Closed Testing Principle in Adaptive Designs

Gernot Wassmer

BBS Training Course
Basel, 13 September 2022

Clue of the Adaptive Test

- Do not pool the data of the stages, combine the stage-wise p-values.
- Then the distribution of the combination function under the null does not depend on design modifications and the adaptive test is still a test at the level α for the modified design.
- In the two stages, different hypotheses H_{01} and H_{02} can be considered, the considered global test is a test for $H_0 = H_{01} \cap H_{02}$.
- Or there are multiple hypotheses at the beginning of the trials and maybe some selected.
- Or there will be even hypotheses to be added at an interim stage (not of practical concern).
- The rules for adapting the design need not to be prespecified!

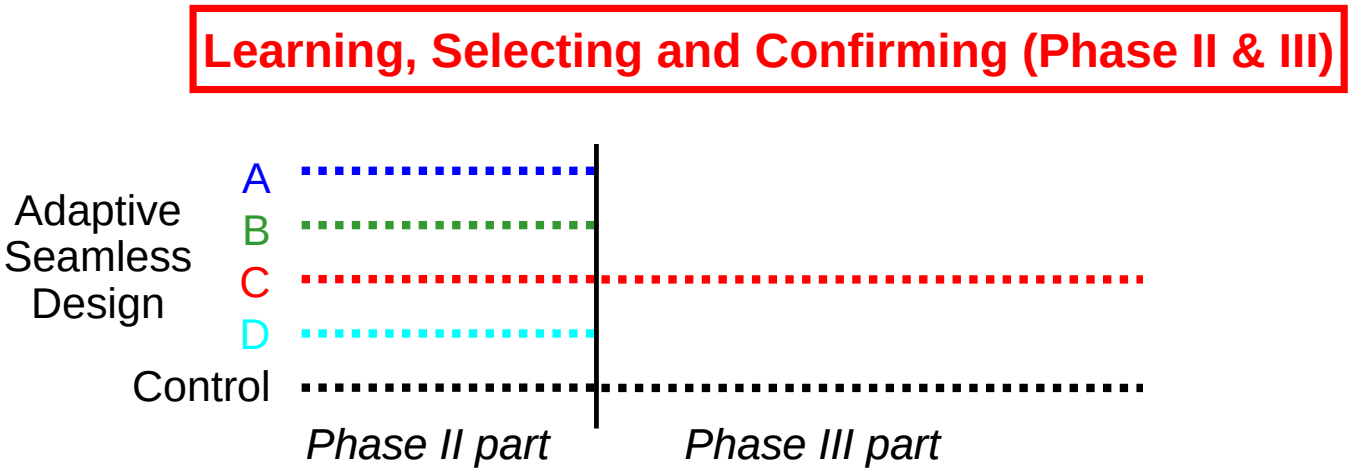
Possible Data Dependent Changes of Design

Examples of data dependent changes of design are

- Sample size recalculation
- Change of allocation ratio
- Change of test statistic
- Flexible number of looks
- Treatment arm selection (seamless phase II/III)
- Population selection (population enrichment)
- Selection of endpoints

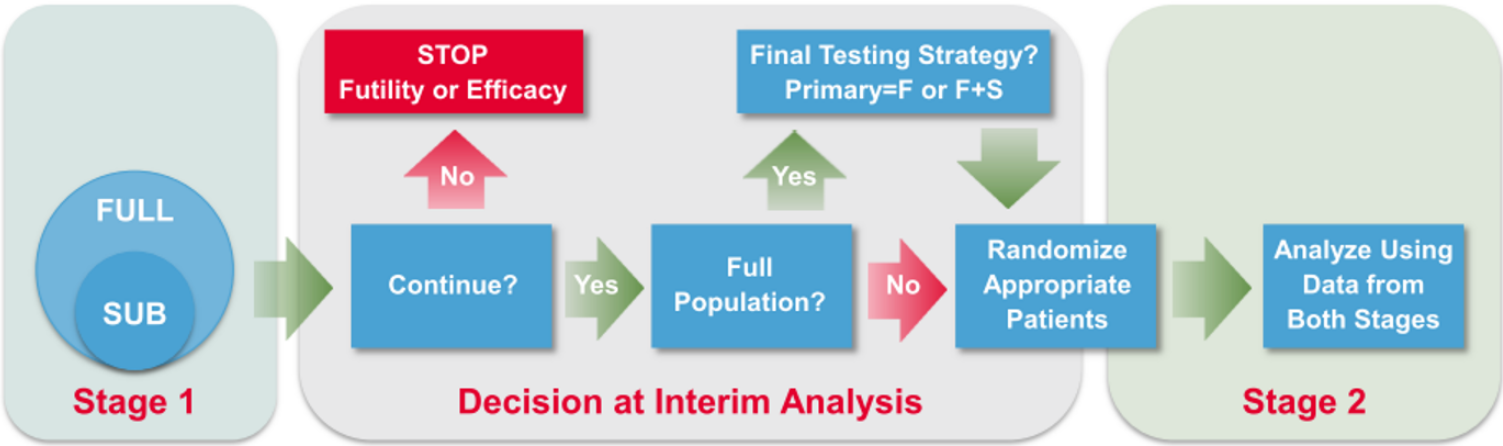
For the latter three, in general, multiple hypotheses testing applies and a closed testing procedure can be used in order to control the experimentwise error rate in a strong sense.

Seamless Phase II/III Trials: Treatment Arm Selection



- Conduct phase II trial as internal part of a combined trial
- Plan phase III trial based on data from phase II part
- Conduct phase III trial as internal part of the same trial
- Demonstrate efficacy with data from phase III + II part

Enrichment: Phase 2/3 Study in HER2- Early Stage BC



- Stage 1 objective
  - Stop for futility/efficacy
  - To continue with HER2- (Full) population – Broad Label (F) or Enhanced Label (F+S)
  - To confirm greater benefit in TNBC Subpopulation – Restricted Label (S)
  - To adjust the sample size
- Stage 2 data and the relevant groups from Stage 1 data combined

Sources for alpha Inflation

- Interim analysis
- Sample size reassessment
- Multiple hypotheses

The proposed adaptive procedure fulfills the regulatory requirements for the analysis of adaptive trials as it strongly controls the prespecified multiple Type I error rate (strong control of familywise error rate).

Multiple Type I Error Rate

Multiple Type I error rate =

Probability to reject **at least one** true null hypothesis.

(Probability to declare at least one ineffective treatment as effective).

Strong control of multiple Type I error rate:

Regardless of the number of true null hypotheses (ineffective treatments):

$$\text{Multiple Type I error rate } \leq \alpha$$

Methods

- **Methods for predefined selection rules**
  (Stallard & Todd 2003, . . .)

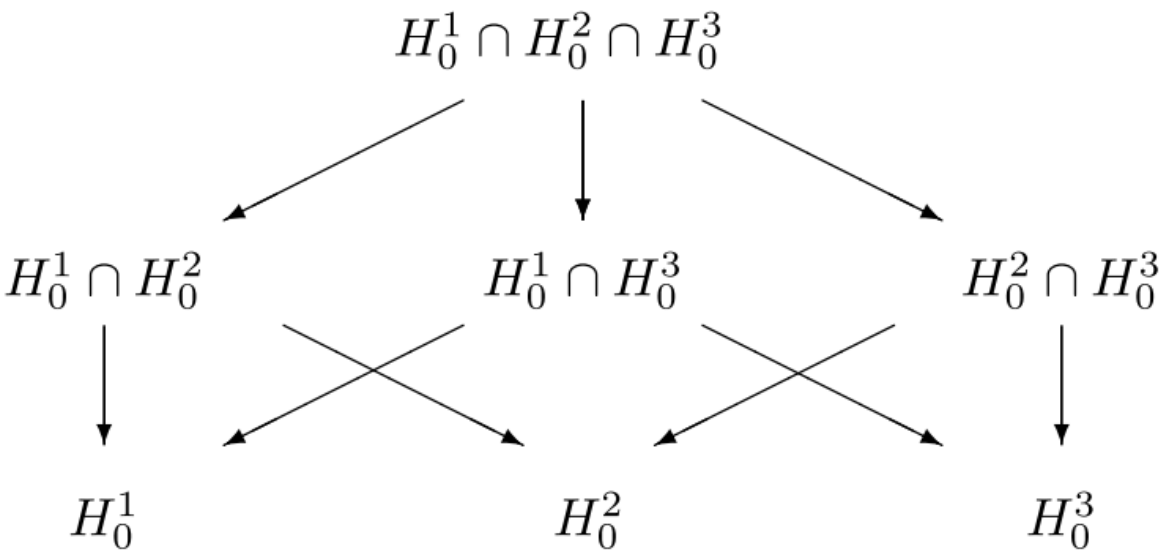
- **Flexible Two Stage Closed Tests**
  (Bauer & Kieser 1999; Hommel 2001; ...)

  – Do not require a predefined treatment and sample size selection rule.

  – Combine two methodology concepts:

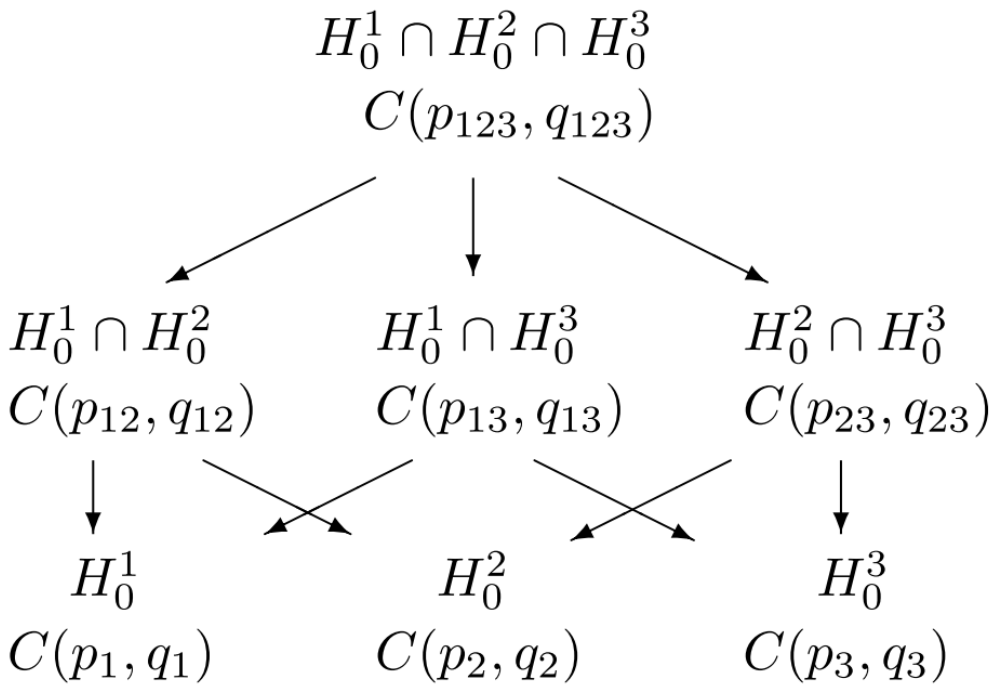
  Combination Tests and Closed Testing Principle.

Closed Testing Principle, 3 Hypotheses



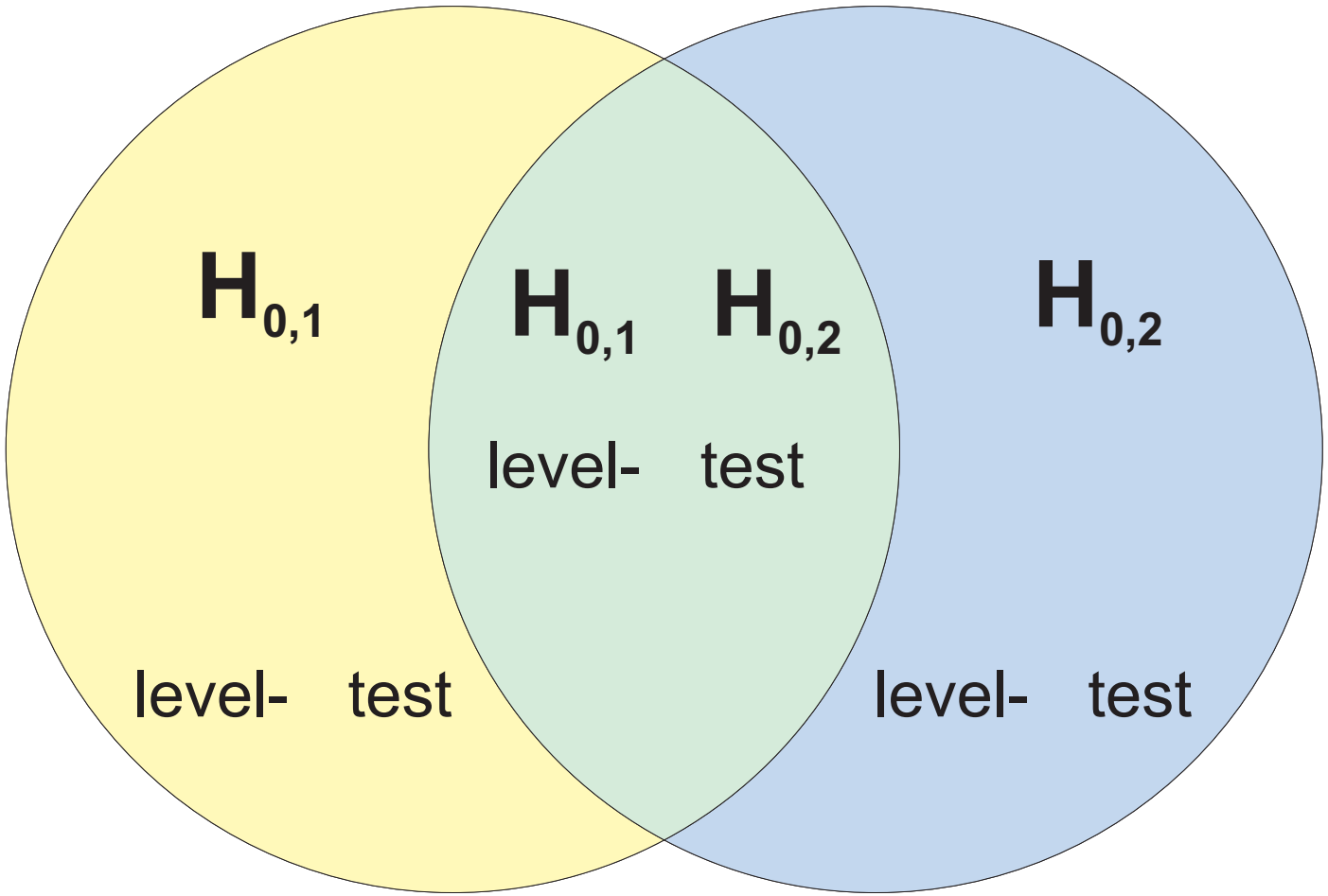
Closed test procedure for three null hypotheses H_0^1, H_0^2, and H_0^3. Arrows point in the direction of the next hypothesis that can be tested if we reject the current null.

Closed Testing Principle, 3 Hypotheses

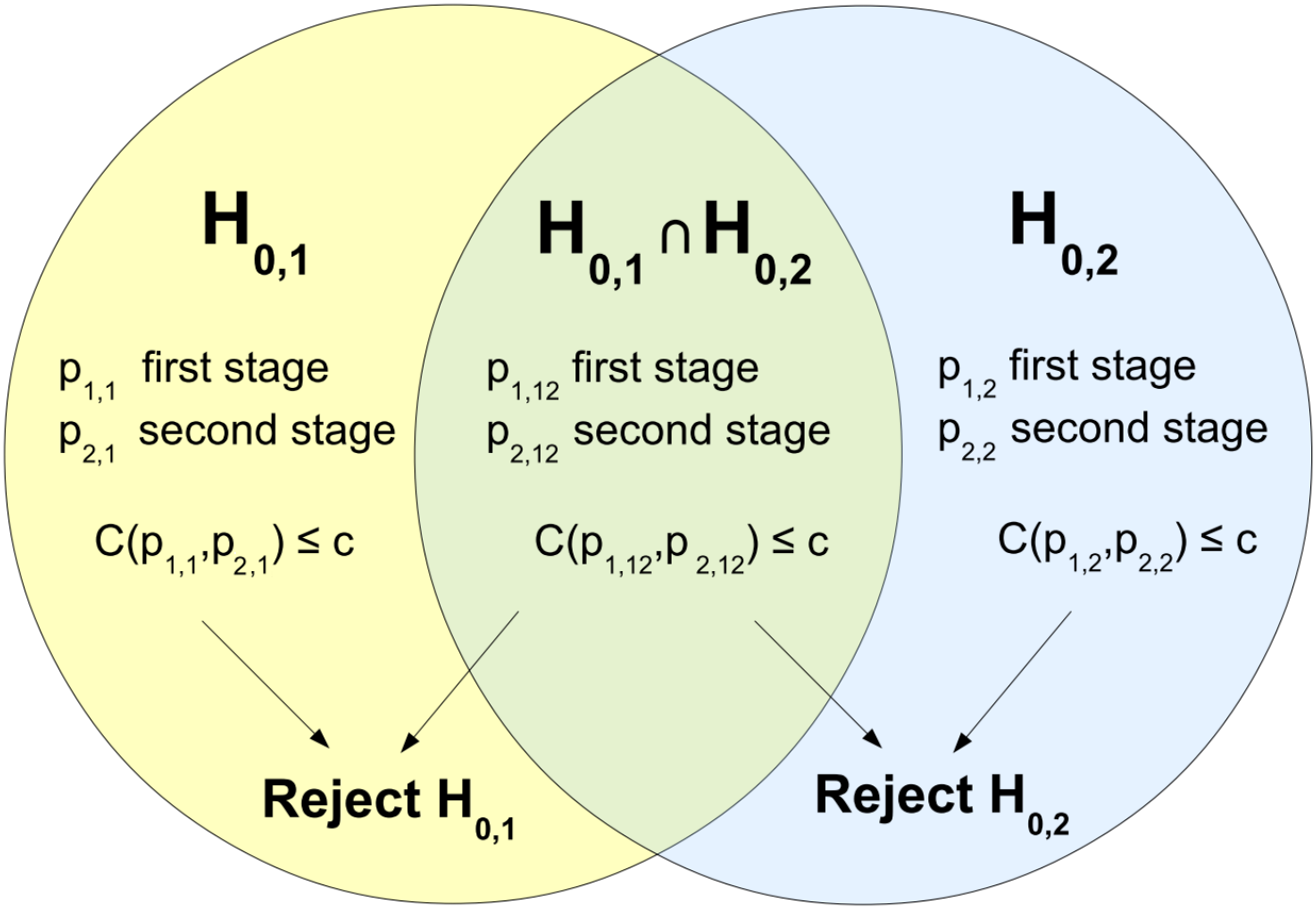


Combination tests to be performed for the closed system of 3 hypotheses.

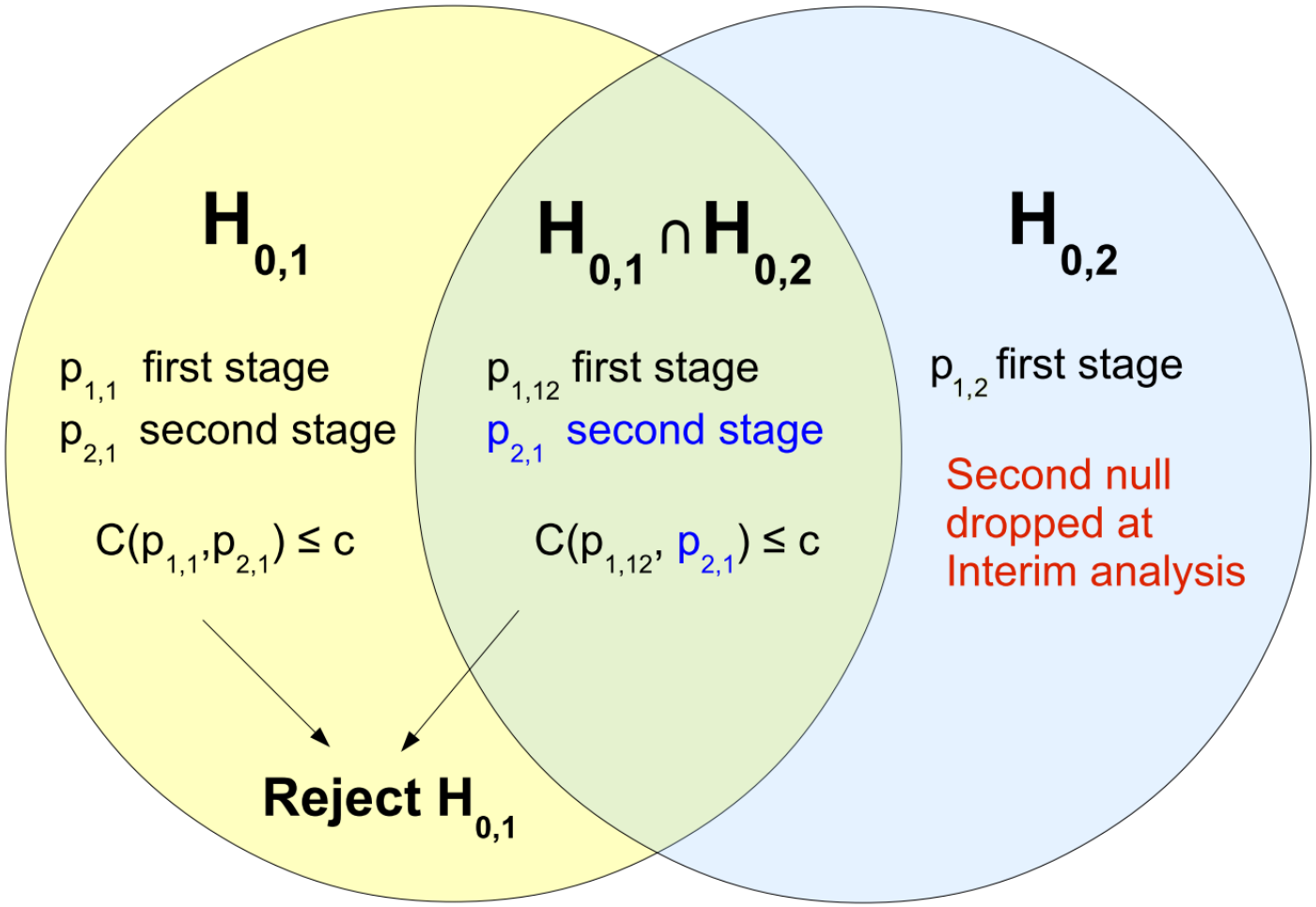
Closed Testing Principle, 2 Hypotheses



Adaptive Closed Test - Selecting Both Hypotheses



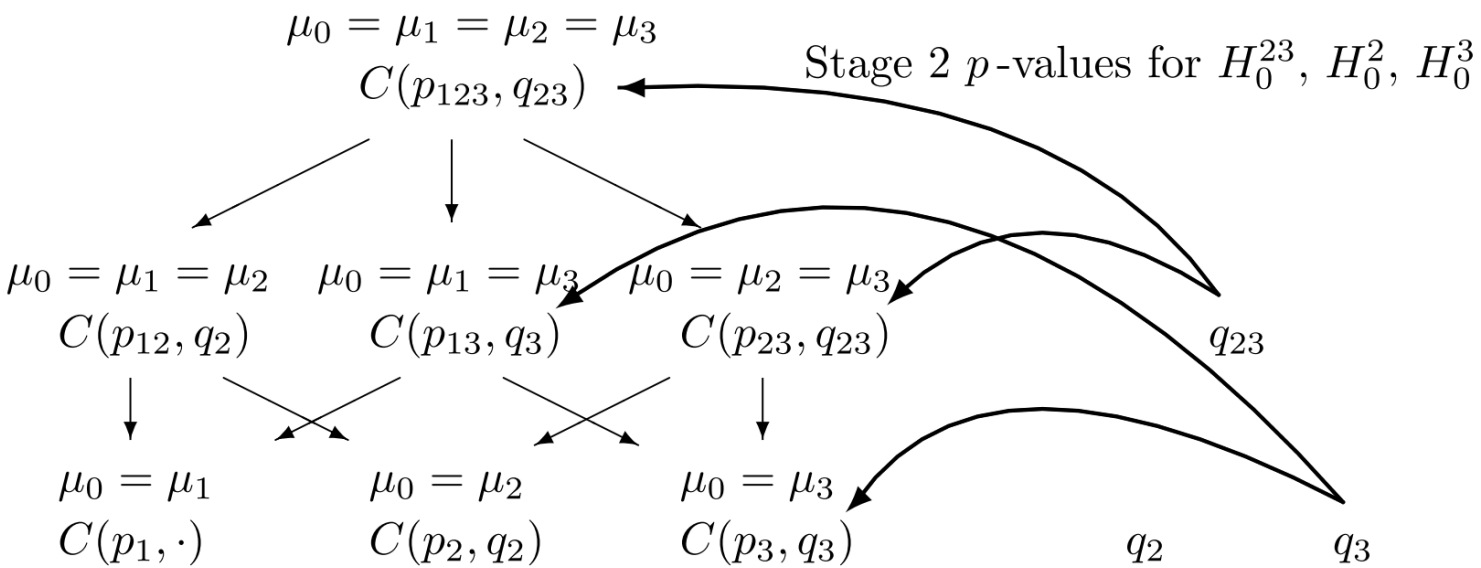
Adaptive Closed Test - Selecting Hypothesis 1 Only



Application of Adaptive Closed Tests

- Adaptive designs with treatment arm selection
  - Can easily extended to selection of more than one treatment arm. The number of selected arms needs not to be preplanned.
  - Choice of intersection tests is free. You can choose between Dunnett, Bonferroni, Simes, Šidák, etc.
  - For two-stage designs, the CRP principle can be applied: adaptive Dunnett test (König et al, 2008).
  - The procedure may become inconsonant and, hence, conservative. I.e., you can reject the global hypothesis, but no single hypothesis.
  - Confidence intervals based on stepwise testing are difficult to construct. This is a specific feature of multiple testing procedures and not of adaptive testing. Posch et al. (2005) proposed to construct repeated confidence intervals based on the single step adjusted overall p-values.

Three Treatment Arms - Selecting Hypotheses 2 and 3

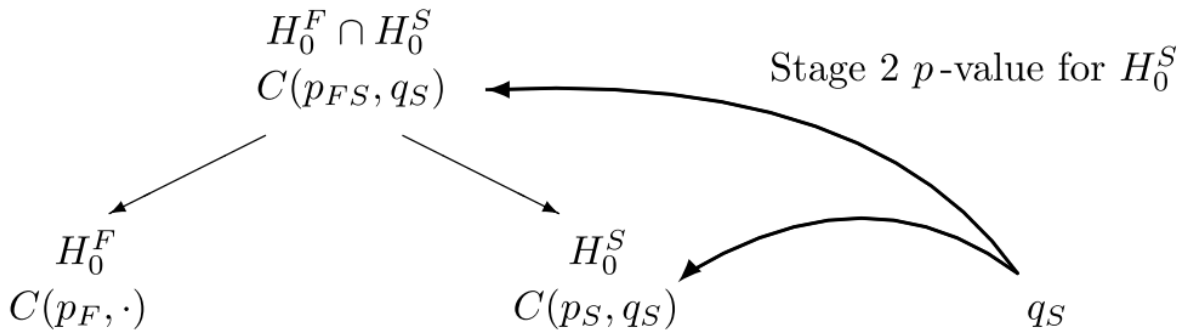


Combination tests to be performed for the closed system of hypotheses ($G = 3$) for testing hypothesis H_0^3 if treatment arms 2 and 3 are selected for the second stage

Application of Adaptive Closed Tests

- Adaptive enrichment designs: data driven selection of the target population
  - Choice of intersection tests in principle as above (except Dunnett).
  - The CRP principle can be applied for one subgroup: conditional bivariate test (Friede et al, 2012).
  - Confidence intervals and overall p-values can be defined analogously to the multi-armed case.
- Selection of endpoints (work to be done)

Population enrichment design



Combination tests to be performed for the closed system of hypotheses ($G = 2$) if subpopulation S referring to hypothesis H_0^S is selected for the second stage.

Population Enrichment Design

Closed Combination Test, G = 3

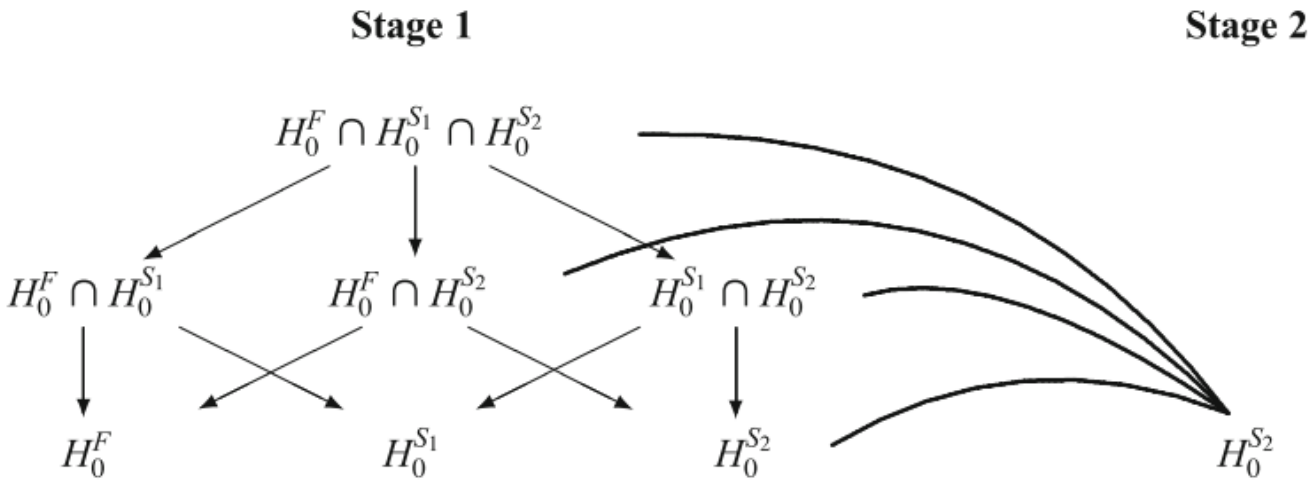


Fig. 11.3 Closed system of hypotheses for $G = 3$ if subpopulation S_2 referring to hypothesis $H_0^{S_2}$ is selected for the second stage. The *arrows* indicate logical implications for hypotheses, the *solid curves* indicate combination tests to be performed to show significance for $H_0^{S_2}$ (from Wassmer and Dragalin 2015)