MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Adaptive Methods in Clinical Research

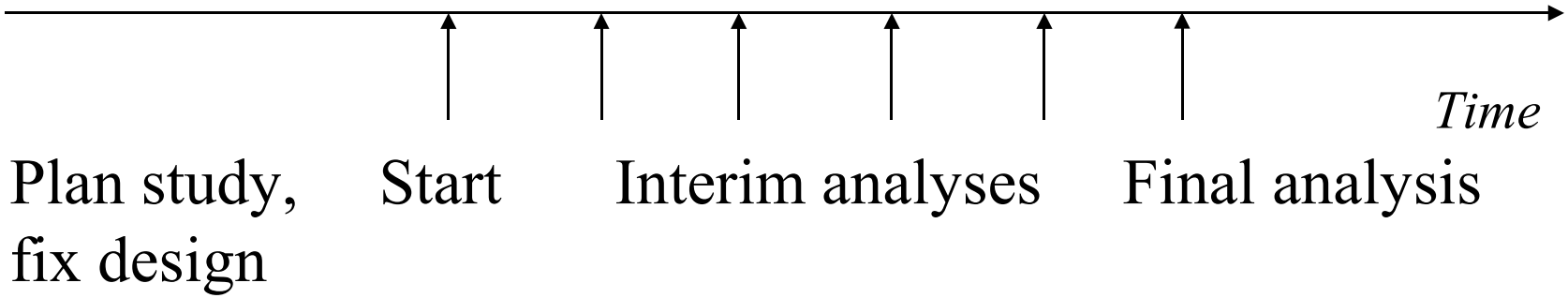
Lecture 7: Fully adaptive designs

Thomas Jaki

MRC Biostatistics Unit

19 September 2024

Adaptive Designs



At each interim

- decide whether or not to stop
- change sample size
- drop or add a dose
- change the endpoint
- change the question

Fisher's combination method

	First stage	Second stage
Observations	n_1	n_2
one-sided p-value	p_1	p_2

- p-values need to be independent
- Combine the evidence from the two stages via taking product of p-values from both stages
- Under H_0, $-2\log(p_1 \times p_2) \sim \chi^2_4$ (Fisher, 1932)

Proof

under H_0

$$p_1 \sim U(0,1) \quad \text{and independently} \quad p_2 \sim U(0,1)$$

Hence

$$-\log(p_1) \sim Exp(1) \quad \text{and} \quad -\log(p_2) \sim Exp(1)$$

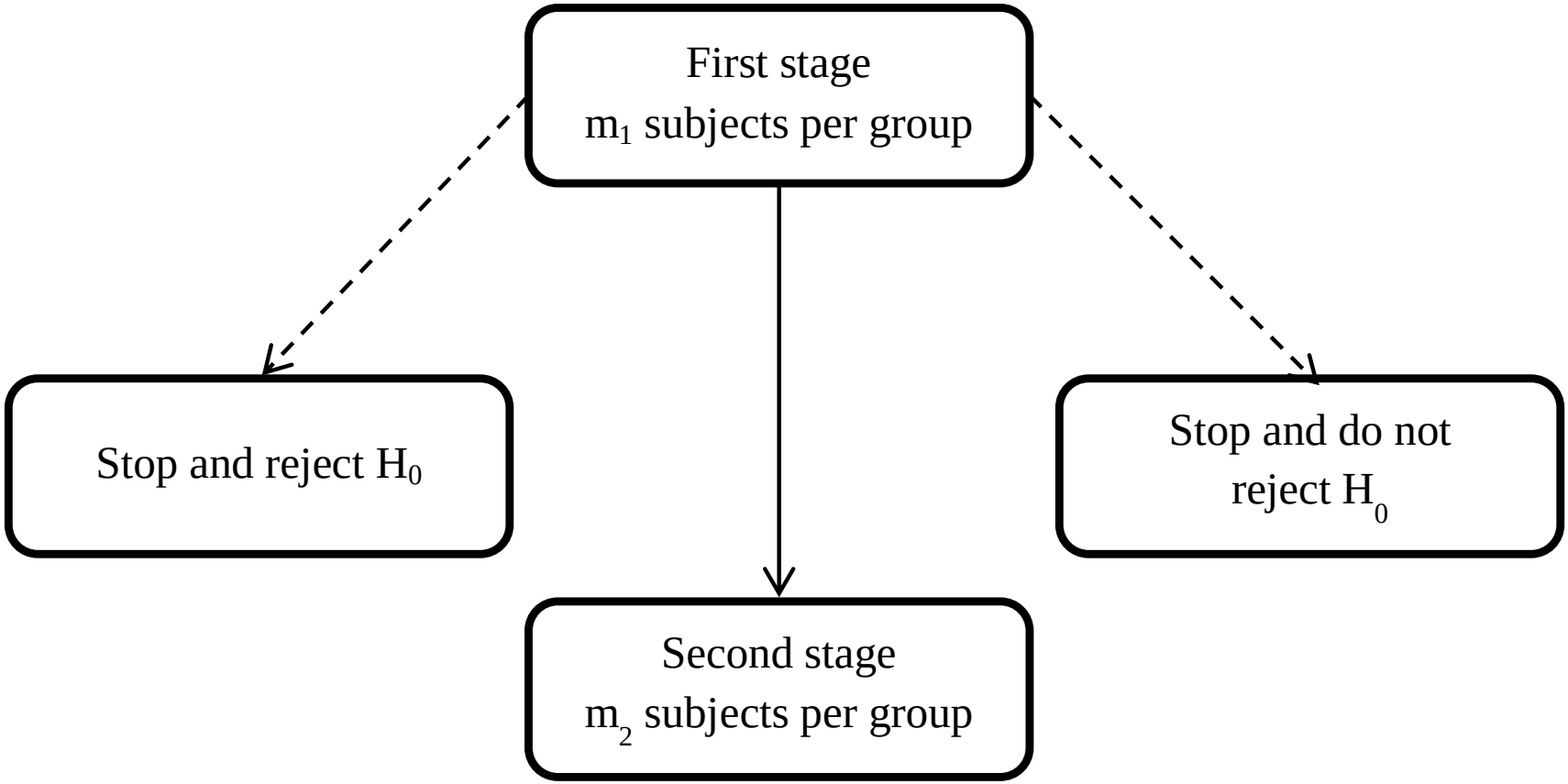
and so

$$-\log(p_1 p_2) = -\log(p_1) - \log(p_2) \sim Ga(2,1)$$

and therefore

$$-2\log(p_1 p_2) \sim Ga\left(\frac{4}{2}, \frac{1}{2}\right) = \chi^2_4$$

Two-stage design with early stopping



- Proposed by Bauer & Köhne (1994)
- Stage 1 and Stage 2 need to be independent

Two-stage design with early stopping

The upper 0.975 point of the χ^2_4 distribution is 11.14. Setting $\alpha = 0.025$ (one-sided), we will PROCEED to claim that E $>$ C if

$$-2\log(p_1 p_2) \geq 11.14$$

or equivalently if

$$p_1 p_2 \leq 0.0038$$

If after the first stage, we already know that

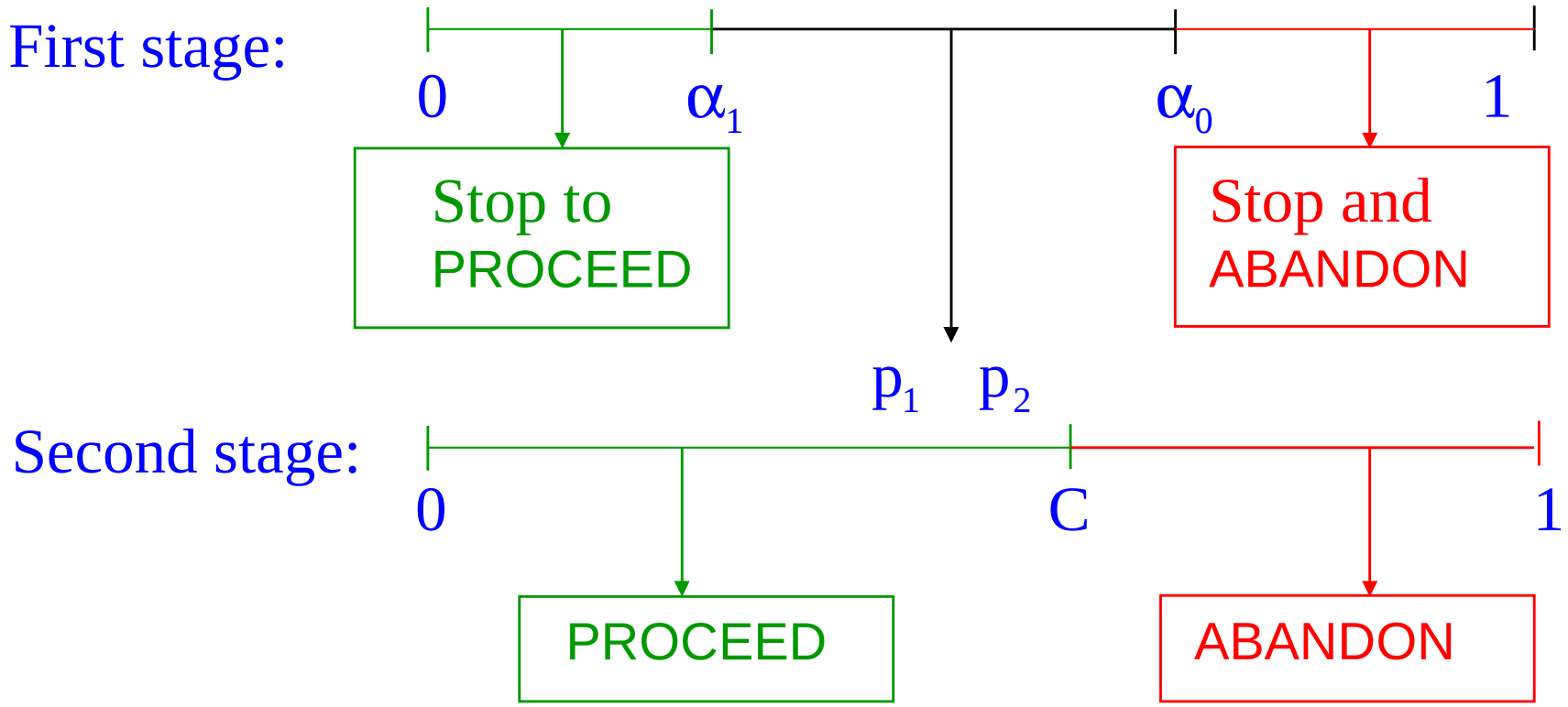
$$p_1 \leq 0.0038$$

then there is no need to conduct the second stage: we will PROCEED to claim that E $>$ C
$-$ curtailed sampling

Two-stage design with early stopping

The full procedure is as shown:

– for $\alpha = 0.025$, $\alpha_1 = 0.0038$

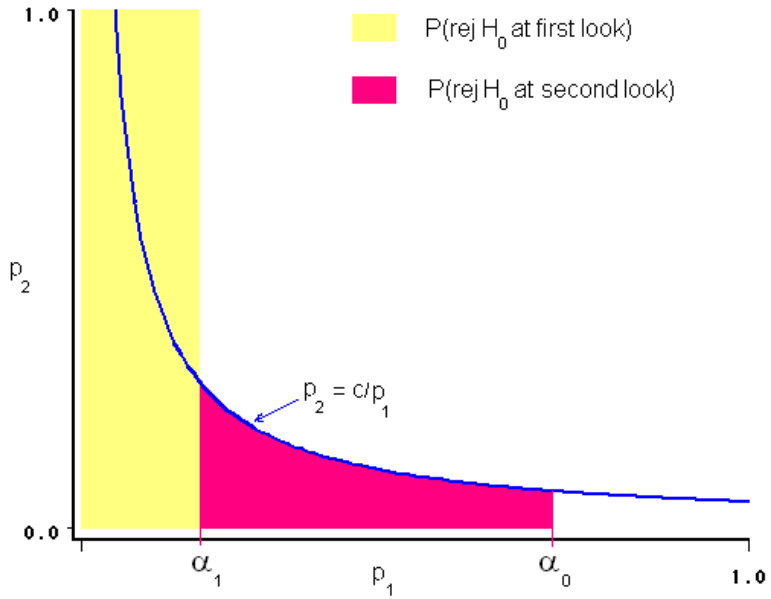


Two-stage design with early stopping

If the early PROCEED boundary is set at α_1 and the early ABANDON boundary is set at α_0, then in order to achieve an overall one-sided type I error rate of α,

$$c = \frac{\alpha - \alpha_1}{\log(\alpha_0) - \log(\alpha_1)}$$

found by integration



Two-stage design with early stopping

Hence

$$\begin{aligned}\alpha &= \alpha_1 + \int_{\alpha_1}^{\alpha_0} \frac{c}{p_1} dp_1 \\ &= \alpha_1 + c\Big[\log(p_1)\Big]_{\alpha_1}^{\alpha_0} \\ &= \alpha_1 + c\Big[\log(\alpha_0) - \log(\alpha_1)\Big]\end{aligned}$$

so that

$$c = \frac{\alpha - \alpha_1}{\log(\alpha_0) - \log(\alpha_1)}$$

Beyond 2 stages

- Wassmer (1999) extended this approach to any number of stages
- At the i^{th} interim analysis calculate the product of p-values
  1. Stop and PROCEED if $p_1 \times \cdots \times p_i \leq c_i$
  2. Stop and ABANDON if $p_i \geq \alpha_{0i}$
  3. Use a recursive method to find $c_1, \ldots, c_k$ and $\alpha_{01}, \ldots, \alpha_{0k}$ satisfying some chosen constraints
- Calculations become more complex as number of stages increases
- All stages equally weighted

Inverse normal method

For group sequential trials we need test statistics B_i and V_i which, under the null hypothesis, satisfy

- $B_i \sim N(0, V_i)$
- increments $(B_i - B_{i-1})$ between interims are independent

Regardless of where these statistics come from

Note:

$$Z = \frac{B}{\sqrt{V}}$$

where B is the score statistic and V is the Fisher information.

Transforming the p-value

Let

$$Z = \Phi^{-1}(1 - P)$$

where Φ denotes the $N(0, 1)$ distribution function

Then

$$\begin{aligned}
\mathbb{P}(Z \leq z|\theta = 0) &= \mathbb{P}(\Phi^{-1}(1 - P) \leq z|\theta = 0) \\
&= \mathbb{P}((1 - P) \leq \Phi(z)|\theta = 0) \\
&= \mathbb{P}(P \geq 1 - \Phi(z)|\theta = 0) \\
&= \Phi(z)
\end{aligned}$$

so that $Z \sim N(0, 1)$.

Combining the p-values

Consider tests of $H_{i0}: \theta_i = 0$ vs $H_{i1}: \theta_i > 0$, based on independent, sequentially available data sets $\boldsymbol{x}_i$, with corresponding one-sided p-values, $p_i,\ i = 1, \ldots, k$

Then

$$Z_i = \Phi^{-1}(1 - P_i),\ i = 1, \ldots, k$$

are independent $N(0, 1)$ random variables, and

$$Y_i = W_i \Phi^{-1}(1 - P_i),\ i = 1, \ldots, k$$

are independent $N(0, W_i^2)$ random variables.

Now put

$$B_i = Y_1 + \cdots + Y_i \qquad V_i = W_1^2 + \cdots + W_i^2$$

Then if all null hypotheses $H_{i0} : \theta_i = 0$ are true,

- $B_i \sim N(0, V_i)$
- increments $(B_i - B_{i-1})$ between interims are independent

So, if these statistics are plotted and compared with sequential stopping boundaries, then the required type I error will be achieved.

Applications

1. Hypotheses H_{i0} could all be the same: $\theta = 0$, based on independent data $\mathbf{x}_i$, each comprising the new data only observed between the ith and the (i-1)th interim analyses

   $\Rightarrow$ the sample size, allocation ratio or other design features concerning the ith dataset can depend on previous data

2. The hypotheses H_{i0} could concern different endpoints (mortality, time to progression, tumour shrinkage), or different test statistics (logrank, Wilcoxon, binary) based on independent groups of patients

   $\Rightarrow$ the endpoint or test statistic could be changed between interim analyses, provided that H_0: "the treatments are identical" is to be tested

Closed testing

Treatments: k experimentals $T_1, \ldots, T_k$
one control T_0

Adaptive designs tend to use pairwise comparisons i.e. individually test each $H_{0,i}$: $\theta_i = 0$

To strongly control the type I error rate for testing multiple hypotheses closed testing procedures are used (Marcus et al., 1976)

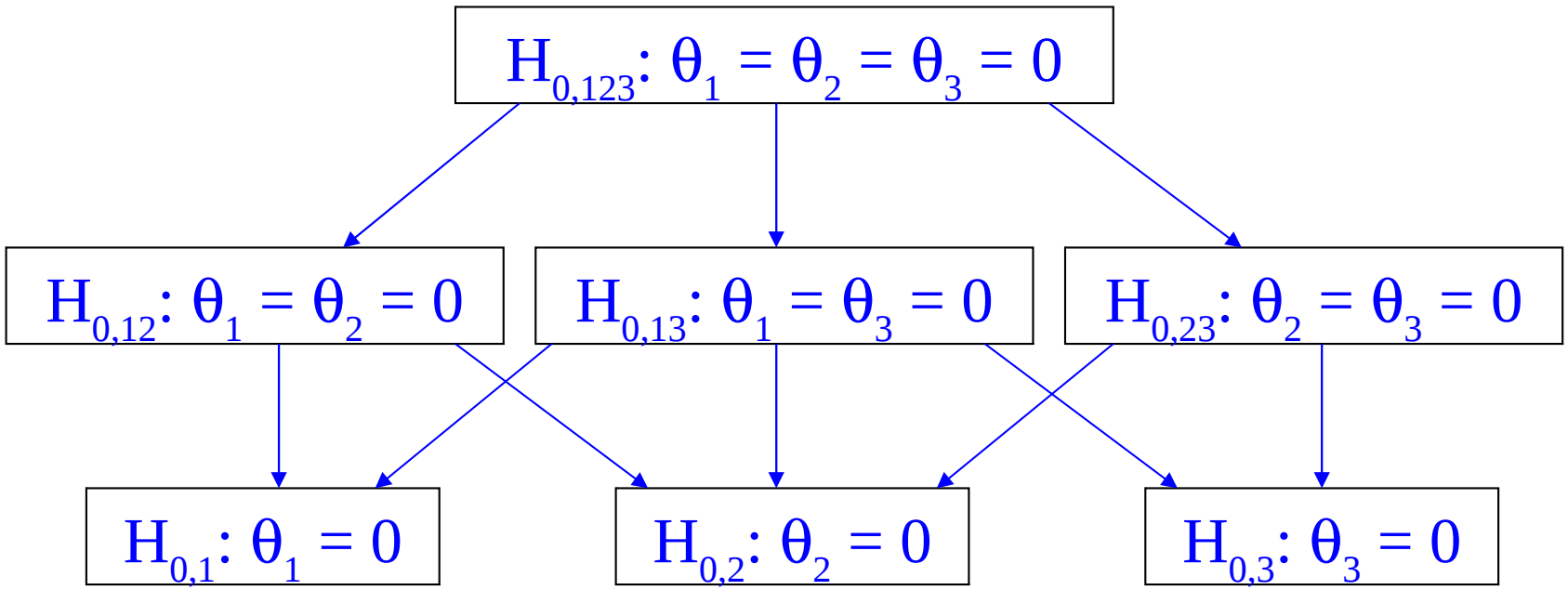
Consider one stage only

Closed testing procedure

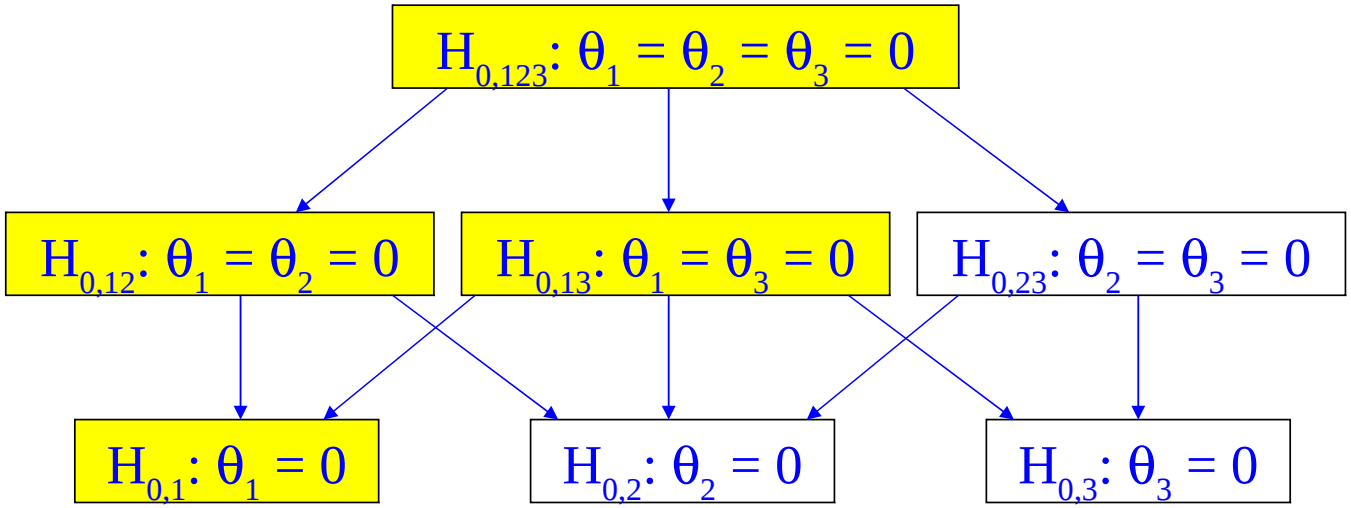
Form a family of H_0's from all possible intersections of $H_{0,i}$'s

Reject $H_{0,i}$ at level α, iff all H_0's that imply $H_{0,i}$ are rejected at level α.

e.g. 3 experimental treatments



Reject $H_{0,1}: \theta_1 = 0$, iff we reject



Let I be a set of m individual hypotheses from k ($m \leq k$)
$H_{0,I}$ denote the intersection hypothesis

$$H_{0,I} = \bigcap_{i \in I} H_{0,i}$$

e.g. if $I = \{1, 2\}$ then $H_{0,I} = H_{0,12}$

p-values for intersection hypotheses

We need to calculate p-value p_I to test $H_{0,I}$

Suppose we calculate a p-value, p_i, for each $H_{0,i}: \ \theta_i = 0$

Bonferroni correction

$$p_I = m \times \min(p_i)$$

where m is the number of indices in I

e.g. $k = 3, \ p_1 = 0.2, \ p_2 = 0.05, \ p_3 = 0.03$
then for $I = \{2, 3\}$ we find $m = 2, \ p_I = 0.03 \times 2 = 0.06$

Multi-arm multi-stage (Bauer & Kieser 1999; Bretz et al, 2006)

Using closed testing procedure we therefore need to conduct tests for $(2^k - 1)$ hypotheses.

For each hypothesis, $H_{0,I}$ calculate a p-value from data collected at stage j, $p_{I,j}$

Take p-values for $H_{0,I}$ convert into a test statistic
e.g. using the inverse normal method as

$$Y_{I,j} = \frac{1}{\sqrt{j}} \sum_{i=1}^{j} \Phi^{-1}\left(1 - p_{I,j}\right)$$

Compare test statistic for $H_{0,I}$ using the **same boundaries** calculated as for the two treatments (T_1 and T_0) case, as

adaptive approach adjusts for the multiple treatments via the p-values, not via the boundaries.

Stopping rules

Stop for efficacy

When conclude at least one T_i is superior to T_0 via the closed testing procedure
i.e when all test statistics for those H_0's that imply $H_{0,i}$ crosses the upper boundary, u_j

Stop for futility

When test statistic for global H_0 crosses the lower boundary, l_j or when maximum number of looks is reached

Example: Alzheimer's disease

- Primary endpoint: Alzheimer's disease assessment scale (ADAS) – cognitive portion. Assumed to be normal.
- Conduct 5 equally spaced looks
- 3 doses of an experimental treatment plus placebo
- Dose-response relationship could not be assumed

Therefore, 3 individual hypotheses of interest giving a total of 7 hypotheses to be tested

Can use the same boundaries as the 2 treatments case, e.g.

	1	2	3	4	5
u	3.03	2.37	2.19	2.15	2.16
l	-0.90	0.61	1.48	2.05	2.16

Consider testing $H_{0,123}: \ \theta_1 = \theta_2 = \theta_3 = 0$

Stage 1

Comparisons	P-value
18 mg/day vs Placebo	$p_{1,1} = 0.106$
24 mg/day vs Placebo	$p_{2,1} = 0.008$
36 mg/day vs Placebo	$p_{3,1} = 0.081$

Bonferroni adjusted test

$I = \{1, 2, 3\}, \ m = 3, \ p_{123,1} = 3 \times 0.008 = 0.024$
$Y_{123,1} = \Phi^{-1}(1 - p_{123,1}) = 1.98$
$l_1 \leq Y_{123,1} \leq u_1$

therefore continue to Stage 2.

Stage 2: Suppose only continue with first 2 doses and 1 control

Comparisons	P-value
18 mg/day vs Placebo	$p_{1,2} = 0.2$
24 mg/day vs Placebo	$p_{2,2} = 0.005$

Bonferroni adjusted test

$I = \{1, 2\},\ m = 2,\ p_{123,2} = 2 \times 0.005 = 0.01$

$Y_{123,2} = \frac{\Phi^{-1}(1-p_{123,1}) + \Phi^{-1}(1-p_{123,2})}{2^{\frac{1}{2}}} = 3.04$

$Y_{123,2} \geq u_2$ therefore reject $H_{0,123}: \ \theta_1 = \theta_2 = \theta_3 = 0$ at end of stage 2.

Now continue testing for other intersection hypotheses:

Comparisons	P-values	
18 mg/day vs Placebo	$p_{1,1} = 0.106$	$p_{1,2} = 0.2$
24 mg/day vs Placebo	$p_{2,1} = 0.008$	$p_{2,2} = 0.005$

Bonferroni adjusted test

$I = \{1, 2\},\ p_{12,1} = 2 \times 0.008 = 0.016,\ p_{12,2} = 2 \times 0.005 = 0.01$

$Y_{12,2} = \frac{\Phi^{-1}(1-p_{12,1})+\Phi^{-1}(1-p_{12,2})}{2^{\frac{1}{2}}} = 3.16$

$Y_{12,2} \geq u_2$ therefore reject $H_{0,12} :\ \theta_1 = \theta_2 = 0$ at end of stage 2.

Now would continue testing $H_{0,23}$.

Comments

- Sample size calculations
  - difficult under complex rules
  - larger then specialised methods
- Can do other adaptations, e.g.
  - select populations
  - Changing the primary endpoint
  - Changing the trial objective (such as switching from non-inferiority to superiority)
  - sample size re-estimation
- Ordering in treatment effects could be incorporated into closed testing procedure