# Confirmatory subgroup analyses: Case Studies

Frank Bretz, Gerd Rosenkranz, Emmanuel Zuber

EMA expert workshop on “Subgroup analysis”
London, November 18, 2011

# Subgroup analyses

- **Exploratory** subgroup analyses are often used to:
  - assess internal consistency of study results
  - rescue a failed trial by assessing the expected risk-benefit compared to the whole trial population in a post-hoc manner
- **Confirmatory** subgroup analyses
  - pre-specify one (or more subgroups) in the trial protocol (based on demographic, genomic or disease characteristics)
  - control Type I error rate for the pre-specified multiple hypothesis test problem and fulfill other standard requirements for confirmatory trials

# Case Study 1

## Treatment of Hep B in **HBeAg+/– patients**

**Design options under discussion, each with advantages / limitations**

1. **Two separate studies**
   - \+ flexibility in conducting each study on its own; if staggered study begin, second study design may benefit from first study results;
   - – costs
2. **One singly study with two strata (or cohorts)**
   - \+ one protocol; better estimation of relative efficacy/safety profile between subgroups; allows estimation of overall treatment effect (of interest here?)
   - – need for harmonized endpoint(s), no learning phase, independent timelines
3. **Two studies under one umbrella protocol**
   - \+ one protocol; retain flexibility through separate randomization schemes
   - – less rigorous in some aspects (pooled analysis, relative efficacy/safety, ...)

# Case Study 2

## New treatment as add-on to **background therapy**

**Primary objective:**

To demonstrate efficacy of at least one of two regimen as add-on therapy despite stable **treatment with X**

**Secondary objective:**

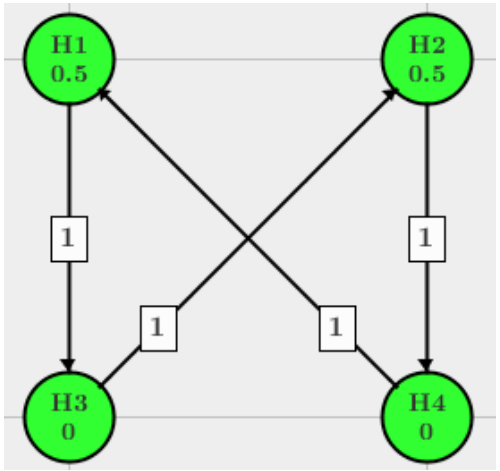
To demonstrate efficacy of at least one of two regimen as add-on despite stable **treatment with X or other drugs of the same class**

**Design:**

Randomization to be **stratified** by **X** or **not X**, enrollment such that 100p% of patients are on X.

**Regimen 1** **Regimen 2**

**X**




**All**

NOVARTIS

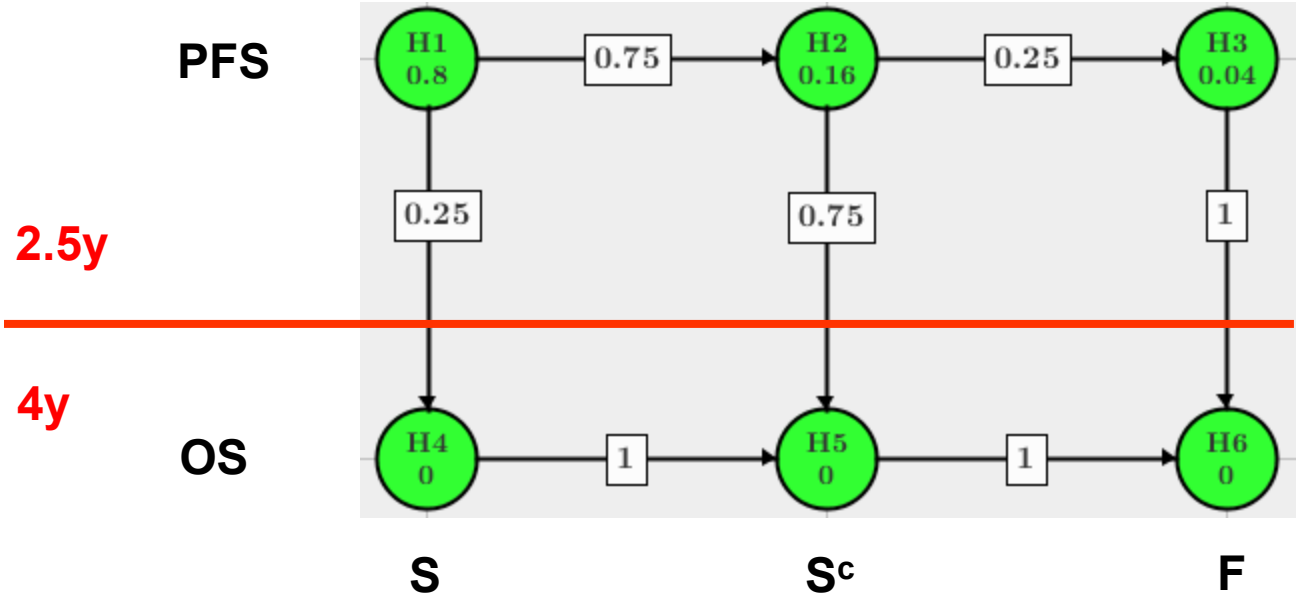
# Case Study 3

## New treatment in **naive/pre-treated patients** for PFS and OS

**Structured hypotheses with two levels of multiplicity**

1. Two-armed trial comparing with six hypotheses: novum vs. verum for
   - three populations (S = naive, $S^c$ = pre-treated, F = full population)
   - two hierarchical endpoints: PFS (after 2.5 years) ➔ OS (after 4 years)
2. Important clinical considerations
   - conditional approval envisaged if PFS significant (study then continued until OS analysis)
   - avoid significance in S and F, but no significance in $S^c$ (otherwise difficulties with label)

**How to construct decision strategy that reflects such requirements?**




NOVARTIS

# Case Study 4

## Confirmatory studies for **China**

**Population:**

~80% patients from mainland China (S) and ~20% not ethnic Chinese ($S^c$)

**Randomization:**

Stratification by mainland Chinese and other

**Requirements:**

- Stand alone report on **mainland Chinese** population with **significant** result
- Report on **full population** as **supportive** analysis
- **Multiplicity** adjustment **not necessary**

**Remark:**

- Multiplicity adjustment **useful** if full study contributes to **submission outside China**
- Alternative **option:** Primary objective on Chinese population, secondary on full population (hierarchical testing)

# **Case study 5** (Brannath et al., 2009)

## Confirmatory **adaptive design** for a targeted therapy in oncology

**Targeted therapy** might primarily benefit a subpopulation

**Evidence of activity**

- Preclinically & Clinically
- But requires better definition of biological characteristics of benefiting patients

**Traditional approach** to identify & confirm a sensitive subpopulation:

- Exploratory trial(s) to identify subpopulation with greater benefit
- Phase II to confirm greater benefit in identified subpopulation
- Phase III trial in the chosen target population (full or subpopulation)

**Ethical and strategic** relevance of allowing

- Focus as early as possible on subpopulation, if it can be defined
- Efficient use of data from patients needed to confirm the subpopulation

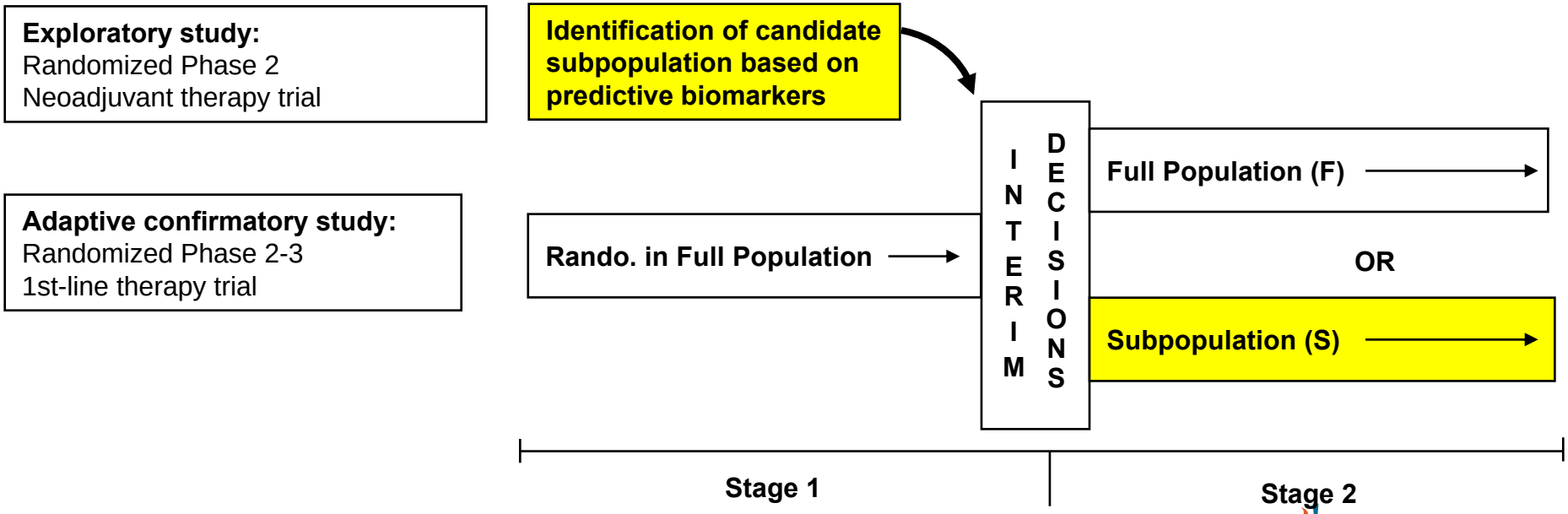
➔ **Integrate Phase II & III objectives in a single adaptive trial**

# Clinical development outline

**Exploratory trial:** large randomized phase II, baseline markers, response rate

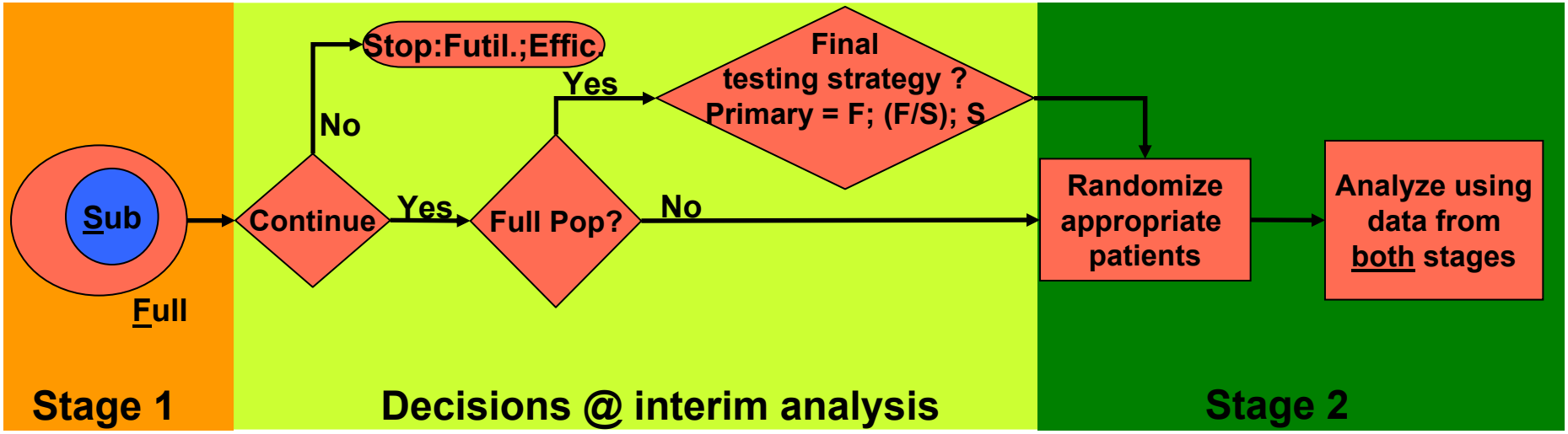
**Adaptive trial:** two stages, with an interim analysis, to simultaneously meet

- Phase II objectives
  - to confirm greater benefit in independently identified subpopulation
  - to decide whether or not to adapt trial to focus on that subpopulation
- Phase III objective
  - to demonstrate superiority on time to event (phase III) endpoint




NOVARTIS

# Confirmatory phase III adaptive design




- **Stage 1: Futility stop or subpopulation selection (Bayesian tools)**
  - Subpopulation defined prior to interim analysis (external to trial)
  - Probabilities of false positive and false negative decisions described a-priori via simulations
- **Stage 2: Confirmation of treatment benefit while maintaining integrity**
  - Combining evidence from first and second stage
  - False positive rate controlled by method, simulation used to explore power

# Methodology for Type I error rate control

- **Multiplicity issues**
  - Testing in 2 populations, group sequential testing (2 stages)
  - Stage 2 adapted based on stage 1 data
- **Adaptive design methodology**
  - Independent p-values from 2 stages combined: inverse normal method
  - Time to event: Independent p-values based on logrank asymptotic independent increments property
- **O'Brien-Fleming α-spending function**
- **Closed testing procedure**

# Adaptation decisions: Bayesian tools and rules

- **Bayesian tools:**
  - Predictive power:
    - Probability of success in each of the possible stage 2 situations (F or S)
  - Posterior probability:
    - Probability that the patients in $S^c$ (outside the subpop.) do not benefit
- **Decision rules:**
  - Predictive power in F and in S **<** threshold(s) $\pi\{F, S\}$
    ⇨ stop for futility
  - Only the predictive power in S **>** threshold $\pi\{S\}$
    or
    Probability (treatment effect in $S^c$ < target) **>** threshold $\pi\{S^c\}$
    ⇨ go with subpopulation
  - Otherwise
    ⇨ go with full population

# Power simulations (selected results)

Assume **no subpopulation effect** (all patients benefit from treatment):

- **Conventional** phase III (no interim analysis): 98% power
- **Conventional** phase III **with interim** (effic./futility): 88% power
- **Adaptive design** phase III: 87% power
  (across a variety of values of subpopulation prevalence)

If **only S benefits**:

| | Overall power | | |
|---|---|---|---|
| **S prevalence** | **Adaptive ph. III** | **Conventional sequential ph. III** | **Conventional seq. ph. III, test in F+S** |
| 30% | 57% | 16% | 39% |
| 40% | 65% | 28% | 52% |
| 50% | 71% | 41% | 62% |

[ with $\pi\{F, S\} = 35\%$, $\pi\{S^c\} = 90\%$ ]

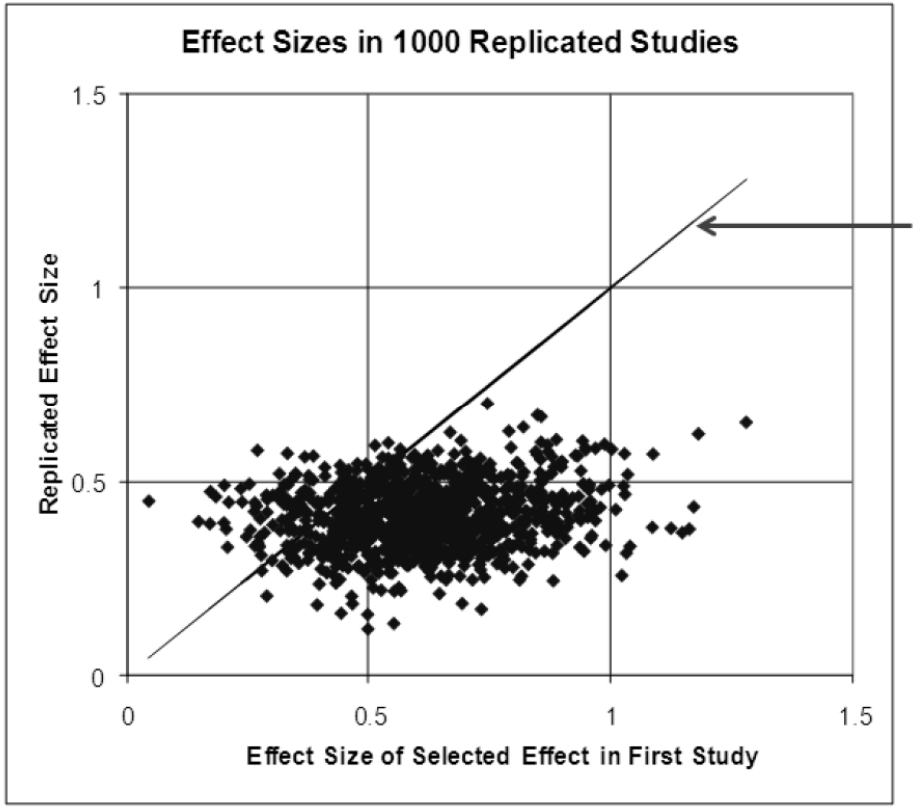
NOVARTIS

# Scientific concern: Reproducibility (selection bias)

**Assume 2 independent studies:**

- Study I – novum vs. verum for 2 subgroups
- Study II – select "best" subgroup from Study I and compare novum vs. verum for that subgroup

**Simulation results** (1000 trials, assuming equal effect in both subgroups):




(adapted from a presentation with Peter Westfall)

# Conclusions

- **Applications** involving confirmatory subgroup analyses **very diverse**
- Selection of **population of interest** (S / $S^c$ / F) **not always clear** and depends on context
- **Adaptive designs** logistically more complex (trial integrity!), but have the potential for more efficient drug development
- **Enriching the subpopulation** may lead to **interpretation problems**
- **Lack of reproducibility** is a major concern, **even more in retrospective analyses than in studies with prospectively defnied subgroups**