On the use of conventional tests in flexible, multiple test designs

Franz König

Peter Bauer Werner Brannath

Section of Medical Statistics
Medical University of Vienna, Austria

MCP, 2007

Flexible (Adaptive) Versus Frequentist Trial

Classical frequentist trials

- details of design and analysis must be prefixed in advance (population, treatments, doses, main and secondary outcome variable(s), analysis strategy, sample sizes,...)
- lack of flexibility to react to information from inside or outside the trial

Flexible (adaptive) design (Bauer & Kieser 99; Hommel 01; ...)

- allow for mid-trial design modifications based on all internal and external information gathered at interim analyses without compromising the type I error rate
- To control the type I error rate, the design modifications need **not** be specified in advance.

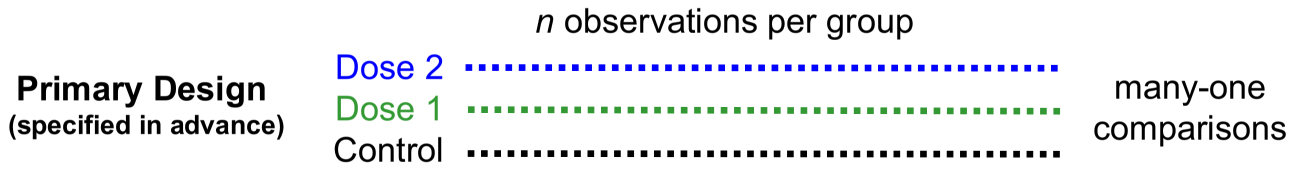
The use of conventional tests in flexible designs

Flexible Designs...

... were criticized because patients may be unequally weighted after adaptations.

Can we ignore the unplanned adaptations and follow the *"one patient = one vote"* philosophy and analyze the data in the conventional way by keeping full flexibility?

Dose selection and efficacy testing



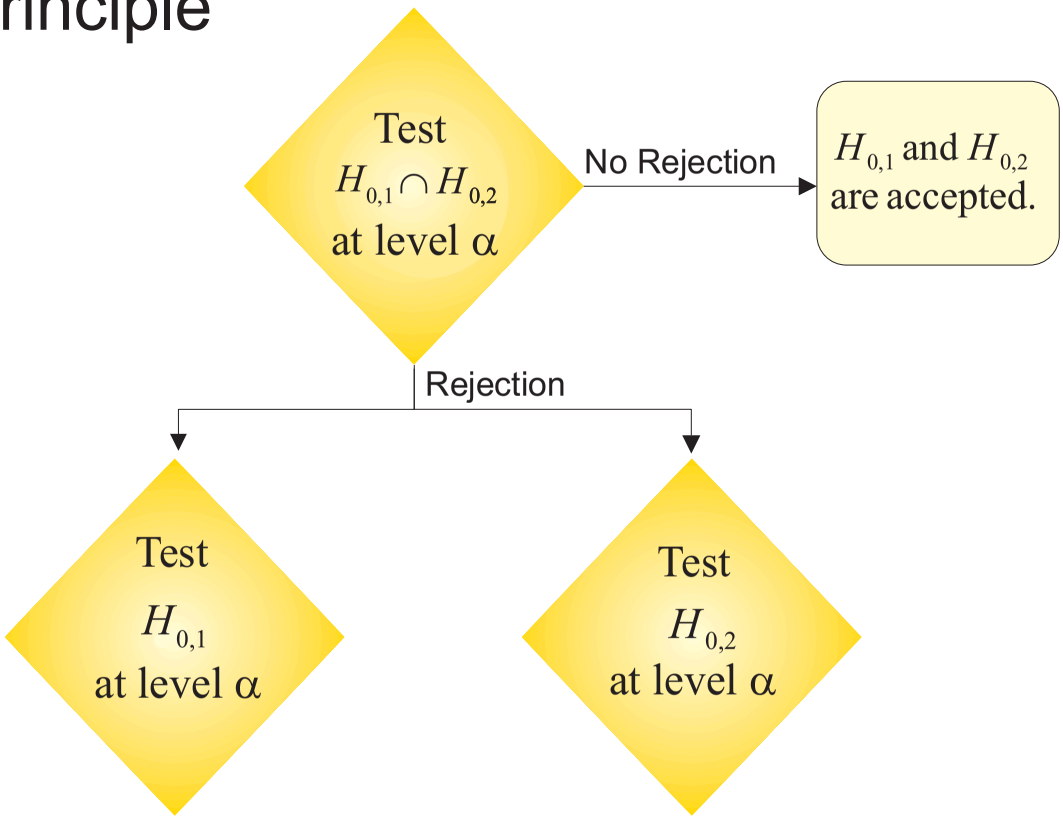
- Parallel group design with 2 dose groups and a control group.
- Testing the one sided hypotheses (common known σ)

$$\text{dose 1:} \quad H_{0,1} : \mu_1 \leq \mu_0 \quad \text{vs.} \quad H_{1,1} : \mu_1 > \mu_0$$

$$\text{dose 2:} \quad H_{0,2} : \mu_2 \leq \mu_0 \quad \text{vs.} \quad H_{1,2} : \mu_2 > \mu_0$$

- Regulatory bodies ask for a control of the multiple type I error rate at level $\alpha = 0.025$.

The Closed Testing Principle



Closed Testing Principle - Two Examples

Notation: Z_i is z-statistics for comparing dose i to control.

Hierarchical test ($H_{0,2} \rightarrow H_{0,1}$)

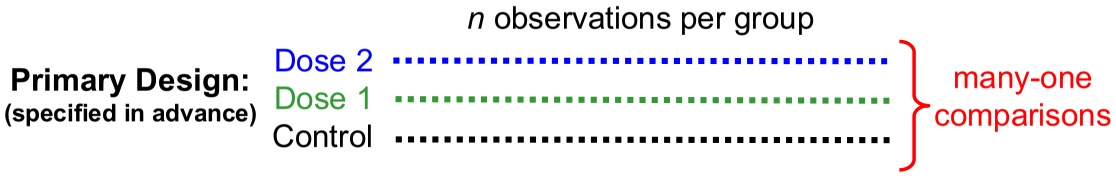
- Proc.: Z-test for $H_{0,2}$ $\overset{\text{rejection}}{\Longrightarrow}$ Z-test for $H_{0,1}$

- Test for $H_{0,1} \cap H_{0,2}$: $Z_2 \geq z_{1-\alpha}$
- Test for $H_{0,i}$: $Z_i \geq z_{1-\alpha}$

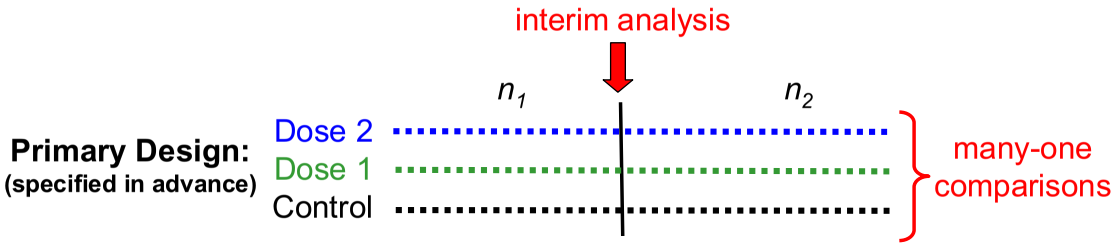
Step-down Dunnett test

- Test for $H_{0,1} \cap H_{0,2}$: $\max(Z_1, Z_2) \geq d$
  where d is Dunnett critical boundary ($d > z_{1-\alpha}$)
- Test for $H_{0,i}$: $Z_i \geq z_{1-\alpha}$

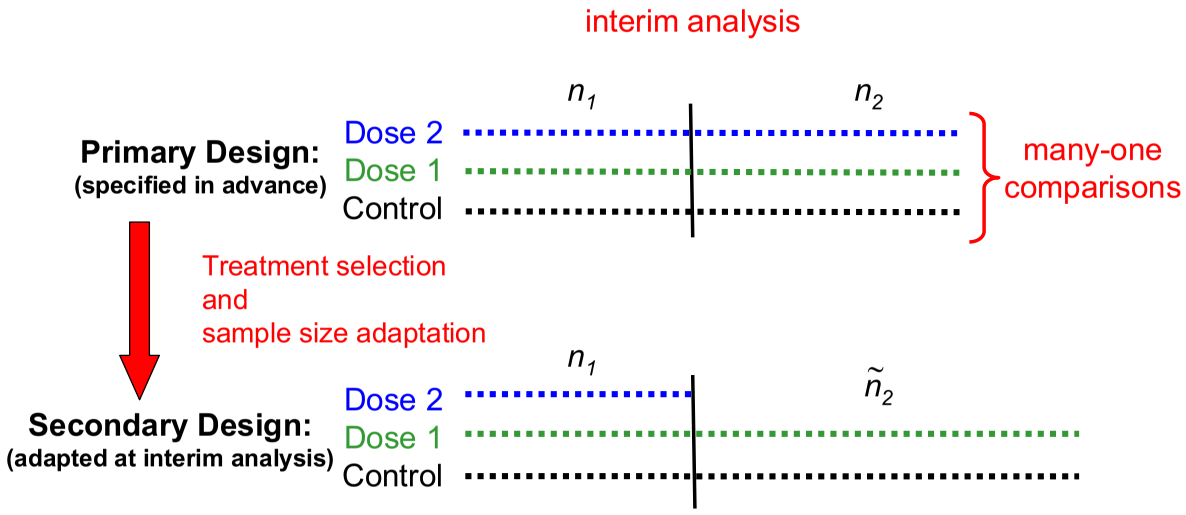
Dose selection and sample size adaptation



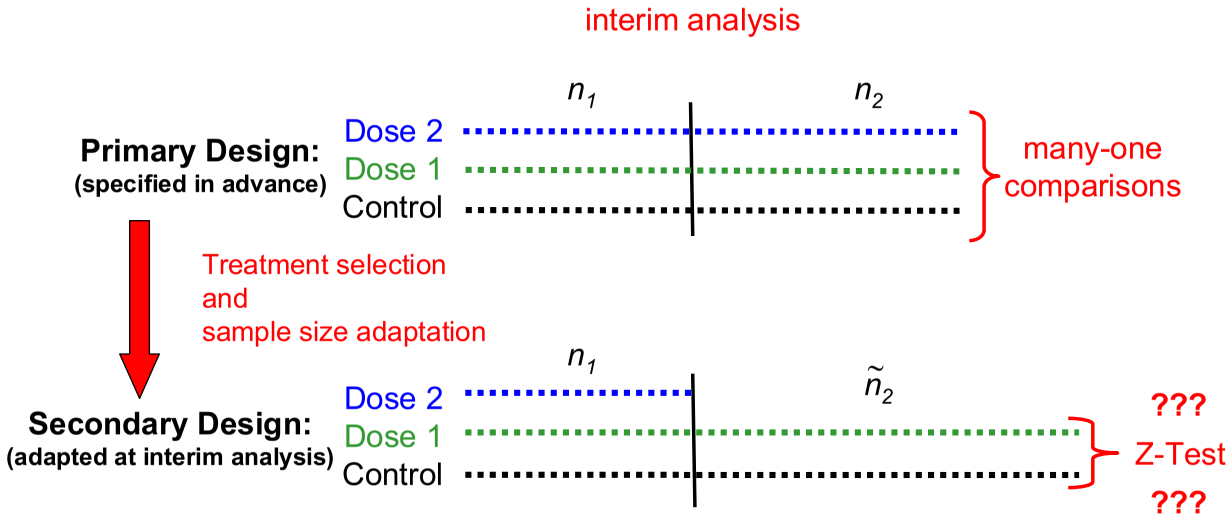
Dose selection and sample size adaptation



Dose selection and sample size adaptation



Dose selection and sample size adaptation



Conditional error rate principle

(MÜLLER & SCHÄFER 2001, 2004 . . .)

Primary test for H (specified in advance) $\longmapsto$ **Secondary test for H (adapted at stage 1)**

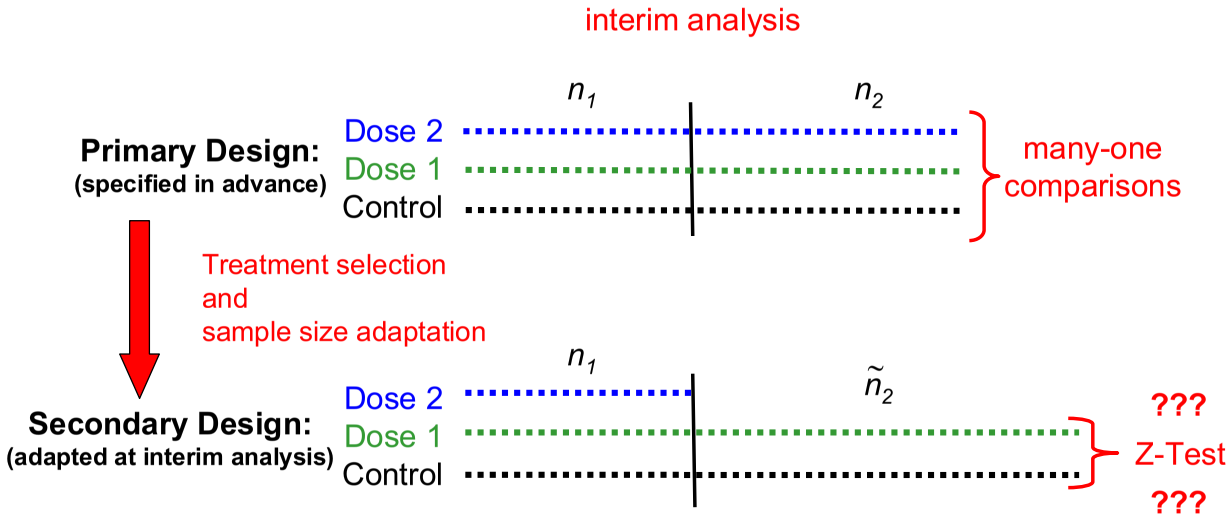
$$A = P_0(\text{reject H} \mid \text{interim data}) \qquad \widetilde{A} = P_0(\text{reject H} \mid \text{interim data})$$

The type I error rate is preserved if always

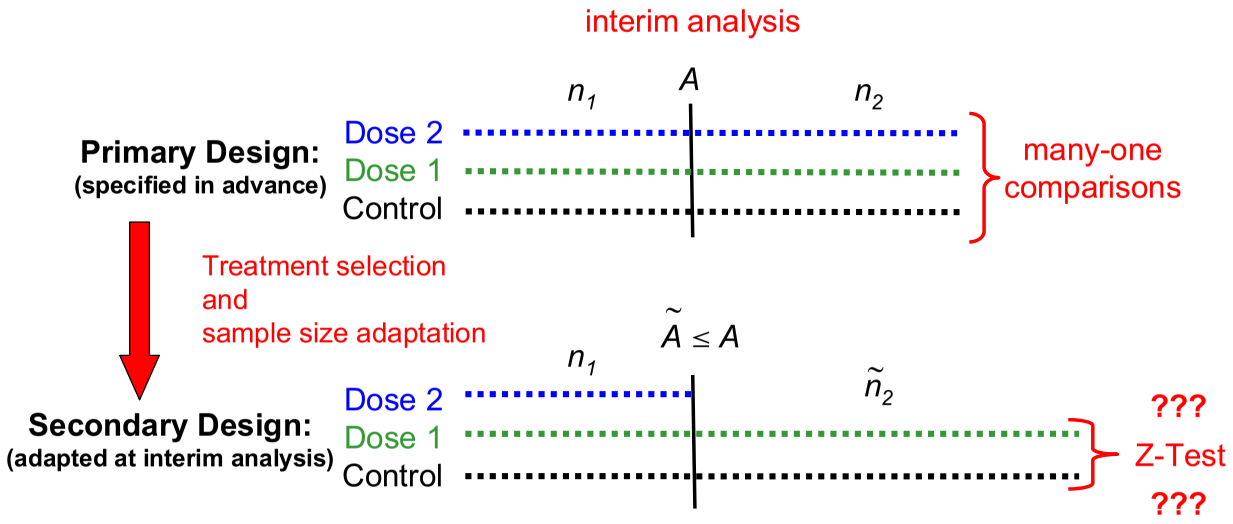
$$A \geq \widetilde{A}$$

$$(\alpha = E_H(A) \geq E_H(\widetilde{A}) = \tilde{\alpha})$$

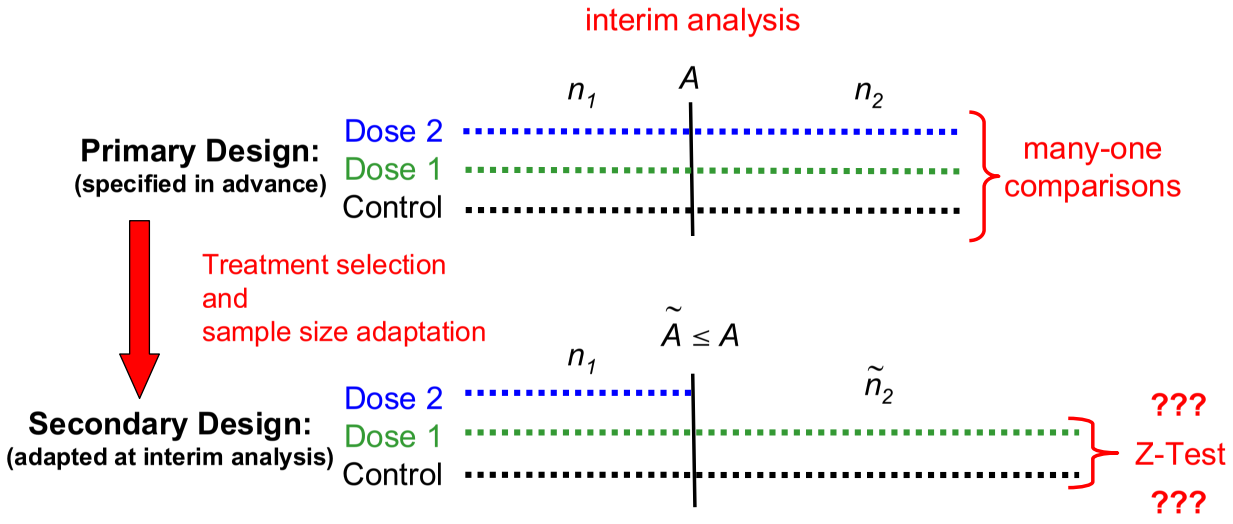
Dose selection and sample size adaptation



Dose selection and sample size adaptation



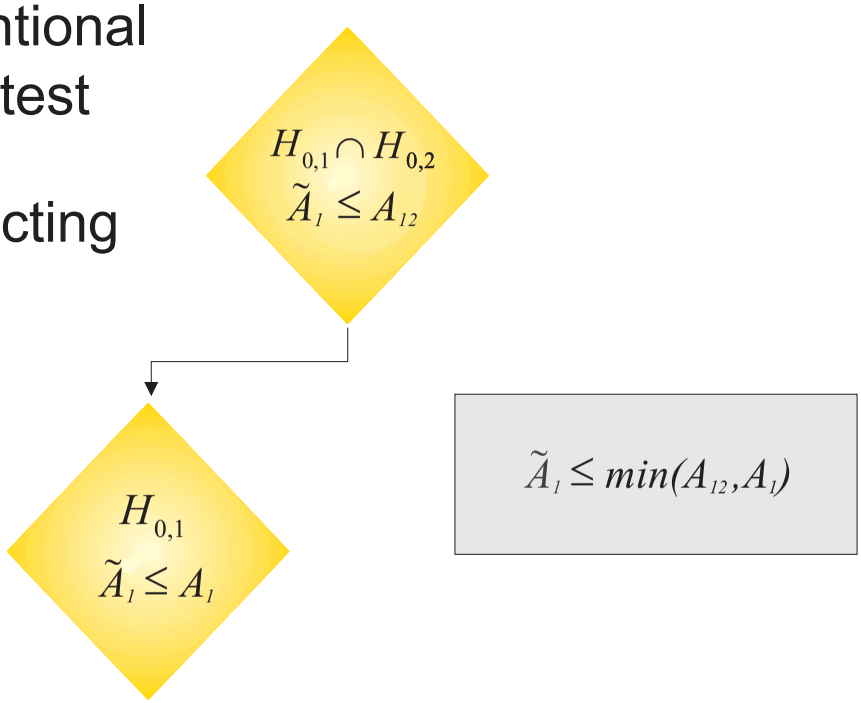
Dose selection and sample size adaptation



In which cases can we apply the conventional level α z-test for the selected dose?

Flexible Switching to conventional level α z-test

after selecting dose 1



Answer: We can switch to conventional *z*-test for dose 1 if $\tilde{A}_1 \leq \min(A_{12}, A_1)$ with $\tilde{A}_1 =$ conditional error rate of conv. *z*-test for dose 1.

Hierarchical test: when can we select the higher dose?

Using the hierarchical test and keeping the preplanned sample size we can always drop the lower dose without inflating the type I error rate.

- What is about sample size reassessment? (SEE POSCH ET AL 2003).
- When can we drop the higher dose and continue with lower dose only and use the conventional *z*-test?

Hierarchical test: when can we select the lower dose?

(BRANNATH, KÖNIG AND BAUER, 2007)

We can select dose 1 and use the conventional z-test if

$$\tilde{A}_1(z_1^{(1)}) \leq \min\{A_2(z_2^{(1)}), A_1(z_1^{(1)})\} \qquad (1)$$

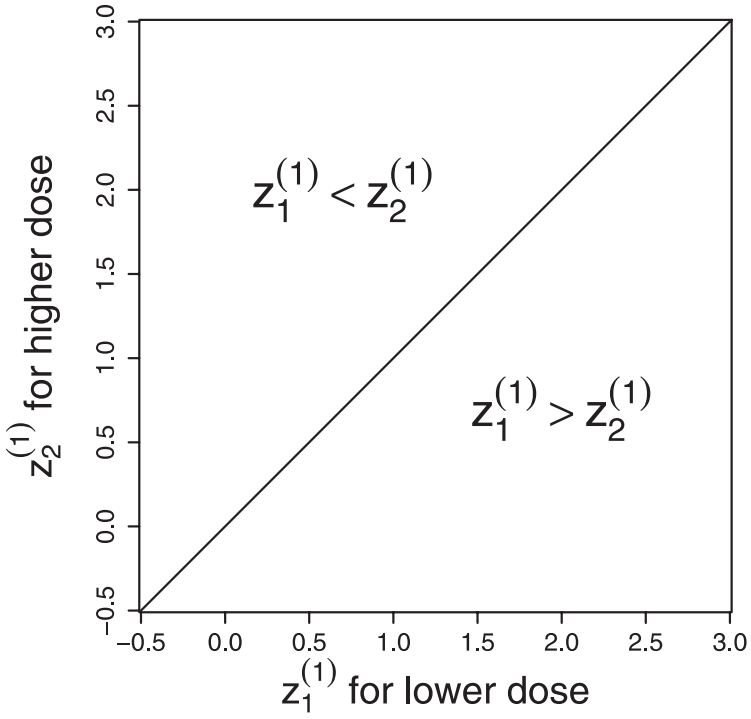
$z_i^{(1)}$ = standardized mean difference at stage 1 for dose i,
$A_i(z_i^{(1)})$ = conditional type I error rate of prim. z-test for dose i.

Numerical investigations:

- Determine $z_1^{(1)}, z_2^{(1)}$ for which (1) is satisfied.

Hierarchical test: when can we select the lower dose?

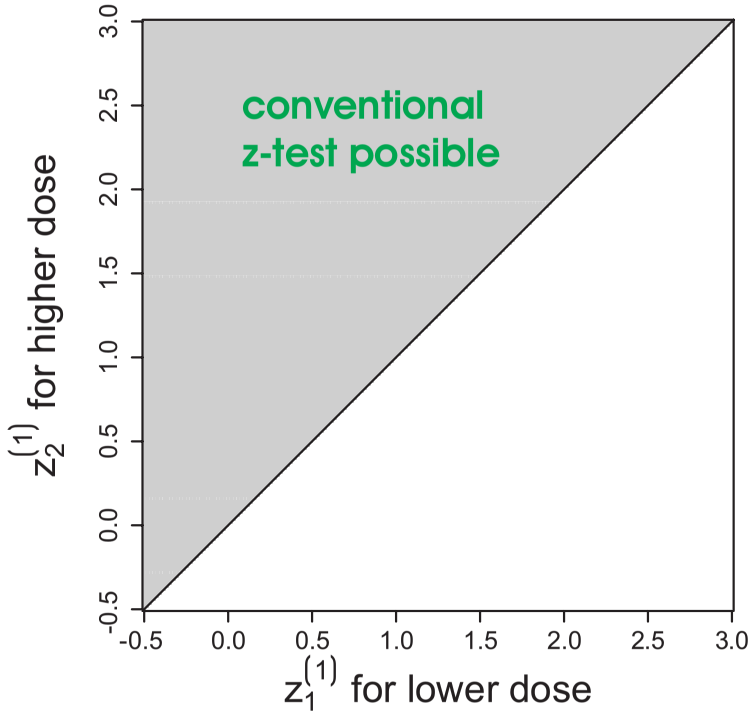
Interim sample space



- $\alpha = 0.025$
- Interim analysis at $n_1 = 0.5\, n$

Hierarchical test: when can we select the lower dose?

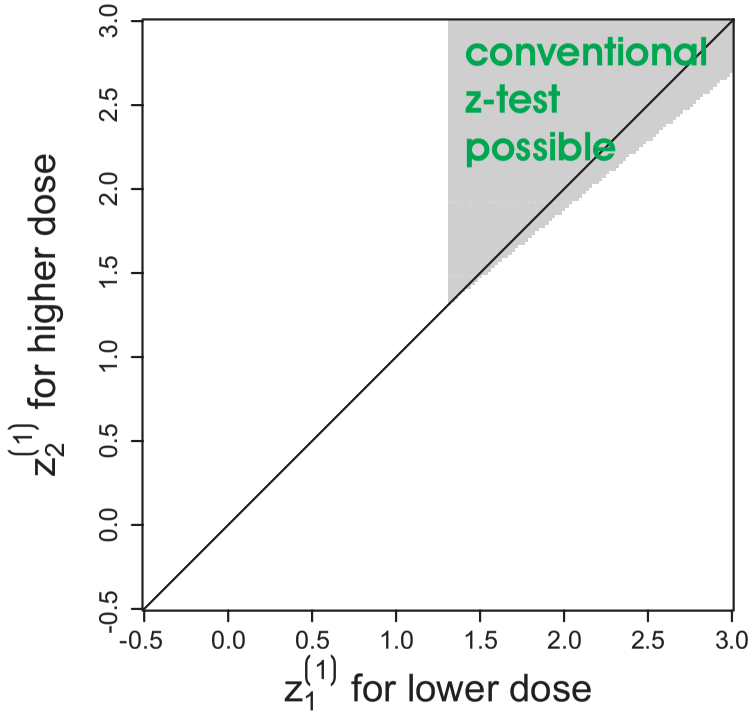
Preplanned sample size



We can select dose 1 if and only if less promising than dose 2.

Hierarchical test: when can we select the lower dose?

Sample size reshuffling



We can select dose 1 (also in cases where it is slightly more promising than dose 2) only if doses 1 and 2 have large interim effects.

Discussion for hierarchical test

- **Preplanned sample size:** If the lower dose has the smaller interim effect, we can always select only the lower dose and apply the conventional level α z-test.
- **Sample size reshuffling:** We can select dose 1 only when observing large interim effects for both doses. In this case dose 1 can be selected even if slightly more promising than dose 2.

Dunnett test: when can we select one dose?

We can select dose s and use the conventional z-test instead of the pre-planned step-down Dunnett test if

$$\tilde{A}_s(z_s^{(1)}) \leq \min\{A_{12}(z_1^{(1)}, z_2^{(1)}), A_s(z_s^{(1)})\} \qquad (2)$$

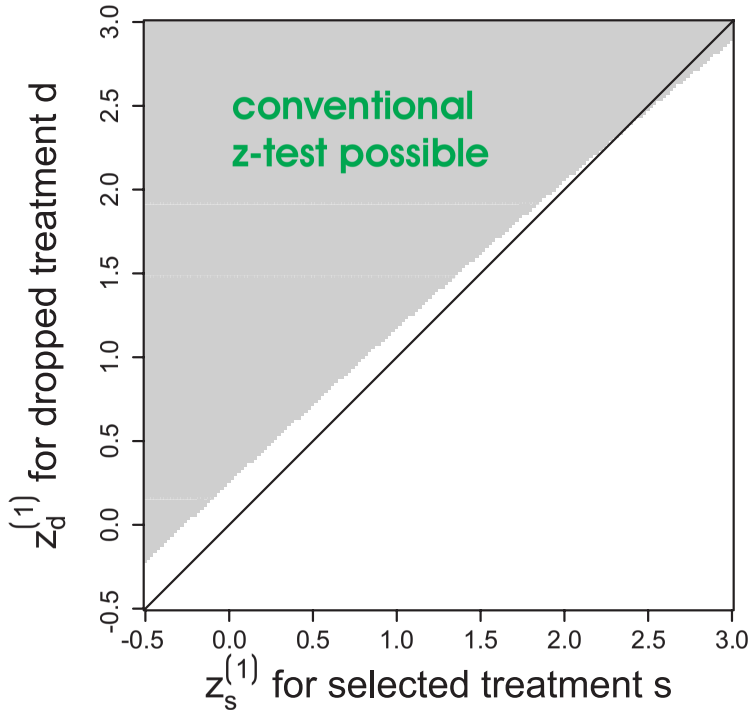
$z_i^{(1)}$ = standardized mean difference at stage 1 for dose i,
$A_s(z_s^{(1)})$ = conditional type I error rate of prim. z-test for dose s
$A_{12}(z_1^{(1)}, z_2^{(1)})$ = cond. type I error rate of prim. Dunnett test.

Numerical investigations:

- Determine $z_s^{(1)}, z_d^{(1)}$ $(d \neq s)$ for which (2) is satisfied.

Dunnett test: when can we select one dose?

Preplanned sample size



We cannot always go with the less promising treatment!

For very large interim effects we can sometimes choose the more promising treatment (if the interim effects are close enough).

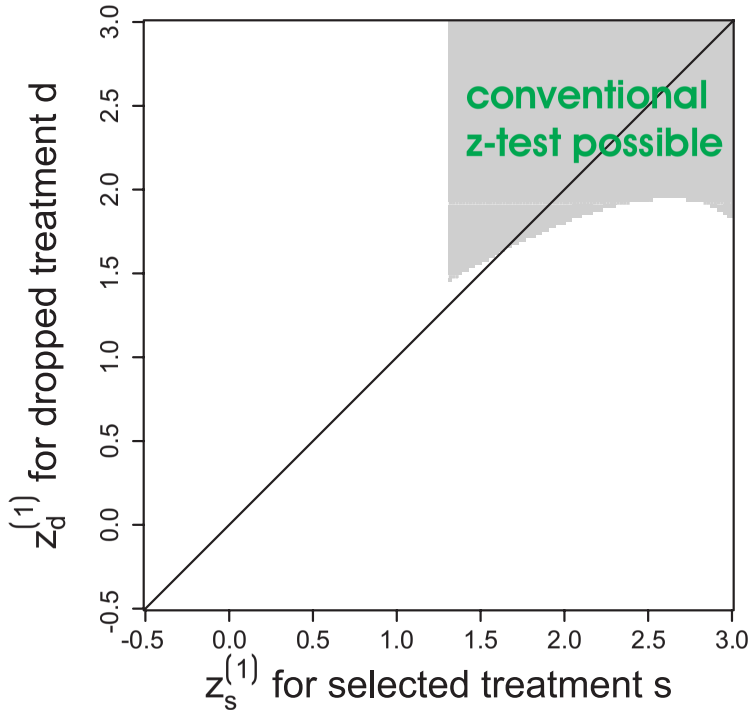
Full Flexibility versus Pre-planned Adaptivity

This example shows the fundamental difference between full flexibility and pre-planned adaptivity

- **Pre-planned adaptivity:** Note that the fixed rule where we always go with the less promising treatment cannot inflate the multiple type I error rate.
- **Full Flexibility:** Utilizing full flexibility and going with the less promising treatment only in the case of two similar and small interim effects (and otherwise staying with both treatments) will inflate the multiple type I error rate.

Dunnett test: when can we select one dose?

Sample size reshuffling



If both interim effects are large enough then we go with any of the two treatments.

Discussion for step-down Dunnett test

- **Preplanned sample size:** Starting with the step-down Dunnett test we CANNOT always select the less promising treatment at stage 1. If both treatments are very promising than we can also select the slightly more promising treatment.
- **Sample size reshuffling:** If both treatments have sufficient large interim effects then we can select any of the treatments.

Summary

- We have discussed the question when we can select one of two treatments and apply the conventional level α z-test.
- We have seen that the answer can be given by comparing conditional type I error rates.
- We have investigated this for the case of sample size reassessments and treatment selection for the hierarchical test and the step-down Dunnett test.

Summary

- Following the *"one patient = one vote"* philosophy and analyzing the data of a flexible design in a conventional way seems to be possible only in exceptional situations.
- Hence, flexible closed tests are, in general, unavoidable if the adaptation rule is not completely pre-specified (e.g., Adaptive Dunnett Tests, see Koenig et al, 2007).

Selected References

- Bauer P, Kieser M. Combining different phases in the development of medical treatments within a single trial. *Statistics in Medicine* 1999; **18**:1833–1848.
- Hommel G. Adaptive modifications of hypotheses after an interim analysis. *Biometrical Journal* 2001; **43**:581–589.
- Müller HH, Schäfer H. Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches. *Biometrics* 2001; **57**:886–891.
- Martin Posch M, Koenig F, Branson M, Brannath W, Dunger-Baldauf C, Bauer P. Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Statistics in Medicine* 2005; **30**:3697-3714.
- König F, Bauer P, Brannath W. An adaptive hierarchical test procedure for selecting safe and efficient treatments. *Biometrical Journal* 2006; **48**:663-678.
- König, F., Brannath, W., Bretz, F., Posch, M. Adaptive Dunnett tests for treatment selection. *submitted*
- Brannath, W., König, F., Bauer, P. Flexibility and multiplicity in clinical trials. *accepted*