

Adaptive Designs for the Development of Targeted Therapies

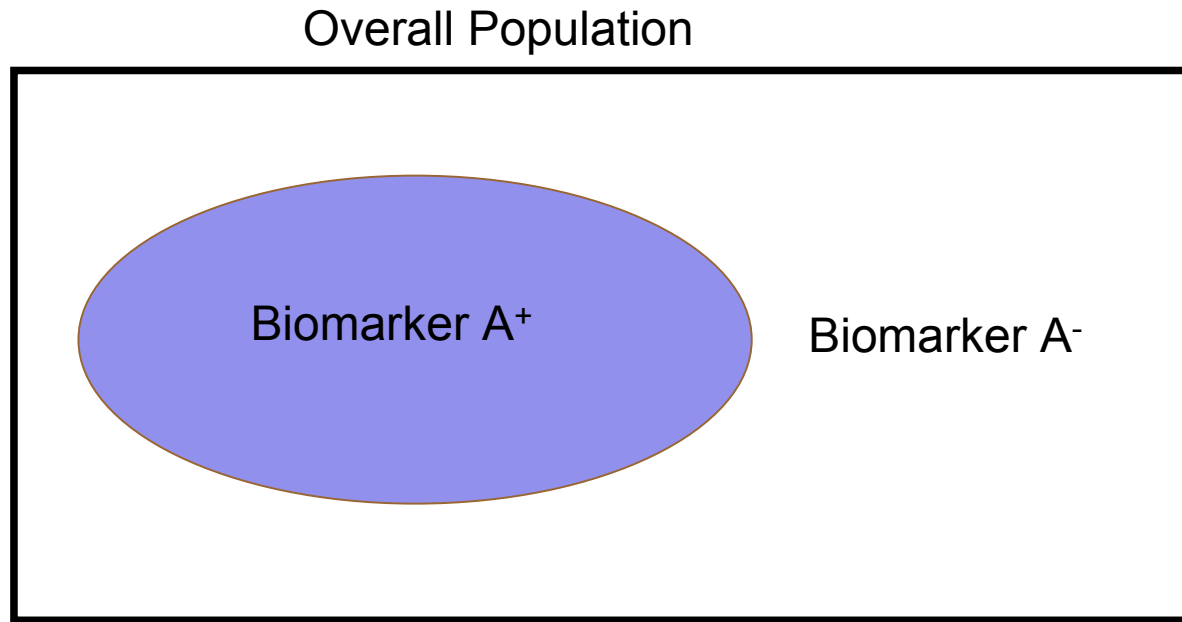
Martin Posch, Alexandra Graf, and Franz König

Institut für Medizinische Statistik, CeMSIIS
Medical University of Vienna
Vienna, Austria

Identifying Target Populations

- The knowledge on the genetic basis of many diseases is increasing rapidly and therapies are developed that target underlying molecular mechanisms.
- Patients' responses are predicted to targeted treatments based on genetic features or other biomarkers.
- **Objective: Identify subgroups based on biomarkers where the treatment has a positive benefit risk balance.**

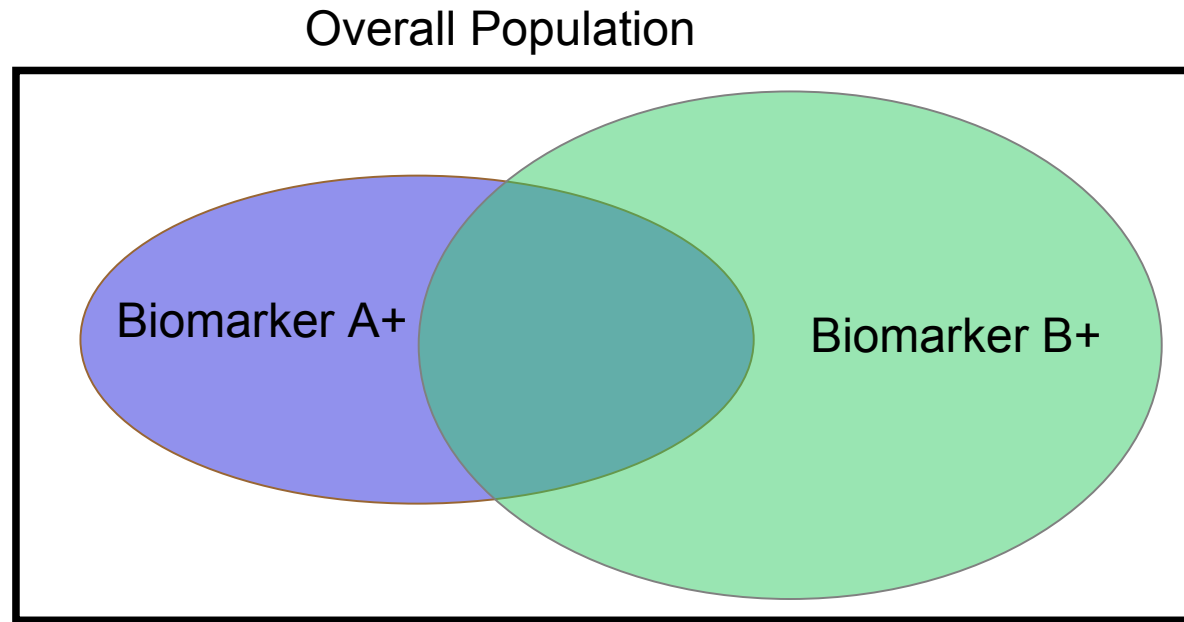
Subgroup Analysis



Demonstration of efficacy is investigated in

- the overall population ($= A^+ \cup A^-$)
- and the A^+ group

More than one Biomarker



Demonstration of efficacy is investigated in

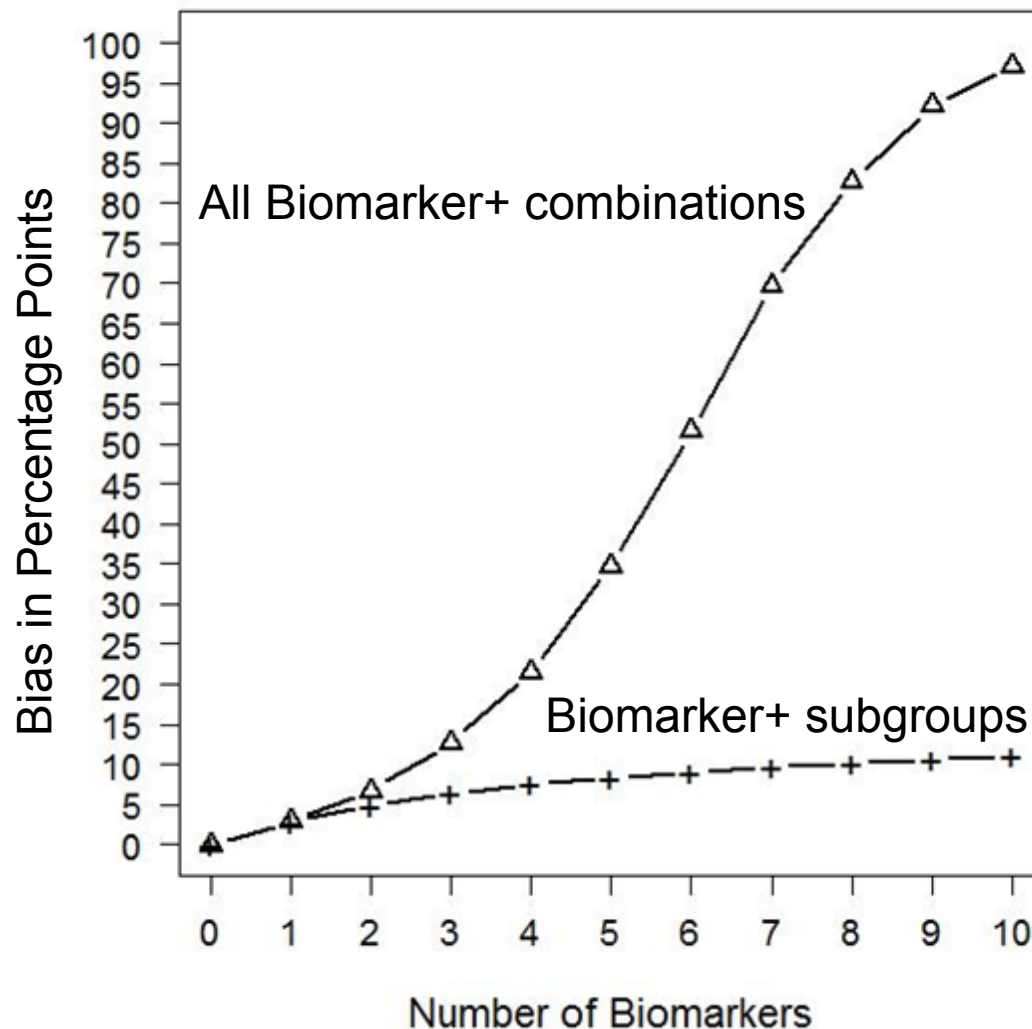
- the overall population
- each biomarker positive group
- (and in the subgroup where both biomarkers are positive)

Multiplicity Issues in Subgroup Analyses

- Several chances to claim significant treatment effect:
 - in the overall population
 - subpopulation(s)
- If subgroups are selected without appropriate adjustment
 - the treatment effect estimates will be biased
 - the false positive rate will be inflated.

Estimation Bias (1)

If the population (either Biomarker + or overall population) with the largest treatment effect is chosen, the effect estimate is biased



Trial Design

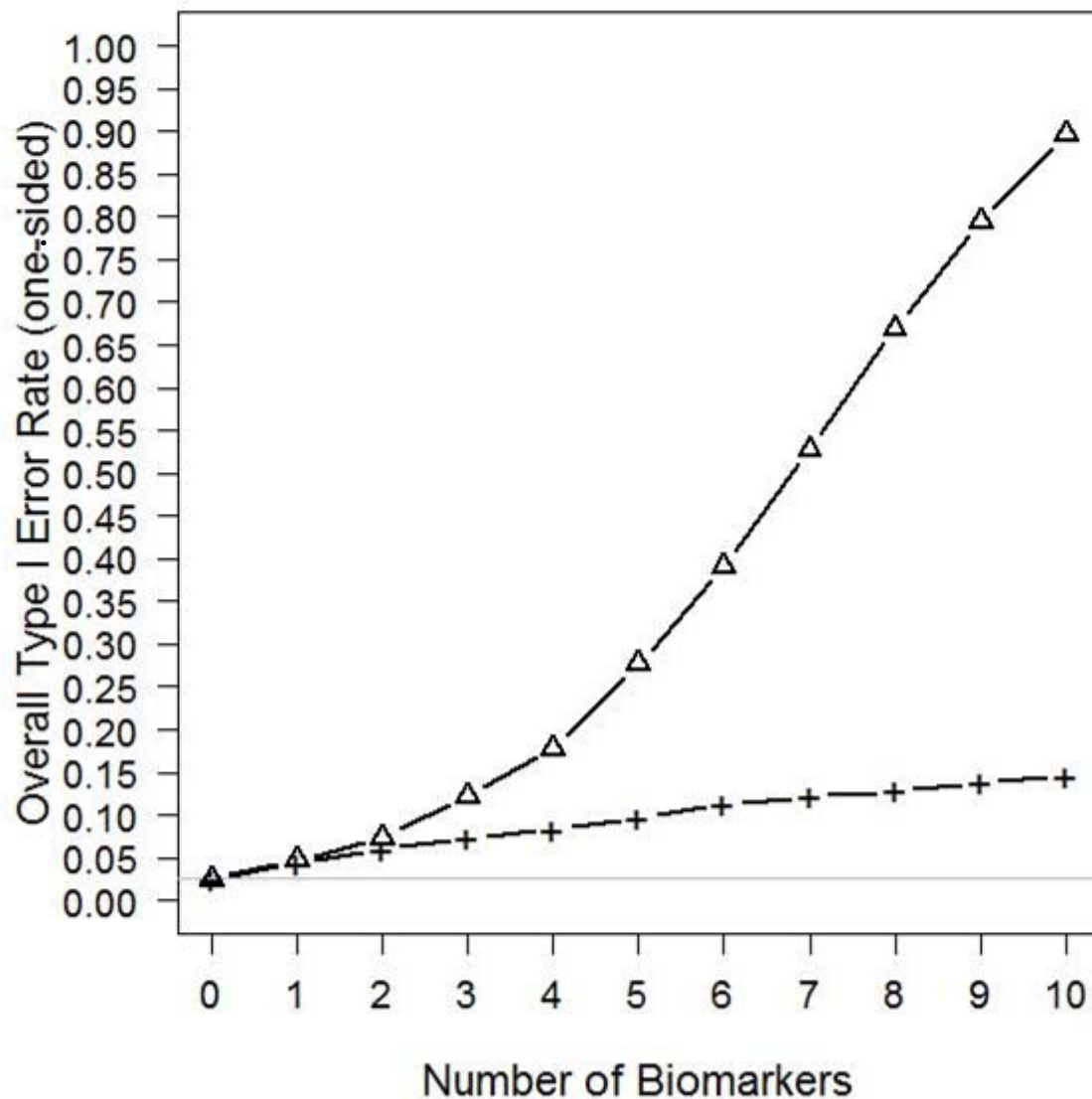
- Parallel Group Design (n=200)
- Test for difference in response rates

Scenario

- Response Rate 50%
- Prevalence of Biomarkers 50%
- Independent Markers
- No efficacy difference

False Positive Rate (2)

Overall false positive rate for unadjusted one-sided hypothesis tests at 2.5%.



All Biomarker+ combinations

Biomarker+ subgroups

Lessons Learned

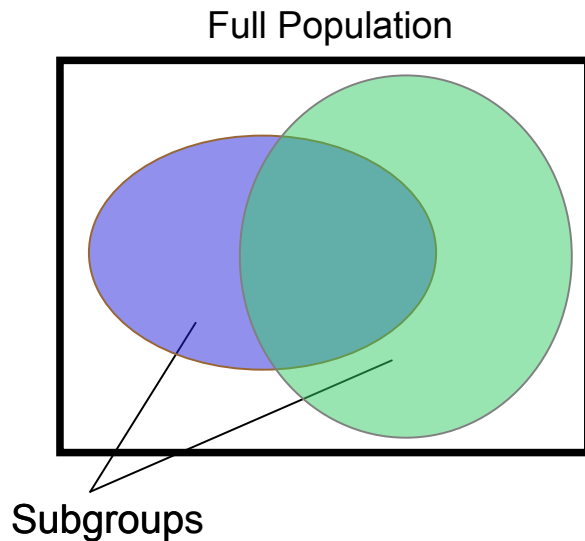
- A **formal adjustment for multiple comparisons** to limit the overall false positive rate (probability to conclude efficacy in a subgroup, when in fact there is none) at the usual significance level (e.g., 2.5% one-sided) **is required**

Points to consider on multiplicity issues in clinical trials, CPMP 2002

Large trials needed for sufficient sample sizes in subgroups and to achieve adequate power (or alternatively hoping for true extreme effect sizes in small subgroups).

Two Trials: Learn and Confirm

Learning (Phase II)



Planning Phase III
Subgroup Selection

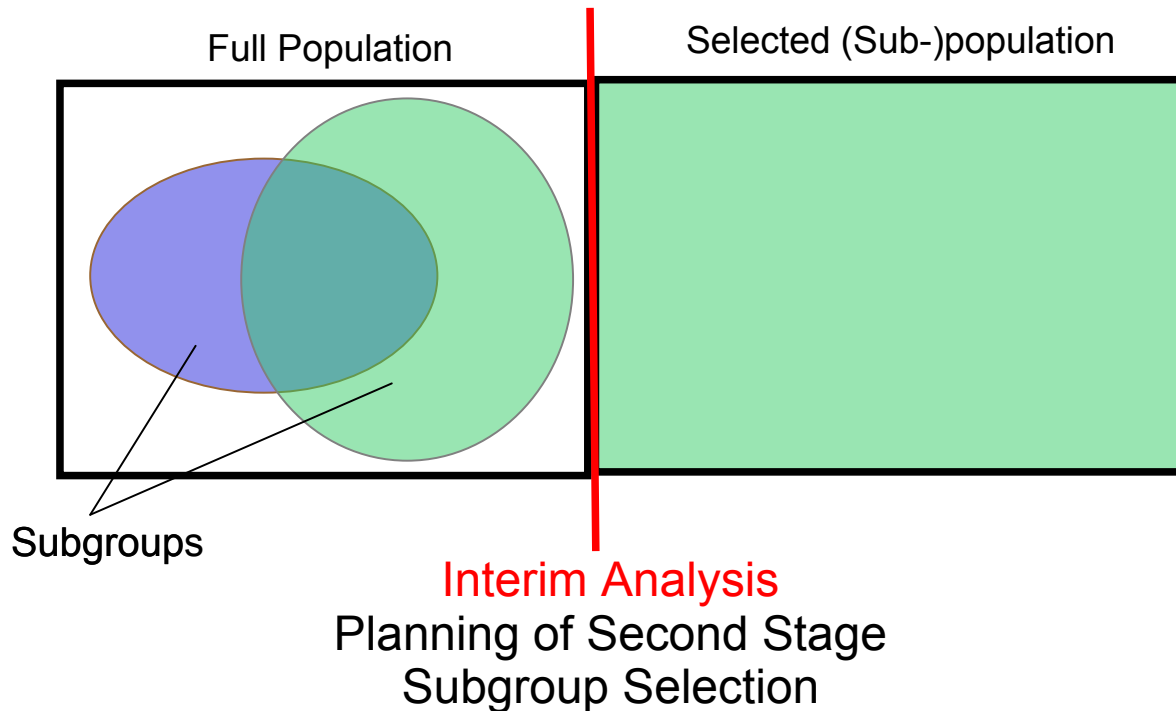
Confirming (Phase III)



- Phase II trial objective: subgroup identification
- Efficacy shown ONLY based on Phase III data (independent replication in the target (sub-)population).

Adaptive Phase II/III Design

Learning & Confirming



- The Phase II data is used for subgroup selection
- Efficacy is demonstrated with phase II + III data (adjusting for multiplicity).

Case Study: Development of Targeted Therapy

(Brannath et al. '09)

Study Setting

- Advanced metastatic disease
- Endpoint: progression free survival
- Biochemical pathways suggest that a specific sub-population of patients are more likely to achieve response to treatment.

The Adaptive Trial Design

First Stage

- Randomization in full population

Decisions in the Interim Analysis based on Bayesian Rules

- Stop the trial for futility
- Continue with the sub-population (A^+)
- Continue with the full population ($F=A^+ \cup A^-$)

Final Frequentist Analysis Based on Both Stages

- Test for efficacy in A^+ .
- If the trial continued with F test also for efficacy in F
- Control of the overall false positive rate with adaptive multiple testing procedures.

Clinical Trial Simulations

- In the planning phase compare the operating characteristics to other (more traditional) strategies
 - Probabilities of Success (evaluate different power definitions)
 - Impact on effect estimates (bias), selection probabilities
 - Average Sample Sizes
 - ...
- What is the impact of the timing of the interim analysis, different effect sizes for F and A^+ , prevalence of A^+ , ...
- Explore different selection rules, e.g., based on
 - absolute observed interim effects in F , A^+ and A^-
 - conditional power arguments
 - Bayesian decision rules, e.g. based on predictive power

Possible Conclusions from the Trial

- (1) Positive effect in the **sub-population**
- (2) Positive effect in the **full population**

The test

- **Controls the false positive rate for conclusions (1) and (2).**
- **Cannot detect if the effect in F is driven by A^+ only.**

A rigid but conservative approach would be to demonstrate efficacy in each subpopulation independently.

Overall Power to show efficacy in F or A⁺

Comparison of 2 Designs

1. Adaptive Design
2. Group Sequential Test

Scenario: Efficacy in A⁺ and A⁻

Power of both designs is 87 – 88% Power.

Scenario: Efficacy only in A⁺:

Prevalence	Adaptive Design	Groupsequential Design
30%	57% (9%)	39% (14%)
50%	71% (24%)	62% (38%)
80%	78% (50%)	79% (70%)

(Probability to conclude efficacy in F)

Summary

- Predefined testing strategy addressing the multiplicity issue is essential to make confirmatory inference on biomarkers.
- Robust operation characteristics of well planned adaptive designs
 - Strict type I error control, regardless of population selection process.
 - Same power as designs without enrichment if there is an effect in the full population.
 - Higher power if the effect is only in the subpopulation and less patients treated for which the drug does not work.
- Increased complexity of adaptive designs