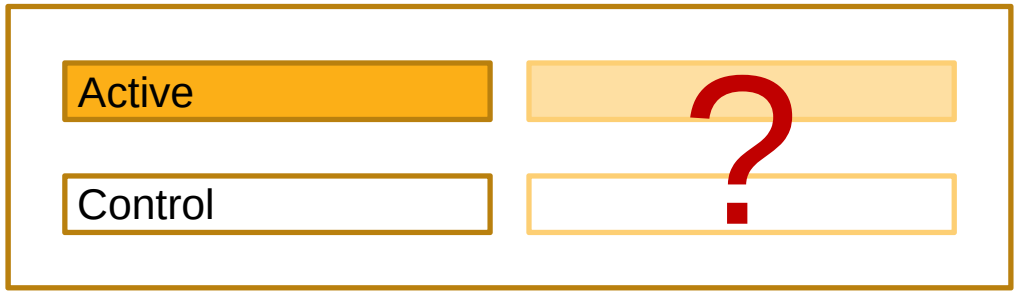


Group sequential and adaptive trials

Marc Vandemeulebroecke

Shanghai Biostatistics Forum

Shanghai, 20 December 2018

Acknowledgements: B Jones, K Rufibach, D Sabanés Bové, R Schwabe, A Stone, M Wolbers

Learning objectives

- Participants should understand:
  - What are group sequential and adaptive trials
    - What are the major subtypes of these designs
  - Why would a researcher choose such a trial design
  - What are the challenges of these trial designs
    - And how to address them
  - Practical considerations

Outline

- Introduction
  - Brief historical overview
  - Example
  - Interim analyses
- Group sequential trials
  - Classical group sequential designs
  - Spending functions
  - Estimation following a group sequential trial
- Adaptive trials
  - Motivation and overview
  - Methods for the confirmatory situation
  - Estimation following an adaptive trial
- Case study and Final words

Outline

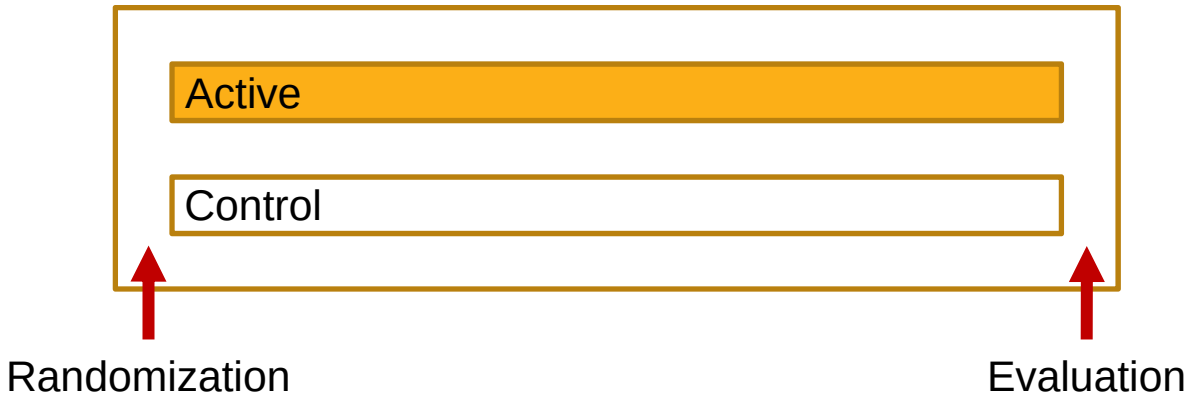
- Introduction
  - Brief historical overview
  - Example
  - Interim analyses
- Group sequential trials
  - Classical group sequential designs
  - Spending functions
  - Estimation following a group sequential trial
- Adaptive trials
  - Motivation and overview
  - Methods for the confirmatory situation
  - Estimation following an adaptive trial
- Case study and Final words

NOVARTIS

Brief historical overview

(Focusing on methods for the confirmatory situation)

- The starting point
  - Randomized, controlled, double-blind clinical trial

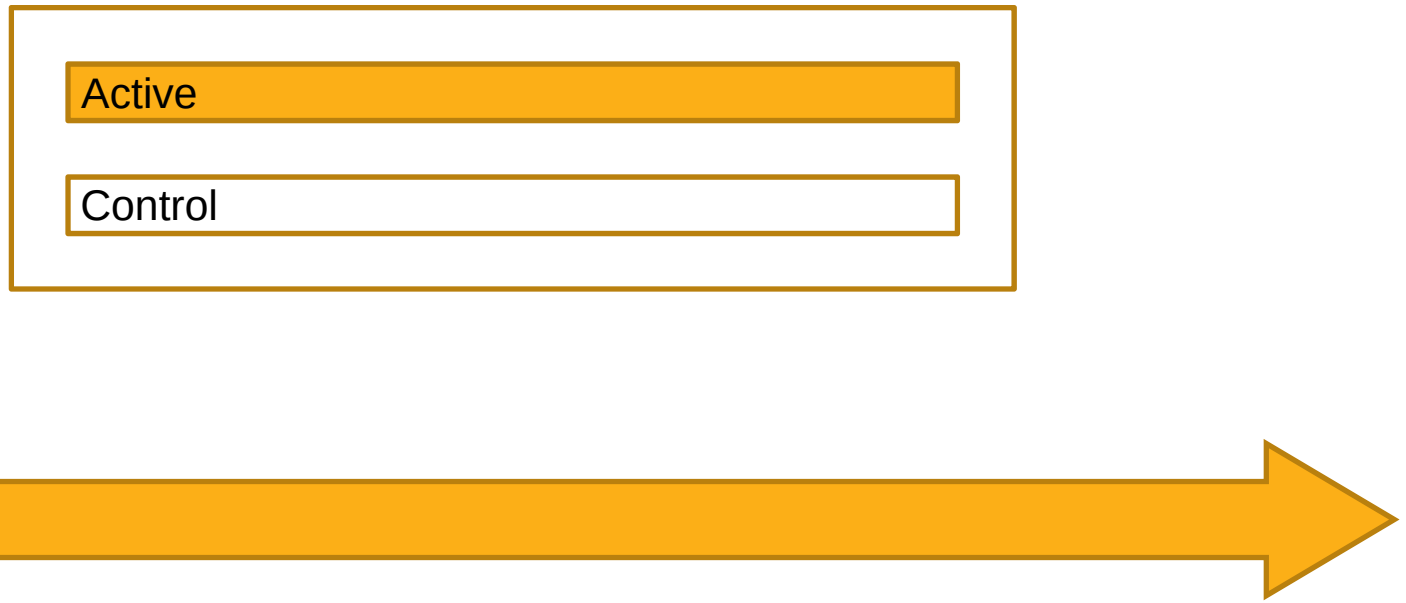


NOVARTIS

Brief historical overview

(Focusing on methods for the confirmatory situation)

- The starting point
  - Randomized, controlled, double-blind clinical trial

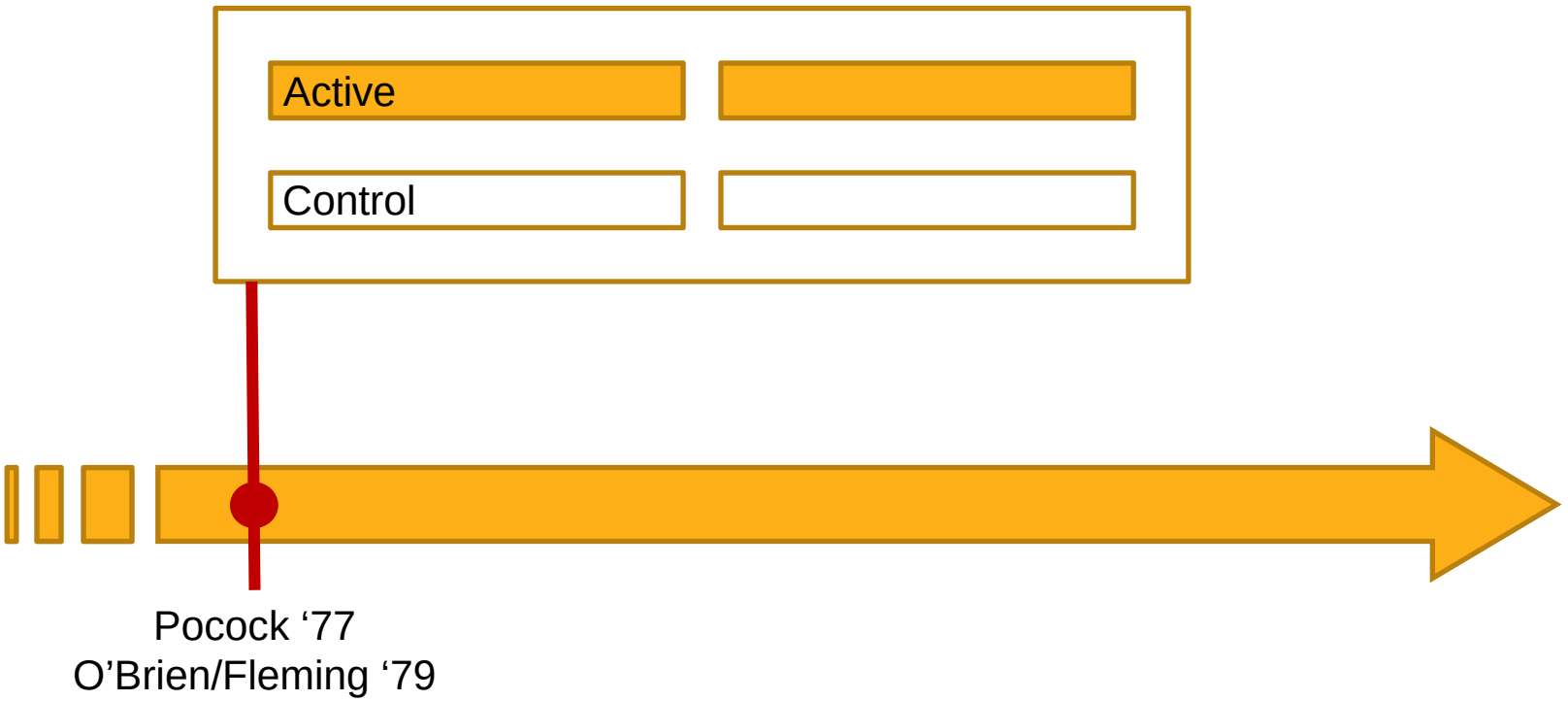


NOVARTIS

Brief historical overview

(Focusing on methods for the confirmatory situation)

- Group sequential design
  - Interim analysis and early stopping

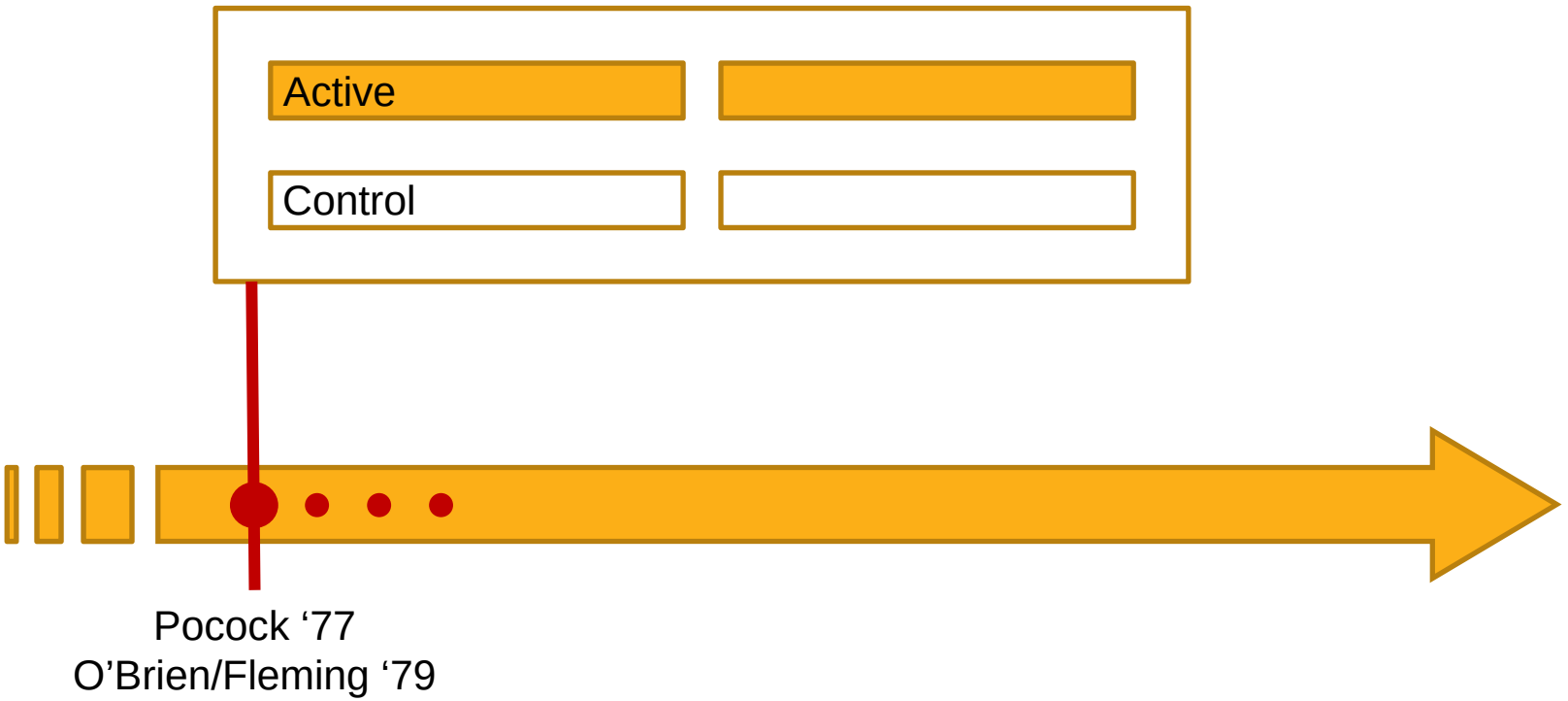


NOVARTIS

Brief historical overview

(Focusing on methods for the confirmatory situation)

- Group sequential design
  - Interim analysis and early stopping

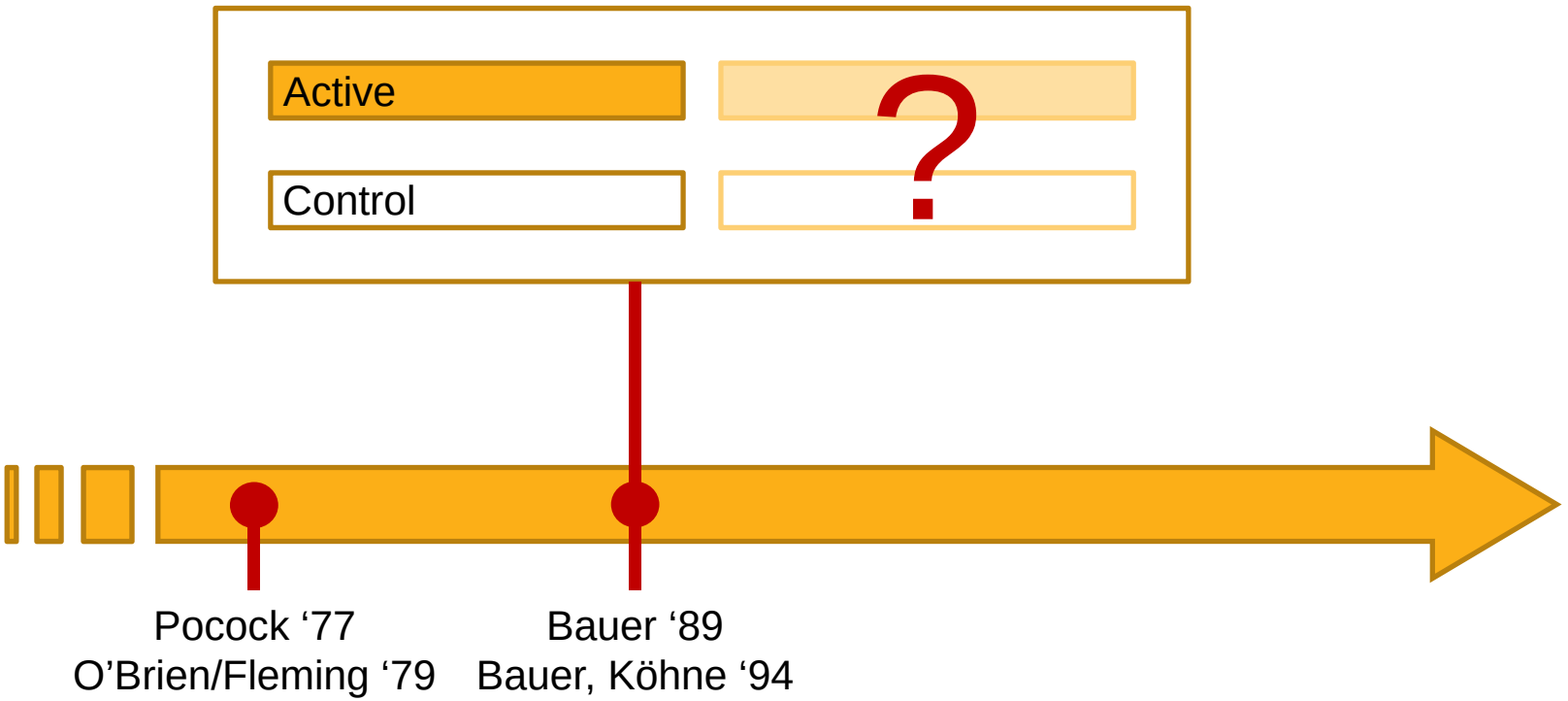


NOVARTIS

Brief historical overview

(Focusing on methods for the confirmatory situation)

- Adaptive design
  - Change trial feature(s) after the interim analysis

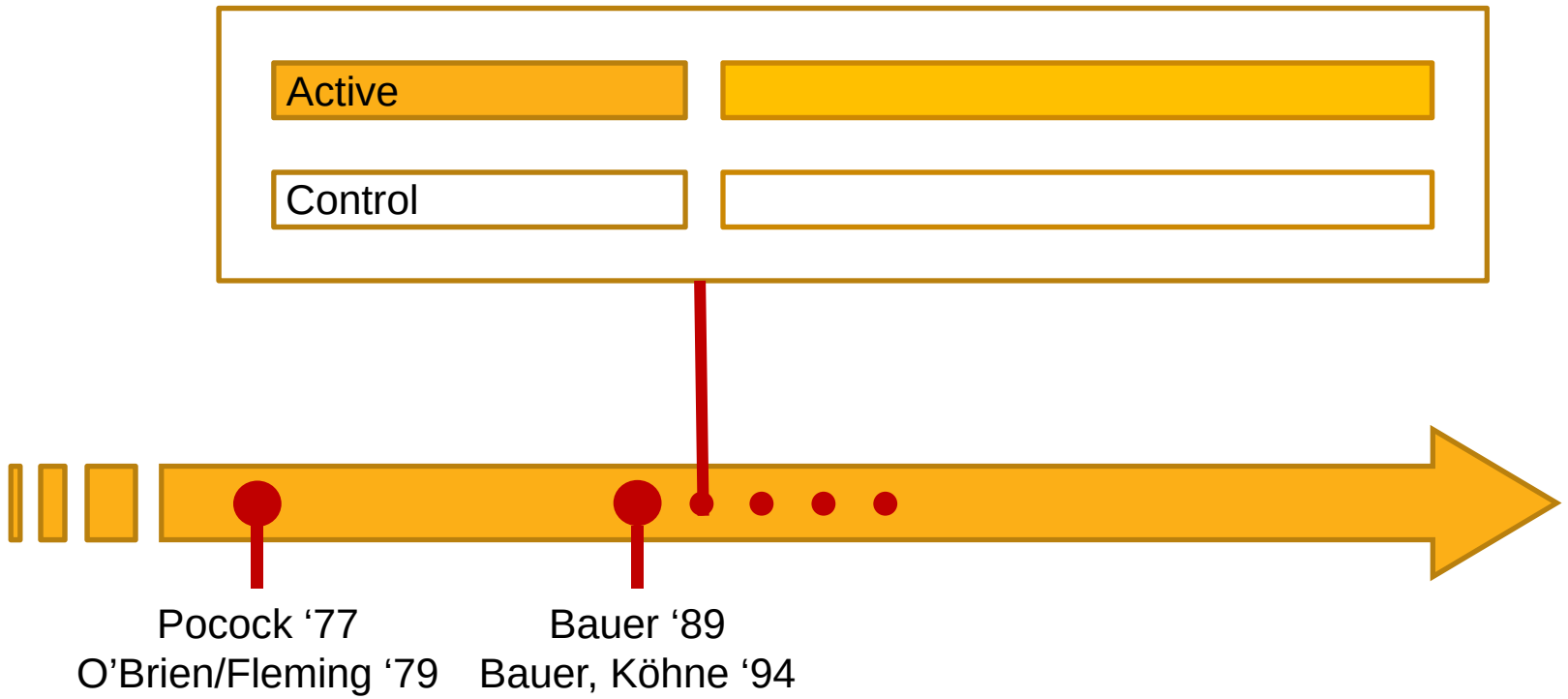


NOVARTIS

Brief historical overview

(Focusing on methods for the confirmatory situation)

- Adaptive design
  1. Change the sample size

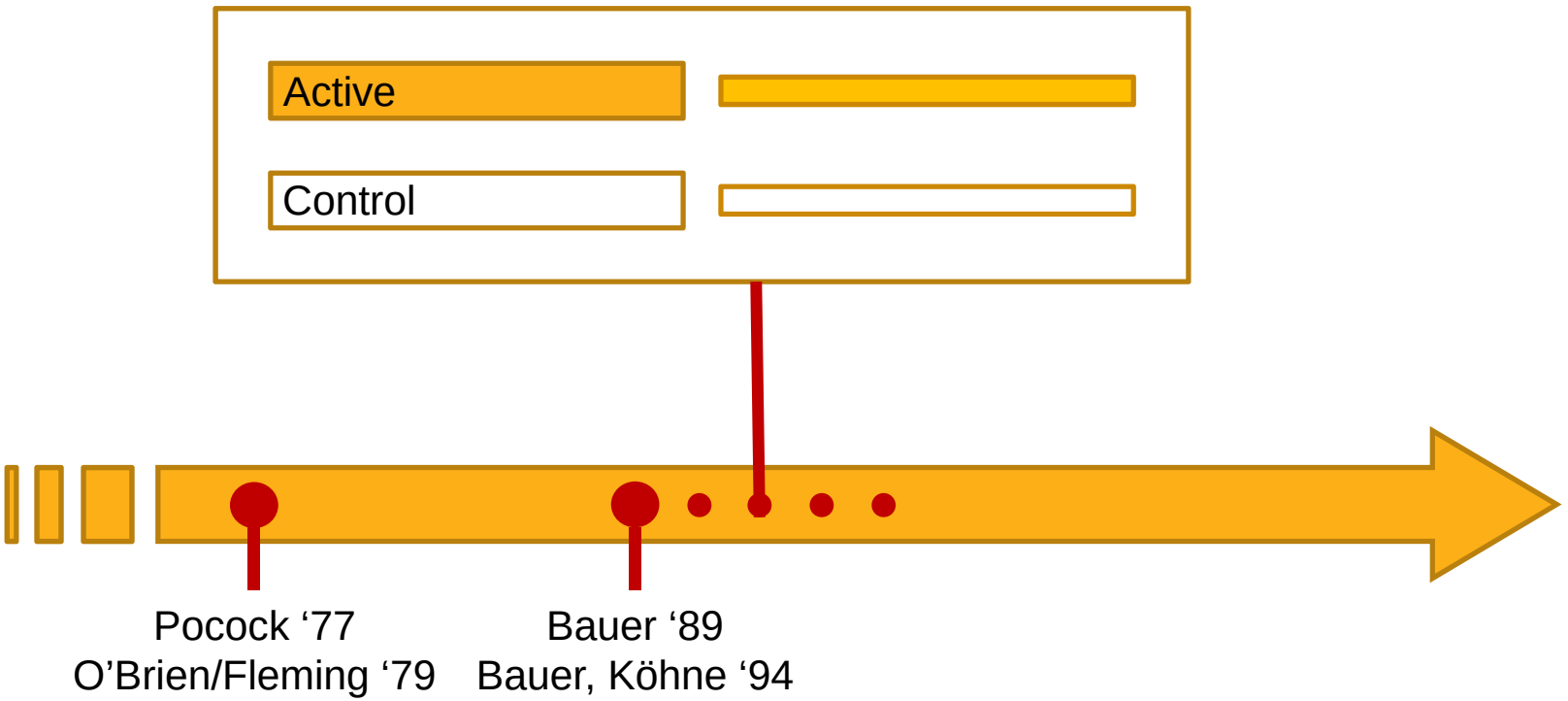


NOVARTIS

Brief historical overview

(Focusing on methods for the confirmatory situation)

- Adaptive design
  2. Adjust the study population («enrichment design»)

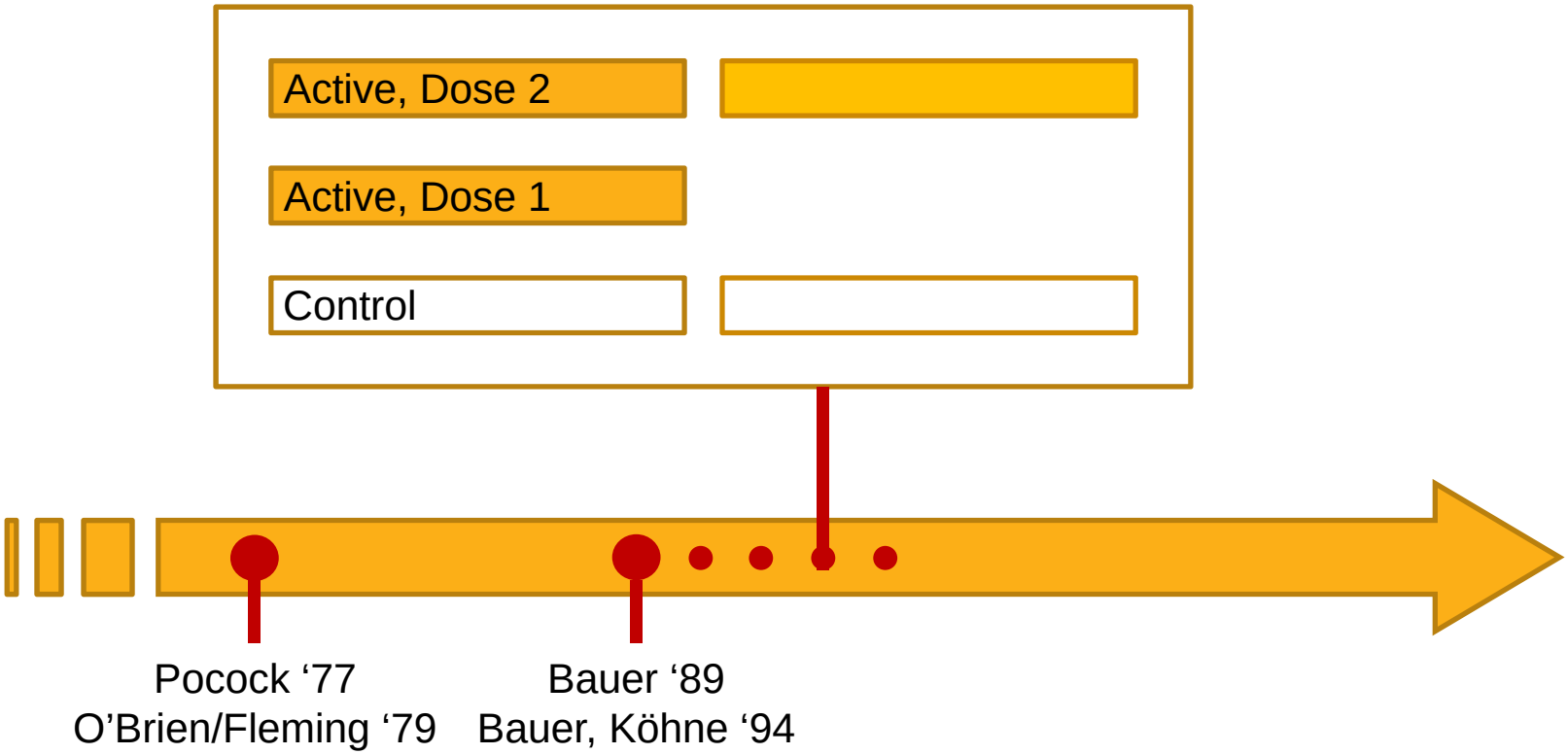


NOVARTIS

Brief historical overview

(Focusing on methods for the confirmatory situation)

- Adaptive design
  3. Select treatment arms

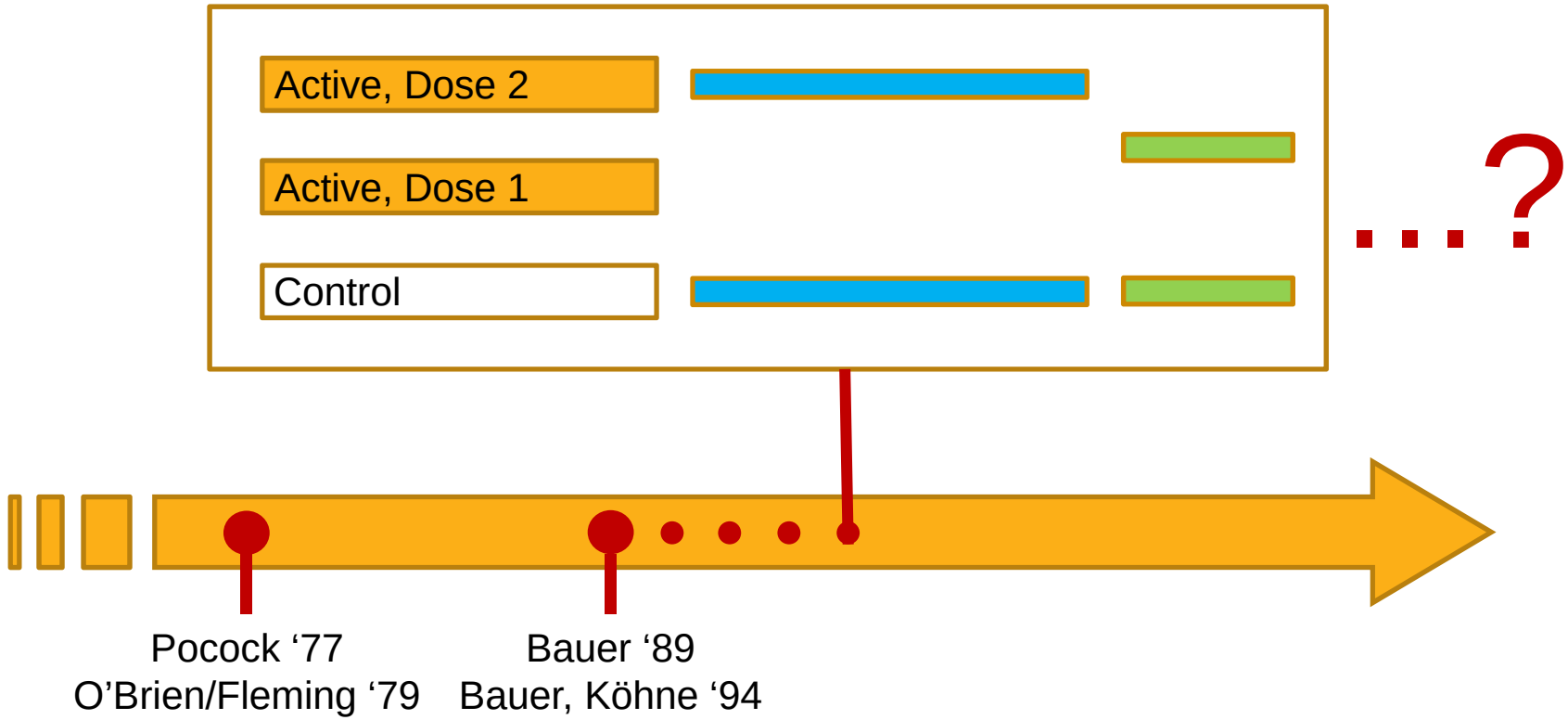


NOVARTIS

Brief historical overview

(Focusing on methods for the confirmatory situation)

- Adaptive design
  4. Change also the null hypothesis, the test statistic, and more?

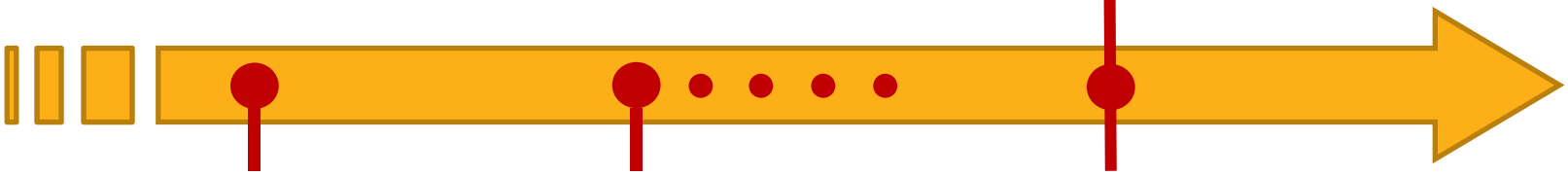


NOVARTIS

Brief historical overview

(Focusing on methods for the confirmatory situation)

- Guidance by Health Authorities



Pocock ‘77
O’Brien/Fleming ‘79

Bauer ‘89
Bauer, Köhne ‘94

EMA ’07
FDA ’10, ‘15, ‘18

NOVARTIS

Brief historical overview

(Focusing on methods for the confirmatory situation)

- Group sequential and adaptive designs today
  - Methodologically developed, regulated by health authorities
  - Effective and accepted – if implemented wisely
  - Emphasis on measures to protect «trial integrity»
  - Main areas of application
    - Interim analyses (early stopping) and sample size adaptations
    - Treatment arm selection
    - Patient enrichment («personalized medicine»)
    - Combination of adaptivity and other sources of multiplicity
    - Exploratory development phases (!)
  - Areas of research
    - Complex combinations (early/late endpoints, multiple sources of multiplicity)
    - Inference after an adaptive design

Outline

- Introduction
  - Brief historical overview
  - Example
  - Interim analyses
- Group sequential trials
  - Classical group sequential designs
  - Spending functions
  - Estimation following a group sequential trial
- Adaptive trials
  - Motivation and overview
  - Methods for the confirmatory situation
  - Estimation following an adaptive trial
- Case study and Final words

Example

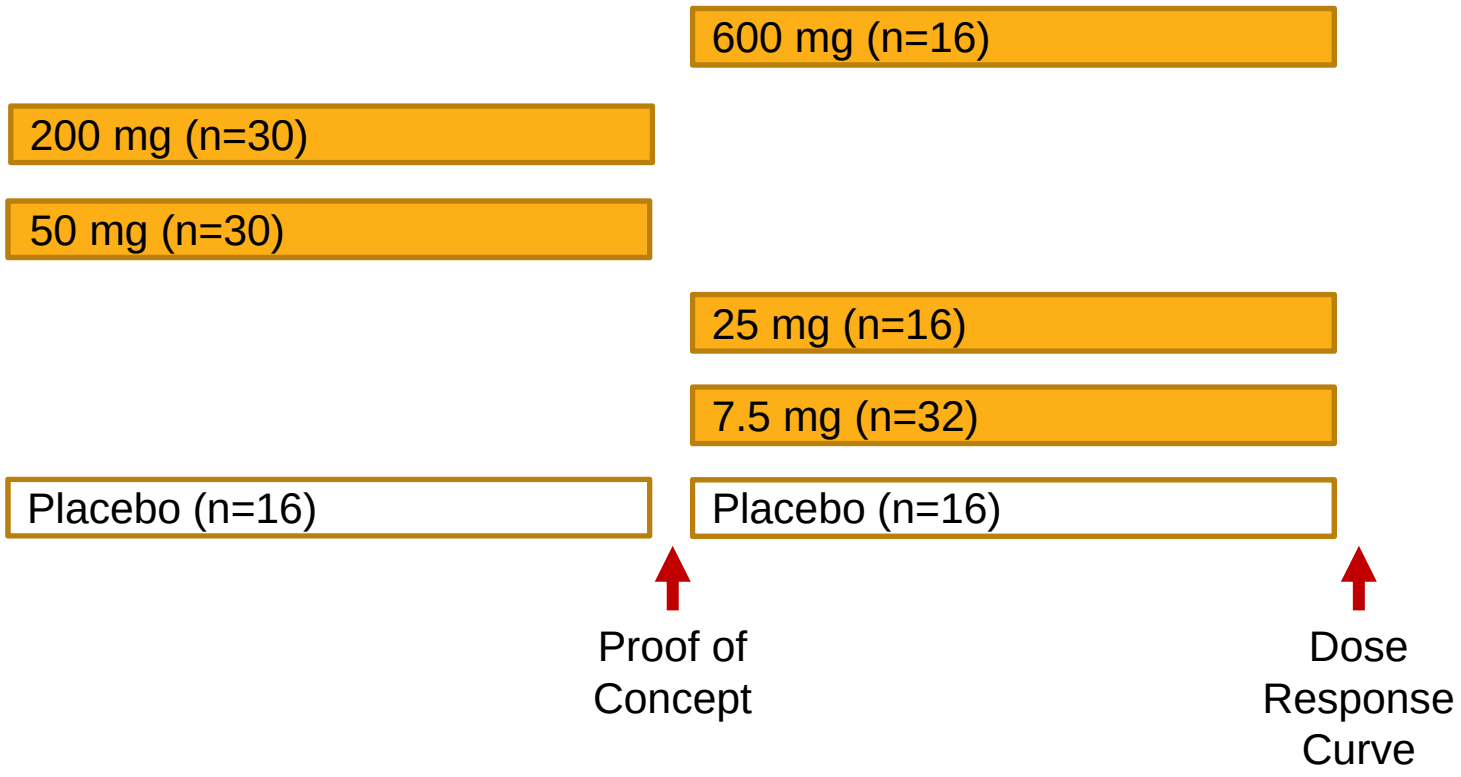
Proof-of-Concept and Dose Finding in Dental Pain

NOVARTIS

Example

Proof-of-Concept and Dose Finding in Dental Pain

- Vandemeulebroecke et al. 2010

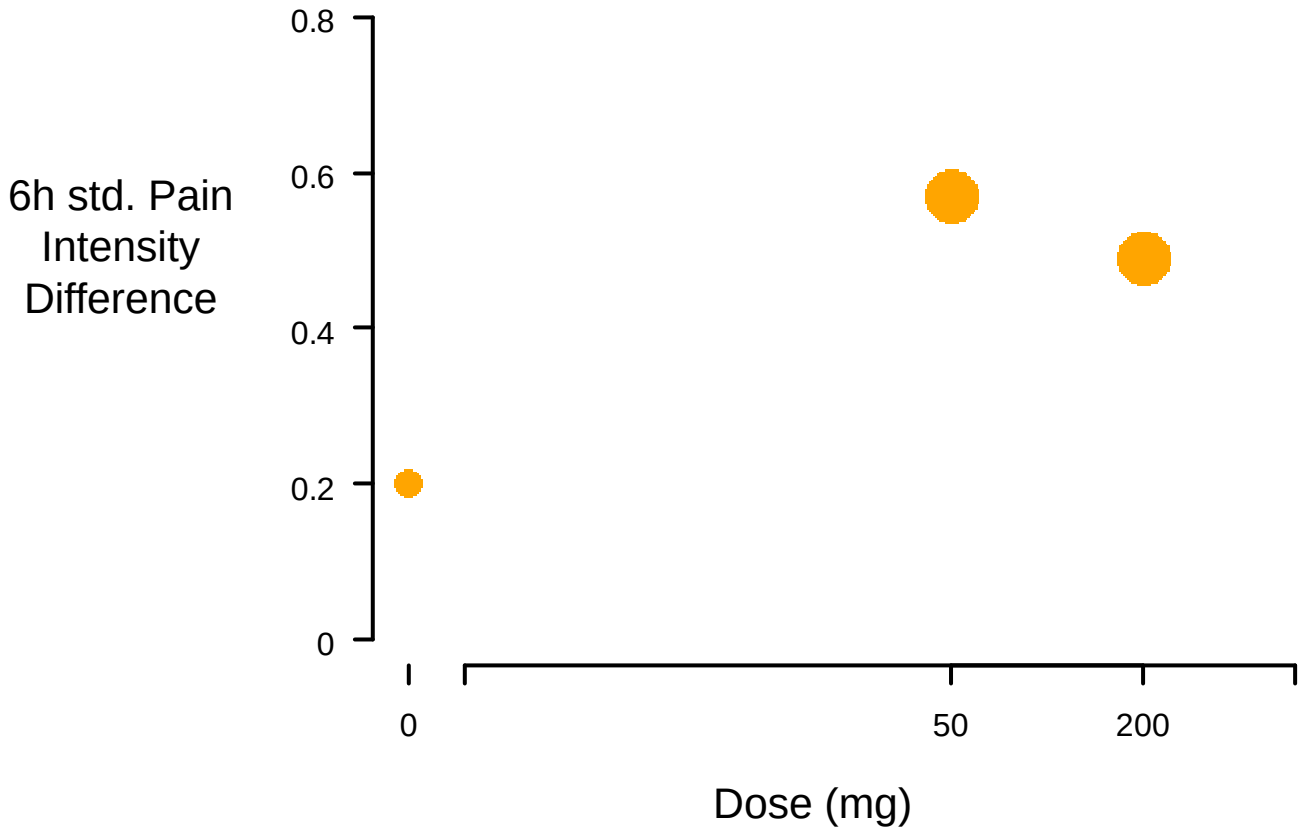


NOVARTIS

Example

Proof-of-Concept and Dose Finding in Dental Pain

- Interim analysis

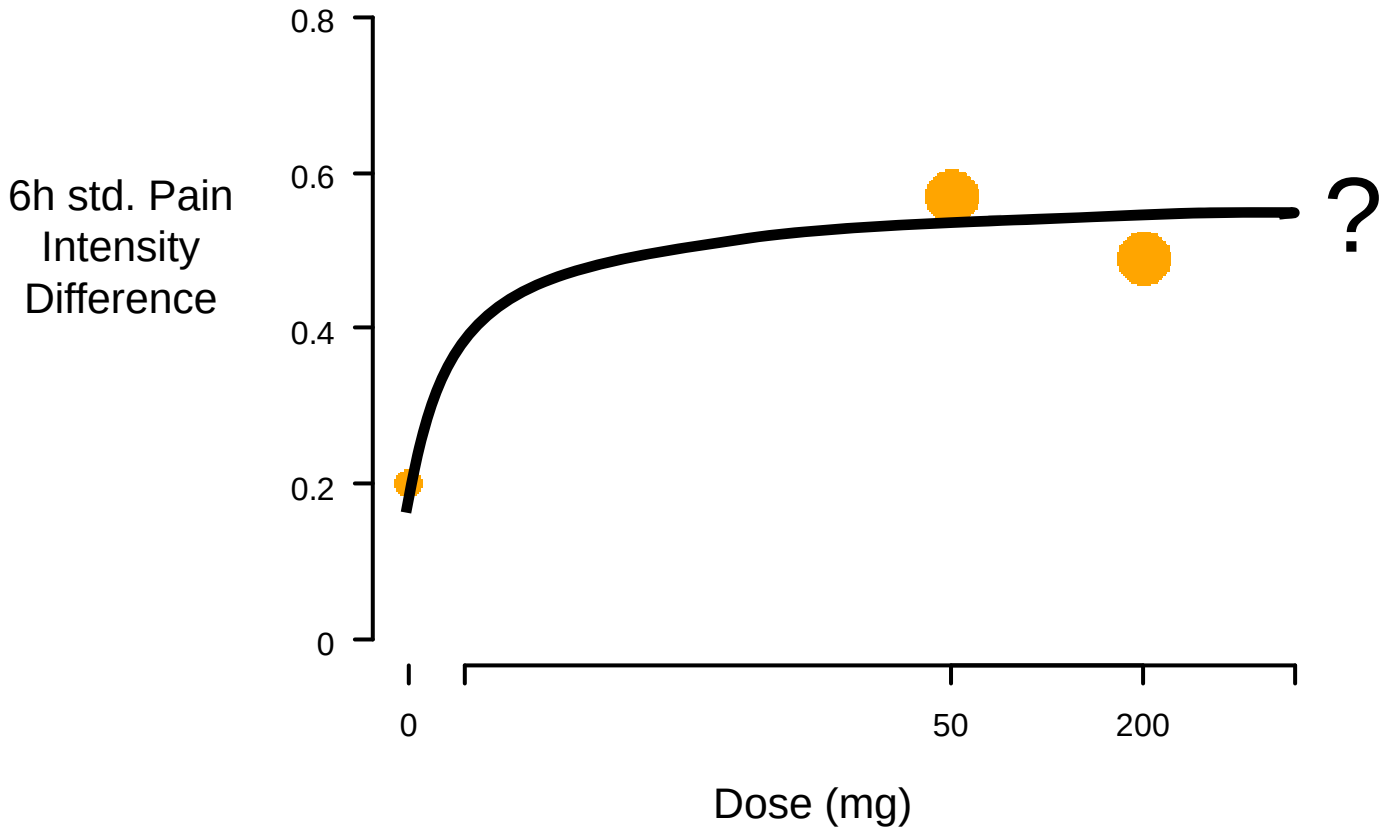


NOVARTIS

Example

Proof-of-Concept and Dose Finding in Dental Pain

- Interim analysis

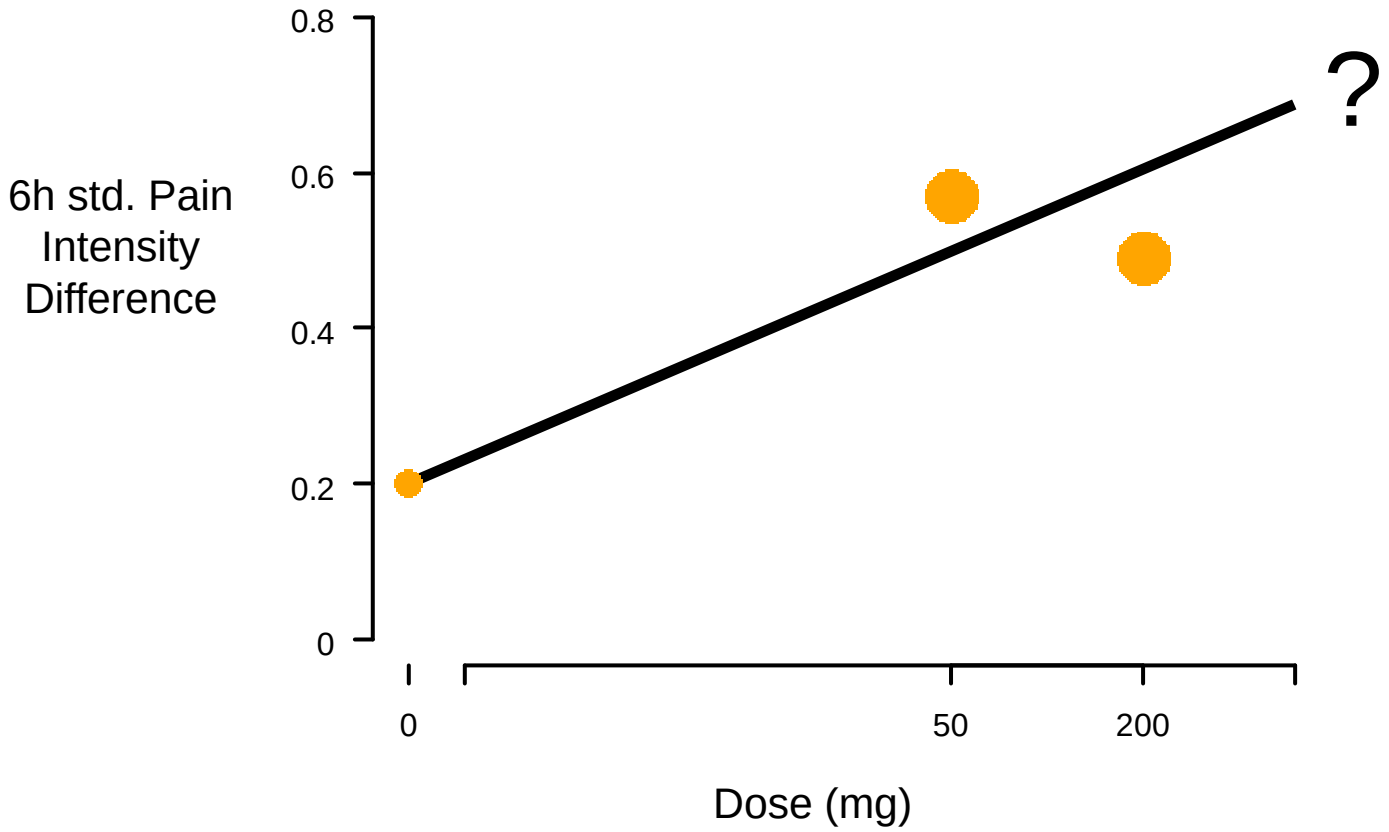


NOVARTIS

Example

Proof-of-Concept and Dose Finding in Dental Pain

- Interim analysis

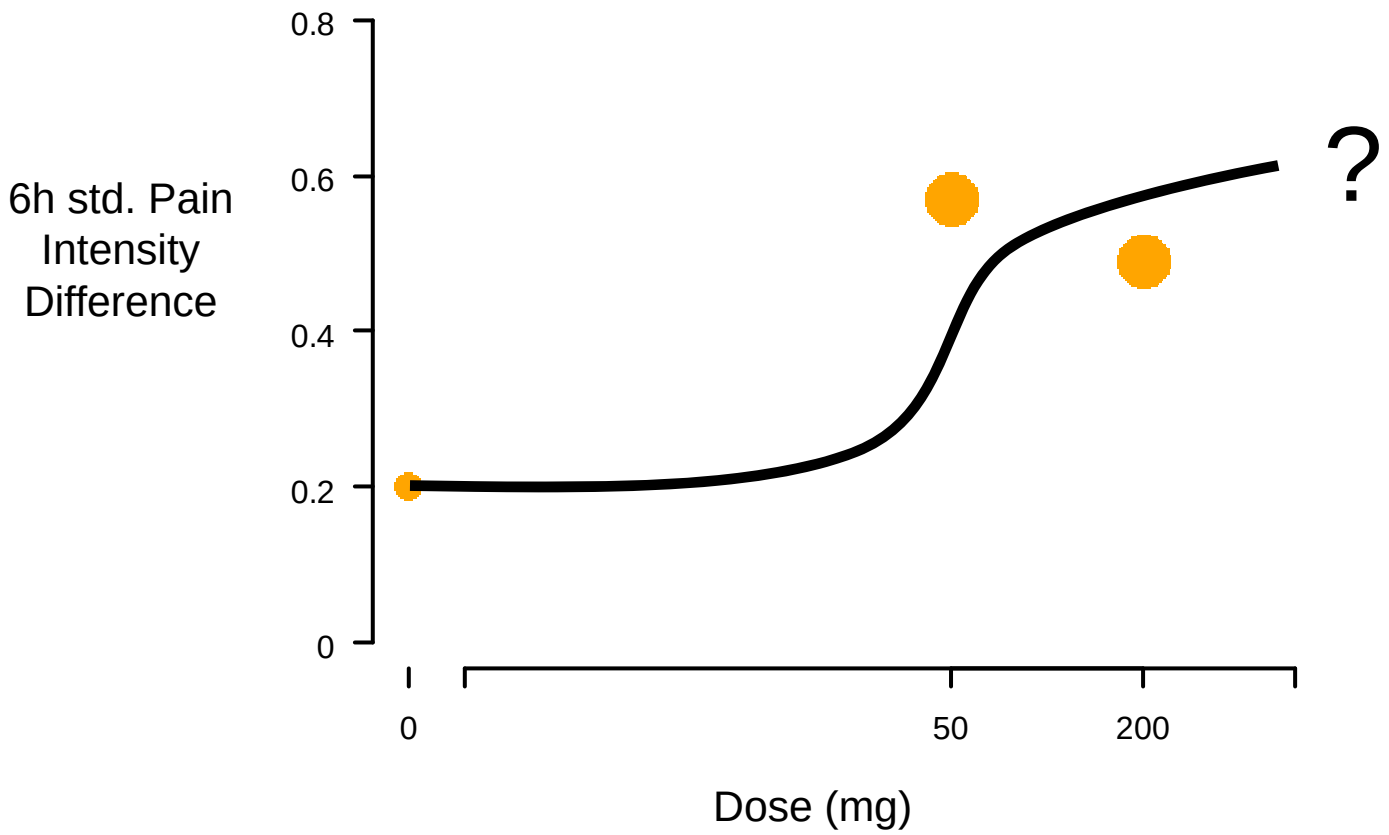


NOVARTIS

Example

Proof-of-Concept and Dose Finding in Dental Pain

- Interim analysis

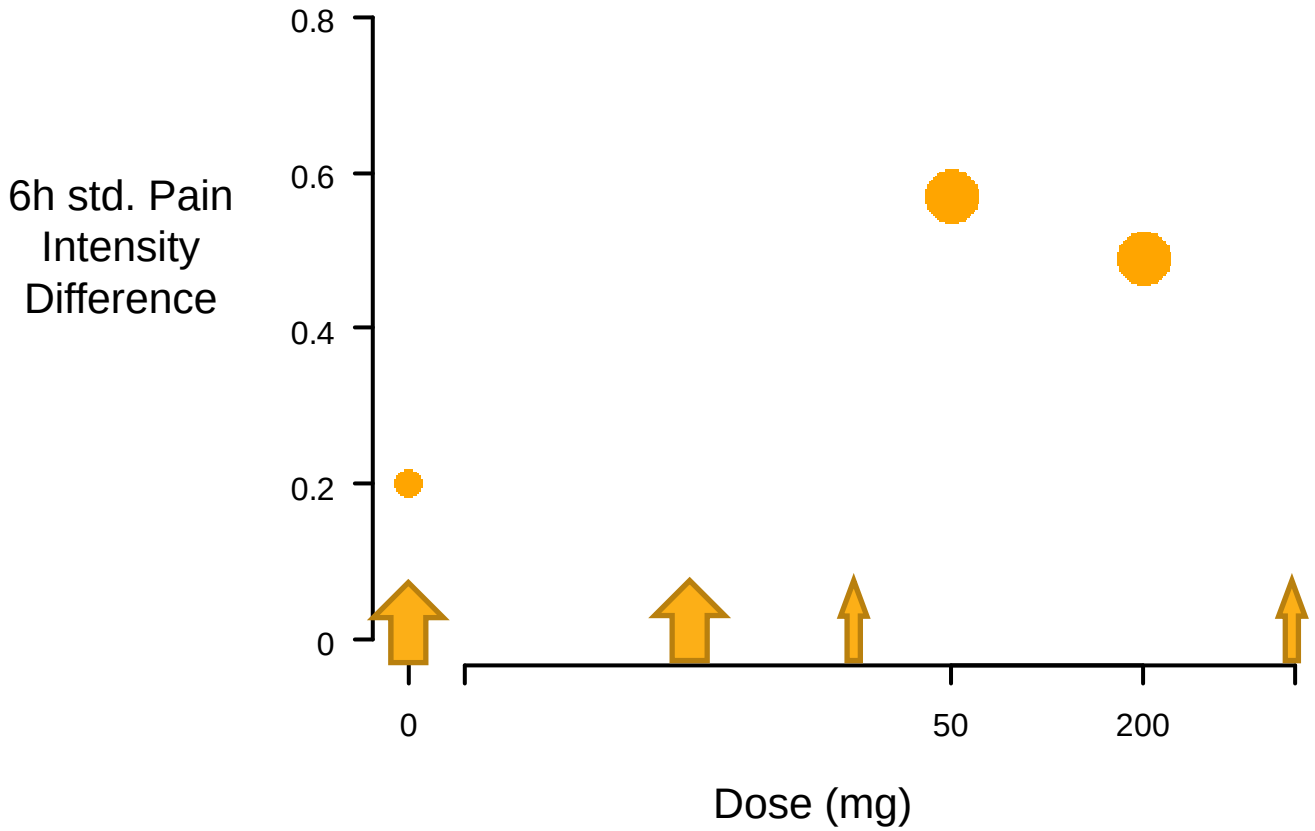


NOVARTIS

Example

Proof-of-Concept and Dose Finding in Dental Pain

- Interim analysis

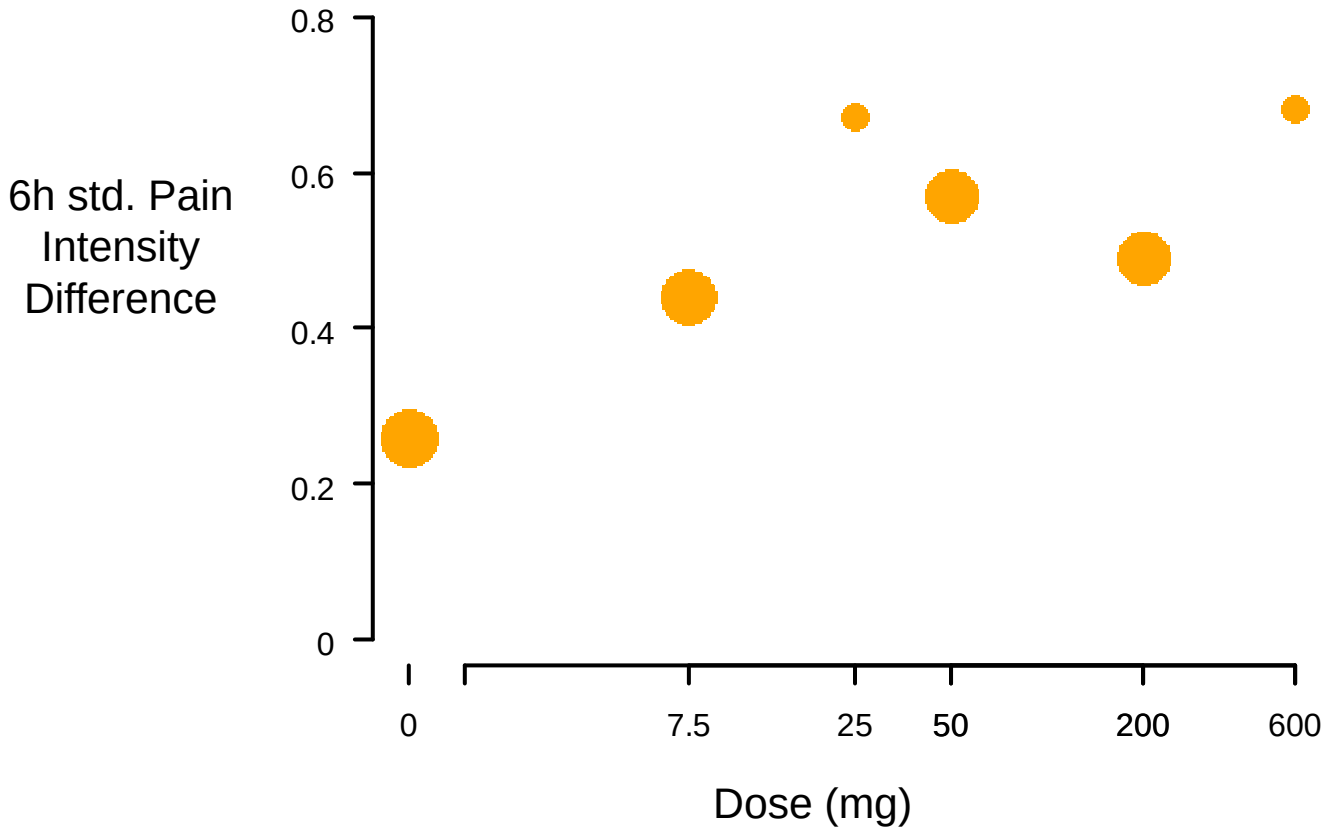


NOVARTIS

Example

Proof-of-Concept and Dose Finding in Dental Pain

- Final analysis

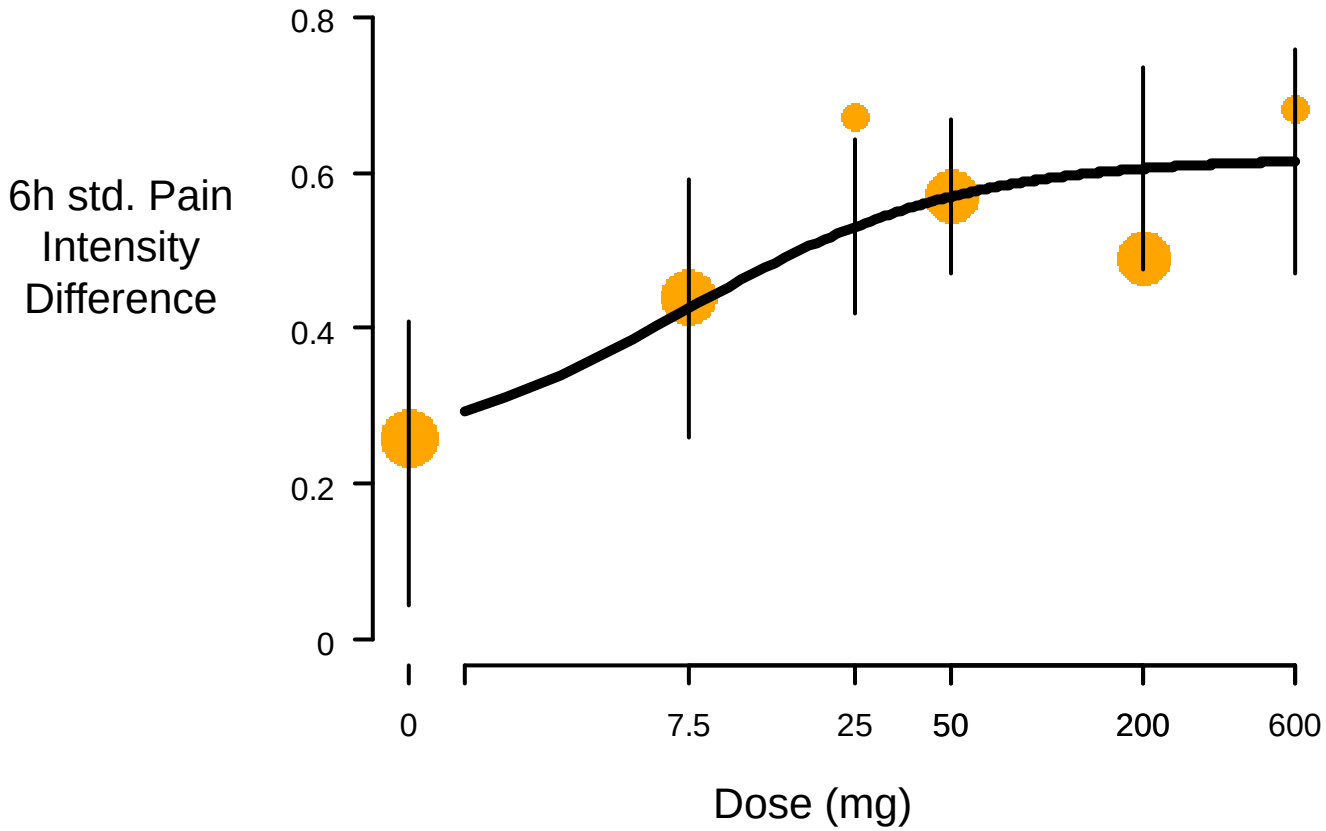


NOVARTIS

Example

Proof-of-Concept and Dose Finding in Dental Pain

- Final analysis



NOVARTIS

Outline

- Introduction
  - Brief historical overview
  - Example
  - Interim analyses
- Group sequential trials
  - Classical group sequential designs
  - Spending functions
  - Estimation following a group sequential trial
- Adaptive trials
  - Motivation and overview
  - Methods for the confirmatory situation
  - Estimation following an adaptive trial
- Case study and Final words

Interim analyses

***What** is an interim analysis?*

- Interim analysis:
  - Any analysis or inspection of unblinded data (individual or summary data) of a trial while it is ongoing

Interim analyses

***Why** would we do an interim analysis?*

1. Opportunity to **stop** the trial (→ early stopping)
   - For safety/tolerability issues
     - Avoid unnecessary exposure of patients to harmful treatment
     - Trials in serious diseases will often require interim analyses
   - For lack of efficacy («futility»)
     - Avoid unnecessary exposure of patients to ineffective treatments
     - Abandon lost cause and save money & resources
   - For overwhelming evidence of efficacy
     - Expedite further development
     - Bring effective treatment to patients earlier
     - Increase profit before patent expiry

Interim analyses

***Why** would we do an interim analysis?*

2. Opportunity to **modify** the trial (→ adaptive design)

- Based on an interim check of planning assumptions
  - Variability greater than expected → increase sample size
  - Effect size smaller than expected (but still of interest) → increase sample size
- Based on new external information
  - Health authority accepts earlier endpoints → focus on these
  - Competitor withdraws from market → targeting smaller effect sizes may suffice
- Select treatment arms, enrich subpopulation etc.
  - Some treatment arms lack efficacy → drop these
  - Efficacy only in pre-defined subgroup → continue recruiting in this subgroup only
- *Continued below...*

Interim analyses

***Why** would we do an interim analysis?*

- Combine development phases to reduce «white space» between them («seamless design»)
  - Dose Finding and Pivotal study → drop treatment arms (as above)
  - Proof of Concept and Dose Finding → expand dose range (dental pain ex.)
  - Caveat: «White space» may help to digest the data and take the right decisions. Combining PoC and DF can be a good idea, but combining DF and Pivotal can be risky

3. Make decisions **external** to the trial

- Sometimes called «administrative interim analysis»
- Planning further trials
  - E.g. sample size calculation based on interim effect estimate
- Trigger other development activities
  - Formulation optimization, manufacturing...

Interim analyses

Challenges

- **Challenges** with interim analyses (particularly in the confirmatory situation)
  - Logistical
    - May need independent Data Monitoring Committee (DMC)
    - Consider recruitment speed vs. time required for the interim analysis
    - Diminishing cost savings: stopping after half the study does not save half the costs
    - Increased work (data analysis, planning efforts)
  - Regulatory
    - Increased burden of proof that «trial integrity» was not compromised
    - More discussions and documentation
  - Statistical
    - Control of Type I error (probability of false positive conclusion)
    - Obtain a reliable treatment effect estimate after the trial
    - Inference on secondary endpoints

Interim analyses

Data Monitoring Committee

- The Data Monitoring Committee (DMC)...
  - ...is a group of experts that review and evaluate interim data to ensure patient safety as well as the validity and scientific merit of the trial
  - ...can be external or internal to the company
    - Recommendation (FDA, 2006) for **external** DMC in late phase trials, particularly in Phase III and/or mortality or high risk trials
    - External DMC members should have no conflict of interest
  - ...would consist of 3-7 experts (often an odd number) covering multiple disciplines (clinicians, one statistician, a safety expert...)
  - ...makes recommendations on interim trial decisions based on:
    - Pre-agreed guidance as per the DMC charter
    - Expert judgement from the totality of available data
    - E.g., would recommend stopping for efficacy if stopping criteria are met **and** overall results are persuasive

Interim analyses

Trial integrity – critical for the confirmatory situation!

- Knowledge of interim results can introduce **unconscious biases** into the conduct of the trial
  - E.g. decision to close certain study sites may be influenced by a low observed treatment effect in these sites («operational bias»)
- Burden of proof is on the sponsor
  - Imagine that males respond better, and more males are recruited late in the trial without clear reason
  - Without interim analysis, then may be fine. With interim analysis, how can you prove that male «enrichment» wasn't done on purpose?
- Failure to adequately protect interim information can:
  - Impair the interpretability of the results, harm the credibility of the trial
  - Downgrade the regulatory status of the trial (confirmatory → supportive)

Interim analyses

Statistical validity – in part. Type I error control – critical for confirm. situat.!

- **Multiple shots on goal**
  - Multiple opportunities to declare success (interim and final analysis)
  - For an ineffective drug: greater chance to declare success **incorrectly**
  - This means an inflation of the Type I error probability
    - E.g., with equally spaced analyses, each at the 5% level:

Number of interim analyses	1	2	5	10
Overall probability of a Type I error	5%	8%	14%	19%

- **Remedy**: Adjust for multiplicity!
  - Adjust thresholds for declaring success such that the overall Type I error probability is as desired (e.g. 5%)
  - This is the main purpose of group sequential trial designs
  - Not needed if no option to stop for efficacy
    - Pure safety monitoring; only futility stopping option; purely administrative interim look

Outline

- Introduction
  - Brief historical overview
  - Example
  - Interim analyses
- Group sequential trials
  - Classical group sequential designs
  - Spending functions
  - Estimation following a group sequential trial
- Adaptive trials
  - Motivation and overview
  - Methods for the confirmatory situation
  - Estimation following an adaptive trial
- Case study and Final words

Classical group sequential designs

Single sample Z-test

- Test Active vs. Control with normally distributed endpoint
- Fixed sample trial (no interim analysis)
- $X_i \sim N(\mu, \sigma^2)$, $Y_i \sim N(\nu, \sigma^2)$, independent
  - $i = 1, \ldots, n$ (subject); σ^2 assumed known
- Null hypothesis: $H_0: \mu = \nu$; alternative hyp.: $H_0: \mu \neq \nu$
- Z-statistic: $Z = \sqrt{\frac{1}{2n\sigma^2}} \left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i\right) \sim_{H_0} N(0,1)$
- Z-test: Reject H_0 if $|Z| > z_{1-\alpha/2}$ ($= 1.96$ for $\alpha = 0.05$)

Classical group sequential designs

Two-stage Z-test

- One interim analysis after half the sample size
- $X_{ki} \sim N(\mu, \sigma^2)$, $Y_{ki} \sim N(\nu, \sigma^2)$, independent
  - $k = 1,2$ (trial stage); $i = 1, \ldots, n$ (subject (within stage)); σ^2 known
- Cumulative Z-statistics:
  - After the first stage: $Z_1^* = \sqrt{\frac{1}{2n\sigma^2}} \left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i\right)$ (same as before)
  - After the second stage: $Z_2^* = \sqrt{\frac{1}{4n\sigma^2}} \left(\left(\sum_{i=1}^{n} X_{1i} + \sum_{i=1}^{n} X_{2i}\right) - \left(\sum_{i=1}^{n} Y_{1i} + \sum_{i=1}^{n} Y_{2i}\right)\right)$
- Two-stage Z-test:
  - Reject H_0 if $|Z_1^*| > c_1$ or $|Z_2^*| > c_2$ → How to set c_1 and c_2?

Classical group sequential designs

Two-stage Z-test

- Pocock (1977) proposed a constant boundary
  - $c_1 = c_2 = c^{(P)}$
  - With $c^{(P)}$ chosen such that the overall Type I error is α
    - $P_0\left(|Z_1^*| > c^{(P)} \text{ or } |Z_2^*| > c^{(P)}\right) = \alpha$
    - $P_0\left(|Z_1^*| \leq c^{(P)} \text{ and } |Z_2^*| \leq c^{(P)}\right) = 1 - \alpha$
    - $\int_{-c^{(P)}}^{c^{(P)}} \int_{-c^{(P)}}^{c^{(P)}} \Phi(z_1, z_2; \rho)\, d(z_1, z_2) = 1 - \alpha$ → Solve for $c^{(P)}$
    - Note: we exploit the (known) correlation ρ of Z_1^* and Z_2^* (here, $\rho = \sqrt{0.5}$)
- Proceed as follows
  - After the first stage, reject H_0 if $|Z_1^*| > c^{(P)}$
    - Otherwise continue to second stage
  - After the second stage, reject H_0 if $|Z_2^*| > c^{(P)}$
    - Otherwise do not reject H_0

NOVARTIS

Here, $\Phi(\cdot, \cdot\,; \rho)$ denotes the bivariate standard normal distribution function with correlation ρ.

Classical group sequential designs

Multi-stage Z-test

- Multiple interim analyses, equally spaced
- $X_{ki} \sim N(\mu, \sigma^2)$, $Y_{ki} \sim N(\nu, \sigma^2)$, independent
  - $k = 1, \ldots, K$ (trial stage); $i = 1, \ldots, n$ (subject (within stage)); σ^2 known
- Cumulative Z-statistics:
  - After each stage: $Z_k^* = \sqrt{\frac{1}{2kn\sigma^2}} \left(\sum_{j=1}^{k} \sum_{i=1}^{n} X_{ji} - \sum_{j=1}^{k} \sum_{i=1}^{n} Y_{ji} \right)$
- Group sequential test:
  - Reject H_0 and stop the trial as soon as $|Z_k^*| > c_k$
  - If $|Z_k| \leq c_k$ throughout the trial (for all k), do not reject H_0
  - With c_k such that the overall Type I error is α

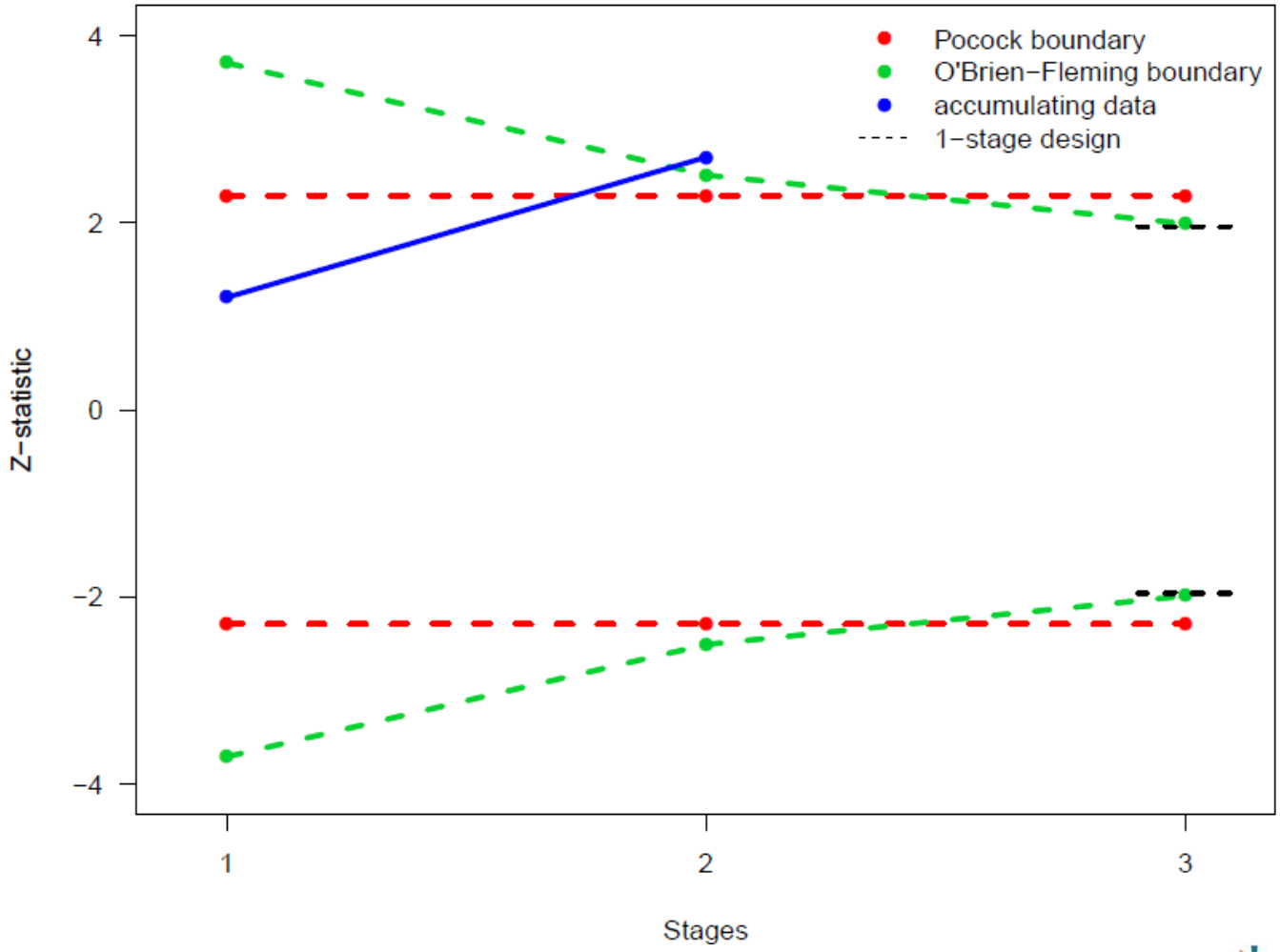
Classical group sequential designs

Variants

- Pocock (1977): constant boundary
  - $c_k = c^{(P)}$ for all k
- O'Brien and Fleming (1979): decreasing boundary
  - $c_k = \frac{c^{(OBF)}}{\sqrt{k}}$
- Wang and Tsiatis (1987): Delta-class
  - $c_k = c^{(WT)} k^{\Delta - 0.5}$
- Etc...
- Always chosen such that the overall Type I error is α

Classical group sequential designs

Illustration



NOVARTIS

Slide from Rufibach, Sabanés Bové, Vandemeulebroecke, Wolbers

Classical group sequential designs

Extensions

- One-sided hypothesis tests
- Interim analyses not equally spaced
  - They should still be pre-specified
- Non-normally distributed endpoints
  - The same principle applies
  - The test statistics should be approximately normally distributed
  - Their stage-wise increments should be independent
  - This covers binary data, count data, time-to-event data, etc.

Classical group sequential designs

Summary and remarks

- Basic group sequential designs provide a valid way to perform interim analyses
  - Allow early stopping for efficacy with full control of the Type I error
  - Number of stages and stage sizes must be pre-specified
- The option to stop for futility can also be accomodated
  - Futility stopping option never increases the Type I error
    - But it decreases the power of the study
  - **Informal** futility guidance always possible (outside of statistical procedure)
  - **Formal** (mandatory) futility rules
    - Lost rejection probability could be recovered by more liberal efficacy bounds – but Health Authorities tend to be skeptical

Classical group sequential designs

Summary and remarks

- Pocock design uses constant boundaries
  - Easy (too easy?) to stop early for efficacy, price to pay at the end
  - Expected sample size is considerably smaller than for single-stage design, but maximal sample size is larger
  - Rarely used in practice
- O'Brien/Fleming design uses decreasing boundaries
  - Requires very strong evidence for early efficacy stop; reserves most α for the end
  - Expected sample size is still notably smaller than that of a single-stage design; maximal sample size is almost the same
  - More often used in practice

Outline

- Introduction
  - Brief historical overview
  - Example
  - Interim analyses
- Group sequential trials
  - Classical group sequential designs
  - Spending functions
  - Estimation following a group sequential trial
- Adaptive trials
  - Motivation and overview
  - Methods for the confirmatory situation
  - Estimation following an adaptive trial
- Case study and Final words

NOVARTIS

Spending functions

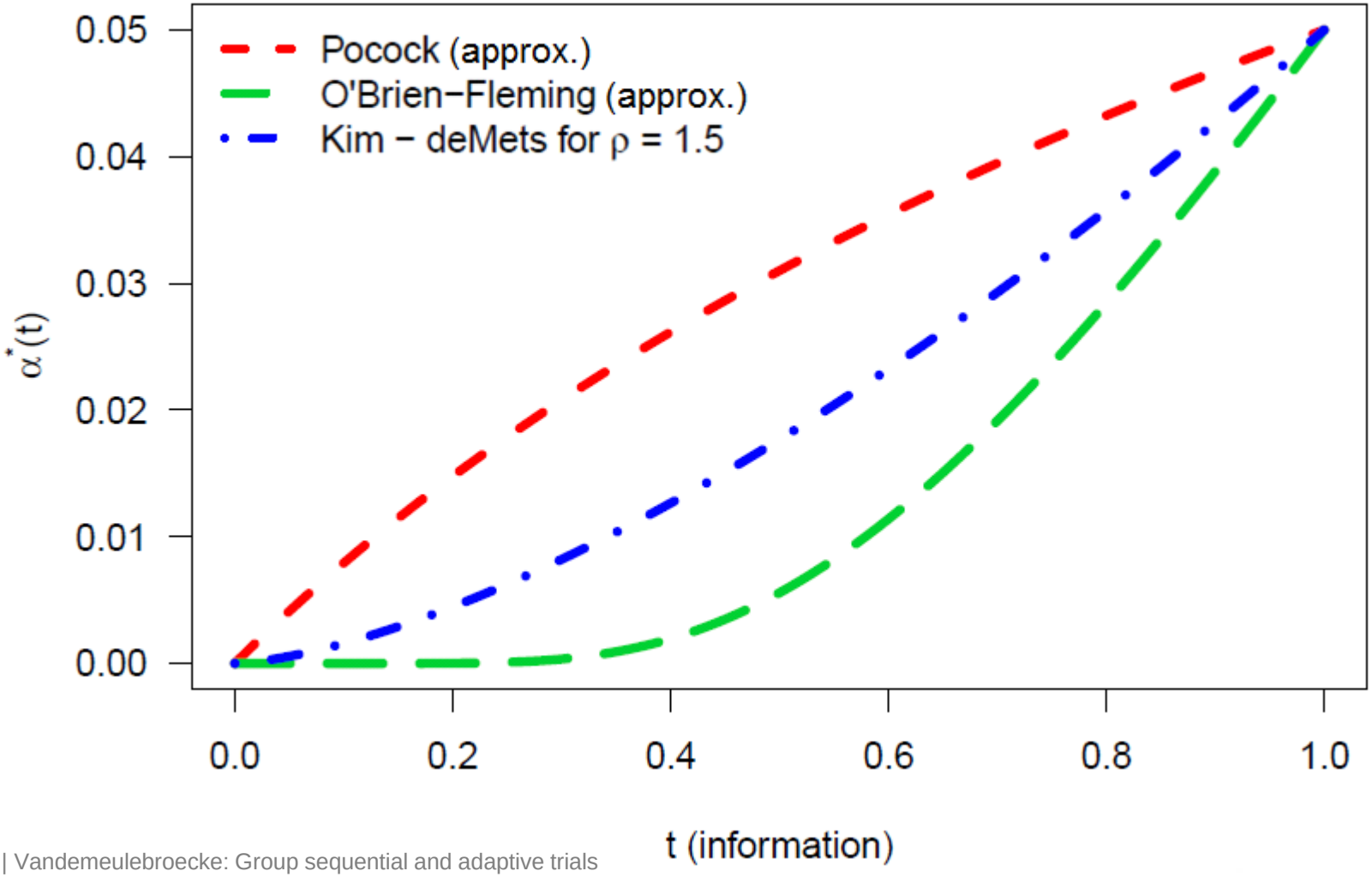
- Error spending function
  - Probability of making a Type I error «now or earlier»
  - Expressed as a function of the «information time»
  - E.g., for Pocock after stage k
    - «Information time» is fraction of planned total sample size: $t = \frac{2kn}{2Kn} = \frac{k}{K}$
    - Error «spent» is $\alpha^*(t) = P_0\left(|Z_1^*| > c^{(P)} \text{ or } |Z_2^*| > c^{(P)} \text{or } \ldots \text{ or } |Z_k^*| > c^{(P)}\right)$
    - Note: $\alpha^*(0) = 0$, $\alpha^*(1) = \alpha$
  - In general, any monotonic function $\alpha^*: [0,1] \rightarrow [0, \alpha]$
- Lan and DeMets (1983) proposed to define a group sequential test through such a function.

Spending functions

- Interim analyses need not be specifed a priori
  - Stage sizes per arm: n_k
  - $N = \sum_{j=1}^{K} n_j$ fixed, but K and n_k not!
  - Information fraction after stage k: $t_k = \frac{1}{N}\sum_{j=1}^{k} n_j$
  - Cumulative Z-statistics similarly as before: Z_k^*
- Choose spending function $\alpha^*: [0,1] \rightarrow [0, \alpha]$
- Successively determine boundary values c_1, ..., c_K from:
  - $k = 1$: $P_0(|Z_1^*| > c_1) = \alpha^*(t_1)$
  - $k = 2, \ldots, K$: $P_0(|Z_1^*| \leq c_1, \ldots, |Z_{k-1}^*| \leq c_{k-1}, |Z_k^*| > c_k) = \alpha^*(t_k) - \alpha^*(t_{k-1})$

Spending functions

- Commonly used spending functions



RTIS

Slide from Rufibach, Sabanés Bové, Vandemeulebroecke, Wolbers

Spending functions

Summary and remarks

- Spending functions provide a flexible way to define group sequential designs
  - Maximal total sample size must be pre-specified
  - Number of stages and stage sizes can be chosen ad-hoc
  - But they must not depend on interim results
- In practice, we never use «pure» classical group sequential designs. We use spending functions that approximate them to retain flexibility.
  - E.g., if we plan an interim analysis after half the trial, but for logistical reasons it happens after 60% of the patients, then we can simply compute the corresponding stopping bound based on the spending function.

Outline

- Introduction
  - Brief historical overview
  - Example
  - Interim analyses
- Group sequential trials
  - Classical group sequential designs
  - Spending functions
  - Estimation following a group sequential trial
- Adaptive trials
  - Motivation and overview
  - Methods for the confirmatory situation
  - Estimation following an adaptive trial
- Case study and Final words

NOVARTIS

Estimation following a group sequential trial

- Naive treatment effect estimates at the end of a group sequential trial will generally be biased
  - Because we may be stopping at a «random high», truncating trajectories of the effect estimate that would otherwise even out again
  - The same holds for «random lows» in the case of futility stopping
- There are methods to adjust for this effect
  - E.g., by subtracting a bias estimate from the treatment effect estimate (Whitehead 1986; Section 4.2 in Wassmer and Brannath 2016)
  - Also, bias is rather small if interim anlaysis happens late in the trial
- Always balance the bias-variance trade-off
  - Effect estimates should not only have little bias, but also good precision
  - Otherwise we could just use first stage data only!

Outline

- Introduction
  - Brief historical overview
  - Example
  - Interim analyses
- Group sequential trials
  - Classical group sequential designs
  - Spending functions
  - Estimation following a group sequential trial
- Adaptive trials
  - Motivation and overview
  - Methods for the confirmatory situation
  - Estimation following an adaptive trial
- Case study and Final words

Motivation and overview

- Motivations for adaptive trial designs
  - Increase probability to prove clinical benefit, to identify best dose, etc.
  - Reduce cost by more efficient study designs
  - We already mentioned the main examples (see Introduction)
    - Revise planning assumptions if they were off
    - React to new external information
    - Select treatment arms, enrich subpopulation etc.
    - Combine development phases to reduce «white space» between them
  - Not a remedy for sloppy planning!
    - Planning adaptive trials requires more effort than planning standard trials

Motivation and overview

- Challenges with adaptive trials
  - We already mentioned the main ones (see Introduction)
    - Logistical, Data Monitoring Committee
    - Regulatory, protect trial integrity and avoid operational bias
    - Statistical, protect Type I error and obtain valid treatment effect estimates
  - In a pivotal trial, protecting the Type I error is mandatory
  - In exploratory development, there is more flexibility
    - Main risk is on the developer's side, not to «fool oneself»

Motivation and overview

- Definitions of adaptive study designs
  - Dragalin (PhRMA), 2006:
    - A multistage study design that uses accumulating data to decide how to modify aspects of the study without undermining the validity and integrity of the trial. [...] preplanning, as much as possible, based on intended adaptations.
  - EMA reflection paper, 2007:
    - A study is called ‘adaptive’ if statistical methodology allows the modification of a design element [...] at an interim analysis with full control of the type I error.
  - FDA draft guidance, 2010:
    - A study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study.

Motivation and overview

- There are many different types of adaptive trial designs
  - (Group sequential designs, early stopping)
  - Adaptive randomization
  - Adaptive dose escalation
  - Adaptive dose finding
  - Sample size re-estimation
  - Treatment arm selection
  - Enrichment designs
  - ...
- Various «schools» of adaptive designs have developed in parallel, depending on the application area

Motivation and overview

- It’s easy to lose orientation
- Be clear about the situation
  - Exploratory or confirmatory?
  - Adaptations of which trial features?
  - Using unblinded data?
  - Pre-planned adaptations or ad-hoc?
  - Based on interim data or external information?
  - Based on efficacy and/or safety information?

Motivation and overview

Adaptive randomization

- Treatment allocation probabilities are modified during the trial. Main types:
  - Allocation depends on past assignments (e.g. biased coin design, Efron 1971)
    - Goal: Balance treatment allocation over time
  - Allocation depends on covariates and past assignments (e.g. Minimization, Pocock and Simon 1975)
    - Goal: Balance treatment allocation within subgroups
  - Allocation depends on prior responses (e.g. Play-the-winner, Wei and Durham 1978)
    - Goal: Assign patients with greater likelihood to efficacious treatments
  - Combinations of the above...

Motivation and overview

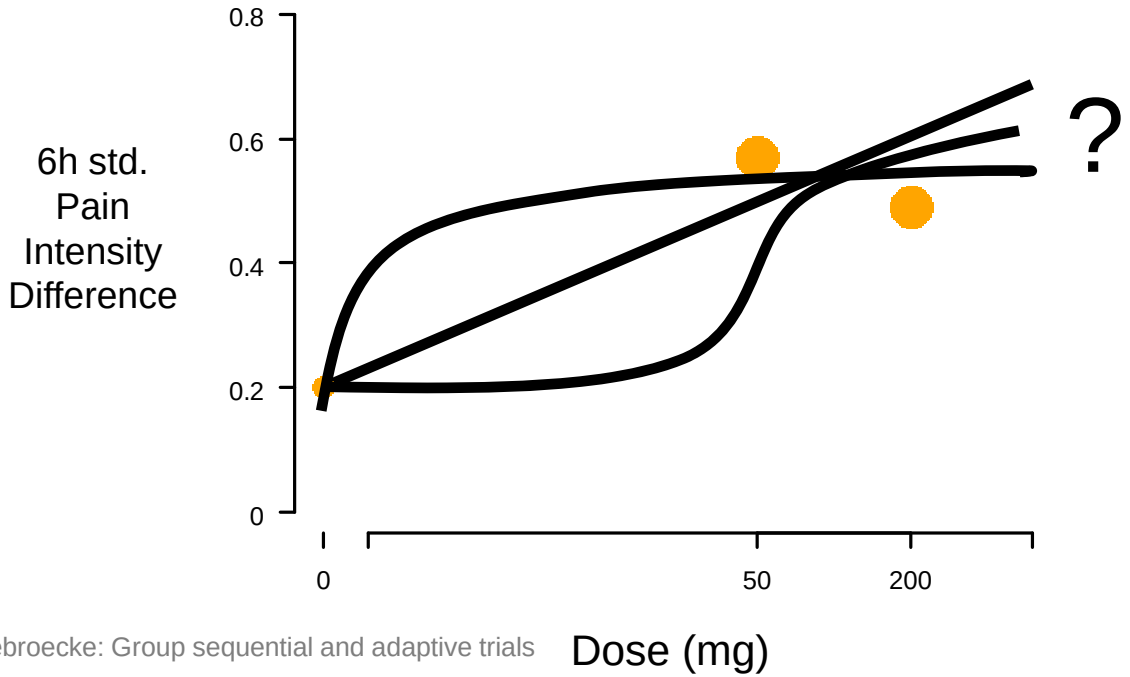
Adaptive dose escalation

- Mainly used in early Oncology studies with high toxicity
- Goals:
  - Estimate the Maximum Tolerated Dose (MTD)
  - Minimize patient exposure to toxic doses
- Traditional algorithmic designs (e.g. «3+3 rule») have given way to designs with better properties
- Continual Reassessment Method (O'Quigley et al. 1990)
- Bayesian adaptive dose escalation
  - Cycle between (i) updating a probability model for dose-limiting toxicities, and (ii) allocating the next cohort of patients

Motivation and overview

Adaptive dose finding

- Find minimum effective dose (MED) – but where to look?
- First allocate broadly, then refine the search with additional patients and/or additional doses
- Remember the motivating example from the Introduction:

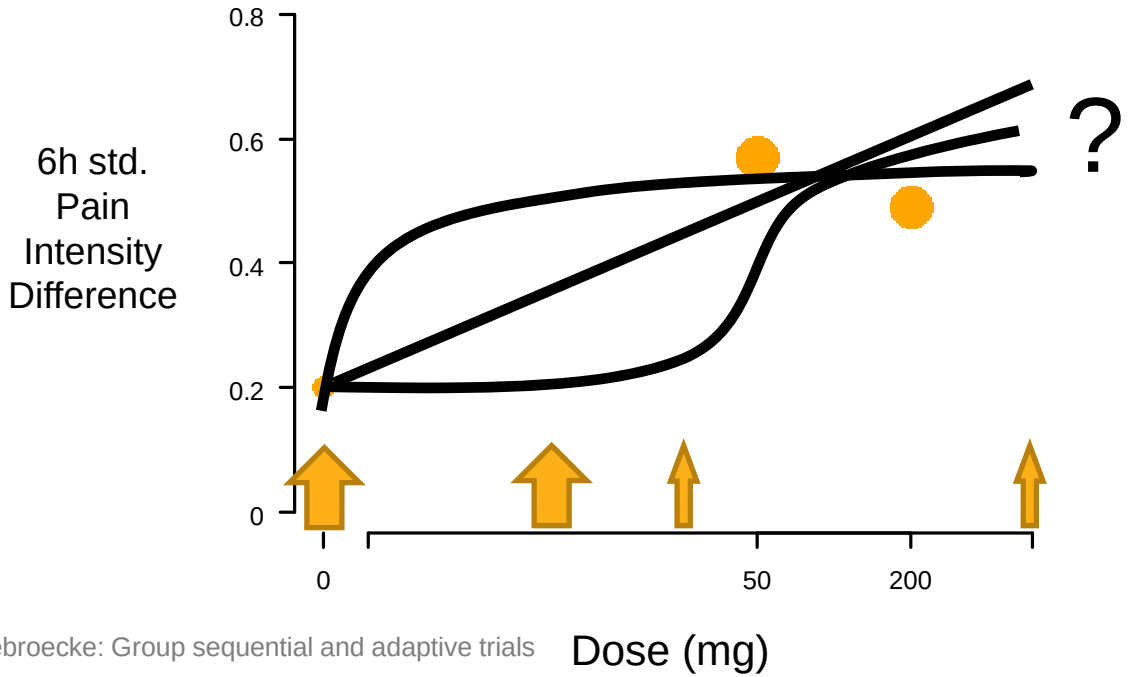


NOVARTIS

Motivation and overview

Adaptive dose finding

- Find minimum effective dose (MED) – but where to look?
- First allocate broadly, then refine the search with additional patients and/or additional doses
- Remember the motivating example from the Introduction:

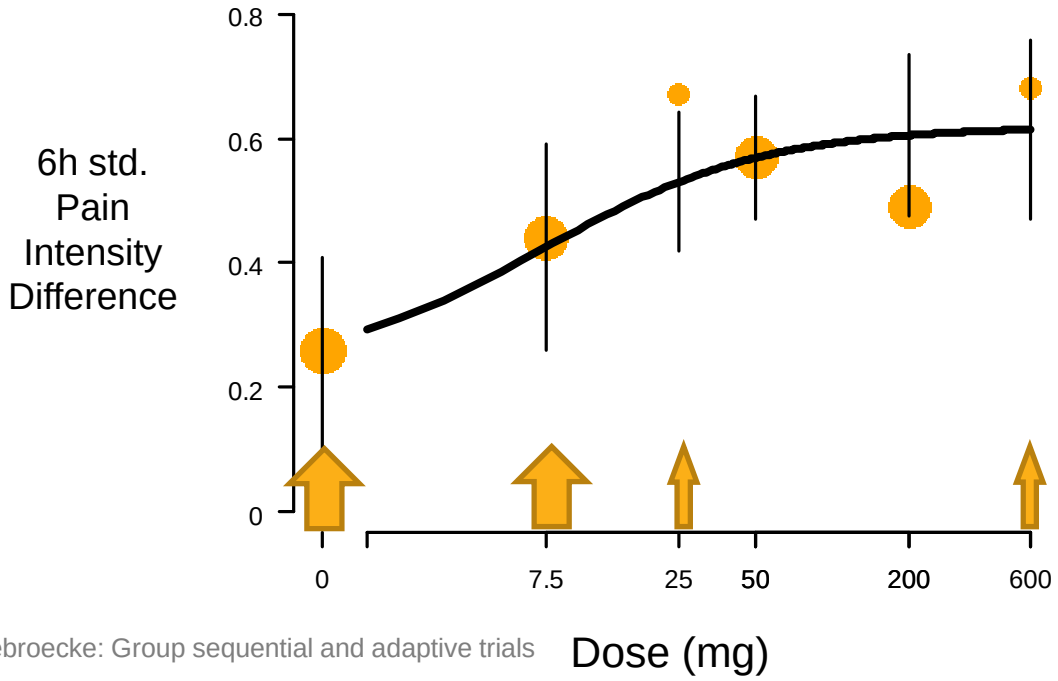


NOVARTIS

Motivation and overview

Adaptive dose finding

- Find minimum effective dose (MED) – but where to look?
- First allocate broadly, then refine the search with additional patients and/or additional doses
- Remember the motivating example from the Introduction:



NOVARTIS

Outline

- Introduction
  - Brief historical overview
  - Example
  - Interim analyses
- Group sequential trials
  - Classical group sequential designs
  - Spending functions
  - Estimation following a group sequential trial
- Adaptive trials
  - Motivation and overview
  - Methods for the confirmatory situation
  - Estimation following an adaptive trial
- Case study and Final words

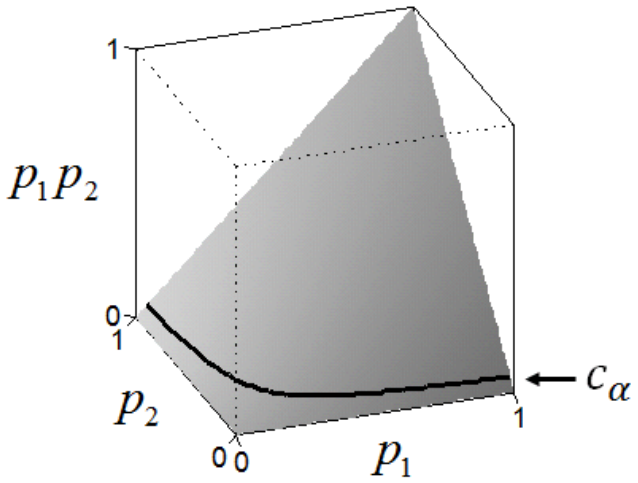
Methods for the confirmatory situation

- In the 1980's, group sequential designs grew popular as they provided a rigorous framework for performing interim analyses and stopping early – but for no other adaptations
- In practice, however, adaptations of running trials were sometimes needed and done. Their impact on the inference was unclear and often ignored.
- Bauer (1989) and Bauer/Köhne (1994) ignited the development of the necessary theory
  - Idea is borrowed from meta-analysis, which combines the inference from separate trials
  - Now: combine the inference from separate stages of one trial
  - This also allows adapting later stages based on information from earlier stages

Methods for the confirmatory situation

Two trial stages

- Combine the p-values of two separate trial stages
  - E.g., in the Product Test, reject H_0 when $p_1 p_2 \leq c_\alpha$
  - With c_α chosen such that the Type I error is α (see next slide)
  - p-value combination



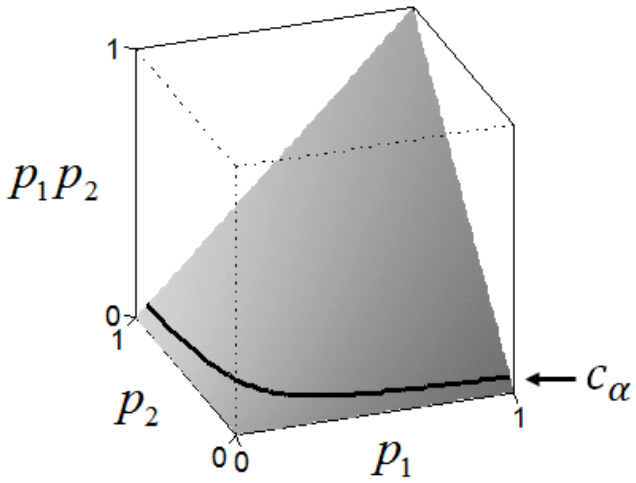
Reject if $p_1 p_2 \leq c_\alpha$

Methods for the confirmatory situation

Two trial stages

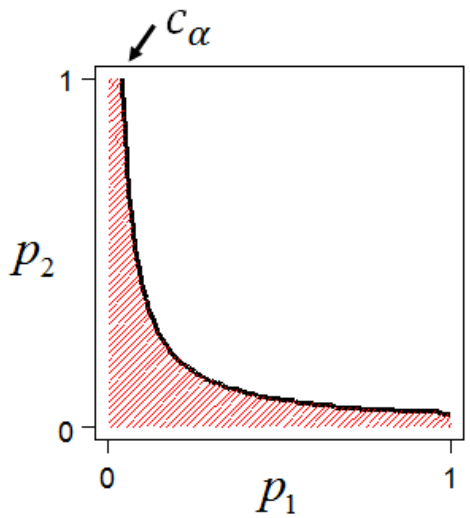
- Combine the p-values of two separate trial stages
  - E.g., in the Product Test, reject H_0 when $p_1 p_2 \leq c_\alpha$
  - With c_α chosen such that the Type I error is α (see next slide)

- p-value combination



Reject if $p_1 p_2 \leq c_\alpha$

- Projection onto the plane



Reject if $p_2 \leq c_\alpha / p_1$

NOVARTIS

Methods for the confirmatory situation

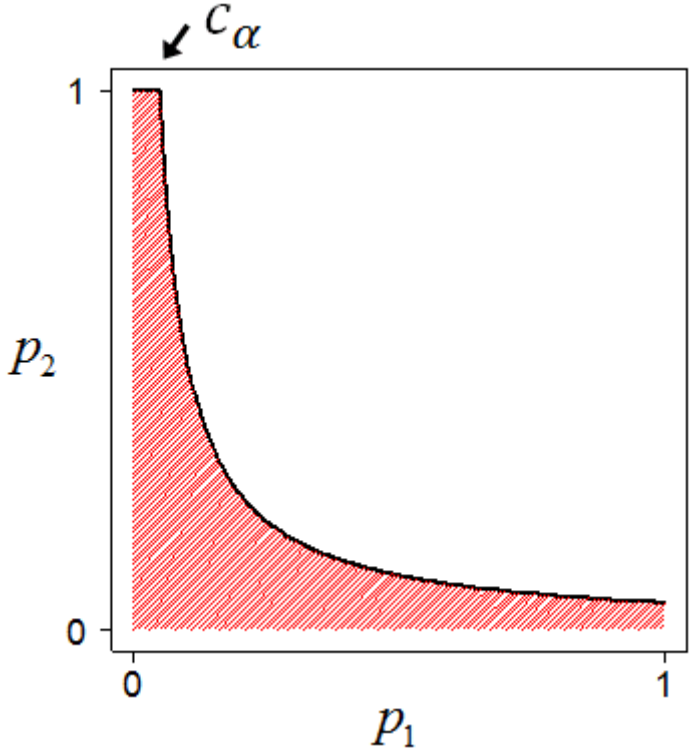
Two trial stages

- Derivation of c_α (Fisher 1932)
  - $p_1, p_2 \sim_{H_0} U[0,1]$ iid
  - $-2\ln(p_1), -2\ln(p_2) \sim_{H_0} \chi^2{}_2$ iid
  - $-2(\ln(p_1) + \ln(p_2)) \sim_{H_0} \chi^2{}_4$
  - Rejecting H_0 when $-2(\ln(p_1) + \ln(p_2)) \geq \chi^2{}_{4,1-\alpha}$ is a level α test
  - Equivalently, rejecting H_0 when $p_1 p_2 \leq c_\alpha = \exp\left(-\frac{1}{2}\chi^2{}_{4,1-\alpha}\right)$
- The area of the rejection region is α (the Type I error)
  - $\int_0^{c_\alpha} 1\, dp_1 + \int_{c_\alpha}^1 c_\alpha / p_1\, dp_1 = c_\alpha - c_\alpha \ln(c_\alpha) = \alpha$
  - This equation is called the level condition and can also be used to set c_α

Methods for the confirmatory situation

Two trial stages

- The conditional error function



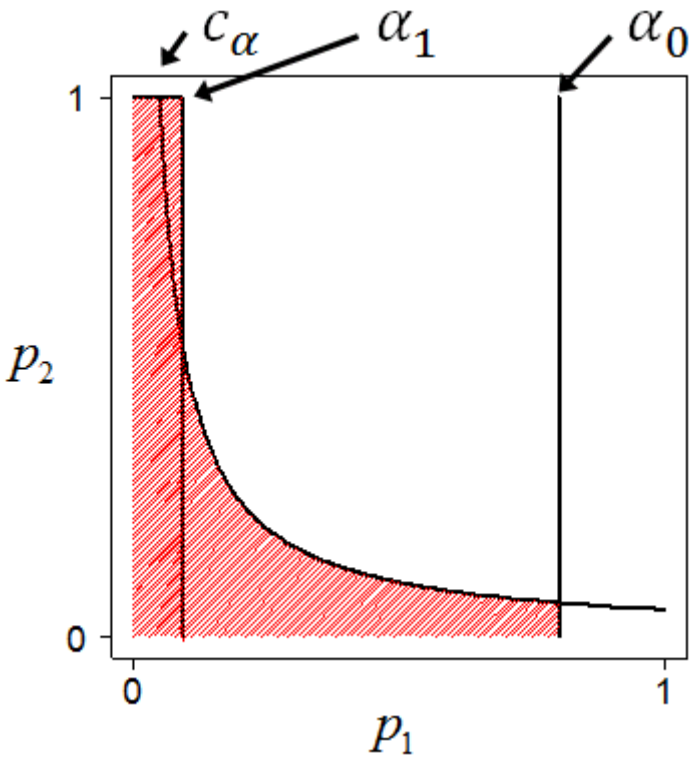
- Reject H_0 if $p_2 \leq c_\alpha\,/p_1$
  - If $p_1 \leq c_\alpha$, stop for efficacy
- The area of the red region is α
- The black curve is the level curve of the p-value combination function; it defines a level α test
- It is called conditional error function

NOVARTIS

Methods for the confirmatory situation

Two trial stages

- Choose stopping bounds



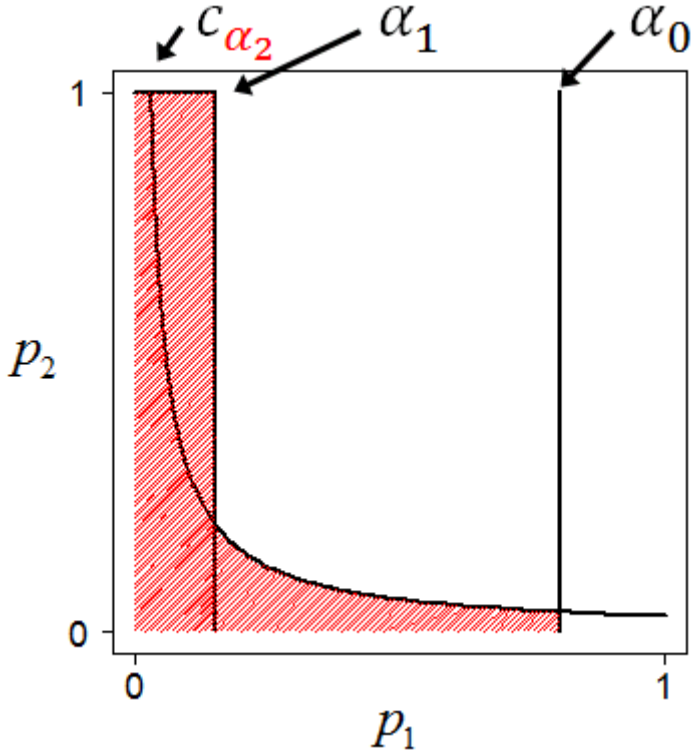
- Impose bounds α_1 and α_0
  - Red area must remain α
- Reject H_0 in the red area
  - If $p_1 \leq \alpha_1$, stop for efficacy
  - If $p_1 > \alpha_0$, stop for futility
  - Otherwise, reject H_0 if $p_2 \leq c_\alpha / p_1$
- α_1 is the local significance level of the test after the first stage

NOVARTIS

Methods for the confirmatory situation

Two trial stages

- Change height of curve

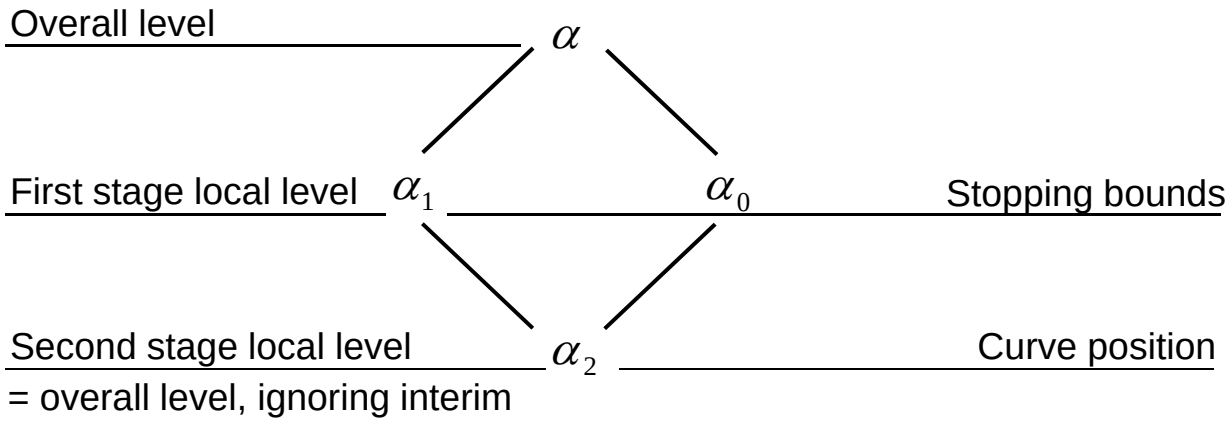


- Choose different level curve
  - Red area must remain α
- Reject H_0 in the red area
- α_2 is the local significance level of the test after the second stage

Methods for the confirmatory situation

Two trial stages

- The «alpha calculus»
  - Four parameters are interdependent



  - Level condition, general form for this test:

$$\alpha_1 + c_{\alpha_2}\left(\ln(\alpha_0) - \ln(\alpha_1)\right) = \alpha$$

Methods for the confirmatory situation

Two trial stages

- Another way to combine p-values:
  - $\frac{1}{\sqrt{2}}\left(\Phi^{-1}(1-p_1)+\Phi^{-1}(1-p_2)\right) \sim_{H_0} N(0,1)$
  - This is the Inverse Normal Method (Lehmacher, Wassmer '99)
- The same mechanism can be applied
  - Impose early stopping bounds α_1 and α_0
  - Choose a local second stage level α_2
  - Such that the overall test level remains α
- Now, the level condition takes this form:
  - $\alpha_1 + \int_{\alpha_1}^{\alpha_0} \underbrace{\Phi\left(\sqrt{2}\Phi^{-1}(\alpha_2) - \Phi^{-1}(p_1)\right)}_{\text{Conditional error function for the inverse normal method}} dp_1 = \alpha$

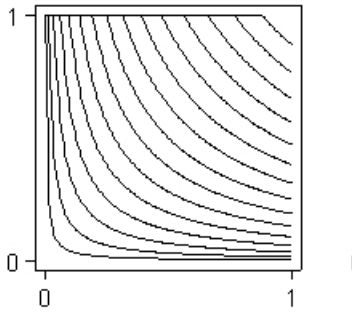
NOVARTIS

Here, Φ denotes the standard normal distribution function.

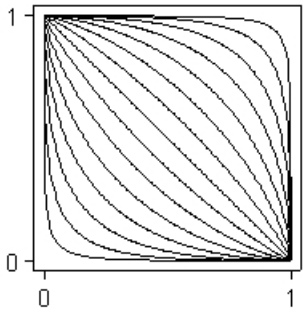
Methods for the confirmatory situation

Two trial stages

- It's exactly the same as with the product test, just using a different family of conditional error functions



Product test



Inverse normal method

- General correspondence (Vandemeulebroecke 2006)
  - Every p-value combination function defines a family of conditional error functions that fills the unit square, and vice versa
  - In principle, any such family of conditional error functions defines an adaptive test by this mechanism
  - In practice, mainly the two families above are used

NOVARTIS

Methods for the confirmatory situation

Adaptations

- How do trial adaptations fit into this?
  - The conditional error function provides the conditional Type I error probability given the first stage data
  - In theory (!), we could change *anything* after the interim analysis, as long as we don't increase the conditional error probability of the remainder of the trial («Conditional Rejection Principle» (CRP))
    - Simply use the p-value p_2 from this new remainder of the trial
    - Adaptations need not be pre-planned and can depend on unblinded interim data
      - One could even replace the second stage of a two-stage trial by another two-stage design, etc → multistage design by «recursive combination»
    - **Note**: The need to retain interpretability of results and credibility of the trial imposes strong constraints on adaptations to be viable in practice!
      - Adaptations should generally be pre-planned and limited to 1-2 trial features

Methods for the confirmatory situation

Applications and extensions

- Sources of multiplicity to control for
  1. Repeated hypothesis testing at IAs → group sequential methods
  2. Adaptations of trial design features → adaptive design methods
  3. Dealing with multiple hypotheses → multiple comparison methods
- Adaptive designs are most commonly used for prospectively planned...
  - ...sample size adaptation
  - ...treatment arm selection
  - ...subgroup enrichment
- Note that we often deal with several sources of multiplicity simultaneously!

Methods for the confirmatory situation

Applications and extensions: three common cases

- Sample size adaptation
  - Often based on conditional power: the power of the trial, given the interim data
  - Adjust the sample size of the remainder of the trial to achieve a desired conditional power in light of the interim results
  - Stop for futility if conditional power is too low
- Treatment arm selection
  - For individual hypotheses (treatment arms), combine p-values across trial stages
  - Then combine inference across treatment arms (often with the «closed testing principle» (Marcus et al. 1976), not covered here)
- Subgroup enrichment works similarly

Methods for the confirmatory situation

Summary and remarks

- The advent of confirmatory adaptive designs has injected great flexibility into clinical trial design, while protecting the Type I error rate
- The foundational methods rely on p-value combinations or (equivalently) conditional error functions
- The startling flexibility of the statistical theory has been harnessed by Health Authorities
  - Trial integrity must be protected and operational bias avoided
  - Adaptations should be limited and pre-planned (prior to unblinding)

Methods for the confirmatory situation

Summary and remarks

- Adaptive designs require careful planning; there are logistical, regulatory and statistical challenges
  - The benefits of implementing and adaptive design should be weighed against the challenges
- Major applications include sample size adaptation, treatment arm selection and subgroup enrichment

Outline

- Introduction
  - Brief historical overview
  - Example
  - Interim analyses
- Group sequential trials
  - Classical group sequential designs
  - Spending functions
  - Estimation following a group sequential trial
- Adaptive trials
  - Motivation and overview
  - Methods for the confirmatory situation
  - Estimation following an adaptive trial
- Case study and Final words

Estimation following an adaptive trial

- Naive maximum likelihood estimates are generally biased after an adaptive design
  - Because of early stopping (as in group sequential designs)
  - Because of adaptations
- For pre-specified adaptation rules, unbiased estimates can be constructed (but often are not)
- For ad-hoc adaptations, bias correction is not possible. Pragmatic solutions have been proposed (next slide).
- If naive estimates are reported, then mark clearly as «unadjusted»

Estimation following an adaptive trial

- Pragmatic estimates after ad-hoc adaptations often average stage-wise estimates, e.g.:
  - $\hat{\vartheta} = \tau\hat{\vartheta}_1 + (1-\tau)\hat{\vartheta}_2$ with pre-specified τ (e.g., $\tau = n_1/(n_1+n_2)$)
  - $\hat{\vartheta} = \tilde{\tau}\hat{\vartheta}_1 + (1-\tilde{\tau})\hat{\vartheta}_2$ with $\tilde{\tau} = \frac{w_1/se_1}{w_1/se_1 + w_2/se_2}$
    - Where se_k is the standard error of $\hat{\vartheta}_k$, and w_k are pre-specified weights such that ${w_1}^2 + {w_1}^2 = 1$
- Open field of research

Outline

- Introduction
  - Brief historical overview
  - Example
  - Interim analyses
- Group sequential trials
  - Classical group sequential designs
  - Spending functions
  - Estimation following a group sequential trial
- Adaptive trials
  - Motivation and overview
  - Methods for the confirmatory situation
  - Estimation following an adaptive trial
- Case study and Final words

Case study: HORIZON-III

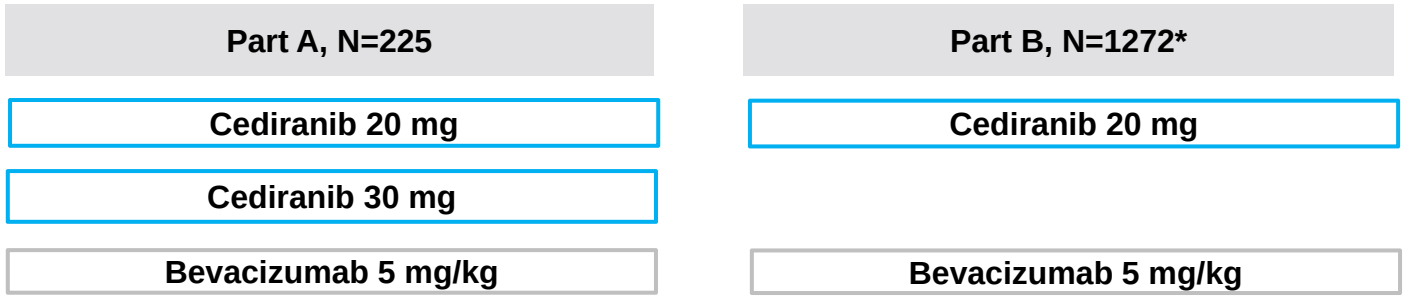
(Schmoll et al. 2012, Cuffe et al. 2014)

- Pivotal program
  - Cediranib under development for Colorectal Cancer (1st line). Promising phase I data and external evidence for Mode of Action argued for start of pivotal program
  - HORIZON-II: Conventional phase III, superiority vs. Placebo (1st line)
  - HORIZON-III: Seamless phase II/III, noninferiority vs. bevacizumab (Standard of Care) (1st line)
  - HORIZON-I: Phase II vs. bevacizumab (*2nd line, started earlier, faster event rate)*
  - All studies used the same 2 active doses, all plus chemotherapy, and Progression-free Survival (PFS) as primary endpoint

Case study: HORIZON-III

(Schmoll et al. 2012, Cuffe et al. 2014)

- HORIZON-III
  - Seamless phase II/III, noninferiority vs. bevacizumab (1st line)



  - At the interim analysis, DMC to select 1 dose based on interim analysis of HORIZON-III (sponsor blinded) *plus final HORIZON-I results (open)*, using predefined criteria for PFS and Relapse Rate
  - At the final analyis, test noninferiority of the selected dose vs. bevacizumab based on PFS, using both stages and controlling the overall Type I error

* Plus N=117 recruited to cediranib 30 mg during the interim analysis

NOVARTIS

Case study: HORIZON-III

(Schmoll et al. 2012, Cuffe et al. 2014)

- Results for cediranib 20 mg and 30 mg vs. bevacizumab
  - HORIZON-I
    - Hazard ratio (95% CI): 1.28 (0.85 - 1.95) and 1.17 (0.77 - 1.70)

  - HORIZON-III
    - 20 mg appeared good at the interim analysis and was selected for Part B

  - HORIZON-III
    - Hazard ratio at final analysis 1.10 (0.97 - **1.25**); **noninf margin: 1.20**
    - The effect had **shrunk**, compared to that observed at the IA

  - HORIZON-I revealed some tolerability issues with cediranib. In HORIZON-III, these were often wrongly charged to chemotherapy, leading to reduction of chemotherapy dose in the cediranib arms, particularly in centers that only participated in Part B. This can have contributed to the shrinking treatment effect.

Case study: HORIZON-III

(Schmoll et al. 2012, Cuffe et al. 2014)

- Lessons learned
  - A shrinking treatment effect from phase II to phase III is common even if all conditions (population etc.) remain equal: regression to the mean. But it is of particular concern in a «lean» development program with little data to corroborate interim findings.
  - Condensation of development program across Phases II and III did not leave time to learn about – or react to – tolerability issues
  - This is an example for the adaptive approach making a development program **less** flexible
    - Often risky to pre-specify Phase III before starting Phase II ...!

Final words

- The social event trial (Koch 2006)
  - «Let's come together, let's start, let's see and then adapt until significance»
- Better:
  - Group sequential trial: early stopping with control of Type I error
  - Exploratory adaptive trial: (many...)
  - Confirmatory adaptive trial:
    - Multistage study design that uses accumulating data to decide how to modify aspects of the study without undermining the validity and integrity of the trial
    - Preplanning, as much as possible, based on intended adaptations
    - Correct statistical inference; avoid operational bias
    - Use external monitoring boards (charter!); restrict access to interim results
    - Bears challenges and requires careful planning

Final words

Burman, Sonnesson (2006):

“Are flexible designs sound?”

Bauer answers:

“Is wine sound? It is the way to drink it that matters”

Many thanks for your attention

Questions?

Contact: marc.vandemeulebroecke@novartis.com

NOVARTIS

References (1)

(Ordered chronologically within each section)

Reference books

- Jennison, Turnbull (2000): Group sequential methods with applications to clinical trials. Chapman&Hall, Boca Raton
- Wassmer, Brannath (2016): Group Sequential and Confirmatory Adaptive Designs. Springer

Reviews and Tutorials

- Todd (2007): A 25-year review of sequential methodology in clinical studies. Statistics in Medicine 26, 237-252
- Vandemeulebroecke (2008): Group Sequential and Adaptive Designs – A Review of Basic Concepts and Points of Discussion. Biometrical Journal 50, 541-557
- Bretz, König, Brannath, Glimm, Posch (2009). Tutorial in biostatistics: Adaptive designs for confirmatory clinical trials. Statistics in Medicine 28, 1181-1217
- Bauer et al. (2016): Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. Statistics in Medicine 35, 325-347

Discussion & interpretation

- Dragalin (2006): Adaptive Designs: terminology and classification. Drug Information Journal 40, 425-435
- Koch (2006): Confirmatory clinical trials with an adaptive design. Biometrical Journal 48, 574-585. Rejoinder 616-622
- Burman, Sonesson (2006): Are flexible designs sound? (with discussion). Biometrics 62, 664-683
- Vandemeulebroecke (2008): *see above*

References (2)

(Ordered chronologically within each section)

Methodological contributions

- Fisher (1932): Statistical methods for research workers. Oliver & Boyd, London
- Efron (1971). Forcing a sequential experiment to be balanced. Biometrika 58, 403-417
- Marcus, Peritz, Gabriel (1976): On closed testing procedures with special reference to ordered analysis of variance. Biometrika 63, 655-660
- Pocock, Simon (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 31, 103-115
- Pocock (1977): Group sequential methods in the design and analysis of clinical trials. Biometrika 64, 191-199
- Wei, Durham (1978). The randomized play-the-winner rule in medical trials. JASA 73, 840-843
- O'Brien, Fleming (1979): A multiple testing procedure for clinical trials. Biometrics 35, 549-556
- Lan, DeMets (1983): Discrete sequential boundaries for clinical trials. Biometrika 70, 659-663
- Whitehead (1986): On the bias of maximum likelihood estimation following a sequential test. Biometrika 73, 573-581
- Kim, DeMets (1987): Design and analysis of group sequential tests based on the type I error spending rate function. Biometrika 74, 149-154
- Wang, Tsiatis (1987): Approximately optimal one-parameter boundaries for group sequential trials. Biometrics 43, 193-199
- Bauer (1989): Multistage testing with adaptive designs (with discussion). Biometrie und Informatik in Medizin und Biologie 20, 130-148
- O'Quigley, Pepe, Fisher (1990): Continual Reassessment Method: A Practical Design for Phase 1 Clinical Trials in Cancer. Biometrics 46, 33-48
- Bauer, Köhne (1994): Evaluation of experiments with adaptive interim analyses. Biometrics 50: 1029-1041 Correction in Biometrics 52 (1996): 380

References (3)

(Ordered chronologically within each section)

Methodological contributions (continued)

- Bauer, Röhmel (1995): An adaptive method for establishing a dose-response relationship. Statistics in Medicine 14, 1595-1607
- Proschan, Hunsberger (1995): Designed extension of studies based on conditional power. Biometrics 51, 1315-1324.
- Posch, Bauer (1999): Adaptive two stage designs and the conditional error function. Biometrical Journal 41, 689-696
- Wassmer (1999): Statistische Testverfahren für gruppensequentielle und adaptive Pläne in klinischen Studien. Verlag Alexander Mönch, Köln
- Lehmacher, Wassmer (1999): Adaptive sample size calculations in group sequential trials. Biometrics 55, 1286-1290
- Brannath et al. (2002): Recursive combination tests. JASA 97 (457), 236-244
- Müller, Schäfer (2004): A general statistical principle for changing a design any time during the course of a trial. Statistics in Medicine 23, 2497-2508
- Fan, DeMets, Lan (2004). Conditional bias of point estimates following a group sequential test. Journal of Biopharmaceutical Statistics 14 505-530
- Bauer, König (2006): The reassessment of trial perspectives from interim data – a critical view. Statistics in Medicine 25, 23-36
- Vandemeulebroecke (2006): An investigation of two-stage tests. Statistica Sinica 16, 933-951
- Freidlin, Korn (2009). Stopping clinical trials early for benefit: impact on estimation. Clinical Trials 6, 119-125
- Glimm, Maurer, Bretz (2010). Hierarchical testing of multiple endpoints in group-sequential trials. Statistics in Medicine 29, 219-228
- Wang, Rosner, Goodman (2016). Quantifying over-estimation in early stopped clinical trials and the "freezing effect" on subsequent research. Clinical Trials 13, 621-631
- Shimura, Gosho, Hirakawa (2017). Comparison of conditional bias-adjusted estimators for interim analysis in clinical trials with survival data. Statistics in Medicine 36, 2067-2080

References (4)

(Ordered chronologically within each section)

Regulatory guidance

- FDA (2006): Establishment and Operation of Clinical Trial Data Monitoring Committees. Rockville, MD
- EMA (2007): Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. European Medicines Agency, CHMP/EWP/2459/02
- FDA (2010): Draft guidance for industry. Adaptive design clinical trials for drugs and biologics. FDA CDER/CBER, Rockville, MD
- Brannath et al. (2010): Comments on the Draft Guidance on “Adaptive Design Clinical Trials for Drugs and Biologics” of the U.S. Food and Drug Administration. Journal of Biopharmaceutical Statistics 20, 1125-1131
- FDA (2015): Draft guidance for industry and food and drug administration staff. Adaptive designs for medical device clinical studies. FDA CDRH/CBER, Rockville, MD
- FDA (2018): Adaptive Designs for Clinical Trials of Drugs and Biologics. Draft Guidance. FDA CDER/CBER, Rockville, MD

Clinical trial examples

- Vandemeulebroecke, Bornkamp, Bretz, Pinheiro (2010): Adaptive dose-ranging studies. Chapter 11 in Pong and Chow (eds.): Handbook of Adaptive Designs for Pharmaceutical and Clinical Development. Chapman & Hall
- Barnes et al. (2010): Integrating indacaterol dose selection in a clinical study in COPD using an adaptive seamless design. Pulmonary Pharmacology & Therapeutics 23: 165-171
- Schmoll et al. (2012): Cediranib with mFOLFOX6 versus bevacizumab with mFOLFOX6 as first-line treatment for patients with advanced colorectal cancer: A double-blind, randomized phase III study (HORIZON III). Journal of Clinical Oncology 30, 3588-3595
- Cuffe, Lawrence, Stone, Vandemeulebroecke (2014): When is a seamless study desirable? Case studies from different pharmaceutical sponsors. Pharmaceutical Statistics 13, 229-237
- Marcus et al. (2017): Obinutuzumab for the First-Line Treatment of Follicular Lymphoma. NEJM 377, 1331-1344

References (5)

Software reviews

- Wassmer, Vandemeulebroecke (2006): A brief review on software developments for group sequential and adaptive designs. Biometrical Journal 48: 732-737
- Tymofyeyev (2014): A Review of Available Software and Capabilities for Adaptive Designs. Chapter in: Practical Considerations for Adaptive Trial Design and Implementation. Springer

Commercial software

- ADDPLAN: http://www.iconplc.com/innovation/addplan/
- EastAdapt and EastSurv: http://www.cytel.com/software/east

R packages

- groupSeq: https://cran.r-project.org/web/packages/GroupSeq/index.html
- gsDesign: https://cran.r-project.org/web/packages/gsDesign/index.html
- ldbounds: https://cran.r-project.org/web/packages/ldbounds/index.html
- seqmon: https://cran.r-project.org/web/packages/seqmon/index.html
- adaptTest: https://cran.r-project.org/web/packages/adaptTest/index.html
- AGSDest: https://cran.r-project.org/web/packages/AGSDest/index.html
- asd: https://cran.r-project.org/web/packages/asd/index.html
- RPACT: https://www.rpact.com/