An adaptive confirmatory trial with interim treatment selection: practical experiences and unbalanced randomisation

Caroline C. Morgan, Cardinal Systems, Paris

Joint work with Stephane Heritier and Serigne N. Lô, The George Institute, Sydney

Clinical Trials Methodology Conference, Bristol
Tuesday 4th October 2011

Overview

1. Adaptive Phase II/III designs with treatment selection
   - Basic principles
   - Ongoing Ph II/III trial in Europe, the Americas & Oceania
2. Posch *et al.* (2005) approach to treatment selection
   - Principles to combine Phases II and III
   - Combination tests for intersections
3. Unbalanced randomisation
4. Pre-study simulations
   - Type I error
   - Power
5. Conclusions
6. References

Combining exploratory and confirmatory phases

- A two-stage design consisting of:
  - a learning stage (Phase II)
  - a confirmatory stage (Phase III)
- Powered for the Phase III analysis
- Information from the learning stage is:
  - used to make design choices (*e.g.* select one or more treatments)
  - combined with information from the confirmatory stage for the Phase III analysis

Key points to consider

- Statistical methodology
- Logistics
- Operational bias

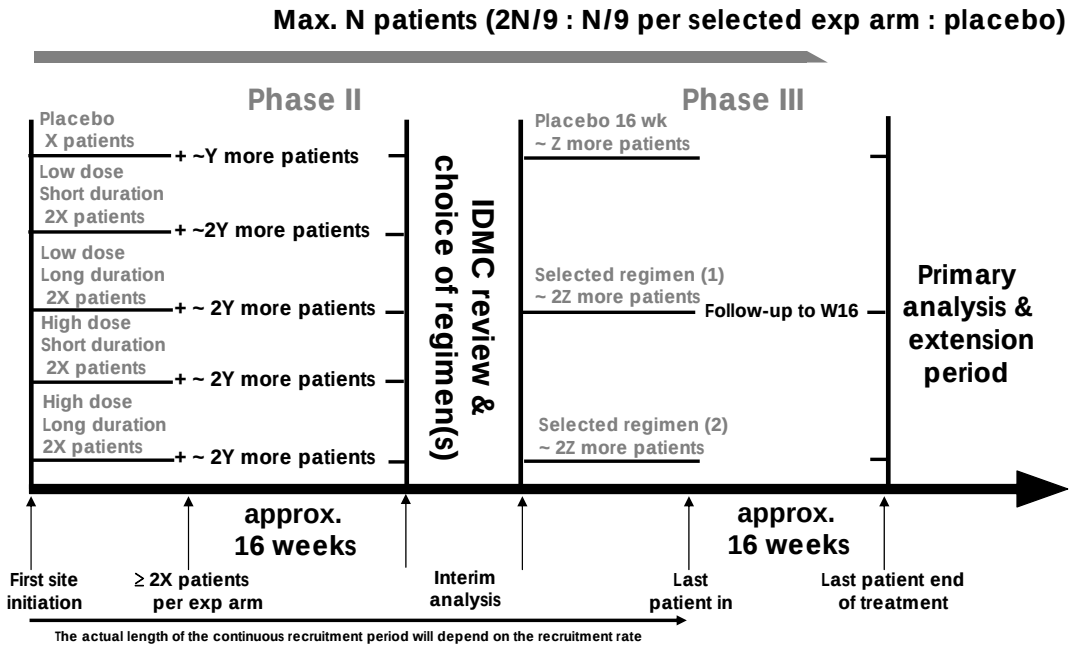
Case study

Ongoing Ph II/III trial in Europe, the Americas & Oceania

- Five arm study: 4 experimental regimens (high/low dose, long/short duration), 1 placebo
- Randomisation ratio: 2:2:2:2:1
- Treatment period: 16 weeks
- 90% power and 0.025 one-sided α
- Interim analysis: 40% of patients followed up for 16 weeks
- Primary analysis: all patients (both stages) followed up for 16 weeks
- Primary endpoint: success at 16 weeks (binary)
- Overrun: recruitment continued during 16-week follow-up and interim analysis period

Adaptive Phase II/III schema (unbalanced randomisation)

Recruitment and follow-up



Statistical principles

Posch *et al.* (2005) approach to treatment selection

- Very flexible methodology due to Bauer and Kieser (1999)
- Treatment selection does not inflate the Type I error rate
- Design based on 2 key statistical principles:
  1. Fisher's method of combining *p*-values (R.A. Fisher, 1932)
  2. The closure principle (closed testing procedure) for testing multiple hypotheses (Hochberg and Tamhane, 1987)

These principles may apply to all sorts of adaptations

Fisher's method for combining *p*-values

- Suppose p is the p-value from experiment 1 (here Phase II) and q is the p-value from an independently conducted experiment 2 (here Phase III).
- Then, under their respective null hypotheses, p and q are uniformly distributed on $[0, 1]$
- Fisher showed that $-2log(pq)$ follows a χ^2 distribution with 4 d.f. under the combined null hypothesis

Inverse χ^2 combination function used in adaptive designs
by Bauer and Köhne (1994), among others

The closure principle

Given a set of null hypotheses that are closed under intersection and an α-level test for each null hypothesis, the closure principle states that any null hypothesis can be rejected if:

1. it is rejected at level α
2. all other intersection test hypotheses that contain it are also rejected at level α

Closure principle applied to adaptive Phase II/III designs

One-sided (elementary) null hypothesis: $H_i : \theta_i \leq 0$, where θ_i is the difference (treatment i - placebo) in success rates.

In the case study described here, 4 experimental treatment regimens are investigated.

Suppose only Regimen 4 is selected at the interim analysis. To declare statistical significance at the end of the trial we must

- reject $H_4 : \theta_4 \leq 0$ with level α
- reject every intersection null hypothesis that contains H_4 at level α, *i.e.*
  - reject $H_{1,4} : \theta_1 \leq 0, \theta_4 \leq 0$, and $H_{2,4}$ and $H_{3,4}$ and
  - reject $H_{1,2,4} : \theta_1 \leq 0, \theta_2 \leq 0, \theta_4 \leq 0$, and $H_{1,3,4}$ and $H_{2,3,4}$ and
  - reject $H_{1,2,3,4} : \theta_1 \leq 0, \theta_2 \leq 0, \theta_3 \leq 0, \theta_4 \leq 0$.
- create an α-level combination test for all 8 hypotheses

Closed testing procedure applied using Simes' test and a combination function

Combine p-values from individual hypotheses within a stage using Simes' (1986) method

Combine p-values from the two stages using a combination function

Only Regimen 4 selected at the interim analysis:

- reject H_4 if $C(p_4, q_4) \leq C(\alpha)$
- reject $H_{1,4}$ if $C(p_{1,4}, q_4) \leq C(\alpha)$, similarly for $H_{2,4}$ and $H_{3,4}$
- reject $H_{1,2,4}$ if $C(p_{1,2,4}, q_4) \leq C(\alpha)$, similarly for $H_{1,3,4}$ and $H_{2,3,4}$
- reject $H_{1,2,3,4}$ if $C(p_{1,2,3,4}, q_4) \leq C(\alpha)$

Note: If two or more doses are carried forward, Simes' test is also applied to Stage 2

Two methods to calculate $C(p, q)$

Inverse χ^2 (Fisher):
$C(\alpha) = (1 - \alpha)$ quantile of the χ^2 distribution with 4 d.f.
$C(p, q) = -2log(pq)$

Weighted inverse normal function:
$C(\alpha) = \alpha$
$C(p, q) = 1 - \Phi(w_1\Phi^{-1}(1 - p) + w_2\Phi^{-1}(1 - q))$, where $w_1^2 + w_2^2 = 1$.

For any intersection test, the combination test statistic is
$Z = w_1\Phi^{-1}(1 - p) + w_2\Phi^{-1}(1 - q)$

Z is a weighted combination of 2 one-sided Z statistics
$Z_1 = \Phi^{-1}(1 - p)$ and $Z_2 = \Phi^{-1}(1 - q)$

$w_1 = (n_1/(n_1 + n_2))^{0.5}$ and $w_2 = (n_2/(n_1 + n_2))^{0.5}$

Combined *p*-values

For the weighted inverse normal function, we reject the null hypothesis that the selected dose has no effect if all intersection hypotheses are rejected, *i.e.*

$$\max(C(p_4, q_4), C(p_{1,4}, q_4), ...C(p_{2,3,4}, q_4), C(p_{1,2,3,4}, q_4) \leq \alpha$$

Notes

$\alpha = 0.025$, as one-sided tests are used

The combined test *p*-value = maximum above

Additional adaptations that may be incorporated at interim analysis include unblinded SSR - weights must remain unchanged for valid procedure

Unbalanced randomisation

Consider randomisation in the ratio r:1

Continuous endpoint: use weights based on the square root of the planned information increments

Binary endpoint: use the same weights so that

$w_1 = (n_{1P}/(n_{1P} + n_{2P}))^{0.5}$ and $w_2 = (n_{2P}/(n_{1P} + n_{2P}))^{0.5}$

where n_{1P} and n_{2P} are respectively the Stage 1 and 2 placebo sample sizes.

Other issues

- Small sample size for placebo arm in Stage 1
- Small placebo success rate expected

Unpooled Z-test inflates the Type I error rate when $r > 1$ and the Stage 1 placebo sample size is small to moderate (Eberhardt and Fligner, 1977)

Choice of test statistic

The Type I error rate was studied for the two combination tests (inverse normal and Fisher's product criterion) for various allocation ratios (r:1) and 3 stage-wise testing procedures:

Pooled Z-test

$$Z_{P,k} = \frac{\hat{p}_k - \hat{p}_0}{\sqrt{\tilde{p}_k(1-\tilde{p}_k)(1/n_{1E} + 1/n_{1P})}}$$

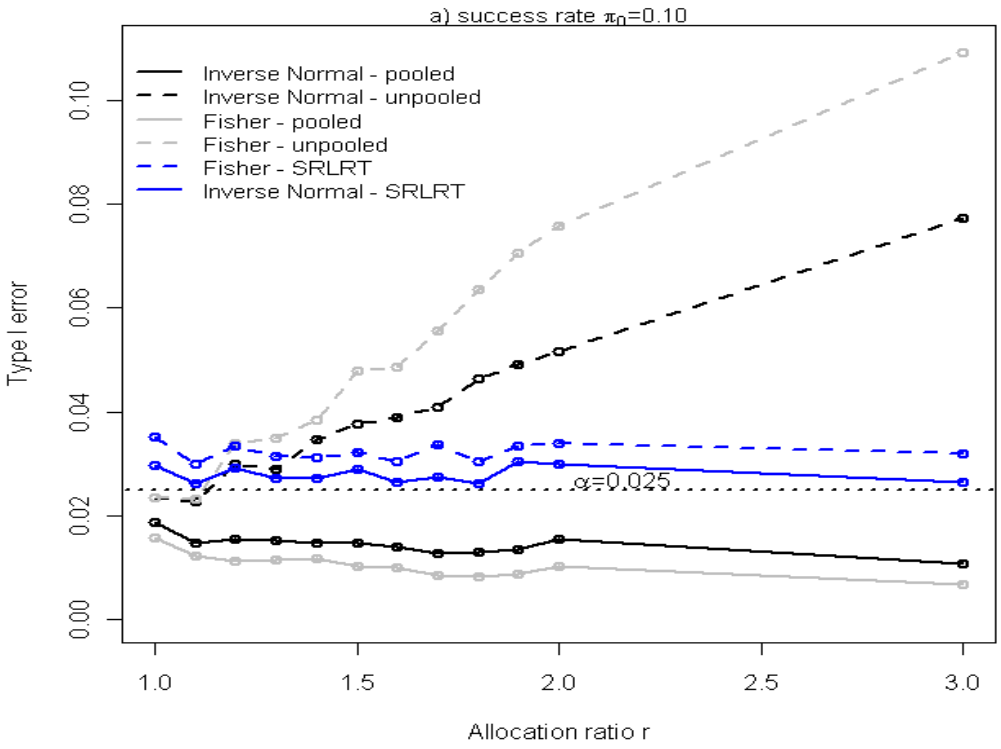
Unpooled Z-test

$$Z_{U,k} = \frac{\hat{p}_k - \hat{p}_0}{\sqrt{\hat{p}_k(1-\hat{p}_k)/n_{1E} + \hat{p}_0(1-\hat{p}_0)/n_{1P}}}$$

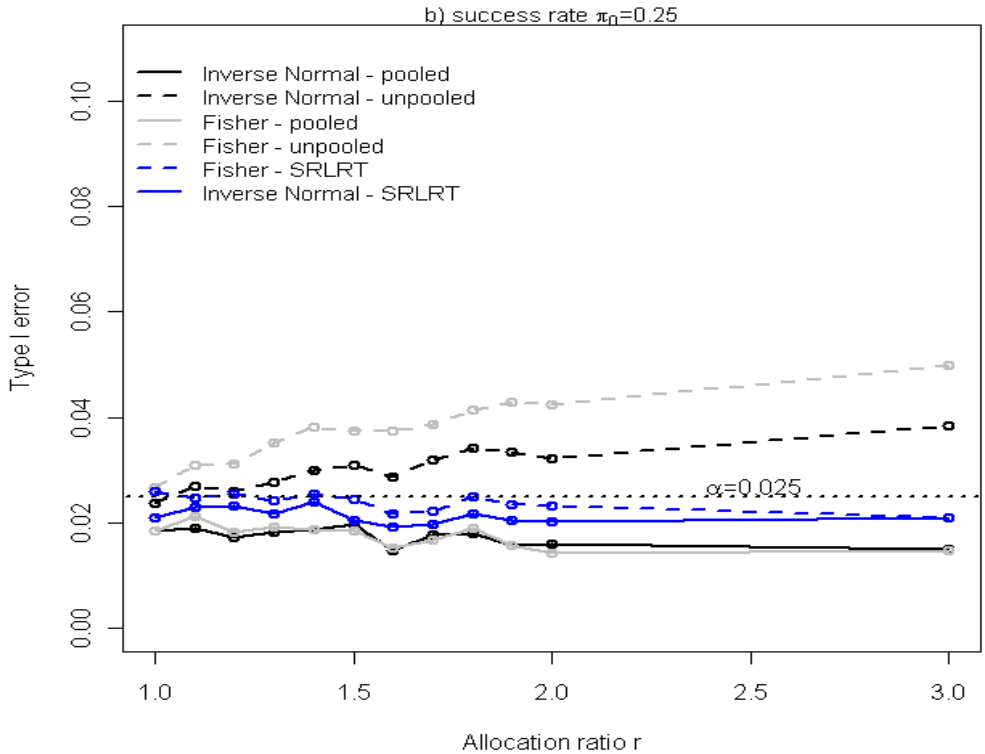
Signed root likelihood ratio test

$$SRLRT = \text{sign}(\hat{\beta}_k)\sqrt{2(LL(\hat{\alpha}_k, \hat{\beta}_k) - LL(\tilde{\alpha}_k, 0))}$$

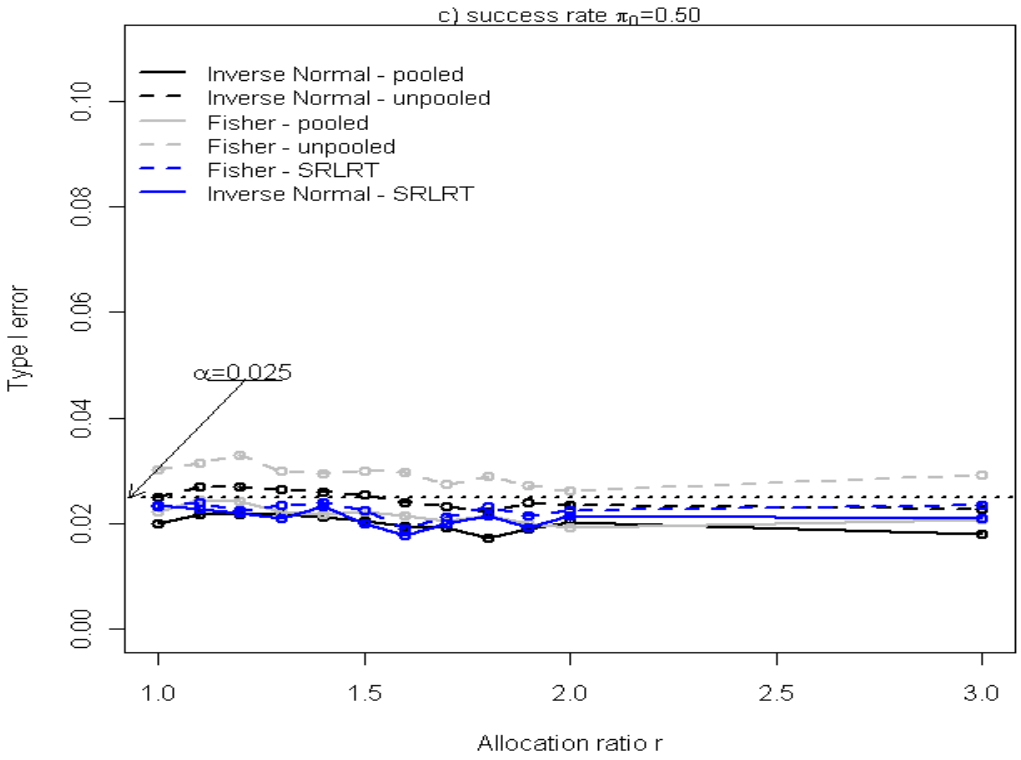
Type I error based on 10 000 simulations n_{1P}=25, n_{2P}=35, selection of most effective regimen



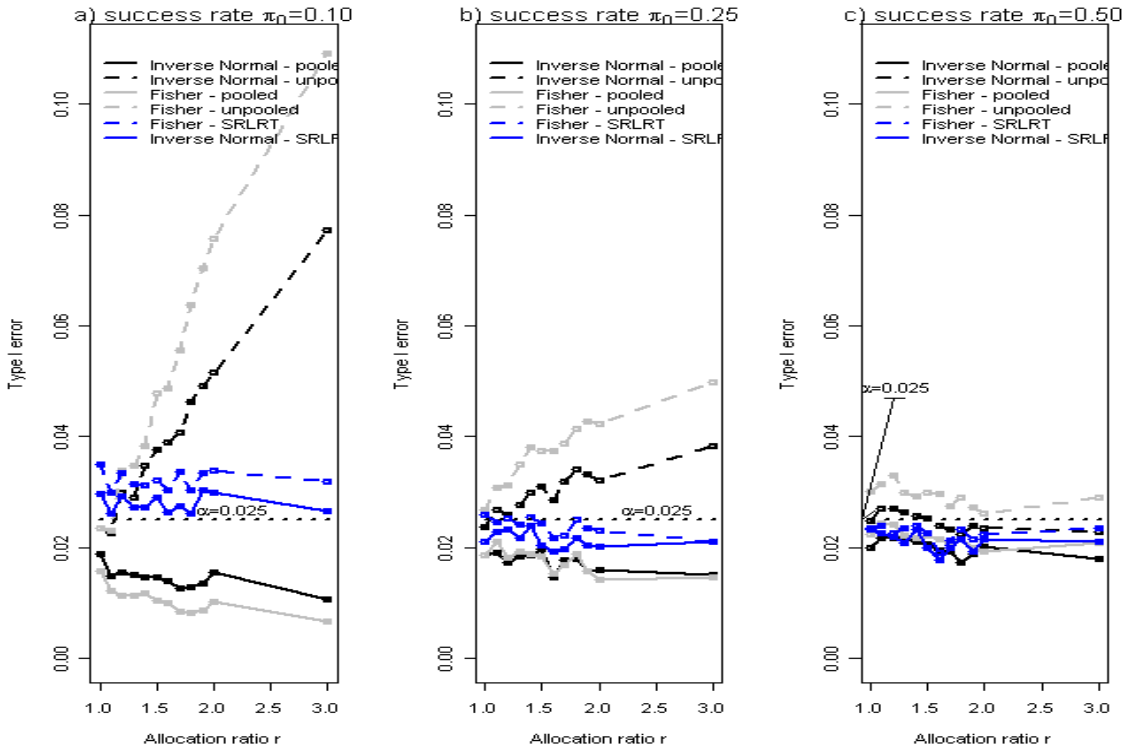
Type I error based on 10 000 simulations n_{1P}=25, n_{2P}=35, selection of most effective regimen



Type I error based on 10 000 simulations n_{1P}=25, n_{2P}=35, selection of most effective regimen



Type I error based on 10 000 simulations n_{1P}=25, n_{2P}=35, selection of most effective regimen



Assumptions

- Assumptions on success rates at the end of the treatment period (Week 16):
  - Placebo: 10%
  - Regimen 1 (low dose short duration): 25%
  - Regimen 2 (low dose long duration): 35%
  - Regimen 3 (high dose short duration): 40%
  - Regimen 4 (high dose long duration): 50%
- One-sided Type I error rate: α = 0.005
- Power > 80% for the case where the most effective lower dose is selected at the interim analysis

Power simulations

Power (superiority)					Sample size		
Most effective regimen selected					n_{1P}	n_{2P}	N
Any (SE)	Regimen 1	Regimen 2	Regimen 3	Regimen 4			
0.965 (0.0006)	0.001	0.050	0.152	0.762	20	25	255
0.978 (0.0005)	0.002	0.054	0.158	0.764	20	30	270
0.994 (0.0003)	0.001	0.041	0.144	0.807	25	35	330
0.996 (0.0002)	0.001	0.043	0.143	0.809	25	40	345
0.997 (0.0002)	0.001	0.043	0.144	0.809	25	45	360
Most effective 5 mg/kg/day regimen selected							
Any 5 mg (SE)	Regimen 1	Regimen 2	Regimen 3	Regimen 4			
0.636 (0.0015)	0.057	0.579	0	0	20	25	255
0.705 (0.0014)	0.067	0.637	0	0	20	30	270
0.813 (0.0012)	0.072	0.741	0	0	25	35	330
0.850 (0.0011)	0.078	0.772	0	0	25	40	345
0.881 (0.0010)	0.082	0.799	0	0	25	45	360

Second-order accuracy procedures

Wald-type tests and the SRLRT give *p*-values that are 1st order accurate, *i.e.* $O(n^{-1/2})$

Such tests (and particularly Wald-type tests) may inflate the Type I error rate, particularly for small sample sizes. Simulation is a must!

Second-order accuracy procedures with O(1/n) have started to emerge in the last decade:

- Accurate bootstrap *p*-values proposed by Lloyd in 2010 for logistic regression (absolute error O(1/n))
- Saddlepoint approximation for GLMs (relative err. O(1/n)), see Lô and Ronchetti 2009

Promising but computational challenges remain!

Conclusions

- Adaptive designs offer new perspectives but create new challenges logistically and methodologically
- Phase II/III adaptive designs with unbalanced randomisation are feasible
- Small Stage 1 sample sizes may be an issue and appropriate testing procedures must be used
- To construct persuasive flexible trial designs, careful planning and extensive simulation studies are required - early discussions with the regulators are a must for confirmatory trials
- The specific advantages and disadvantages of using an adaptive design must be studied on a trial by trial basis prior to deciding on whether or not it is suitable

References (1/3)

- Bauer P, Kieser M. Combining different phases in the development of medical treatments within a single trial. Stat Med 1999; 18:1833-48.
- Bauer P, Köhne K. Evaluation of experiments with adaptive interim analyses. Biometrics 1994; 50:1029-41. Correction Biometrics 1996; 52:380.
- Bretz F, König F, Brannath W, Glimm E, Posch M. Tutorial in Biostatistics: Adaptive designs for confirmatory clinical trials Stat Med 2009; 28:1181-217.
- Bretz F, Schmidli H, König F, Racine A, Mauer W. Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: General concepts. Biom J 2006; 48:623-34.
- Eberhardt KR, Fligner MA. A comparison of two tests for the equality of two proportions. Am Stat 1977; 31:151-55.
- Fisher RA. Statistical Methods for Research Workers. 4th edition, 1932. Edinburgh: Oliver & Boyd.

References (2/3)

- Heritier S, Lô SN, Morgan CC. An adaptive confirmatory trial with interim treatment selection: Practical experiences and unbalanced randomisation. Stat Med 2011; 30:1541-54.
- Hochberg Y, Tamhane AC. Multiple Comparison Procedures. Wiley, New York 1987.
- Jennison C, Turnbull BW. Mid-course sample size modification in clinical trials based on the observed treatment effect. Stat Med 2003; 22:971-93.
- Jennison C, Turnbull BW. Adaptive seamless designs: Selection and prospective testing of hypotheses. J Biopharm Stat 2007; 17:1135-61.
- Lloyd CJ. Bootstrap and second-order tests of risk difference. Biometrics 2010; 66:975-82.
- Lô SN, Ronchetti E. Robust and accurate inference for generalized linear models. J Mult Anal 2009; 100:2126-36.

References (3/3)

- Posch M, König F, Branson M, Brannath W, Dunger-Baldauf C, Bauer P. Testing and estimation in flexible group sequential designs with adaptive treatment selection. Stat Med 2005; 24:3697-714.
- Simes RJ. An improved Bonferroni procedure for multiple tests of significance. Biometrika 1986; 73:751-4.
- Schmidli H, Bretz F, Racine A, Maurer W. Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: Applications and practical considerations. Biom J 2006; 48:635-43.
- European Agency for the Evaluation of Medicinal Products Evaluation of Medicines for Human Use. Points to Consider on Methodological Issues in Confirmatory Clinical Trials with Flexible Design and Analysis Plan. CPMP/EWP/2459/02, 2006. London, UK.
- FDA Draft Guidance ‘Adaptive Design Clinical Trials for Drugs and Biologics’, February 2010.