Adaptive designs with subgroup selection in oncology

Werner Brannath, MUW, Vienna

Joint work with E. Zuber, M. Branson, F. Bretz, P. Gallo, M. Posch, A. Racine-Poon

Joint Meeting of MHRA and EFSPI, London, 2010

Study Setting

- Novel therapy for advanced metastatic disease
- Biochemical pathways suggest that a specific sub-population of patients is more likely to achieve response to treatment.
- Should the therapy be targeted?

Development of targeted therapy

Traditional Approach

1. Exploratory study to identify a sub-population S.
2. Phase II study to confirm sensitivity of sub-population S.
3. Phase III study for formal claim of efficacy in the previously identified and confirmed sub-population S.

Adaptive Trial

- Phase II and III are integrated into a single trial
- Confirmatory test is based on data from both stages
- Efficacy is tested with a survival endpoint (PFS).

The Adaptive Trial Design

First Stage

- Randomization in full population F

Sub-population Definition

- Based on a different (exploratory) study

Decisions in the Interim Analysis

- Stop the trial for futility
- Continue with the sub-population (S)
- Continue with the full population (F)

The Adaptive Trial Design

Final Analysis Based on Both Stages

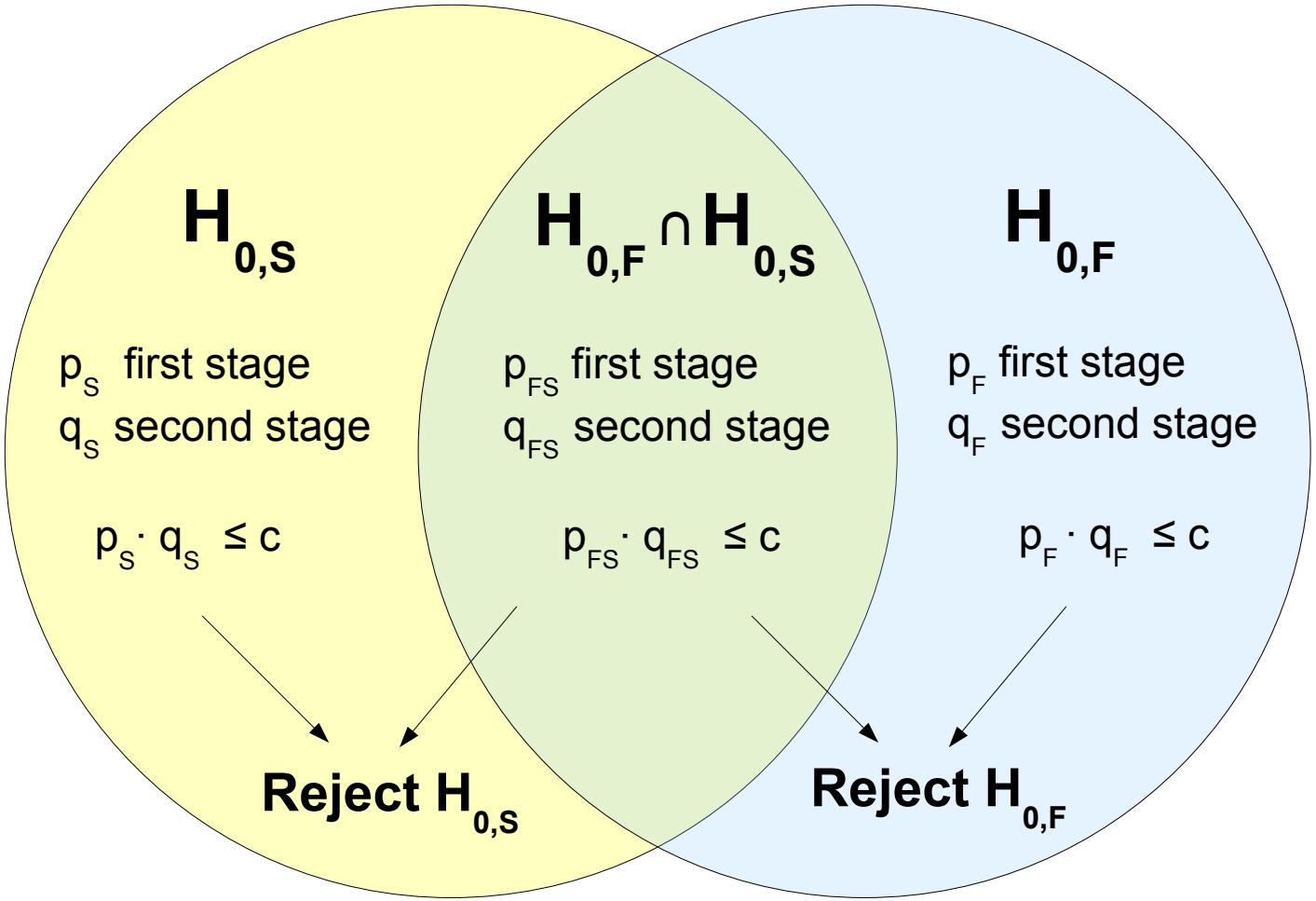
- If the trial continues with S, we test $H_{0,S}$ (no efficacy in S).
- If the trial continues with F, we test $H_{0,S}$ and $H_{0,F}$.

We combine two inferential methods

(BAUER & KIESER 1999, HOMMEL 2001, . . .):

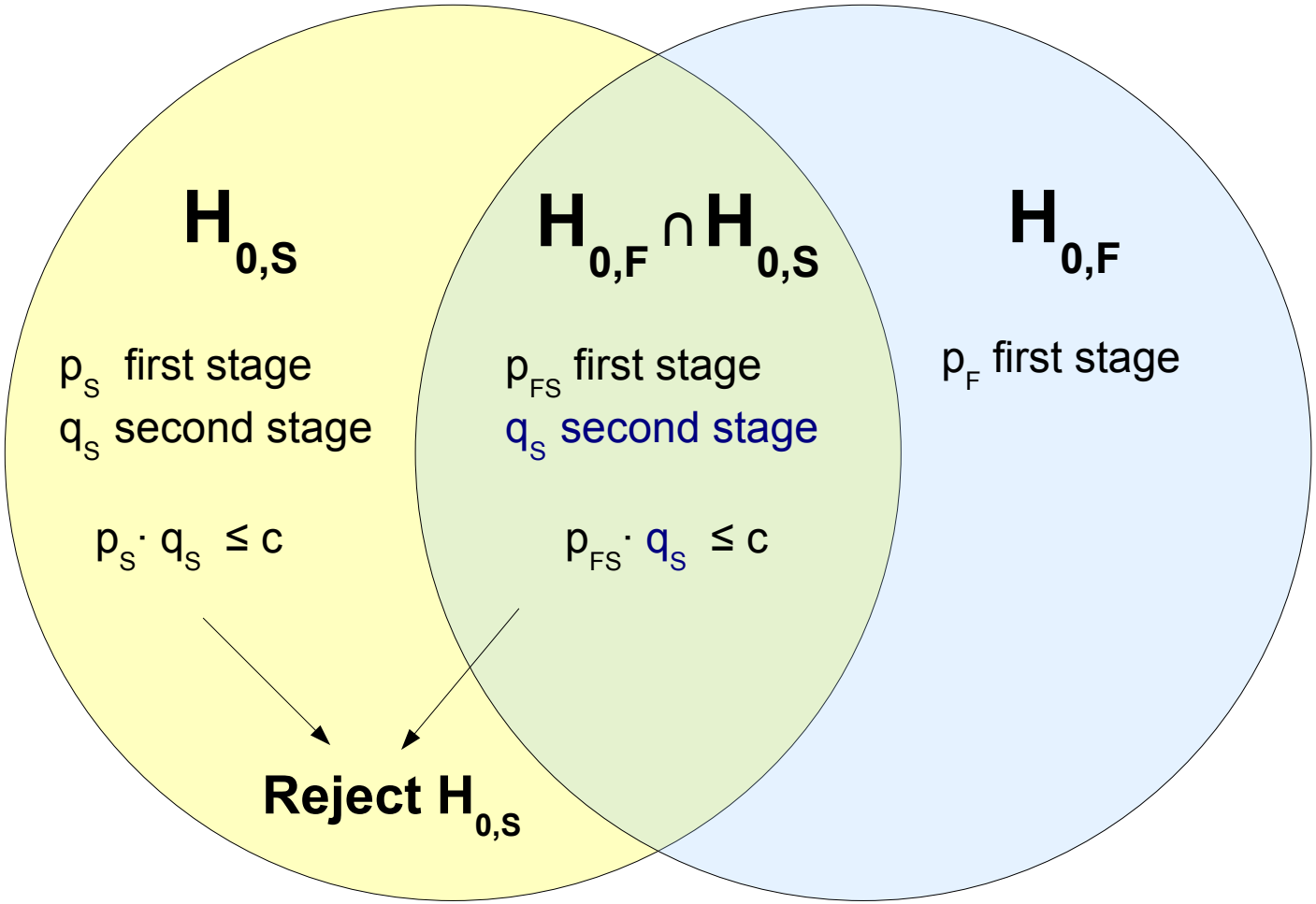
- Combination tests, to deal with the data driven choice of the event number per population.
- Closed testing principle, to deal with the multiplicity issue.
- We can use Hochberg's test for $H_{0,F} \cap H_{0,S}$

If the trial continues in *F*



$p_{FS} = \min\{1, p_F, p_S, 2\min(p_F, p_S)\}$, q_{FS} is defined similarly

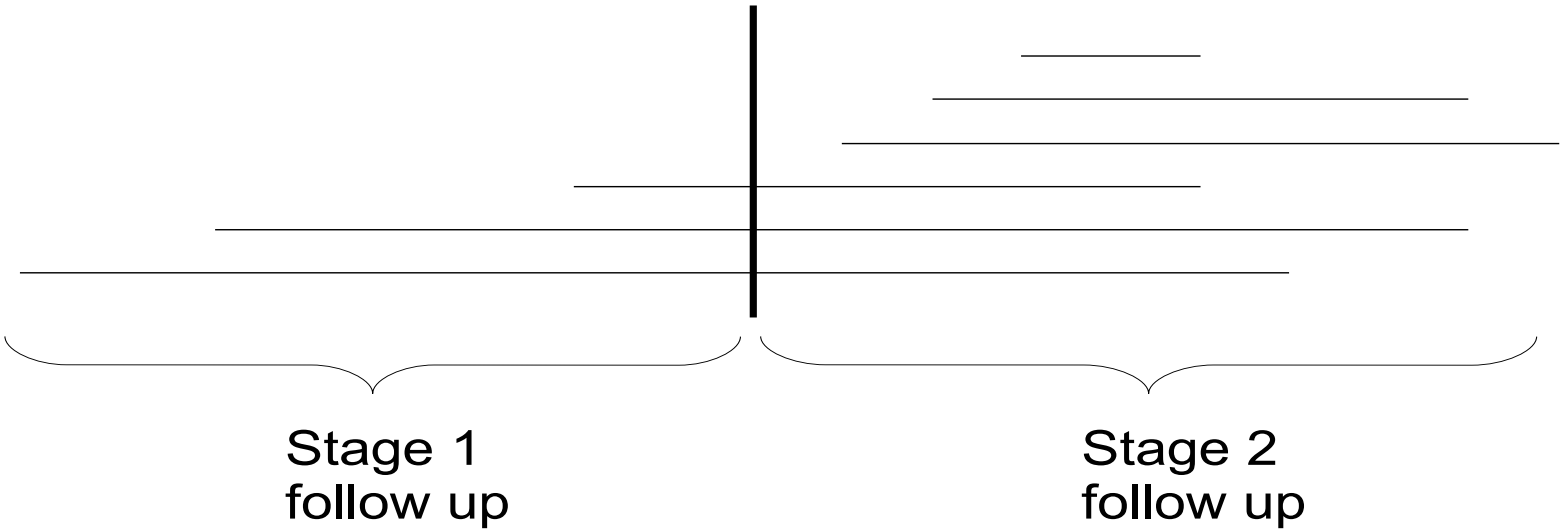
If the trial continues in S



$p_{FS} = \min\{1, p_F, p_S, 2\min(p_F, p_S)\}$, q_{FS} is defined similarly

Combination test with follow-up wise stages

(Müller & Schäfer '01, Bauer & Posch '01, Wassmer '06, ..)



Stage 1: p_F, p_S from — log-rank tests, follow-up **till** time of IA, i.e., **right** censoring at time of IA

Stage 2: q_F, q_S from — log-rank tests, follow-up **from** time of IA, i.e., **left** censoring at IA or increments.

Combination test with follow-up wise stages

- Same type of stage-wise data splitting as in group sequential designs (Jennison and Turnbull, 2001).
- Strict type I error rate control **only** if IA-decisions depend solely on PFS (Bauer & Posch '01),
- Reason: When using further interim information (e.g. sub-group membership) from censored patients then first and second stage log-rank statistics may fail to be assym. independent.

$\Longrightarrow$

No strict type I error rate control when sub-group selection is based sub-group membership information!

Combination test with stratified log-rank test

(Zuber et al. '06, Brannath et al. '08)

For H_F we stratify the log-rank test for S and $S^c = F \setminus S$.

$\Longrightarrow$

- Strict type I error rate control when selection is based also on sub-group membership.
- Reason: Test statistic is linear combination of log-rank test statistics within S and S^c.

Bayesian Decision Tools at Interim

Predictive Power (PP) and Posterior Probability (P)

- Predictive Power to show efficacy in F: PP^F
- Predictive Power to show efficacy in S: PP^S
- Posterior Probability that the hazard ratio in S^c (the complement of S) is below a fixed threshold: P^{S^c}

Predictive Power

The probability that the second stage p-value leads to a rejection, taking into account the closed test and posterior probabilities.

Bayesian Decision Rules at Interim

Prior and Posterior Distributions

- Normal priors for the log hazard ratios in S and S^c.
- Normal approximation of the logrank statistics. $\Rightarrow$
- In each subgroup: the posterior for the log hazard ratio is normal with known mean and variance.

Decision Rules

- PP^F and PP^S are small $\Rightarrow$ **stop for futility**
- PP^F is large $\Rightarrow$ **go with *F***
  (unless PP^S is large and P^{S^c} is small $\Rightarrow$ go with S).
- PP^F is small and PP^S is large $\Rightarrow$ **go with *S***.

Case Study

Study Characteristics

- Endpoint: Progression Free Survival
- Overall Number of Patients: 1200 (fixed)
- Target number of events: 918 in F or 640 in S
- First interim at 170 events in F
- Second interim (efficacy analysis, no adaptations)
  551 events in F or 384 in S, if S is selected.
- Inverse normal combination function with O'Brien-Fleming boundaries and weights according to event numbers in F.

Overall Power

Thresholds: 35% for PP^F and PP^S, 25% for P^{S^c}

1. AD: Adaptive Design
2. GFS: Group Sequential Test with Hochberg Test in F and S
3. GF: Group Sequential Test testing only in F

Scenario: Efficacy in S and S^c (for both HR $= 0.77$)

AD, GFS,GF $\approx 87 - 88\%$ Power.

Scenario: Efficacy only in S (HR $= 0.77$)

Probability to reject in S or F (in F)

Prevalence	AD	GFS	GF (only in F)
30%	64% (6%)	39% (14%)	16%
50%	75% (16%)	62% (38%)	41%
80%	78% (34%)	79% (17%)	73%

Operation Characteristics of Decision Rules

Scenario: No Efficacy, neither in S nor in S^c

Prevalence	Futility	Go with F	Go with S
30%	50%	19%	31%
50%	60%	16%	23%
80%	65%	20%	15%

Scenario: Efficacy only in S (HR $= 0.77$)

Prevalence	Futility	Go with F	Go with S
30%	20%	22%	58%
50%	15%	28%	57%
80%	15%	39%	46%

Operation Characteristics of Decision Rules

Scenario: Efficacy in S and S^c (for both HR $= 0.77$)

Prevalence	Futility	Go with F	Go with S
30%	8%	72%	20%
50%	8%	71%	21%
80%	10%	69%	21%

Remark

- Power and operation characteristics depend on hazard ratios, thresholds and prevalence.
- A decrease in the threshold for P^{S^c} would increase the chance to go with F under all scenarios.

Summary

- Subgroup (determined in an external study) can be confirmed in the first part of a phase II/III trial.
- Bayesian tools for decision making at interim
  - incorporates predictive power calculations in F and S,
  - incorporates posterior probabilities of efficacy in S^c,
  - a priori information (e.g. from other studies) may be incorporated via an informative prior.
- Strict control of multiple type I error rate, also when deviating from the pre-specified rule.

- No strict control of a false efficacy claim in S^c (via a claim in F), however, the design a provides way to balance false negative decisions in F and false positive decisions in S^c.

Selected References

- Schäfer H., Müller H.-H. Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections *Statistics in Medicine* 2001, **20**:3741–3751
- Bauer B., Posch M. Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections by H. Schäfer and H.-H. Müller, Statistics in Medicine 2001; 20: 3741-3751. *Statistics in Medicine*, **23**:1333–1334
- Wassmer G. Planning and Analyzing Adaptive Group Sequential Survival Trials *Biometrical Journal* 2006, **48**:714–729
- Brannath W., Zuber E., Branson M., Bretz F.,Gallo P., Posch M., Racine-Poon A. Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Statistics in Medicine* 2009, **28**:14451463.
- Zuber E., Brannath W., Branson M., Bretz F., Gallo P., Posch M., Racine-Poon A. Phase II/III seamless adaptive designs with bayesian decision tools for an efficient development of a targeted therapy in oncology. Technical Report 2006-05, Department of Statistics and Decision Support Systems, University of Vienna.