_1^{(j)} - \bar{Y}_0^{(j)})\sqrt{\mathcal{I}_j}$, where $\mathcal{I}_j = \frac{n_j}{2\sigma^2}$
- $(Z_1, \ldots, Z_J)$ is multivariate normal (since linear combination of independent normals), and marginally $Z_j \sim N(\theta\sqrt{\mathcal{I}_j}, 1)$

Example: normally distributed endpoint

- $(Z_1, \ldots, Z_J)$ is multivariate normal (since linear combination of independent normals), and $Z_j \sim N(\theta\sqrt{\mathcal{I}_j}, 1)$

- For $j_1 \leq j_2$,

$$\begin{aligned}\mathrm{Cov}(Z_{j_1}, Z_{j_2}) &= \mathrm{Cov}\left(\bar{Y}_1^{(j_1)} - \bar{Y}_0^{(j_1)}\,,\ \bar{Y}_1^{(j_2)} - \bar{Y}_0^{(j_2)}\right)\sqrt{\mathcal{I}_{j_1}}\sqrt{\mathcal{I}_{j_2}} \\ &= \frac{2}{n_{j_1} n_{j_2}} n_{j_1} \sigma^2 \sqrt{\mathcal{I}_{j_1}}\sqrt{\mathcal{I}_{j_2}} = \sqrt{\mathcal{I}_{j_1} / \mathcal{I}_{j_2}}\end{aligned}$$

MRC Biostatistics Unit UNIVERSITY OF CAMBRIDGE

- Using the multivariate normal distribution of $(Z_1, \ldots, Z_J)$, we can derive the probability of the trial stopping at each stage.

Example: normally distributed endpoint

- Using the multivariate normal distribution of $(Z_1, \ldots, Z_J)$, we can derive the probability of the trial stopping at each stage.

- For example, the probability of stopping for efficacy in the second stage is:

$$\int_{l_1}^{u_1} \int_{u_2}^{\infty} \phi_2((y_1, y_2), \left(\theta\sqrt{\frac{n_1}{2\sigma^2}}, \theta\sqrt{\frac{n_2}{2\sigma^2}}\right), \begin{pmatrix} 1 & \sqrt{n_1/n_2} \\ \sqrt{n_1/n_2} & 1 \end{pmatrix}) dy_2 dy_1$$

where $\phi_2(y, \mu, \Sigma)$ is the density of the bivariate normal distribution with mean μ and covariance Σ at y.

- To get type I error rate, add up probabilities of stopping for efficacy at each analysis when $\theta = 0$
- To get power, add up probabilities of stopping for efficacy at each analysis when $\theta = \delta$
- E.g. for three-stage design with $\boldsymbol{u} = (2.5, 2, 1.5)$, $\boldsymbol{l} = (0, 0.75, 1.5)$, $\boldsymbol{n} = (20, 40, 60)$, $\delta = 0.5$, $\sigma^2 = 1$:

	$\theta = 0$		$\theta = 0.5$	
Analysis	Prob stop for futility	Prob stop for efficacy	Prob stop for futility	Prob stop for efficacy
1	0.500	0.006	0.057	0.179
2	0.299	0.019	0.042	0.420
3	0.137	0.038	0.049	0.253

$$\text{So type I error rate} = 0.006 + 0.019 + 0.038 = 0.063;$$

$$\text{power} = 0.179 + 0.42 + 0.253 = 0.852.$$

MRC Biostatistics Unit UNIVERSITY OF CAMBRIDGE

Finding a design with given error rates

- We want to find a design with given type I error rate and power.
- Generally it is assumed that analyses are equally spaced in terms of patients, so that $(n_1, \ldots, n_J) = (n, 2n, \ldots, Jn)$
- First find boundary that gives correct type I error rate, then find sample size n to give correct power.

- We want to find a design with given type I error rate and power.
- Generally it is assumed that analyses are equally spaced in terms of patients, so that $(n_1, \ldots, n_J) = (n, 2n, \ldots, Jn)$
- First find boundary that gives correct type I error rate, then find sample size n to give correct power.
- One approach: choose $\boldsymbol{u}$ and $\boldsymbol{l}$, then find c such that $(cl_1, \ldots, cl_J)$ and $(cu_1, \ldots, cu_J)$ gives correct type I error rate.

Example:

- Assume normally distributed endpoint as before
- For $\boldsymbol{l} = (0, 0.75, 1.5)$ and $\boldsymbol{u} = (2.5, 2, 1.5)$, $c = 1.081$ gives type I error rate equal to 0.05

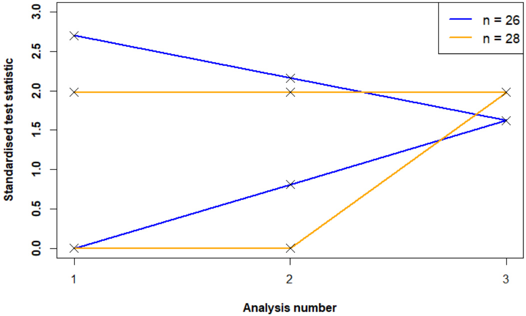
Example (continued):

- Stopping boundaries $\boldsymbol{l} = (0, 0.811, 1.622)$ and $\boldsymbol{u} = (2.703, 2.162, 1.622)$
- In this case, $n = 26$ is needed for 90% power when $\theta = 0.5$.
- So $\boldsymbol{n} = (26, 52, 78)$, $\boldsymbol{l} = (0, 0.811, 1.622)$ and $\boldsymbol{u} = (2.703, 2.162, 1.622)$ is a group sequential design with type I error rate 0.05 and power 0.9.

Example (continued):

- Stopping boundaries $\boldsymbol{l} = (0, 0.811, 1.622)$ and $\boldsymbol{u} = (2.703, 2.162, 1.622)$
- In this case, $n = 26$ is needed for 90% power when $\theta = 0.5$.
- So $\boldsymbol{n} = (26, 52, 78)$, $\boldsymbol{l} = (0, 0.811, 1.622)$ and $\boldsymbol{u} = (2.703, 2.162, 1.622)$ is a group sequential design with type I error rate 0.05 and power 0.9.
- However, $\boldsymbol{n} = (28, 56, 84)$, $\boldsymbol{l} = (0, 0, 1.98)$, $\boldsymbol{u} = (1.98, 1.98, 1.98)$ is another design with the same characteristics

Example (continued):



- In fact there are an *infinite* number of possible designs
- How to pick between them?

Expected sample size

- One way to distinguish between designs with the same type I error rate and power is through *expected sample size* (ESS).

MRC Biostatistics Unit UNIVERSITY OF CAMBRIDGE

- One way to distinguish between designs with the same type I error rate and power is through *expected sample size* (ESS).

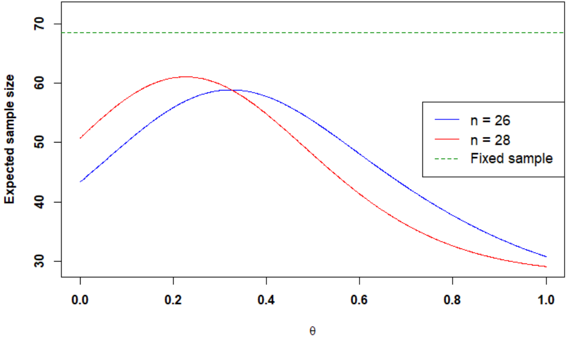
- Can be calculated using the stopping probabilities. For example, from previous example on slide 14 when $\boldsymbol{n} = (20, 40, 60)$ and $\theta = 0$:

Analysis	Prob stop for futility	Prob stop for efficacy	Total prob of stopping
1	0.500	0.006	0.506
2	0.299	0.019	0.318
3	0.137	0.038	0.176

- The ESS is: $(0.506 \times 20) + (0.318 \times 40) + (0.176 \times 60) = 33.4$

- More generally, the ESS is $\sum_{j=1}^{J} n_j p_j$, where p_j is the probability of stopping at analysis j

For example, expected sample size for the two designs mentioned on slide 16.



Which is preferable?

Expected and maximum sample size

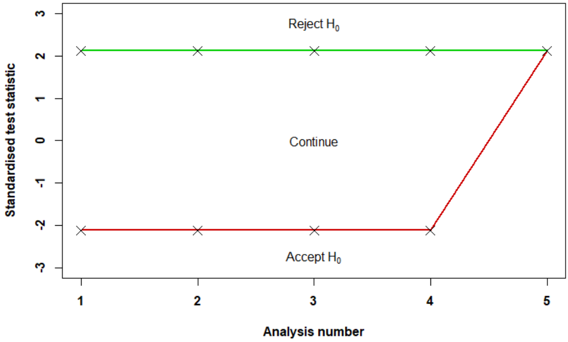
- If trial is powered to detect difference $\theta = 0.5$, the true θ may be smaller than this.
- Thus design in blue may be preferable in practice – however, depends on prior beliefs on θ
- Another advantage of the blue design is that its *maximum sample size*, MSS, is lower (78 vs 84).
  - This is the number of patients recruited if the trial does not stop until the last analysis
  - For fixed sample size trial, need 69 per arm.

Choosing stopping boundary shapes

- The shape of the stopping boundaries will determine the ESS and MSS.
- Two main ways to choose between the infinite number of possible shapes:
  1. Use some 'fixed' boundary shape (specified through a simple functional form);
  2. Search for an optimal design to (e.g.) minimise ESS for some value of θ.
- First method is much quicker, but second method allows greater control over the ESS properties of the design.
- Common boundary shapes to choose from:
  1. Pocock
  2. O'Brien-Fleming
  3. Triangular test

Pocock

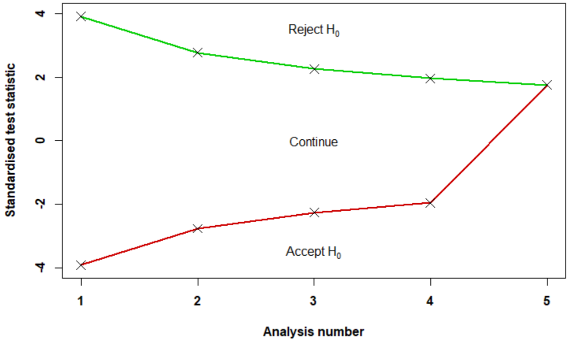
- $u_1 = \cdots = u_J = C$
- $l_1 = \cdots = l_{J-1} = -C$, $l_J = C$
- C chosen to ensure correct type I error rate and power



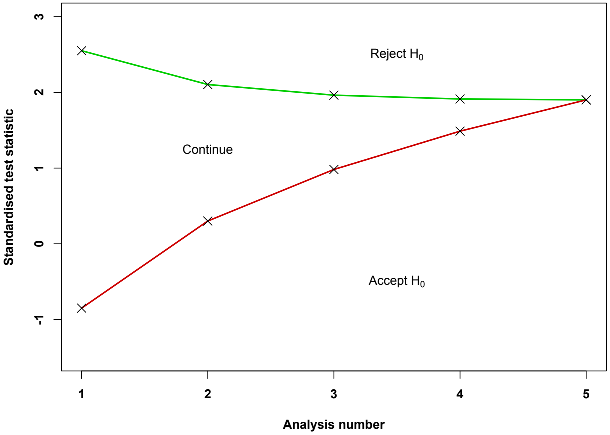
MRC Biostatistics Unit
UNIVERSITY OF CAMBRIDGE

O'Brien-Fleming

- $u_j = C\sqrt{J/j}, j = 1, \ldots J$
- $l_j = -C\sqrt{J/j}, j = 1, \ldots J - 1$ and $l_J = C$
- C chosen to ensure correct type I error rate and power



Triangular test



Optimal designs

- Firstly, define a *feasible design* as a group sequential design which meets some type I error rate and power constraints.
- An optimal design is feasible *and* minimises the expected sample size for some value of θ.
- Infinite number of optimal designs; some common ones:
  - Null-optimal design; optimal for $\theta = 0$;
  - Alternative-optimal design; optimal for $\theta = \delta$
  - θ-minimax design; has the lowest maximum expected sample size.

(Near-)Optimal boundaries

- To aid finding optimal design, can consider a constrained set of possible boundary shapes:
- Stopping boundaries are

$$
\begin{aligned}
u_j &= C_u (j/J)^{\Delta_u - 1/2} \\
l_j &= \delta\sqrt{\mathcal{I}_j} - C_l (j/J)^{\Delta_l - 1/2}
\end{aligned}
$$

- The constants C_u and C_l are chosen to ensure correct type I error and power
- (Δ_u, Δ_l) chosen to minimise ESS
- This is the approach taken in the `OptGS` package (see Practical)

Advantages and disadvantages of group sequential trials

- Advantages:
  - Fewer patients required on average compared to fixed sample-size designs.
  - If one treatment is ineffective, fewer patients on average will be exposed to it.
  - Reduces time to get effective treatment to market.
- Disadvantages:
  - If trial continues to the end, more patients will be used compared to fixed sample-size design.
  - Interim analyses can introduce practical issues
  - Complicates the analysis (see Lecture 6 . . .)

References

1. Jennison, C. and Turnbull, B. (1997). Group-Sequential Analysis Incorporating Covariate Information. *Journal of the American Statistical Association*, 92:1330–1341

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

3. Wason, J. Mander, A. and Thompson, S. (2012). Optimal multi-stage designs for randomised clinical trials with continuous outcomes. *Statistics in Medicine*, 31:301–312.