AptivSolutions℠
Accelerating the Possibilities

Adaptive Designs in Drug Development

Professor Andy Grieve
SVP Clinical Trial Methodology,
Aptiv Solutions, Cologne (Andy.Grieve@aptivsolutions.com)

PSDM | OSS, NL | November 8, 2012

Outline

- Introduction and taxonomy of clinical trial designs
  - Pre-1990's
- Basic principles of adaptive designs
  - Allocation rule
  - Sampling rule
  - Stopping rule
  - Decision rule
- Phases of development
- Adaptive designs for the learn phase of drug development
  - First-in human / MTD
  - Two-stage designs
  - Adaptive dose-ranging designs
  - Bayesian adaptive randomisation
- Adaptive designs for the confirmatory phase of drug development
  - Sample size re-assessment
  - Adaptive group sequential designs
  - Seamless phase II/III designs
  - Population enrichment designs
- Practical aspects of adaptive design implementation
- Discussion

Outline

- Introduction and taxonomy of clinical trial designs
  - Pre-1990's
- Basic principles of adaptive designs
  - Allocation rule
  - Sampling rule
  - Stopping rule
  - Decision rule
- Phases of development
- Adaptive designs for the learn phase of drug development
  - First-in human / MTD
  - Two-stage designs
  - Adaptive dose-ranging designs
  - Bayesian adaptive randomisation
- Adaptive designs for the confirmatory phase of drug development
  - Sample size re-assessment
  - Adaptive group sequential designs
  - Seamless phase II/III designs
  - Population enrichment designs
- Practical aspects of adaptive design implementation
- Discussion

Palmer – Classification of designs

Stat. Meth. in Med. Res., 2002

- Parallel Group, Fixed Sample Size
  - Eg Bradford-Hill : streptomycin & treating pulmonary tuberculosis (*Br. Med. J*, 1948)
- Data-Dependendent Designs (DDD)
  - Sequential (Abraham Wald , 1940's)
  - Group sequential (Armitage et al, 1960's)
  - Adaptive Interim Designs (Bauer et al, 1990's)
  - Response-adaptive designs
  - Bayesian decision theoretic designs

Issues in Clinical Trials - Palmer

- Many trials struggle to recruit the required # of patients
- Informed consent needs to be improved
- Inadequate consent processes -> increasing # law suits
- Patient advocacy groups more involved
- Patient/doctor relationships act against participation
- Drug development times too long / costs too high
- Future trials will need to detect smaller effects
- Ethics committees less able to monitor ongoing trials
- Recent technological / methodological advances under utilised
- Investigators might not be willing to randomise relatives (Uncle Test)
- Some trials driven by round numbers (α=0.05,1-β=0.2, δ, n etc)

Adaptive designs - Pre 1990s

- An Old Idea Resurrected (1930s)
- Up-and-Down Designs (1940s)
- Play-The-Winner Designs (1960s)
- Randomized Play-The-Winner Designs (1970s)

The Earliest Adaptive Idea

William R Thompson
Biometrika (1933)

ON THE LIKELIHOOD THAT ONE UNKNOWN PROBABILITY EXCEEDS ANOTHER IN VIEW OF THE EVIDENCE OF TWO SAMPLES.

By WILLIAM R. THOMPSON. From the Department of Pathology, Yale University.

Thus, if, in this sense, P is the probability estimate that one *treatment* of a certain class of individuals is *better* than a second, as judged by data at present available, then we might take some monotone increasing function of P, say $f_{(P)}$, to fix the fraction of such individuals to be treated in the *first manner*, until more evidence may be utilised, where $0 \leqslant f_{(P)} \leqslant 1$; the remaining fraction of such individuals $(1 - f_{(P)})$ to be treated in the *second manner*; or we may establish a probability of treatment by the two methods of $f_{(P)}$ and $1 - f_{(P)}$, respectively. If

- Ethical Design – concentrating on delivering the best treatment to the most patients

Simple Idea

- Suppose at some point in a trial we have data following kind:

Treatment A	Treatment B
1	0
0	0
0	1
1	0
.	.
.	.
1	0
r_A / n_A	r_B / n_B

The Evidence in Favour of Treatment B

- If π_A and π_B are the response rates of each treatment then

$$prob(\pi_A < \pi_B | Data)$$

measures the "superiority" of B over A.

- Thompson suggested that patients be randomised to treatments A and B in the ratio

$$\frac{1 - prob(\pi_A < \pi_B | Data)}{prob(\pi_A < \pi_B | Data)}$$

2 x 2 Contingency Table Data Structure

	Response	No Response
Treatment A	$r_1\ (\pi_A)$	$n_1\text{-}r_1(1\text{-}\pi_A)$
Treatment B	$r_2\ (\pi_B)$	$n_2\text{-}r_2\ (1\text{-}\pi_B)$

Likelihood $$\propto \pi_A^{r_1}\left(1-\pi_{1A}\right)^{n_1-r_1}\pi_B^{r_2}\left(1-\pi_B\right)^{n_2-r_2}$$

Prior $$\propto \pi_A^{\alpha_1-1}\left(1-\pi_A\right)^{\beta_1-1}\pi_B^{\alpha_2-1}\left(1-\pi_B\right)^{\beta_2-1}$$

Posterior $$\propto \pi_A^{r_1+\alpha_1-1}\left(1-\pi_A\right)^{n_1-r_1+\beta_1-1}\pi_B^{r_2+\alpha_2-1}\left(1-\pi_B\right)^{n_2-r_2+\beta_2-1}$$

2x2 Contingency Table - Posterior Inference "Uninformative Priors" : $\alpha_A = \beta_A = \alpha_B = \beta_B = 1$

- The probability of interest is

$$\text{Prob}(\pi_A < \pi_B \mid \text{Data}) = \sum_{k=0}^{n_1 - r_1} \frac{\binom{n_1 + n_2 - r_1 - r_2 - k}{n_2 - r_2}\binom{r_1 + r_2 + 1 + k}{r_2}}{\binom{n_1 + n_2 + 1}{n_1 + 1}}$$

based on the cumulative hypergeometric function as is Fisher's exact test

(Altham JRSSB1969; Raiffa & Schlaifer, Applied Statistical Decision Theory, 1960))

Thompson's Practical Interpretation

- Thompson(1935) proved the identity:

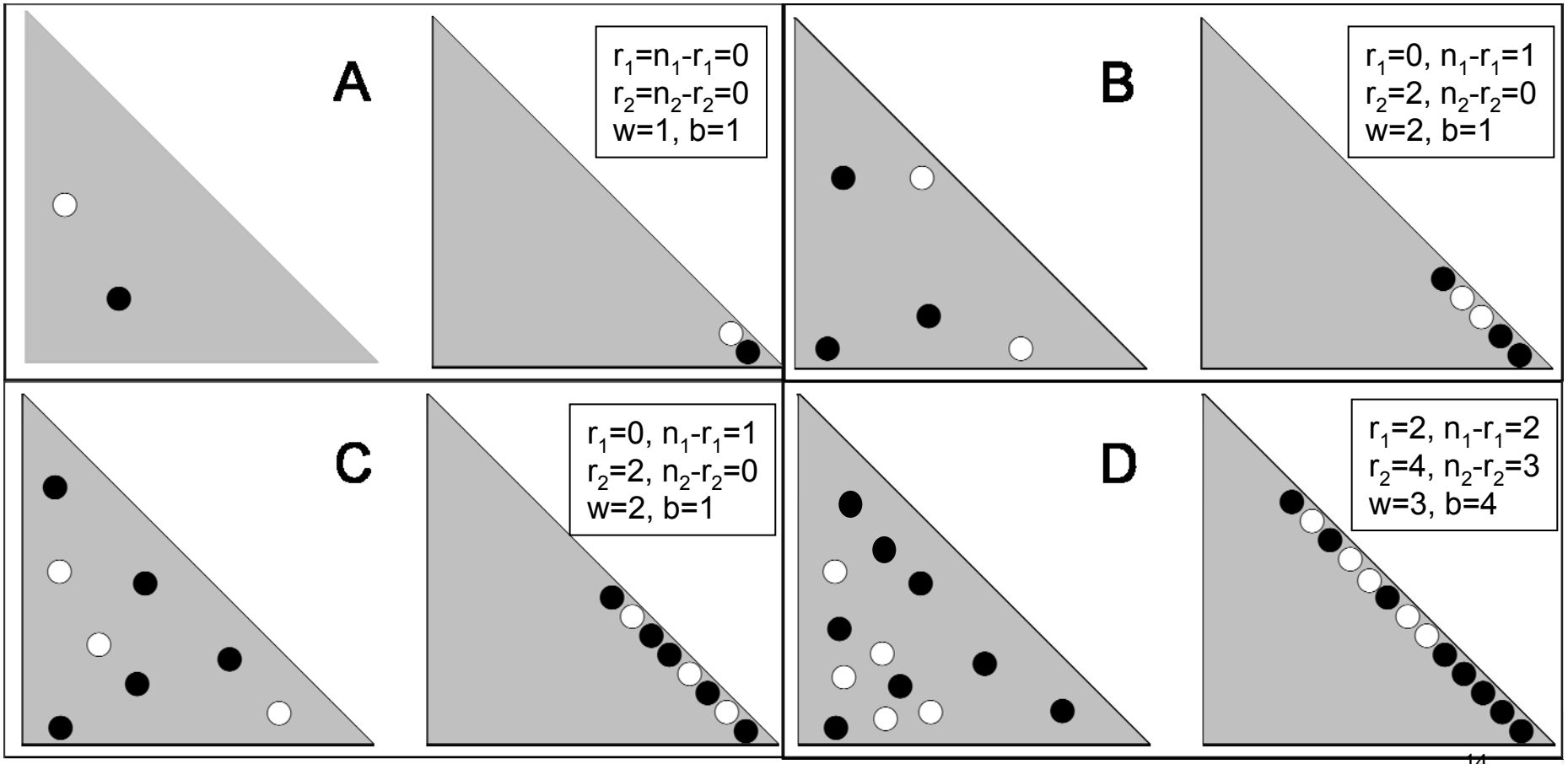
$$\sum_{k=0}^{n_1-r_1} \frac{\binom{n_1+n_2-r_1-r_2-k}{n_2-r_2}\binom{r_1+r_2+1+k}{r_2}}{\binom{n_1+n_2+1}{n_1+1}} = \sum_{k=0}^{\min(b-1,W-w)} \frac{\binom{W}{w+\alpha}\binom{B}{b-1-\alpha}}{\binom{W+B}{w+b-1}}$$

where: $W=n_1+1$, $B=n_2+1$, $w=n_1-r_1$ and $b=n_2-r_2$

- This second term is the probability under sampling without replacement from a mixture of W white balls and B black balls that we will get w white balls before b black balls

Thompson(1935) Mechanical Randomisation & Simulation

- For $W=n_1+1$, $B=n_2+1$: choose A if $w=n_1-r_1+1$ white balls occur before $b=n_2-r_2+1$ black balls

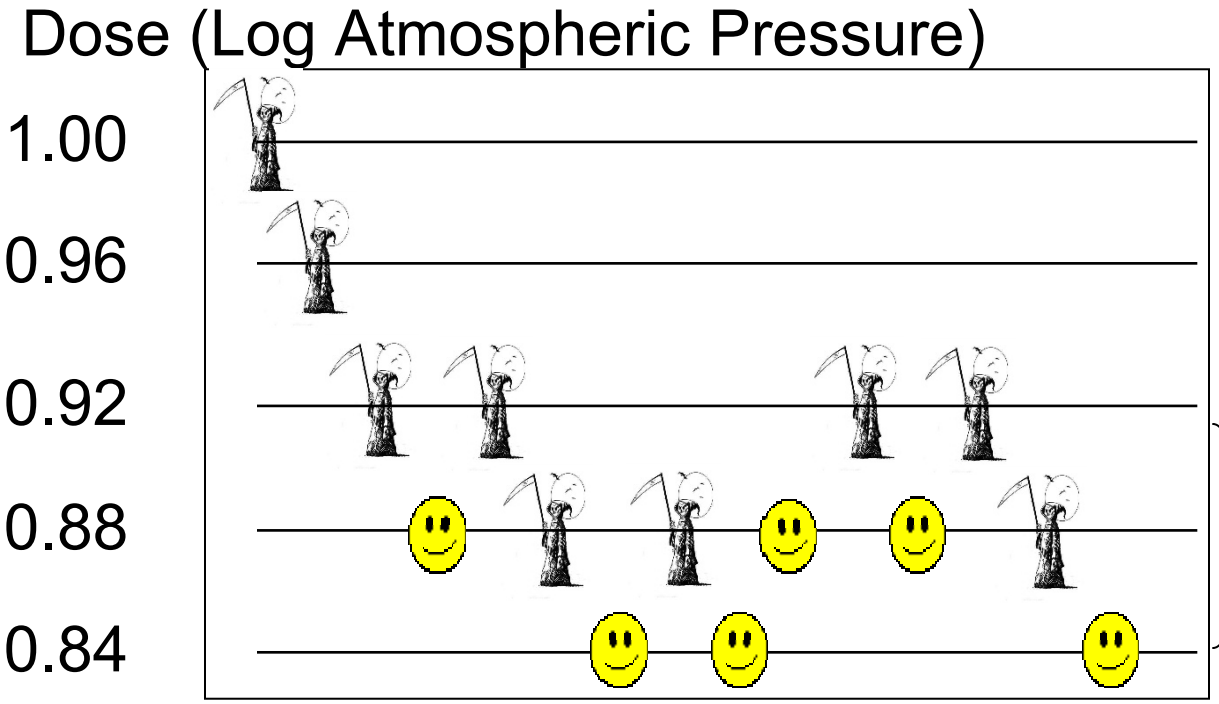


Up-And-Down Designs

An Old Design problem

- Non-linear response function
  - Optimal design available if we know the function
  - We don't know the function
- Solution :
  - Do some experiments
  - Learn a bit
  - Optimise
  - Learn a bit more
  - Optimise
  - etc
- Up-and-Down Design
  - Allocates patients to dosing groups (usually unequally)
  - Dose finding process
  - Nth patient gets allocated to dose depending on response of (N-1)th patient
    - Success : lower dose
    - Failure : increase dose

Up-and-Down Design
Impact of Mechanical Head Trauma
Choi (Biometrics, 1990)



Random Walk
LD_{50}=E(limiting distribution)

LD_{50}
?

1) Start at d_0
2) If rat dies at d_i : $d_{i+1}=(d_i-\Delta)$
   else $d_{i+1}=(d_i+\Delta)$

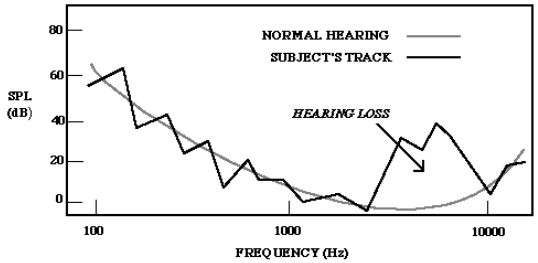
- Died

- Survived

Efficiency depends on starting d_0 and Δ

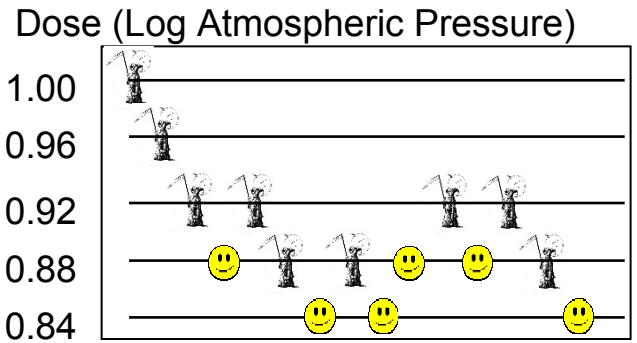
Up-And-Down Experiment

- Up-and-down (staircase) was developed by Dixon & Mood (J. Amer. Statist. Assoc.,1948) to estimate the ED_{50} (γ=0.5)



Aside - similar ideas were developed in sensitivity testing, psychophysics predating Dixon and Mood – Georg v Bekesy (Acta Oto-laryngol.1947).

- Estimate : Determine the number of doses giving successes and the number giving failures. Take the smaller total. Take the average dose for the smaller total - A. Then the estimated ED_{50} =A+Δ/2 for successes and ED_{50} =A-Δ/2 for failures. (se can also be determined)



9 failures , 6 successes

A=0.86

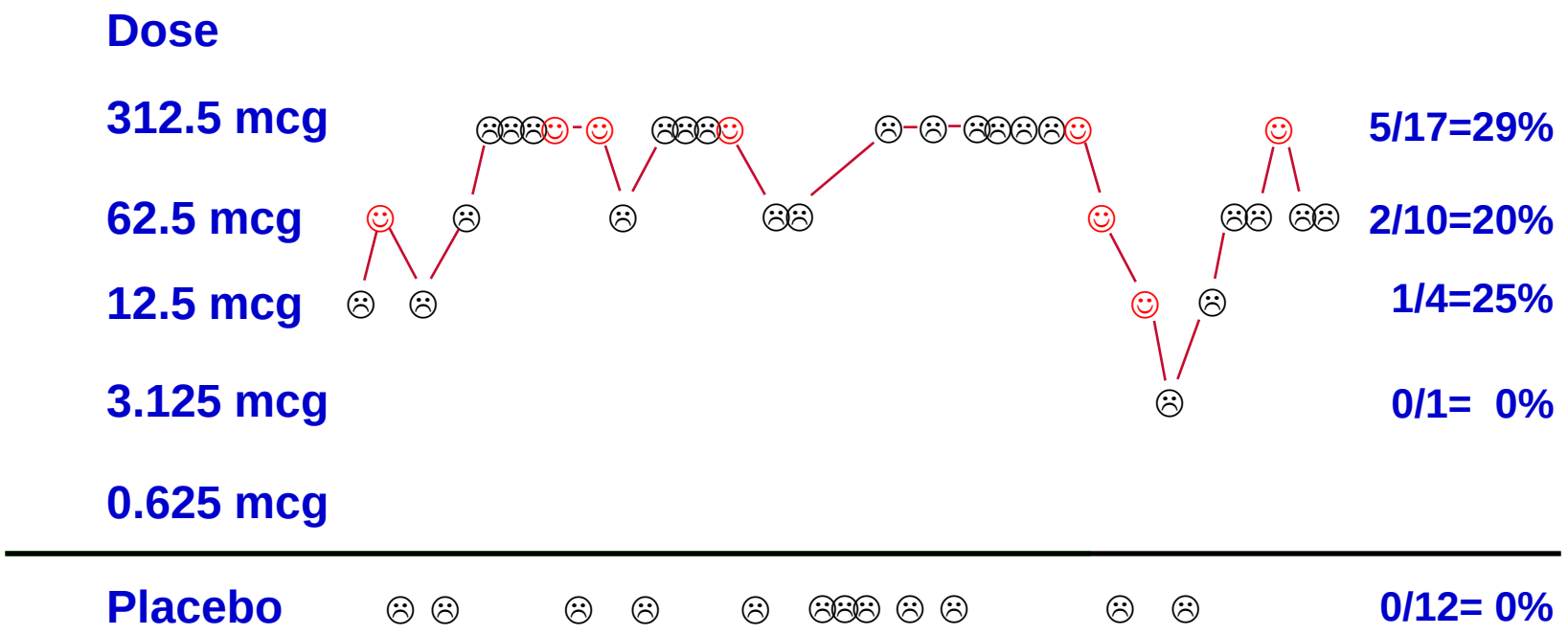
Ed50= 0.86+0.04/2=0.88

Up-and-Down Design
Background : Hoon et al (Ann. Neurol. ,2000)

- New compound - anti migraine
- Activity from 0.5 mcg
- Different mode of action from 500 mcg - more like elitriptan/sumitriptan
- Dose range is therefore 0.5 mcg - 500 mcg
- Need to reduce this range before conducting a dose response study
- Window of opportunity
- Placebo, 0.625, 3.125, 12.5, 62.5, 312.5 mcg - limited number because of dose form - intravenous : syringe sizes

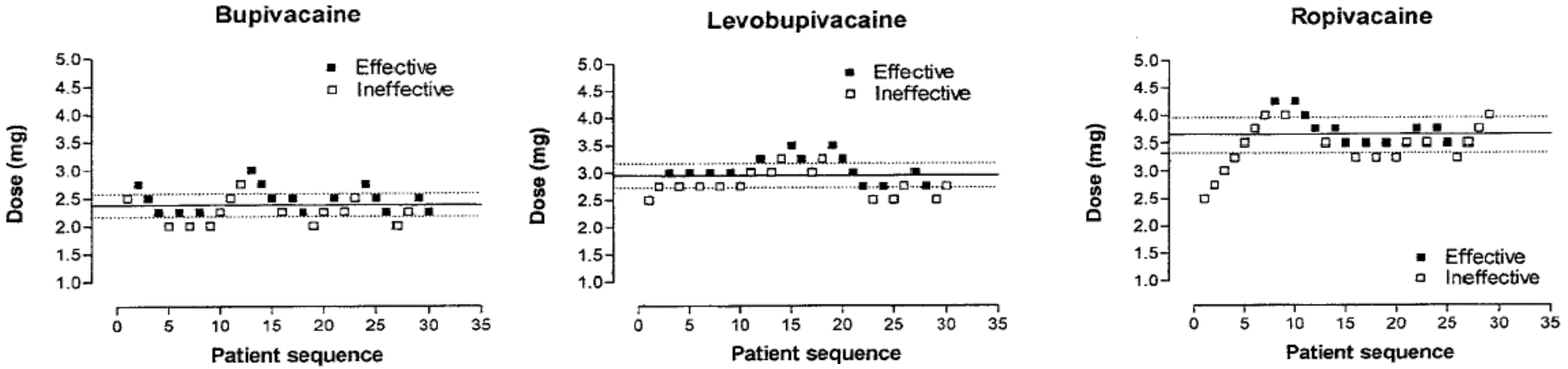
- What is dose at which 50% of patients respond ? Seen as 20% > than placebo rate (30%)
- Response :
  - Change within 2 hours from severe or moderate headache to mild or no headache - Glaxo definition.
- Need enough patients around optimum dose to have confidence in estimate
- May not achieve this with standard parallel group (equal n) design

Up-and-Down Design
Result from Hoon et al



Examples of Up-and-Down Designs Many in Anaesthetics

- Camorcia et al(Anesthesiology, 2004)
  - Ropivacaine, Levobupivacaine & bupivacaine in intrathecal labor



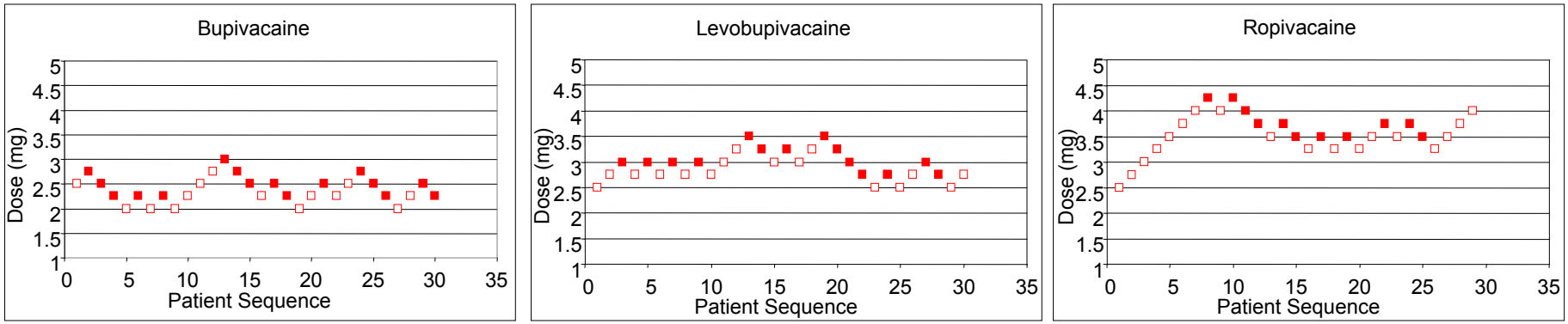
Analgesic Potency Ratio	Dixon and Massey Method	Probit Regression	*P* Value
Bupivacaine: levobupivacaine	0.81 (0.69–0.94)	0.79 (0.70–0.88)	< 0.01
Bupivacaine: ropivacaine	0.65 (0.56–0.76)	0.62 (0.55–0.69)	< 0.001
Levobupivacaine: ropivacaine	0.80 (0.70–0.92)	0.79 (0.70–0.88)	< 0.01

One-way analysis of variance with Bonferroni correction.

?

Example of Up-and-Down Design Camorcia et al(Anesthesiology, 2005)

- Ropivacaine, Levobupivacaine & bupivacaine in intrathecal labor



Drug	ED50 (mg) (95% CI)
Bupivicaine	2.37 (2.17-2.58)
Levobupivicaine	2.94 (2.73-3.16)
Ropivicaine	3.64 (3.33-3.96)

Play-The-Winner Design

Play-the the-Winner Rule
Zelen (J Am Statis Ass, 1969)

- Treatment assignment depends on the outcome of previous patients - Response adaptive assignment
- When response is determined quickly
- 1st subject: toss a coin, H = Trt A, T = Trt B
- For subsequent patients
  - assign previous treatment if it was successful
  - Otherwise, switch treatment assignment

- Advantage: Potentially more patients receive the better treatment
- Disadvantage: Investigator knows the next assignment

TRT A : S S F S S S F

TRT B : S F

Patient 1 2 3 4 5 6 7 8 9

.....

- Analysis based on sequence lengths

An Example of a PTW Study

Scand J Clin Lab Invest 1998; 58: 241–250

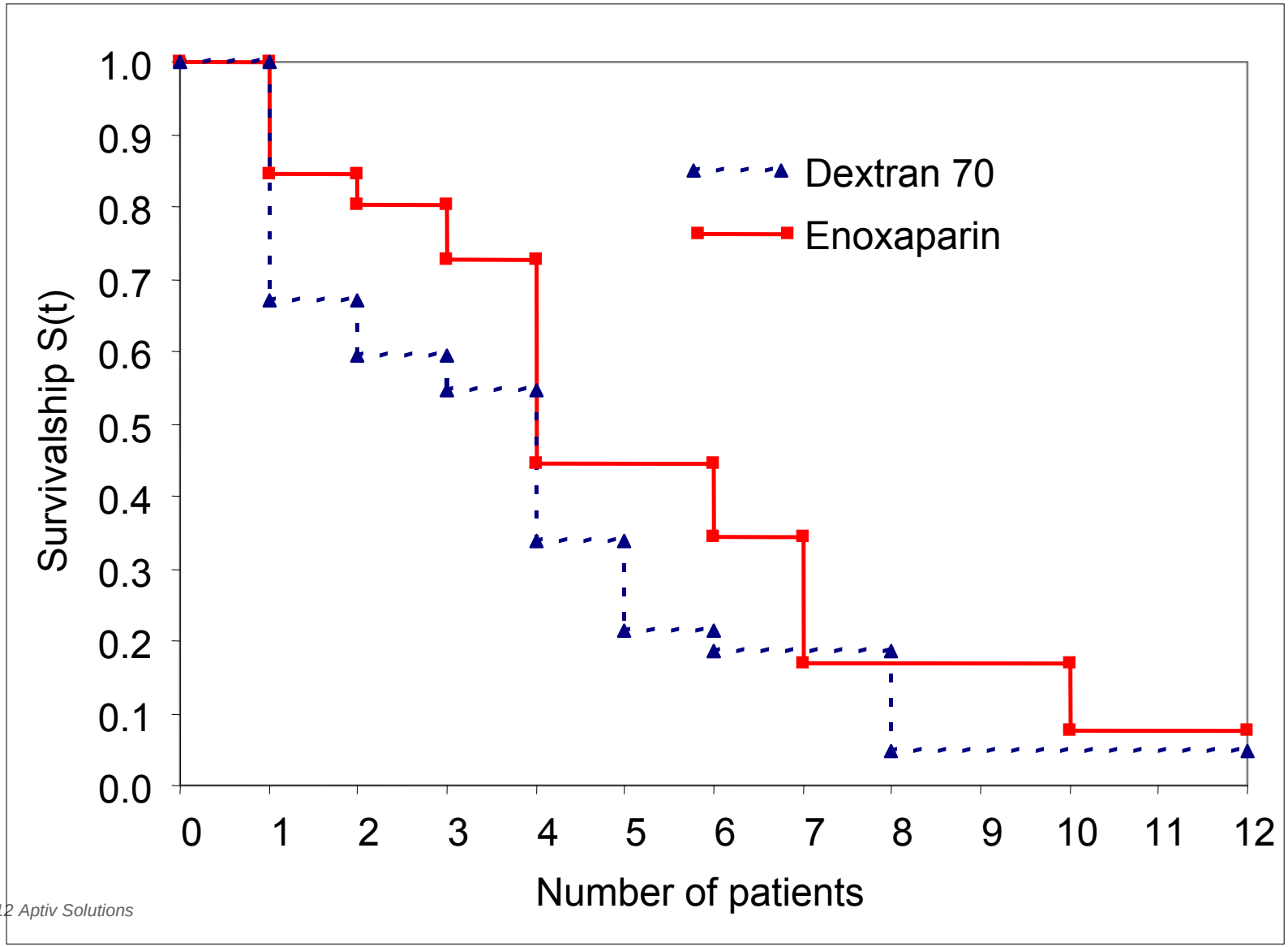
Development of play-the-winner design: weighting and accumulation of information

O. REIERTSEN,* S. LARSEN† & J. H. SOLHAUG‡

*Department of Surgery, Akershus Central Hospital, Nordbyhagen, †Medstat Research, Centre for Design, Administration and Statistical Analysis, Lillestrøm, ‡Department of Surgery, Diakonhjemmet Hospital, Oslo, Norway

- comparison of enoxaparin and dextran 70 for the prevention of venous thrombo-embolism following digestive surgery.
- modified version of the basic PTW design
  - following 15 consecutive successes a change of treatment automatically took place. The treatment sequences were regarded as non-ended and handled as censored.
- In total 231 patients were included in a PTW design.
  - The design allocated 140 patients to enoxaparin and 91 to dextran-70.
  - A survival analysis detected a significant difference {p<0.05) in favour of enoxaparin.

Kaplan-Meier Estimates of Sequence Lengths



Randomised Play-The-Winner Design

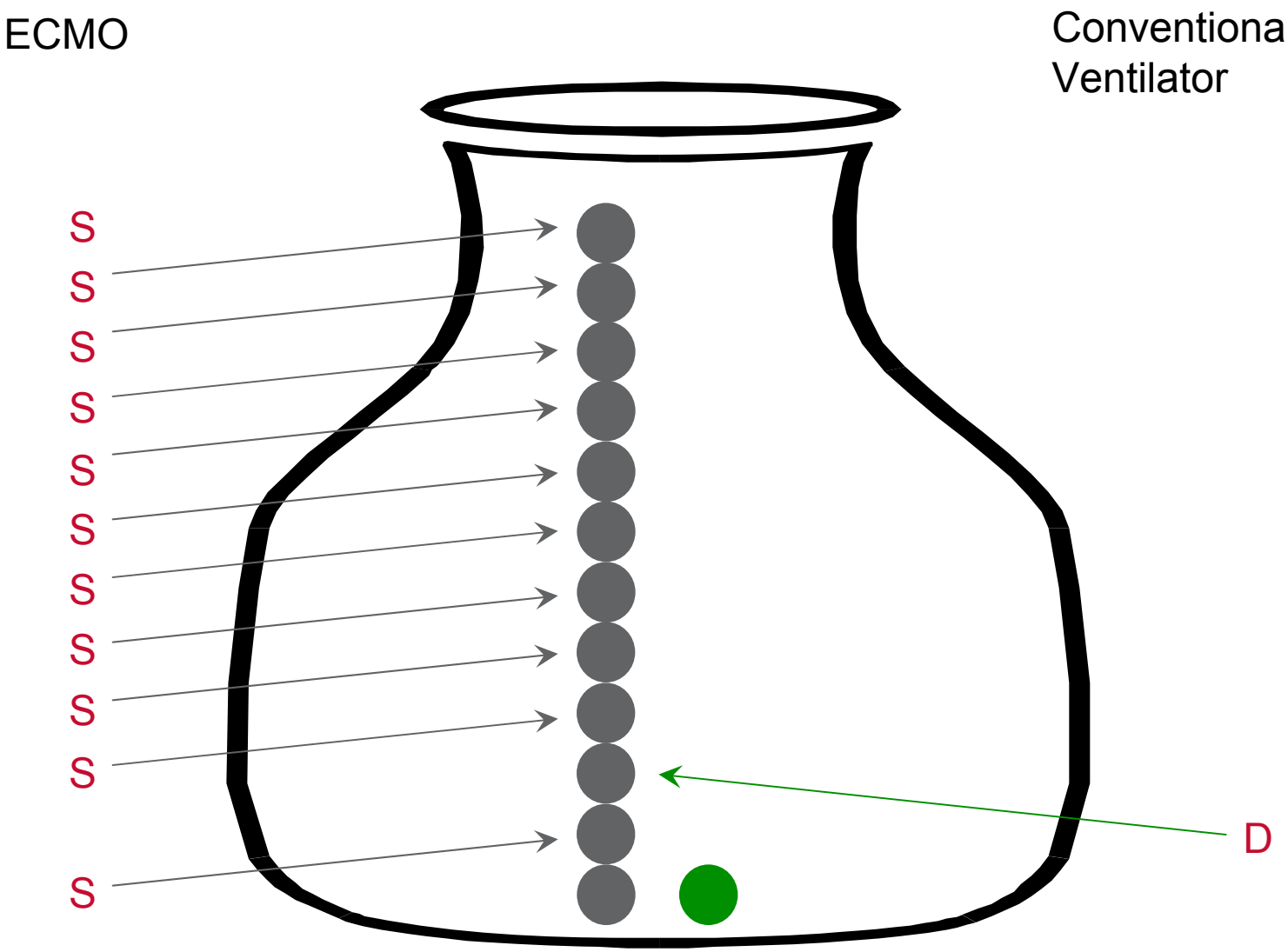
Randomise Play the Winner (RPW) Design Wei LJ, Durham SD (J. Am Statis Ass, 1978

- Urn model
- At beginning of trial
  - Urn contains α balls of each of two colours (W&R) representing 2 treatments
  - When a patient is to be treated a ball is chosen at random (with replacement)
  - When the response is known the urn content is updated as follows:
- If the patient was allocated to treatment t and responds
- positively, β balls of colour t are added to the urn otherwise γ of colour s (the complement of t) are added.
- In time the urn will contain a higher proportion of colored balls associated with the more successful treatment
- RPW(α,β,γ) design

Extra Corporeal Membrane Oxygenation Bartlett et al (Paediatrics, 1985)

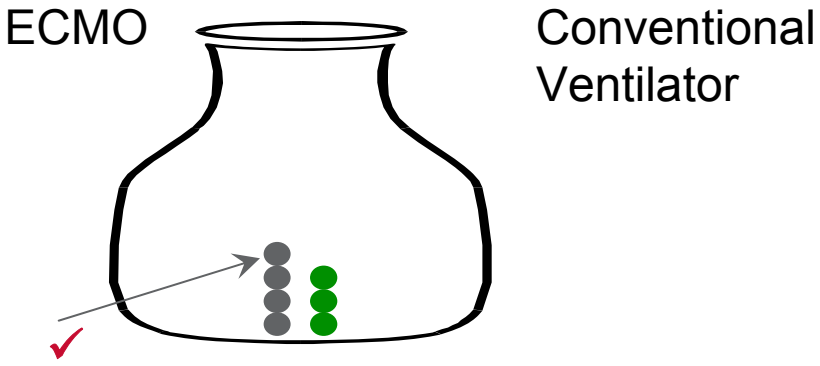
- Neonates with severe respiratory failure - Mortality
- ECMO vs Traditional Ventilator
- Phase I trials >50% survival on ECMO
- Optimal Therapy : survival < 20 %
- Chose Randomised Play-the-Winner (RPW)
  - speedy outcome - anticipated response diff -> small sample size - scientific/ethical dilemma

Results from Bartlett et al (Paediatrics, 1985)



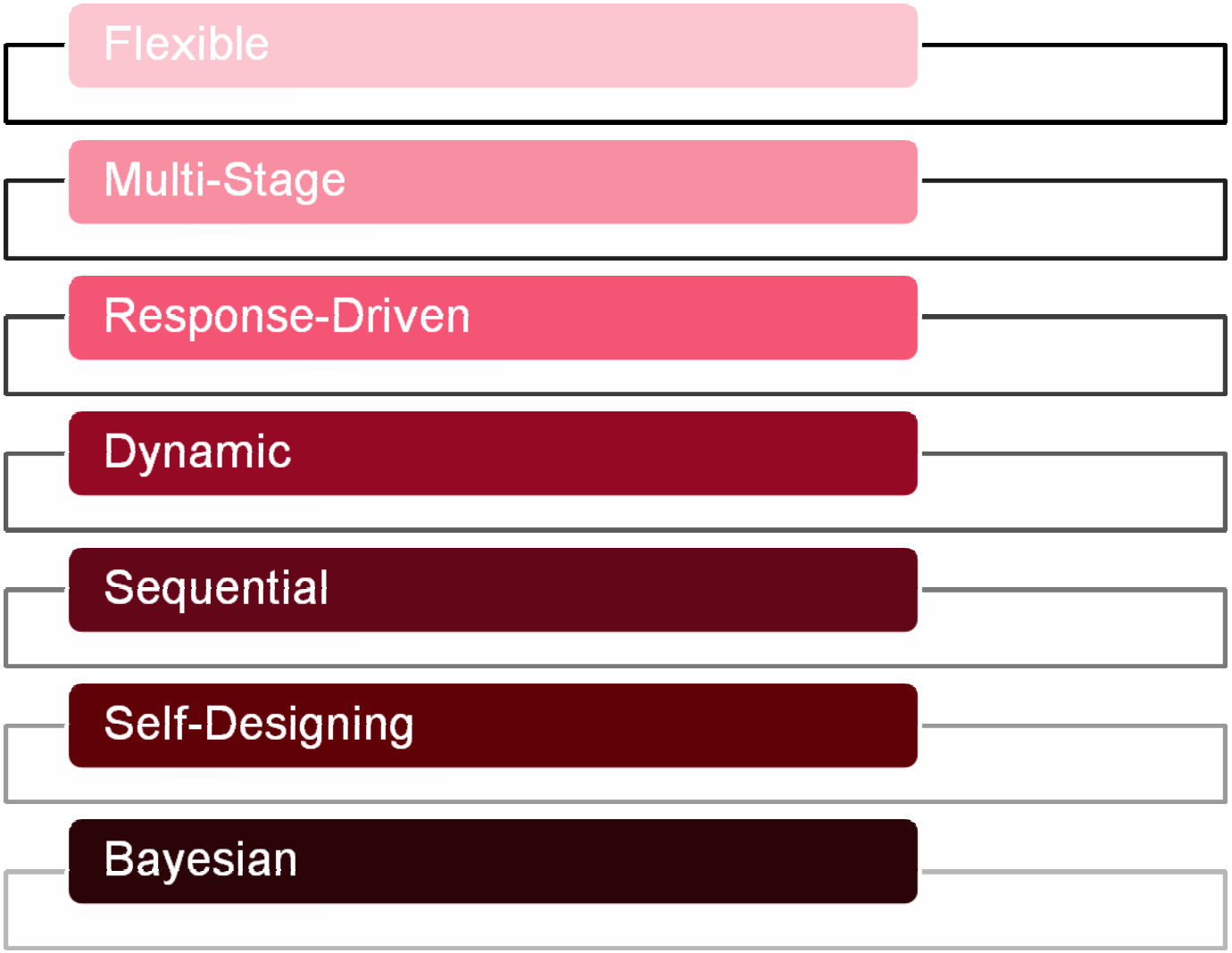
Randomised Play-the-Winner Urn Model (ECMO) : Issues

- Was the urn model sensible ?
  - Other parameters
  - Begin with randomised block
- How reliable are the results - 11/11 vs 0/1 ?
  - Ranking and selection procedure
  - Minimum number of patients



- Ethics
- Tamura et al (J Am Statist Ass,1994)

Terminology



What?

- ***Adaptive Design*** is one that uses accumulating data from the ongoing trial to modify aspects of the study without undermining the validity and integrity of the trial

  PhRMA ADWG, Gallo et al (2006)

- ***Adaptive design clinical study*** is defined as 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

  FDA Guidance on AD (2010)

- Validity
  - providing correct statistical inference: adjusted p-values, estimates, confidence intervals
  - providing convincing results to a broader scientific community
  - minimizing statistical bias
- Integrity
  - preplanning based on intended adaptations
  - maintaining confidentiality of data
  - assuring consistency between different stages of the study
  - minimizing operational bias

Adaptive by Design

Aspects of the Study to be Modified

- Number of Subjects
- Study Duration
- Endpoint Selection
- Treatment Duration
- Patient Population
- Number of Treatments
- Number of Interim Analyses
- Hypotheses

Outline

- Introduction and taxonomy of clinical trial designs
  - Pre-1990's
- Basic principles of adaptive designs
  - Allocation rule
  - Sampling rule
  - Stopping rule
  - Decision rule
- Phases of development
- Adaptive designs for the learn phase of drug development
  - First-in human / MTD
  - Two-stage designs
  - Adaptive dose-ranging designs
  - Bayesian adaptive randomisation
- Adaptive designs for the confirmatory phase of drug development
  - Sample size re-assessment
  - Adaptive group sequential designs
  - Seamless phase II/III designs
  - Population enrichment designs
- Practical aspects of adaptive design implementation
- Discussion

Fixed Sample vs Planned Adaptive Designs
Dragalin (Drug Information J, 2006)

4 Rules define an adaptive design –

	Planned Adaptive Designs	Fixed Sample
● Allocation Rule	– Defines how pts are allocated to arms. Can be fixed but can change based on accruing data	– Randomisation remains fixed throughout study
● Sampling Rule	– How many subjects sampled at next stage (cohort size)	– Only one stage
● Stopping rule	– When to stop a trial: efficacy, futility	– No stopping
● Decision Rule	– Final analysis or interim changes not covered by the above 3 (eg dropping arms)	– No Changes

Examples

- Group Sequential Designs: only **Stopping Rule**
- Response Adaptive Allocation: only **Allocation Rule**
- Sample Size Re-assessment: only **Sampling Rule**
- Flexible Designs:
  - Adaptive AR: changing the randomization ratio
  - Adaptive SaR: the timing of the next IA
  - Stopping Rule
  - Adaptive DR: changing the target treatment difference; changing the primary endpoint; varying the form of the primary analysis; modifying the patient population; etc

Outline

- Introduction and taxonomy of clinical trial designs
  - Pre-1990's
- Basic principles of adaptive designs
  - Allocation rule
  - Sampling rule
  - Stopping rule
  - Decision rule
- Phases of development
- Adaptive designs for the learn phase of drug development
  - First-in human / MTD
  - Two-stage designs
  - Adaptive dose-ranging designs
  - Bayesian adaptive randomisation
- Adaptive designs for the confirmatory phase of drug development
  - Sample size re-assessment
  - Adaptive group sequential designs
  - Seamless phase II/III designs
  - Population enrichment designs
- Practical aspects of adaptive design implementation
- Discussion

The late Lew Sheiner
Learning and Confirming (1997)

CLINICAL PHARMACOLOGY & THERAPEUTICS
VOLUME 61 NUMBER 3 MARCH 1997

COMMENTARY

Learning versus confirming in clinical drug development

Lewis B. Sheiner, MD *San Francisco, Calif.*

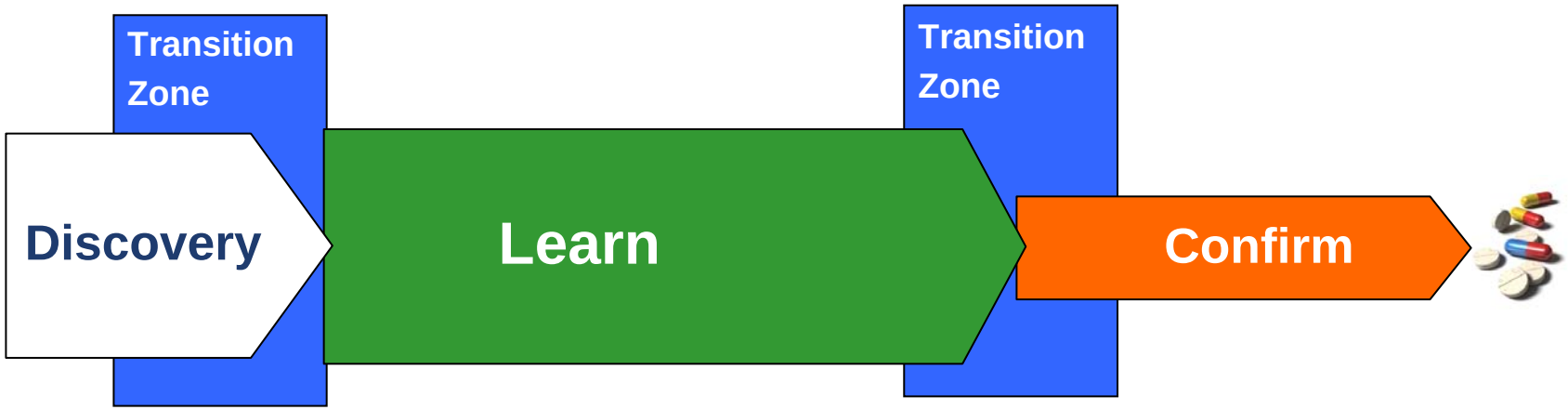
“Learn and Confirm” Paradigm

LEARN: Learning/exploratory activities

- Phase I already adaptive, but could be smarter:
  - better estimate of safety
  - better understanding of PK-PD
- Phase IIa & IIb (possibly combined) – sweet spot of current adaptive design:
  - “Quick win/Quick kill”
  - test more doses
  - test more hypotheses
- During the LEARN phase we can modify the following characteristics:
  - Population
  - Endpoints
  - Dosing regimen

CONFIRM: Confirming/ confirmatory activities

- Phase III limited adaptation possible, but useful insurance:
  - start with more doses
  - sample size reassessment
- During the CONFIRM phase we can still modify the following characteristics:
  - Drop doses
  - Focus on a sup-population
  - Re-assess Sample size
  - Early Stopping



Types of Adaptive Designs: Learn

First-in Human	• Single ascending dose escalation designs • Up-and-Down and CRM to find MTD • Establish Proof-of-Mechanism or Proof-of-Target Modulation

MAD and PoC	• Two-stage adaptive approach in patients
	  • 1st stage – to identify MTD
	  • 2nd stage – to select dose and exposure levels (necessary cond.)

PoC and ADRS	• Start with the highest feasible tolerated dose and placebo • If a pre-specified futility condition is satisfied => stop • Otherwise, open enrollment to lower doses
Seamless Phase I/II Design	• SAD or MAD combined with Biomarker-based Efficacy • To identify the Optimal Safe Dose
Adaptive Dose Ranging Design	• Finding a target dose (MED, EDp) • Response Adaptive Allocation • Covariate Adjusted Response Adaptive Allocation

CRM: Continual Reassessment Method; **MTD:** Maximum Tolerated Dose; **MAD**: Multiple Ascending Dose; **SAD**: Single Ascending Dose; **MED**: Minimum Effective Dose; **EDp:** Dose achieving 100p% of maximum effect

Types of Adaptive Designs: Confirm

Sample Size Reassessment	• Sample size adjustment based on blinded or unblinded data:
	  • Using nuisance parameter estimate
	  • Using treatment effect estimate

Adaptive Group Sequential Design	• Early stopping for efficacy, futility, harm or safety • Adjusting the number and/or timing of interim analyses • Increasing the maximum sample size
Seamless Phase II/III Design	• Design combining the objectives of Phase II dose ranging study and confirmatory Phase III trial in a single protocol • Dose selection at the interim analysis
Population Enrichment Design	• Placebo run-in; Active control run-in; Dose titration • Adaptively enrich the population at the interim analysis • Enrich based on biomarker or clinical endpoint response
Adaptive TRT Switching	• Randomized withdrawal design • Re-randomization of responders or/and non-responders

Why? Need to Improve Drug Development

- Pharmaceutical industry pipeline problem: fewer approvals, escalating development costs, high late phase attrition, tougher regulatory environment, expiring patents in blockbusters
- Traditional development paradigm not sustainable
- Failure rate in Phase III estimated at 50%
- Innovative designs and analysis methods are key priority for improving clinical development practice

Motivation for Adaptive Design

- Opportunity to calibrate initial assumptions used at trial design based on partial observed information
- Improved knowledge efficiency vs. conventional (non-adaptive) designs
  - Faster
  - Less expensive
  - More information for same investment
- Increase likelihood of success, or reliable early termination (e.g., futility rule)
- Improved understanding of treatment effect

Challenges for Adaptive Design

- Adaptive designs (AD) offer considerable opportunities for improving drug development, but come with risks and costs
- Industry mindset favoring traditional development approaches ⇨ change management
- Regulatory concerns with new approaches, especially in confirmatory studies: FDA draft guidance on AD quite helpful in that regard
- Need adequate operational infrastructure: recruitment, data management, drug supply, etc
- Resource needs: increased planning, more people with proper expertise; adequate commercial software for design and implementation; hardware for intensive computing

How? Implementation Process

Define study objectives. Endpoints, treat. effects, target for adaptation, study pop, base case enrollment

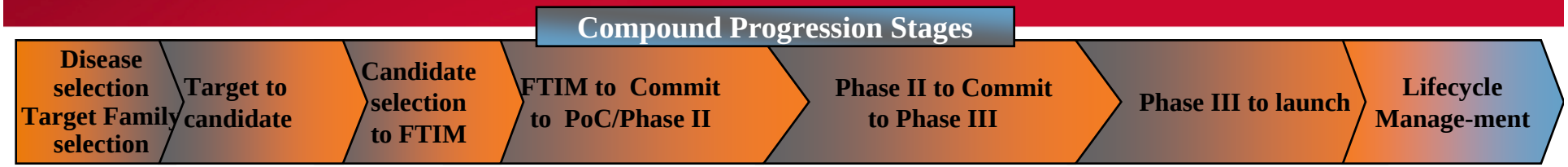
Identify design options (including conventional design)

Simulation to compare design Performance (within context of CDP) Recommend design & evaluate assumptions

Simulation Report Quantify decision process Assess alternative recruitment speed & site numbers to achieve optimal recruitment

Ensuring appropriate firewalls in place, more complex logistics

Classification



SINGLE ARM TRIALS
Two-stage Designs
Screening Designs
TWO-ARM TRIALS
Group Sequential Designs
Information Based Designs
Adaptive GSD (Flexible Designs)
MULTI-ARM TRIALS
Bayesian Designs
Group Sequential Designs
Flexible Designs
DOSE-FINDING STUDIES
Dose-escalation designs
Dose-finding designs (Flexible Designs)
Adaptive Model-based Dose-finding
SEAMLESS DESIGNS
Dose-escalation based on efficacy/toxicity
Learning/Confirming in Phase II/III

Two-Stage Designs

- **Objective:** single-arm studies using short-term endpoints; hypothesis testing about some minimal acceptable probability of response
- Gehan design: early stopping for futility; sample size of the 2nd stage gives a specified precision for response rate
- Group sequential designs: Fleming (1982), Simon (1989)
- Adaptive two-stage design: Banerjee&Tsiatis (2006)
- Bayesian designs: Thall&Simon (1994)

Screening Designs

- **Objective:** adaptive design for the entire screening program
  - Minimize the shortest time to identify the “promising” compound
  - Subject to the given constraints on type I and type II risks for the entire screening program
    - type I risk = Pr(screening procedures stops with a FP compound)
    - type II risk= Pr(any of the rejected compounds is a FN compound)
- Two-stage design (Yao&Venkatraman, 1998)
- Adaptive screening designs (Stout and Hardwick, 2002)
- Bayesian screening designs (Berry, 2001)

Classification

Group Sequential Designs

- **Objective:** testing two hypotheses with given significance level and power at the specified alternative, prefixed maximum sample size

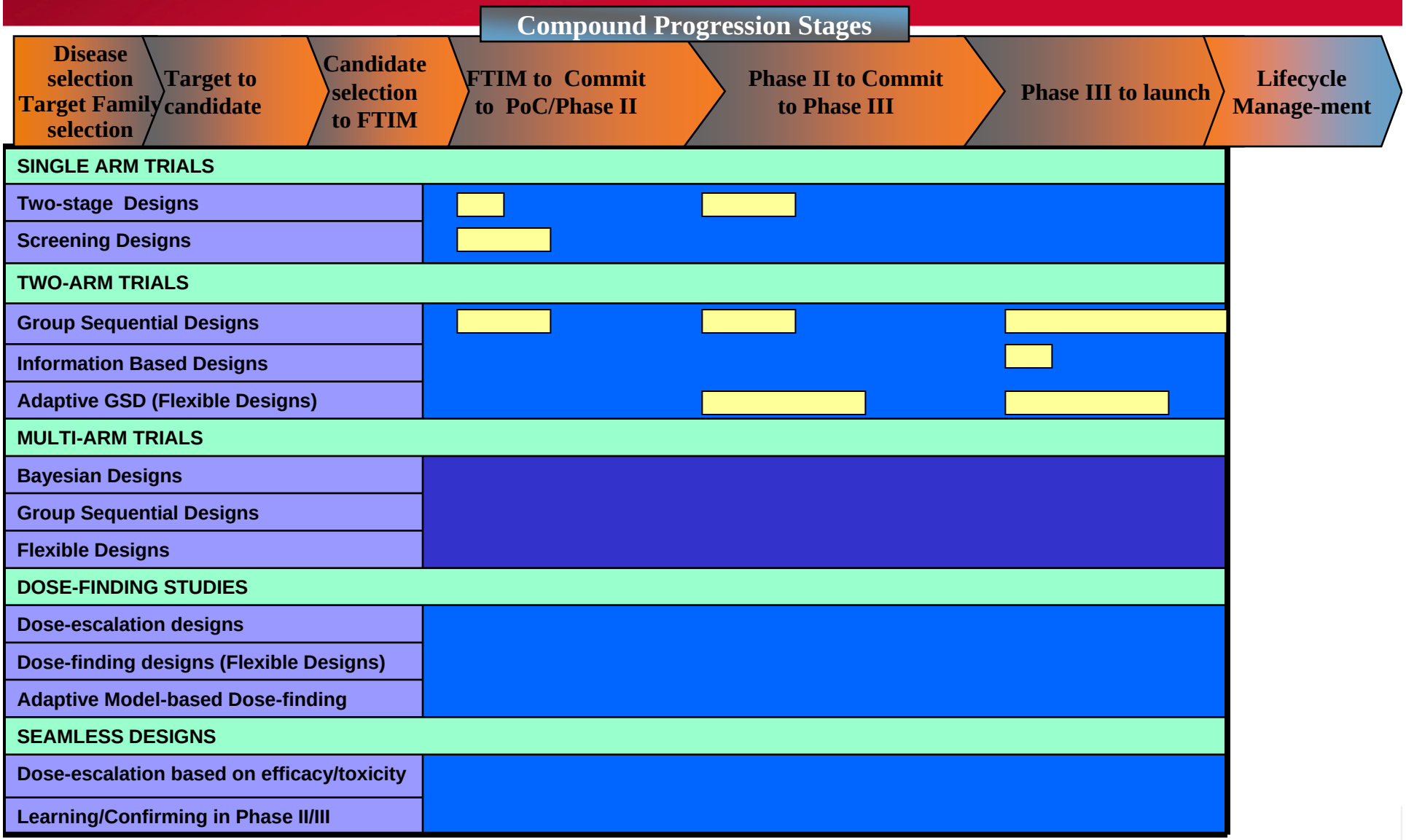
  - **AR:** fixed randomization
  - **SaR:** after a fixed number (a group) of observations,
    - or using error-spending function,
    - or using “Christmas-tree” adjustment
  - **StR:** boundary crossing
    - Haybittle, Pocock, O’Brien-Fleming type
    - linear boundaries
    - error-spending families
    - conditional power, stochastic curtailment
  - **DR:** final decision - to accept or reject the null hypothesis

- **References:** Jennison&Turnbull (2000); Whitehead (1997)

Adaptive GSD (Flexible Designs)

- **Objective:** testing two hypotheses with given significance level and power at the specified alternative or adaptively changing the alternative at which a specified power is to be attained
  - **AR:** fixed or adaptive randomization
  - **SaR:** sample size of the next stage depends on results at the time of interim analysis
  - **StR:** p-value combination, conditional error, variance-spending
  - **DR:** adapting alternative hypothesis, primary endpoint, test statistics, inserting or skipping IAs
- **References:** Bauer; Brannath et al; Müller&Schäfer; Fisher

Classification



Bayesian Designs

- **Objective:** to use the posterior probabilities of hypotheses of interest as a basis for interim decisions (*Proper Bayesian*) or to explicitly assess the losses associated with consequences of stopping or continuing the study (*Decision-theoretic Bayesian*)
  - **AR:** equal randomization or *play-the-winner* (next patient is allocated to the currently superior treatment) or *bandit designs* (minimizing the number of patients allocated to the inferior treatment)
  - **SaR:** not specified
  - **StR:** not formally pre-specified stopping criterion, or using a *skeptical prior* for stopping for efficacy and an *enthusiastic prior* for stopping for futility, or using *backwards induction*
  - **DR:** update the posterior distribution; formal incorporation of external evidence; inference not affected by the number and timing of IAs
- **References:** Berry (2001, 2004); Berry et al. (2001); Spiegelhalter et al. (2004).

Pairwise comparisons with GSD

- **Objective:** compare multiple treatments with a control; focus on type I error rate rather than power

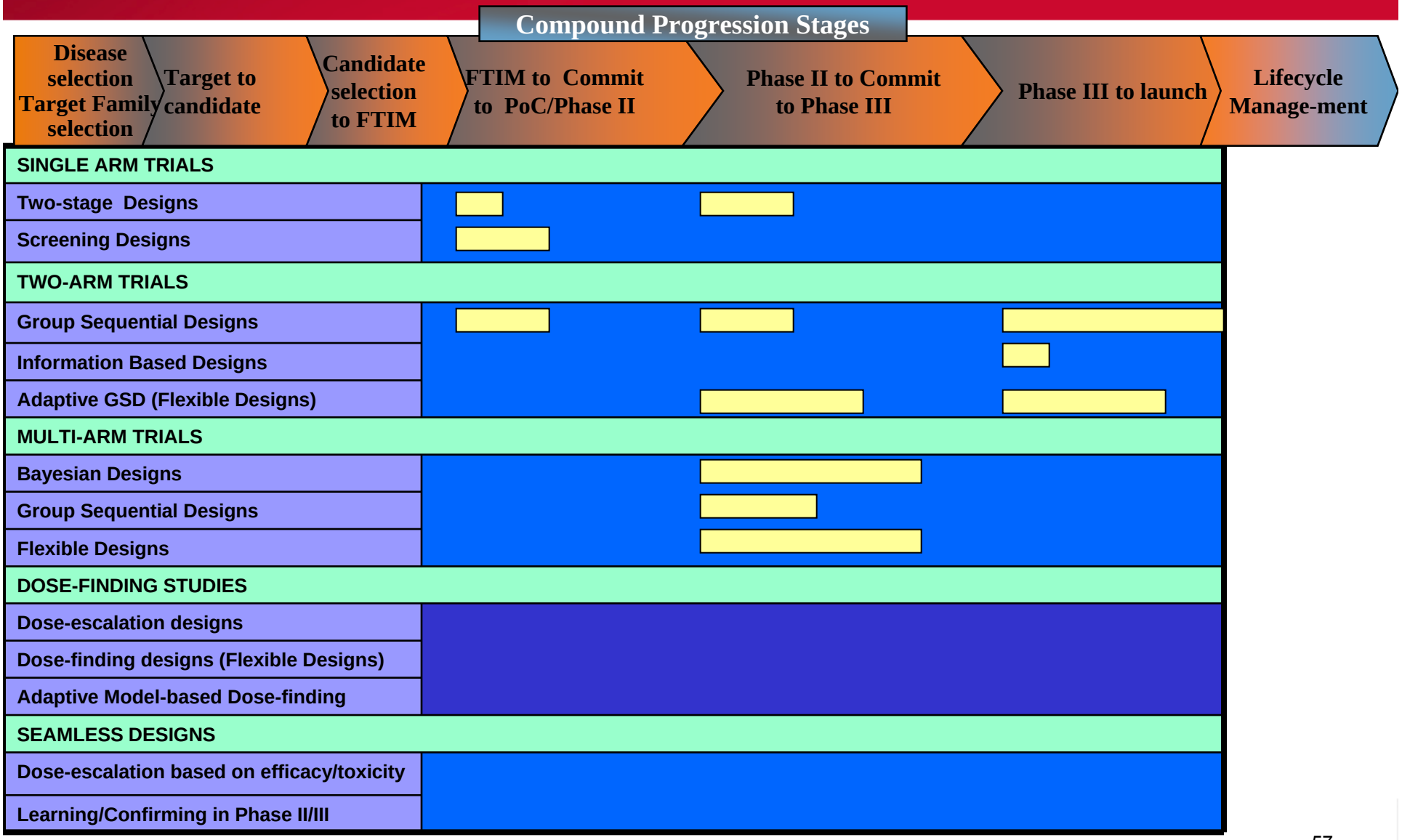
  - A simple Bonferroni approximation is only slightly conservative
  - Treatments may be dropped in the course of the trial if they are significantly inferior to others
  - "Step-down" procedures allow critical values for remaining comparisons to be reduced after some treatments have been discarded

- **References:** Follmann et al (1994)

p-value combination tests

- **Objective:** compare multiple treatments with a control in a two-stage design allowing integration of data from both stages in a confirmatory trial
- **Focus:** control of multiple (familywise) Type I error level
- **Great flexibility:**
  - General distributional assumptions for the endpoints
  - General stopping rules and selection criteria
  - Early termination of the trial
  - Early elimination of treatments due to lack of efficacy or to safety issues or for ethical/economic reasons
- **References:** Bauer&Kieser (1994)

Classification



Dose-escalation designs

- **Objective:** target the MTD (Phase I) or the best safe dose (Phase I/II) or find the therapeutic window
  - **AR:** non-parametric (3+3 rule, up-and-down)
    - or model-based (Continual Reassessment Methods)
    - or Escalation With Overdose Control (EWOC)
    - or Bayesian Decision Design
    - or Bayesian Optimal Design
    - or Penalized Adaptive D-optimal Design
  - **SaR:** cohorts of fixed size or in two stages (Storer design)
  - **StR:** no early stopping or stopping by design (e.g. 3+3 rule)
  - **DR:** update model parameters (for model-based AR)
- **References:** O'Quigley et al.; Babb et al., Dragalin (2010)

Adaptive Model-based Dose-finding

- **Objective:** find the optimal dose; working model for the dose-response; dose sequence identified in advance
  - **AR:** *Bayesian* (based on predictive probabilities: smallest average posterior variance) or *frequentist* (based on optimal experimental design: maximum information per cost)
  - **SaR:** cohorts of fixed size or after each observation
  - **StR:** stopping for futility or when the optimal dose for confirmatory stage is sufficiently well known (estimation!)
  - **DR:** update model parameters, Bayesian predictions of long-term endpoint using a longitudinal model
- **References:** Berry et al. (2001); Dragalin & Fedorov (2006)

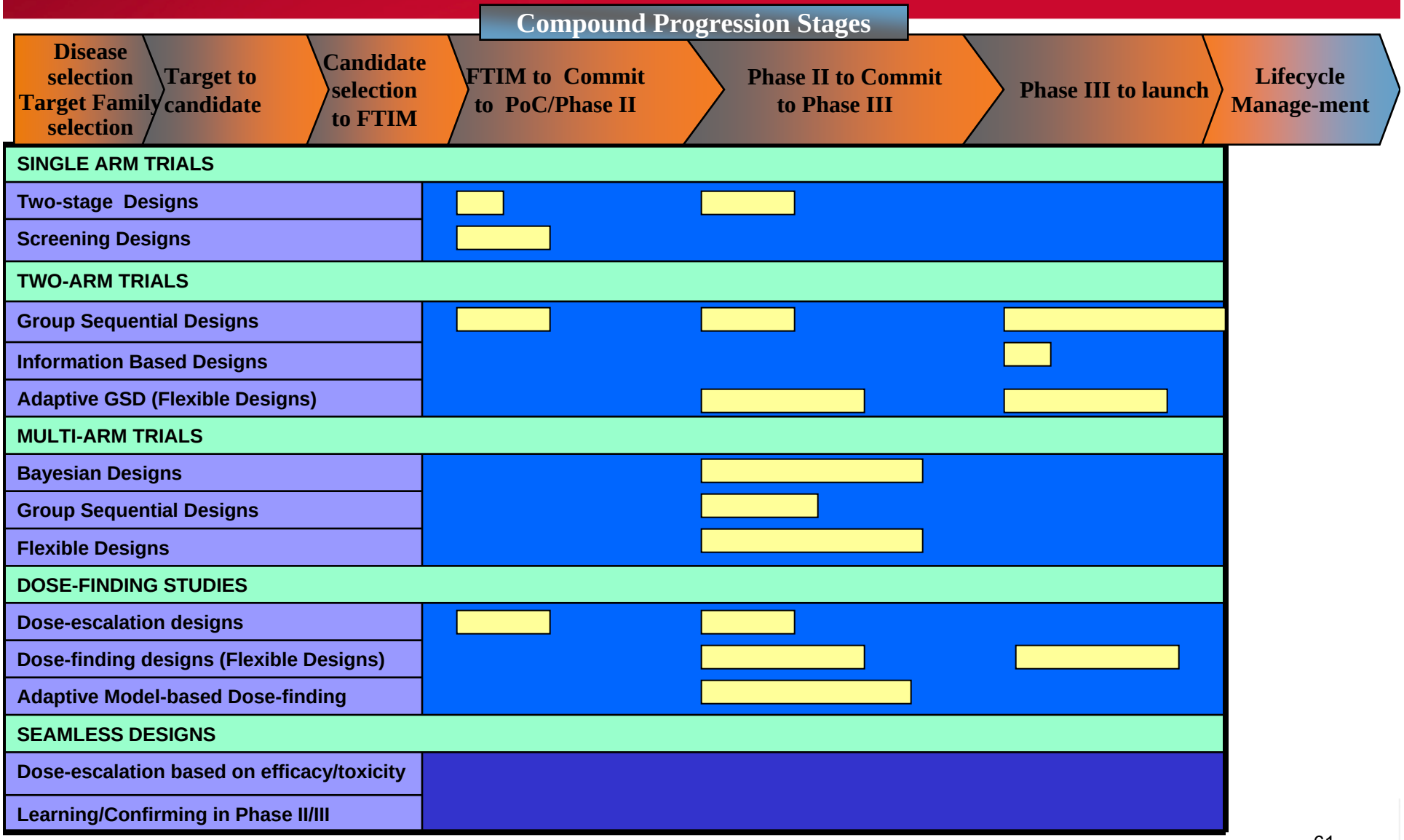
Adaptive Dose-finding (Flexible Designs)

- **Objective:** establishing a dose-response relationship or combining Phase II/III using p-value combination tests

  - **AR:** drop or add doses
  - **SaR:** sample size reassessment for the next stage
  - **StR:** early stopping for futility or early termination of some inferior doses
  - **DR:** adapting hypotheses, primary endpoint, test statistics, inserting or skipping IAs

- **References:** Bauer&Kohne

Classification



Seamless Designs: Definition

Seamless design

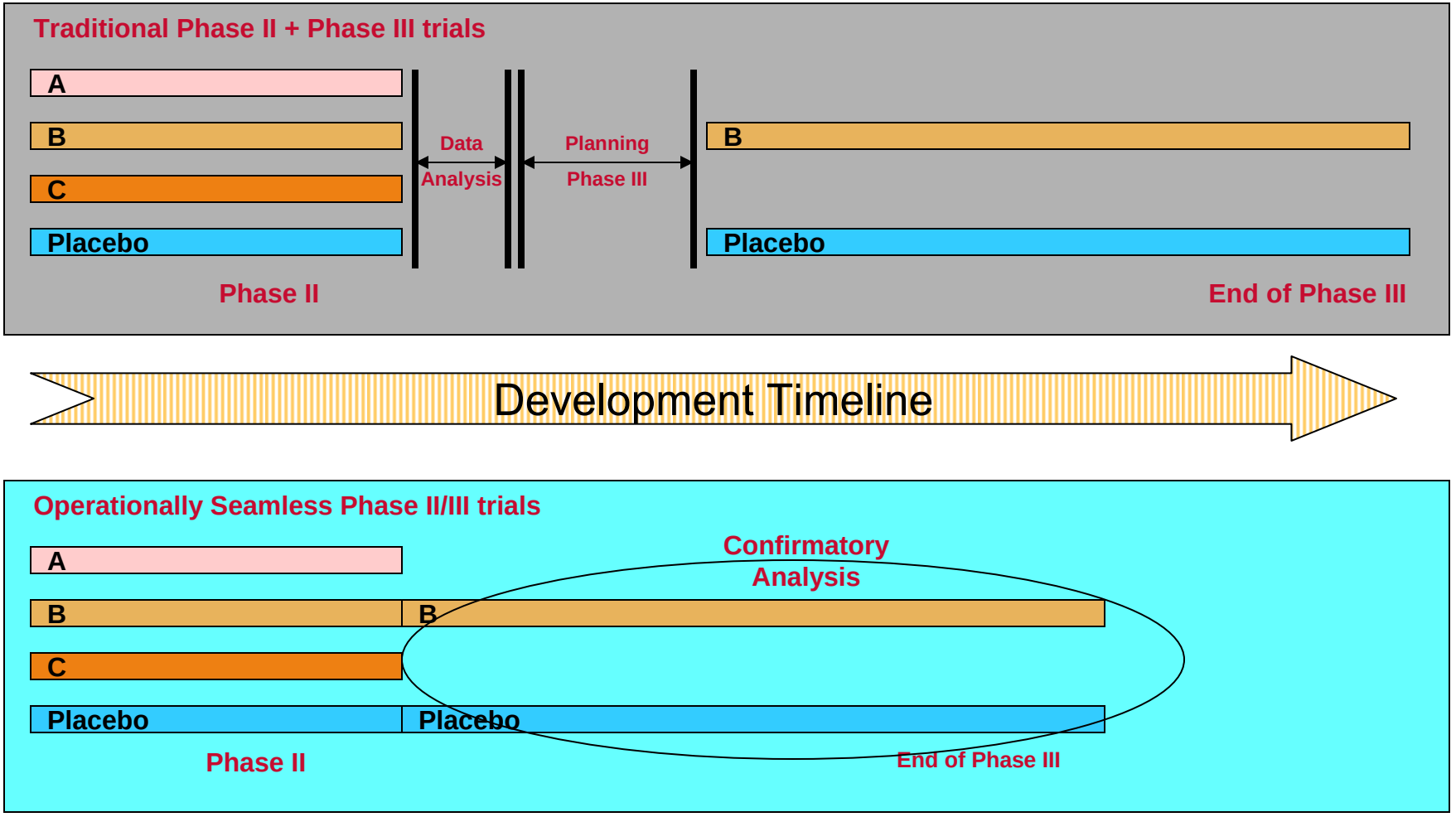
- A clinical trial design that combines into a single trial objectives which are traditionally addressed in separate trials (*operationally* seamless)

Adaptive Seamless design

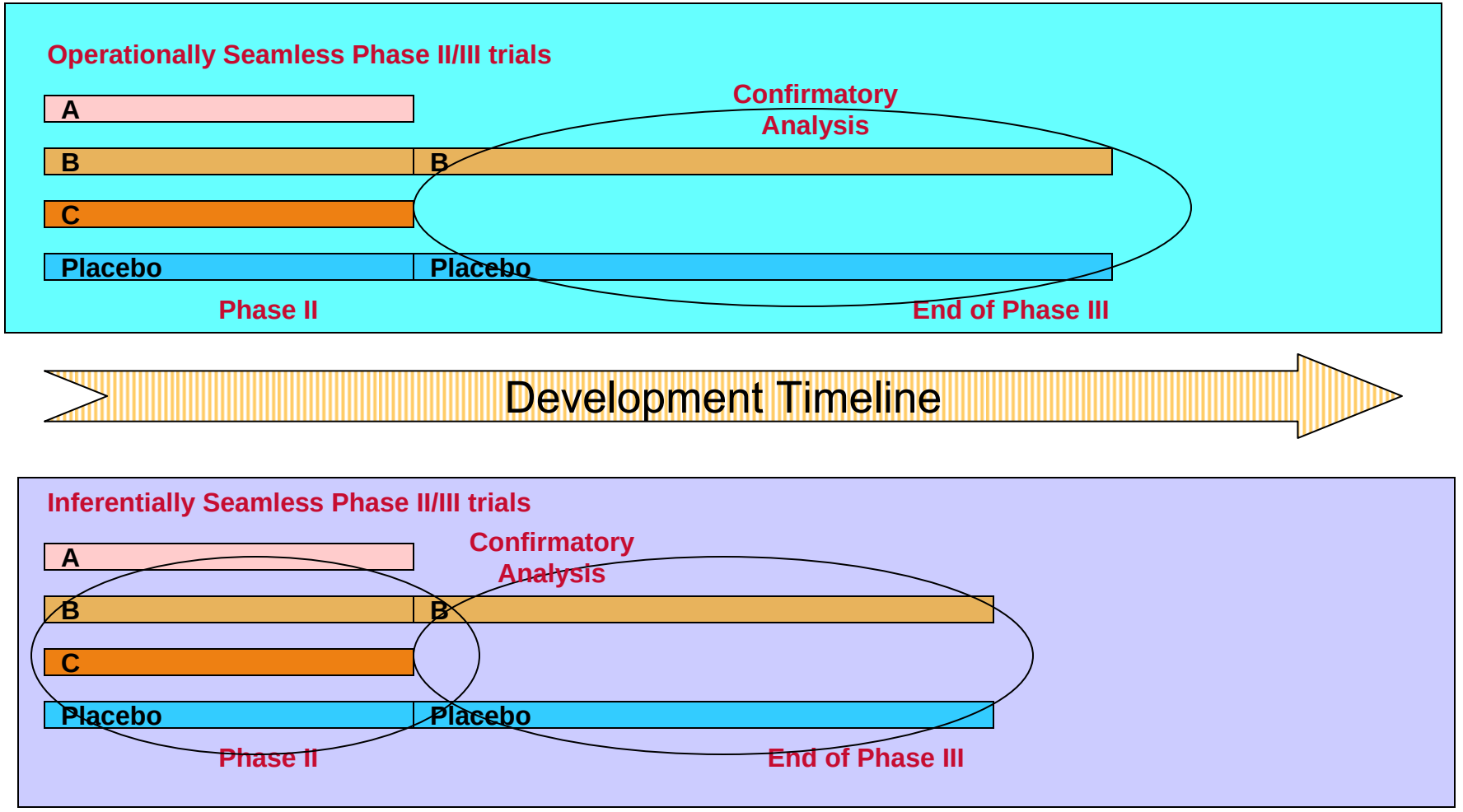
- A seamless trial in which the final analysis will use data from patients enrolled before and after the adaptation (*inferentially* seamless)

References: Maca et al. (2006)

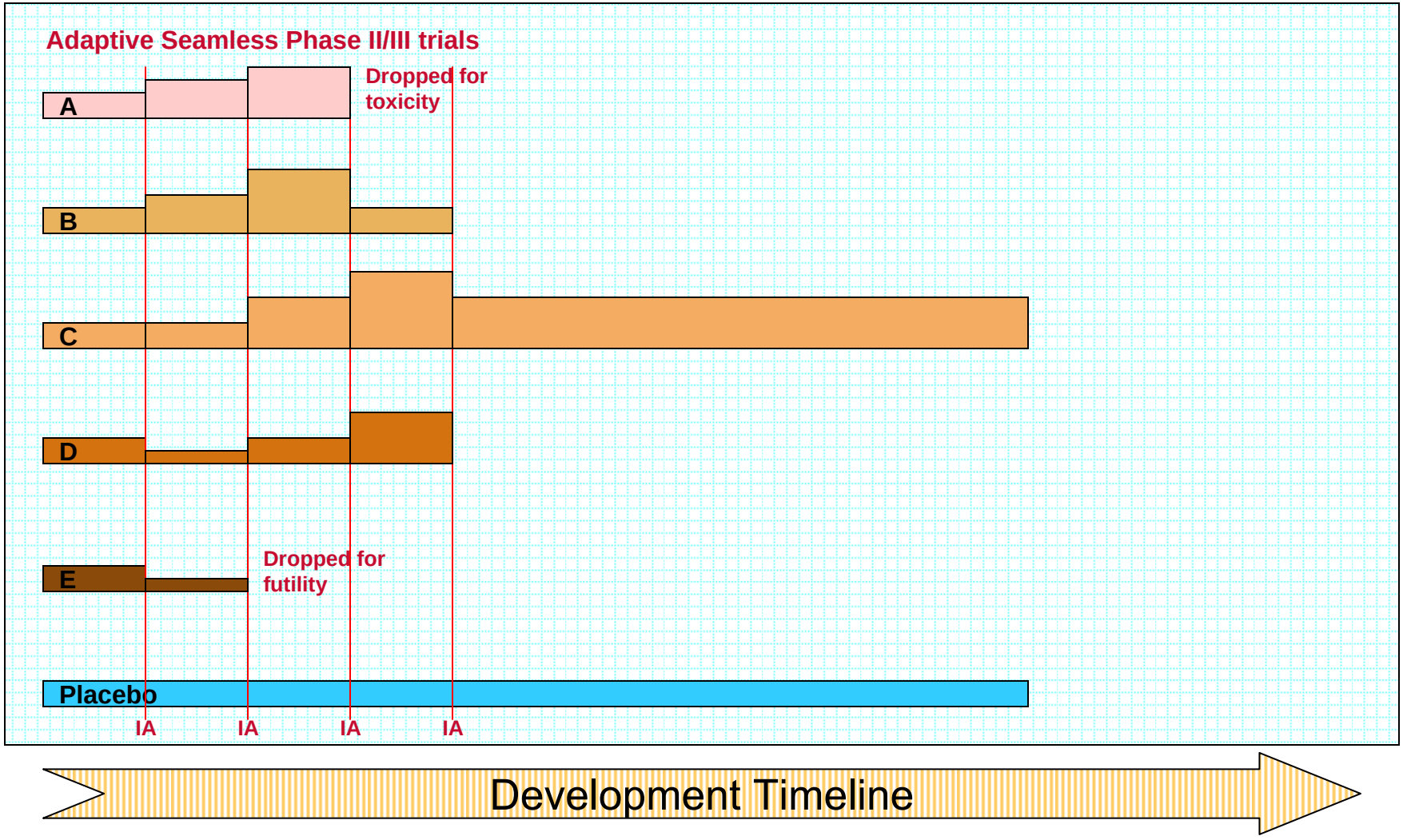
Faster: Operationally Seamless



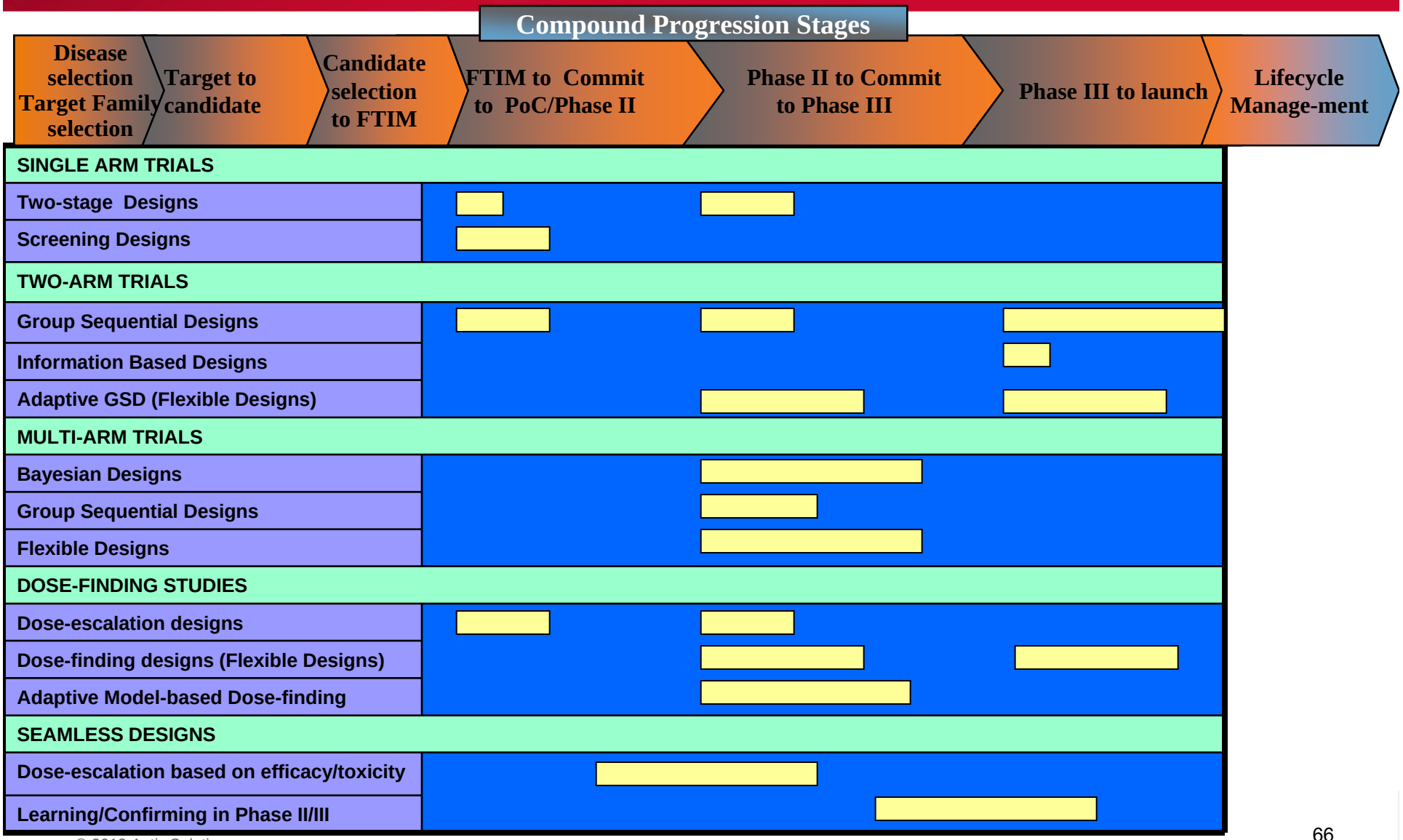
At lower costs: Inferentially Seamless



Better: Adaptive Seamless



Classification



Outline

- Introduction and taxonomy of clinical trial designs
  - Pre-1990's
- Basic principles of adaptive designs
  - Allocation rule
  - Sampling rule
  - Stopping rule
  - Decision rule
- Phases of development
- Adaptive designs for the learn phase of drug development
  - First-in human / MTD
  - Two-stage designs
  - Adaptive dose-ranging designs
  - Bayesian adaptive randomisation
- Adaptive designs for the confirmatory phase of drug development
  - Sample size re-assessment
  - Adaptive group sequential designs
  - Seamless phase II/III designs
  - Population enrichment designs
- Practical aspects of adaptive design implementation
- Discussion

The Continuous Reassessment Method (CRM)

- Goal : identify a dose with the targeted toxicity as quickly as possible and focus experiment at that dose
- Doses are pre-defined : d_1, d_2, …., d_k
- Outcome is binary : DLT / No DLT
- Assumption : There exists a monotone dose-response function $\psi(d;\theta) = \text{Prob}(DLT|d,\theta)$ depending on a single parameter θ
- The number of patients N is fixed in advance

O'Quigley et al (1990)

CRM Original form

- Given the doses : d_1, d_2, …., d_k, define a set of probabilities p_1, p_2, …., p_k
- Define : Prob (DLT|d_j) = $(p_j)^{\theta}$ – this can be thought of as a local model

- Aside – In O'Quigley et al (1990) dose was not necessarily predefined - could be a combination of compounds whose rank order was assumed known
- Given p_1, p_2, …., p_k , d_1, d_2, …., d_k can be defined by

$$d_i = \tanh^{-1}(2p_i - 1)$$

CRM Original form

- A second alternative model looked at by O'Quigley et al specifies the Dose-response model as follows :

$$p(DLT \mid d_j) = \frac{\exp(c + \beta d_j)}{1 + \exp(c + \beta d_j)}$$

  - For some constant *c*.
  - This is a one-parameter logistic model. O'Quigley et al (1990) suggested *c*=3

CRM Original form

- A “vague” prior is assumed for θ, eg g(θ)=exp(-θ) with mean 1
  - Alternatively : Prob (DLT|d_j) = $(p_j)^{\exp(\theta)}$

$$p(\theta) \sim N(0,\sigma^2)$$

- Suppose that you have observed a sequence of doses and response pairs (d_i,y_i={0,1}) i=1, …, N
- Posterior distribution for θ is

$$p(\theta \mid d, y) \propto \prod_{i}^{N} p(y_i \mid d_i, \theta) e^{-\theta} d\theta$$

CRM Original Form

- The Mean of the distribution is available to give information about θ
- Predictive probabilities

$$\pi_i = \int_a p(Y = 1 \mid d_i, \theta)\, p(\theta \mid \underline{d}, \underline{y})\, d\theta$$

- Choose as next dose the one which gives π_i closest to the target π^*
- Or : choose as next dose the one for which $E[(p_j)^{\exp(\theta)}|\text{Data}]$ is closest to π^*
- Continue until a pre-specified number of patients - final dose is the estimate

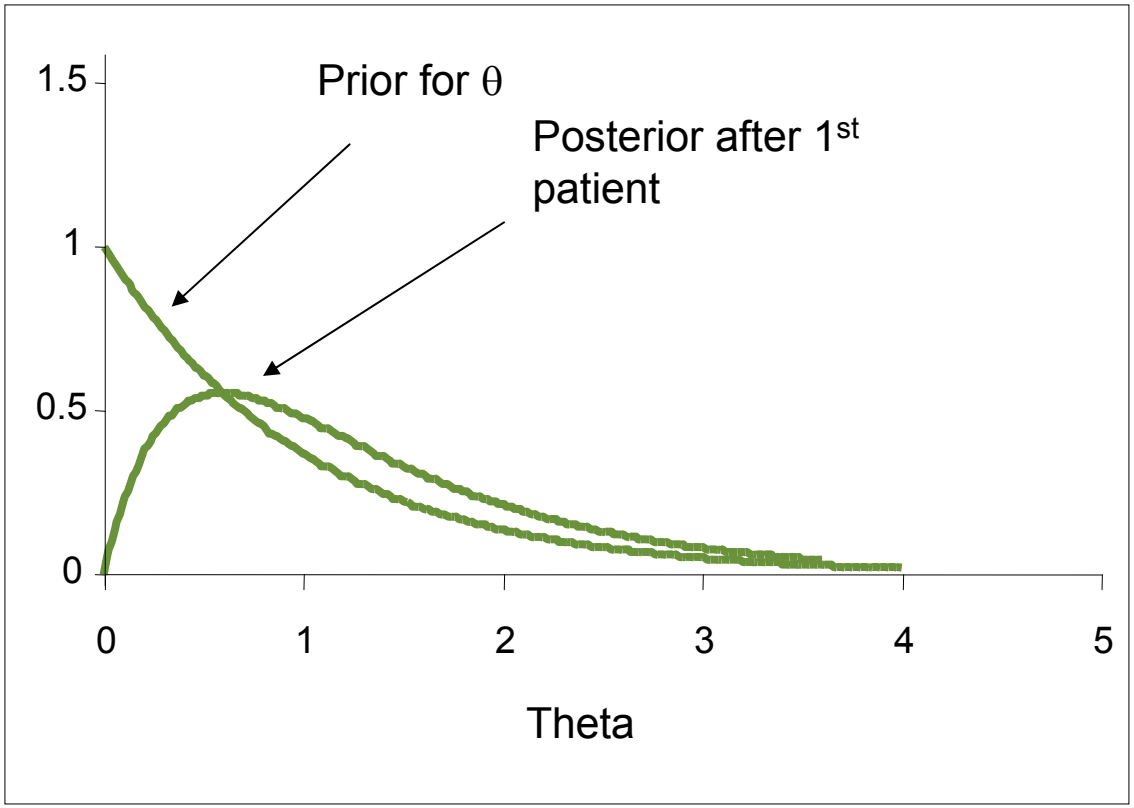
Simulated example: O'Quigley et al (1990)

- O'Quigley et al illustrate the principle of the CRM with the following simulated example
  - The purpose was to estimate the MTD for a target probability of 0.2
  - The prior on θ : g(θ)=exp(-θ) – prior mean 1.0
  - There were 6 doses with the following initial probabilities :

    0.05 0.10 0.20 0.30 0.50 0.70

    The 1st patient was allocated to the 3rd dose – no DLT
  - The posterior for $\theta \propto (1-0.2^{\theta})e^{-\theta}$

Simulated example: O'Quigley et al (1990)



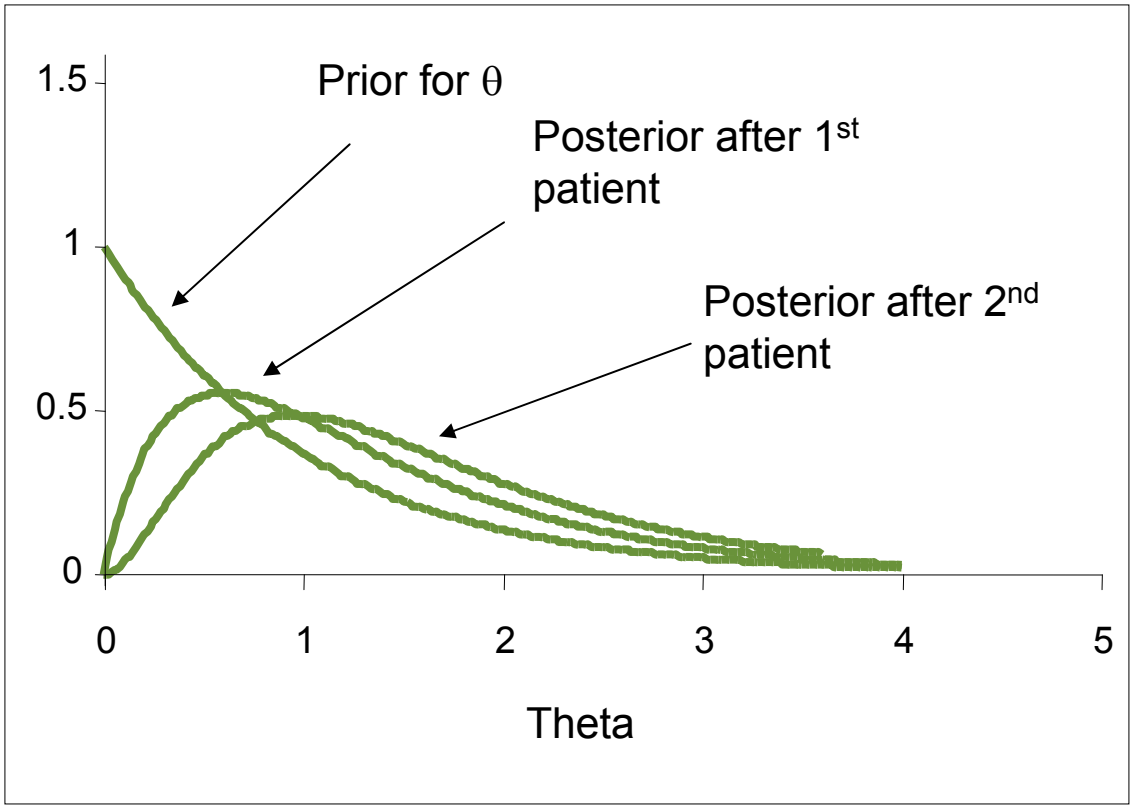
Simulated example: O'Quigley et al (1990)

- The posterior mean : 1.38
- For this value of θ, the estimated probabilities of response are

0.016 0.041 0.108 0.189 0.354 0.611

- 0.189 is closest to 0.2 – choose dose 4
- The 2nd patient was allocated to the 4th dose – no DLT
- The posterior for $\theta \propto (1-0.2^{\theta})(1-0.3^{\theta})e^{-\theta}$

Simulated example: O'Quigley et al (1990)



Simulated example: O'Quigley et al (1990)

- The posterior mean : 1.68
- For this value of θ, the estimated probabilities of response are

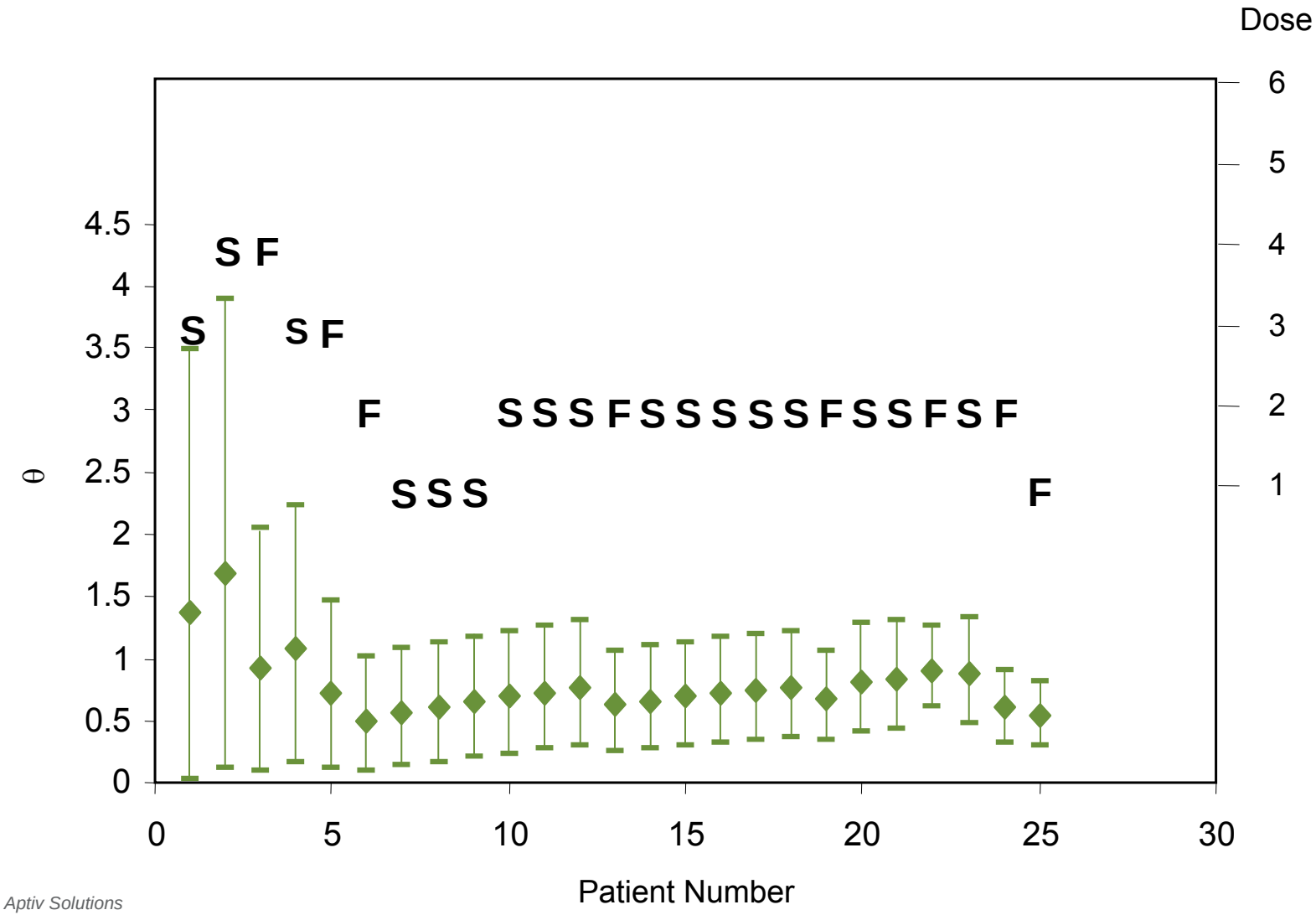
0.006 0.021 0.067 0.132 0.312 0.549

- 0.132 is closest to 0.2 – choose dose 4
- etc

O'Quigley et al (1990) Simulation

Pat	Dose						Posterior	Estimated Expected Response					
	1	2	3	4	5	6	Mean	1	2	3	4	5	6
0	0.05	0.10	0.20	0.30	0.40	0.50	1.00	0.05	0.10	**0.20**	0.30	0.40	0.50
1			0				1.38	0.02	0.04	0.11	**0.19**	0.38	0.61
2				0			1.68	0.01	0.02	0.07	**0.13**	0.31	0.55
3				**1**			0.93	0.06	0.12	**0.22**	0.33	0.52	0.72
4			0				1.07	0.04	0.08	**0.18**	0.27	0.48	0.68
5			**1**				0.72	0.12	**0.19**	0.31	0.42	0.61	0.77
6		**1**					0.50	**0.22**	0.32	0.45	0.55	0.71	0.84
7	0						0.56	**0.19**	0.28	0.41	0.51	0.68	0.82
8	0						0.60	**0.16**	0.25	0.38	0.48	0.66	0.81
9	0						0.64	0.15	**0.23**	0.36	0.46	0.64	0.80
10		0					0.69	0.13	**0.21**	0.33	0.44	0.62	0.78
11		0					0.73	0.11	**0.19**	0.31	0.42	0.60	0.77
12		0					0.77	0.10	**0.17**	0.29	0.40	0.59	0.76
13		**1**					0.63	0.15	**0.23**	0.36	0.47	0.65	0.80
14		0					0.66	0.14	**0.22**	0.34	0.45	0.63	0.79
15		0					0.69	0.13	**0.20**	0.33	0.43	0.62	0.78
16		0					0.72	0.12	**0.19**	0.31	0.42	0.61	0.77
17		0					0.75	0.11	**0.18**	0.30	0.41	0.60	0.77
18		0					0.77	0.10	**0.17**	0.29	0.40	0.59	0.76
19		**1**					0.67	0.13	**0.21**	0.34	0.45	0.63	0.79
20		0					0.69	0.12	**0.20**	0.33	0.43	0.62	0.78
21		0					0.72	0.12	**0.19**	0.32	0.42	0.61	0.77
22		**1**					0.64	0.15	**0.23**	0.36	0.46	0.64	0.80
23		0					0.66	0.14	**0.22**	0.35	0.45	0.63	0.79
24		**1**					0.60	**0.17**	0.25	0.38	0.49	0.66	0.81
25	**1**						0.54	**0.20**	0.29	0.42	0.52	0.69	0.83
	0												

Posterior Distributions for θ: O'Quigley et al (1990)



Issues

- 5 patients are treated above the MTD
  - Is this ethical ? Acceptable ?
- Dose 2 was chosen 15 times consecutively
  - Is this a sufficient indication of the MTD ?
  - Does the posterior for θ itself indicate that stability has been achieved ?

O'Quigley et al (1990) - Simulations

	1	2	3	4	5	6	
Prior	0.05	0.10	0.20	0.30	0.40	0.50	% Toxicity
P(DLT)	0.05	0.10	0.20	0.30	0.40	0.50	
%chosen	5	17	37	34	7	0	22
%final	1	20	45	32	3	0	
P(DLT)	0.09	0.16	0.27	0.38	0.57	0.75	
%chosen	19	29	31	19	2	0	23
%final	14	39	37	11	0	0	
P(DLT)	0.30	0.40	0.52	0.61	0.76	0.87	
%chosen	74	14	9	3	0	0	34
%final	93	7	0	0	0	0	
P(DLT)	0.00	0.00	0.04	0.09	0.25	0.49	
%chosen	0.5	1	9	35	52	2	17
%final	0	0	1	30	67	3	
P(DLT)	0.01	0.03	0.09	0.16	0.35	0.59	
%chosen	1	4	21	50	23	0.2	18
%final	0	0	16	54	20	0	

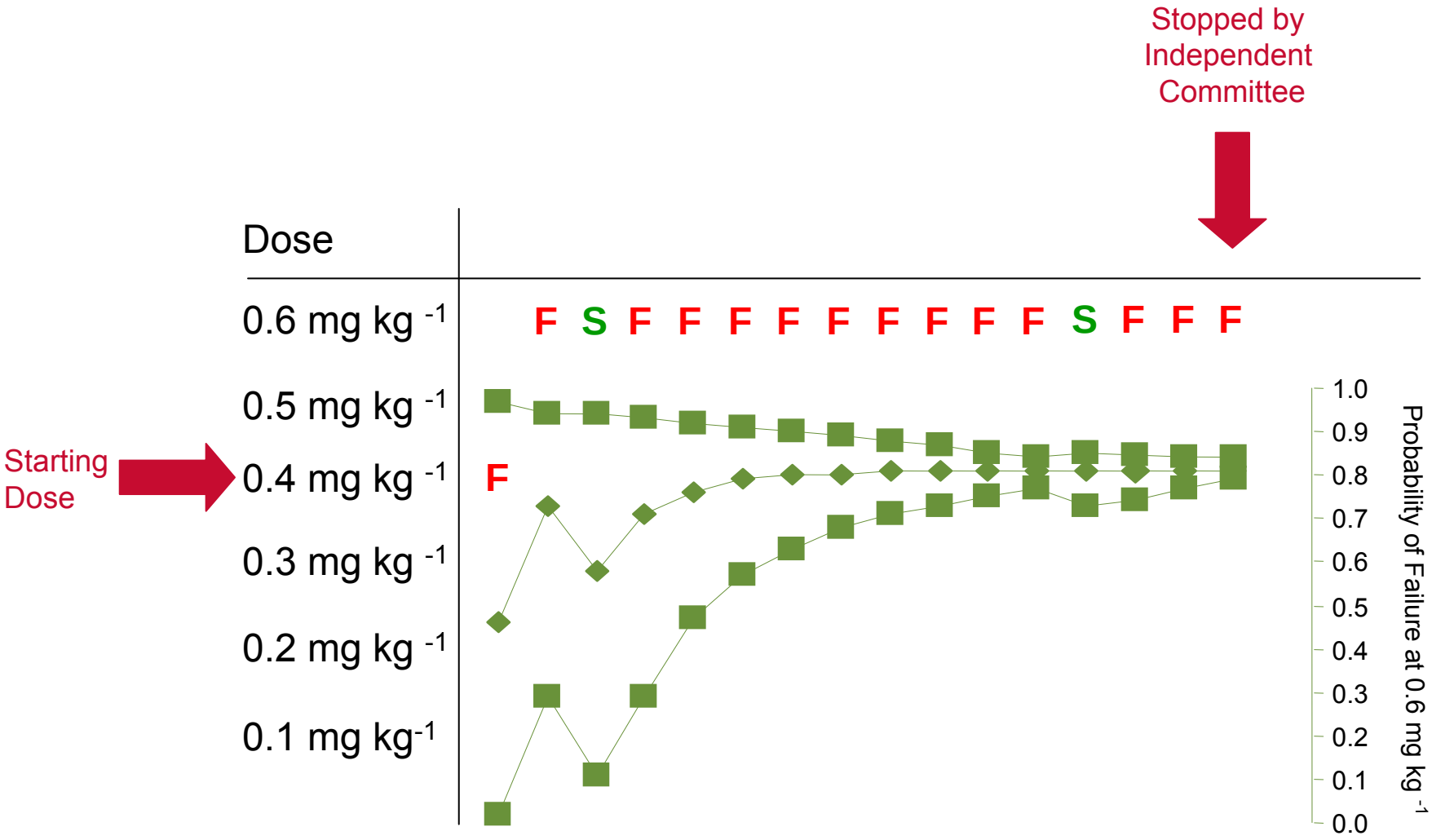
CRM Dose Selection : Sedation of Infants / Cardiac Catheterisation

- Fabre et al (1998) - Br J Clin Pharm
- Aim : Find ED90 (90% sedated)
- Bayesian approach
- One parameter (α) logistic dose response
- Choose dose to "optimise" gain (utility) function
  - predictive probabilities

$$\pi_i = \int_a p(Y = 1 \mid x_i, \alpha) p(\alpha \mid \underline{x}, \underline{y}) d\alpha$$

- Choose as next dose the one which gives π_i closest to the target π (ED90)

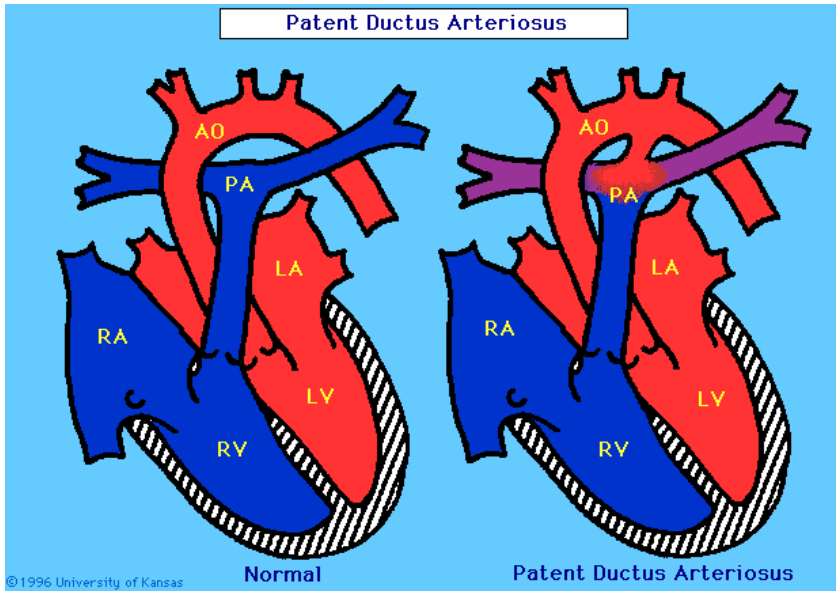
CRM Design Infant Sedation



Dose-finding study of ibuprofen in patent ductus arteriosus

- Study designed to find the minimum effective dose regimen (MEDR) of IBU (one course) required to close ductus arteriosus in preterm infants.

  - Study run in two independent groups (20 per group)
  - PMA 27-29 weeks : 80% closure
  - PMA < 27 weeks
  - 50% closure



Example of Clinical CRM

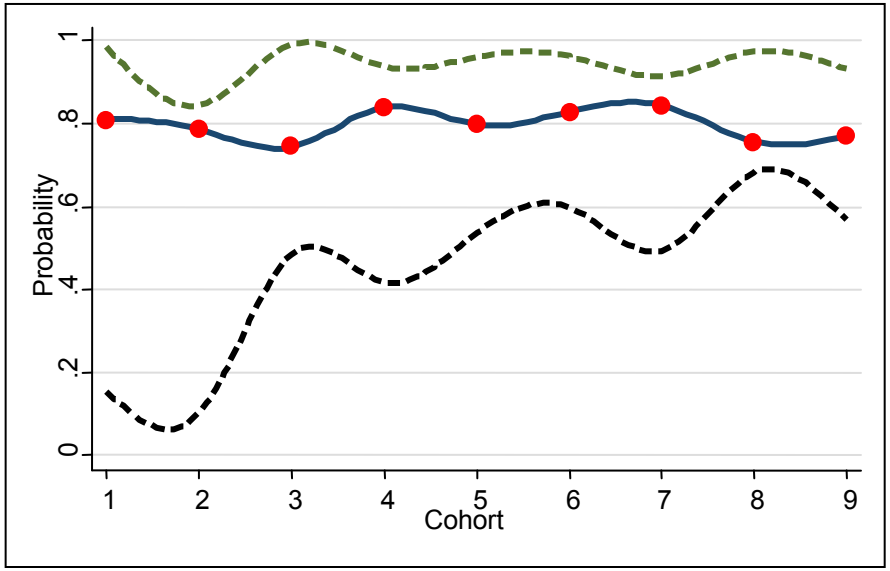
Journal of Clinical Pharmacy and Therapeutics (2005) **30**, 121–132

ORIGINAL ARTICLE

Dose-finding study of ibuprofen in patent ductus arteriosus using the continual reassessment method

L. Desfrere* MD, S. Zohar† PhD, P. Morville‡ MD, A. Brunhes* MD, S. Chevret† MD PhD, G. Pons§ MD PhD, G. Moriette* MD, E. Rey§ MD and J. M. Treluyer§ MD PhD

				Ibuprofen loading dose (mg/kg)			
				5	10	15	20
Cohort	Patients	Allocated	Success/	Prior estimated probabilties of success			
	(n)	Dose	Failure	0.6	0.8	0.9	0.95
1	3	10	2/1	0.481	0.683	**0.812**	0.891
2	1	5	0/1	0.370	0.544	0.682	**0.787**
3	3	15	3/0	0.539	**0.744**	0.861	0.925
4	3	10	2/1	0.512	0.717	**0.840**	0.915
5	3	15	2/1	0.467	0.667	**0.799**	0.882
6	2	15	2/0	0.500	0.703	**0.829**	0.903
7	1	10	1/0	0.519	0.723	**0.845**	0.914
8	3	15	3/0	0.553	**0.757**	0.870	0.931
9	1	10	1/0	0.567	**0.771**	0.880	0.938



Issues with CRM

- Needs more inputs (prior, defined MTD)
- Simulations have shown relatively good performance
- Designed for cancer trials but seems widely applicable
- Critics of design have suggested stepped increments, repeated increments, starting from minimum possible dose
- Critics have suggested logistic curves, non-parametric curves
- Can use cohorts, predefined stopping rules, eg if 6 patients treated with same dose stop.

Stopping Rules

- The use of stopping rules will reduce the amount of experimentation when unnecessary and hence can be regarded as more ethical

- Approaches
  - CI interval width
  - Number of patients at MTD
  - Futility
  - Prediction

Stopping Rules

- CI width
- Heyd and Carlin (1999) proposed that the width of the posterior interval for the free parameter be used to stop
  - Model: the one-parameter logistic
  - They considered 5 rules
    1. at maximum sample size
    2. when β is < 1.0 or maximum sample size reached
    3. when β is < 1.5 or maximum sample size reached
    4. when β is < 2.0 or maximum sample size reached
    5. when β is < 1.5

		Stopping Rule				
Parameter	Truth	1	2	3	4	5
β	4.00	4.00	4.00	4.01	4.03	4.03
P(DLT|MTD)	0.30	0.27	0.30	0.38	0.35	0.38
MTD	2.15	2.12	2.15	2.24	2.20	2.23
Sample Size		24.0	23.4	16.5	13.2	16.5

Stopping Rules

- Number of Patients at MTD
  - Korn et al (1994) proposed that if the next recommended dose has already been used for K patients the trial should stop (cf O'Quigley simulated example).
  - They concentrated on K=6.
  - Results : The stopping rule reduced the number of patients relative to the 3+3
- "Futility"
  - Thall (2001) proposed stopping the trial if

    $$\Pr(\text{DLT at } d_1 | \text{data}) > 0.95,$$

    i.e. there is a very high chance of a DLT even at the lowest dose, the objective of the trial cannot be achieved

Safety Rule
Stop if p(DLT at d_1 | data) > 0.95

	1	2	3	4	5	6	
Prior	0.05	0.10	0.25	0.40	0.50	0.55	% Toxicity
P(DLT)	0.25	0.40	0.50	0.60	0.75	0.80	
CRM	60	35	5	0	0	0	32
Safety	56	36	5	0	0	0	31
P(DLT)	0.40	0.50	0.55	0.60	0.60	0.60	
CRM	87	11	2	0	0	0	41
Safety	50	10	2	0	0	0	26
P(DLT)	0.45	0.50	0.55	0.60	0.60	0.60	
CRM	88	10	2	0	0	0	46
Safety	39	10	2	0	0	0	24
P(DLT)	0.5	0.55	0.60	0.60	0.60	0.60	
CRM	95	4	1	0	0	0	50
Safety	25	4	1	0	0	0	16

Target : Prob(DLT)=0.25
N=26 patients in 12 cohorts, size=3
Starting dose : d_3

Peter Thall
Adaptive Sequential Dose-Finding Methods in Phase I/II Clinical Trials.
Henry Stewart Conference, Washington
July 2001

Other DE Designs (continued)

- **2. Escalation with Overdose Control (EWOC)**
- Model: $p = \dfrac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$
- Reparameterization:

$$\beta_0 = \frac{1}{\gamma - x_{\min}}\left(\gamma \text{logit}(\rho_0) - x_{\min}\text{logit}(p)\right)$$

$$\beta_1 = \frac{1}{\gamma - x_{\min}}\left(\text{logit}(p) - \text{logit}(\rho_0)\right)$$

where:
γ = MTD
ρ_0 = Pr(DLT) at x_{min}

- Marginal posterior cdf of the MTD: $\Pi_k(x)$
- Escalation Scheme: The kth patient is allocated to dose $x_k = \Pi_k^{-1}(\alpha)$ so that the posterior probability of exceeding MTD is equal to the "feasibility bound," α.

References: Babb J, Rogatko A, Zacks S (1998); Chu et al. (2009)

Novartis Example ofe Use of MCRM

Motivating example (from Neuenschwander, et al, 2008)

- Open-label, multicenter, non-comparative, dose-escalation cancer trial designed to characterize the safety, tolerability and PK profile of a drug and to determine its MTD.
- The pre-defined doses were 1, 2.5, 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 150, 200 and 250 mg. Target P(DLT)=.3.
- The first cohort of patients was treated at 1mg. No DLTs were observed for the first four cohorts of patients.

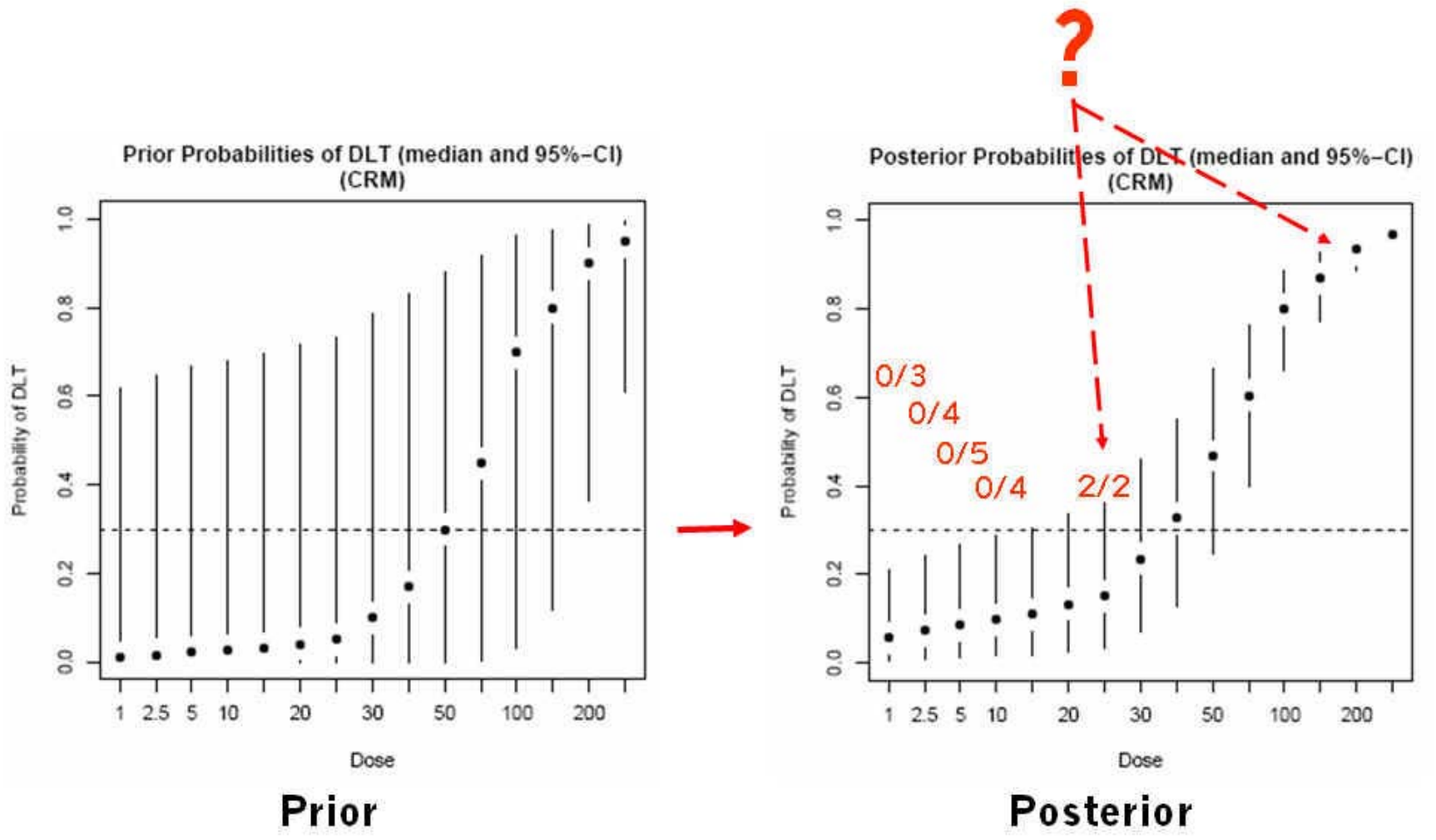
  ➔ clinical team decided to skip 2 doses to 25mg (contradicting the planned MCRM in which doses were not supposed to be skipped)
- Both patients dosed at 25 mg experienced DLT
  - MCRM recommended further escalation,

Case study – CRM Results

	Doses									
	1	2.5	5	10	15	20	25	30	40	50
no. of patients	3	4	5	4	–	–	2	–	–	–
no. of DLT's	0	0	0	0	–	–	2	–	–	–
	Posterior summaries (original skeleton)									
Skeleton (CRM)	0.010	0.015	0.020	0.025	0.030	0.040	0.050	0.100	0.170	0.300
Mean	0.069	0.085	0.099	0.111	0.123	0.144	0.163	0.242	**0.330**	0.465
StDev	0.055	0.062	0.068	0.072	0.076	0.082	0.087	0.101	0.109	0.108
	Posterior summaries (equi-distant skeleton)									
Skeleton (CRM)	0.063	0.125	0.188	0.250	0.313	0.375	0.438	0.500	0.563	0.625
Mean	0.024	0.054	0.090	0.130	0.176	0.226	**0.281**	0.341	0.405	0.475
StDev	0.030	0.051	0.069	0.084	0.097	0.107	0.115	0.119	0.120	0.117

- Recommendation:
  - from original p_i: dose = 40 or 30
  - from equidistant p_i: dose = 25 (questionable)
  - Note: the p_i are structural assumptions, should not be changed!

One Novartis experience with Case study Results

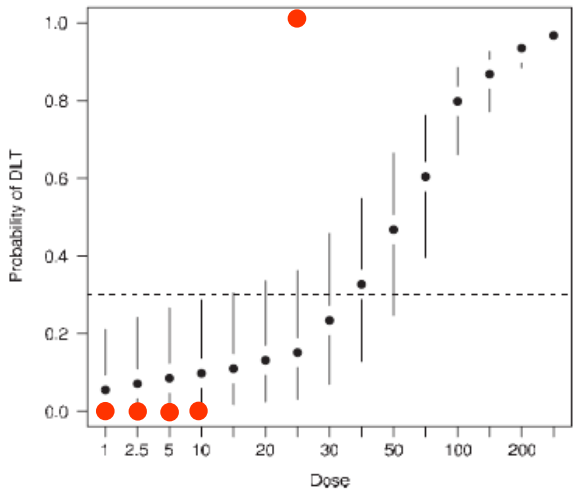


Neuenschwander, Branson &Gsponer SIM, 2008

- Problems with One parameter approaches to MTD estimation
- Two important differences to CRM

  1) Use of two parameter logistic

  2) Change the method determining the dose for next patient (cohort)



- Determine the posterior probability that the DLT probability at each dose is in the range:

Under dosing	: 0.00-0.20
Target	: 0.20-0.35
Excessive	: 0.35-0.60
Unacceptable	: 0.60-1.00

- Choose the dose with the largest posterior probability of being in target and meeting overdose criteria

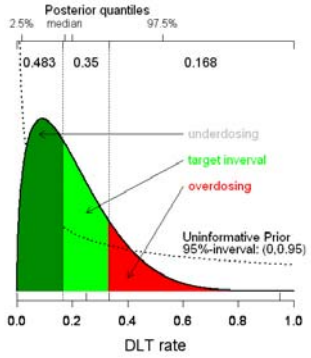
Output

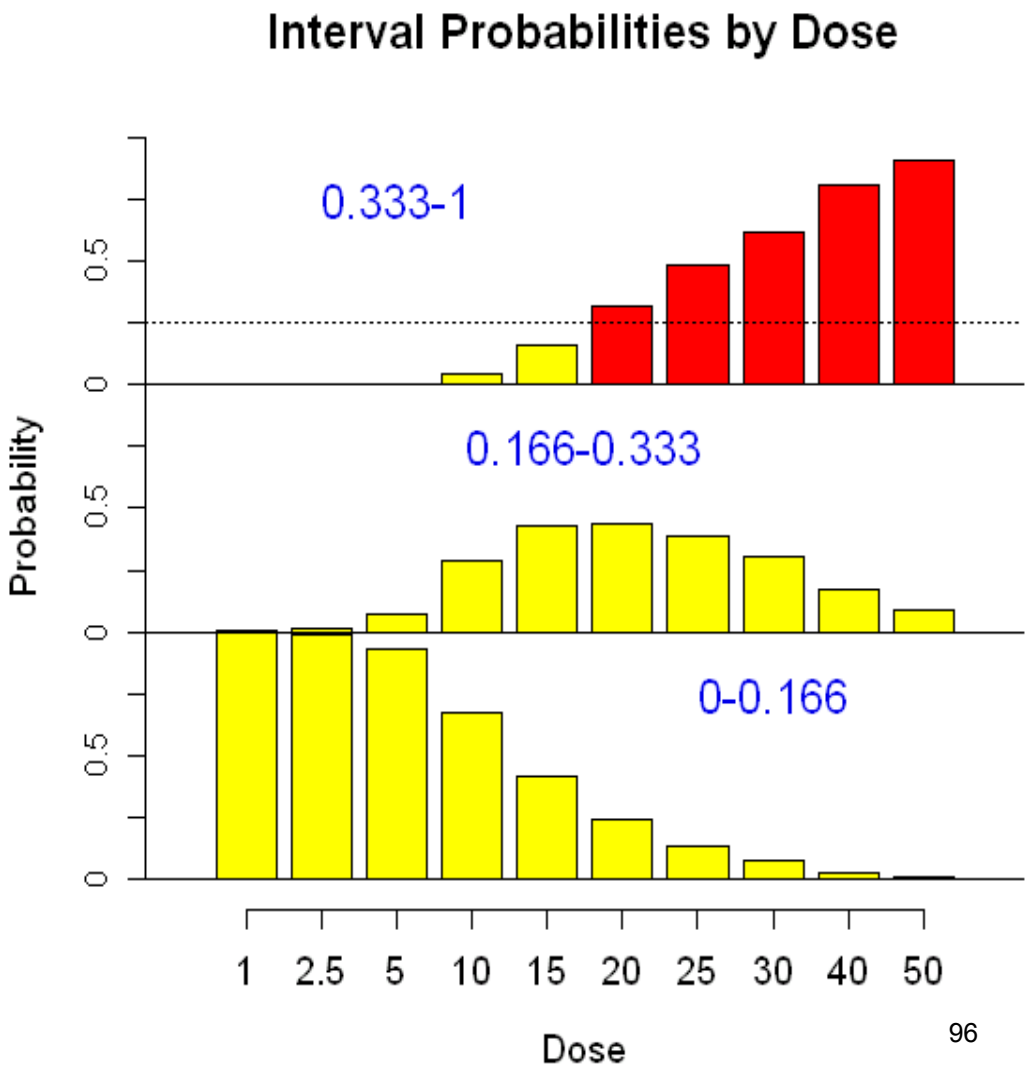
Interval Probabilities: underdosing, targeted toxicity, overdosing

overdosing

targeted toxicity

underdosing





Top Panel

probability of *overdosing*

failed overdose criterion in red!

Pr(true DLT rate p >0.333) > 25%

Middle Panel

probability of *targeted toxicity*

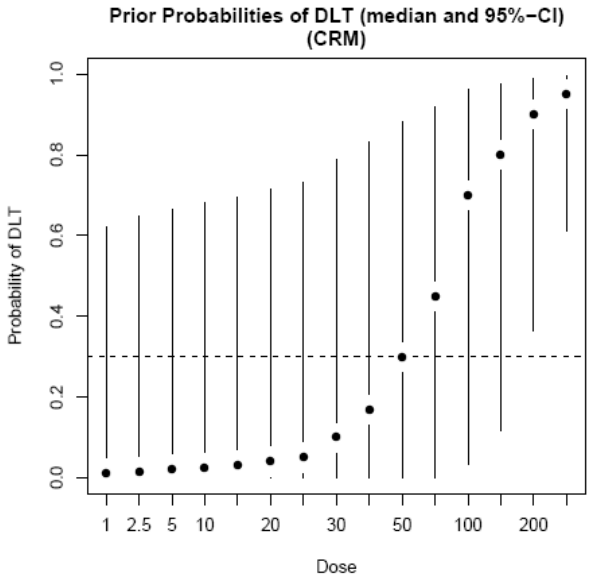
Bottom Panel

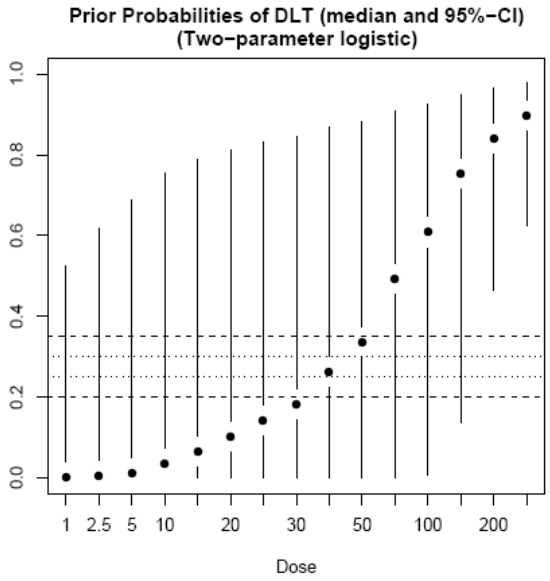
probability of *underdosing*

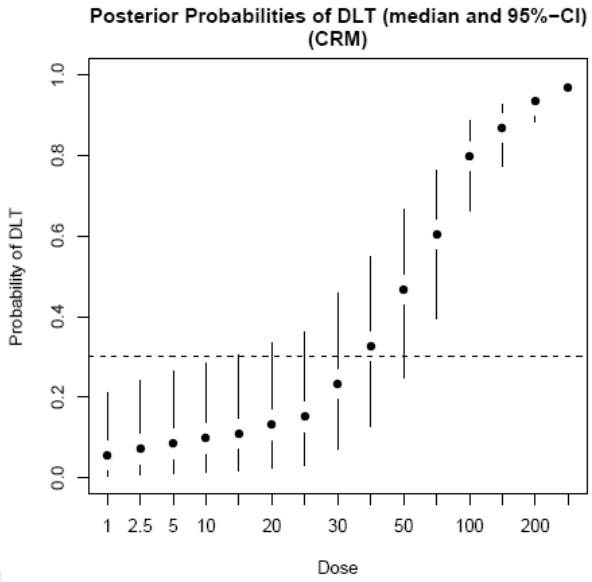
Recommended Dose

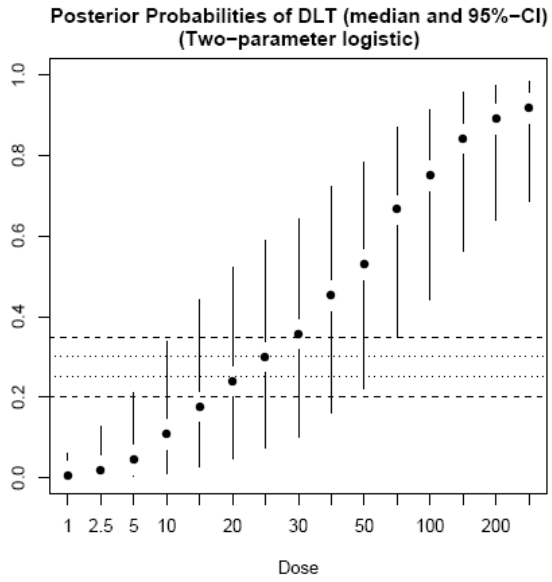
15 (max target w/ overdose<25%)

Case study - comparison









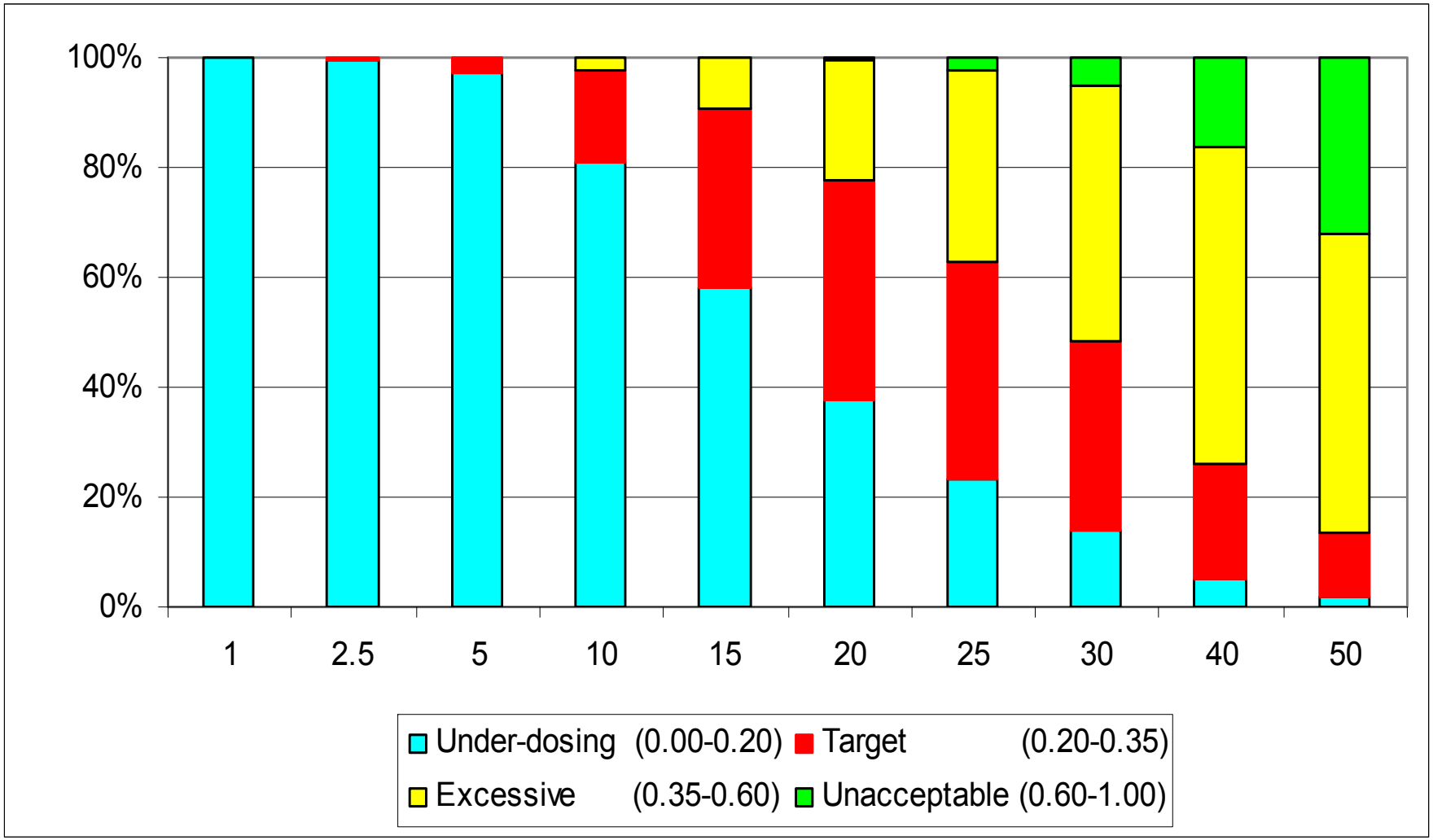
Priors

- Prior for BLR chosen to be similar to prior for CRM

Posteriors

- CRM: “too” narrow intervals for doses where no data have been seen. Similar things happen for other 1-parameter models

Neuenschwander, Branson &Gsponer SIM, 2008



Escalation With Overdose Control (EWOC)

- Model: $p = \dfrac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$
- Reparametrization:

$$\beta_0 = \frac{1}{\gamma - x_{\min}}\left(\gamma \text{logit}(\rho_0) - x_{\min}\text{logit}(p)\right)$$

$$\beta_1 = \frac{1}{\gamma - x_{\min}}\left(\text{logit}(p) - \text{logit}(\rho_0)\right)$$

where:
γ = MTD
ρ_0 = Pr(DLT) at x_{min}

- Marginal Posterior CDF of MTD: $\Pi_k(X)$
- Escalation scheme: The $(k+1)st$ patient is allocated to dose $x_k = \Pi_k^{-1}(\alpha)$ so that the posterior probability of exceeding MTD is equal to the "feasibility bound," α.

References: Babb J, Rogatko A, Zacks S (1998); Chu et al. (2009)

Outline

- Introduction and taxonomy of clinical trial designs
  - Pre-1990's
- Basic principles of adaptive designs
  - Allocation rule
  - Sampling rule
  - Stopping rule
  - Decision rule
- Phases of development
- Adaptive designs for the learn phase of drug development
  - First-in human / MTD
  - Two-stage designs
  - Adaptive dose-ranging designs
  - Bayesian adaptive randomisation
- Adaptive designs for the confirmatory phase of drug development
  - Sample size re-assessment
  - Adaptive group sequential designs
  - Seamless phase II/III designs
  - Population enrichment designs
- Practical aspects of adaptive design implementation
- Discussion

Adaptive Seamless MAD/POC Study in Rheumatoid Arthritis

Key Discussion Points

- Key Study Parameters
- Study Design
- Advantages of the Seamless Design
- Regulatory Agency Review
- Top-Levels Results
- Key Benefits and Savings
- Main Conclusions

Key Study Parameters

- Seamless phase I/II, randomized, double-blind, placebo-controlled, sequential/parallel design
- 6 cohorts (up to 33 subjects each)
  - 5 doses of Drug X (10, 30, 50, 60, and 100 mg)
  - 1 pooled placebo cohort
- Treatment duration: 16 weeks, 4 subcutaneous injections
- Primary endpoint
  - ACR20 response at week 16
- Randomization
  - Stage 1: Initial dose escalation according to traditional MAD sequential format (3:1 active to placebo) using Week 4 DLT endpoint
  - Stage 2: After highest tolerated cohort is open, randomization will proceed in a parallel fashion for all “safe” treatment arms and placebo
  - Enrollment to futile doses can be stopped using Week 4 biomarker

Study Design

Ascending MAD until all doses are open

Safety Decision: Subjects will receive a 2nd dose only after a safety review of the 2nd dose in the preceding cohort.

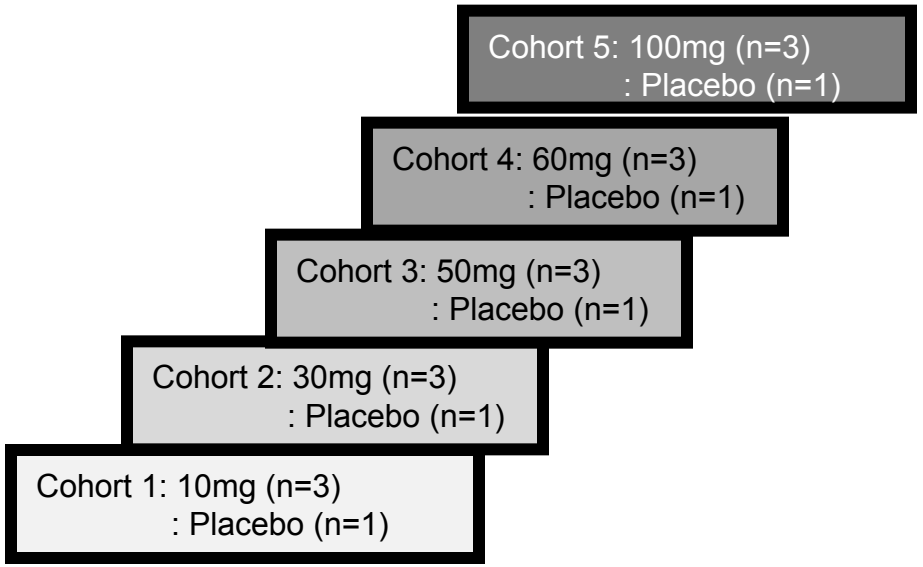
The dose regimen is Q4 Weeks for 4 cycles.

2nd Stage begins after escalating to the maximum tolerated dose

N=33 patients for each of five doses (10, 30, 50, 60, 100 mg) and placebo

STAGE 2

STAGE 1



Futility Decision: Based on ACR20 and 25% reduction in CRP at 4 weeks

To avoid incongruent data, enrollment will not be stopped for futility in a higher dose if a lower dose is still ongoing

Internal DMC for safety & futility decisions:

Unblinded Medical Monitor
Unblinded Biostatistician

Advantages of the Seamless Design

- Increases the utility of information obtained by allowing subjects in both the MAD and POC stages to provide both definitive safety and efficacy data
- Performing the MAD component in subjects with Rheumatoid Arthritis allows for the earliest characterization of safety of Drug X in the clinically relevant population
- Minimizes the number of subjects that are exposed to ineffective doses of Drug X
- Focuses subjects to doses that are most informative for accurate dose selection for subsequent confirmatory trials
- Optimizes the benefit/risk balance for participating subjects via improved efficiency of decision making in relation to the doses of Drug X studied

Regulatory Agencies Review (1)

- Regulatory Agencies consulted
  - FDA, MHRA, MPA, BfArM, Canada Health

Outcomes

- Appropriateness of seamless aspect
  - UK, Sweden, Canada: transition appropriate & efficient given the mechanism of action and safety review in study
  - Germany: efficient; their focus is safety; if seamless transition does not detrimentally impact safety assessment, unlikely to have any objection

Regulatory Agencies Review (2)

Outcomes (cont'd)

- Appropriateness of dose escalation in MAD
  - UK and Canada: no issue with planned multiple dose escalation or safety evaluation procedures
  - Sweden: no issue with planned multiple dose escalation or safety evaluation procedures (on condition that there is real-time safety evaluation)
  - Germany: make escalation procedures specific; ensure safety assessment continuous; add to rationale for 2nd dose criteria
- Appropriateness of futility criteria
  - UK: need to show minimal efficacious dose; accepted that the rule is conservative
  - Sweden: rule conservative
  - Germany and Canada: no objection

Top-Level Results

- Study enrolled and randomized 253 patients in one year
  - 10 Interim Analyses; Maximum time for an Interim Analysis - 3 days
- The criteria for the achievement of clinical POC for Drug X in Rheumatoid Arthritis were met
  - The top dose achieved the primary efficacy endpoint of ACR20 improvement at Week 16
  - Improvement over placebo was also observed for secondary endpoints: DAS28, ACR50, ACR70, and EULAR
- The study was successful and saved time and money

Key Benefits and Savings

- 9 month savings in overall Drug X development time
- Total cost savings by not undertaking a separate MAD study: ~$1.2m
- Site start-up savings (one *vs* two trials)
- IRB/EC/Regulatory Review (one *vs* two trials)
- Study close-out activities (one *vs* two trials)
- The first dose was dropped for futility – savings in total number of patients
- Back-filling in the lower dose levels during the MAD portion avoided enrollment pause
- Allowed flexibility in the strategy for site start-up
  - 22 centers randomized patients in the MAD part
  - 30 additional centers for the POC part

Main Conclusions

- A seamless adaptive phase MAD/POC study was undertaken in rheumatoid arthritis patients to rapidly and efficiently identify the optimum dose to take forward to proof of concept
- Regulatory agencies fully supported the study design
- The study met its objectives and saved a considerable amount of time and resource compared to a conventional study
- This study has significant cost/resource implications for the Sponsor's entire pipeline containing more the 25 programmes
- Adaptive seamless designs with multiple interim analyses are capable of being implemented and can provide substantial benefits
- Current implementation technology (e.g. AptivAdvantage™) would make execution of a complex seamless trials such as this even easier to undertake and provide additional efficiency benefits

Phase II Dose-Finding Study in Treatment of Post-Operative Nausea & Vomiting Breakthrough Following Prophlaxis – Determination of MED

Key Discussion Points

- Key Study Parameters
- Study Design
- Advantages of the Seamless Design
- Regulatory Agency Review
- Top-Levels Results
- Key Benefits and Savings
- Main Conclusions

Key Study Parameters

- Phase II, randomized, double-blind, active-controlled, adaptive, parallel design.
- 6 treatment arms
  - 5 single intravenous doses of Drug X (6, 12, 18, 24 or 36 mg)
  - Control: single intravenous doses of active
- Treatment duration: 24 hours or discharge whichever is sooner
- Primary endpoint
  - Complete response: no emesis/further rescue medication from 10 minutes post-infusion up to 24 hours or discharge from hospital whichever is sooner
- Model: a change-point logistic
  prior information on control arm
  Bayesian analysis
- Randomization
  - Stage 1: 1:1:1:1:1:1 randomisation
  - Interim analysis: includes test for futility
  - Stage 2: unequal depending on shape of dose-response curve

Study Design

STAGE 2

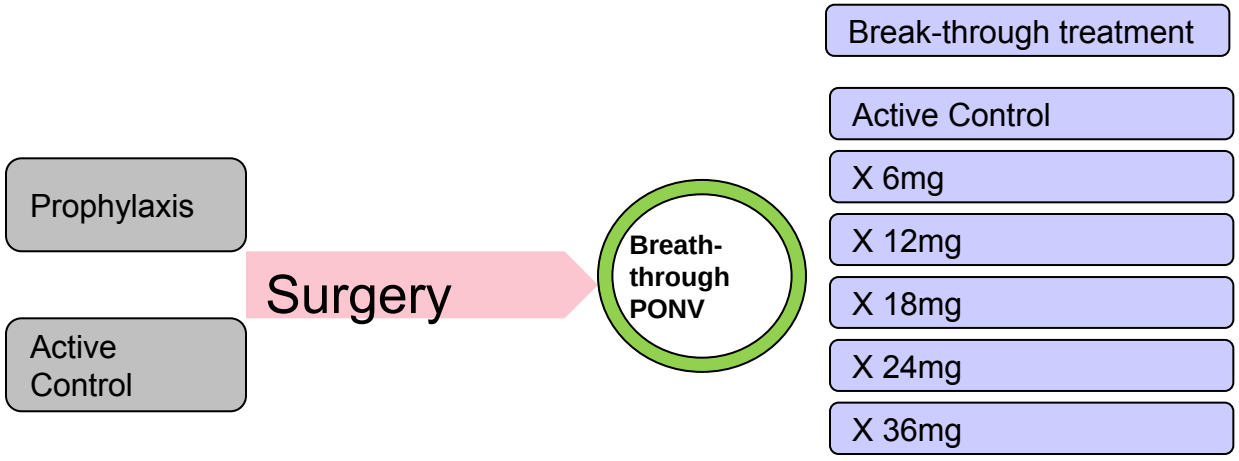
- Active Control
- X ? mg
- X ? mg
- X ? mg

- **Doses and randomisation ratios depend on the shape of the dose-response curve**

Interim

- **when at least 18 / arm have completed the study**
- **enrollment will be stopped if any of the following criteria are met**
  1. **% CR on any dose of X <50% if active > 30%**
  2. **% CR on active > 80%**
  3. **For all doses of X Post Prob ($p_{X(i)} > p_{active}+0.2$) < 0.30**
- **choose doses for stage 2**

STAGE 1



Advantages of the Design

- Original design was a fixed sample design.
- Introduction of an interim:
  - Allows testing of the assumptions.
    - the prior distribution
    - Effect sizes
    - Early stopping
- Minimizes the number of subjects that are exposed to ineffective doses of Drug X
- Focuses subjects to doses that are most informative for accurate determination of the MED.

Regulatory Agencies Review

- Regulatory Agencies consulted
  - FDA, UK, Germany, Poland, Russia, Ukraine.

- European agencies raised questions mainly about CMC, QP related and labeling
- FDA raised some questions about the prior distribution and its impact. They were nit concerned with the adaptive nature od the study.

Top-Level Results

- Study enrolled 121 patients in 4 months
- The 3RD criterion for futility:
  - For all doses of X Post Prob ($p_{X(i)} > p_{active}+0.2$) < 0.30
- The criterion was met and the study was stopped.

Key Benefits and Savings

- 2 month savings in overall Drug X development time
- $800,000 (out of $4.2m) savings made on CRO budget
- $290,000 (35% of budget) savings made on investigator fees

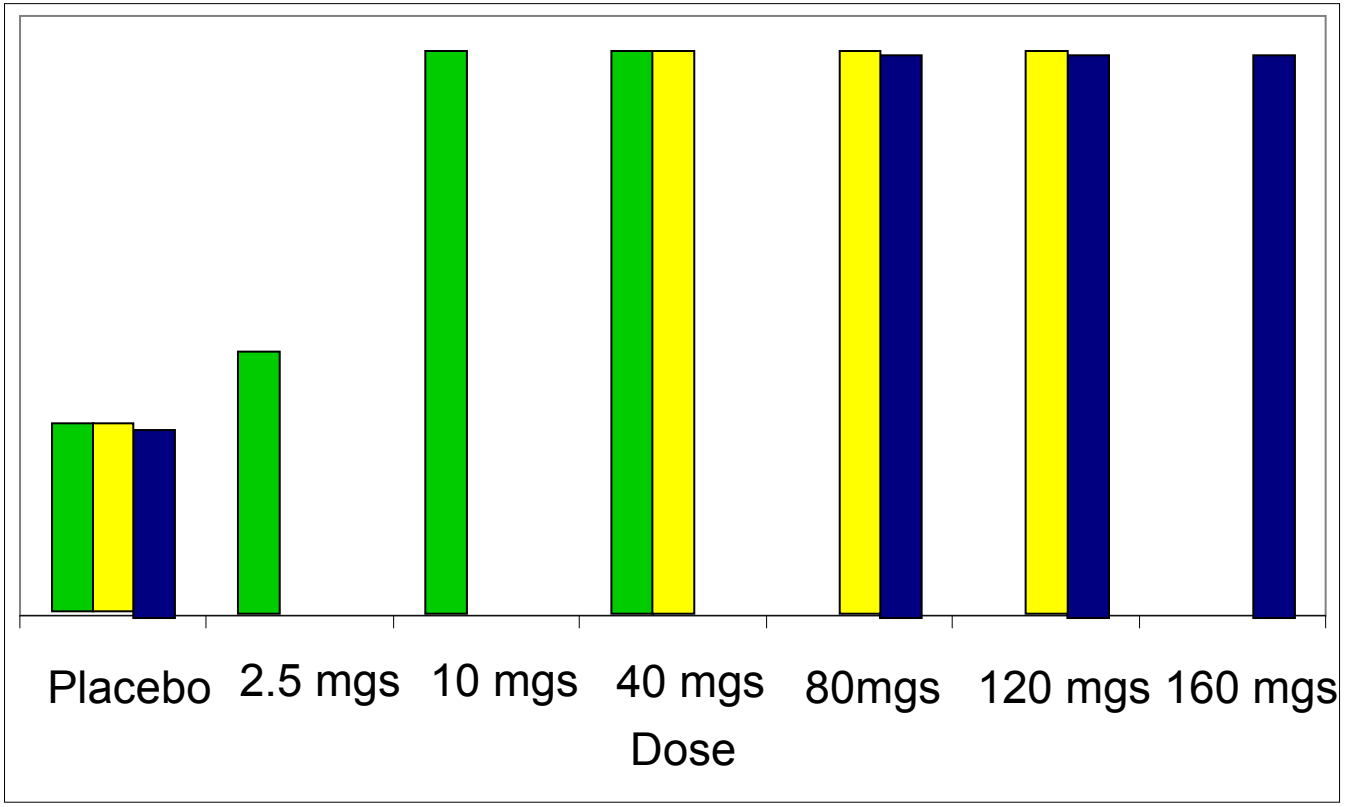
Main Conclusions

- An adaptive phase II was undertaken in Post-Operative Nausea & Vomiting to determine the MED.
- Regulatory agencies raised no issues on the adapgtive nature of the design.
- The study met its objective to stop early in the case that no dose had sufficient efficacy
- The introduction of an interim allowed the early stopping of this study with consequent savings.

Outline

- Introduction and taxonomy of clinical trial designs
  - Pre-1990's
- Basic principles of adaptive designs
  - Allocation rule
  - Sampling rule
  - Stopping rule
  - Decision rule
- Phases of development
- Adaptive designs for the learn phase of drug development
  - First-in human / MTD
  - Two-stage designs
  - Adaptive dose-ranging designs
  - Bayesian adaptive randomisation
- Adaptive designs for the confirmatory phase of drug development
  - Sample size re-assessment
  - Adaptive group sequential designs
  - Seamless phase II/III designs
  - Population enrichment designs
- Practical aspects of adaptive design implementation
- Discussion

Phase 2b Dose Selection Design Circa 1993

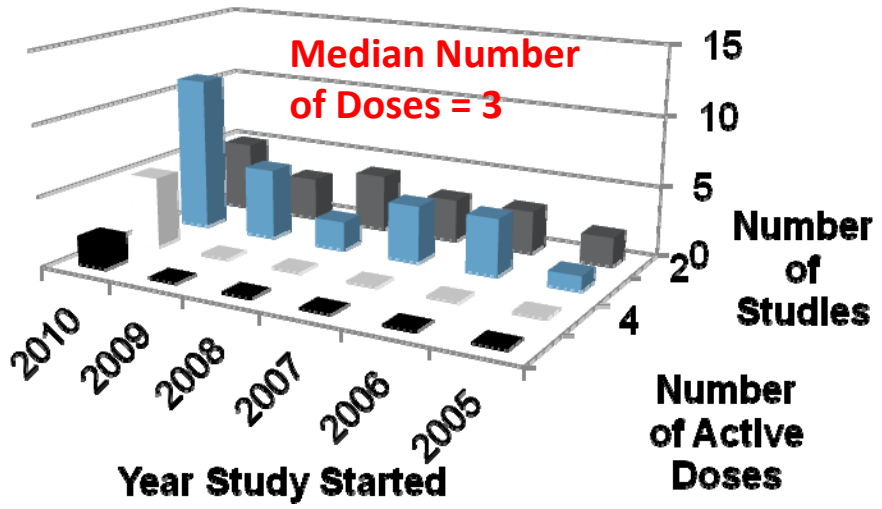


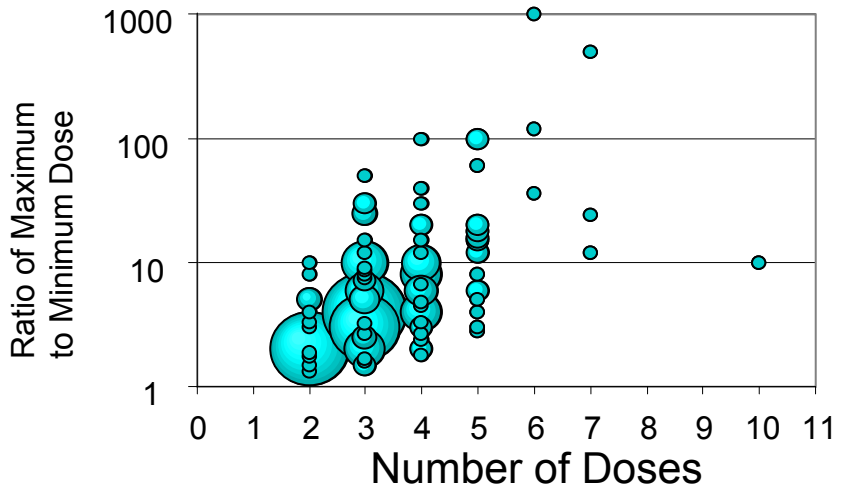
- More Efficient
  - wide range of doses, smaller numbers of patients per group
  - followed by one large parallel group study focusing on the doses showing promise in exploratory study.

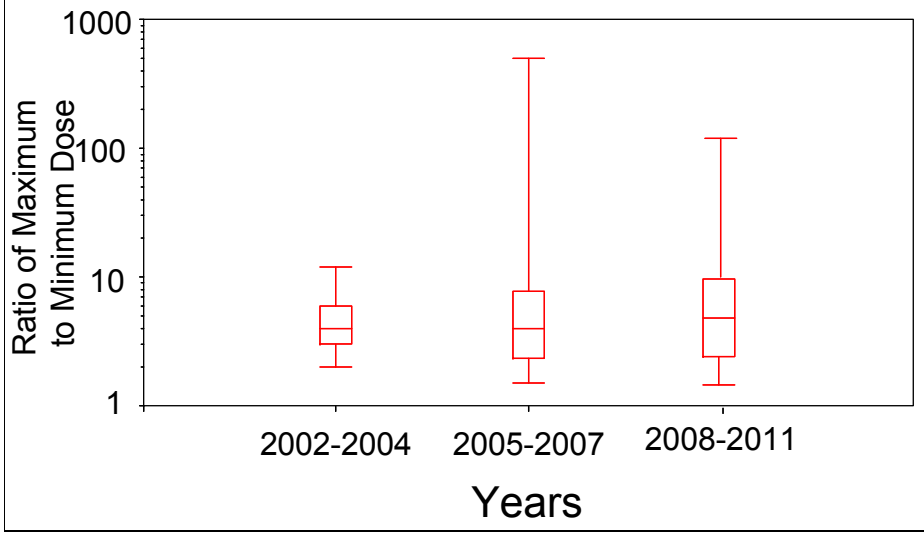
Comparison Between Successful and Unsuccessful Phase II Programs

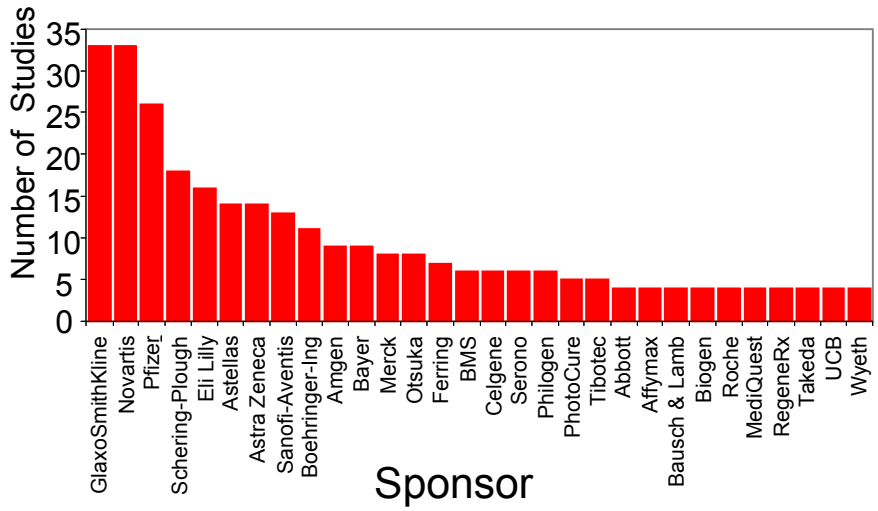
Initial Dose Finding Unsuccessful - More Studies Required			Initial Dose Finding Successful	
Study	Initial Dose Range	Total Dose Range Examined	Study	Dose Range Examined
1	4	64	1	40
2	1	4	2	8
3	6	16	3	4
4	4	8	4	10
			5	4
Median	4	12	Median	8

Phase 2b Dose Selection Design
Data Extracted from www.clinicaltrials.gov

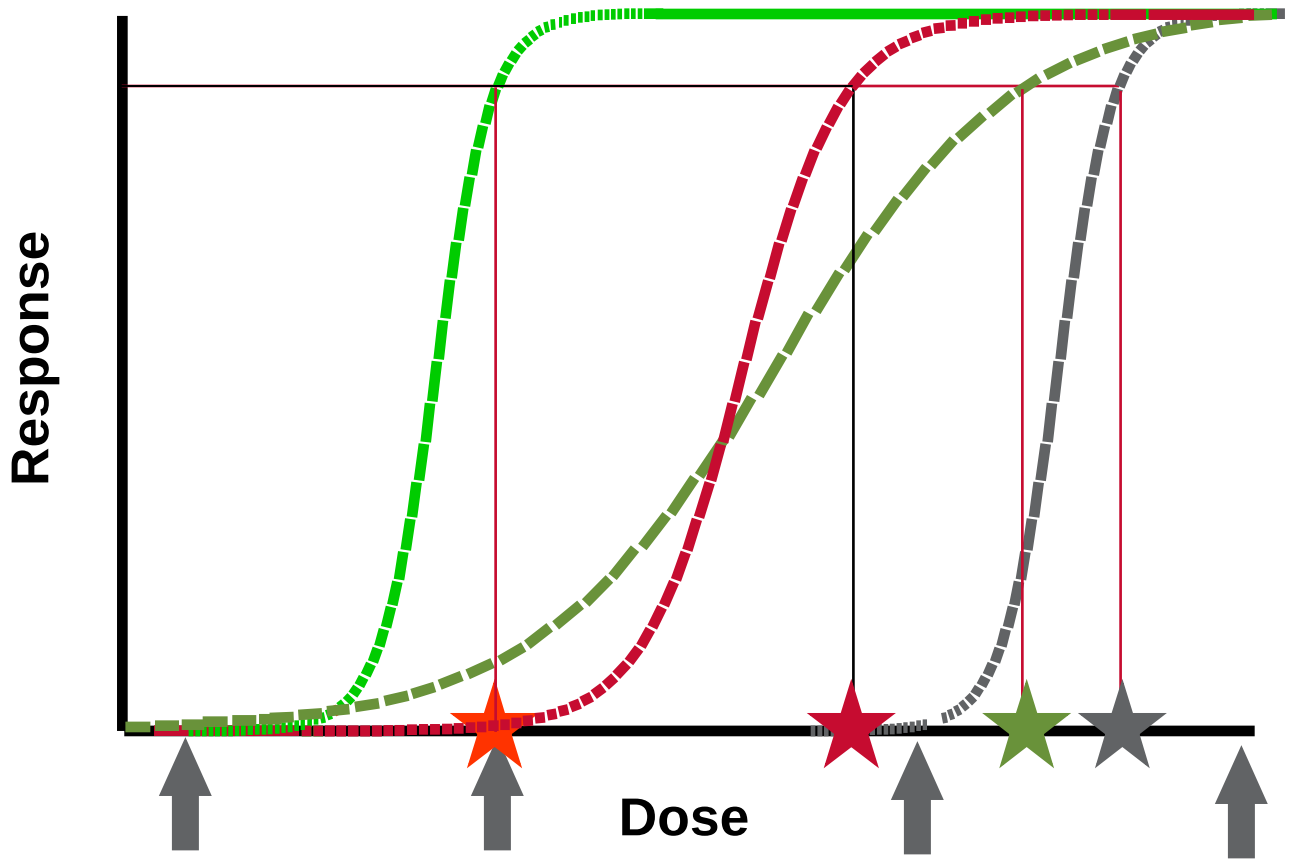




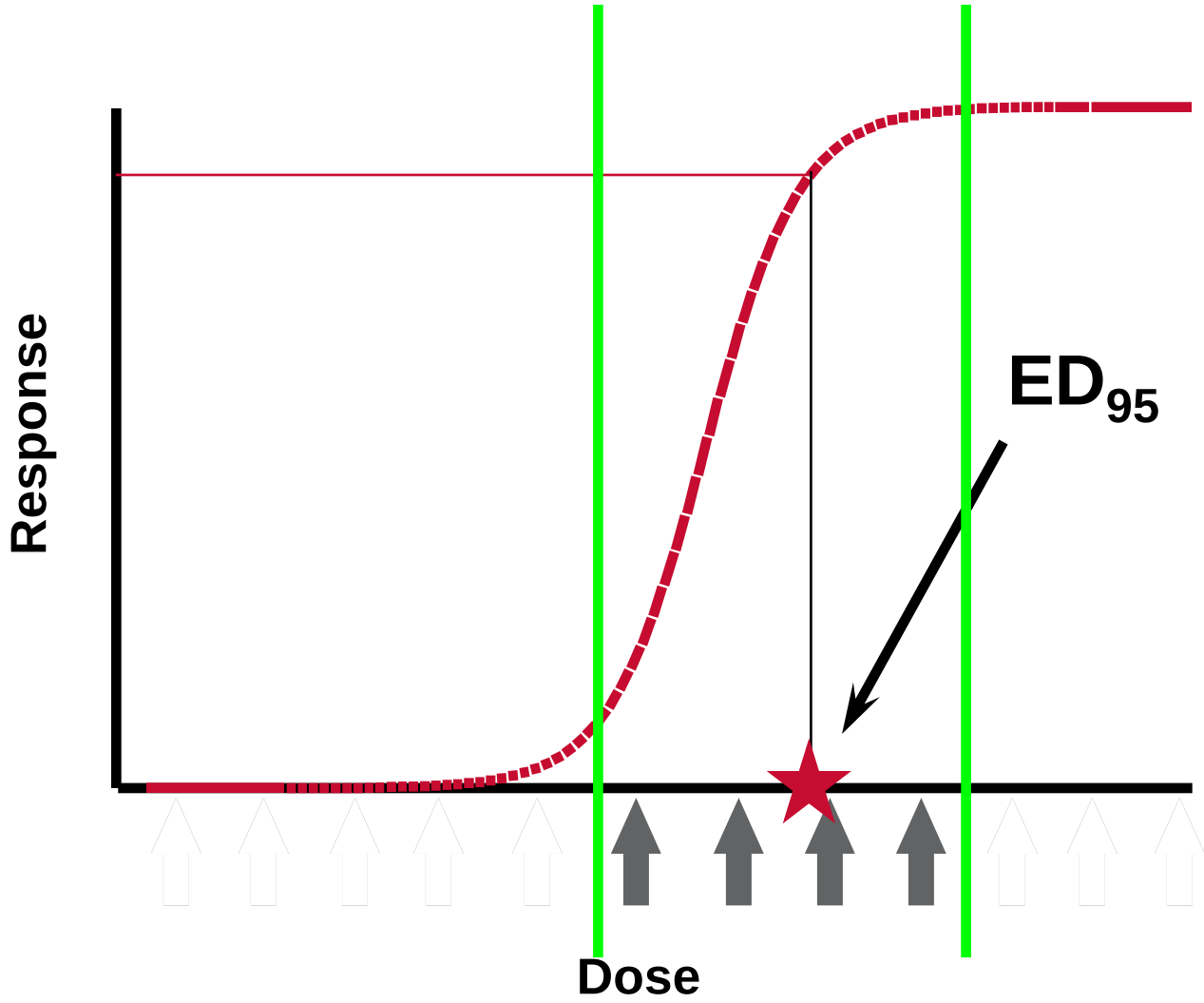




Issues in Dose Selection
Standard design



Issues in Dose Selection Increase Number of



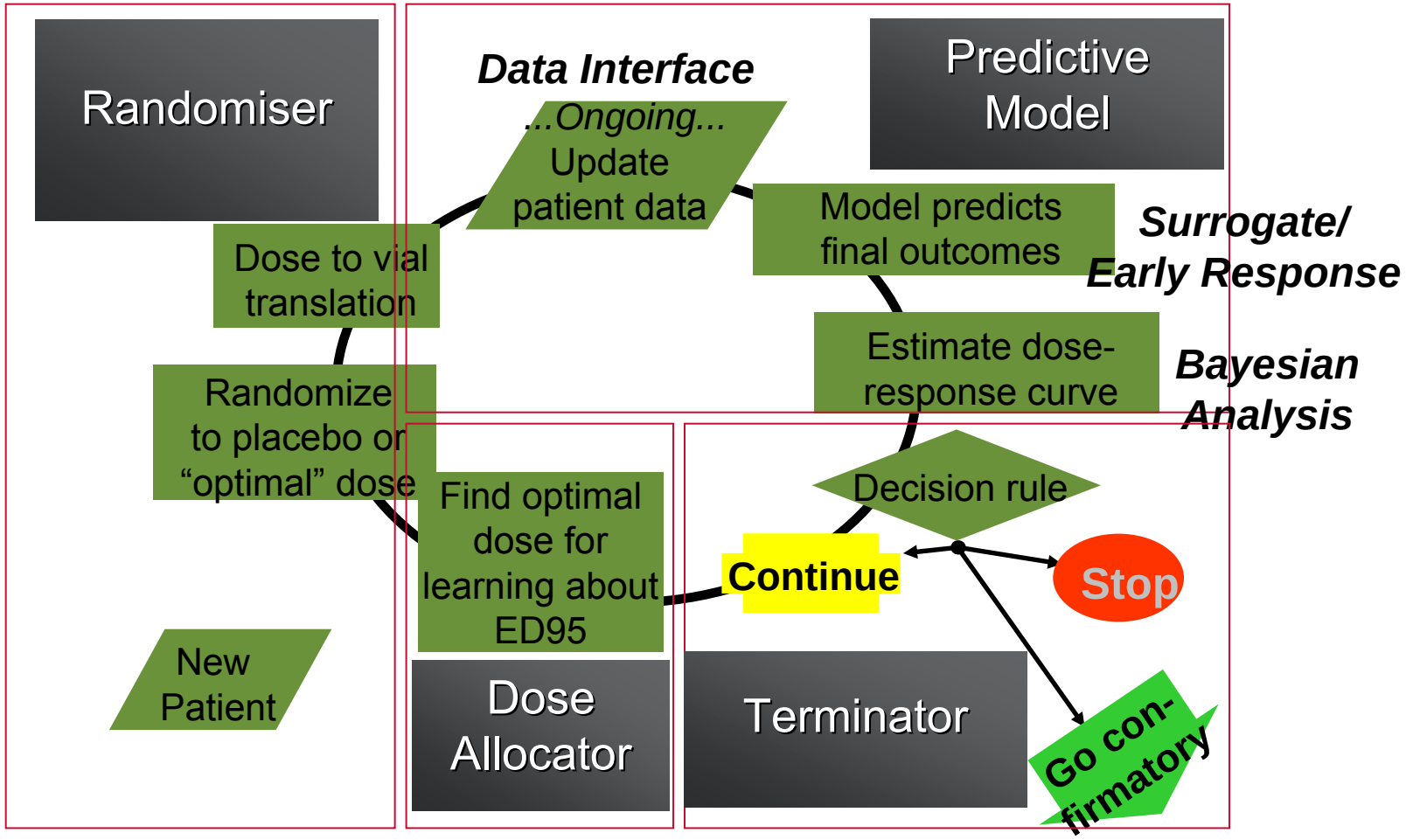
Improvement to Standard Design

- Increase number of doses - considerably
  - Improves chance of successfully learning about dose response
  - May be expensive if numbers of patients allocated to each dose is too large
  - Learning trials (Sheiner)
- Prevent allocating patients to ineffective doses
- Borrow strength from neighbouring doses

Improvement to Standard Design

- Solution - *ADAPTIVE DESIGN*
  - More accurate information with less resources - large resources only when necessary
  - Dose-finding : is there a dose with sufficient efficacy to take into a confirmatory trial ?
  - Go/no-go
  - Futility analysis / early decision making
  - How ?
    - Based on Bayesian statistics

Design Process

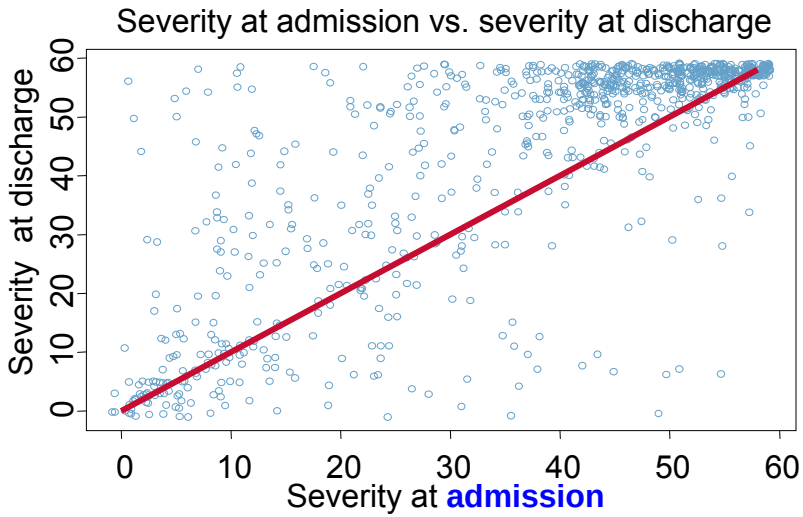


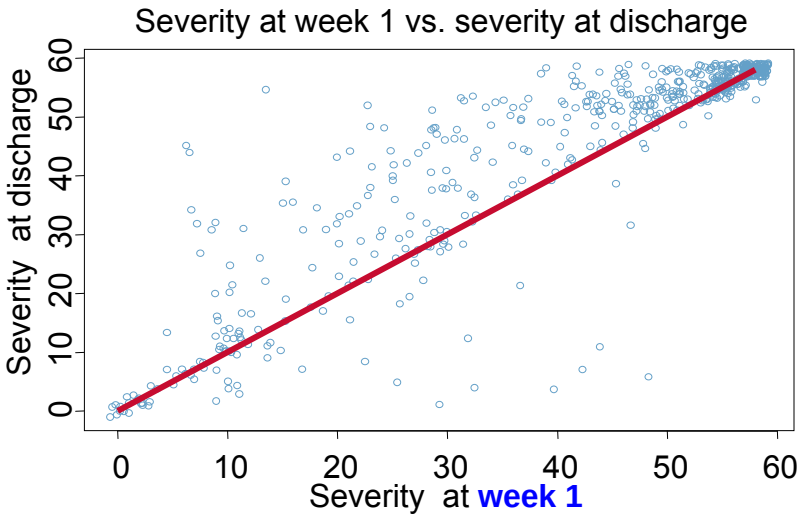
Issues

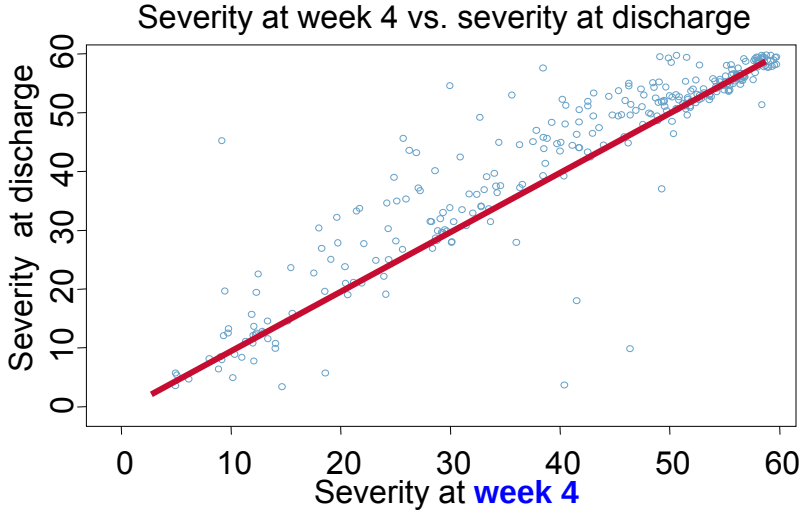
- How do we predict ?
  - Longitudinal model
- How and what do we update ?
  - Dose response
  - Longitudinal model
- How do we model response ?
- Decision Problems
  - How do we choose a dose ?
  - How do we stop ?

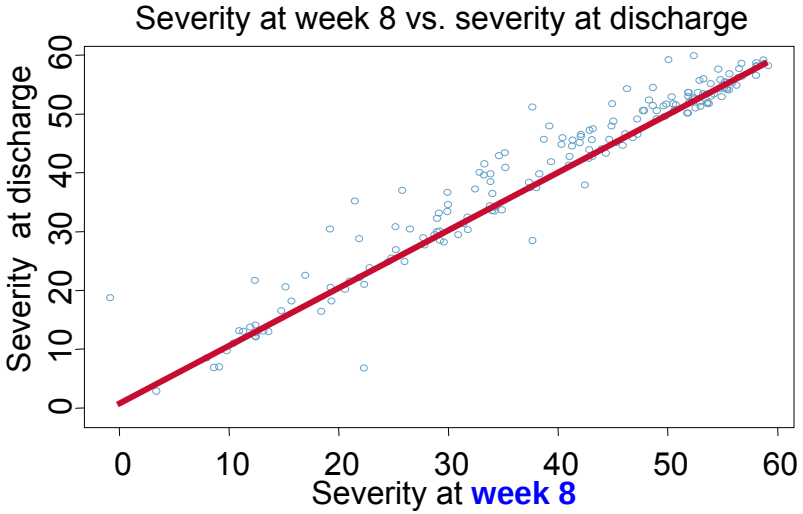
Building Predictive Models

Data from Copenhagen Stroke Data-Base









Two Decision Problems

- Dose Allocation: allocate dose z^0 to next patient to "learn most about the curve."

  (Non-linear sequential optimal design) – utility (gain function) – maximise expected utility (EASY)

- Optimal stopping: at each period decide
  - A0 stop and abandon the drug
  - A2 stop and switch to confirmatory phase
  - A1 continue

HARD

Decision Problem 1 : Dose Allocation

- df (z,θ) advantage over placebo at dose z, using a curve parameterized by θ.
- z^* dose which achieves 95% of possible improvement over placebo (ED95)

- Utility $u(z, \tilde{y}, x, \tilde{D}, D) = -Var[df(z^*, \theta) \mid D, \tilde{D}, \tilde{y}, x, z]$

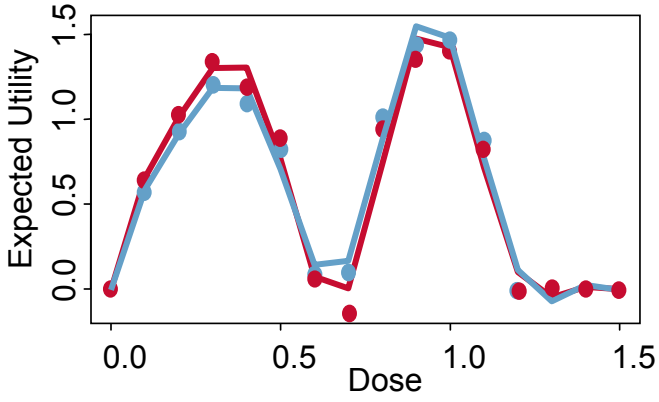
- Where :
  - x : covariates of a new patient
  - z : the assigned dose
  - $\tilde{y}$: predicted response of a new patient
  - D : data
  - $\tilde{D}$: missing data (missing final response)

- Expected Utility :

$$U(z, x, D) = \int_{\tilde{y}, \tilde{D}} u[z, \tilde{y}, x, \tilde{D}, D] \times p(\tilde{D} \mid D)\, p[\tilde{y} \mid D, z]\, d\tilde{D} d\tilde{y}$$

- **Substitute average value:** $x \equiv \bar{x}$

- **Approximate Expected Utilities :**

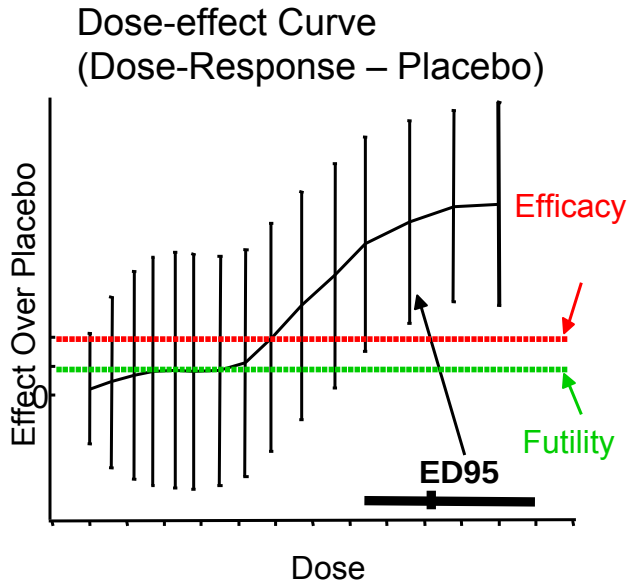


- Maximise expected utilities as a function of dose

Decision Problem 2 : Early Stopping

- Stopping criteria based on posterior probabilities of clinically meaningful events (Thall, Simon and Estey, 1995; Thall and Russell, 1998)
- Bayesian m-step look ahead procedure (Whitehead and Williamson, 1998)
- Vlachos and Gelfand (1998)
- Bayesian decision-theoretic myopic design

- Probabilities of clinically meaningful events



Stopping rules & utilities

- In the ASTIN study we considered a decision theoretic approach to the choice of dose.
  - But this depended on getting agreement >12 months before the study on the utility per point improvement on the endpoint scale and other factors.
  - Also very difficult to deconvolute this for answering “why are we stopping?” type of questions.

A pragmatic approach

- In the end we used a decision criteria based on the limits of an interval estimate.
- Much easier to express.
  - "If X% of the distribution of this parameter is above the clinically meaningful difference then we stop and claim an effect."
  - "If Y% of the distribution is below this level then we stop and conclude that it is unlikely we will ever show that magnitude of effect."
- Choose X, Y (typically 80% and 95%).

The Dose-Response Curve : $\theta_i = f(z_i, \theta)$

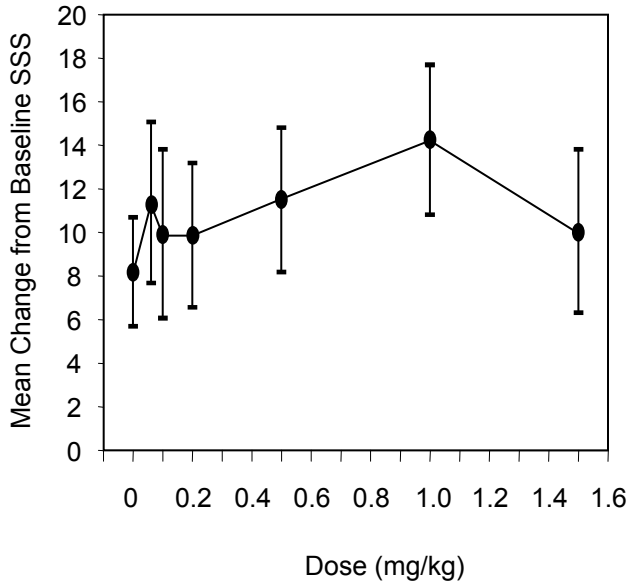
Mean Response

Dose

Parameters

- Requirements
  - To model f (z ,θ) we need :

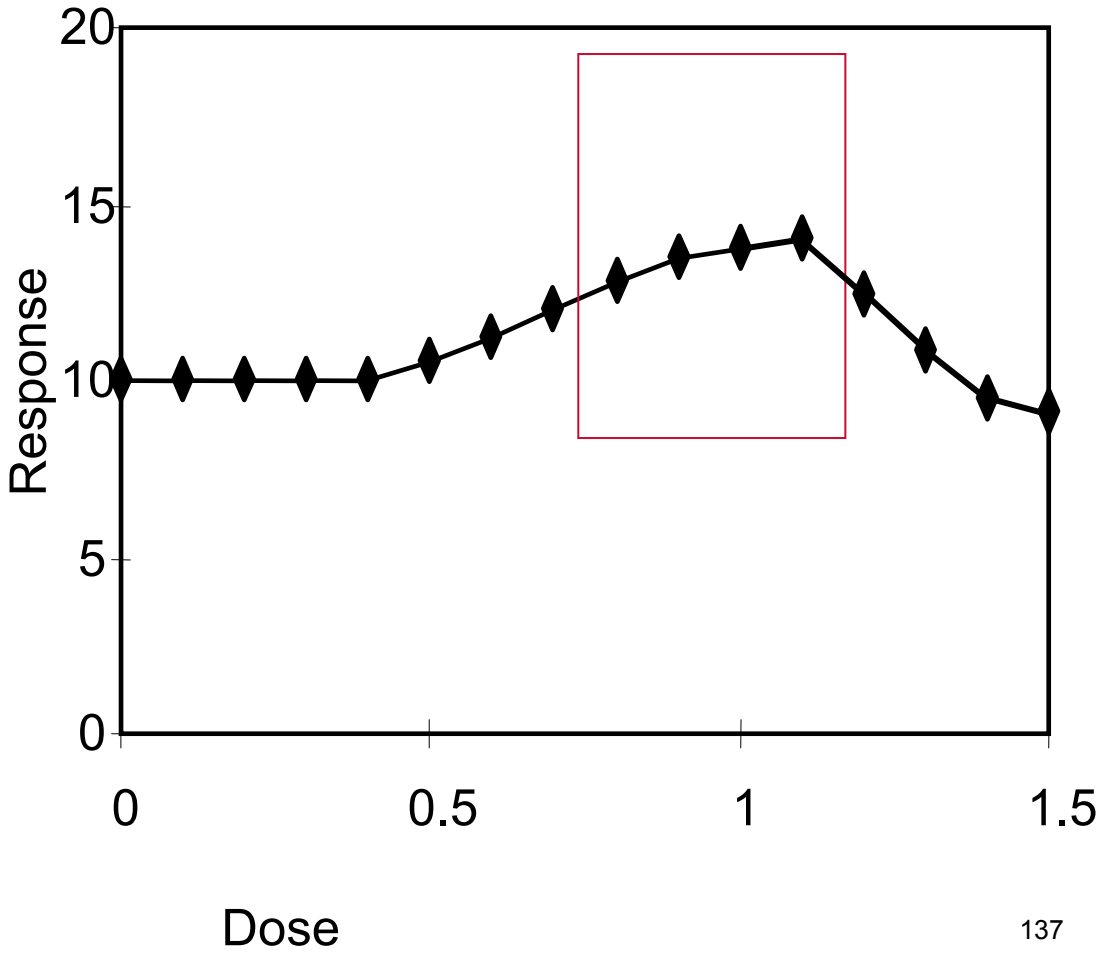
1. a flexible model, allowing non-monotone curves. and allocator)
2. analytical posterior updating (simulation required for terminator and allocator)
3. efficient (analytic) computation of expected utilities



- Possibilities
  1. Splines
  2. Kernel Regression
  3. Normal Dynamic Linear Model

Modelling Dose Response

- We model f (z , θ) as a 2nd order polynomial NDLM (West and Harrison 1997):

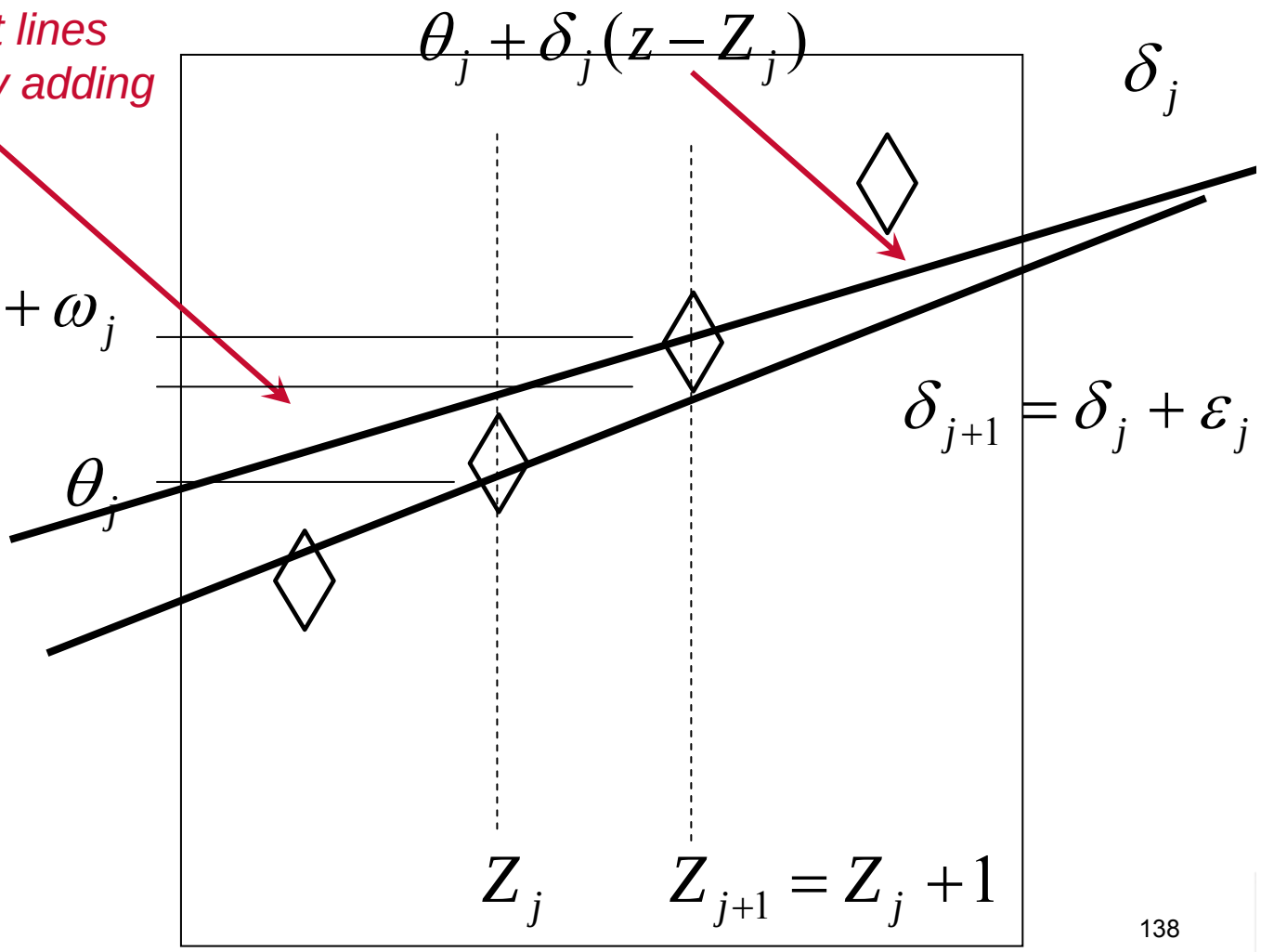


2nd Order Polynomial NDLM

Locally around z = Z_j a straight line with level θ_j and slope δ_j

Parameters of the straight lines change between doses by adding a (small) evolution noise.

$$\theta_{j+1} = \theta_j + \delta_j + \omega_j$$
$$\theta_{j+1} = \theta_j + \delta_j$$

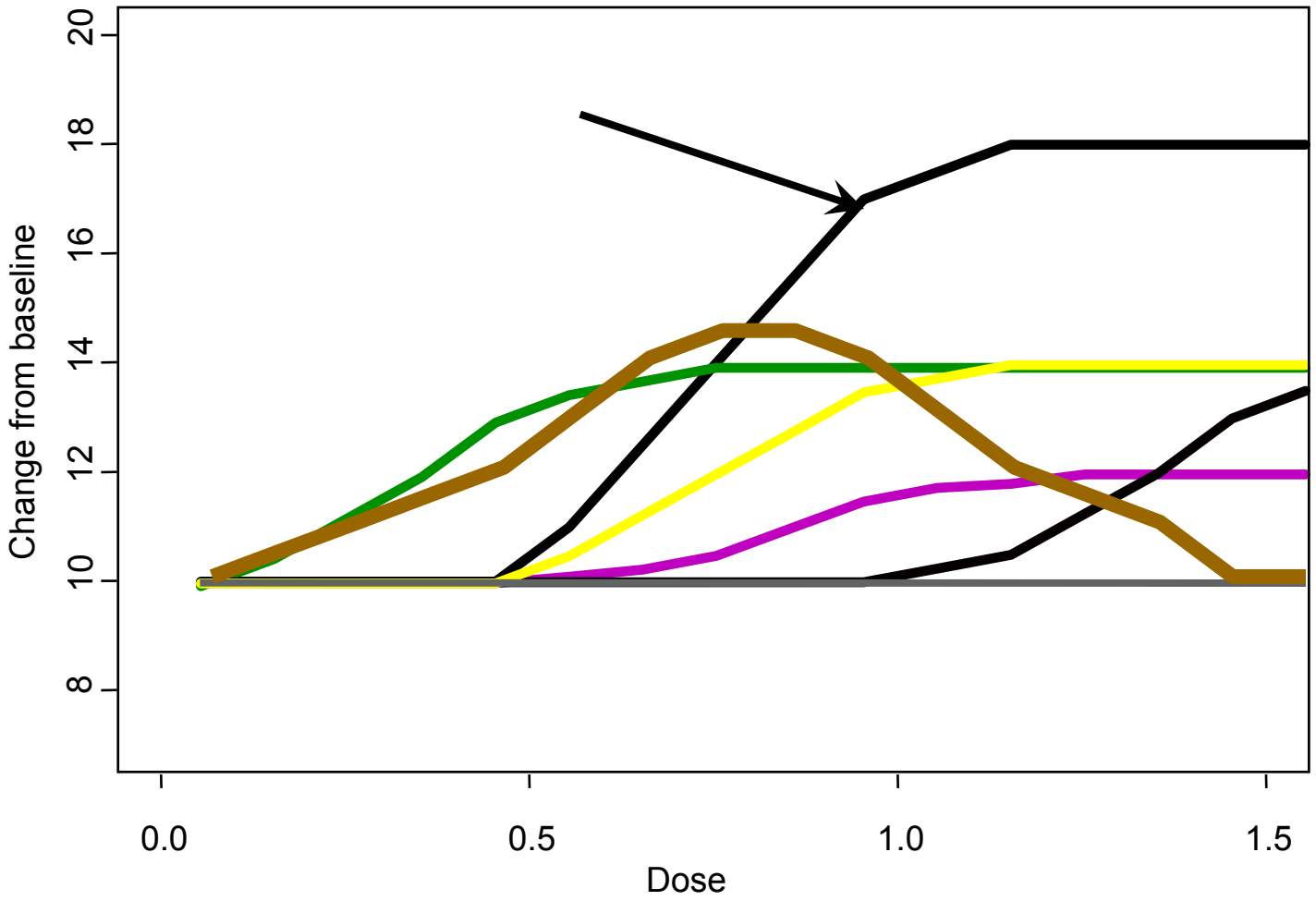


Evolution Variance = Smoother

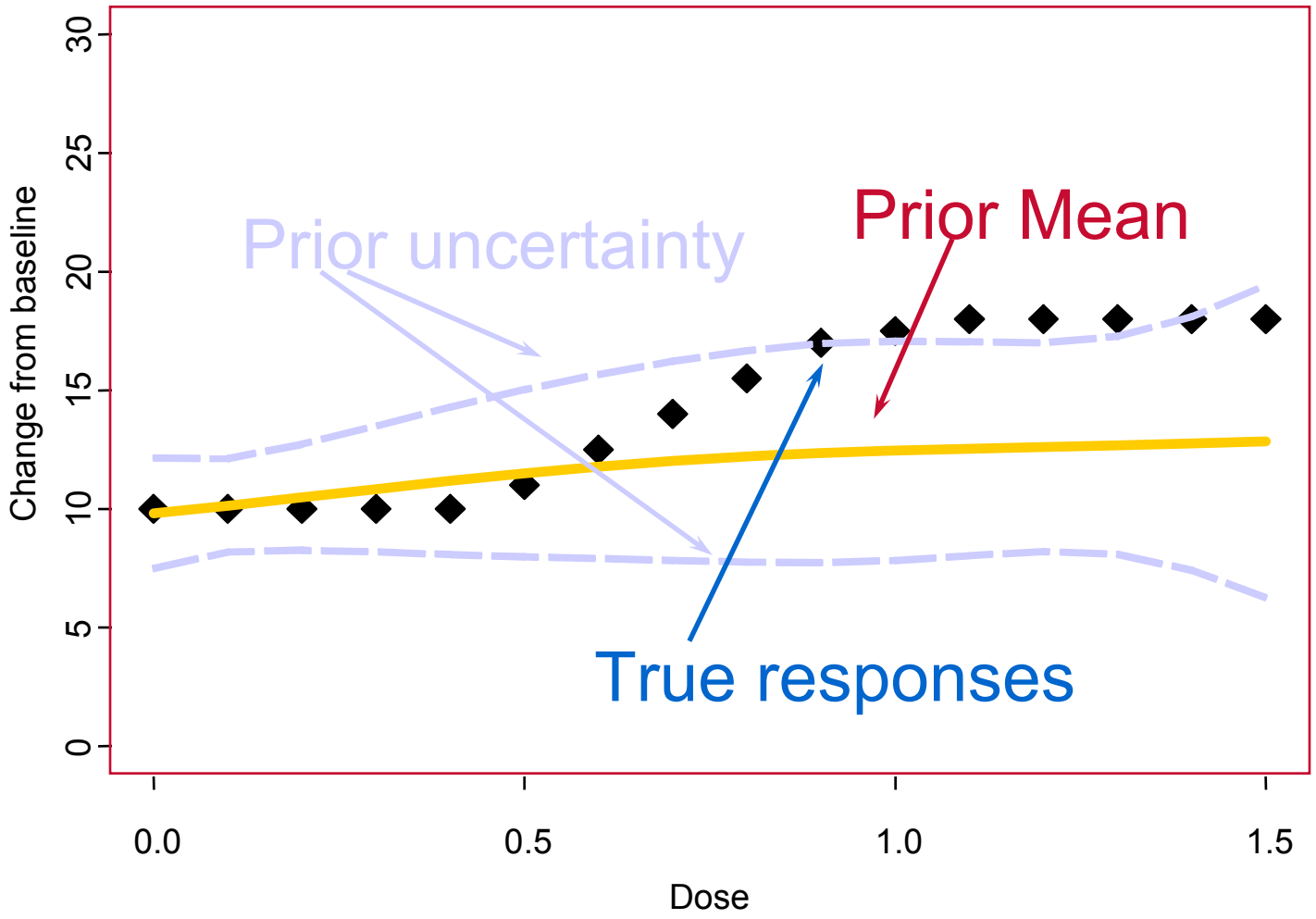
Purpose of Simulations
Key Questions

- Does algorithm accurately estimate d/r curves?
- Does adaptive allocation result in sensible" choices of doses?
- Do we "learn" about the d/r curve?
- How soon can we stop?
  - Ineffectual drug: Stop early
  - Effective drug: Move into confirmatory trial
- Do we benefit over a "traditional" design?
  - Smaller overall sample size
  - In what cases do we not win?
- What aspects are we comparing?
  - Type I error
  - Power to detect clinically meaningful difference
  - Does the algorithm make the "right decision"

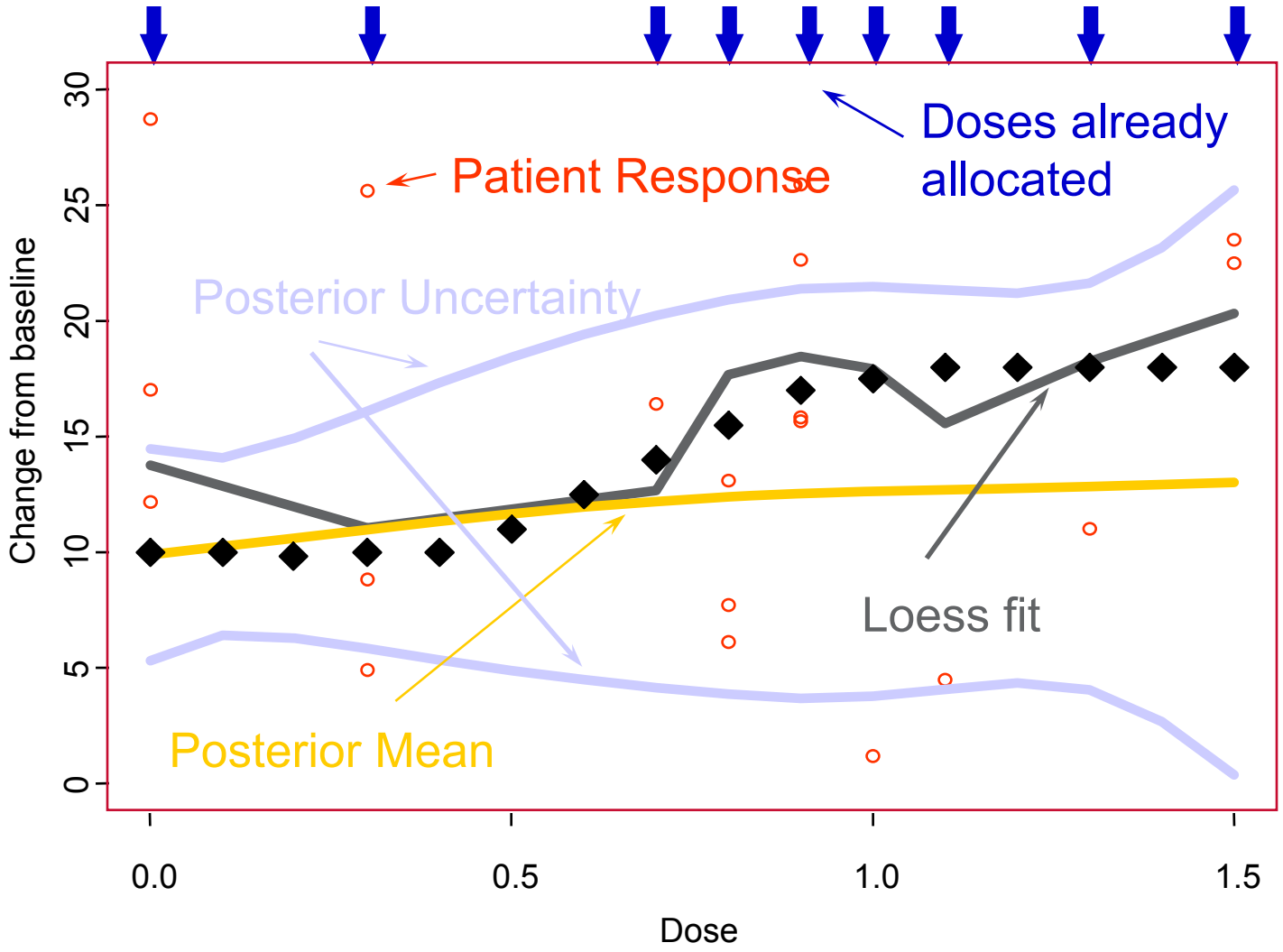
Simulated Dose Response Curves

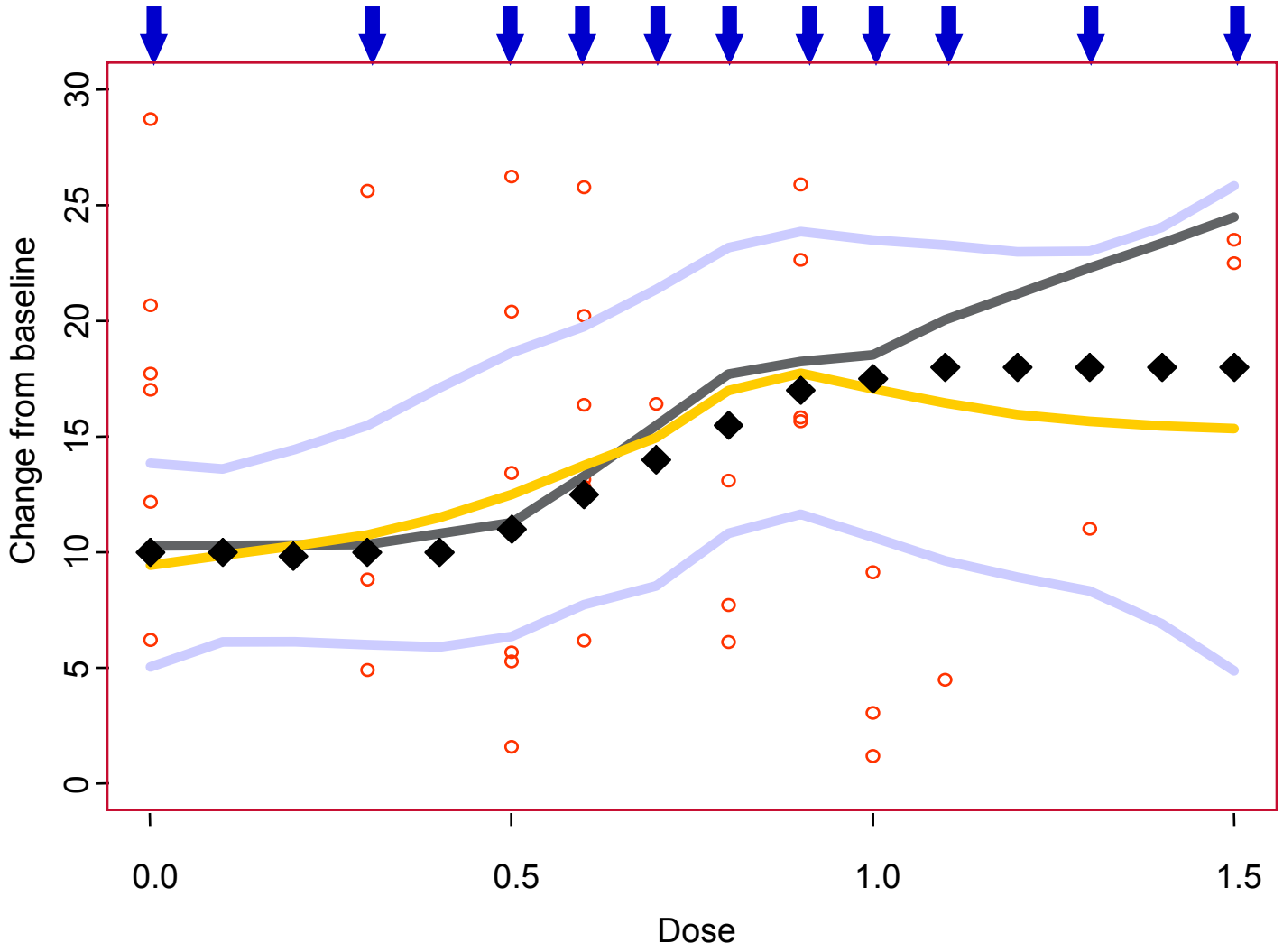


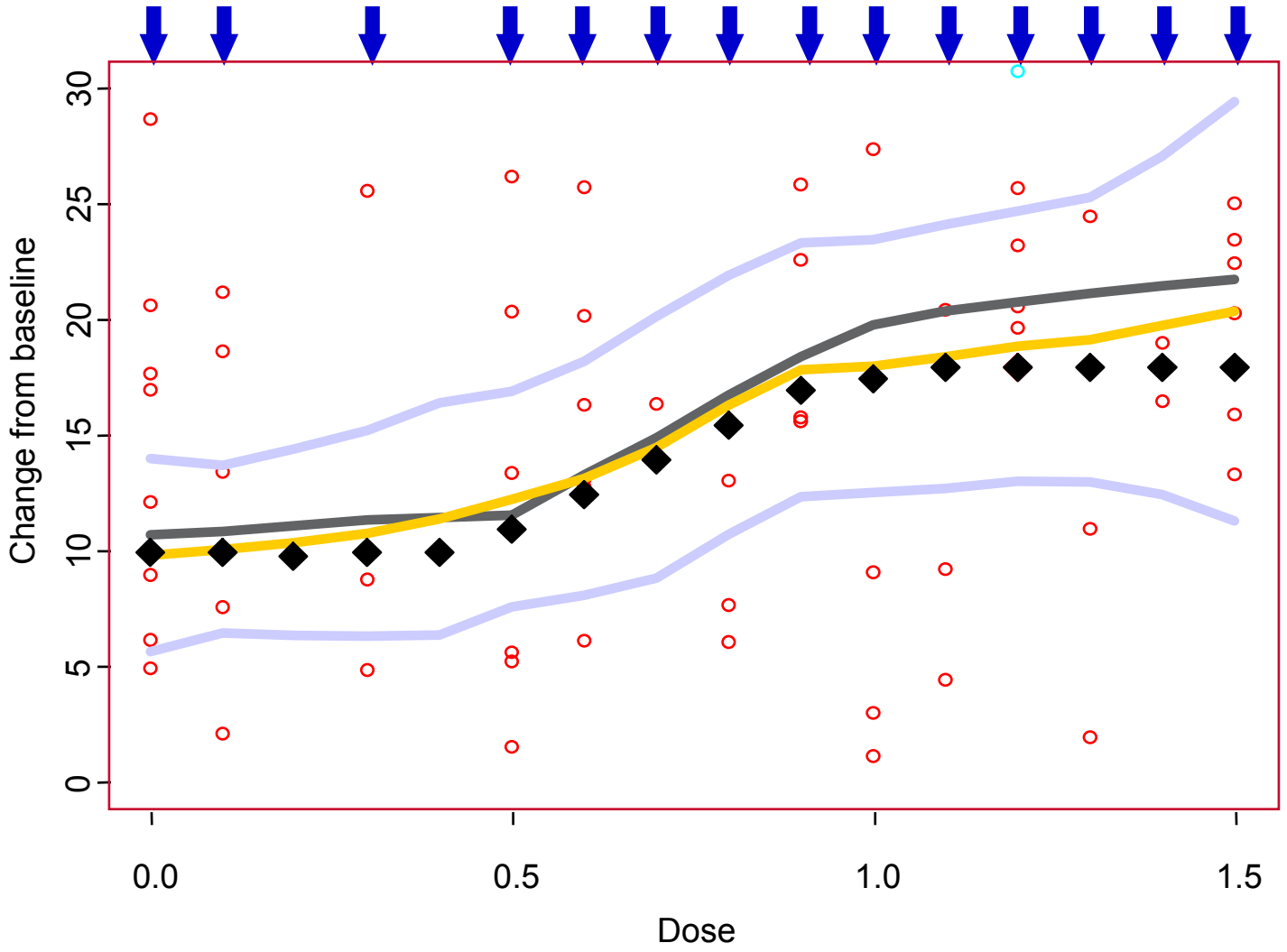
Prior Information

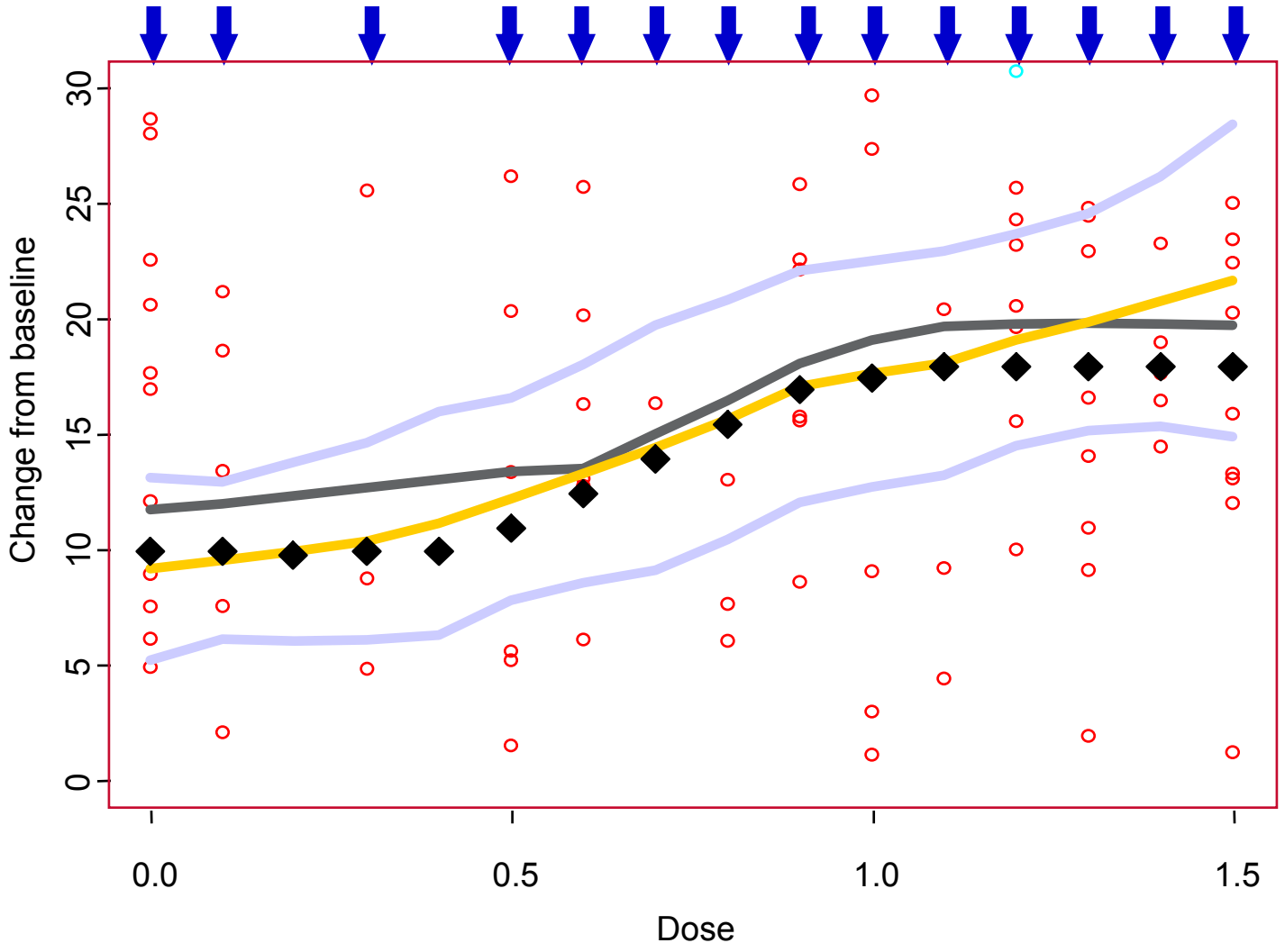


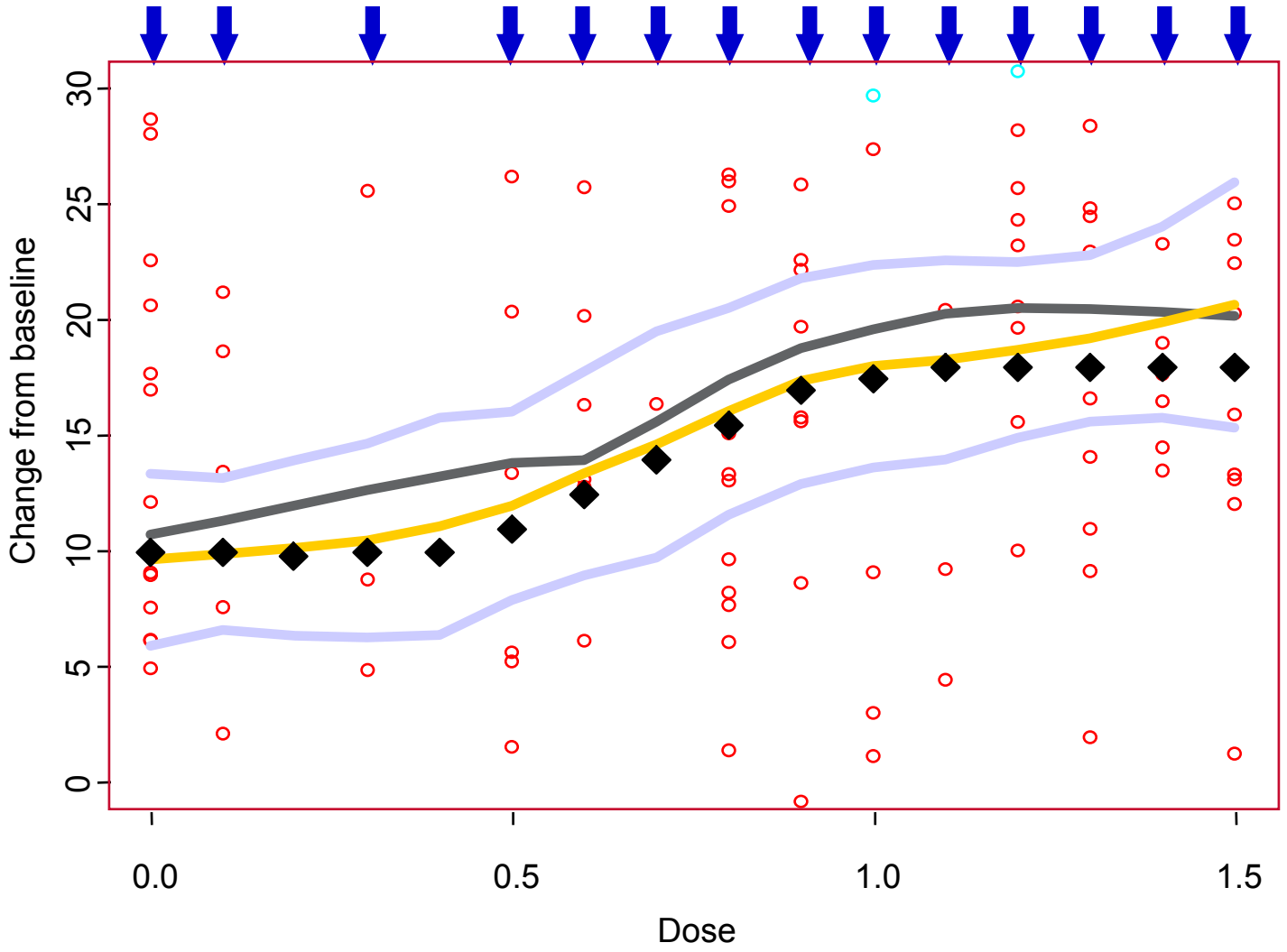
Data from 25 Patients

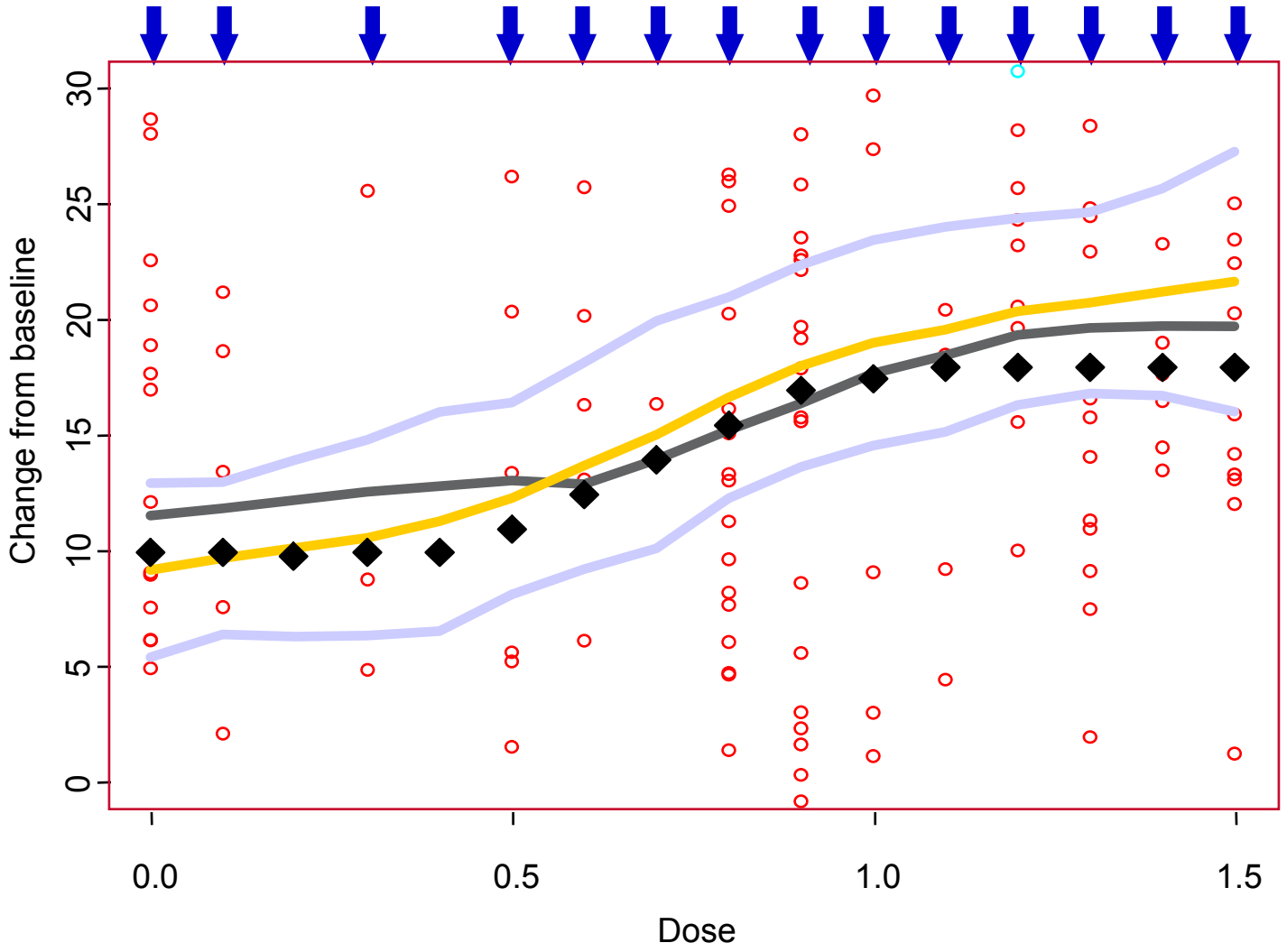


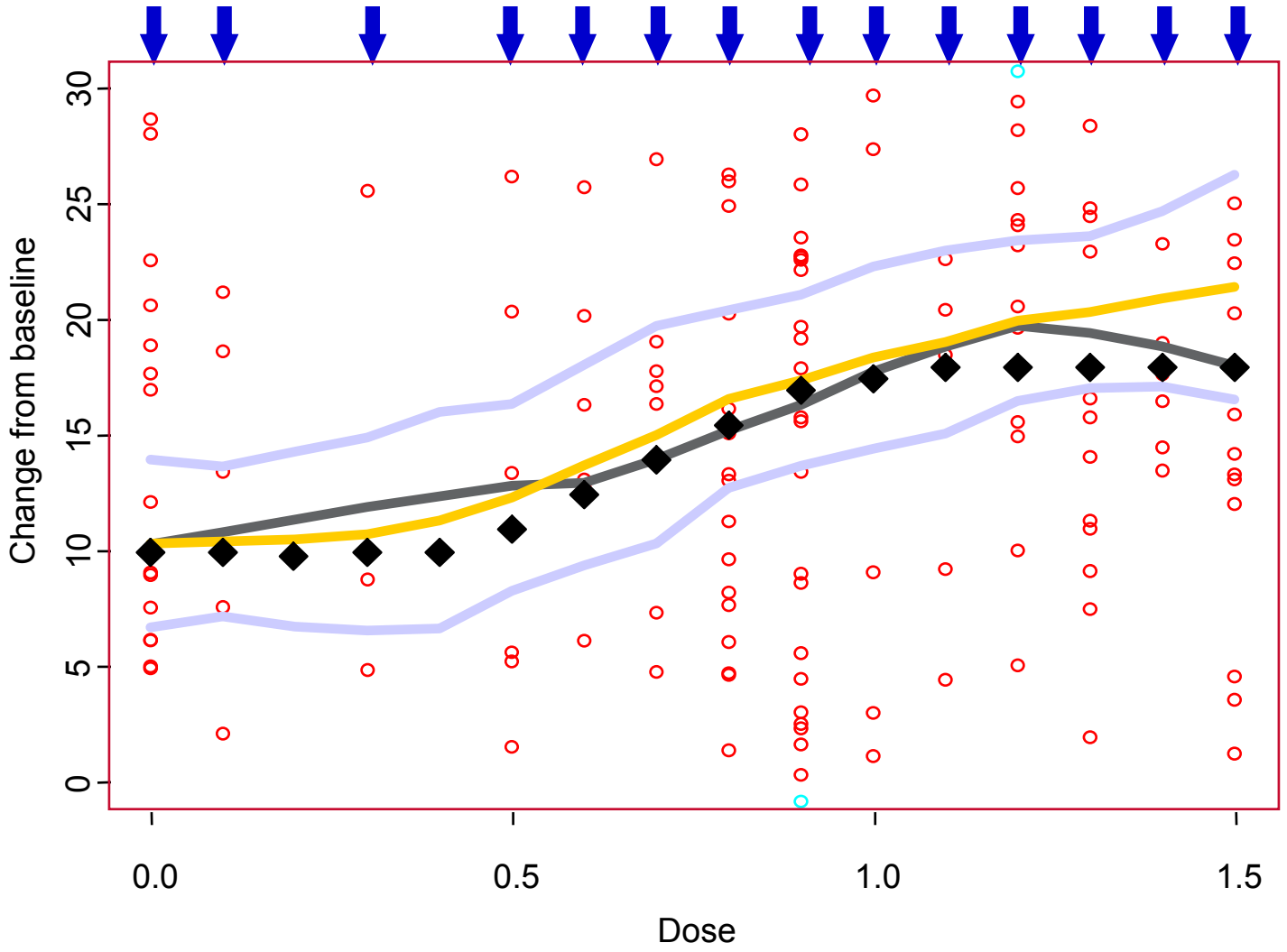


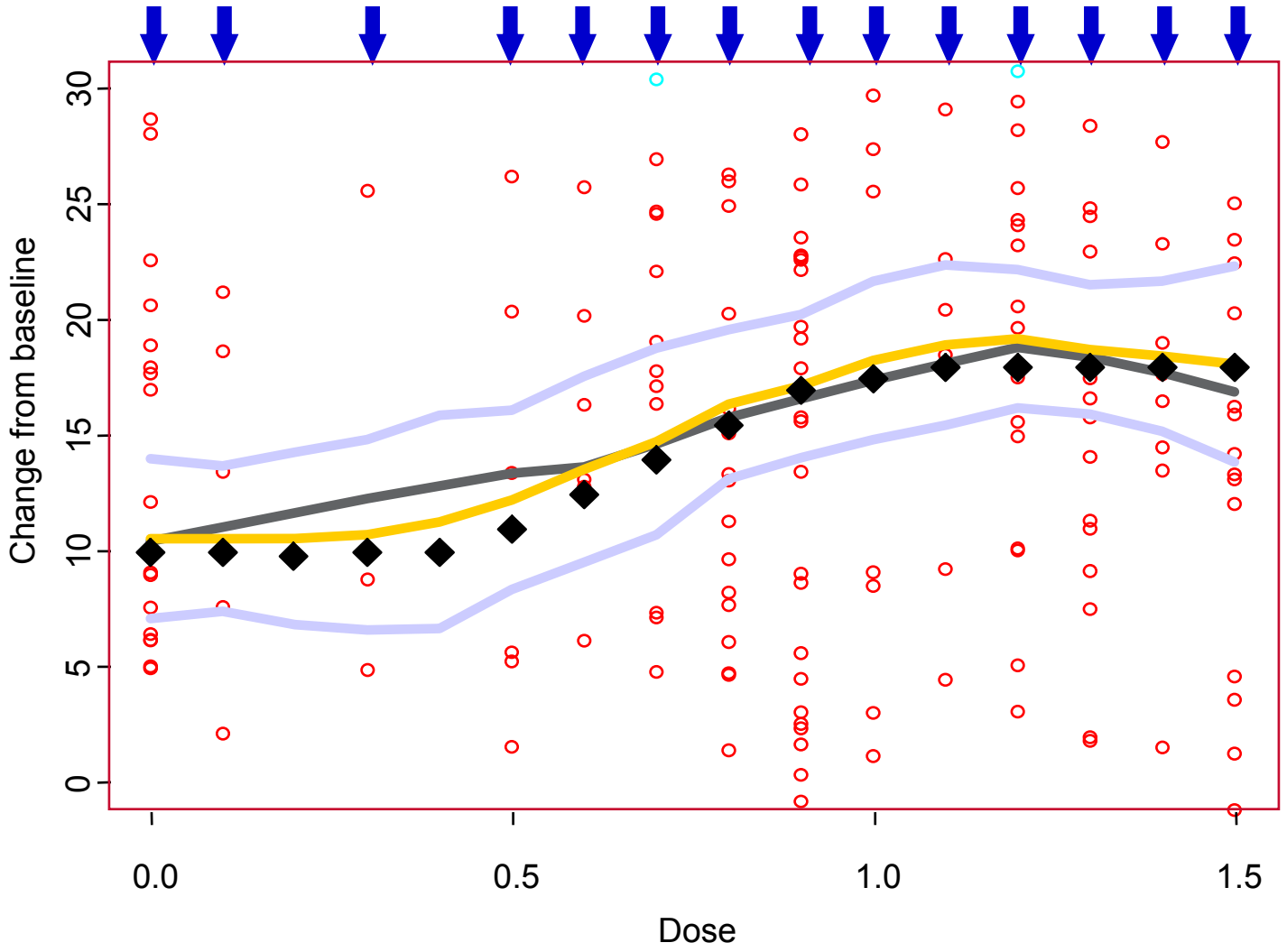


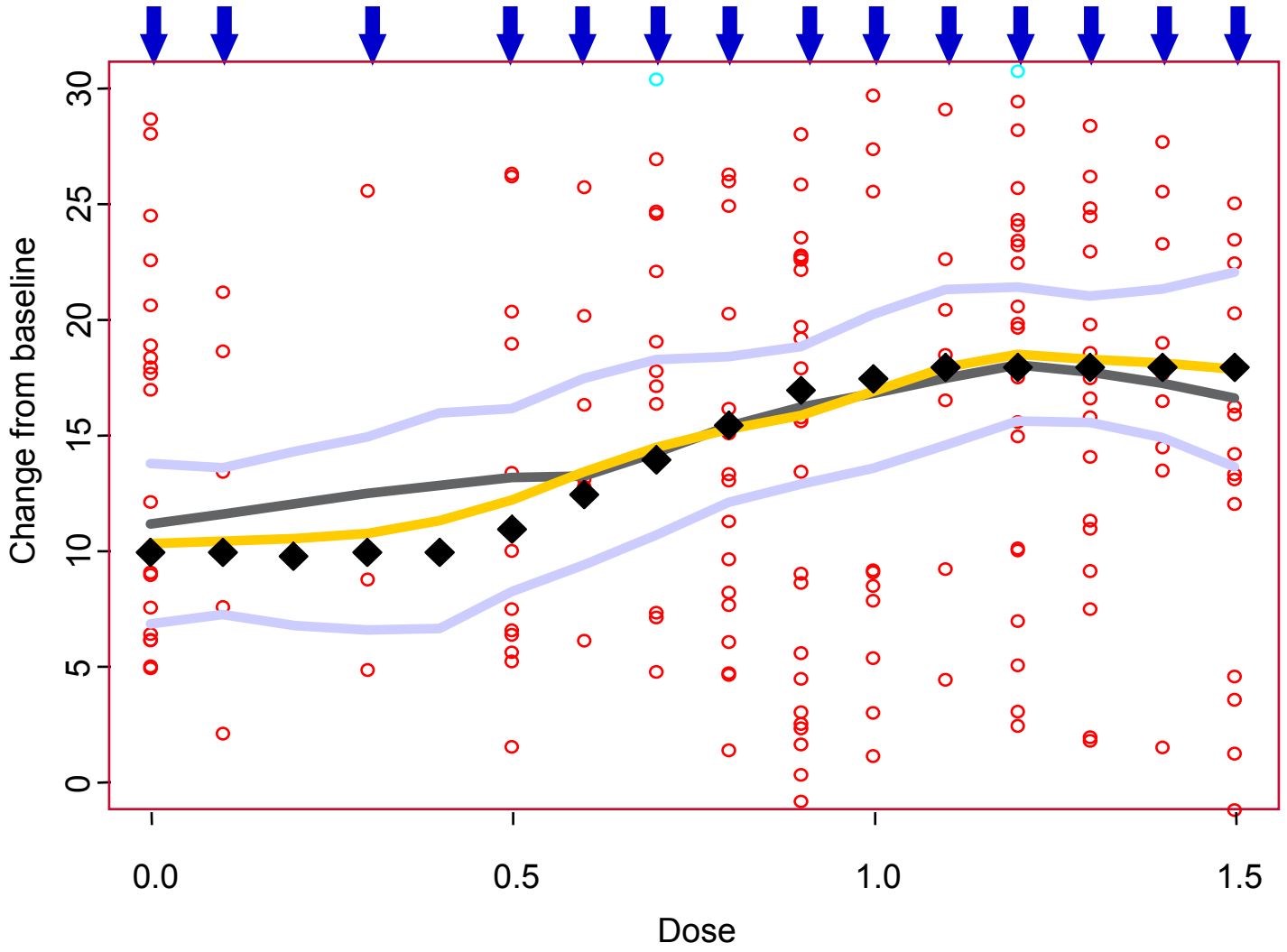


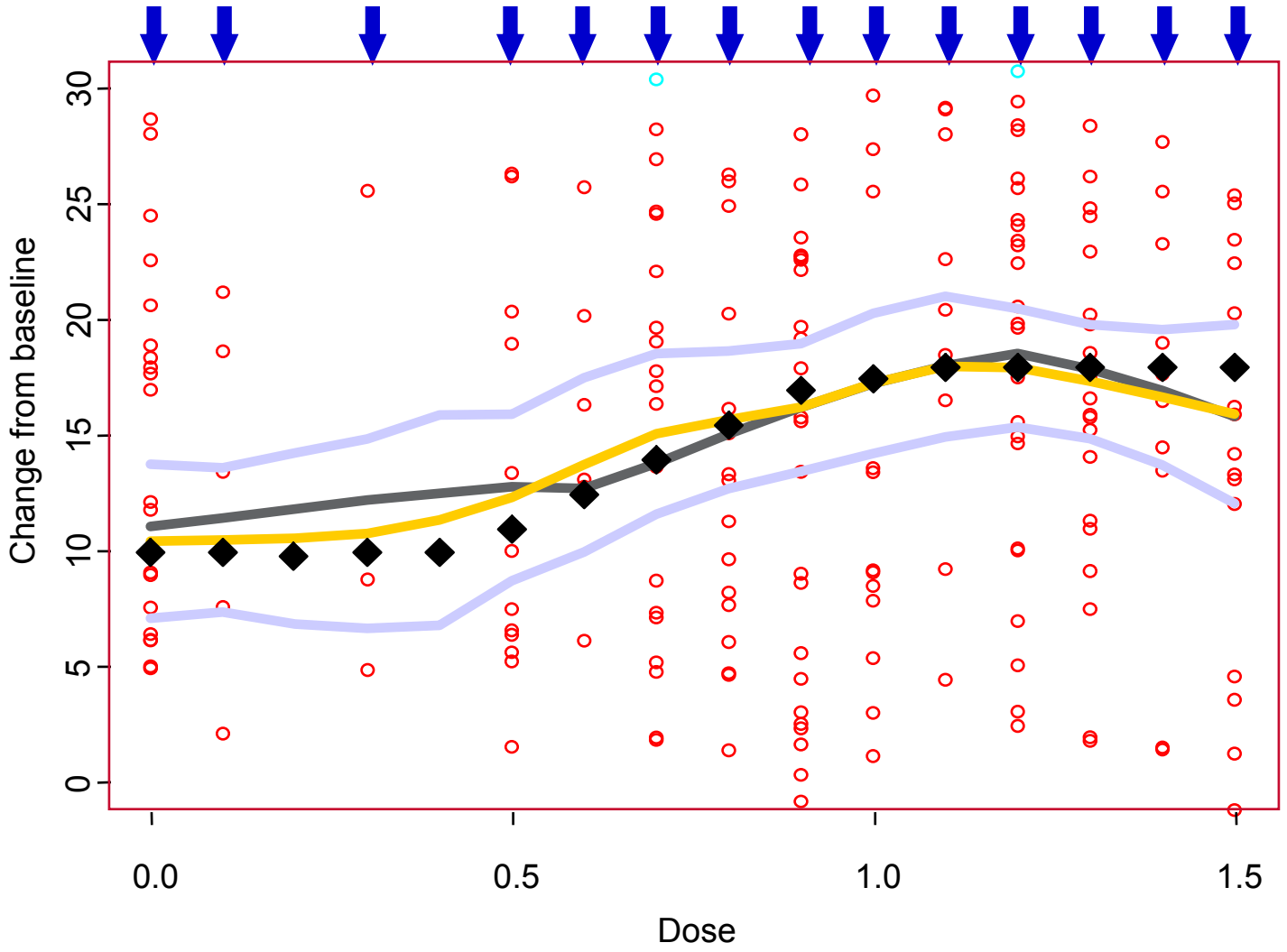


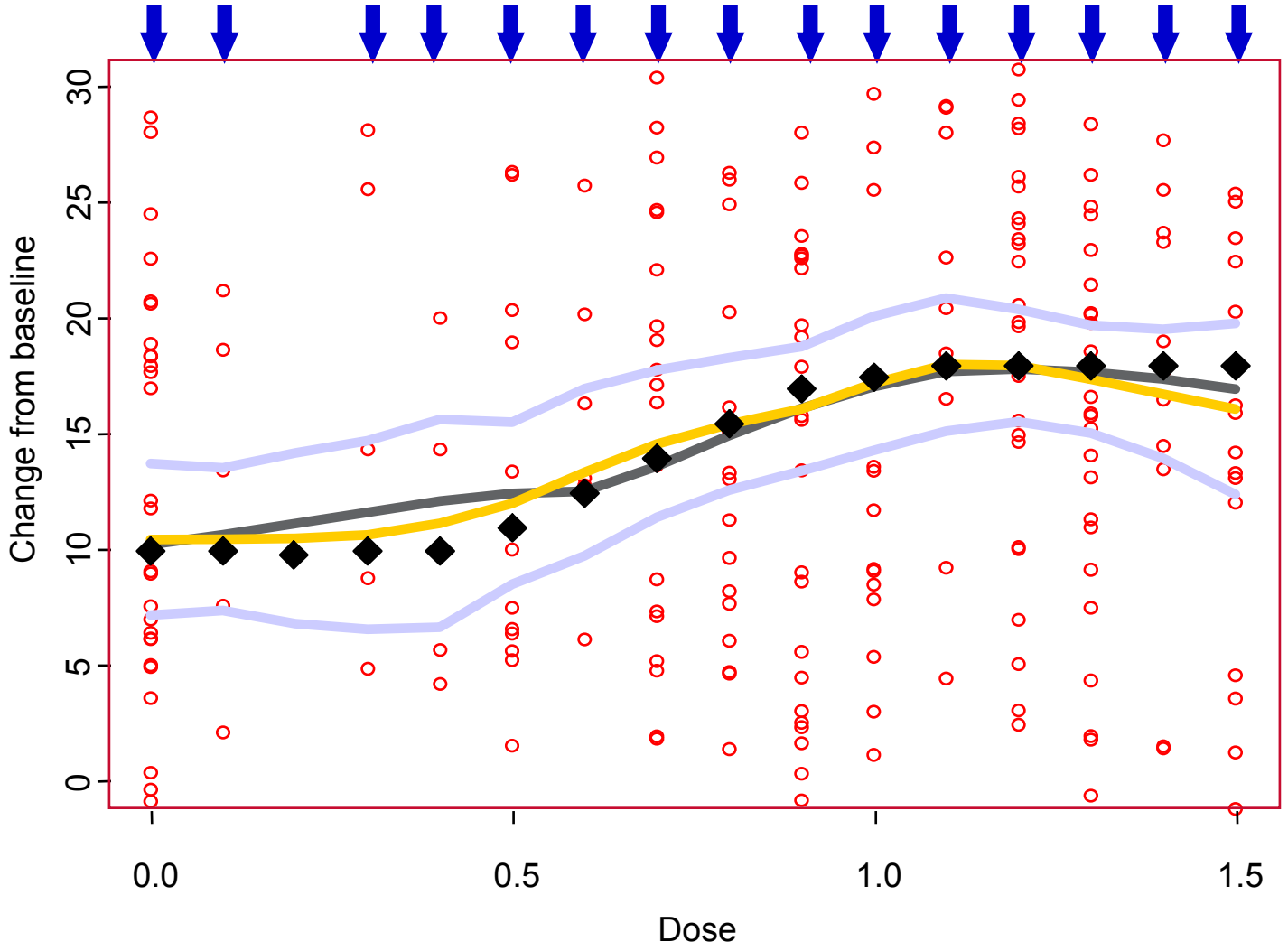


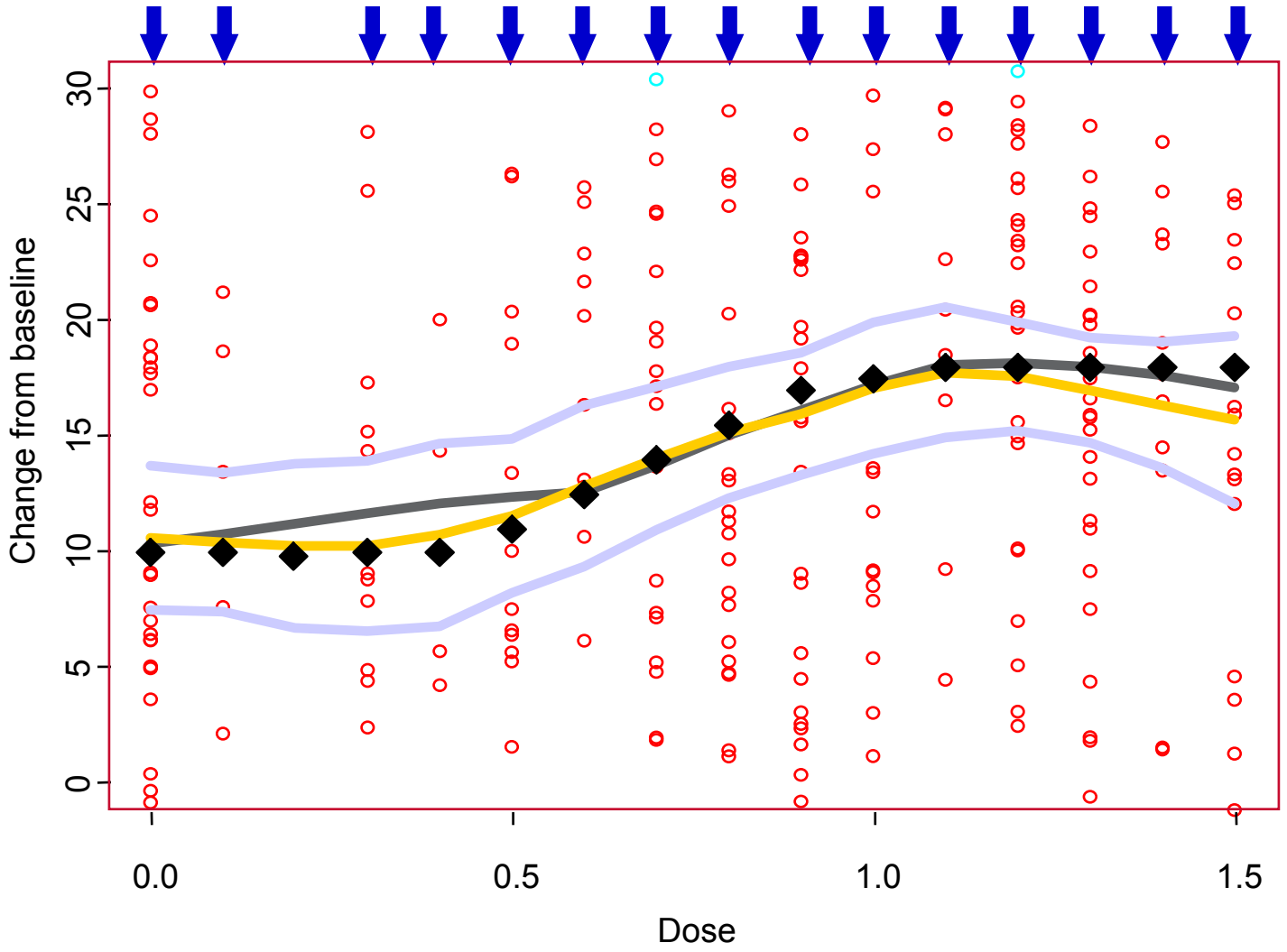


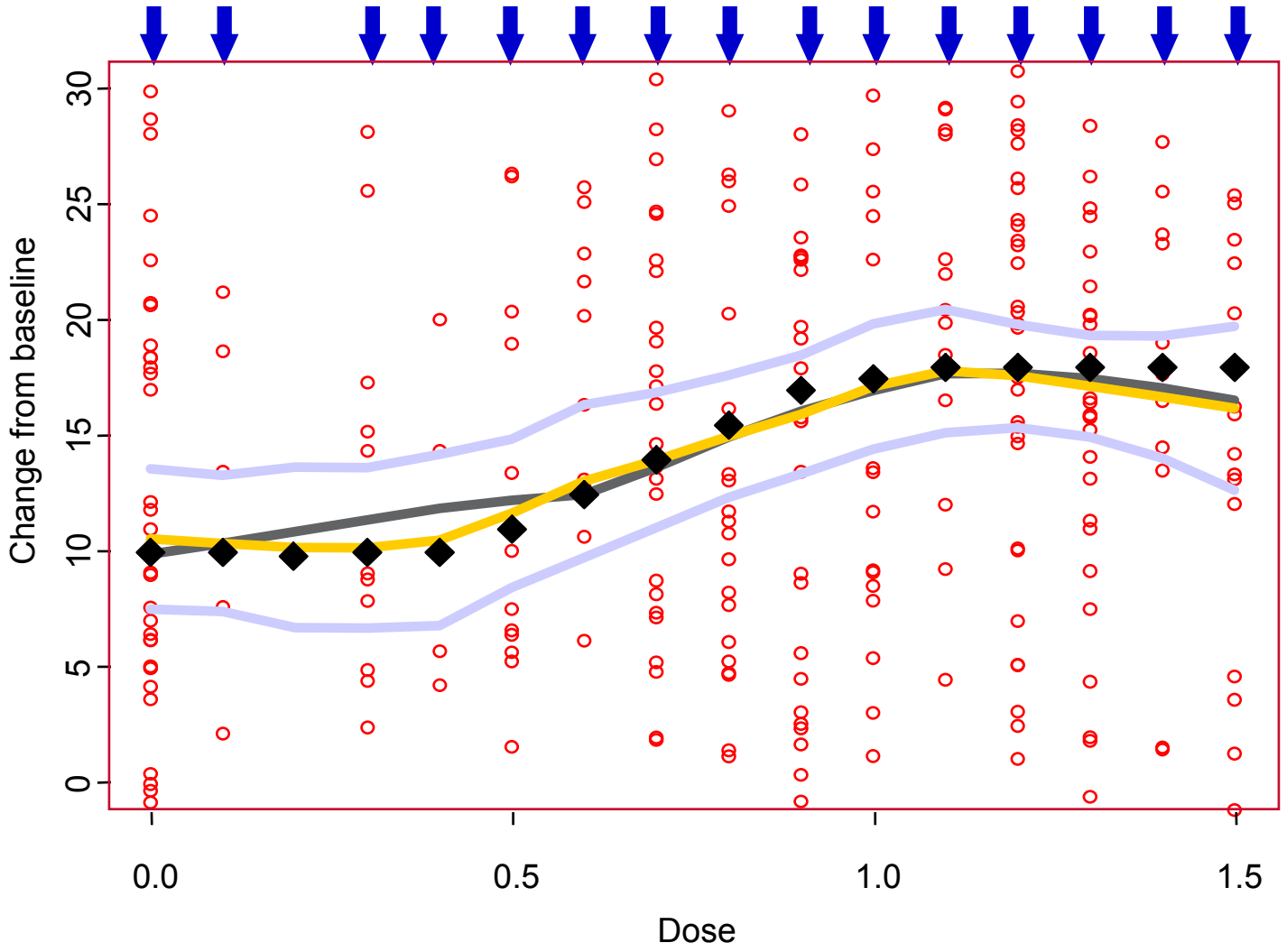


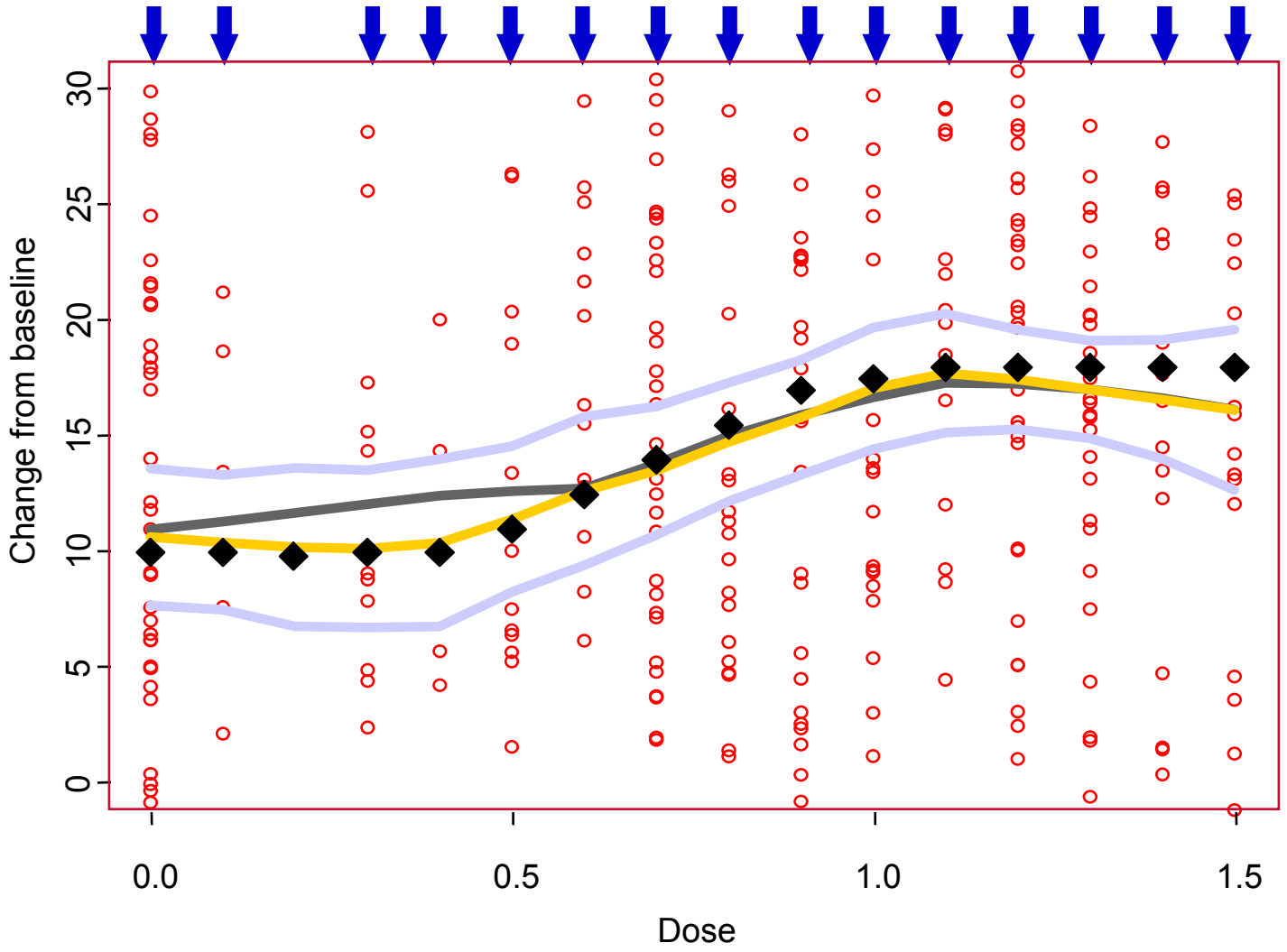


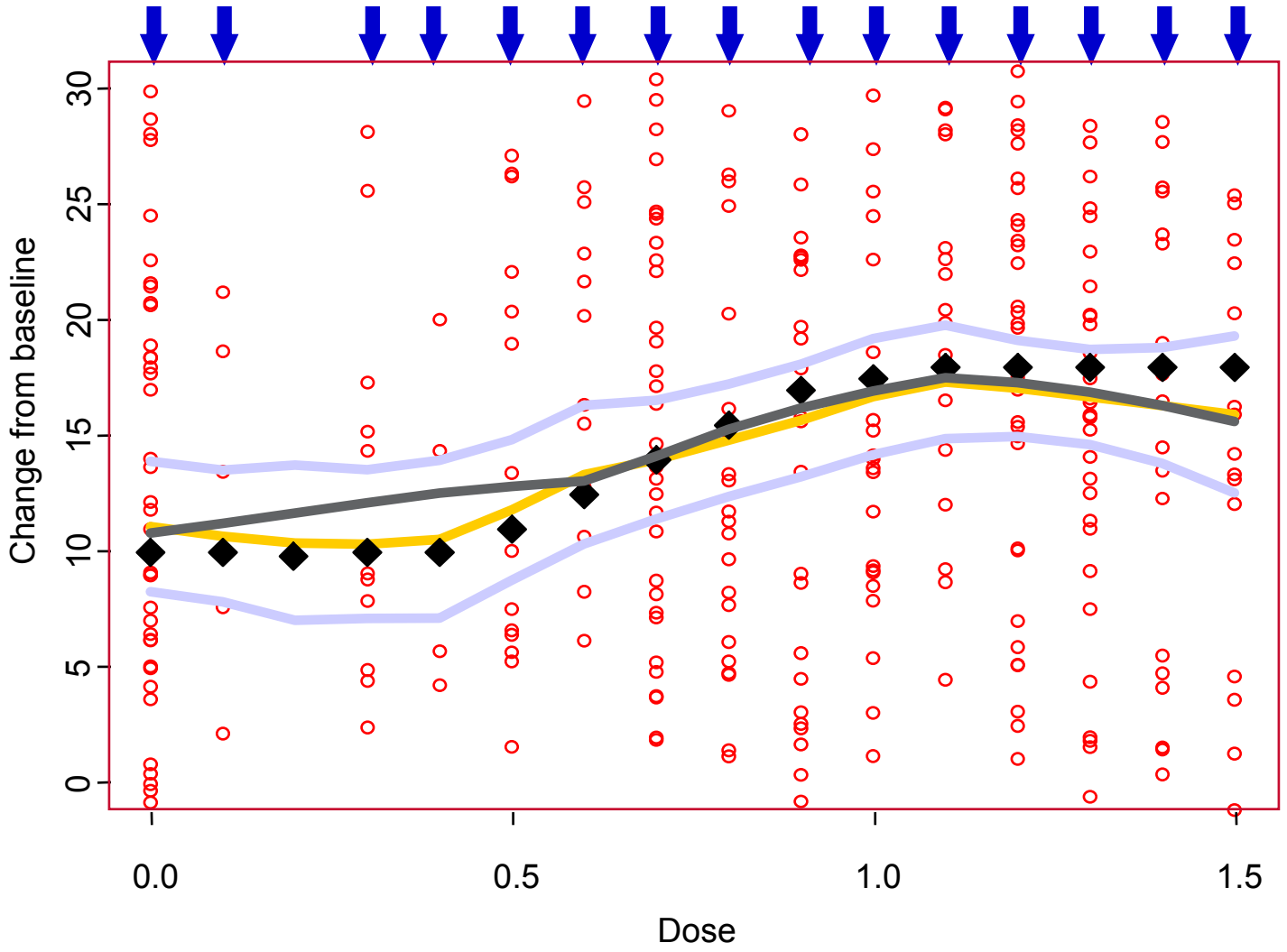


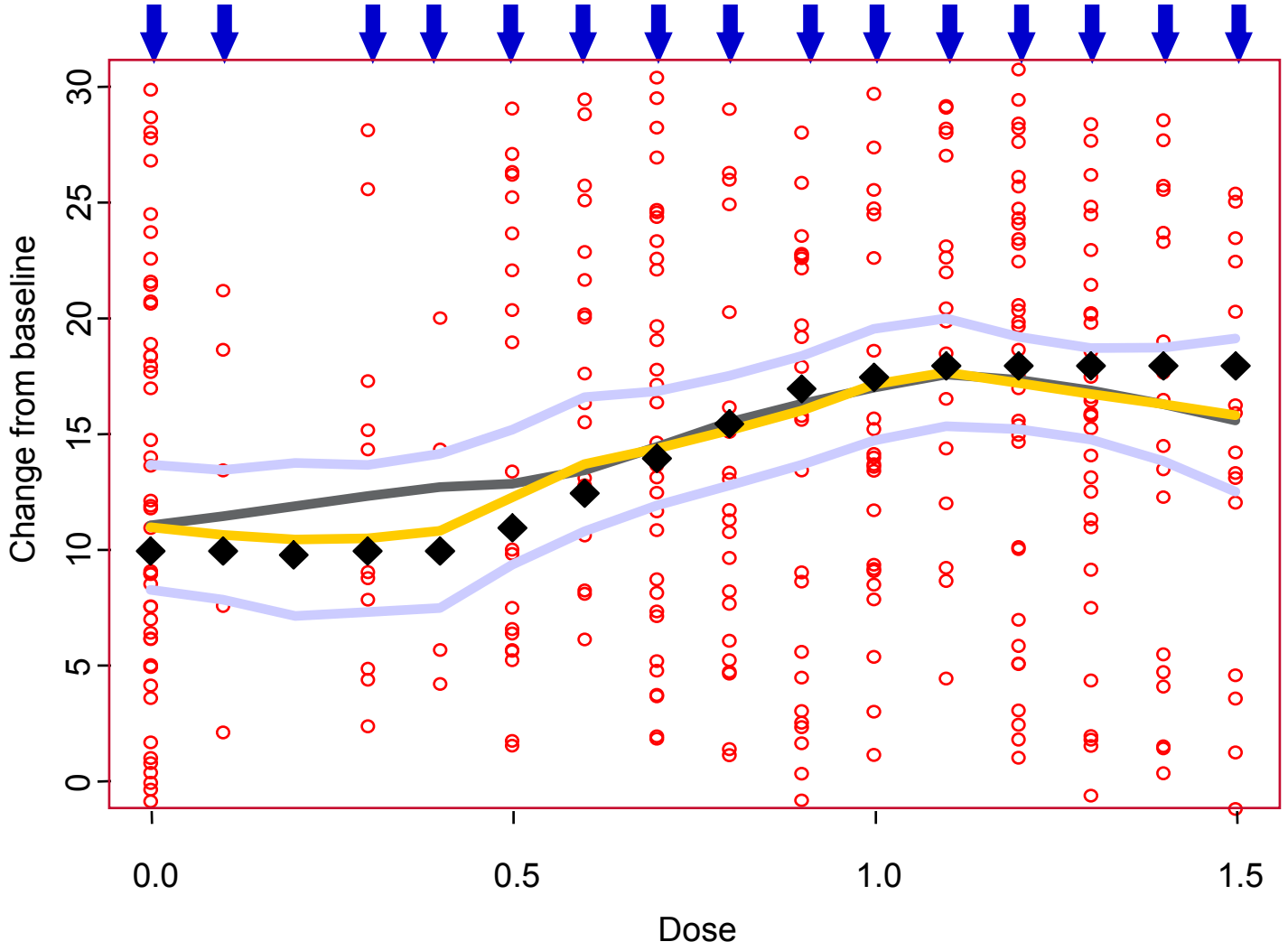


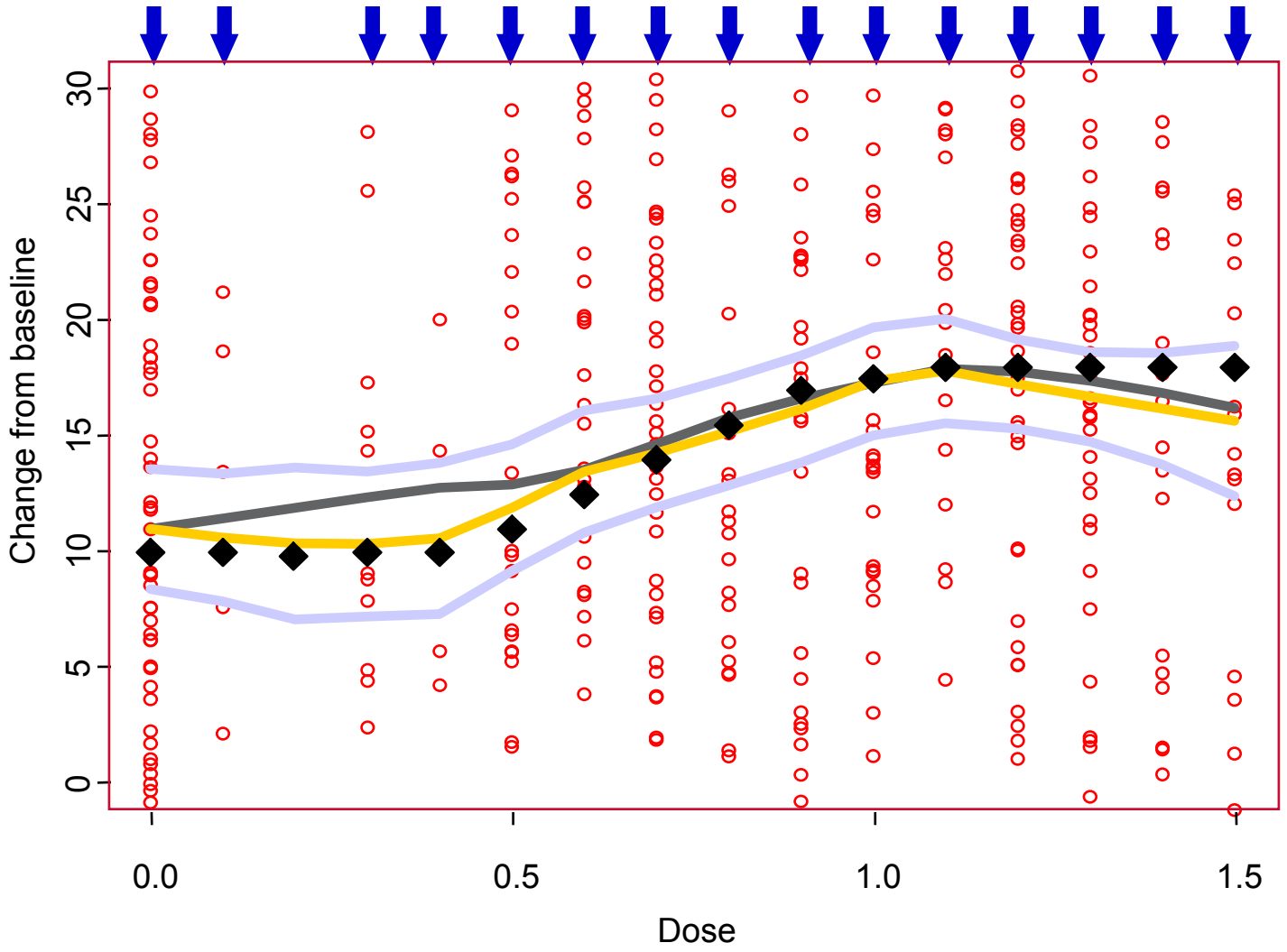


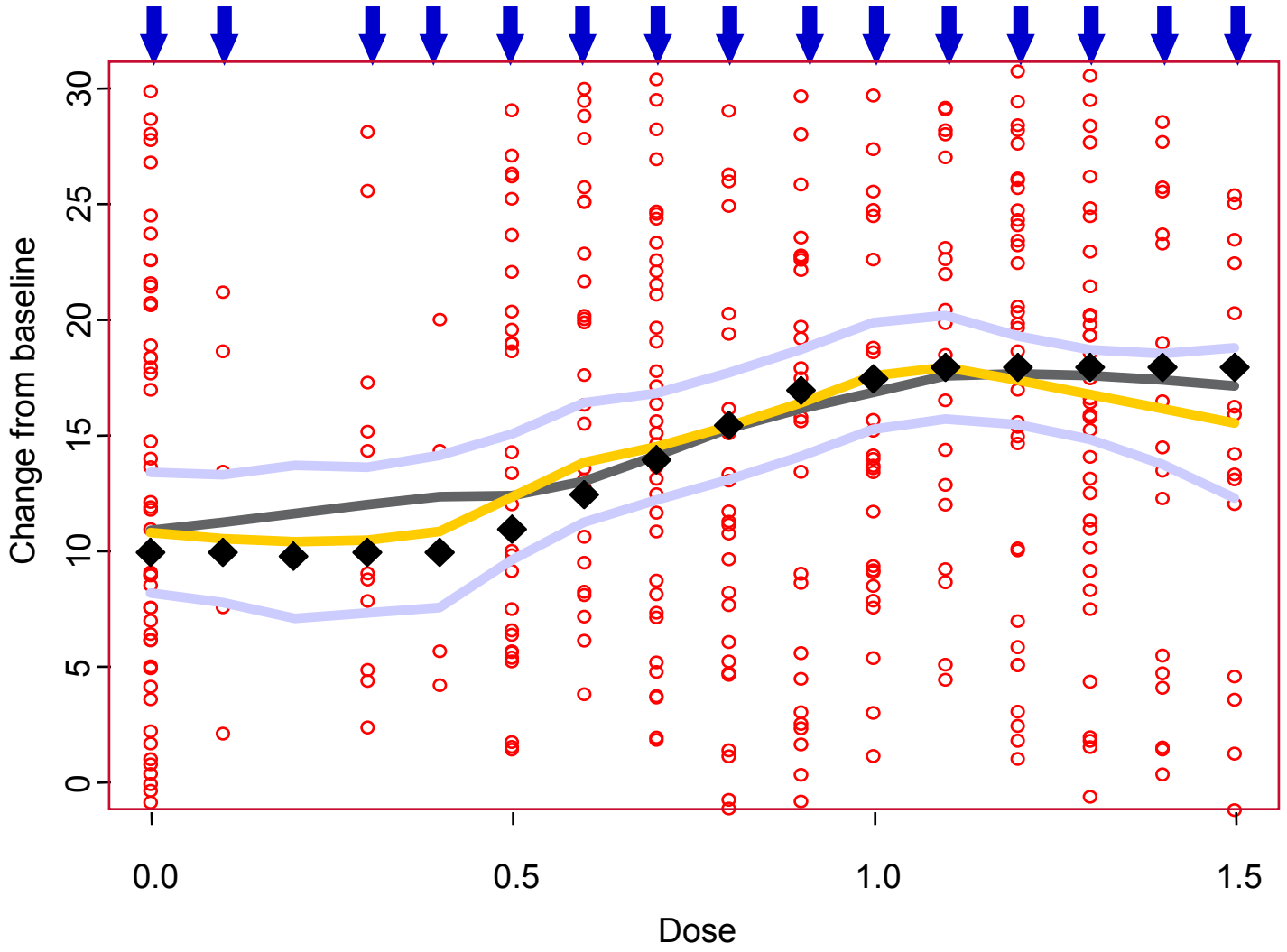


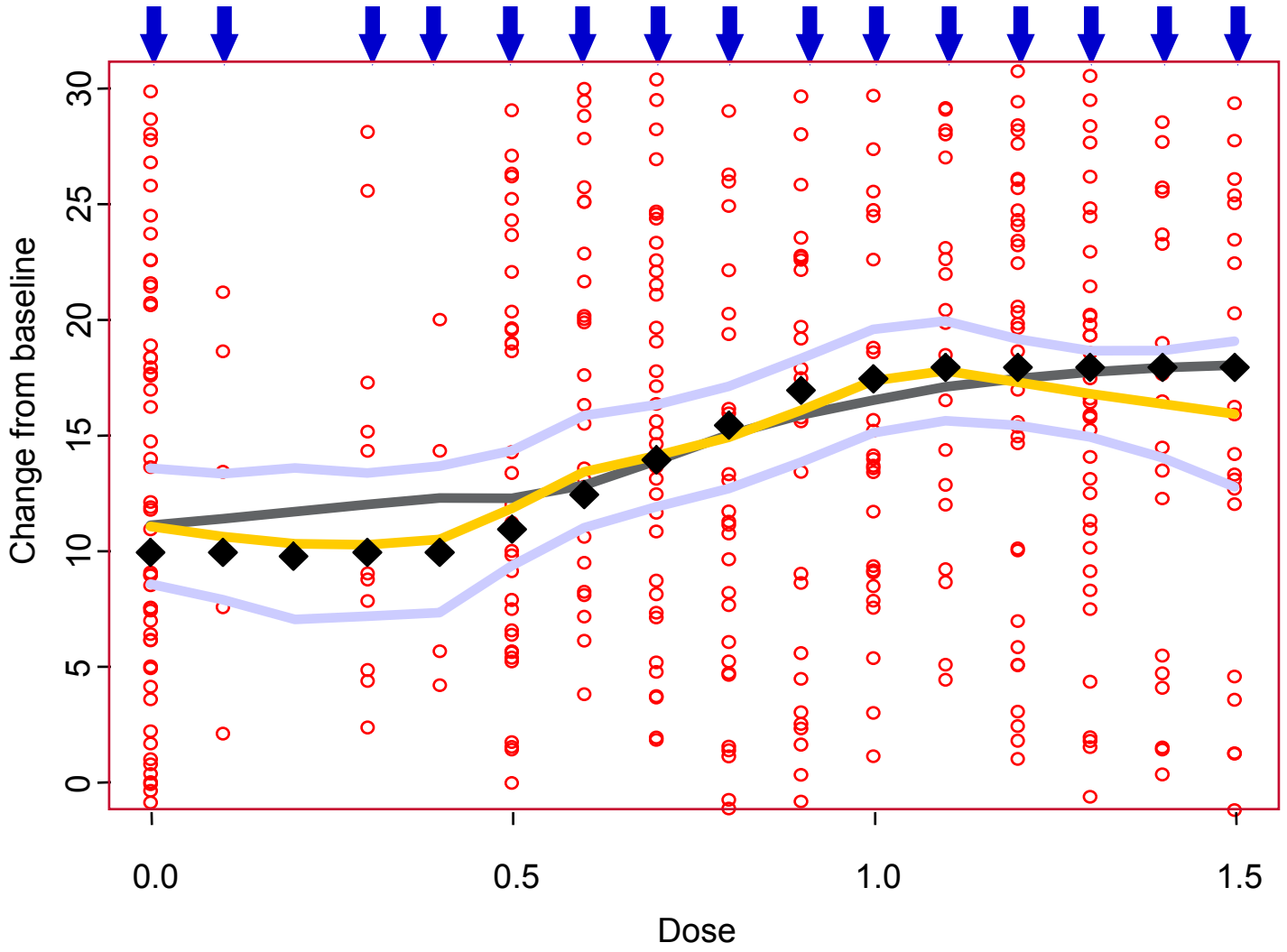




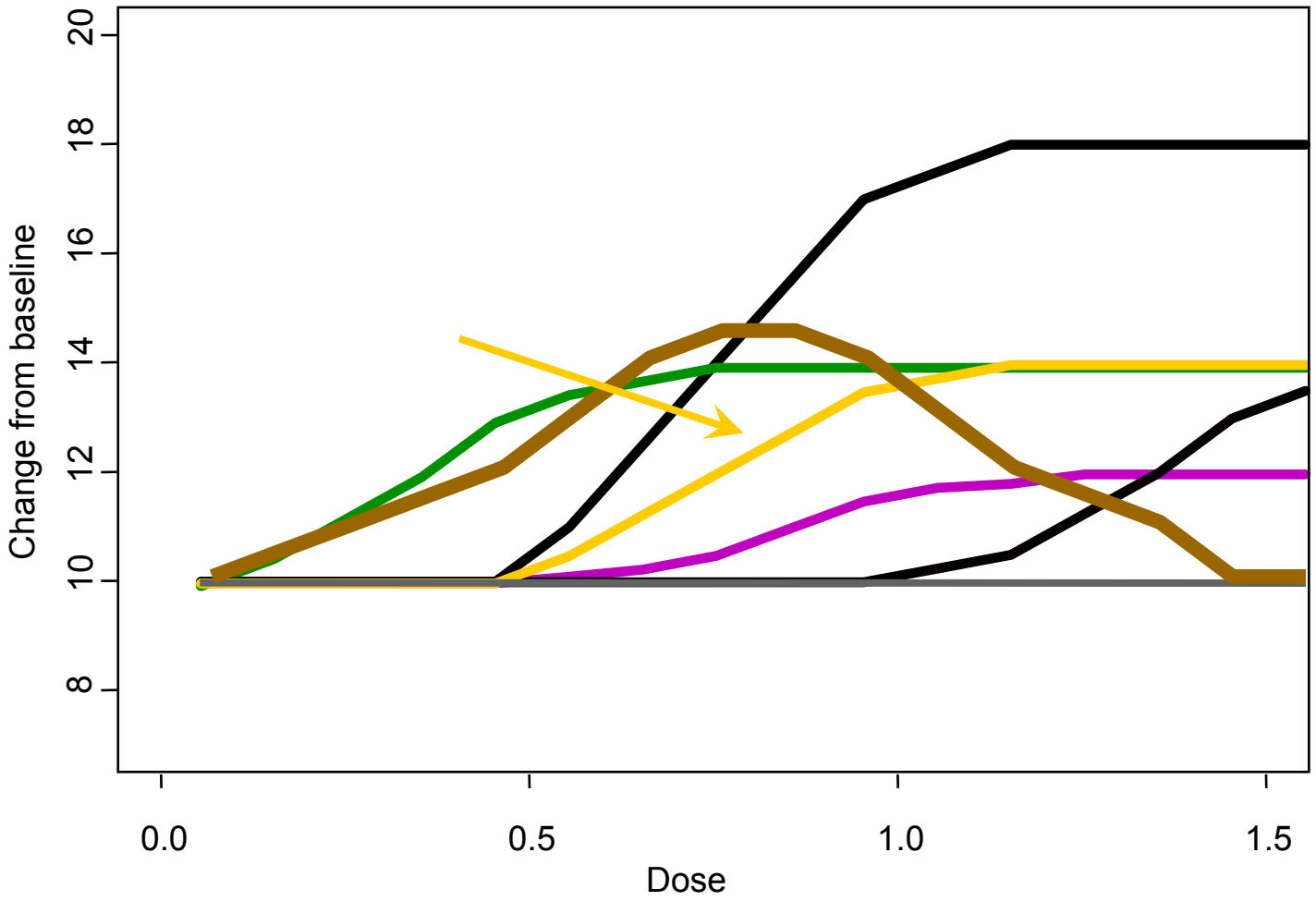




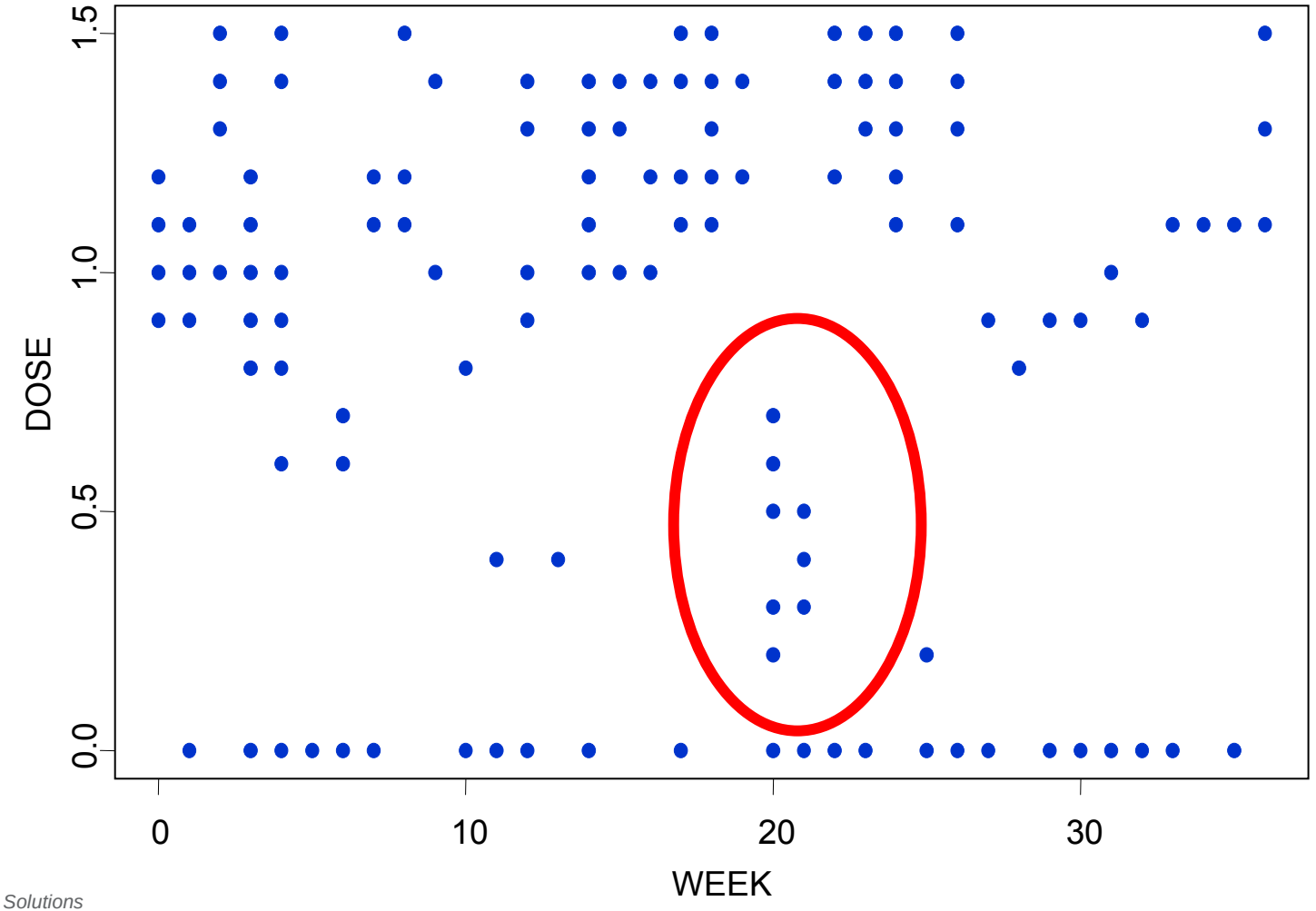




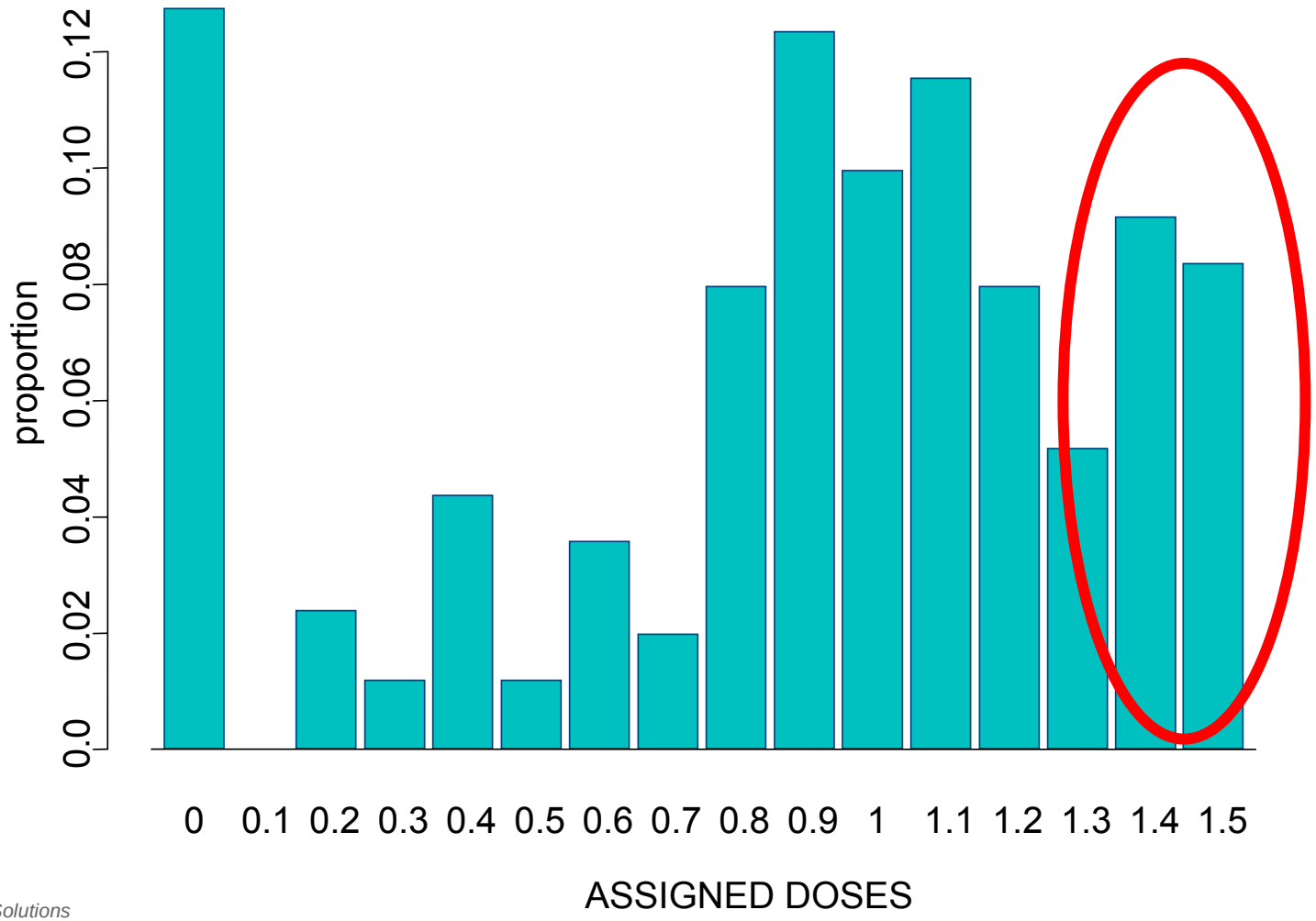
Simulated Dose Response Curves



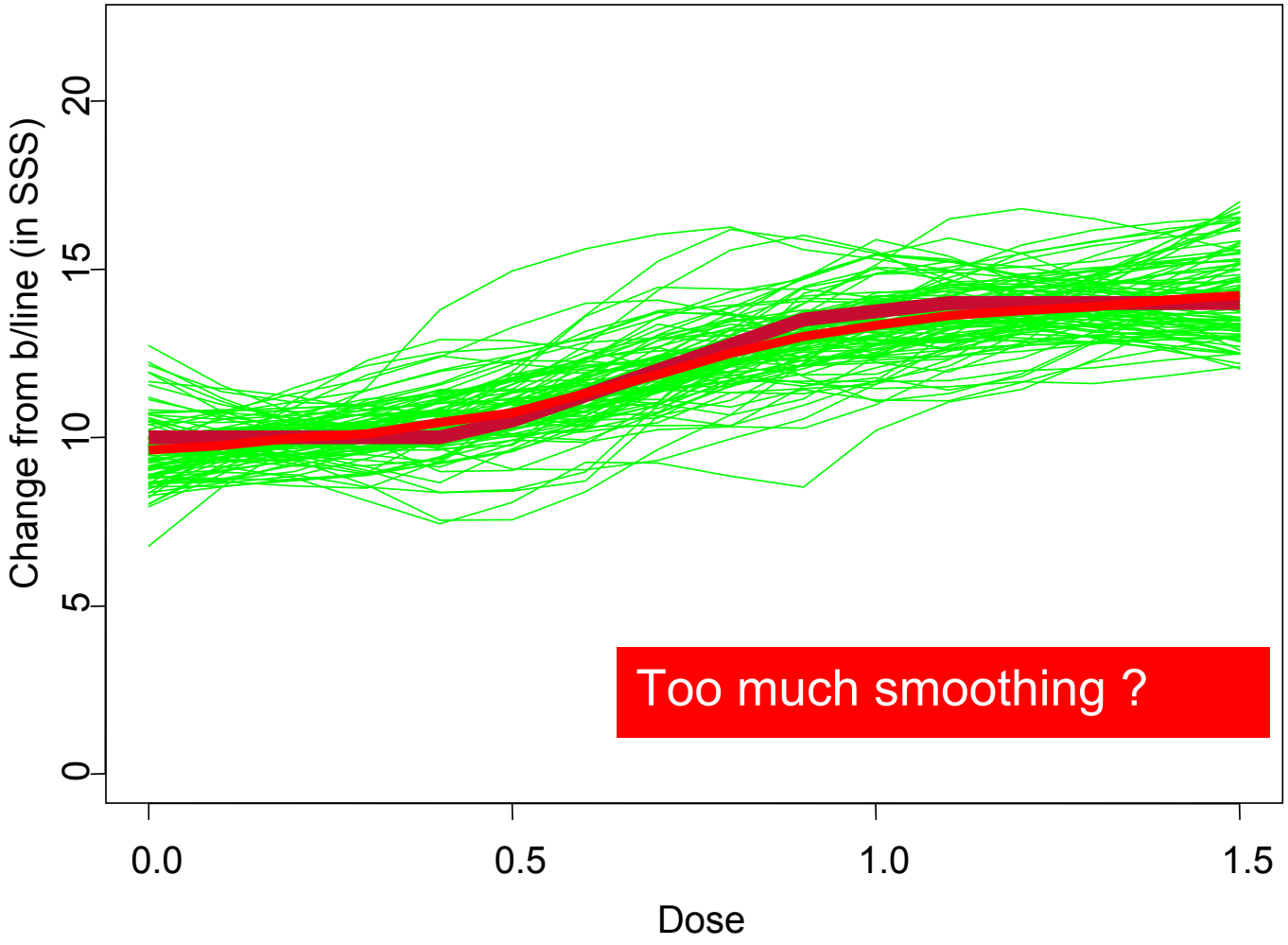
Assigned Doses



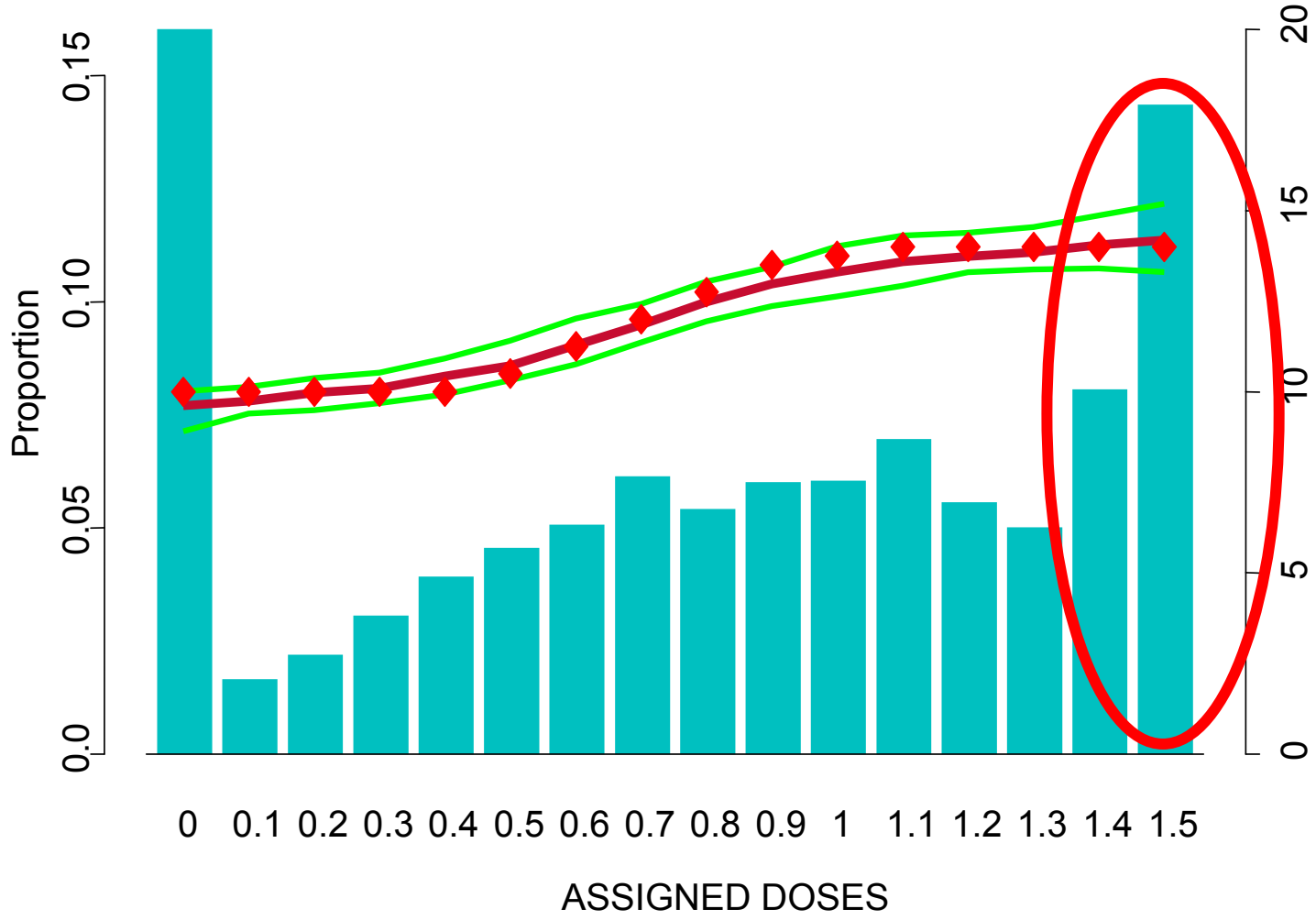
Dose assigned Histogram – 1 simulation



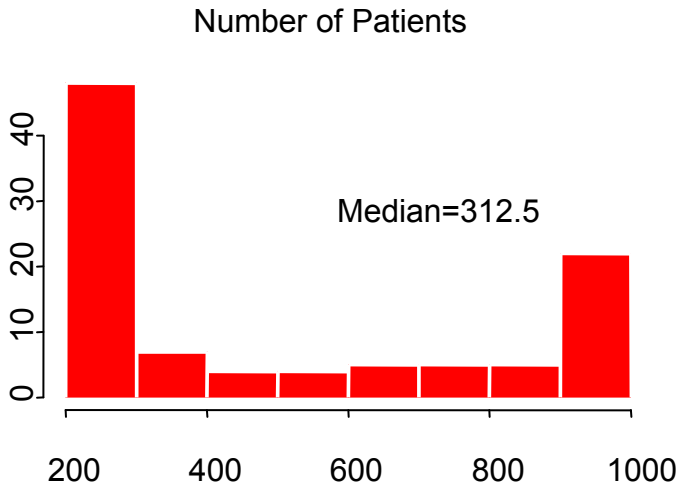
Estimated Dose response Function (x100)

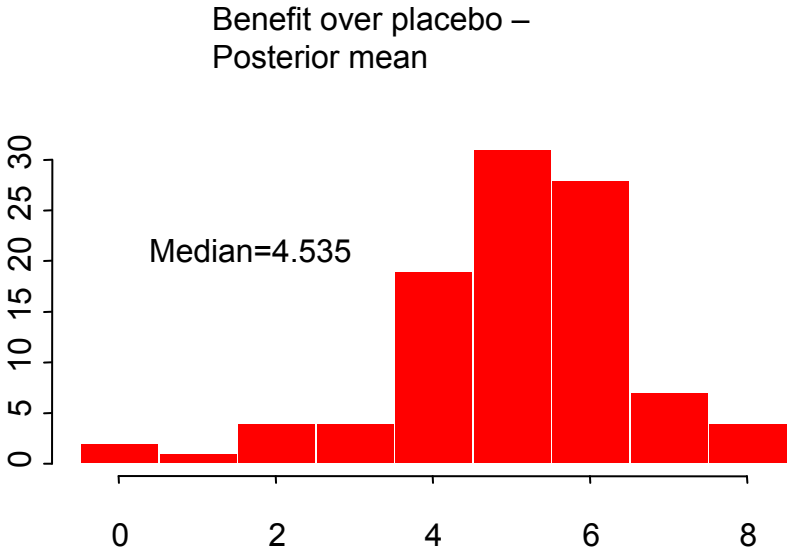


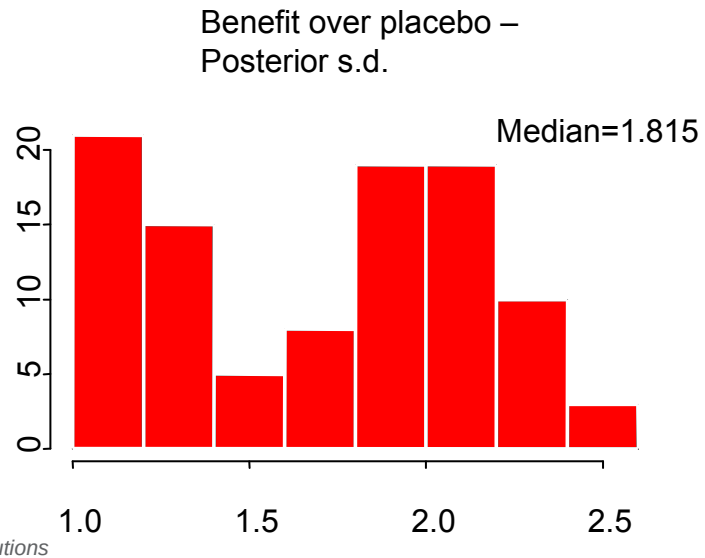
Doses Assigned Across All Simulations

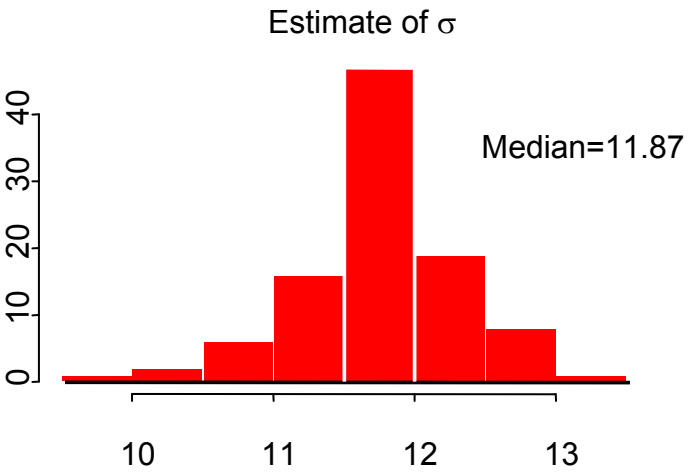


Simulation Summaries

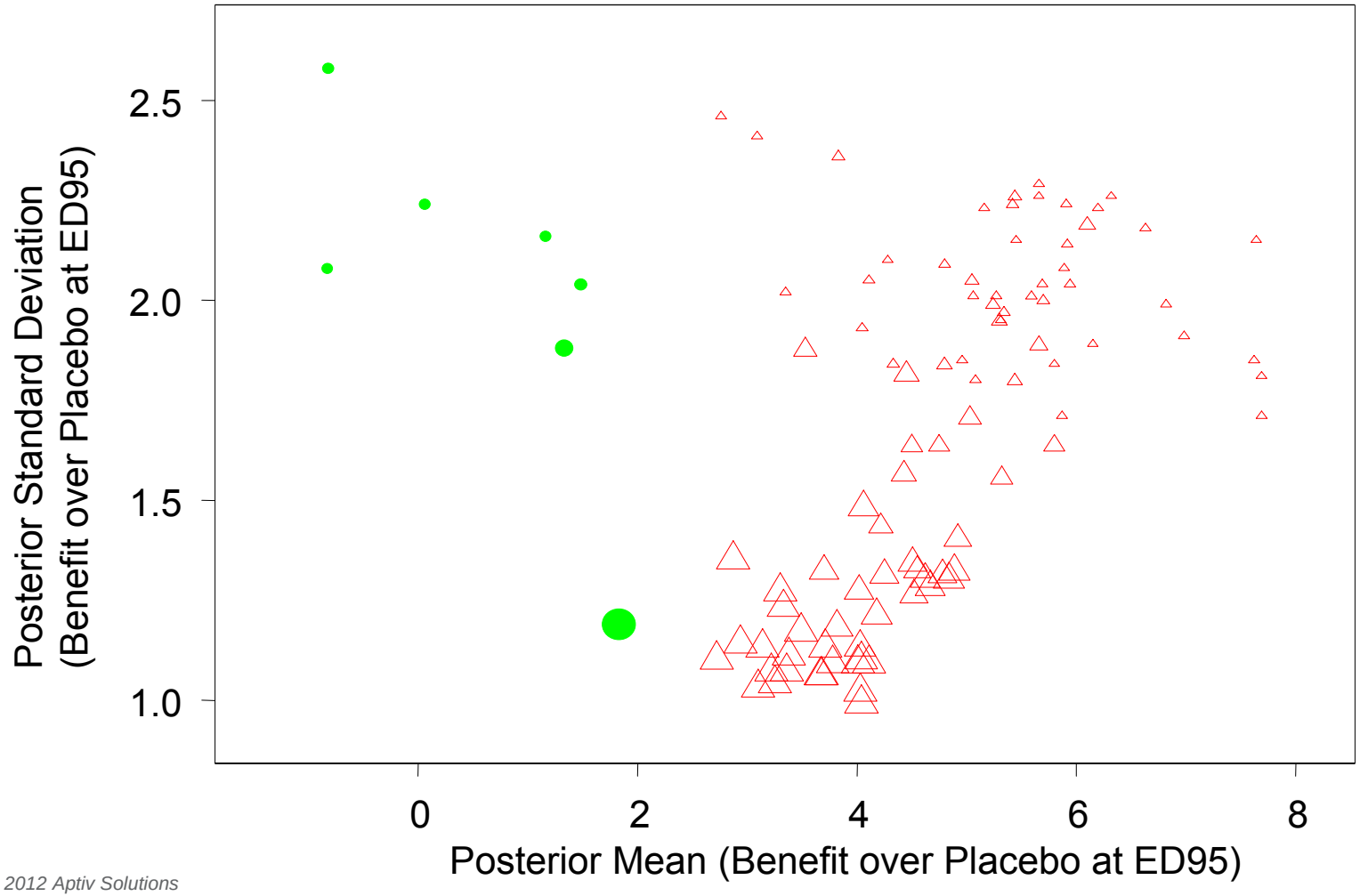




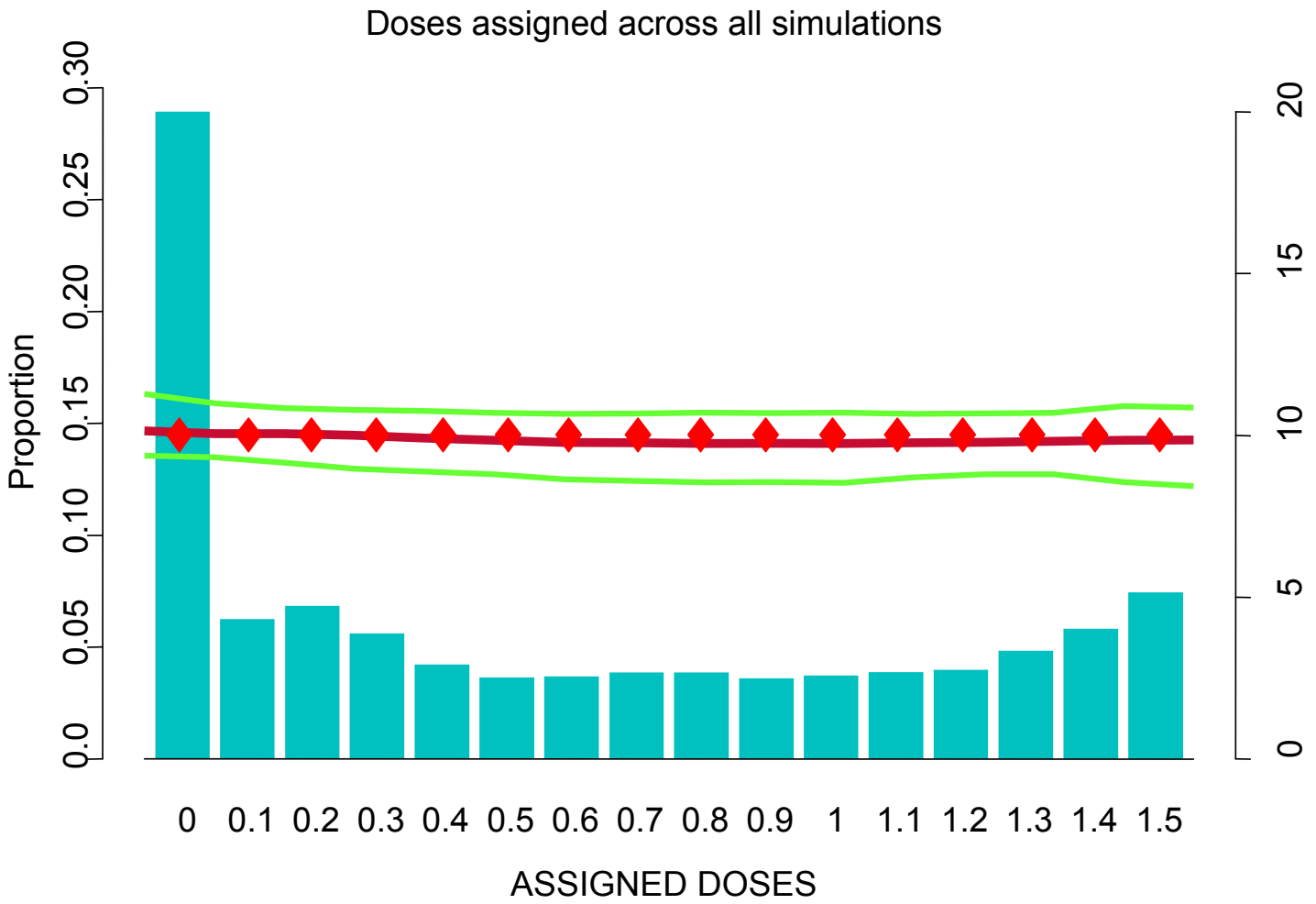




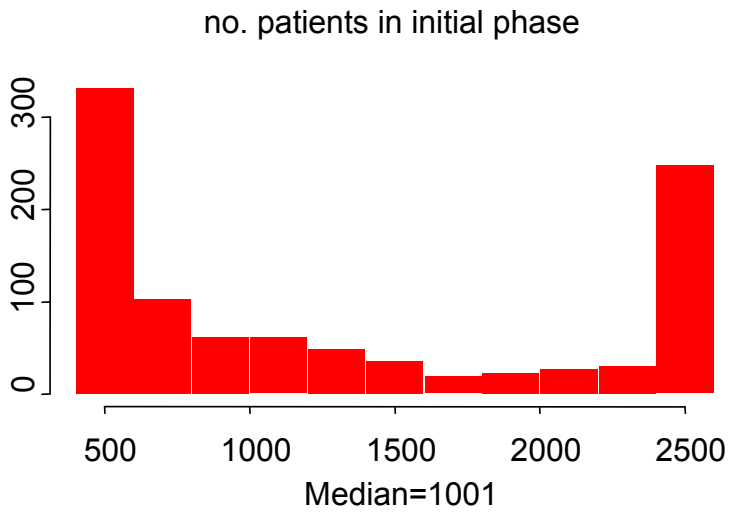
Relationship of Posterior Mean to Posterior SD at Stopping Point

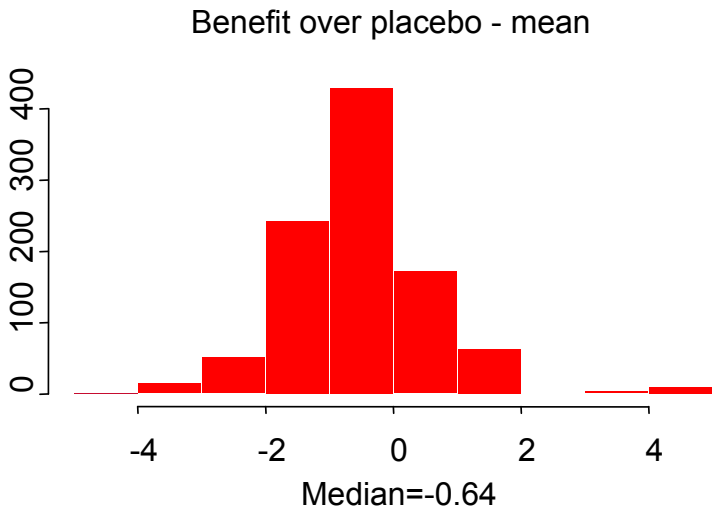


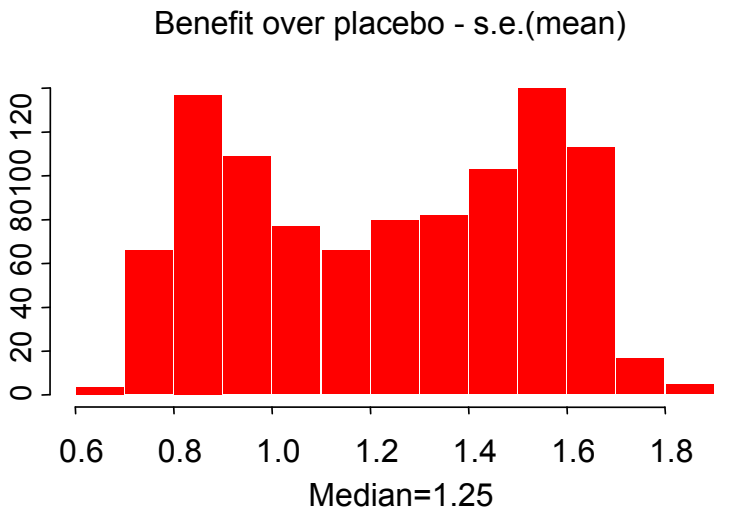
Doses Assigned Across All Simulations
Flat dose response curve

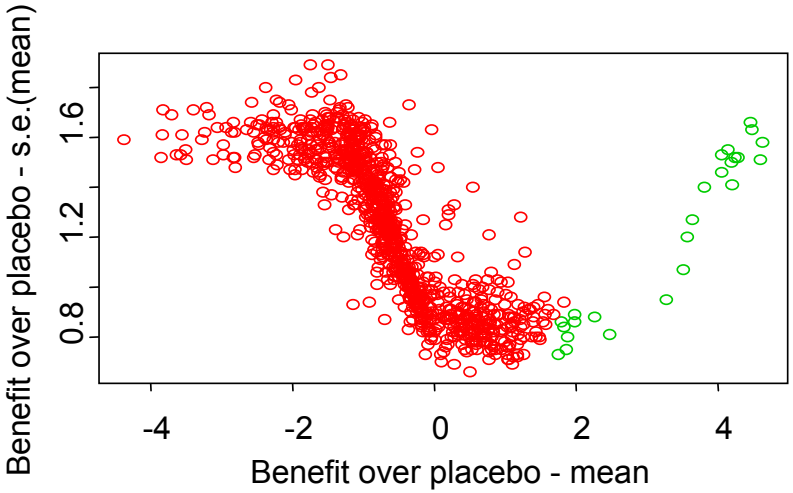


Simulation Summaries
Flat dose response curve









Traditional Design

Sample Size Adaptive Design

Benefit Over Placebo	80% Power	90% Power
0 points	-	-
2 points	2432	3220
3 points	1080	1432
4 points	608	808

Adaptive design max 1000 evaluable patients	
% stop efficacy	Median # pats
0.02	501
0.56	644
0.90	416
0.95	280

ASTIN Study

- Does UK-279,276 improve recovery in ischemic stroke?
- Double-blind, placebo-controlled, dose response finding study
- Placebo and 15 treatment arms (single 15 min iv infusion)
- Adaptive treatment allocation
- Efficacy endpoints
  - Secondary: NIH-stroke scale, mod Rankin, Barthel
  - Primary: Scandinavian Stroke Scale:
    Mean change from baseline to day 90 >=3 points?

ASTIN Study

- Include, if
  - Patients ≥ 50 years with acute stroke (<6 h)
  - Baseline stroke severity 10-40 on Scandinavian Stroke Scale
- Exclude, if
  - Premorbid modified Rankin ≥ 2
  - Women of childbearing potential
  - Coma or reduced level of consciousness on admission
  - Fixed eye deviation with total hemiplegia
  - Seizure since onset of stroke
  - Temperature on admission of ≥ 38°C
  - Signs or symptoms suggestive of concurrent infection

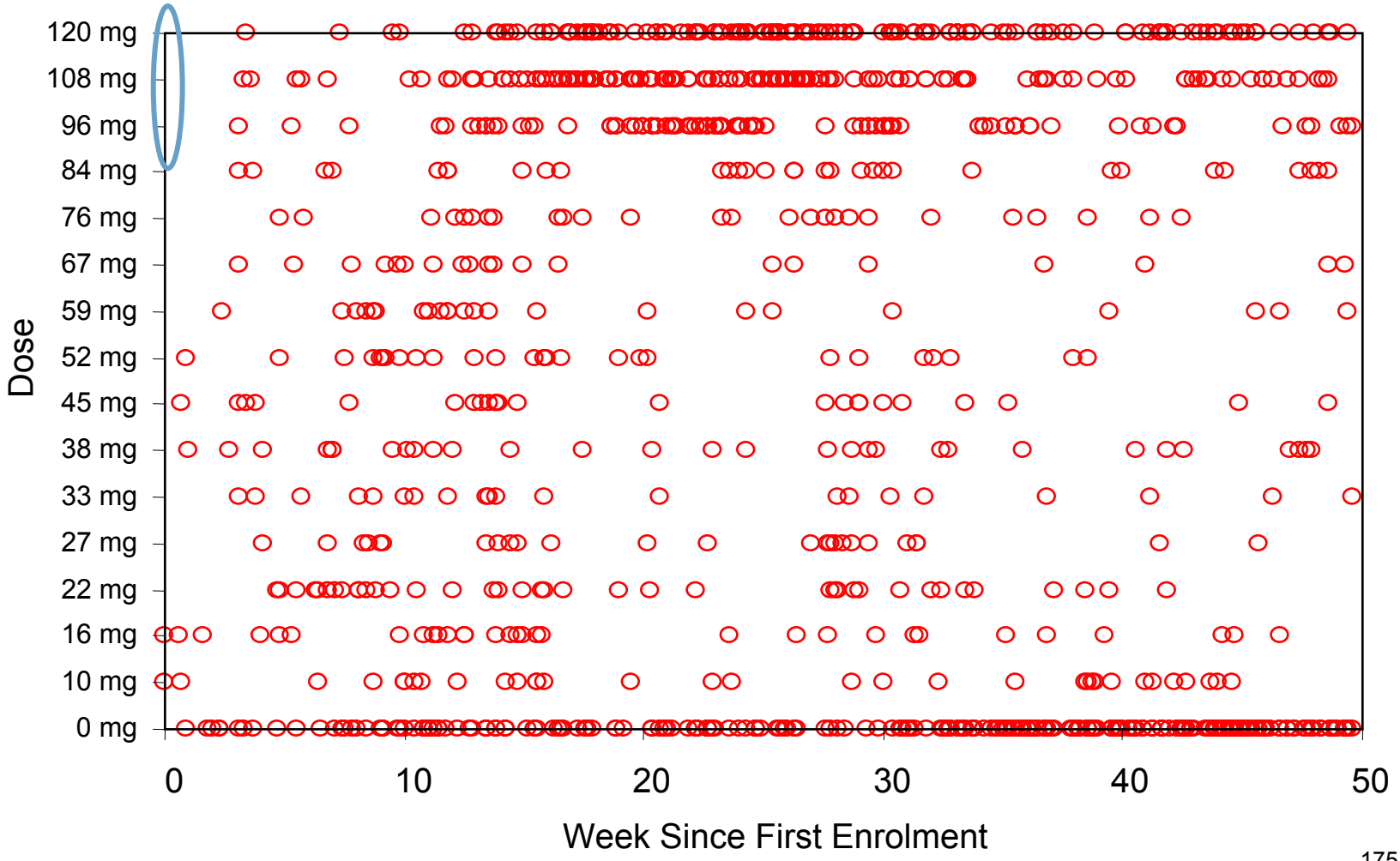
ASTIN Study

- Executive Steering Committee
- Independent Data Monitoring Committee
- Independent statistician preparing reports for IDMC
- Computer system run by Tessella Ltd (UK)

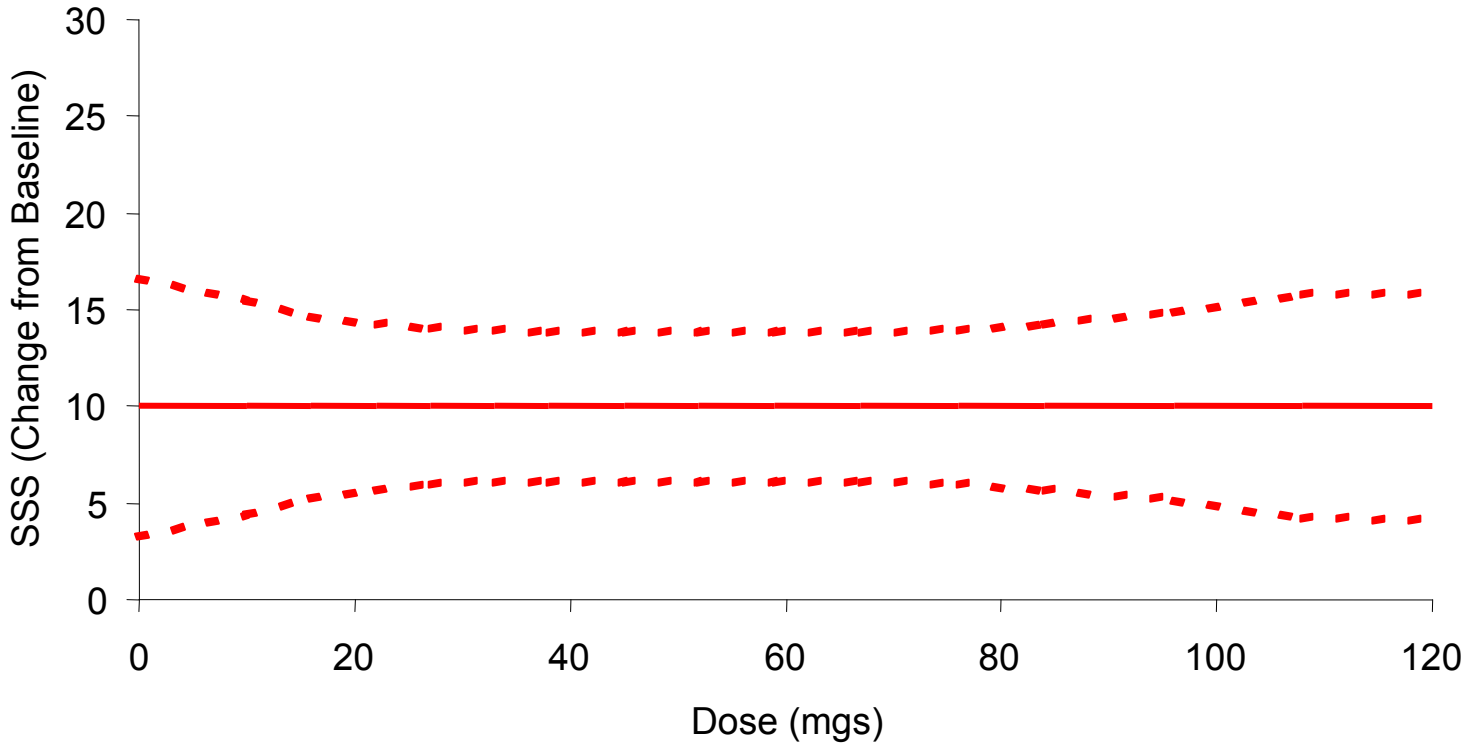
ASTIN Results

- Trial stopped for futility
- 966 patients randomized and treated
- 93% ischemic stroke
  - 21% cotreated with tPA
  - Mean baseline severity ~ 28 points Scand Stroke Scale
  - Demographics comparable across treatment arms
  - Mean onset-to-treatment time: 4 h 08 min
  - Mean door-to-needle time: 2 h 27 min

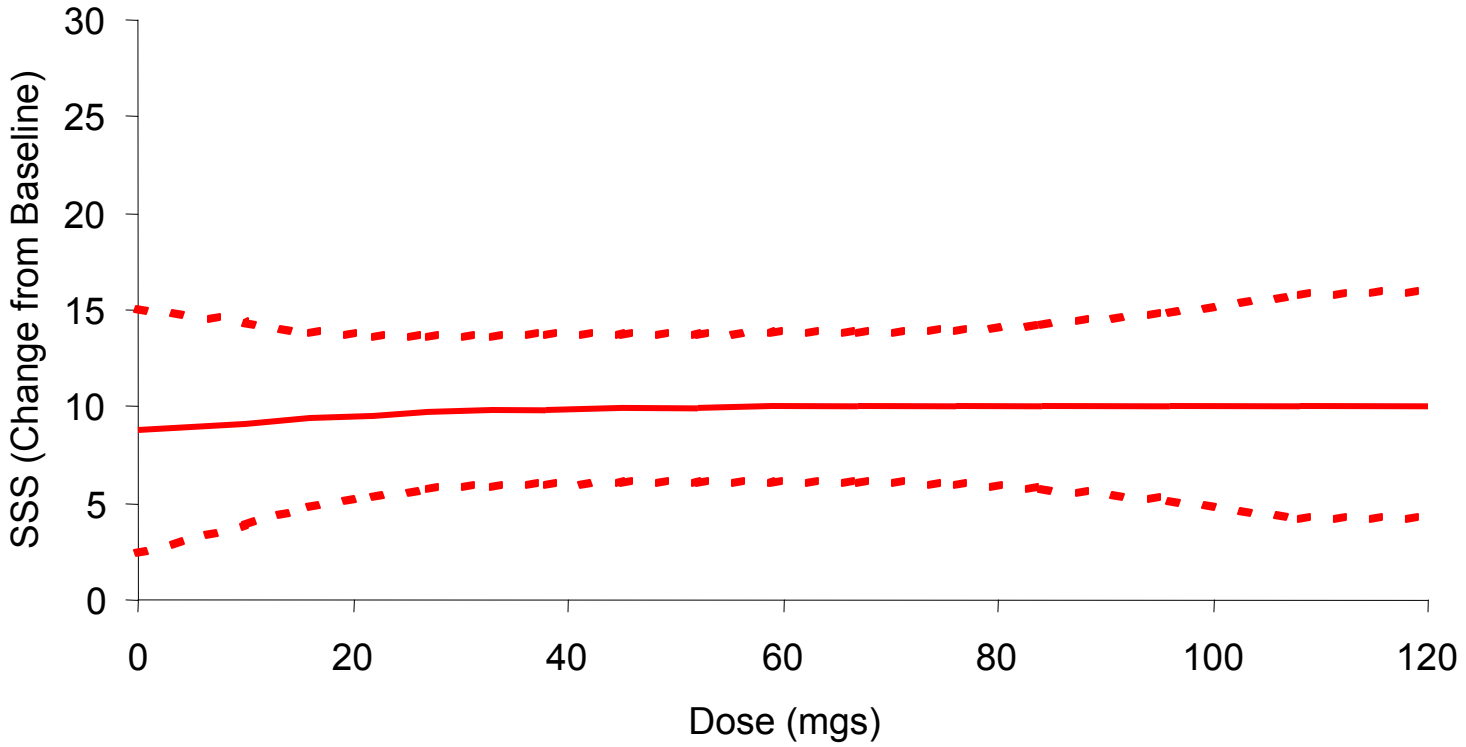
ASTIN – Enrolment Over Time



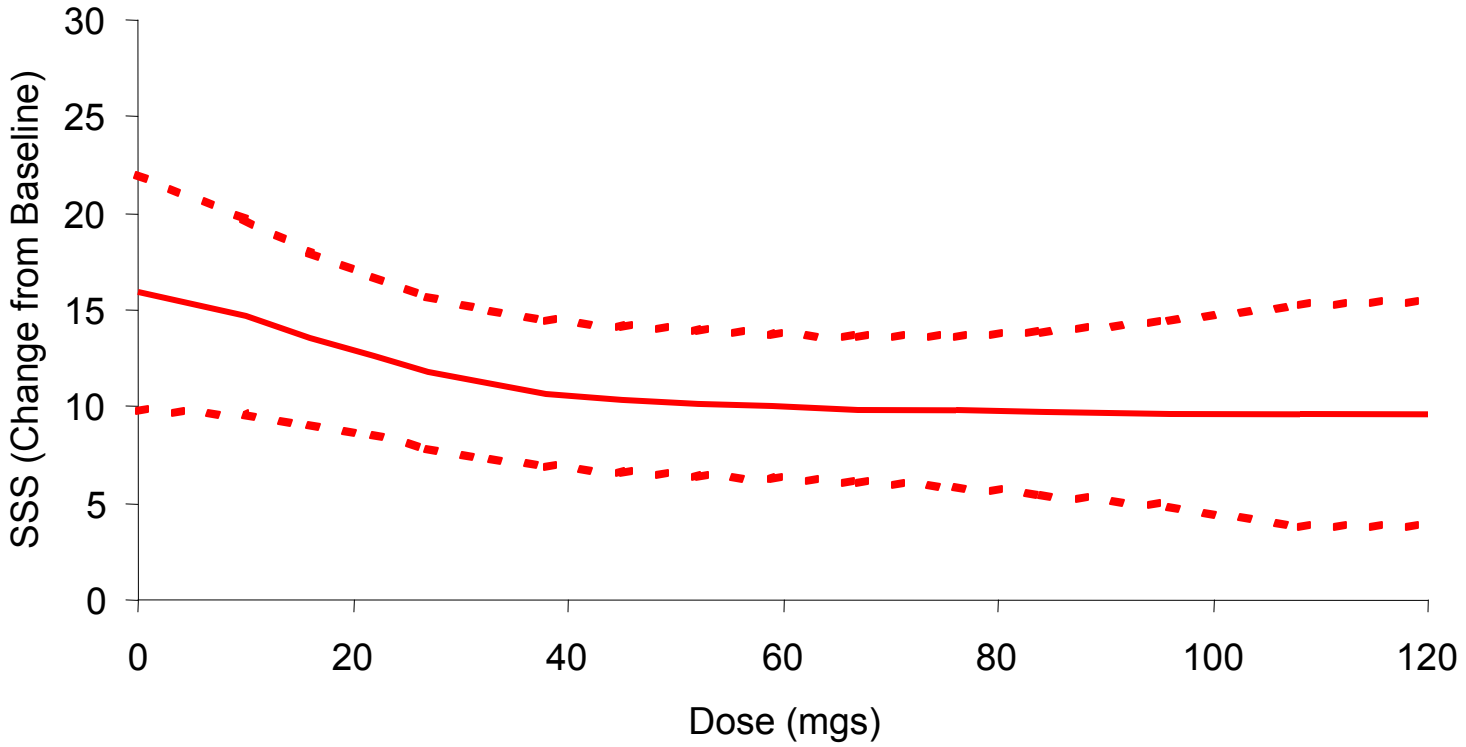
Dose Response : Fortnightly Updates



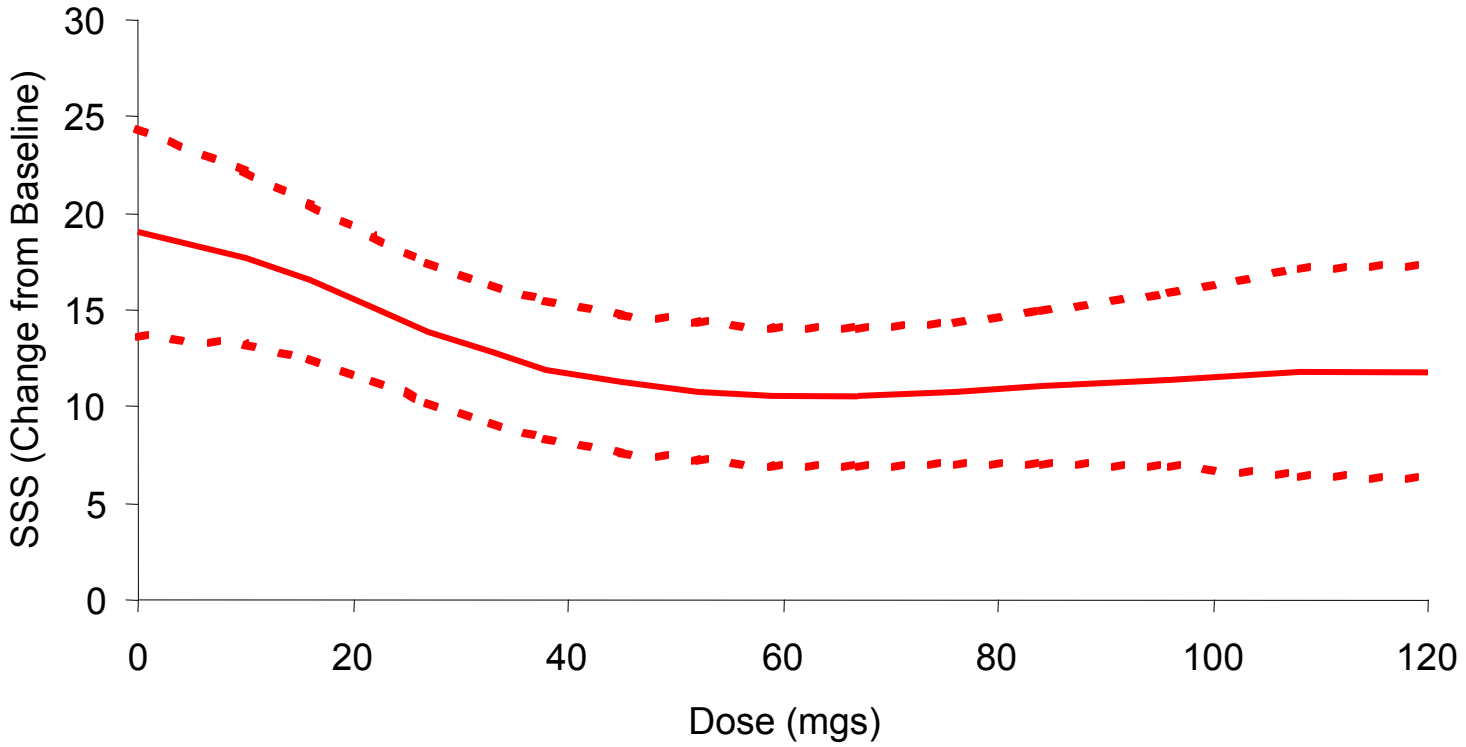
Week - 0



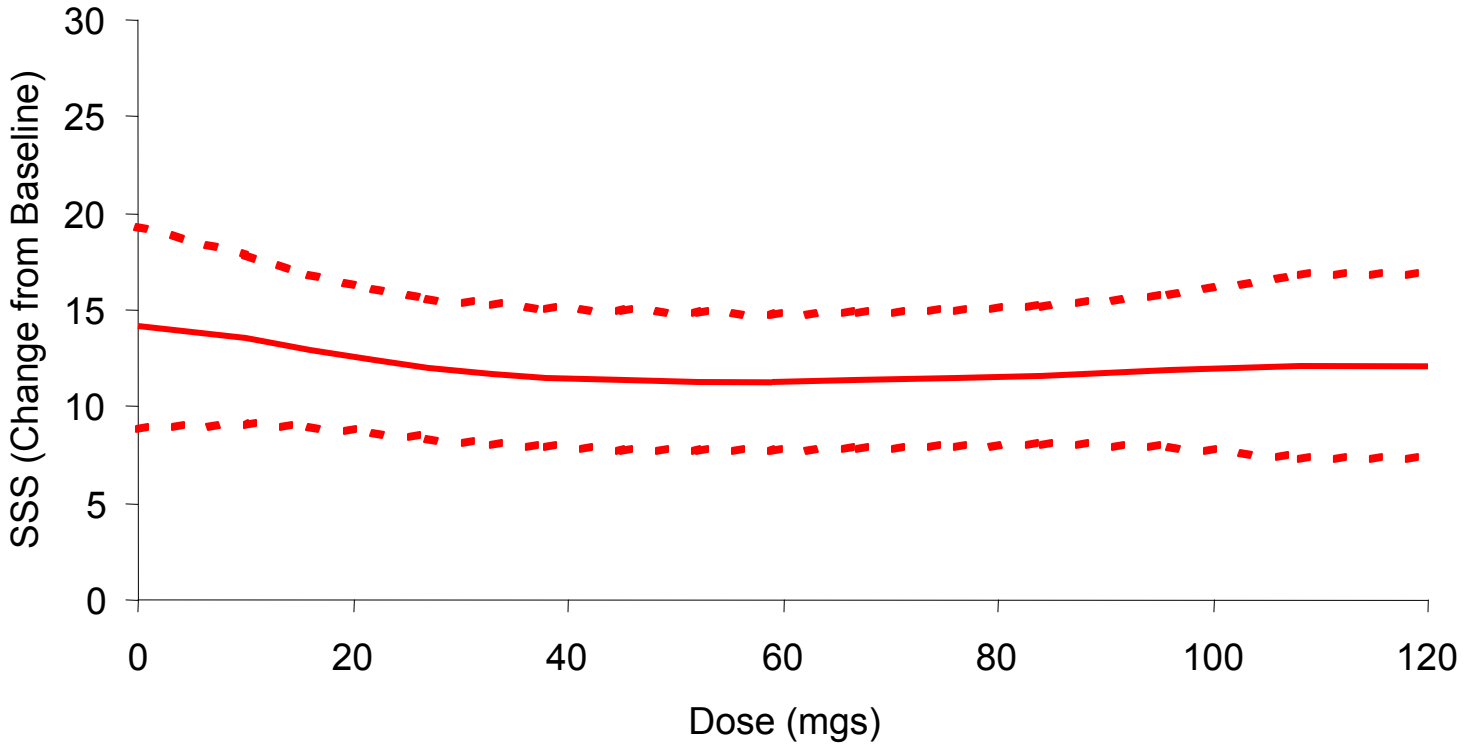
Week - 2



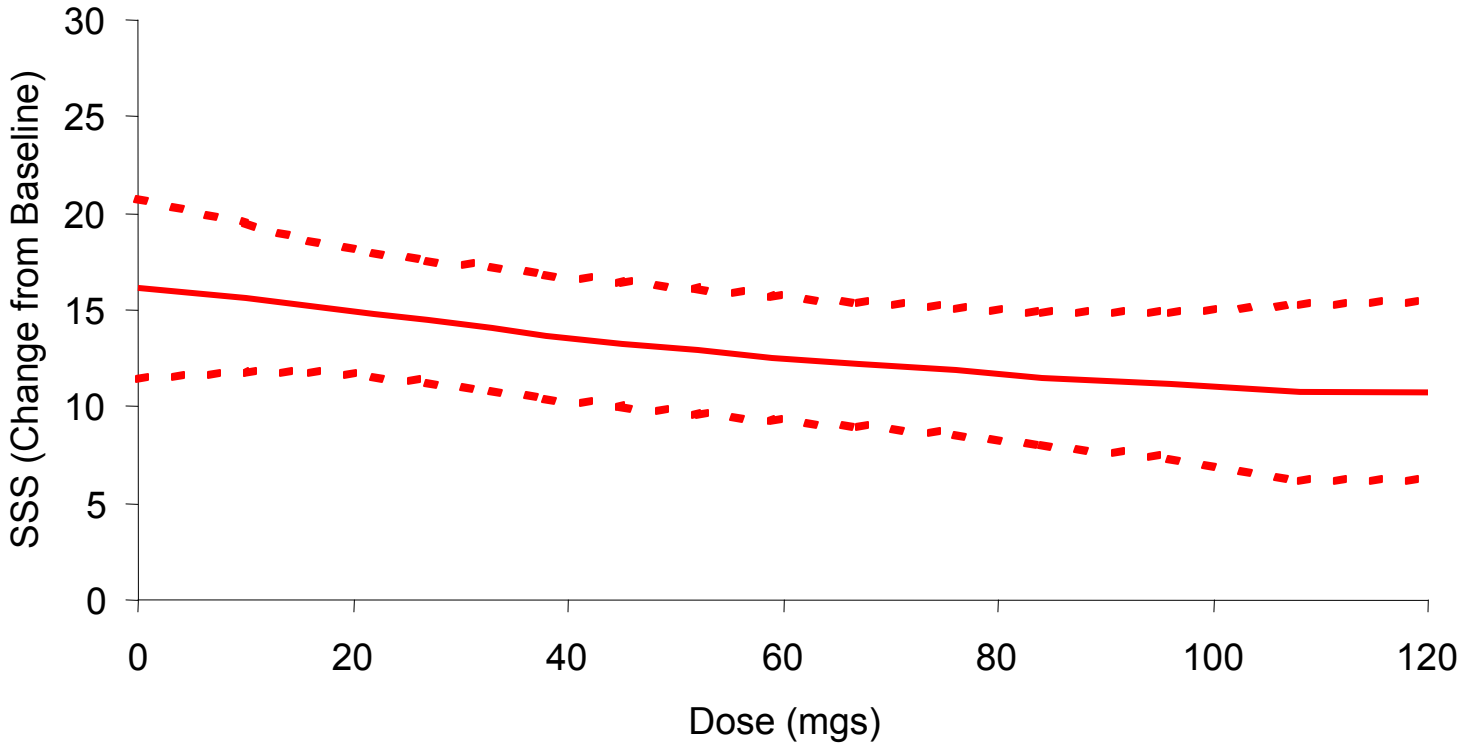
Week - 4



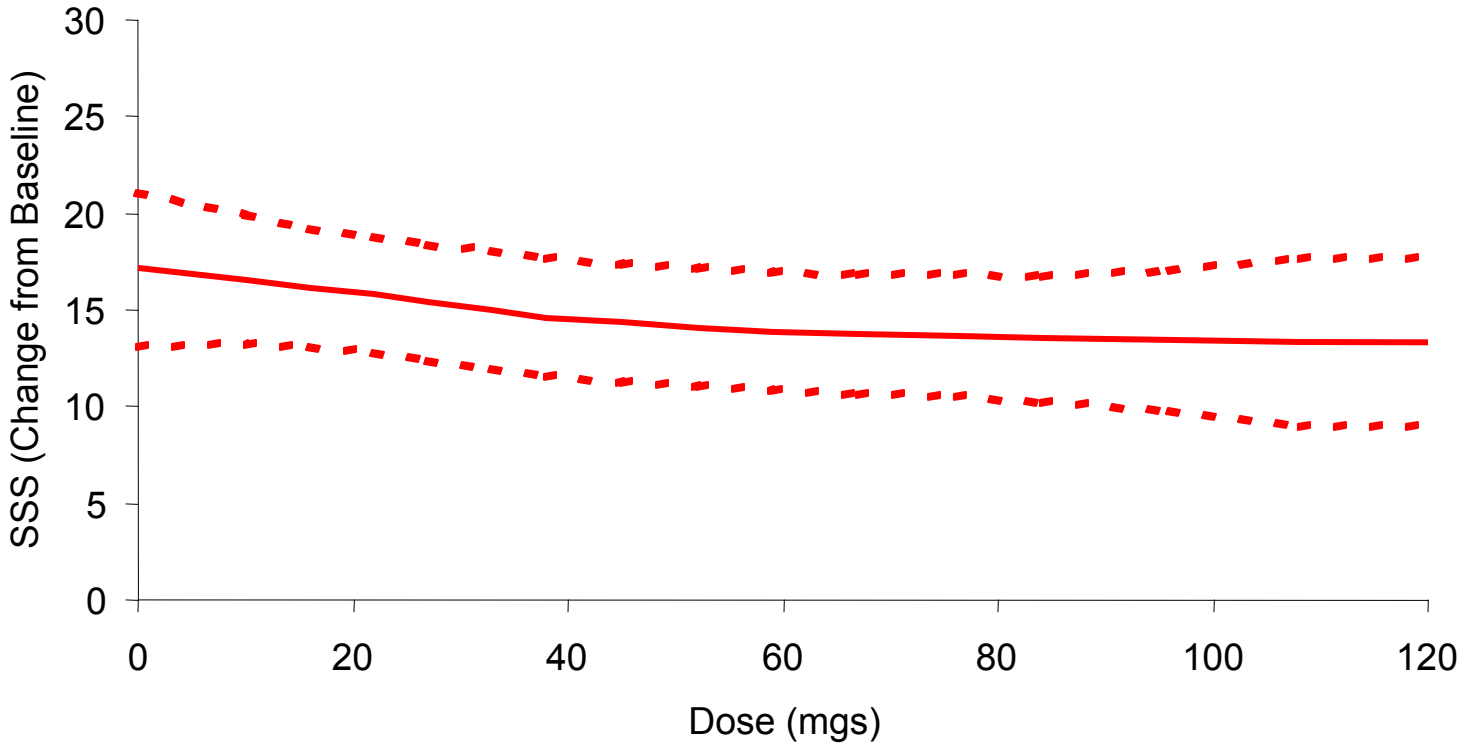
Week - 6



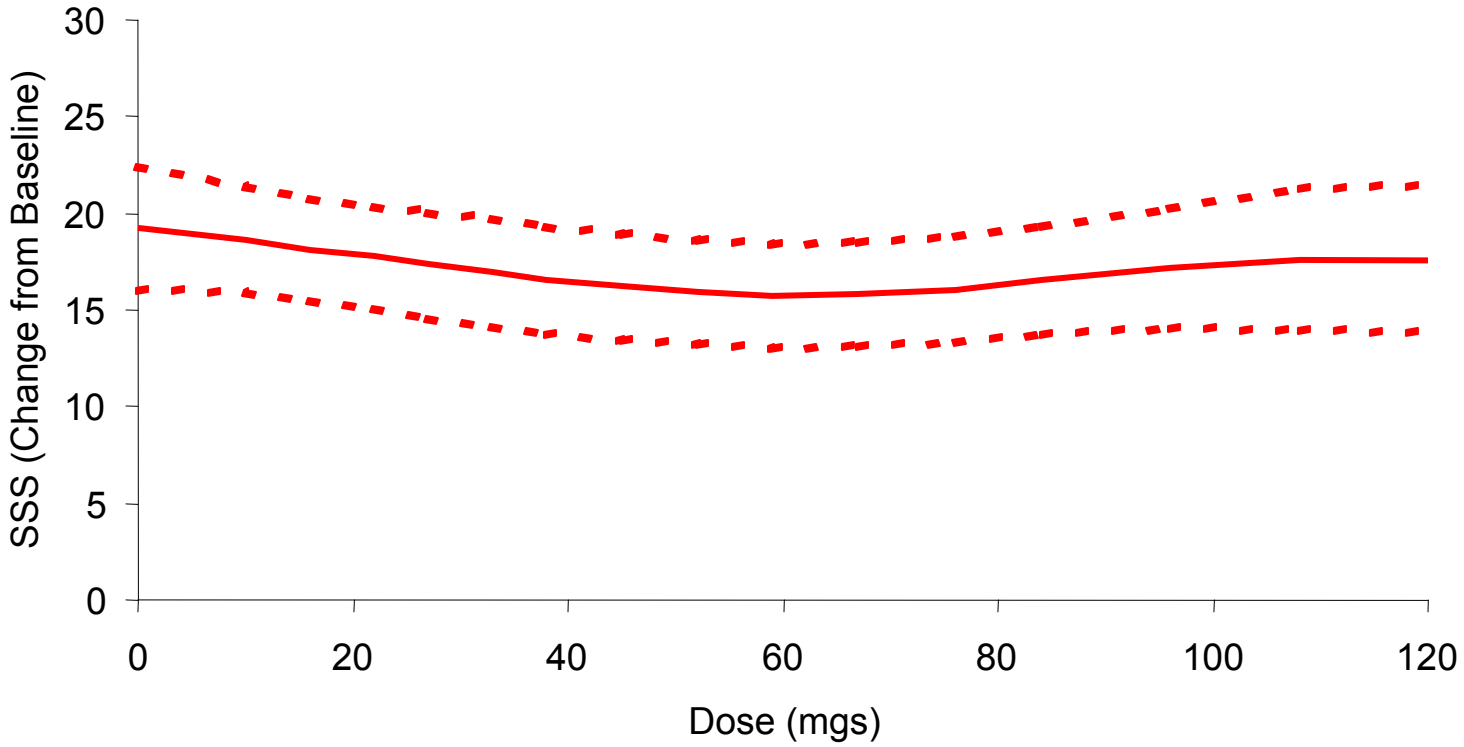
Week - 8



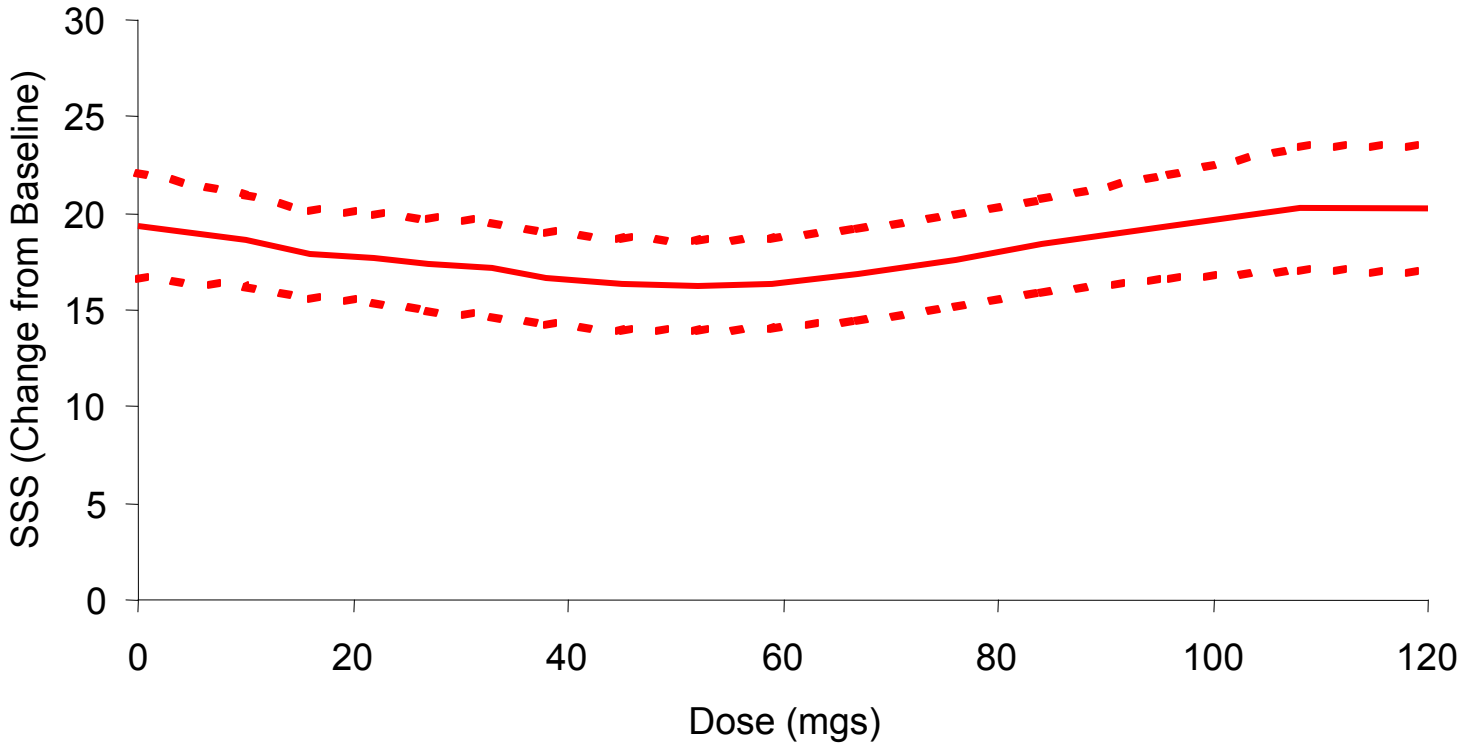
Week - 10



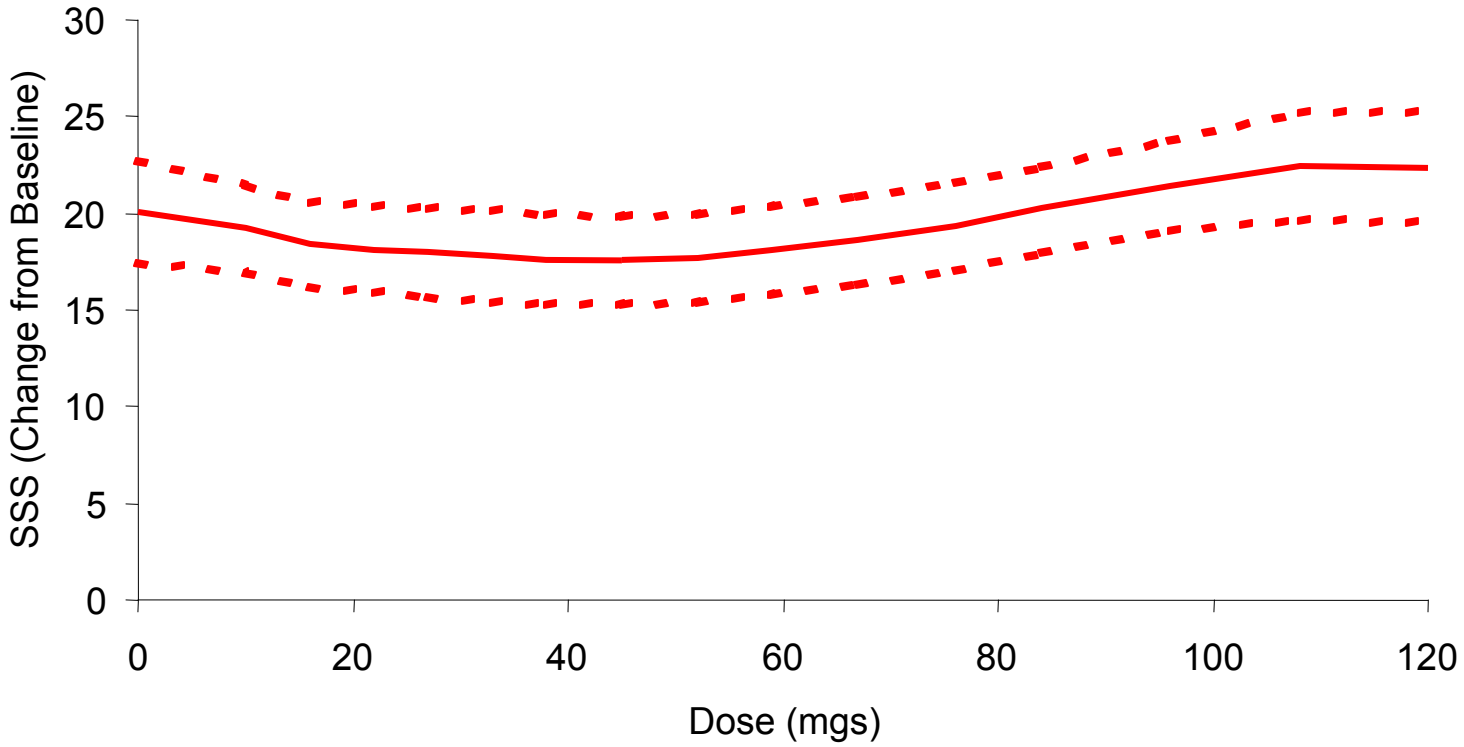
Week - 12



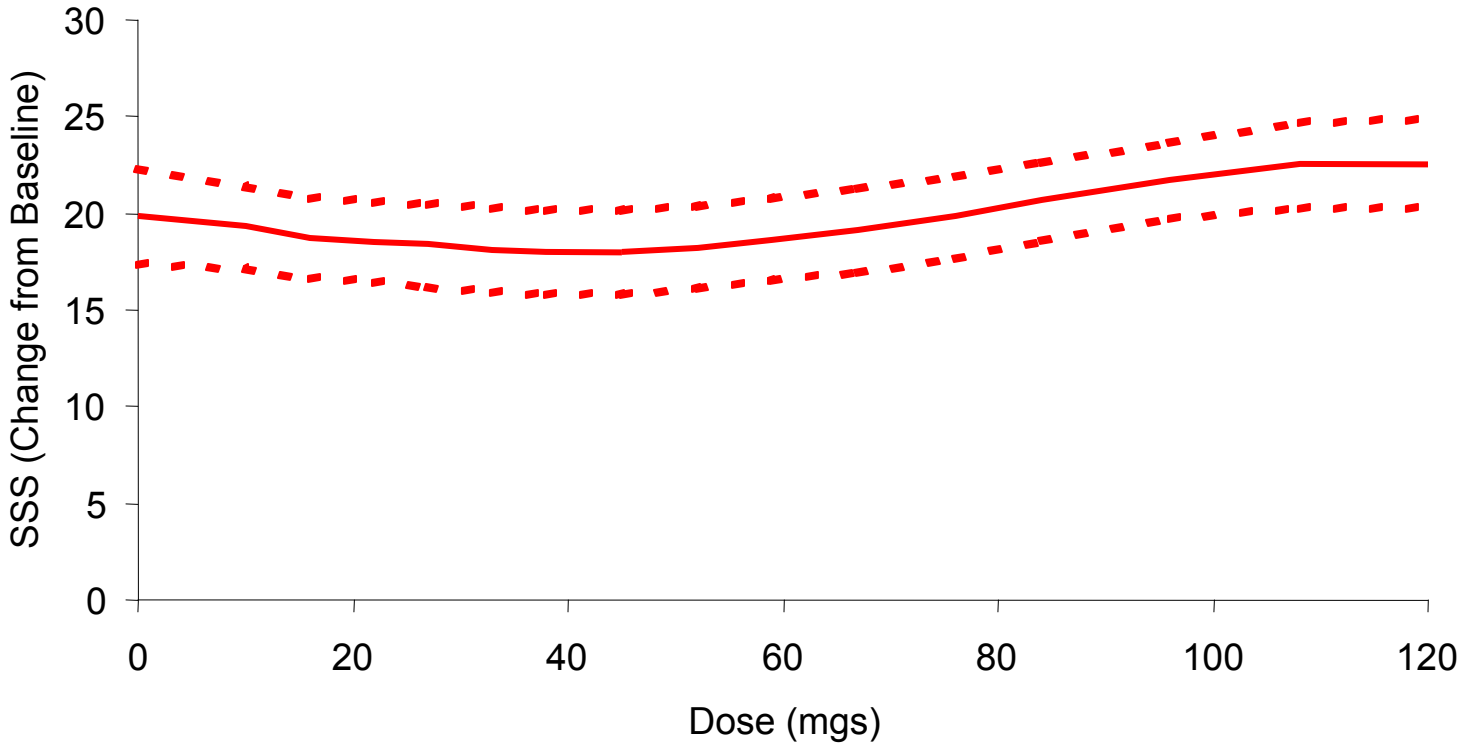
Week - 14



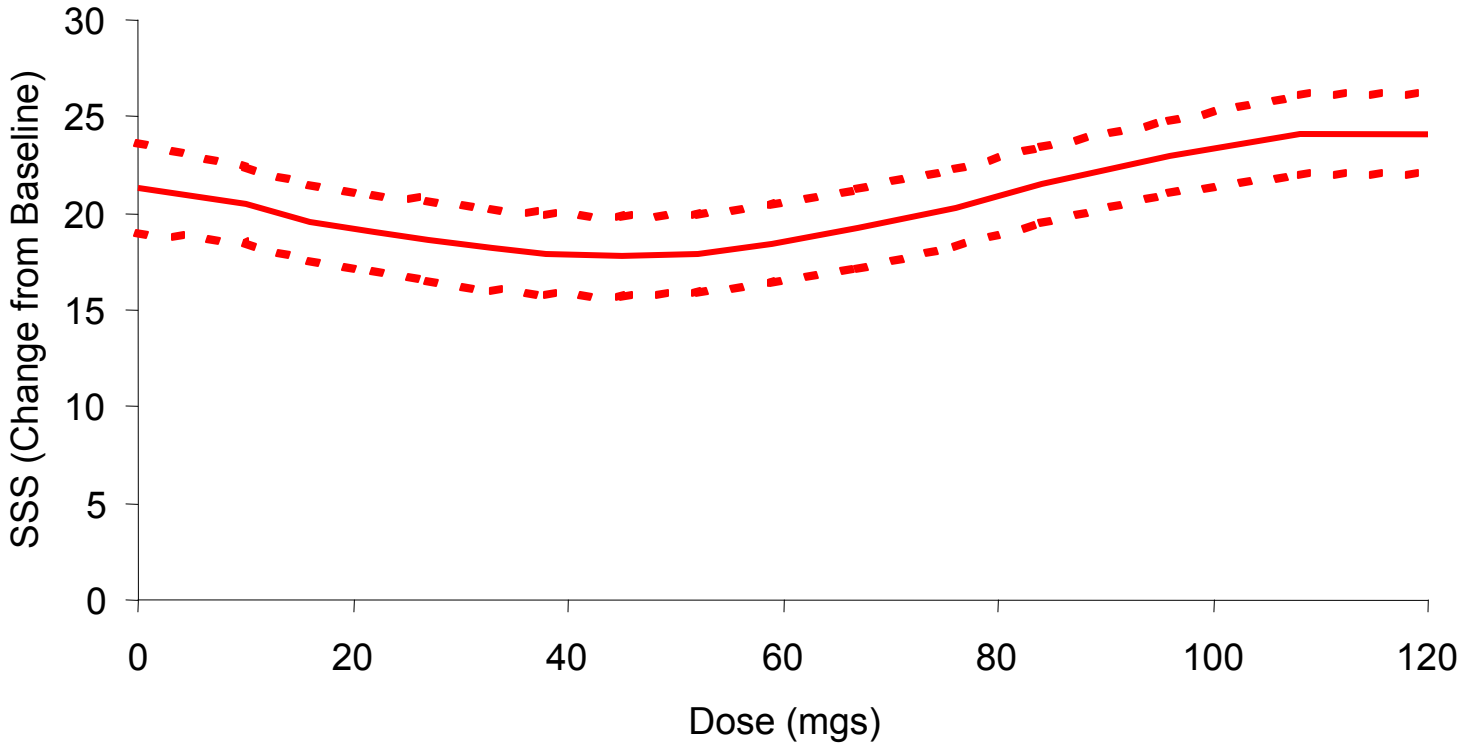
Week - 16



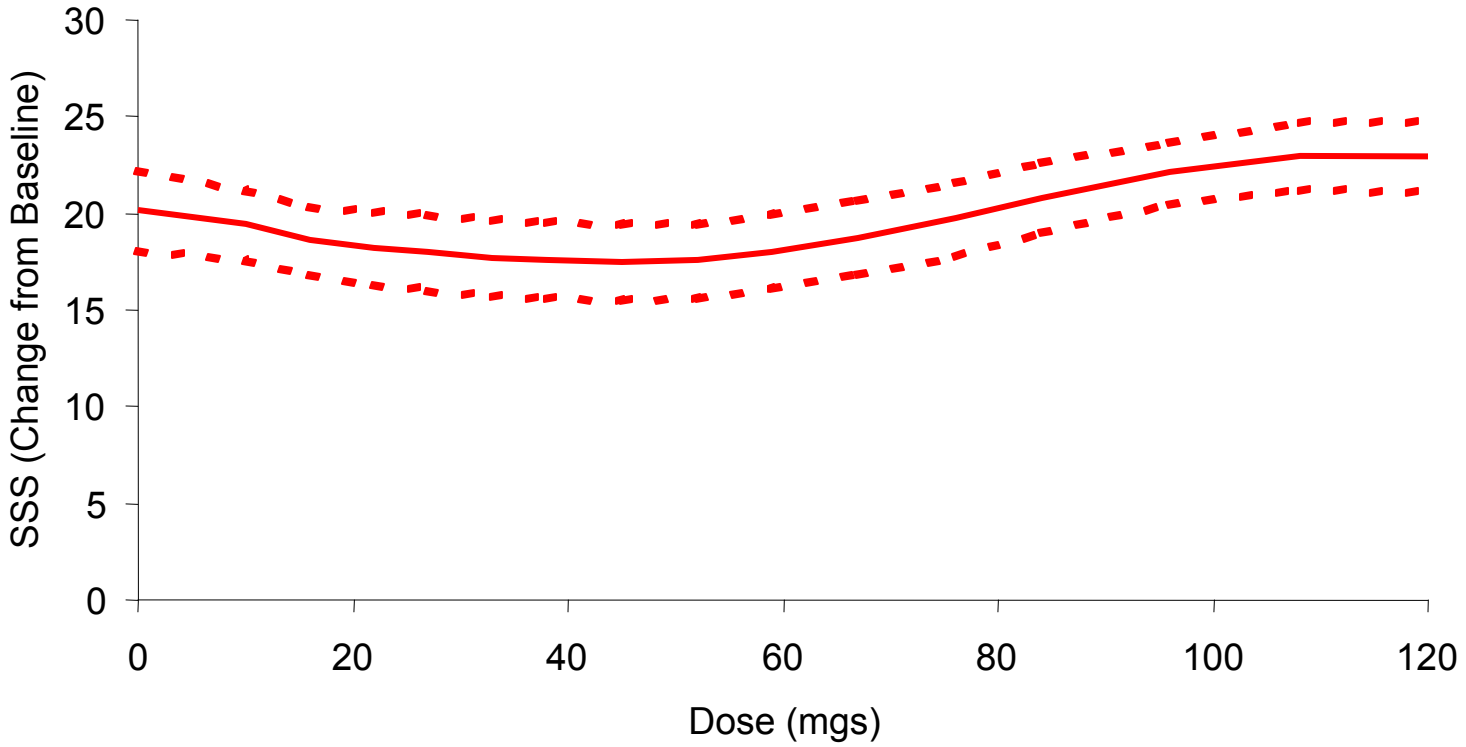
Week - 18



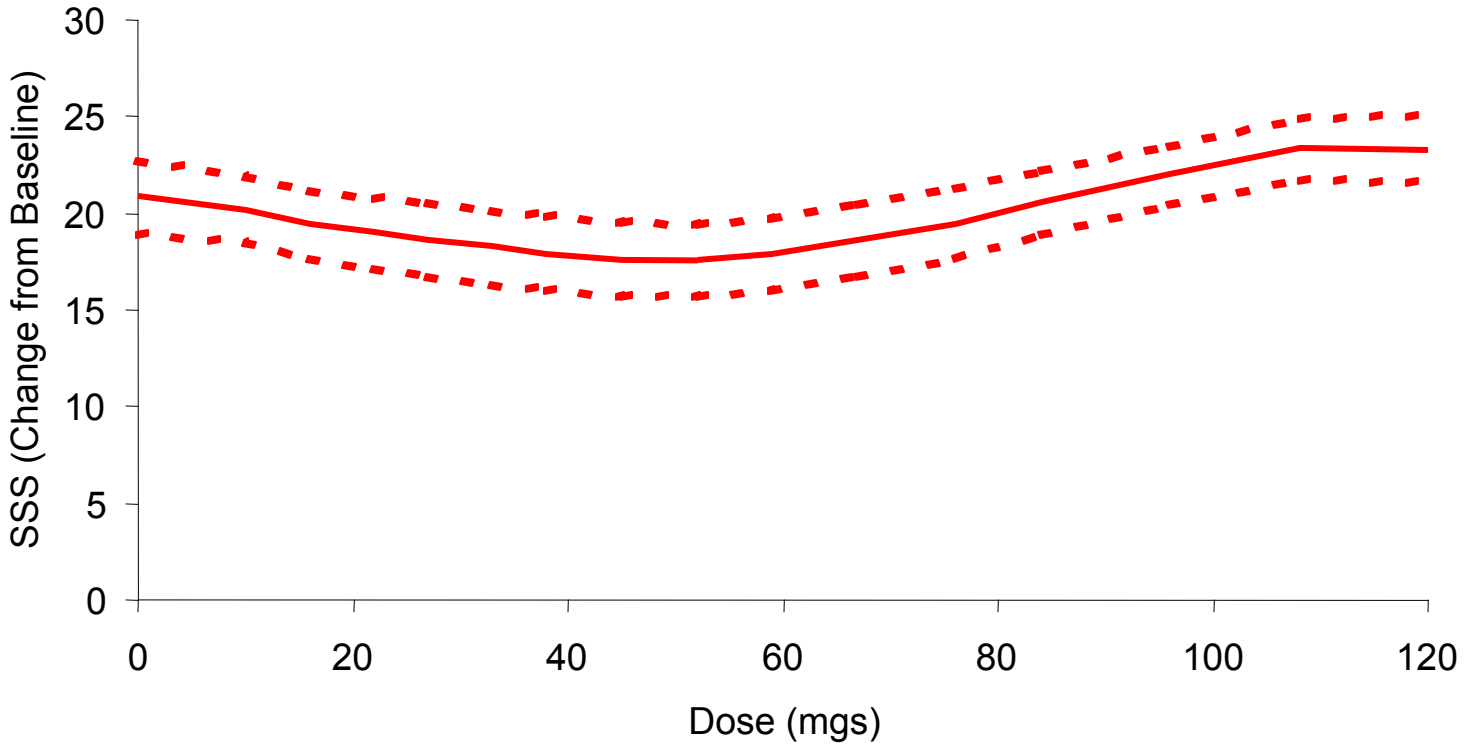
Week - 20



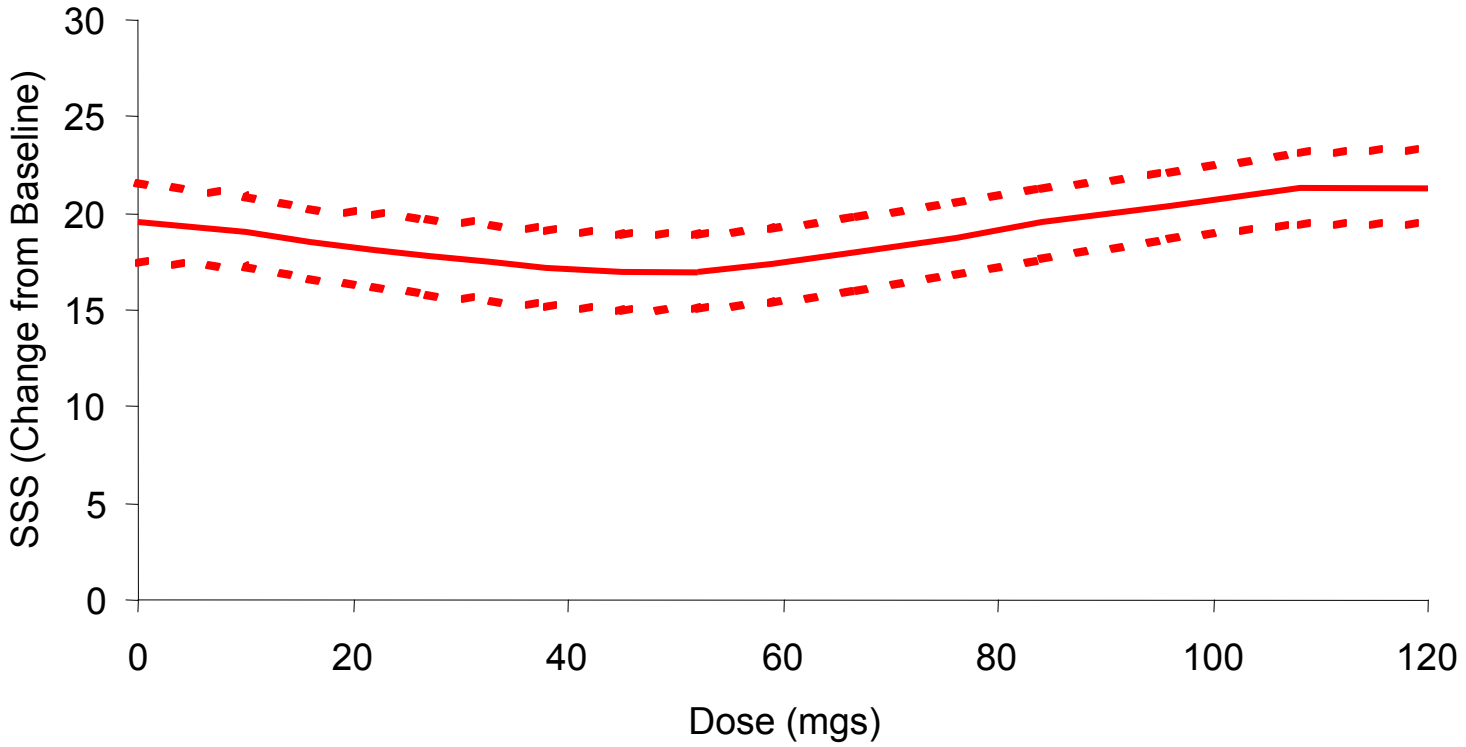
Week - 22



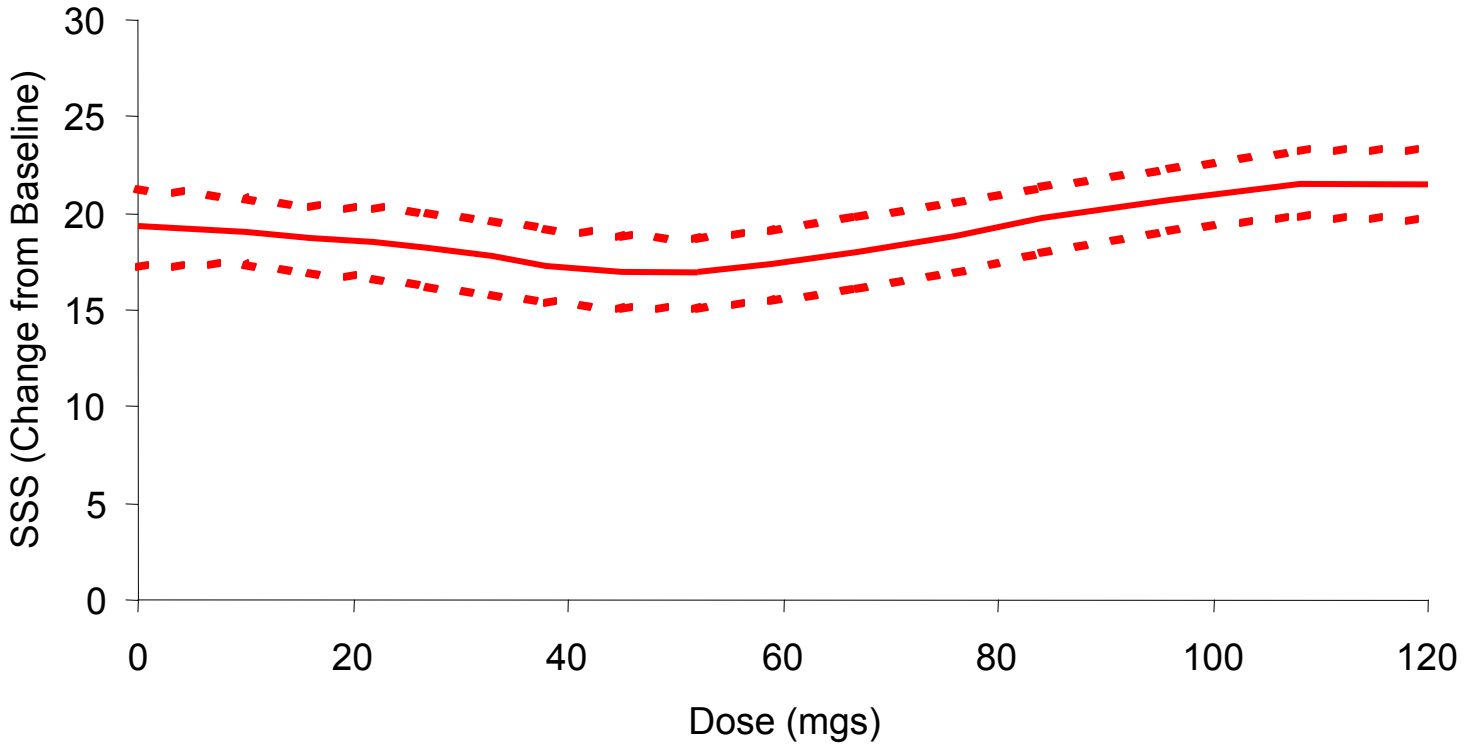
Week - 24



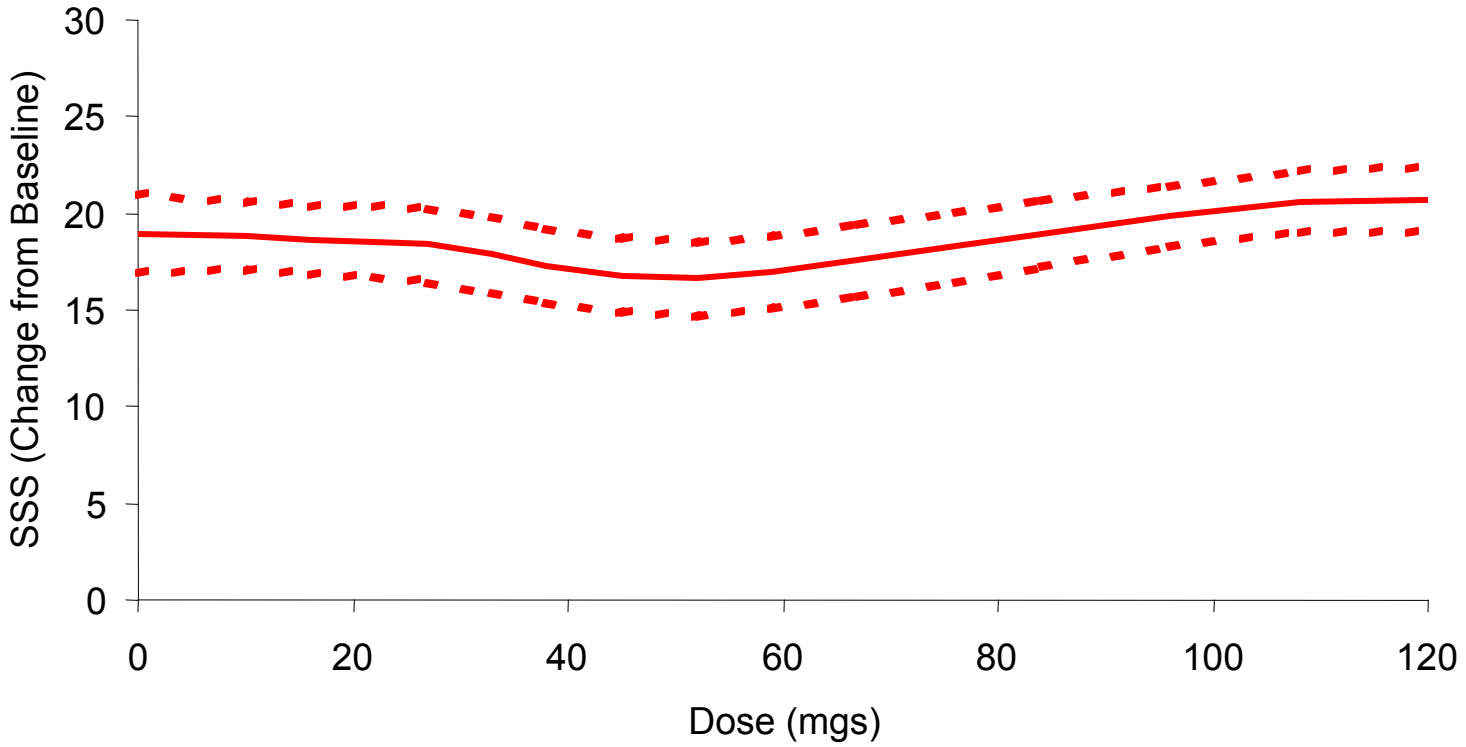
Week - 26



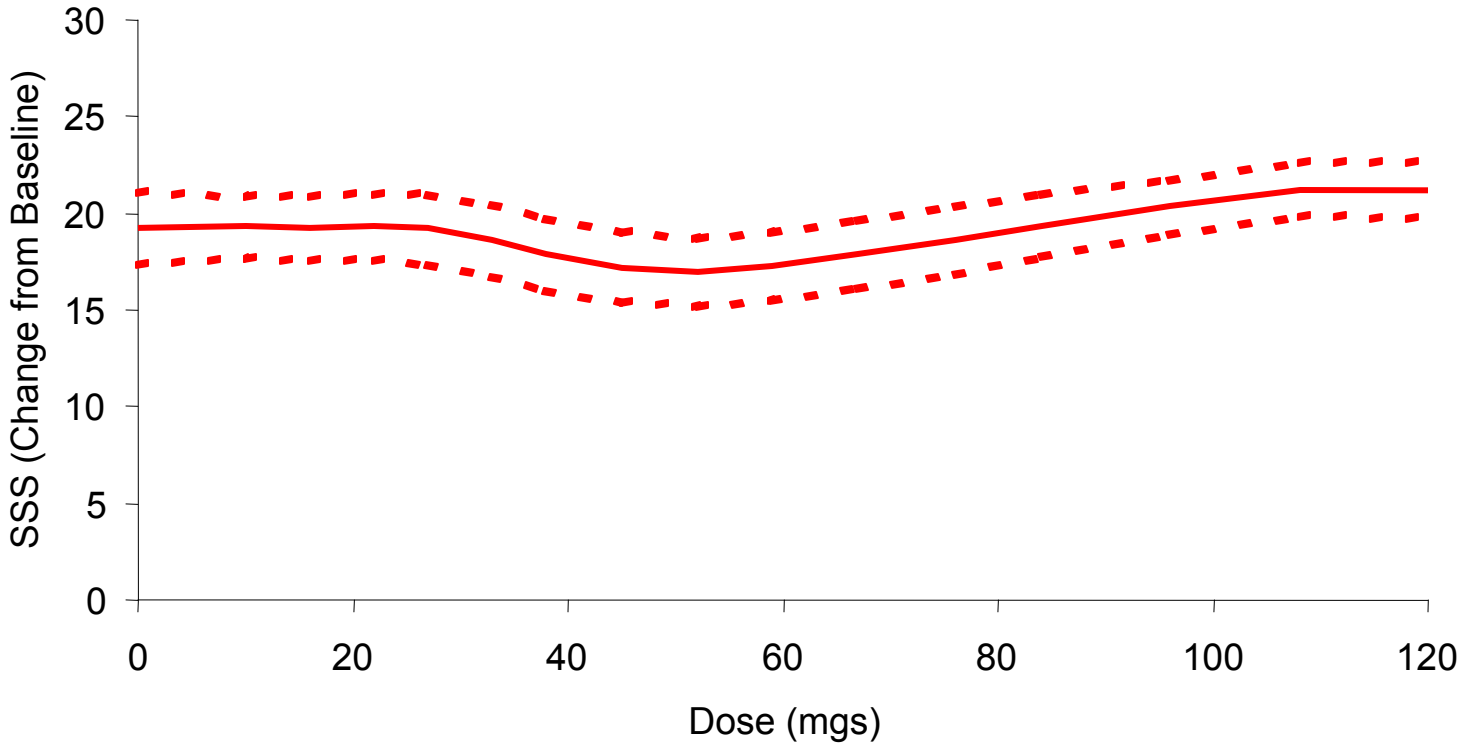
Week - 28



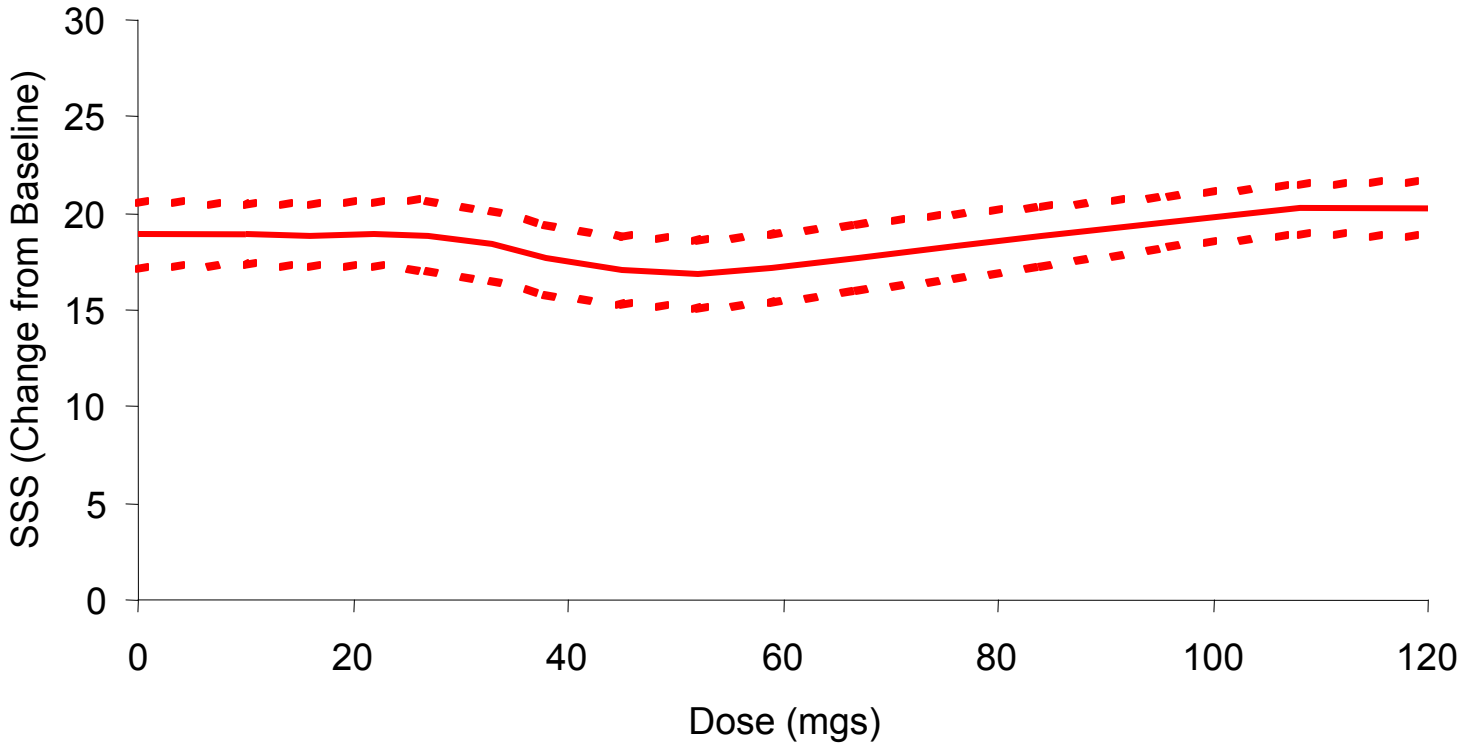
Week - 30



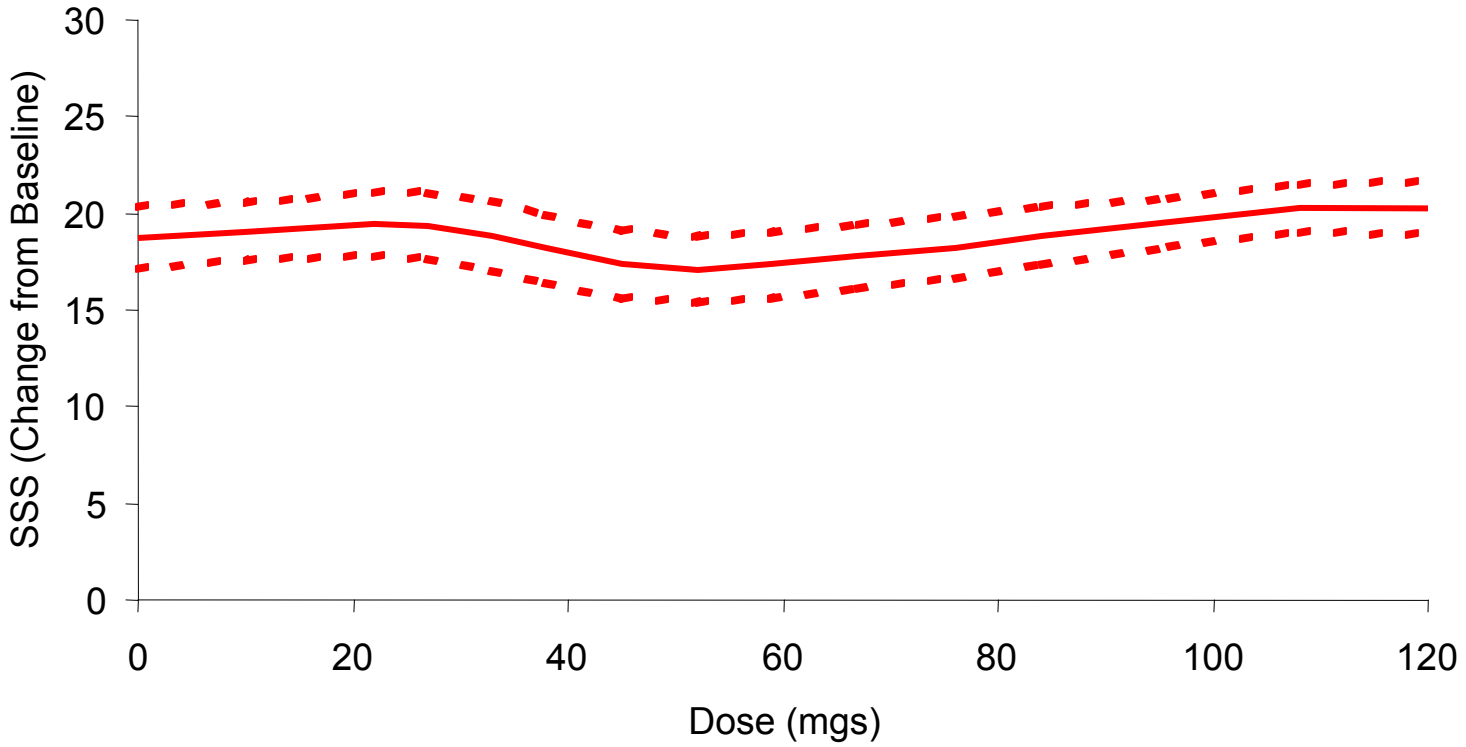
Week - 32



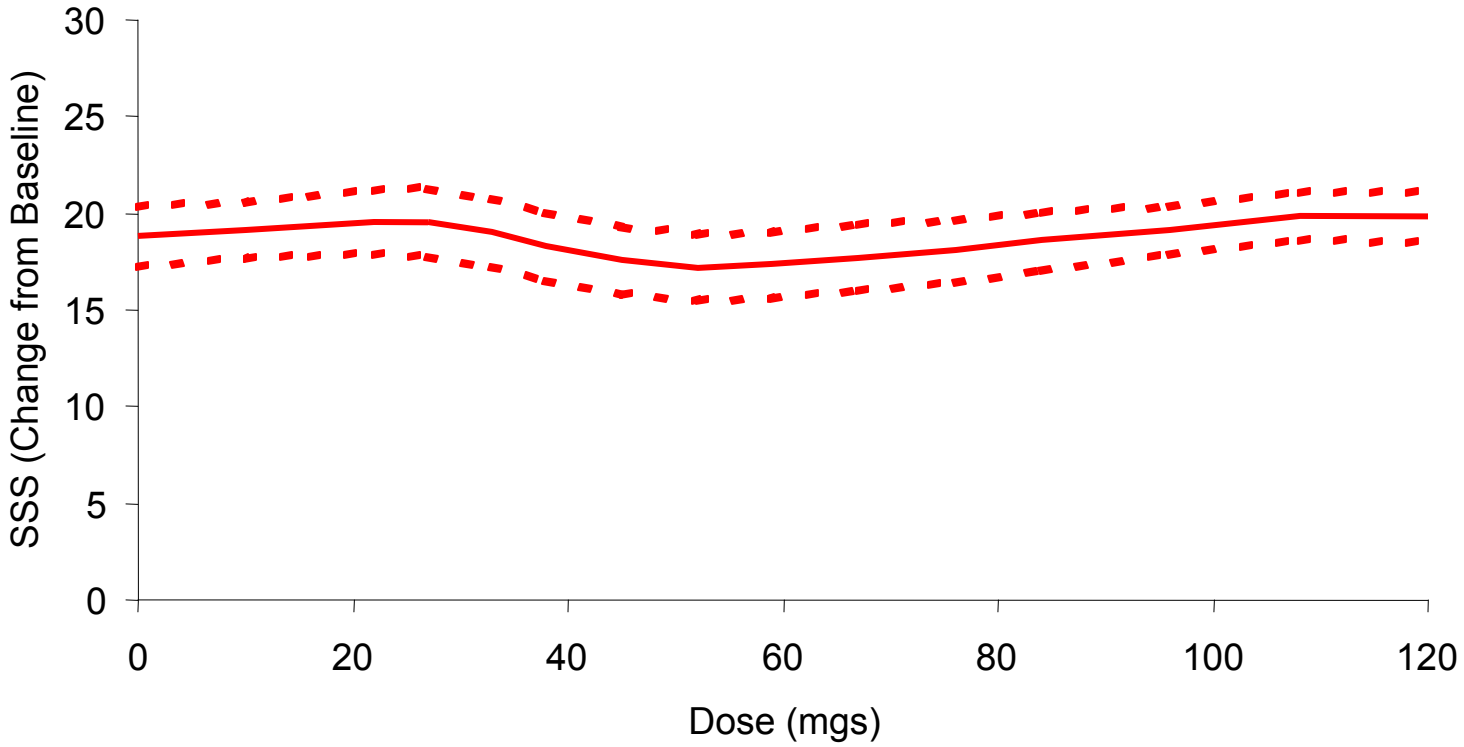
Week - 34



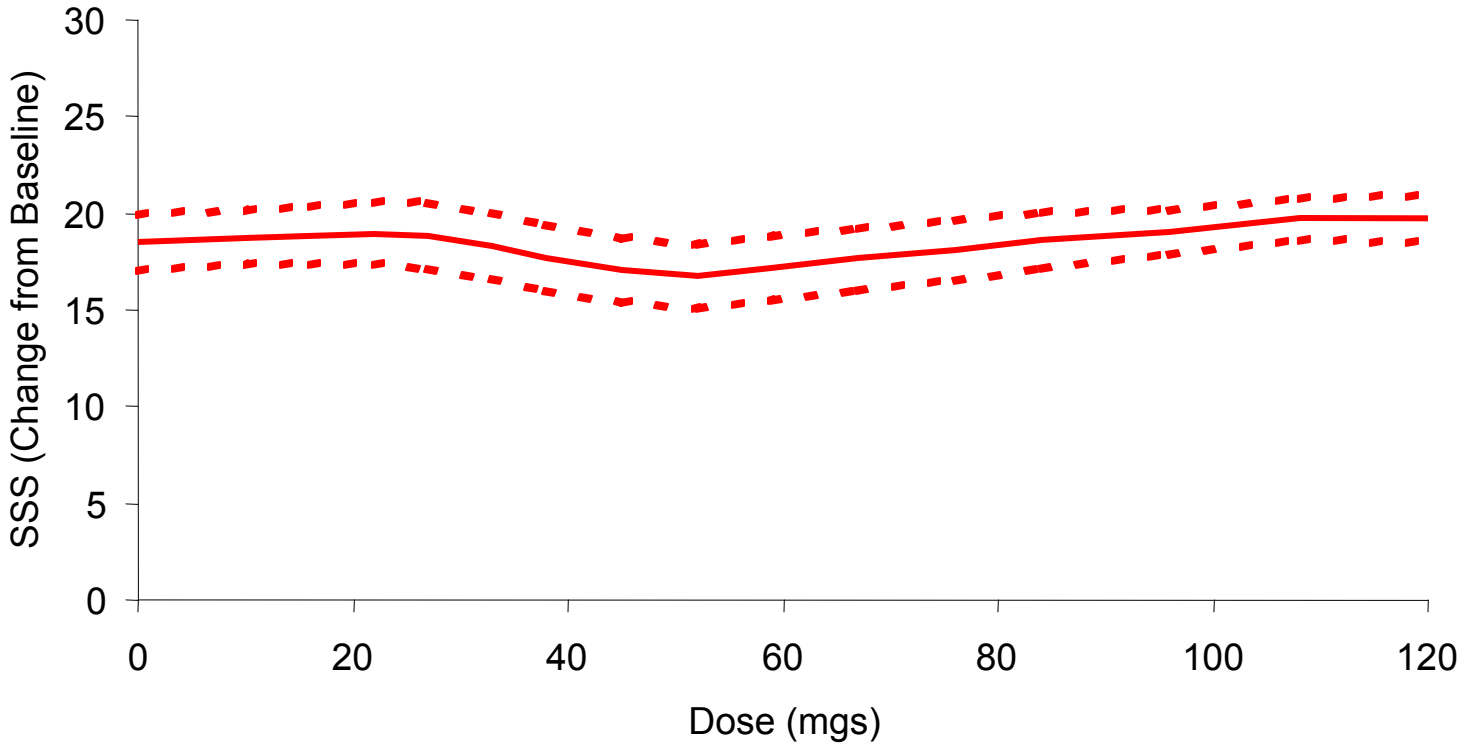
Week - 36



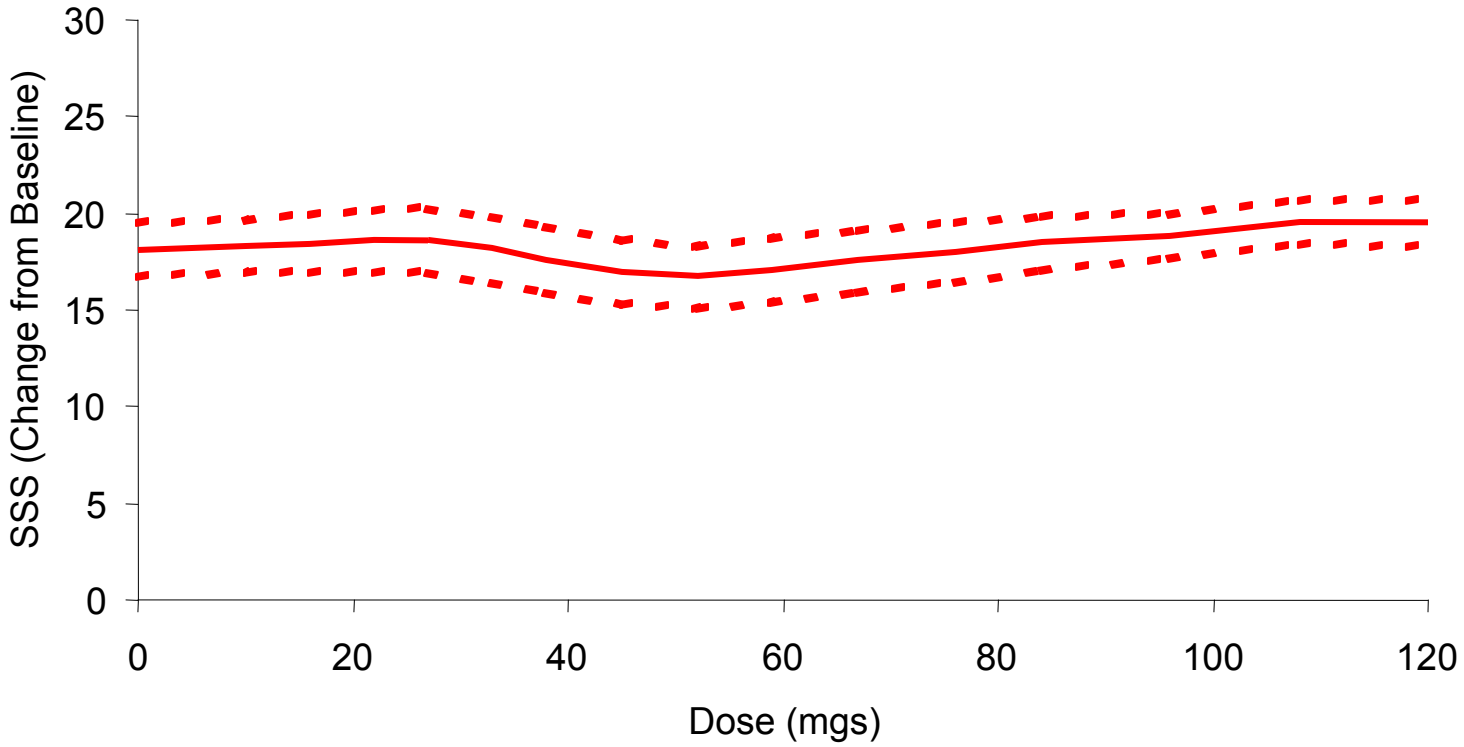
Week - 38



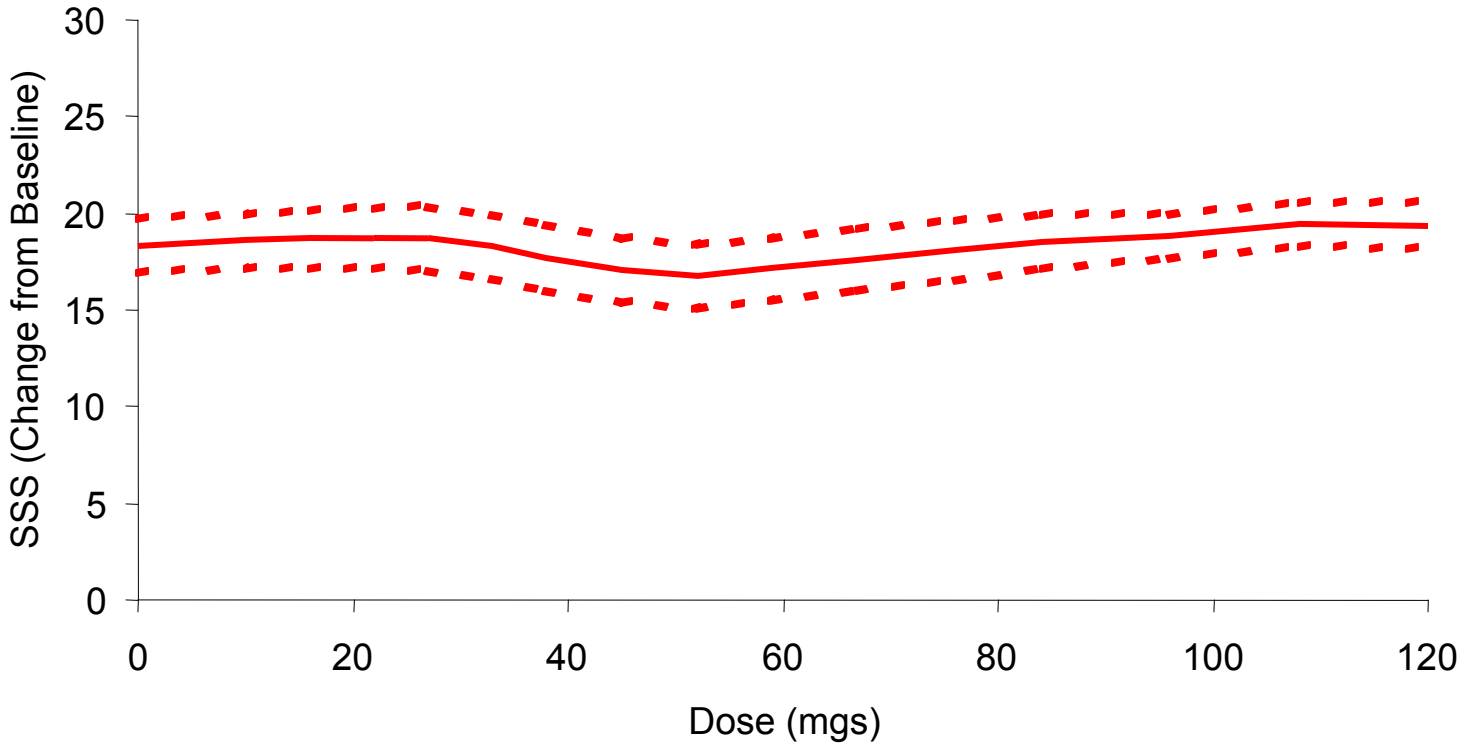
Week - 40



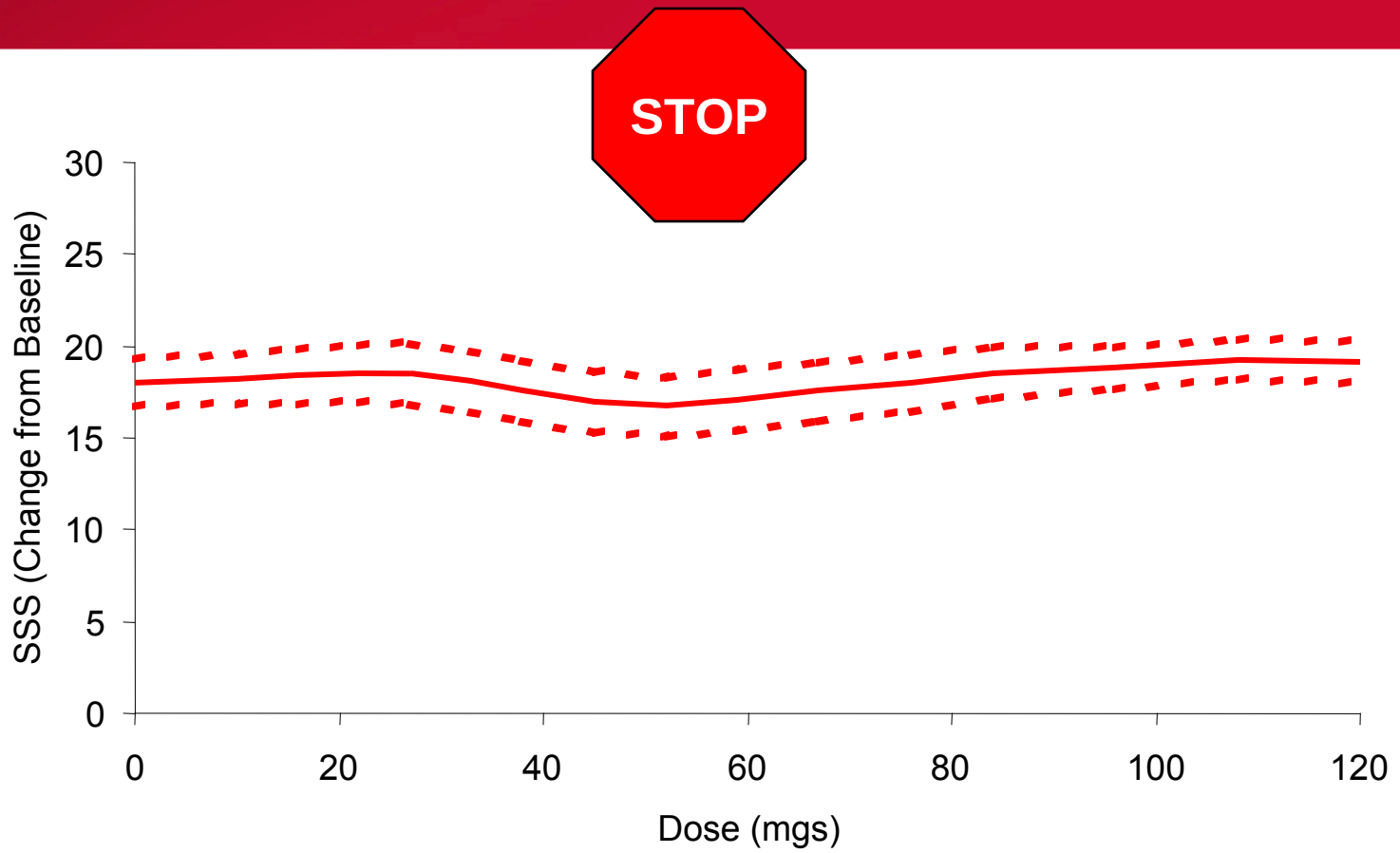
Week - 42



Week - 44

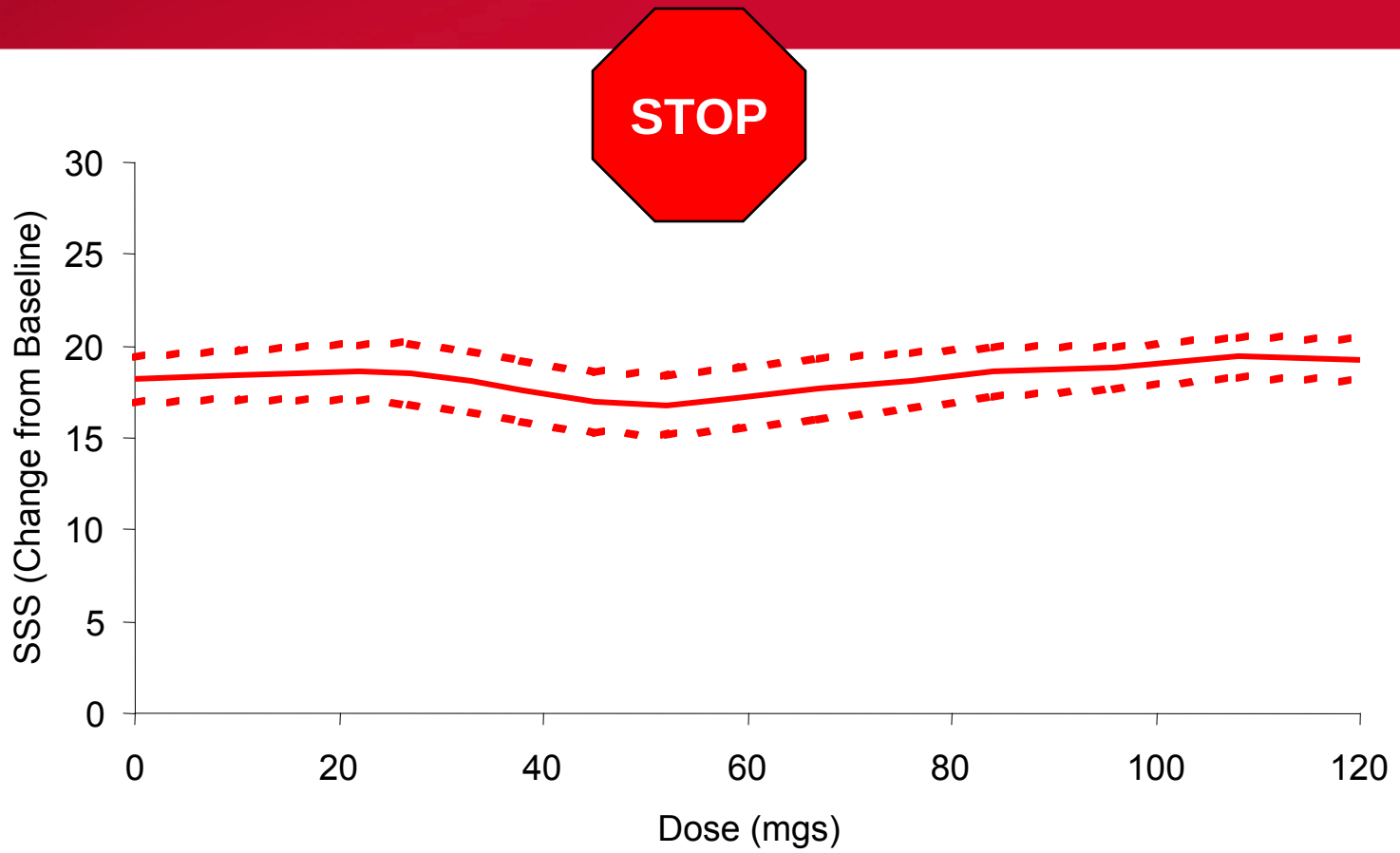


Week - 46

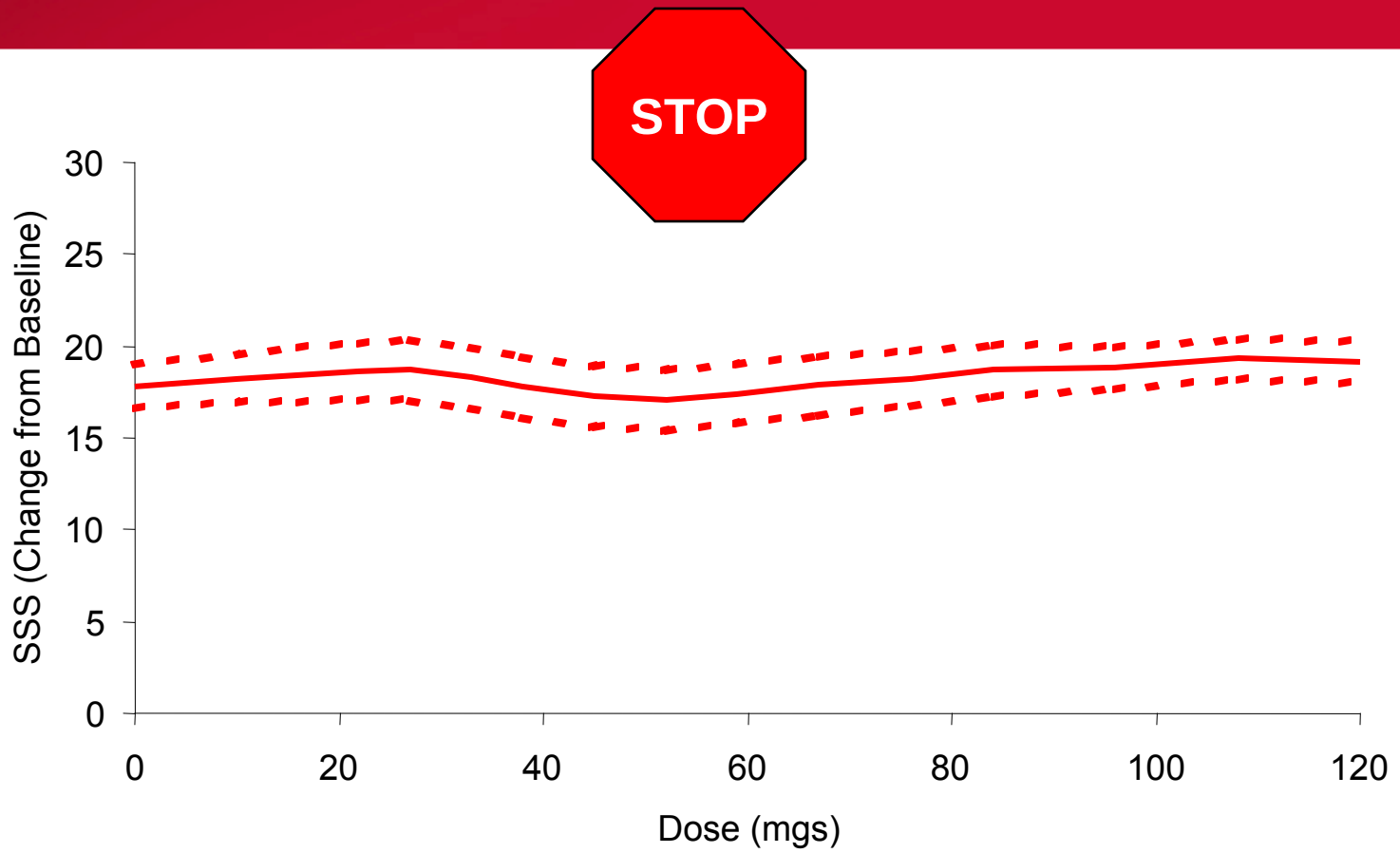


Stopped by IDMC

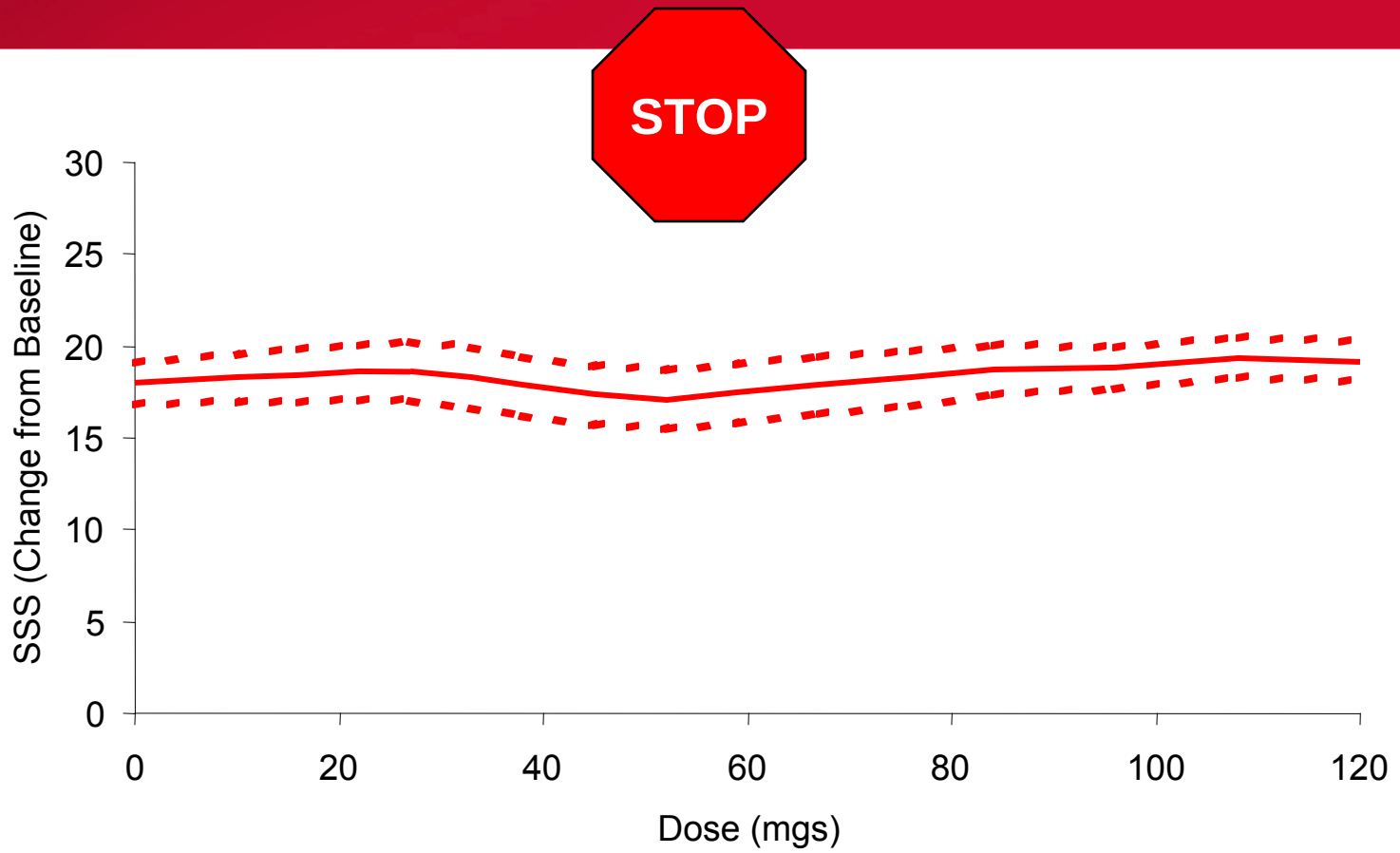
Week - 48



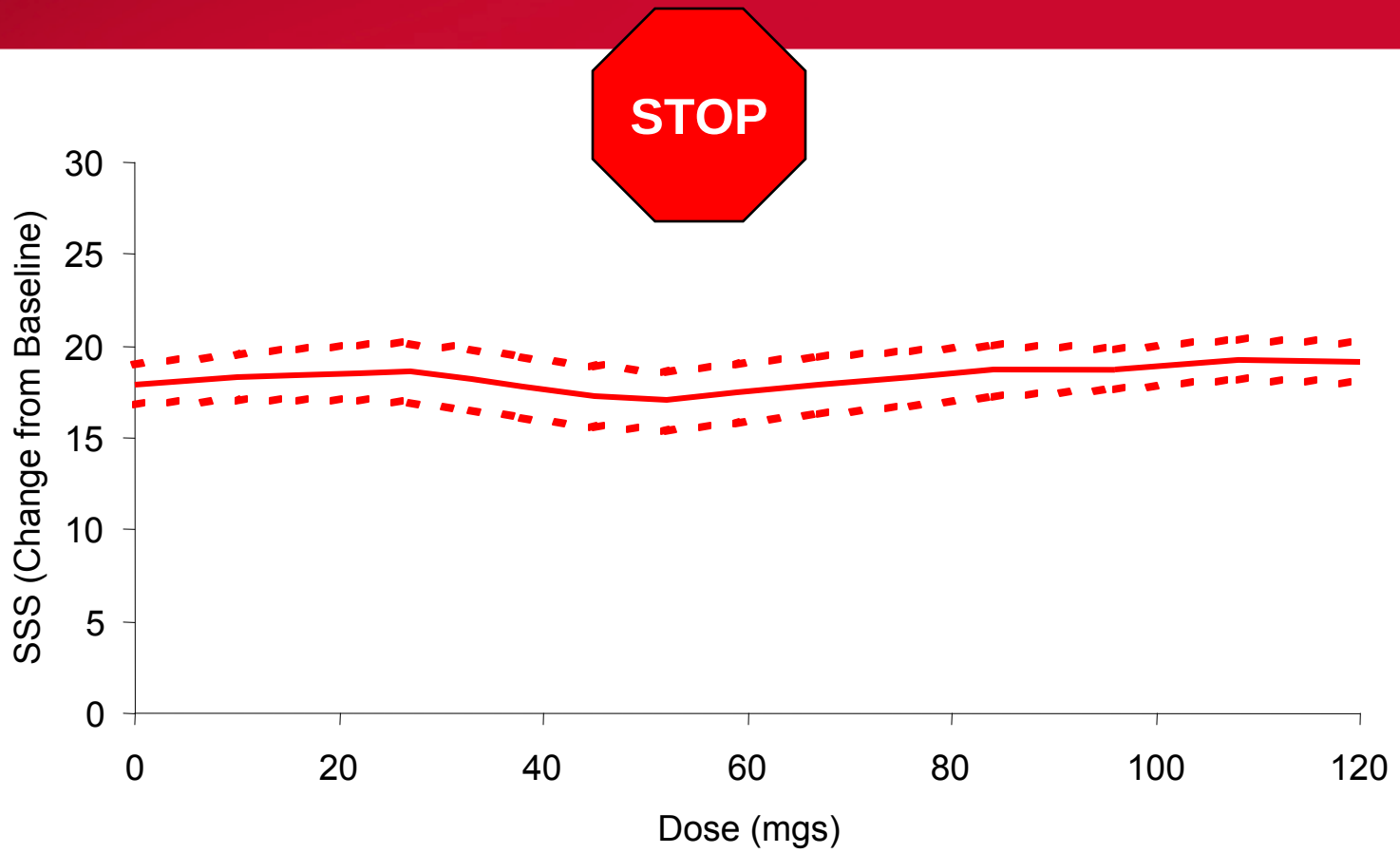
Week - 50



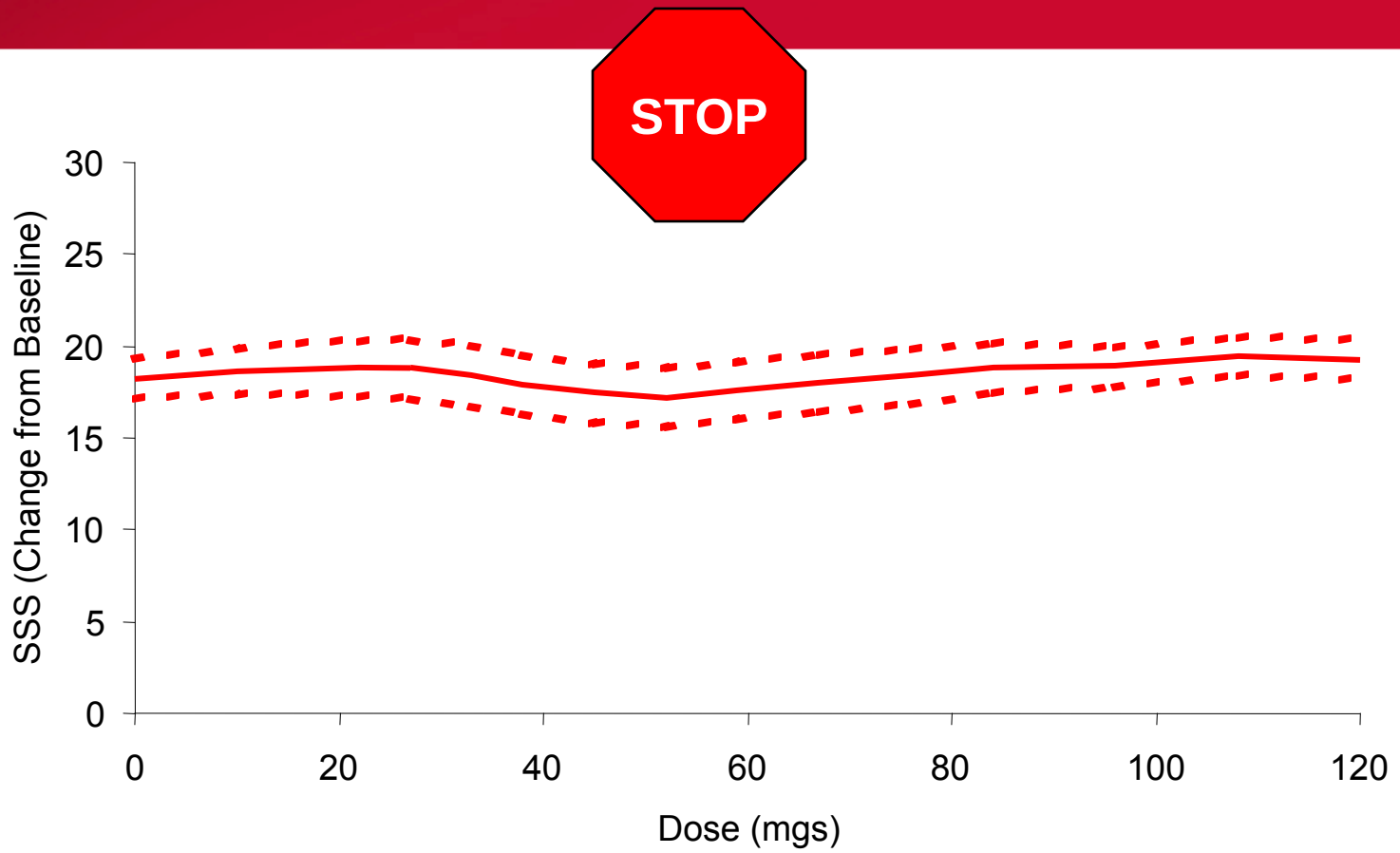
Week - 52



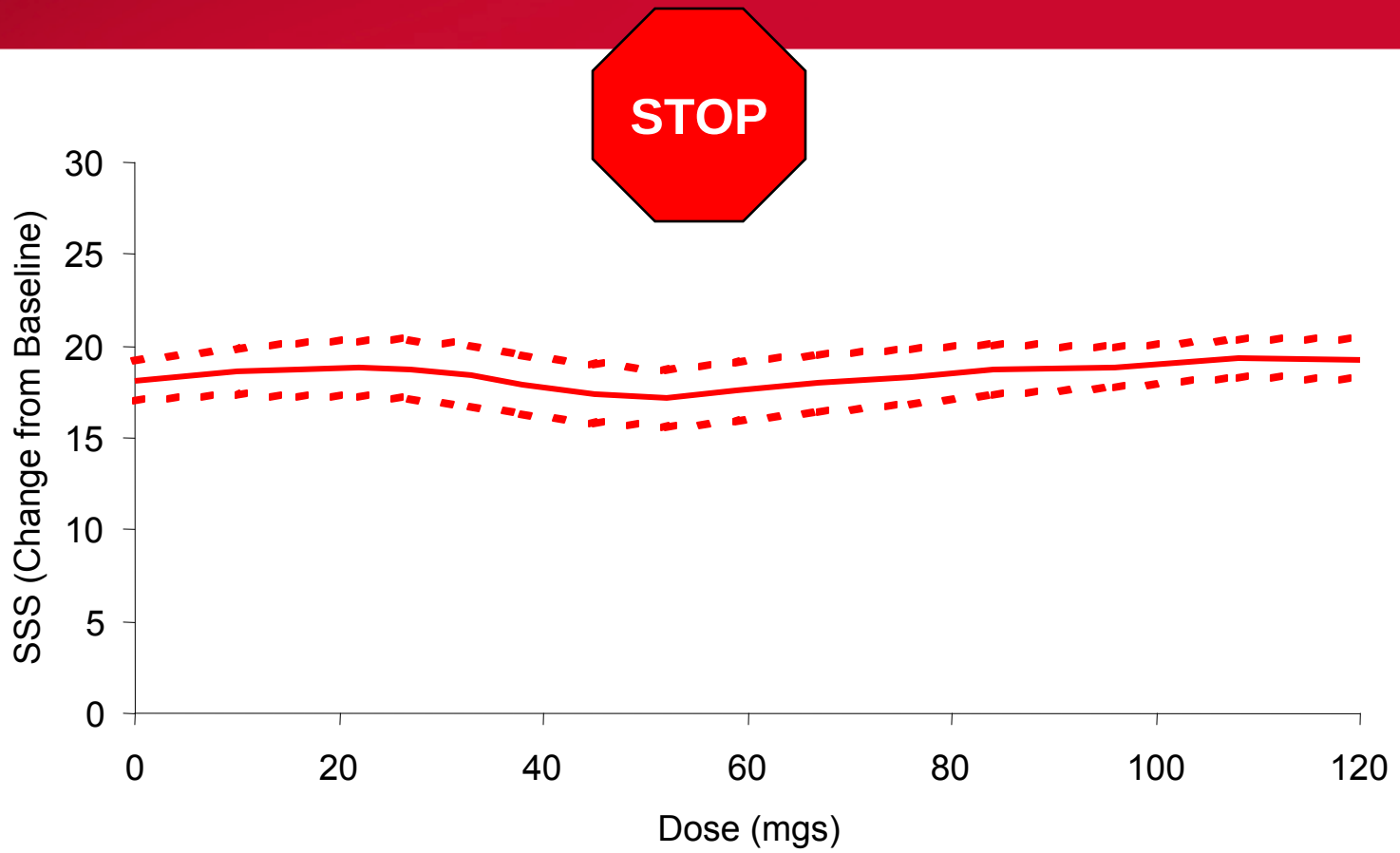
Week - 54



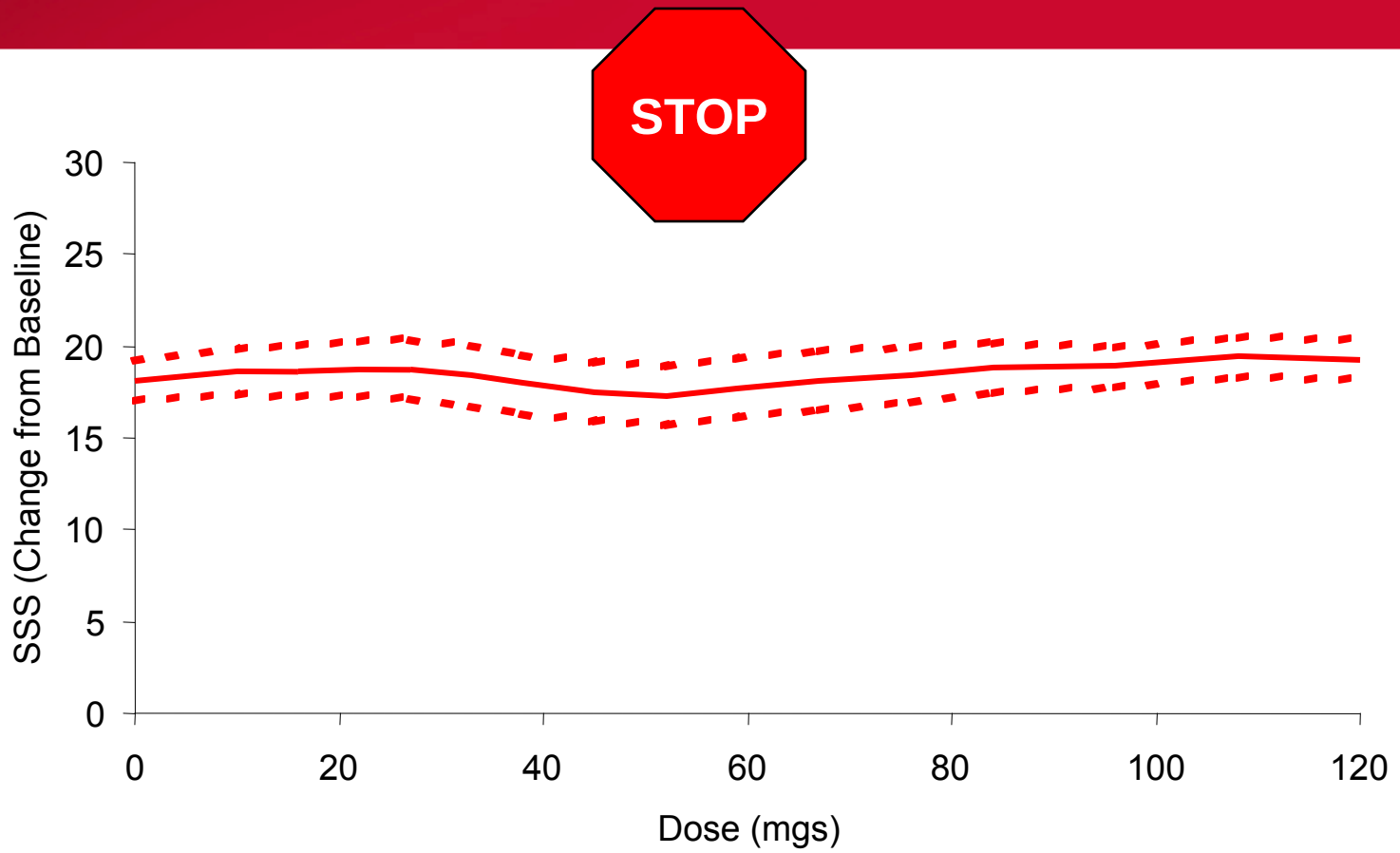
Week - 56



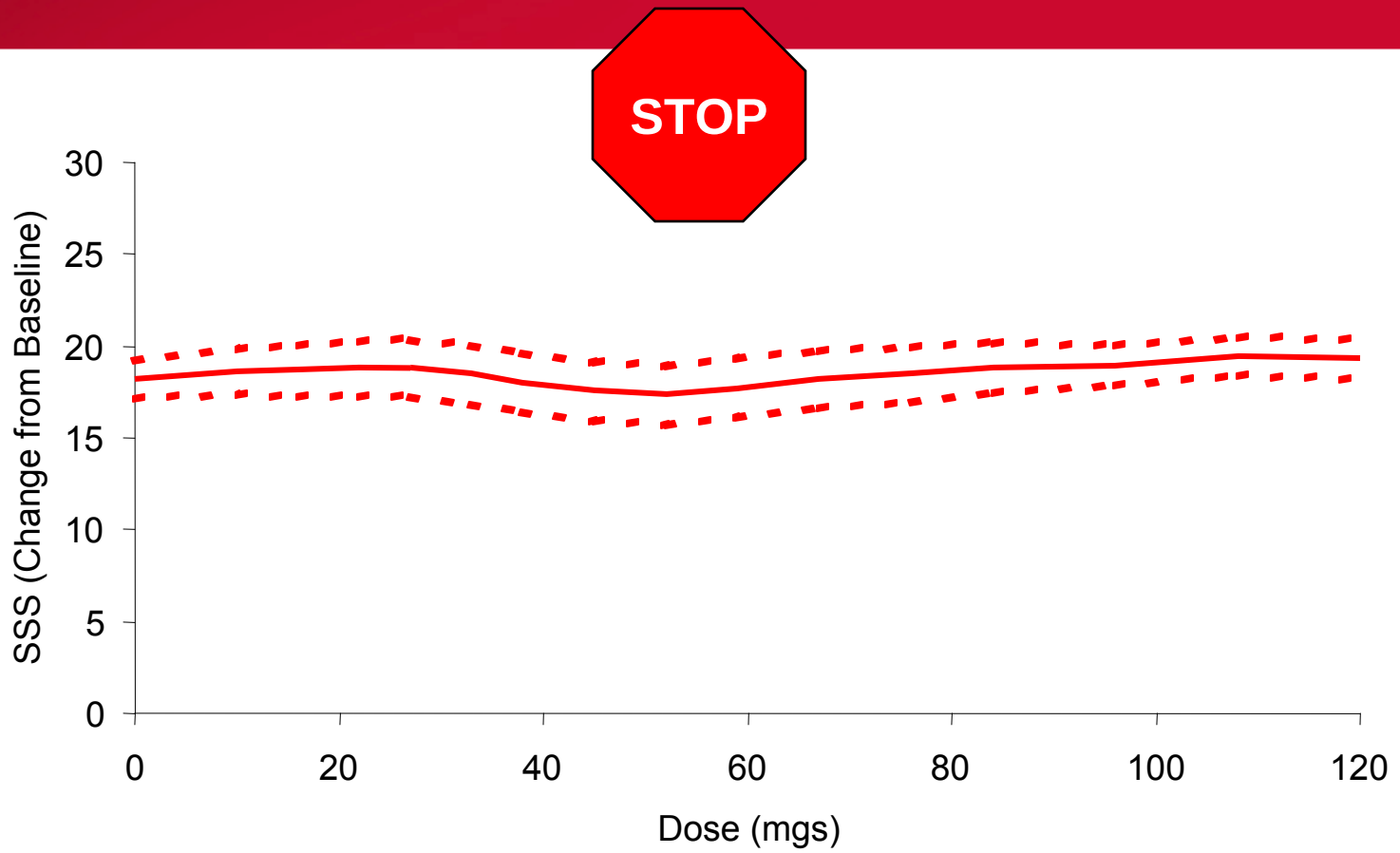
Week - 58



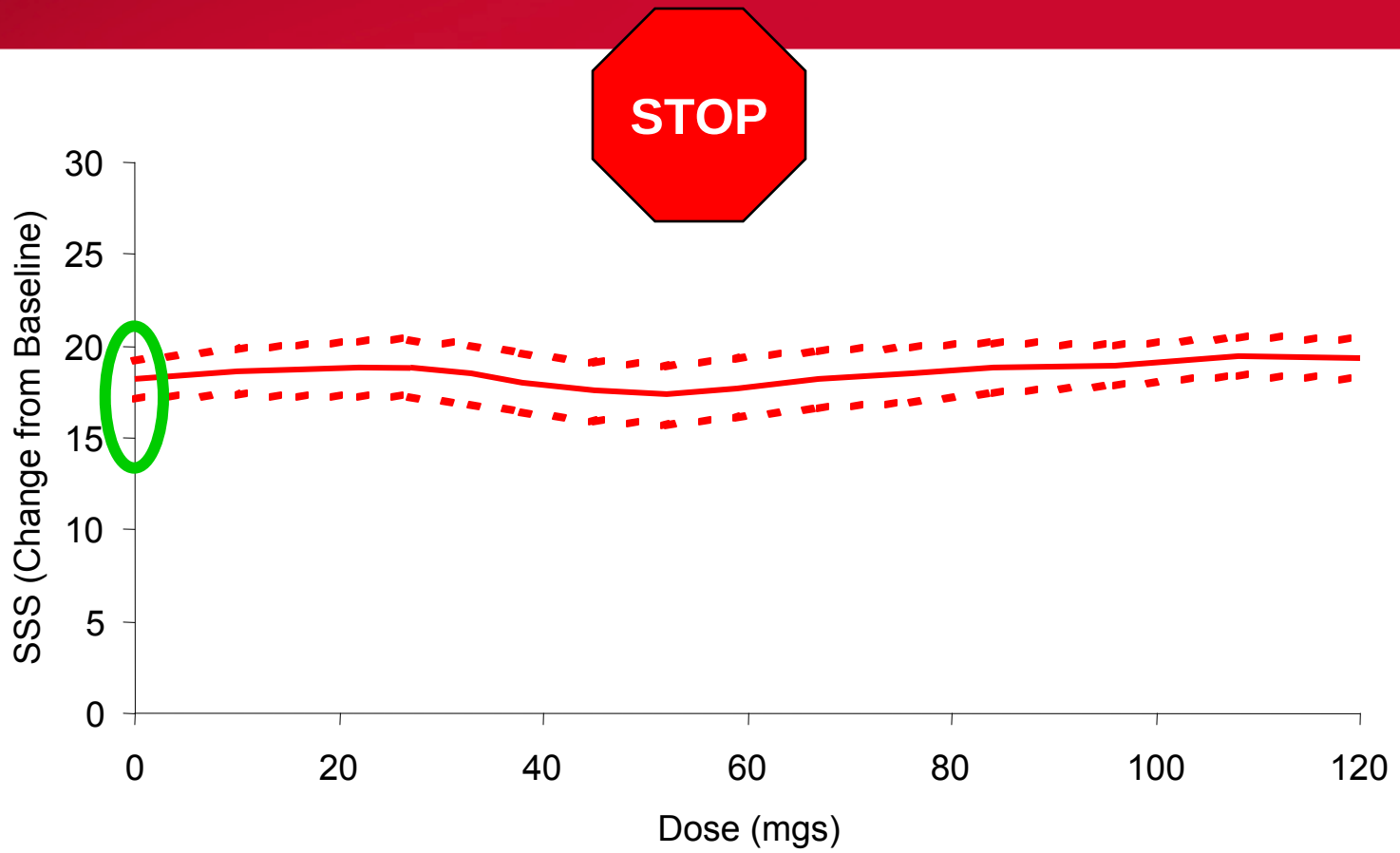
Week - 60



Week - 62

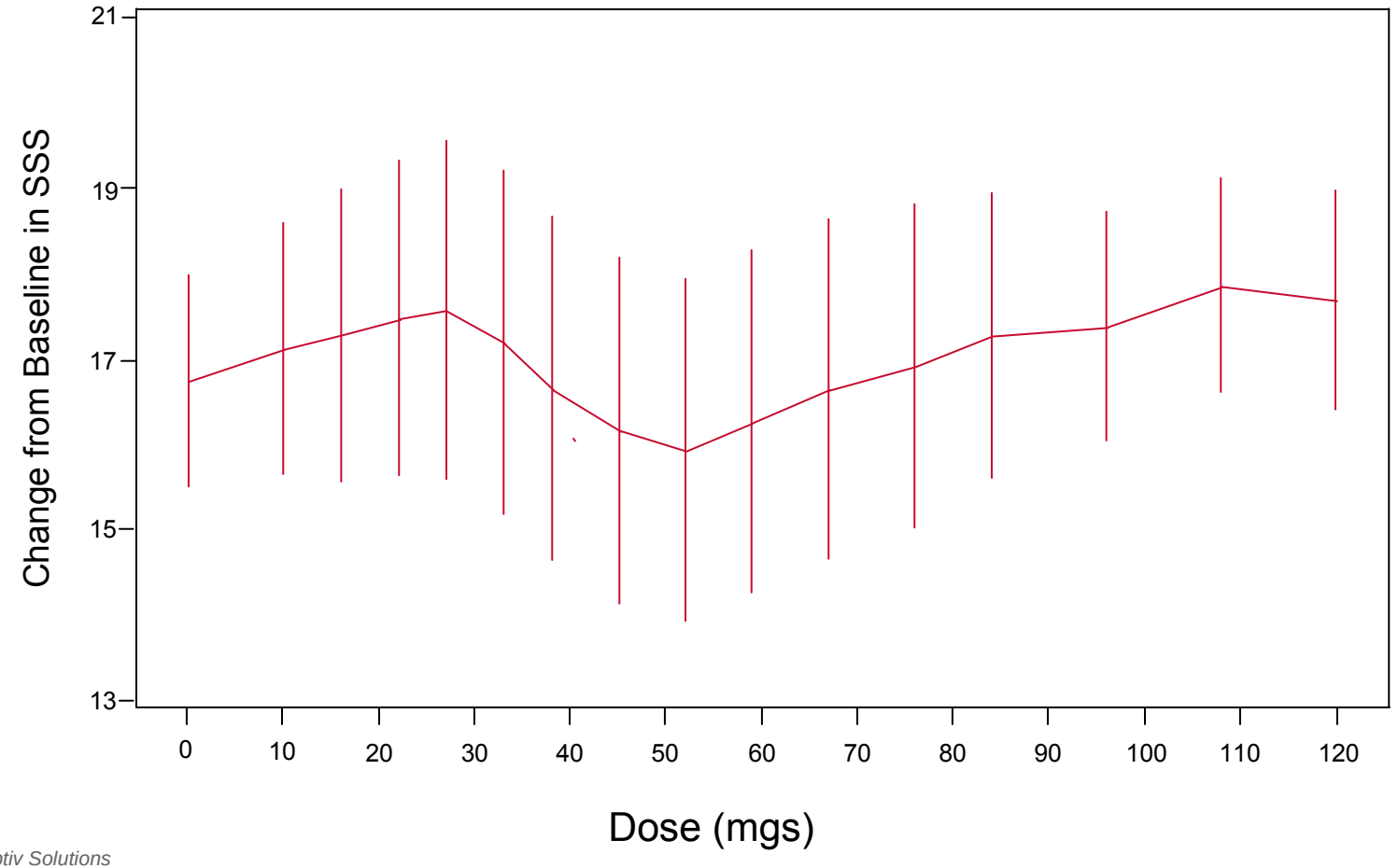


Week - 64

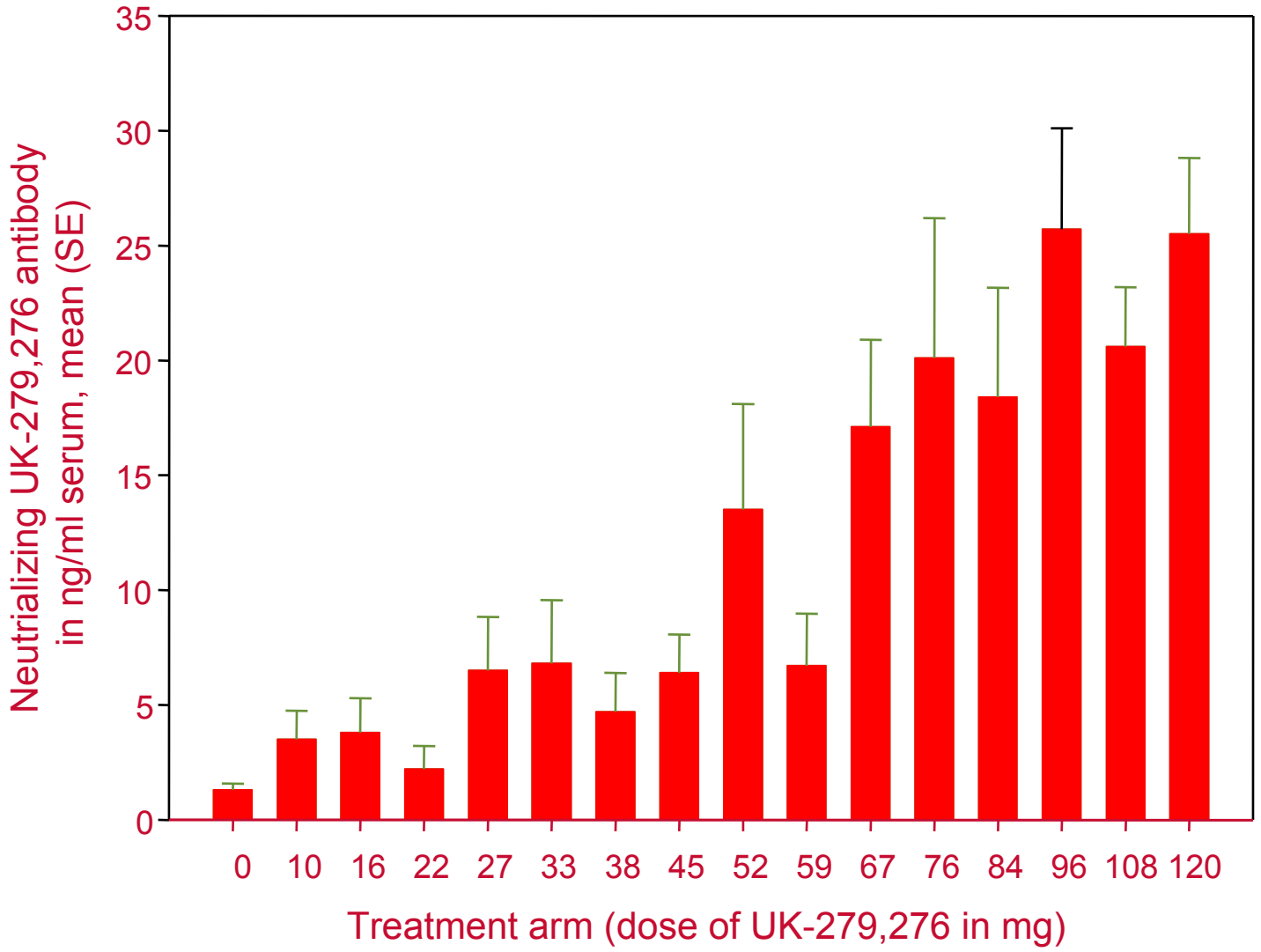


Week - 66

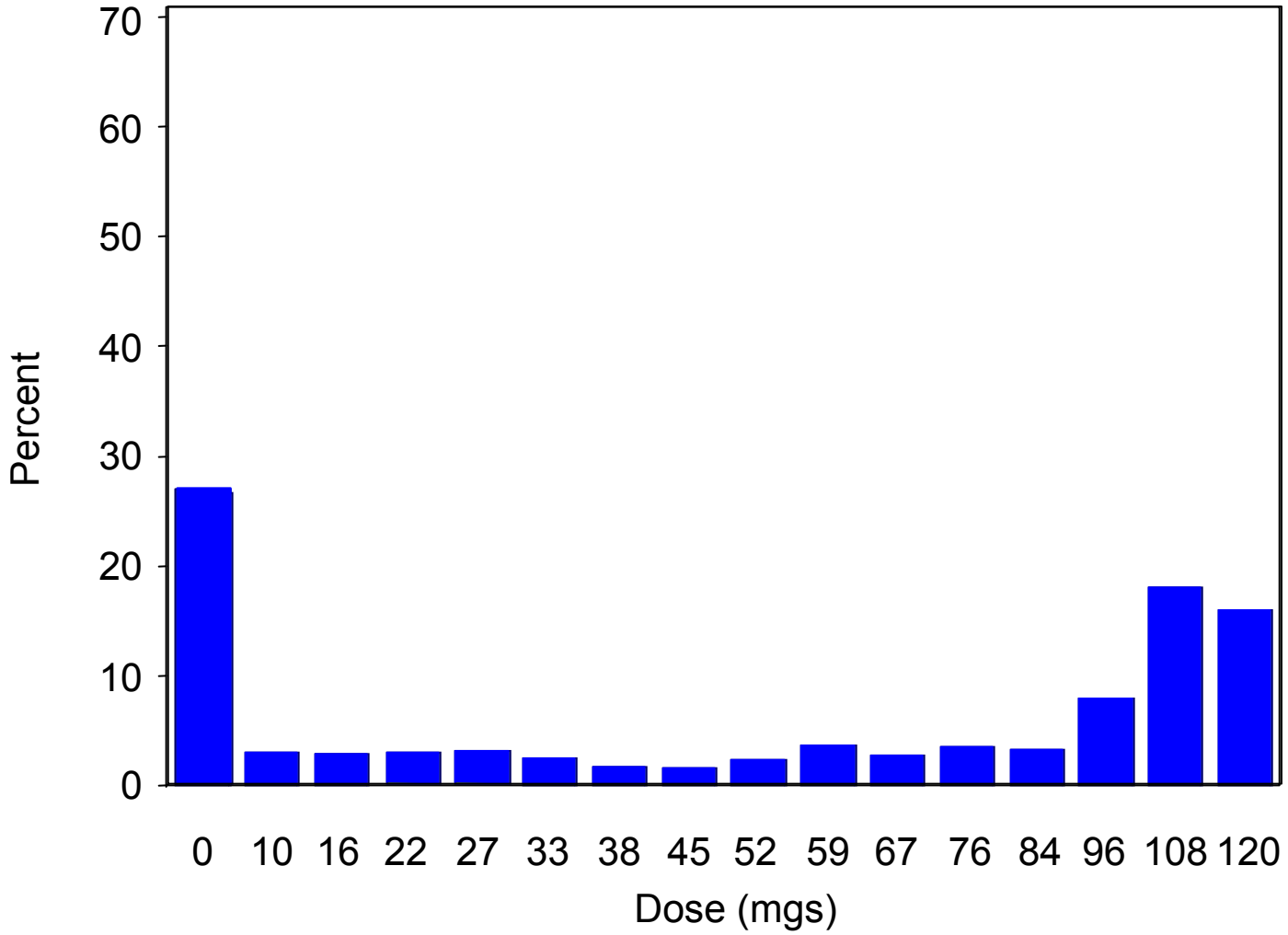
SSS Response Curve – Primary Analysis



Neutrializing antibodies to UK-279,276



Histogram of Allocated Doses



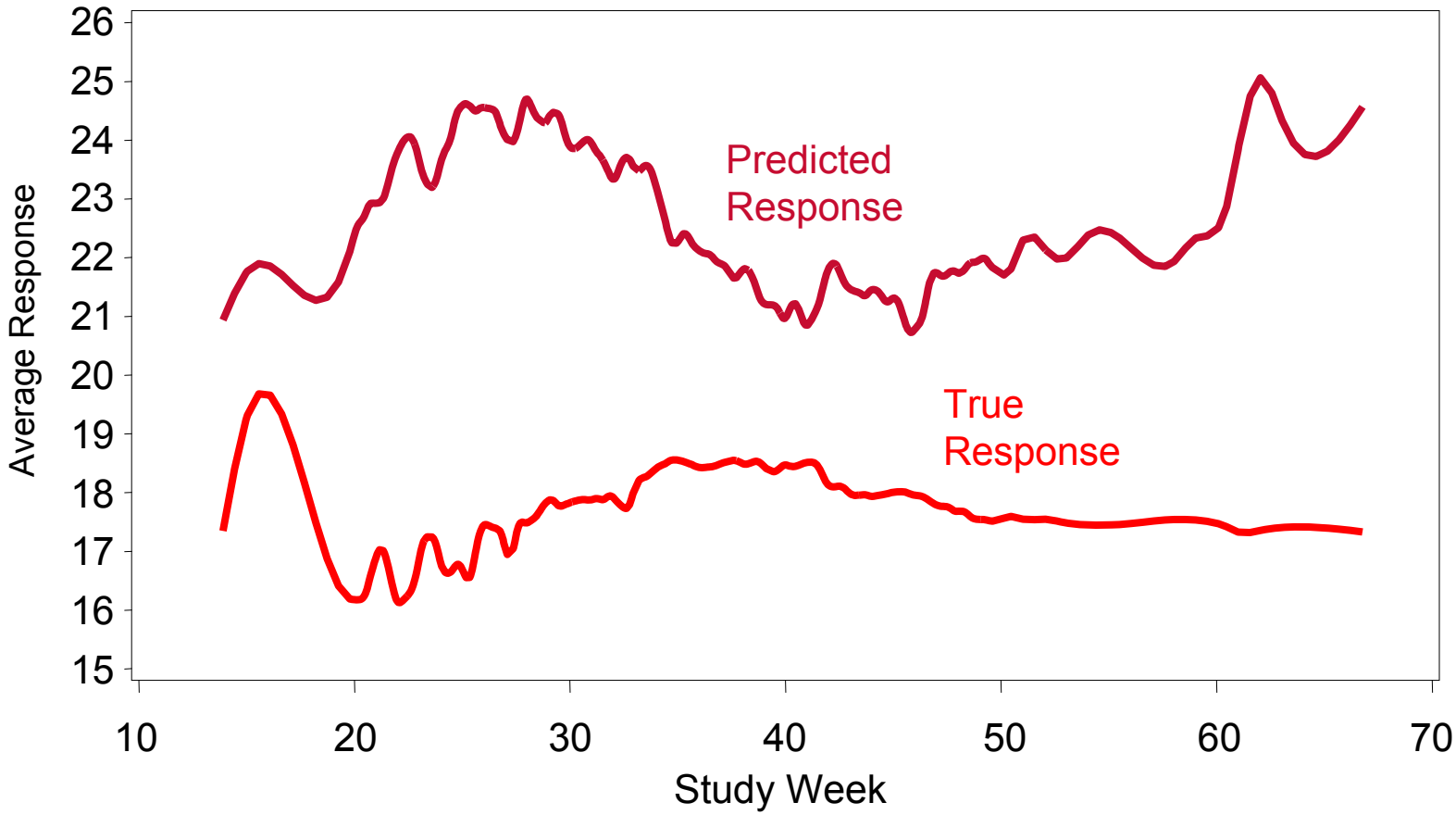
Operational Learnings

- Recruitment speed
- Exchangeability – 100 centres worldwide
- Covariates of interest

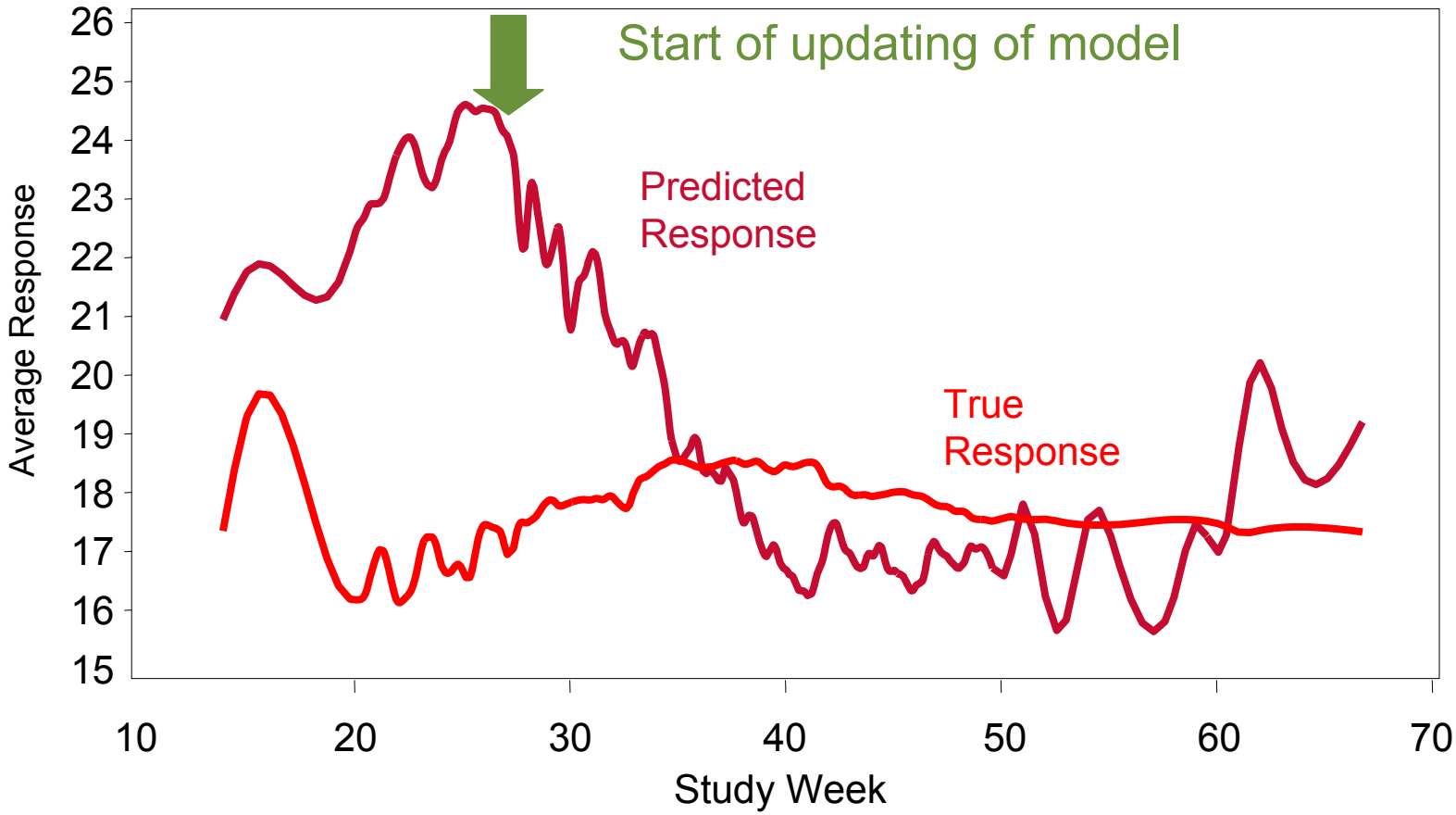
Statistical

- Longitudinal model
  - understand how the longitudinal model worked.
- Allocation Rule
  - look at an issue with choosing the ED95
- The stopping rule
  - Investigate stability of posterior probabilities

Comparison of Imputed & True Responses – Initial longitudinal model



Comparison of Predicted & True Responses – Updated longitudinal model



Longitudinal Model

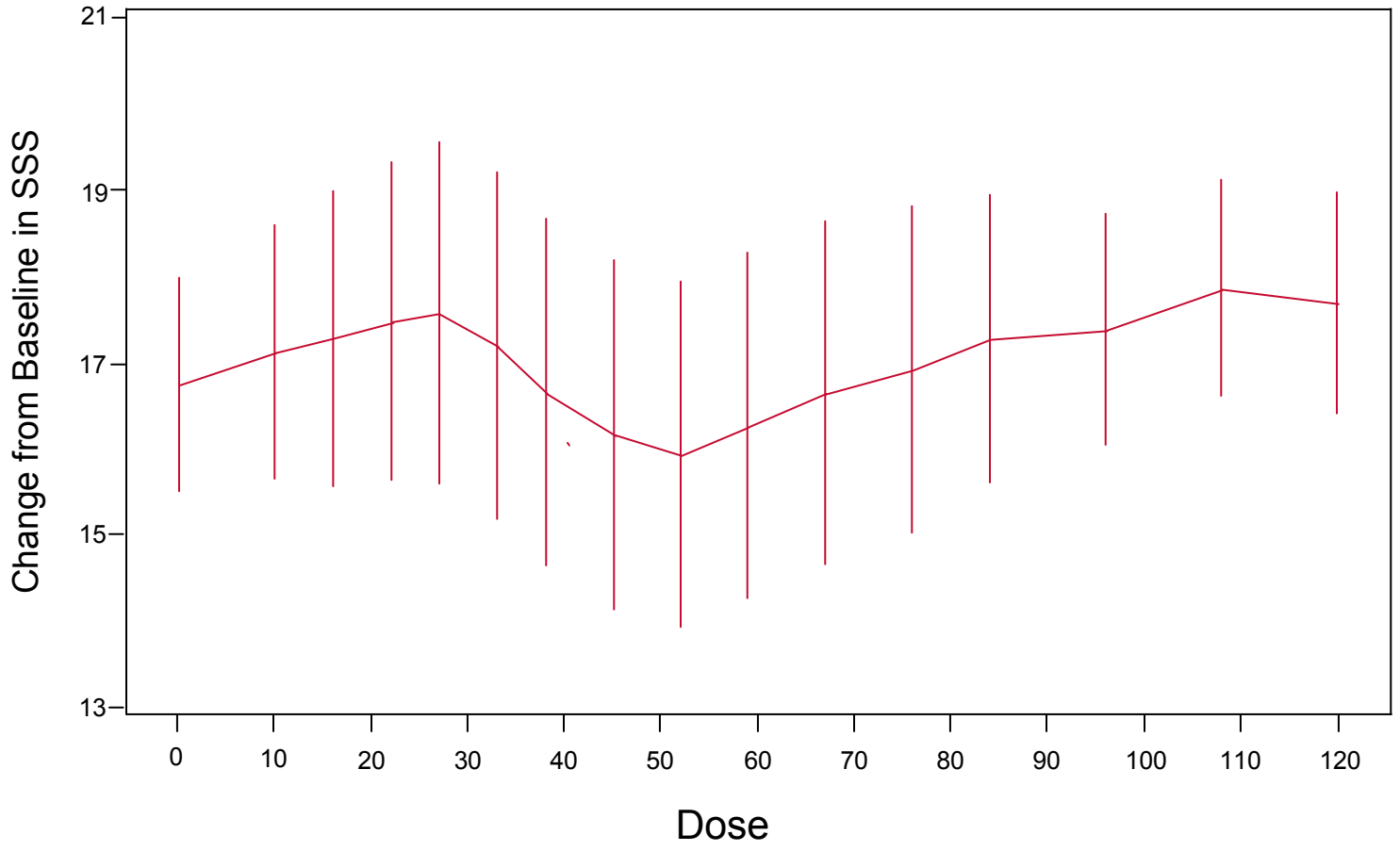
- Parametrisation

$y_i = m_i + \alpha_i y_{i+1}$

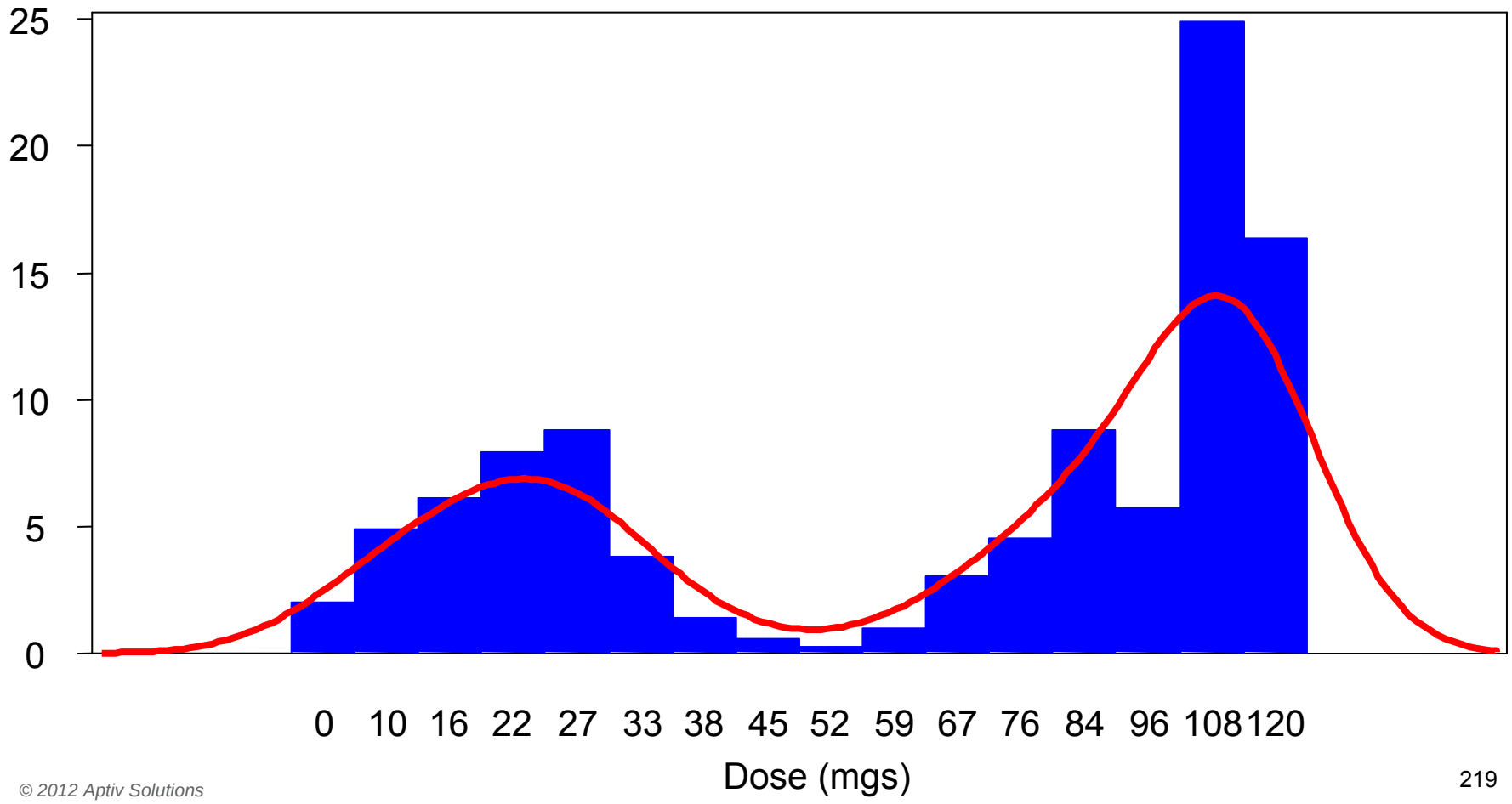
where y_i is the response at week I

1) All the m's and α's are linked – update one and the rest are updated
2) They are estimated in different categories of pts

SSS Response Curve – Primary Analysis



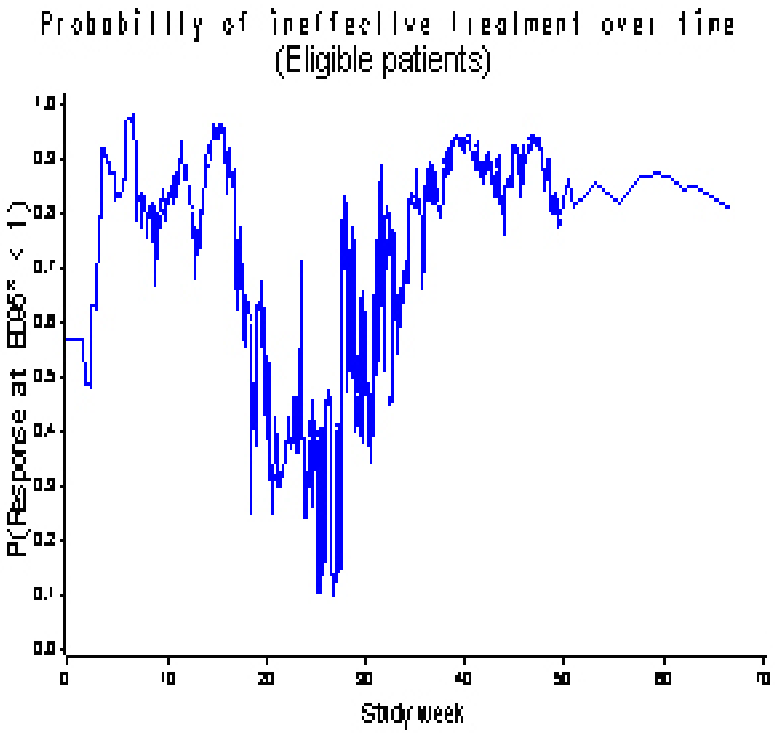
Posterior Distribution of ED95

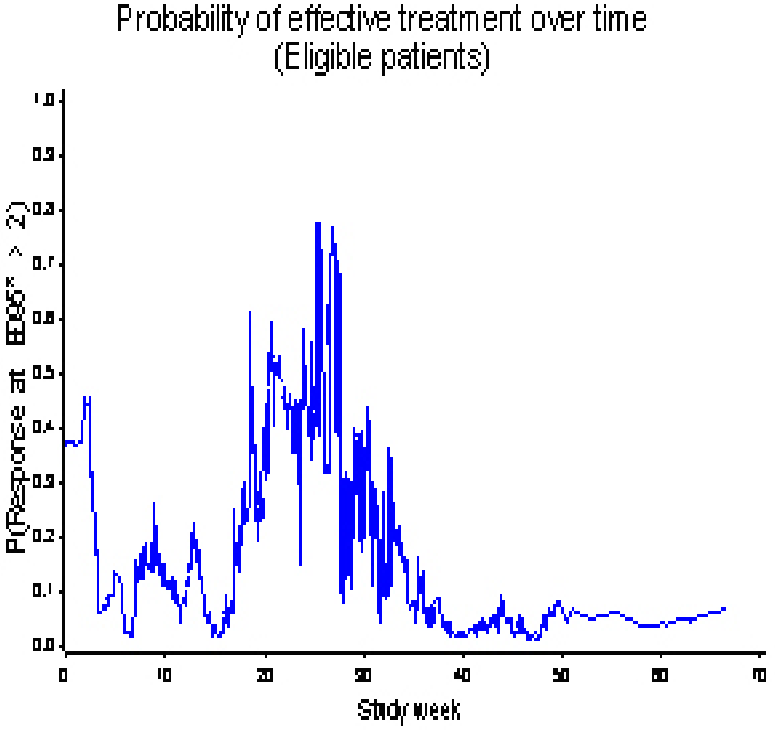


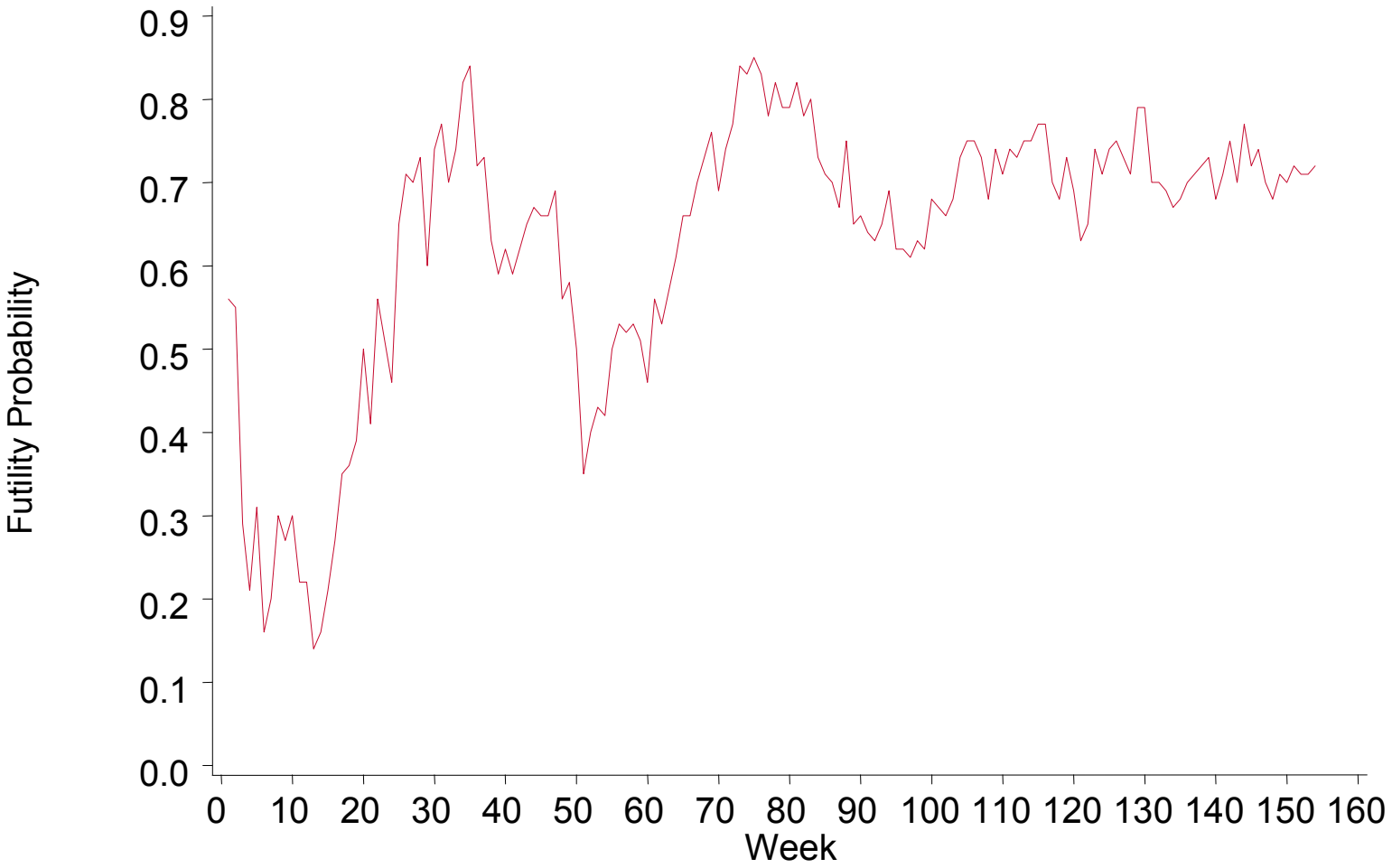
Other Possibilities

- Use posterior mode rather than posterior mean
- Use an estimate based on the posterior expected response curve
  - Investigations carried out for the IDMC show that this has similar properties to the posterior mean ED95

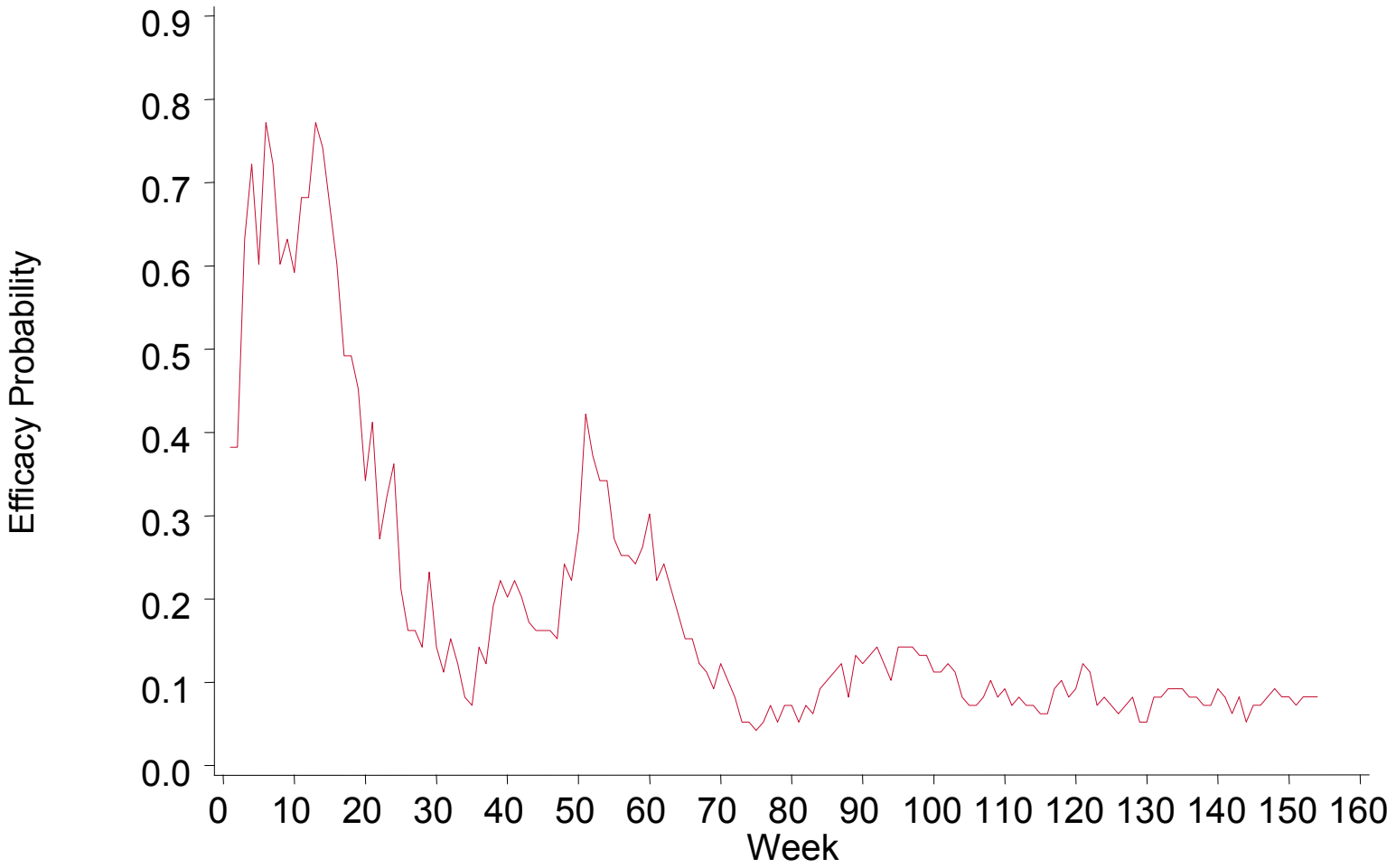
Probability of Futility and Efficacy






Futility Probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
Week

Plot of Futility Probability Over Time Simulation from a Flat Curve



Outline

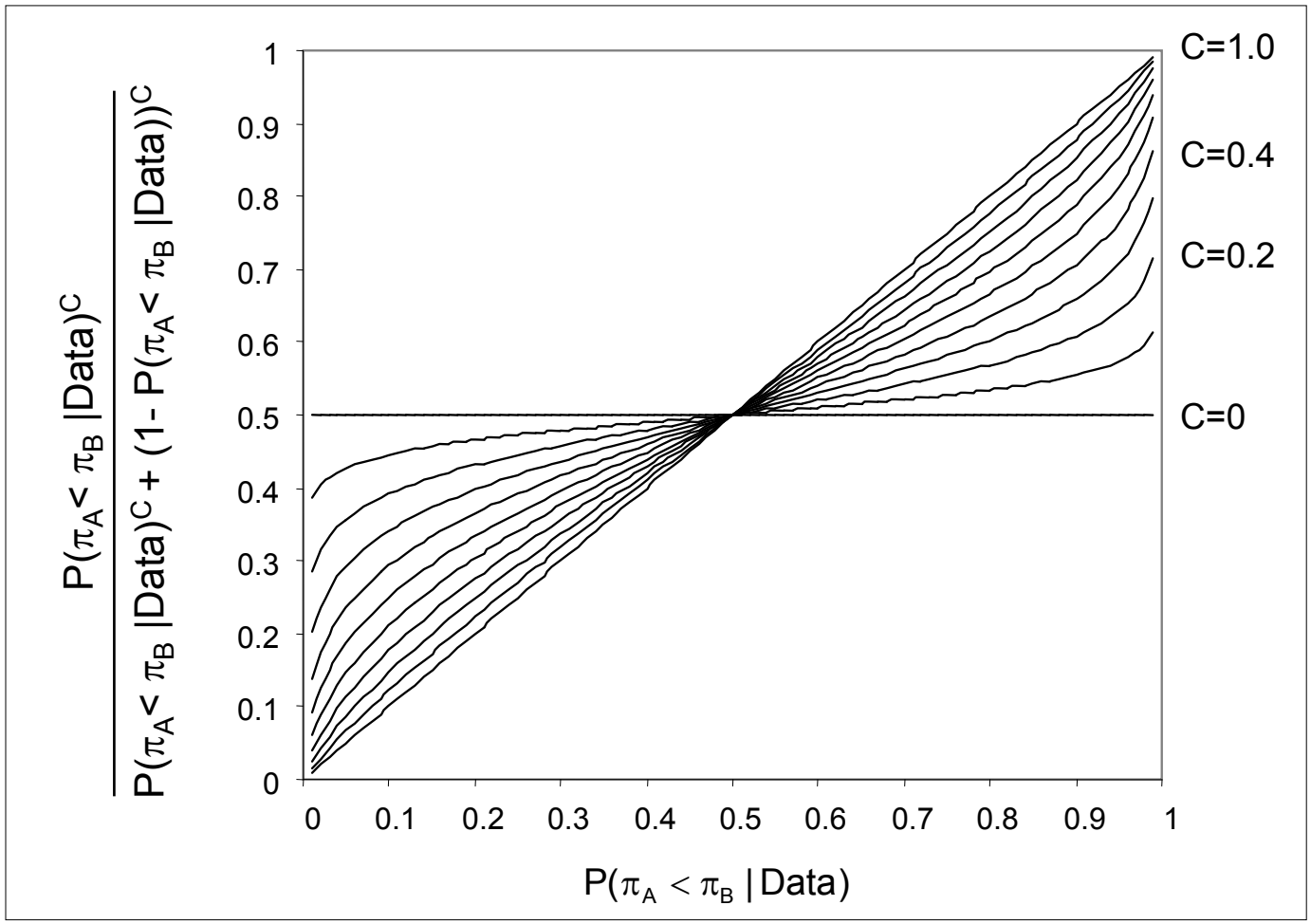
- Introduction and taxonomy of clinical trial designs
  - Pre-1990's
- Basic principles of adaptive designs
  - Allocation rule
  - Sampling rule
  - Stopping rule
  - Decision rule
- Phases of development
- Adaptive designs for the learn phase of drug development
  - First-in human / MTD
  - Two-stage designs
  - Adaptive dose-ranging designs
  - Bayesian adaptive randomisation
- Adaptive designs for the confirmatory phase of drug development
  - Sample size re-assessment
  - Adaptive group sequential designs
  - Seamless phase II/III designs
  - Population enrichment designs
- Practical aspects of adaptive design implementation
- Discussion

Bayesian Adaptive Randomisation
Thall and Wathen (Eur J Cancer, 2007)

- Back to the idea of Thompson (1933)
- Similar to RPW – binary outcome
- Randomisation to treatment A on the basis of a function of $P(\pi_A < \pi_B|Data)$ although in practice Thompson used $P(\pi_A < \pi_B|Data)$.
- Unstable
- Thall and Wathen (2007)

$$\frac{P(\pi_A < \pi_B \mid Data)^C}{P(\pi_A < \pi_B \mid Data)^C + \left(1 - P(\pi_A < \pi_B \mid Data)\right)^C}$$

Bayesian Adaptive Randomisation
Impact of Choice of C



Bayesian Adaptive Randomisation
Impact of Choice of C

- Thall and Whalen recommend C= n/(2N)
  - n=current sample size
  - N=study's maximum sample size
- Begins with C=0, ends with C=1/2
- C=1/2 "works well in many applications"

- Giles et al (J Clin Oncology, 2003)
  - Similar idea – but now with 3 arms (2 experimental, 1 control) using functions of $P(m_1<m_0|data)$, $P(m_2<m_0|data)$, and $P(m_1<m_2|data)$, - m_2, m_1, and m_0 are the median survival times

2 x 2 Contingency Table

- Data structure

	Response	No Response
Treatment A	$n_{11}\ (\pi_A)$	$n_{12}\ (1\text{-}\pi_B)$
Treatment B	$n_{21}\ (\pi_B)$	$n_{22}\ (1\text{-}\pi_B)$

Likelihood $$\pi_A^{n_{11}}\left(1-\pi_{1A}\right)^{n_{12}}\pi_B^{n_{21}}\left(1-\pi_B\right)^{n_{22}}$$

Prior $$\propto \pi_A^{\nu_{11}-1}\left(1-\pi_A\right)^{\nu_{12}-1}\pi_B^{\nu_{21}-1}\left(1-\pi_B\right)^{\nu_{22}-1}$$

Posterior $$\propto \pi_A^{n_{11}+\nu_{11}-1}\left(1-\pi_A\right)^{n_{12}+\nu_{12}-1}\pi_B^{n_{21}+\nu_{21}-1}\left(1-\pi_B\right)^{n_{22}+\nu_{22}-1}$$

2x2 Contingency Table - Posterior Inference

- The probability of interest is

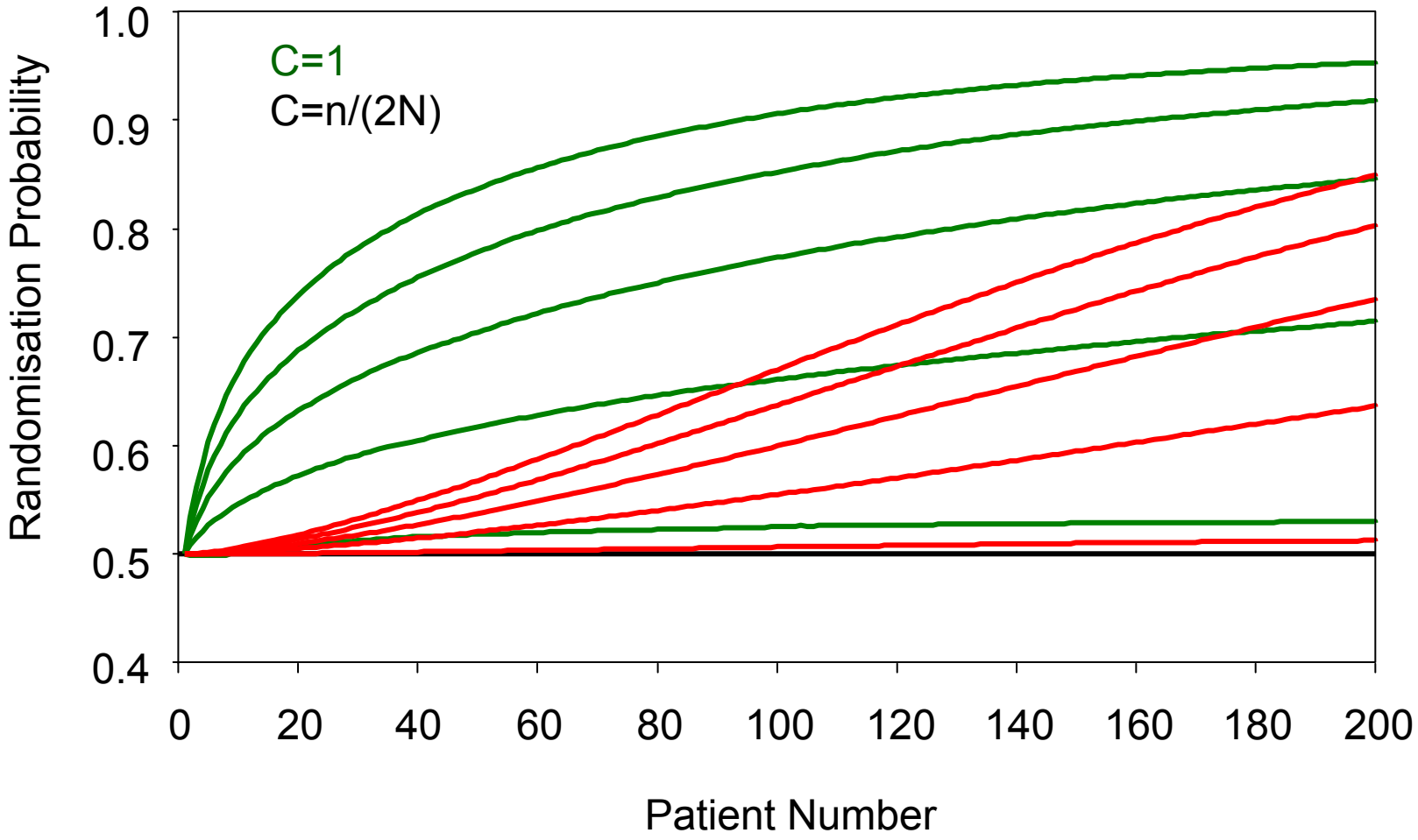
$$\mathrm{Prob}(\pi_A < \pi_B \mid \mathrm{Data}) = \sum_{k=\max(n_{21}+\nu_{21}-n_{12}-\nu_{12},0)}^{n_{21}+\nu_{21}-1} \frac{\binom{n_{11}+\nu_{11}+n_{21}+\nu_{21}-1}{k}\binom{n_{12}+\nu_{12}+n_{22}+\nu_{22}-1}{n_{21}+\nu_{21}+n_{22}+\nu_{22}-1-k}}{\binom{n_{11}+\nu_{11}+n_{21}+\nu_{21}+n_{12}+\nu_{12}+n_{22}+\nu_{22}-2}{n_{11}+\nu_{11}+n_{12}+\nu_{12}-1}}$$

based on the cumulative hypergeometric function as is Fisher's exact test:

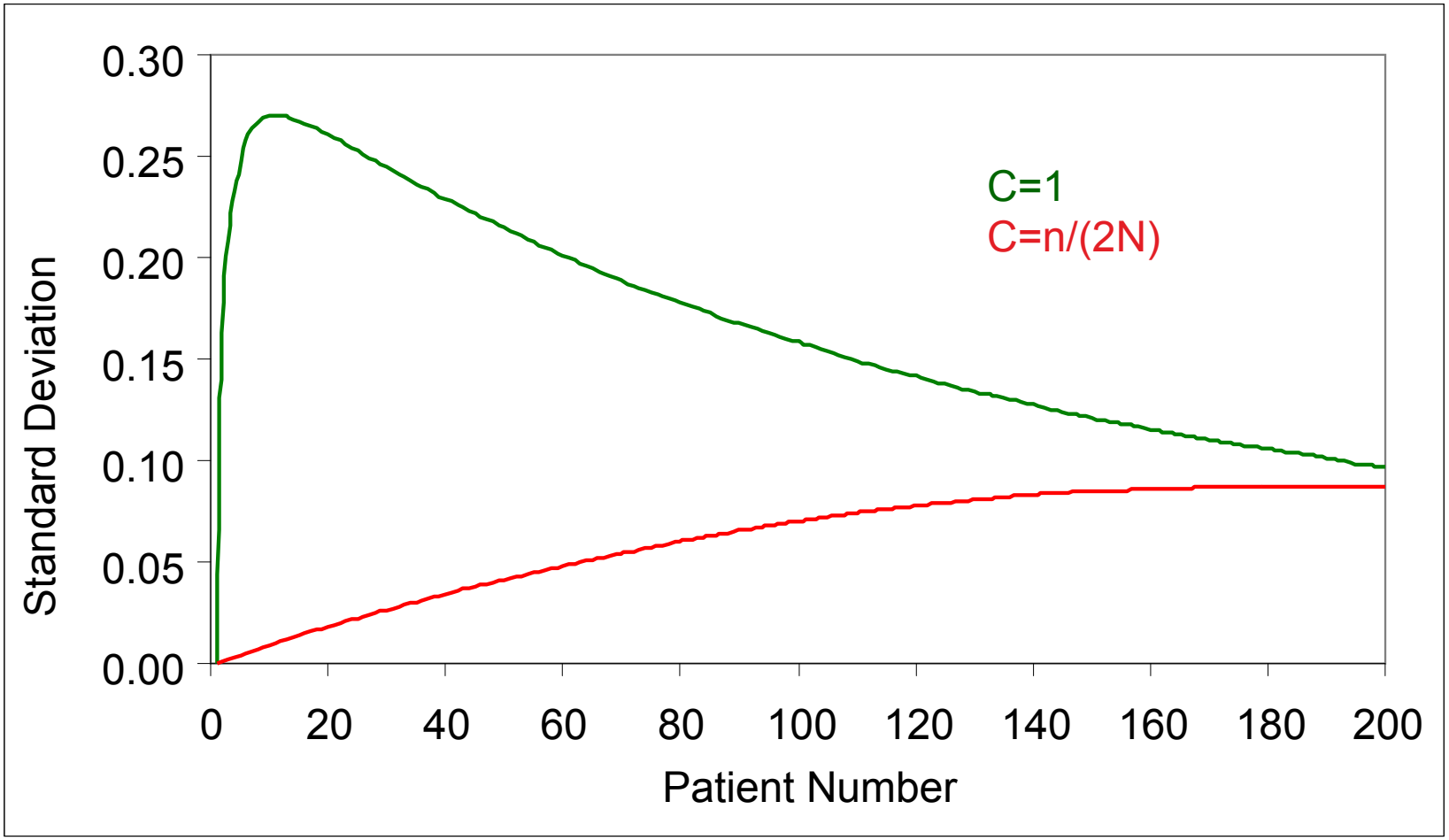
Bayesian AD – Thall & Wathen(EJC,2007) Type-I Error Based on T&W Criterion

- Thall & Wathen illustration is based on:
  - N = 200
  - Stopping Rules
    - If $P(\pi_A<\pi_B|Data) > 0.99$ stop and "choose" B
    - If $P(\pi_A<\pi_B|Data) < 0.01$ stop and "choose" A (futility)
- What does the type I error look like ?
- A complication is that the control rate, π_A, is a nuisance parameter

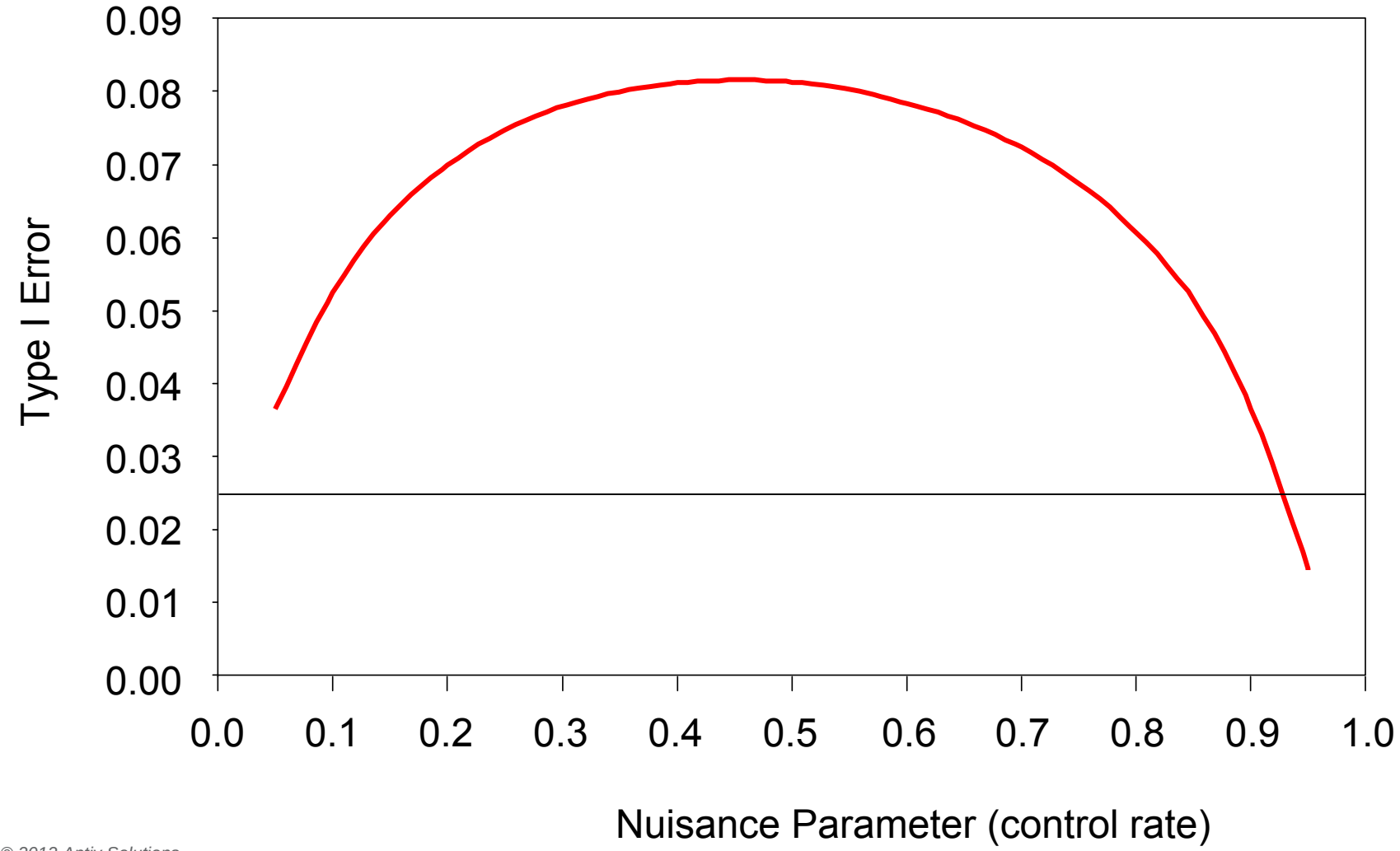
Bayesian AD – Thall & Wathen(EJC,2007) N=200
Randomisation Probabilities (10^5 simulations)
π_A=0.25 , π_B=0.25(0.05)0.45



Bayesian AD – Thall & Wathen(EJC,2007) N=200
Variability of Randomisation Probabilities
π_A=0.25 , π_B=4.25



Bayesian AD – Thall & Wathen(EJC,2007) N=200
Type-I Error Based on T&W Criterion - $P(\pi_A > \pi_B | Data) > 0.99$
10^6 Simulations / control rate



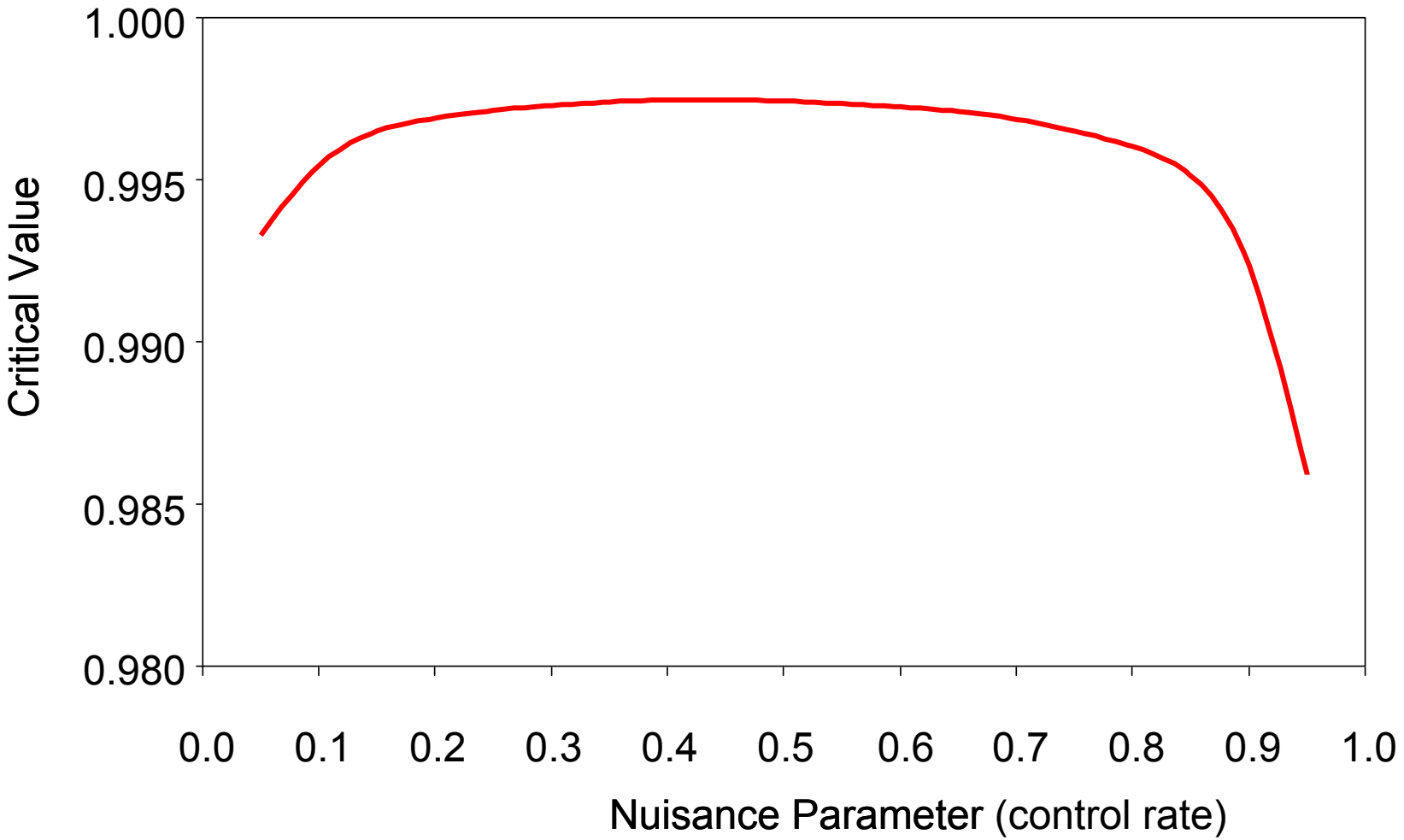
Bayesian AD – Thall & Wathen(EJC, 2007) N=200 Control One-Sided Type-I Error

- The issue is the number of tests being conducted

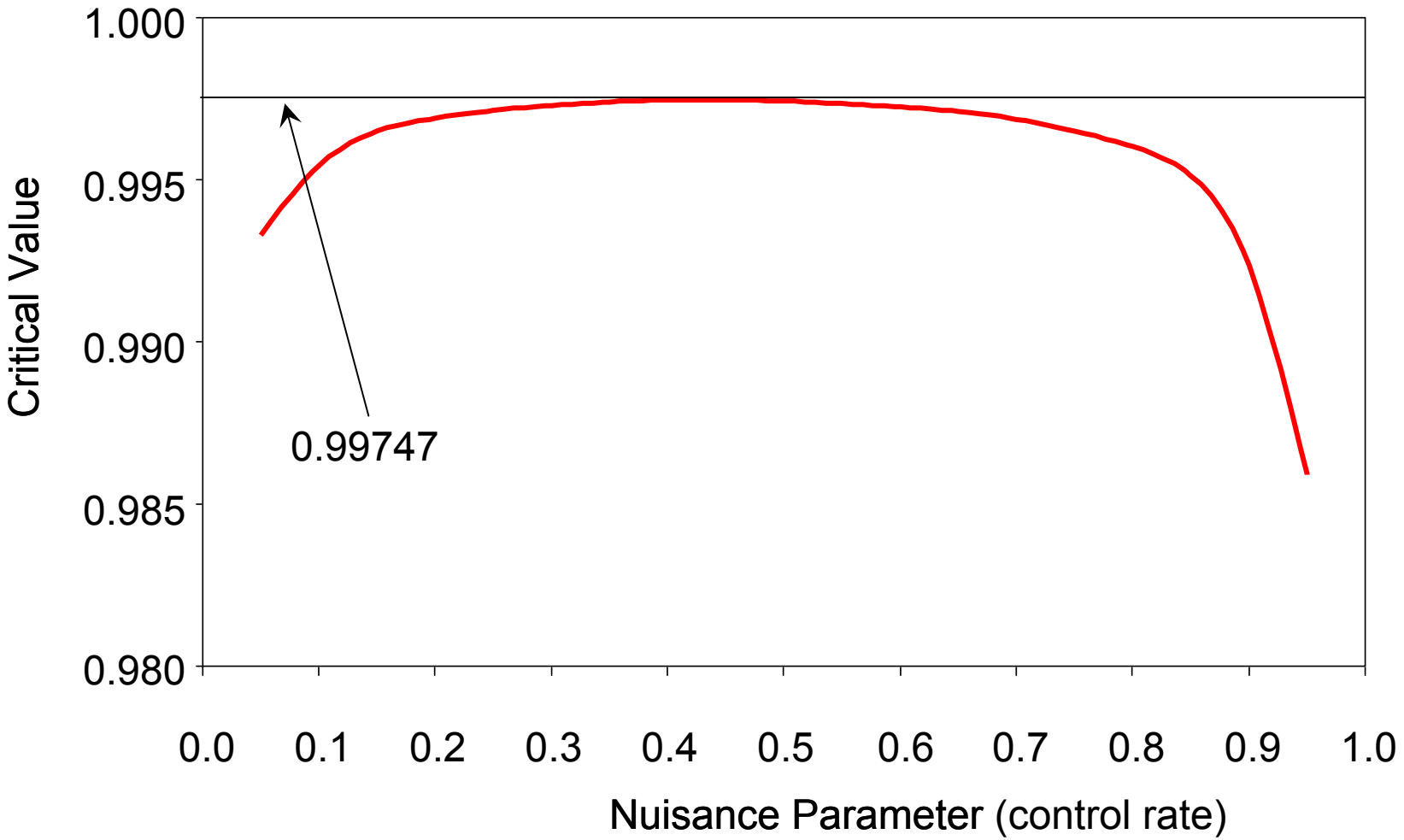
1. Reduce the problem using cohorts (20, 50 ?)

2. Or choose decision criterion $P(\pi_A<\pi_B|Data)>$? to control type-I error

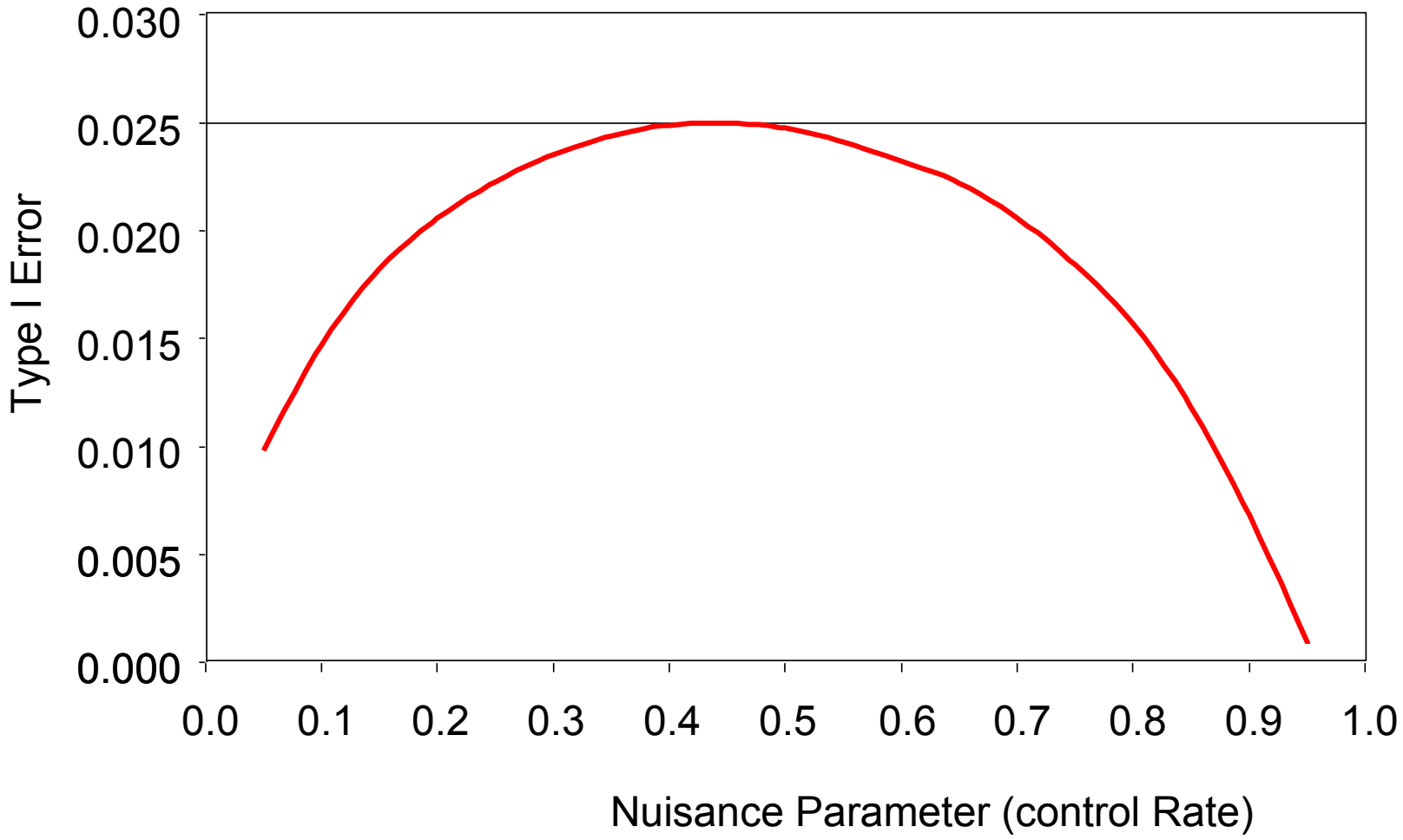
Bayesian AD – Thall & Wathen(EJC, 2007) N=200
Critical Value to Control One-Sided Type-I Error
10^6 Simulations / control rate



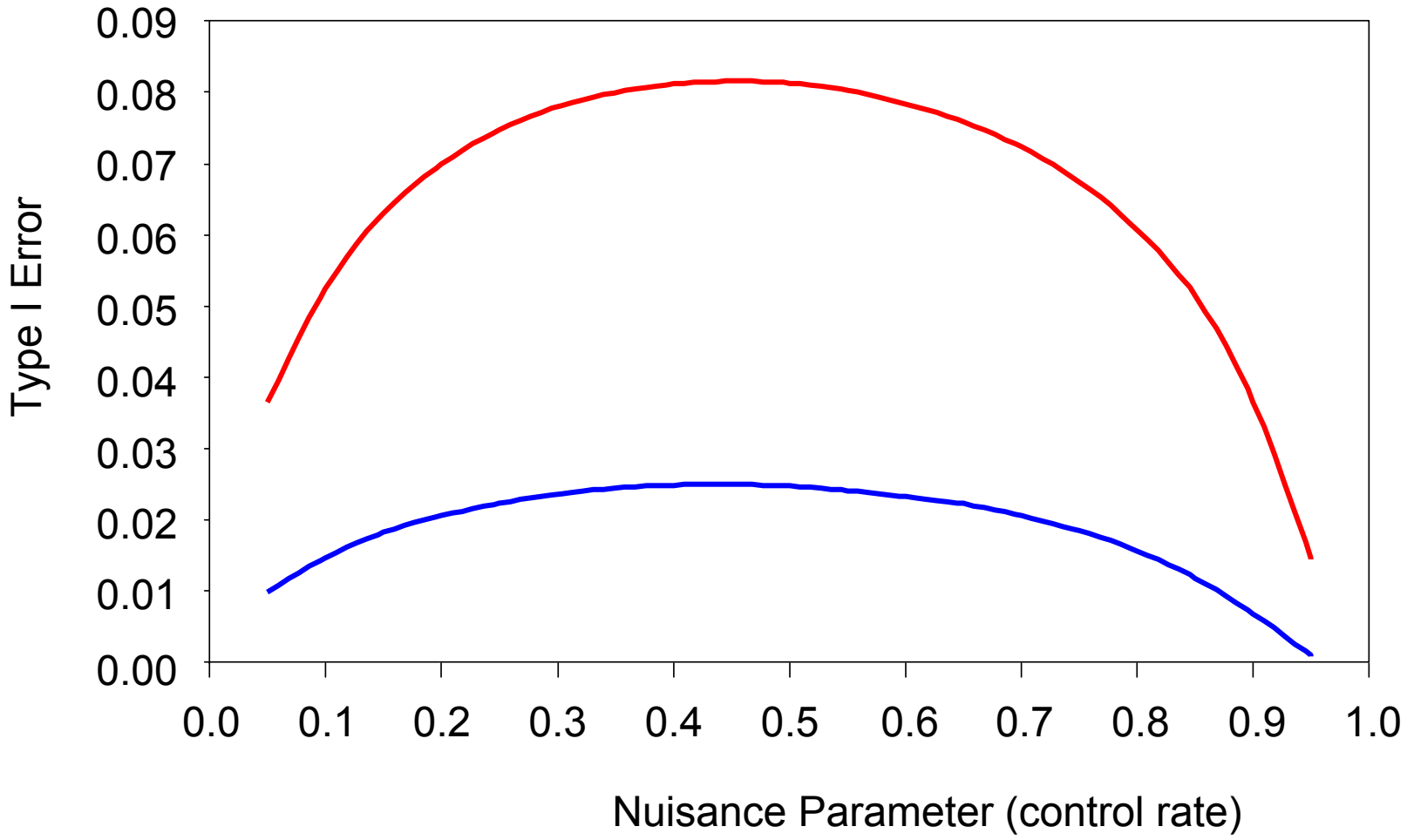
Bayesian AD – Thall & Wathen(EJC, 2007) N=200
Critical Value to Control One-Sided Type-I Error
10^6 Simulations / control rate



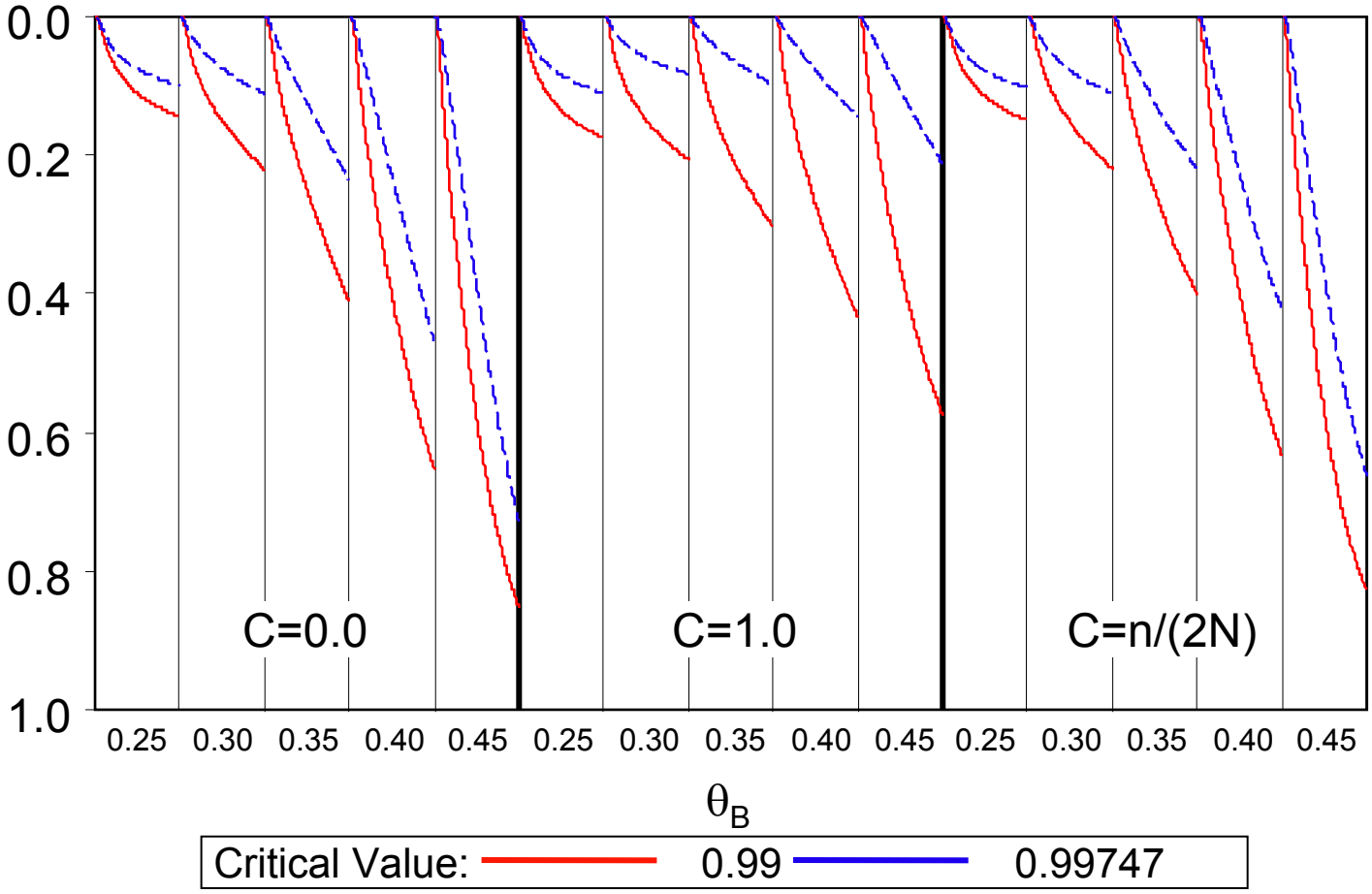
Bayesian AD – Thall & Wathen(EJC, 2007) N=200
Type-I Error Based on $P(\pi_A<\pi_B|Data)>.99747$
10^6 Simulations / control rate



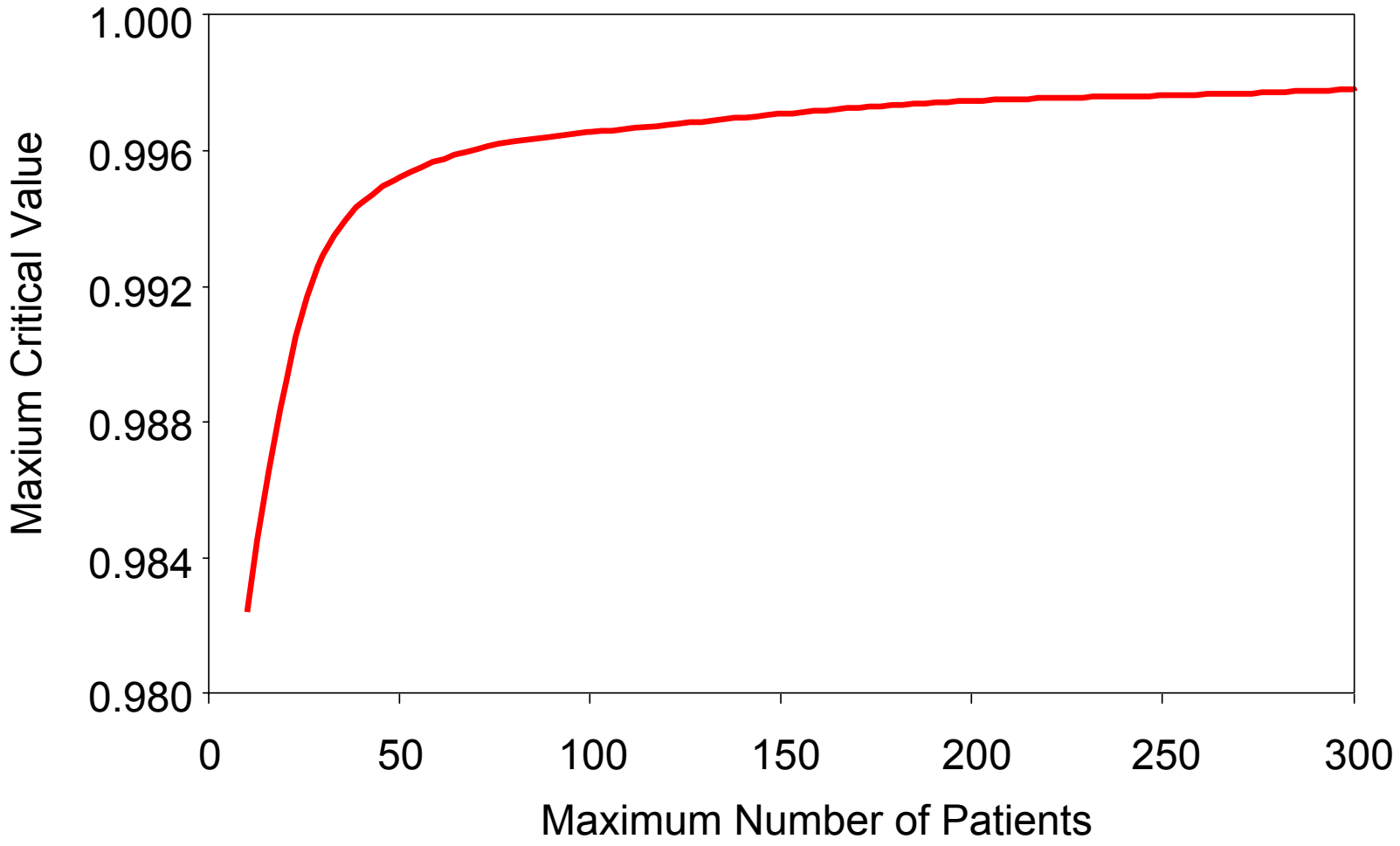
Bayesian AD – Thall & Wathen(EJC, 2007) N=200 Comparison of Type-I Error Based on T&W Criterion & Adjusted



Bayesian AD – Thall & Wathen(EJC, 2007) N=200
Probability of Choosing B as a Function of Patient Accrual
10^5 Simulations



Bayesian AD – Thall & Wathen(EJC, 2007) Maximum Critical Value vs Maximum Patient Number >10^6 Simulations / control rate



Crisiticism of This Approach

- Korn and Freidlin (J Clin Oncol, 2011)
- Their simulations “show”:
  - Thall & Wathen AD inferior to1:1 randomisation in terms of information, benefits to patients in trial

- True
- I agree with Don Berry (J Clin Ocol 2011) that the greatest benefits are likely to accrue for trials with more than 2 arms
- Rather as in the case of T=1 in the group sequential case greater complexity gives more scope for Bayesian designs

Adaptive Randomisation: Giles et al, JCO(2003)

- Troxacitabine (T) in acute myeloid leukemia (AML) combined with cytarabine (A) or idarubicin (I)
- Adaptive randomization to:
  - IA vs TA vs TI
- Max n = 75
- End point: Time to Complete Remission (< 50 days)

Adaptive Randomization

- Assign 1/3 to IA (standard, 0) throughout (unless only 2 arms)
- Adaptive to TA (1) and TI (2) based on current results
  - Time to success : Exponential
  - Prior(Median : m_i)=Gamma(2.001,4.624) (i=0,1,2)
  - Initial randomization : $p_0=p_1=p_2=1/3$
  - Define : $q_1=P(m_1<m_0|data)$, $q_2=P(m_2<m_0|data)$, $r=P(m_1<m_2|data)$
  - Then

$$\pi_1 = \frac{2q_1^2}{3(q_1^2+q_2^2)} \quad , \quad \pi_2 = \frac{2q_2^2}{3(q_1^2+q_2^2)}$$

Adaptive Randomization

- If at any time $q_1>0.85$ or $q_1>0.85$ – either TA or TI were outperforming IA – IA would be dropped
  - If both TA and TI were still in the study randomisation probabilities would be

$$\pi_1 = \frac{r^2}{(r^2+(1-r^2))} \quad , \quad \pi_2 = \frac{1-r^2}{(r^2+(1-r^2))}$$

- If at any time $q_1<0.15$ or $r<0.15$ – TA being outperformed by either TI or IA - TA would be dropped
- If at any time $q_2<0.15$ or $r>0.85$ – TI being outperformed by either TA or IA - TI would be dropped
  - If only IA and one investigational arm *k* remained, randomisation probabilities would be

$$\pi_k = \frac{q_k^2}{(q_k^2+1-q_k^2)} \quad , \quad \pi_0 = 1-\pi_k$$

- An arm that dropped out could be reopened if information (i.e., CR by day 49) became available from patients previously randomly assigned to that arm or if the other arms performed sufficiently poorly, subsequent to closure of the arm in question.
- The operating characteristic of the design was identified by simulation

Study Operating Characteristics

True Probabilities			Prob (choose arm 0 superior)	Prob (choose arm 1 superior)	Prob (choose arm 2 superior)	Mean Sample Sizes			
P_0	P_1	P_2				n_0	n_1	n_2	Sum
0.3	0.3	0.6	0.025 (0.005)	0.178 (0.145)	0.797 (0.740)	11	12	17	40
0.3	0.3	0.5	0.020 (0.007)	0.118 (0.097)	0.862 (0.843)	9	10	5	24
0.3	0.2	0.3	0.101 (0.029)	0.449 (0.321)	0.450 (0.333)	16	18	18	52
0.4	0.2	0.2	0.540 (0.299)	0.238 (0.102)	0.230 (0.102)	25	19	19	63
0.5	0.3	0.5	0.209 (0.157)	0.154 (0.114)	0.637 (0.564)	16	12	17	45
0.3	0.6	0.6	0.005 (0.004)	0.507 (0.501)	0.488 (0.478)	7	12	12	31

Study Results

	Probability Assign to:			Assigned	Outcome
Pat.	IA	TA	TI	Arm	CR<50
1	0.33	0.33	0.33	TI	NOT
2	0.33	0.34	0.32	IA	CR
3	0.33	0.35	0.32	TI	NOT
4	0.33	0.37	0.30	IA	NOT
5	0.33	0.38	0.28	IA	NOT
6	0.33	0.39	0.28	IA	CR
7	0.33	0.39	0.27	IA	NOT
8	0.33	0.44	0.23	TI	NOT
9	0.33	0.47	0.20	TI	NOT
10	0.33	0.43	0.24	TA	CR
11	0.33	0.50	0.17	TA	NOT
12	0.33	0.50	0.17	TA	NOT
13	0.33	0.47	0.20	TA	NOT
14	0.33	0.57	0.10	TI	NOT
15	0.33	0.57	0.10	TA	CR
16	0.33	0.56	0.11	IA	NOT
17	0.33	0.56	0.11	TA	CR

	Probability Assign to:			Assigned	Outcome
Pat.	IA	TA	TI	Arm	CR<50
18	0.33	0.33	0.33	TA	NOT
19	0.33	0.34	0.32	TA	NOT
20	0.33	0.35	0.32	IA	CR
21	0.33	0.37	0.30	IA	CR
22	0.33	0.38	0.28	IA	CR
23	0.33	0.39	0.28	IA	CR
24	0.33	0.39	0.27	IA	CR
25	0.87	0.13	0	IA	NOT
26	0.87	0.13	0	TA	NOT
27	0.96	0.04	0	TA	NOT
28	0.96	0.04	0	IA	CR
29	0.96	0.04	0	IA	NOT
30	0.96	0.04	0	IA	CR
31	0.96	0.04	0	IA	NOT
32	0.96	0.04	0	TA	NOT
33	0.96	0.04	0	IA	NOT
34	0.96	0.04	0	IA	CR

Summary of results

CR < 50 days:

- IA: 10/18 = 56%
- TA: 3/11 = 27%
- TI: 0/5 = 0%

Outline

- Introduction and taxonomy of clinical trial designs
  - Pre-1990's
- Basic principles of adaptive designs
  - Allocation rule
  - Sampling rule
  - Stopping rule
  - Decision rule
- Phases of development
- Adaptive designs for the learn phase of drug development
  - First-in human / MTD
  - Two-stage designs
  - Adaptive dose-ranging designs
  - Bayesian adaptive randomisation
- Adaptive designs for the confirmatory phase of drug development
  - Sample size re-assessment
  - Adaptive group sequential designs
  - Seamless phase II/III designs
  - Population enrichment designs
- Practical aspects of adaptive design implementation
- Discussion

Introduction

- Why consider sample size re-estimation?
  - Minimize number of patients exposed to inferior or highly toxic treatment
  - Right-size the trial to demonstrate efficacy
    - Reduce or increase sample size
  - Stop the trial for futility if insufficient benefit
  - Incorporate new internal or external information into a trial design during the course of the trial

The Problem

- In order to appropriately power a trial, one needs to know:
  - The true effect size you wish to detect
  - Nuisance parameters such as
    - Variability of a continuous endpoint
    - Population event rate for a binary outcome or time to event
  - Other ancillary information (e.g., correlation between co-primary endpoints needed to evaluate study-level power)
- Inappropriate assumptions about any of these factors can lead to an underpowered trial

Consequences of incorrect planning for treatment difference δ and/or standard deviation σ (α=0.05, planned power=90%)

	N planned/ N required	Power
Over-estimate δ or under-estimate σ by 50%	0.44	58%
Under-estimate δ or over-estimate σ by 50%	2.25	99.8%
Over-estimate δ AND under-estimate σ by 50%	0.20	30%
Under-estimate δ AND over-estimate σ by 50%	5.06	>99.9%
Under-estimate δ AND under-estimate σ by 50%	1	90%

Acknowledgement to Keaven Anderson

Solutions to the problem

- Plan a fixed trial conservatively
  - Pro: trial should be well-powered
  - Cons: Can lead to lengthy, over-powered, expensive trial
- Use group sequential design and plan conservatively
  - Pro: can power trial well and stop at appropriate, early interim analysis if your assumptions are too conservative
  - Con: over-enrollment occurs past definitive interim analysis because it takes time to collect, clean and analyze data
- Use adaptive design
  - Pro: can decide to alter trial size based on partial data or new, external information
  - Cons: methods used to adapt must be carefully chosen, regulatory scrutiny over methods and 'partial unblinding,' may not improve efficiency over group sequential design

SSR Strategies

- Update sample size to ensure power as desired based on interim results
  - Internal pilot studies: Adjust for nuisance parameter estimates only
    - Blinded estimation
    - Unblinded estimation
    - Testing strategy: no adjustment from usual test statistics
  - Adjusting for interim test statistic/treatment effect
    - All methods adjust based on unblinded treatment difference
    - Adjust sample size to retain power based on interim test statistic (conditional or predictive power)
      - Assume observed treatment effect at interim
      - Assume original treatment effect
    - Testing strategy:
      - adjust stage II critical value based on interim test statistic
      - p-value combination test

Unblinded SSR

- Advantage
  - Could provide more accurate sample-size estimate.
- Disadvantages
  - Re-estimate sample size in a continuous fashion can reveal interim difference.
  - There could be concerns over bias resulting from knowledge of interim observed treatment effect.
  - Typically require an external group to conduct SSR for registration trials.
    - Interim treatment differences can be misleading
      - Due to random variation or
      - If trial conditions change

P-value combination tests

- Methods for controlling Type I error
- The invariance principle – calculate separate standardized test statistics from different stages and combine them in a predefined way to make decisions.
  - Weighting of a stage does not increase if sample size for that stage is increased, meaning that individual observations for that stage are down-weighted in the final test statistic
    - Efficiency issue (Tsiatis and Mehta, 2003)
  - Many methods available, including
    - Fisher's combination test (Bauer, 1989)
    - Conditional error functions (Proschan and Hunsberger, 1995; Liu and Chi, 2001)
    - Inverse normal method (Lehmacher and Wassmer, 1999)
    - Variance spending (Fisher, 1998)

Combination tests

- Apply combination test method to determine the critical value for the second stage based on the observed data from the first stage.
- Make assumption on treatment effect; options include:
  - Observed effect (highly variable)
  - Original treatment effect used for sample size planning
  - Predictive power: integrate the CP over the likelihood
- Compute next stage sample size based on critical value, set conditional power to originally desired power given interim test statistic and assumed second stage treatment effect
  - Generally, will only raise sample size – not lower

Statistical Penalty for Sample Size Re-estimation?

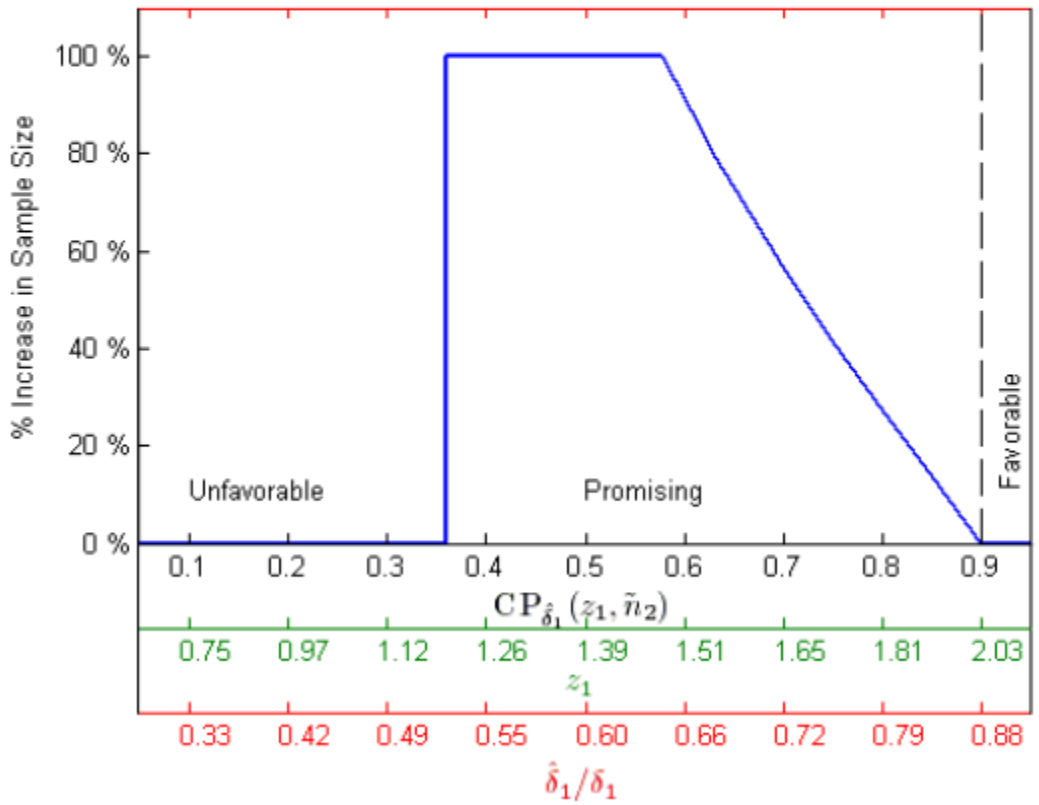
- Only increase sample size when conditional power at interim analysis already exceeds around 30% to 40%

and/or

- One stops for futility at interim analysis if conditional power is less than 10% - Gao et al [J Biopharm Stats 08; 18, 1184]

Mehta and Pocock (SIM, 2008) Promising Zone

Figure 1. Partitioning of Interim Result into Zones[†] and % Sample Size Increase in Each Zone: An Illustrative Example where $n_{max}/n_2 = 2$, $n_1/n_2 = 0.5$, one-sided $\alpha = 0.025$, and $1 - \beta = 0.9$.



Statistical Penalty for Sample Size Re-estimation?

- To preserve the Type I error
- 1) Down-weight the later data? NO
- [Cui et al Biometrics ‟99]
- *
- illogical, need to weight equally
- link to estimation
- 2) Adjust final α ?

Bioequivalence Bayesian Approach
Flühler, Grieve, Mandallaz, Mau and Moser (1981)

- log-normal transformation
- Uniform priors
- Two-period analysis
- Sufficient statistics: $\bar{x}_N - \bar{x}_S, s^2$

$$P(\theta_L < \mu_N / \mu_S < \theta_U \mid Y) = \int_B^A p_\nu(\tau) d\tau \quad \text{where}$$

$$A = \frac{n^{1/2}\left(\bar{x}_N - \bar{x}_S - \log(\theta_L)\right)}{2^{1/2} s} \quad , \quad B = \frac{n^{1/2}\left(\bar{x}_N - \bar{x}_S - \log(\theta_U)\right)}{2^{1/2} s}$$

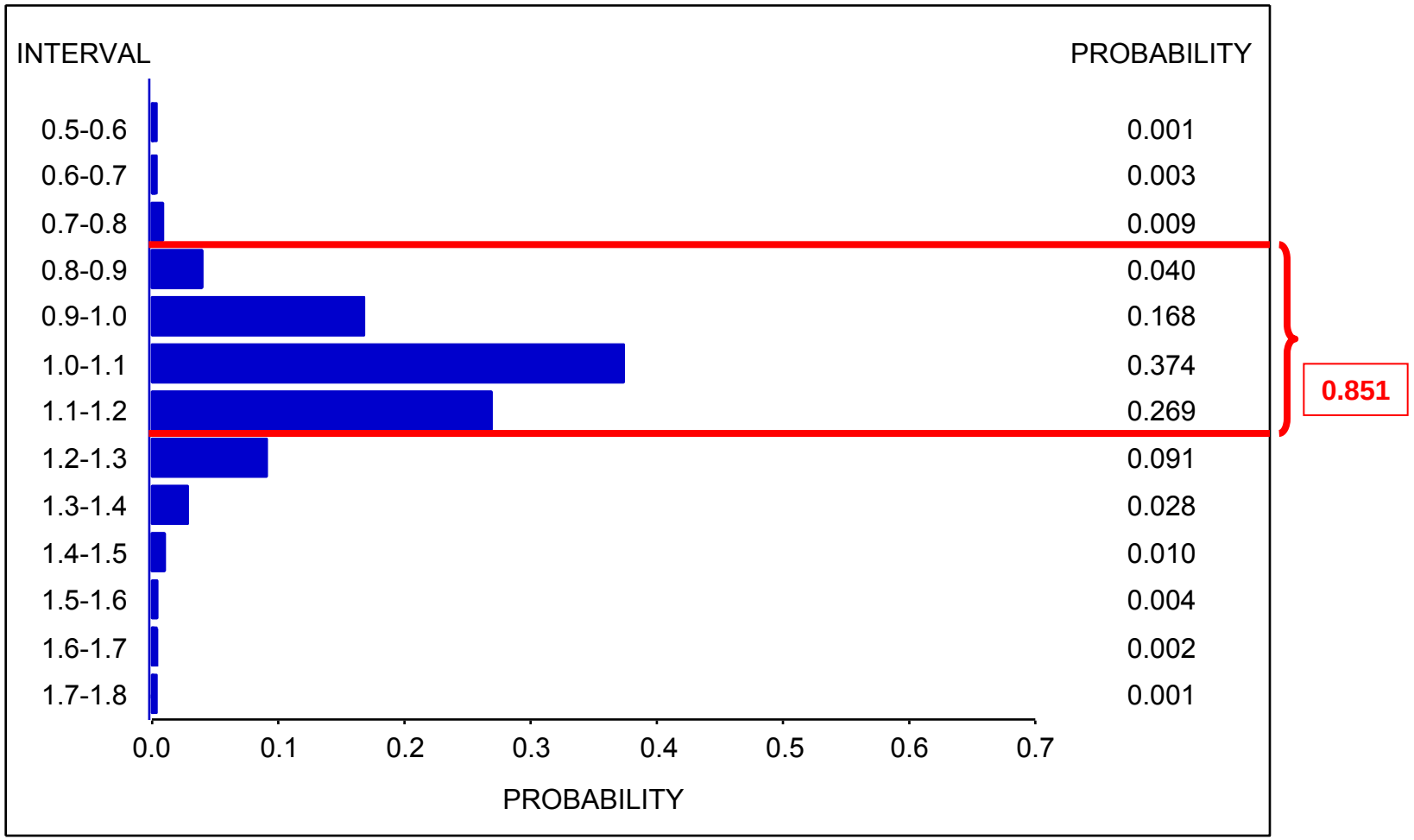
Two-Stage Study of Oxaprotiline HCL

	Film-Coated(N)		Aqueous Soln (S)	
Subject	Period	Ln(AUC)	Period	Ln(AUC)
1	1	0.1415	2	0.1231
2	2	0.7659	1	0.8862
3	2	0.9310	1	0.8442
4	1	0.4637	2	0.4713
5	1	1.1019	2	1.1049
6	2	1.0980	1	0.6318

Posterior Distribution for AUC of Oxaprotiline HCL

New = Film-Coated Standard = Aqueous Solution

Observed Theta = 1.0762 Root Mean Square Error = 14.89%



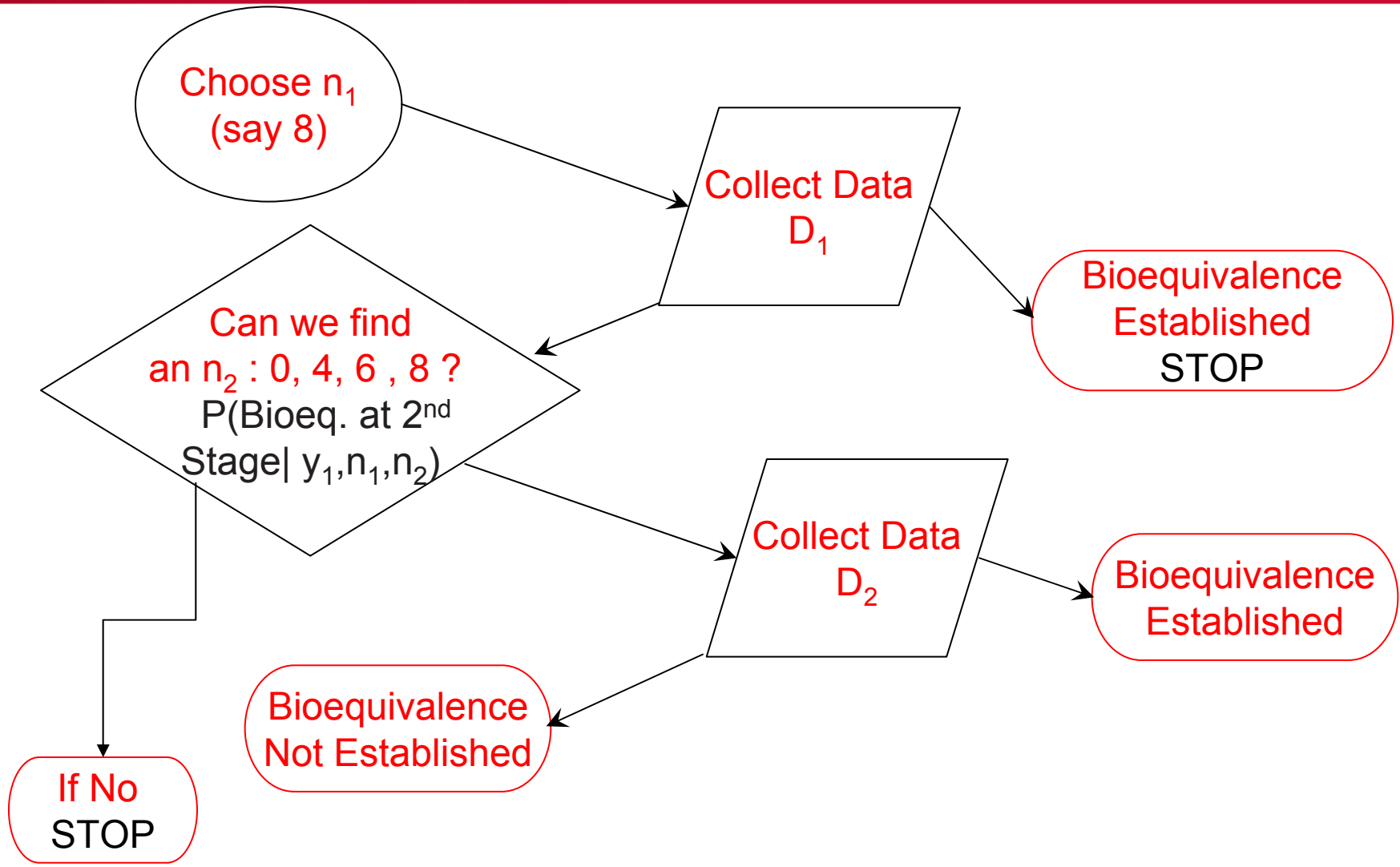
Standard Ciba-Geigy Presentation (ca 1984)

Bioequivalence Design Simulation Results CV=20% , n=12

θ	0.90	0.95	1.00	1.05	1.10
	27.5	38.8	40.0	31.4	19.5

- Single-stage studies wasteful of resources
- Two-stage study
  - abandon “hopeless cases”
  - continue “hopeful cases”

Schematic Representation of a Two-Stage Design (Racine-Poon et al , Biometrics 1987)



Example of a Two-Stage Scheme

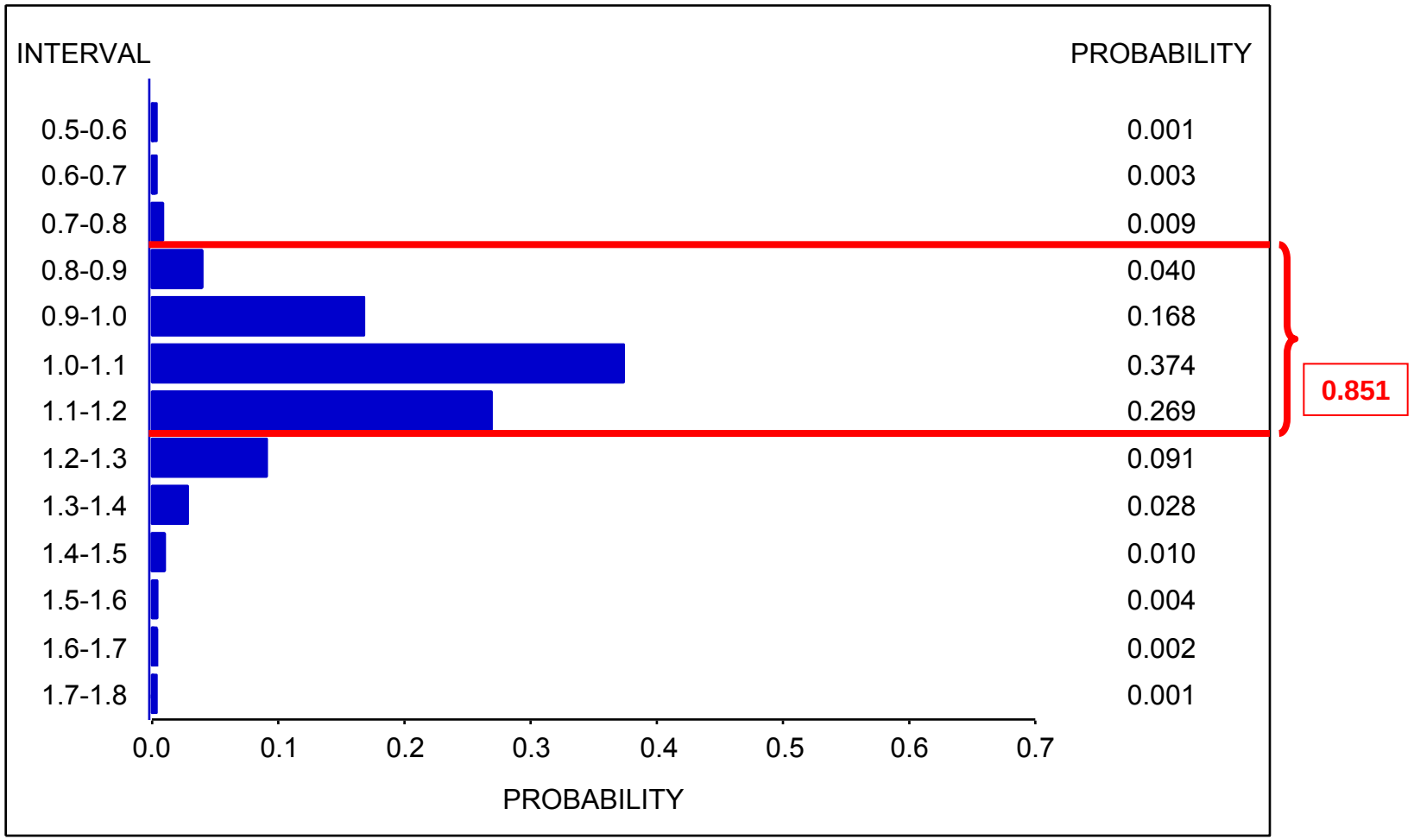
Stage 1	Choose: n_1	=8	
Stage 2	Choose: n_2	=0	if $P(\log(0.8)<\phi<\log(1.2)\|n_1, D_1, s_1) > 0.95$
		=0	if $\|D_1 - 1\| > 0.15$ or $s_1 > 0.25$
		=4	if $s_1 < 0.15$
		=6	if $0.15 < s_1 < 0.20$
		=8	if $0.20 < s_1 < 0.25$

Bioequivalence Designs
Simulation Results - Two Stage Design
% Acceptance , CV= 20%

θ	0.90	0.95	1.00	1.05	1.10
n=12	27.5	38.8	40.0	31.4	19.5
n_1=8	38.8	52.2	58.5	45.8	27.0
Ave # Subs	11.8	11.7	11.9	11.8	11.4

- Improved acceptance rates
- No increase in average numbers of subjects
- Choice of scheme ?

Observed Theta = 1.0762 Root Mean Square Error = 14.89%



Two-Stage Study of Oxaprotiline HCL

Stage 1

	Film-Coated(N)		Aqueous Soln (S)	
Subject	Period	Ln(AUC)	Period	Ln(AUC)
1	1	0.1415	2	0.1231
2	2	0.7659	1	0.8862
3	2	0.9310	1	0.8442
4	1	0.4637	2	0.4713
5	1	1.1019	2	1.1049
6	2	1.0980	1	0.6318

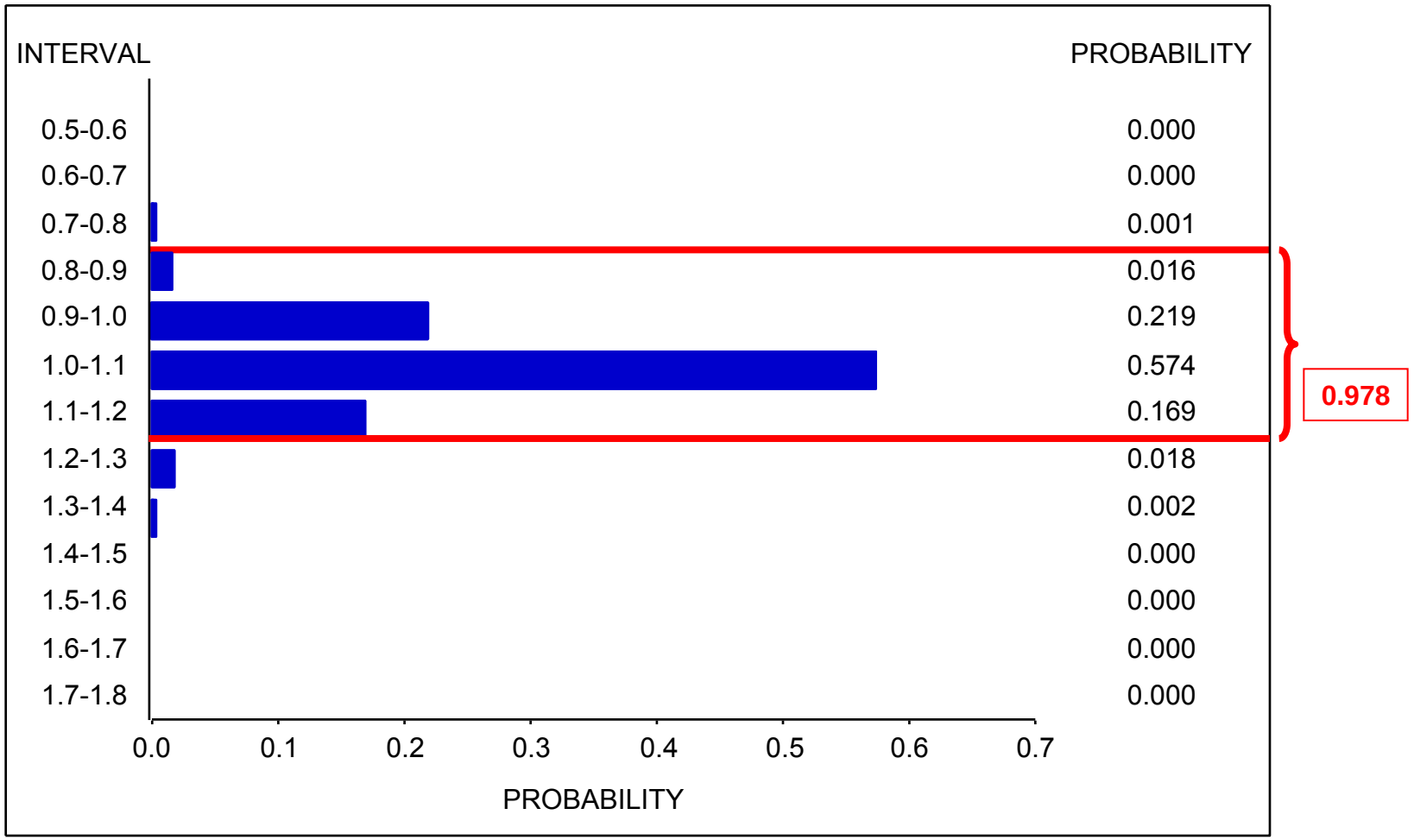
Stage 2

	Film-Coated(N)		Aqueous Soln (S)	
Subject	Period	Ln(AUC)	Period	Ln(AUC)
1	1	0.1415	2	0.1231
2	2	0.7659	1	0.8862
3	2	0.9310	1	0.8442
4	1	0.4637	2	0.4713

Posterior Distribution for AUC of Oxaprotiline HCL (After 2nd Stage)

New = Film-Coated Standard = Aqueous Solution

Observed Theta = 1.0436 Root Mean Square Error = 12.53%



Mehta and Pocock's Promising Zone Approach

- MP's Example 1 concerns a Phase 3 trial of a new treatment for schizophrenia in which a new drug is to be compared to an active comparator.
- The efficacy endpoint is improvement in the Negative Symptoms Assessment score from baseline to week 26.
- Denote responses by
  - Y_{Bi}, i = 1, 2, . . . , on the new treatment,
  - Y_{Ai}, i = 1, 2, . . . , on the comparator treatment.
- Responses are assumed to be normally distributed with variance $\sigma^2=7.5^2$, so each

$$Y_{Ai} \sim N(\mu_A, \sigma^2) \text{ and } Y_{Bi} \sim N(\mu_B, \sigma^2)$$

- The treatment effect is $\theta = \mu_B - \mu_A$.

MP Example (Cont)

- An initial plan is for a total of n_2= 442 patients, 221 on each treatment.
- The final analysis will reject H0: $\theta \leq 0$ if Z2 > 1.96, where

-) are treatment means from a total of n_2 observations.
- This gives a test with one-sided type I error rate 0.025 and power 0.8 at $\theta = 2$. Higher power, e.g., power 0.8 at $\theta = 1.6$, would be desirable. BUT the sponsors will only increase sample size if interim results are "promising". An interim analysis is planned after observing n_1= 208 responses. Due to uniform staggered accrual and the 26 week delay in obtaining a response, another 208 subjects will be treated by this time and await 26 weeks follow up. Recruitment continues. The final data set will contain at least the original 442 subjects: with "promising" data, an increase up to 884 subjects is permitted

Outline

- Introduction and Taxonomy of Clinical Trial Designs
- Basic Principles of Adaptive designs
  - Allocation Rule
  - Sampling Rule
  - Stopping Rule
  - Decision Rule
- Phases of Development
- Adaptive Designs for the Learn Phase of Drug development
  - First-in Human / MTD
  - Two-Stage Designs
  - Adaptive Dose-Ranging designs
- Adaptive Designs for the Confirmatory Phase of Drug Development
  - Sample Size Re-Assessment
  - Adaptive Group Sequential designs
  - Seamless Phase II/III Designs
  - Population Enrichment Designs
- Practical Aspects of Adaptive Design Implementation
- Discussion

Adaptive Confirmatory Designs

Group Sequential Designs

Pocock (1977), O'Brien & Fleming (1979)

The prototype case

- Parallel group design of normally distributed observations with known variance
- Two-sided test

$$H_0: \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1: \mu_1 - \mu_2 \neq 0$$

$k = 1,...,K$ Stages

- Compute standardized test statistic Z_k^* at each stage k of the trial

Group Sequential Designs

- are characterized by specifying decision regions (stopping and rejection boundaries) for the standardized test statistic

$$Z_k^* = \frac{\sqrt{n_1}Z_1 + \sqrt{n_2}Z_2 + \ldots + \sqrt{n_k}Z_k}{\sqrt{n_1 + n_2 + \ldots + n_k}} =$$

$$\frac{\sqrt{t_1}Z_1 + \sqrt{t_2 - t_1}Z_2 + \ldots + \sqrt{t_k - t_{k-1}}Z_k}{\sqrt{t_k}}, \text{ where}$$

$$Z_k = \frac{(\bar{X}_{1k} - \bar{X}_{2k})}{\sigma}\sqrt{\frac{n_k}{2}} \text{ and } t_k = \sum_{\tilde{k}=1}^{k} n_{\tilde{k}} / N, k = 1,\ldots,K.$$

t_k: „Information rates“

Repeated Significance Test

- Compute critical values through

$$P_{H_0}\{|Z_1^*| \geq u_1 \text{ or } |Z_2^*| \geq u_2 \text{ or } \ldots \text{ or } |Z_K^*| \geq u_K) = \alpha$$

or

$$P_{H_0}\{|Z_1^*| < u_1, |Z_2^*| < u_2, \ldots, |Z_K^*| < u_K\} = 1 - \alpha$$

- Recursion formula due to Armitage et al. (1969) enables calculation for arbitrary K.

Example

K = 5, critical values u_k and adjusted levels α_k
α = 0.05, two-sided, equally sized stages

	O'Brien & Fleming		Pocock	
k	u_k	α_k	u_k	α_k
1	4.562	0.000005	2.413	0.0158
2	3.226	0.0013	2.413	0.0158
3	2.634	0.0084	2.413	0.0158
4	2.281	0.0226	2.413	0.0158
5	2.040	0.0413	2.413	0.0158

Applications

The theory can also be applied for many other, practically relevant cases:

- testing means in the t test situation:
  apply „significance level approach“, i.e. reject H_0 if $p_k < \alpha_k$
- testing the difference and ratio of rates
- testing the coefficient in a regression model
- testing the (log) odds ratio
- testing the (log) hazard ratio in a survival design
- …

It is necessary that the process of data accumulation can be represented, al least asymptotically, as a process of normally distributed and independent increments.

Survival Data

- During the stages of the trial, a sequence of accumulated events $d_1,\ldots,d_K$ is observed.
- At each stage k of the test procedure the logrank test statistic

$$LR_k^* = \frac{\sum_{i=1}^{d_k}\left(I_{2i} - N_{2ik}/(N_{1ik}+N_{2ik})\right)}{\sqrt{\sum_{i=1}^{d_k} N_{1ik}N_{2ik}/(N_{1ik}+N_{2ik})^2}}, k = 1,\ldots,K,$$

is calculated, where N_{1ik} and N_{2ik} are the number of patients at risk at stage k in treatment groups 1 and 2, respectively, when the ith event occurred.

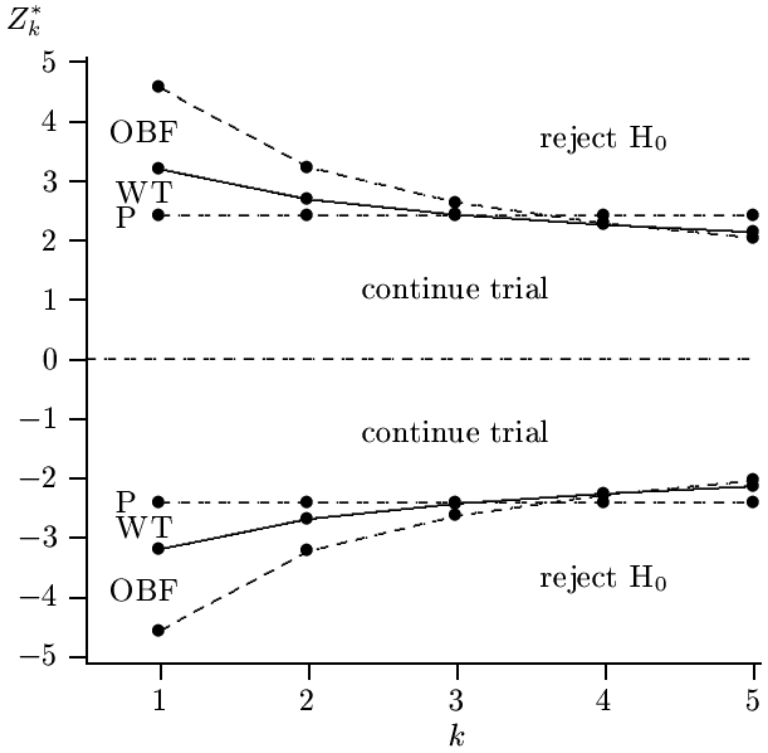
Fundamental Result:

Approximately, the sequence of logrank test statistics $LR_1^*,\ldots,LR_K^*$ has the independent and normally distributed increments structure.

Therefore, the group sequential tests can be applied in the usual way.

Decision Boundaries

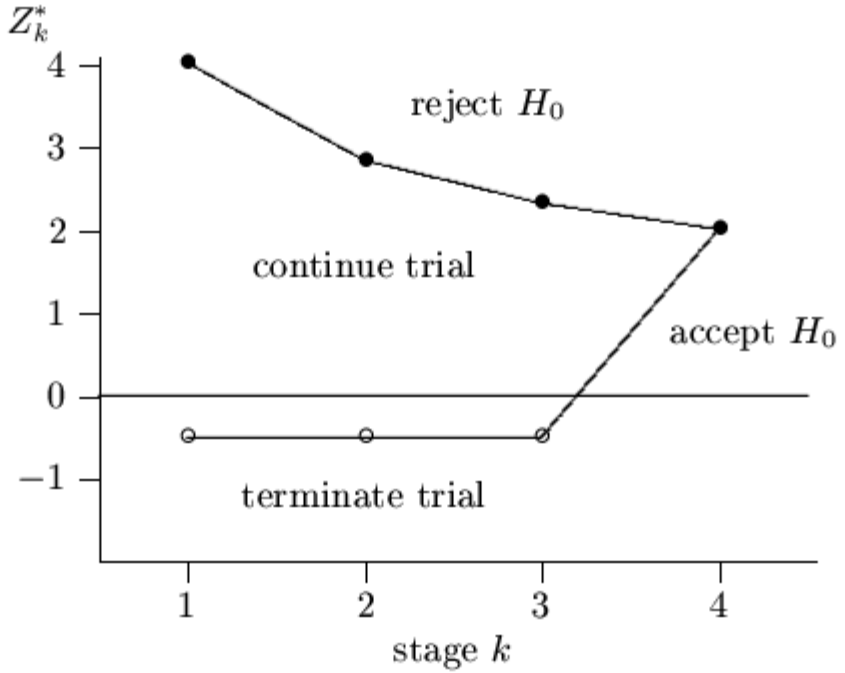
Pocock and O'Brien & Fleming design
Wang & Tsiatis Δ-class



Decision regions of Wang & Tsiatis test (WT) for $\Delta = 0.25$ (solid line) as compared to O'Brien & Fleming's (OBF) and Pocock's (P) design (dashed lines); $K = 5$, $\alpha = 0.05$.

Decision Boundaries

DeMets & Ware design: One-sided

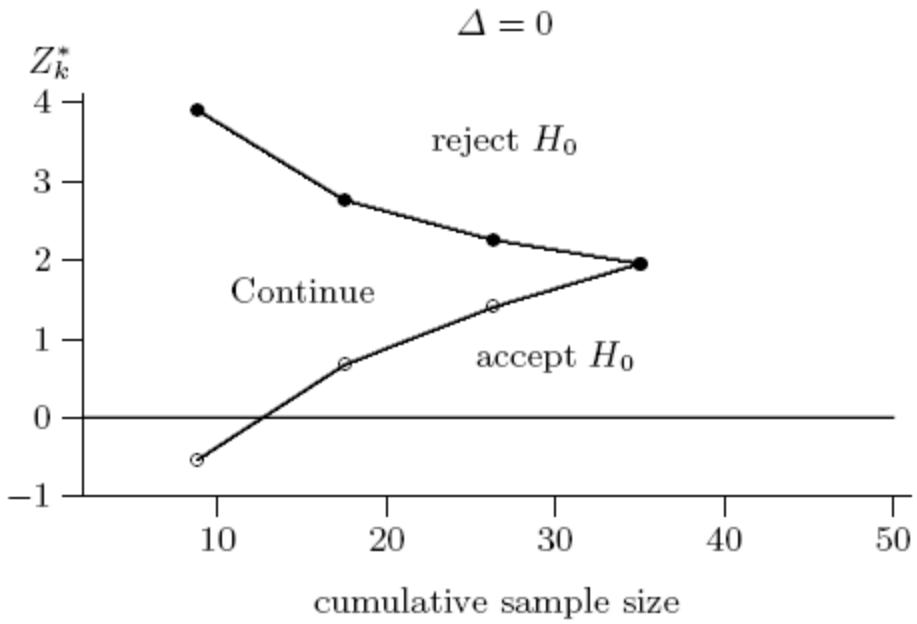


Continuation and decision regions for the one-sided design of DeMets & Ware with O'Brien & Fleming type boundaries; $K = 4$, $\alpha = 0.025$, $u^L = -0.5$.

Note: The futility boundary is binding!

Decision Boundaries

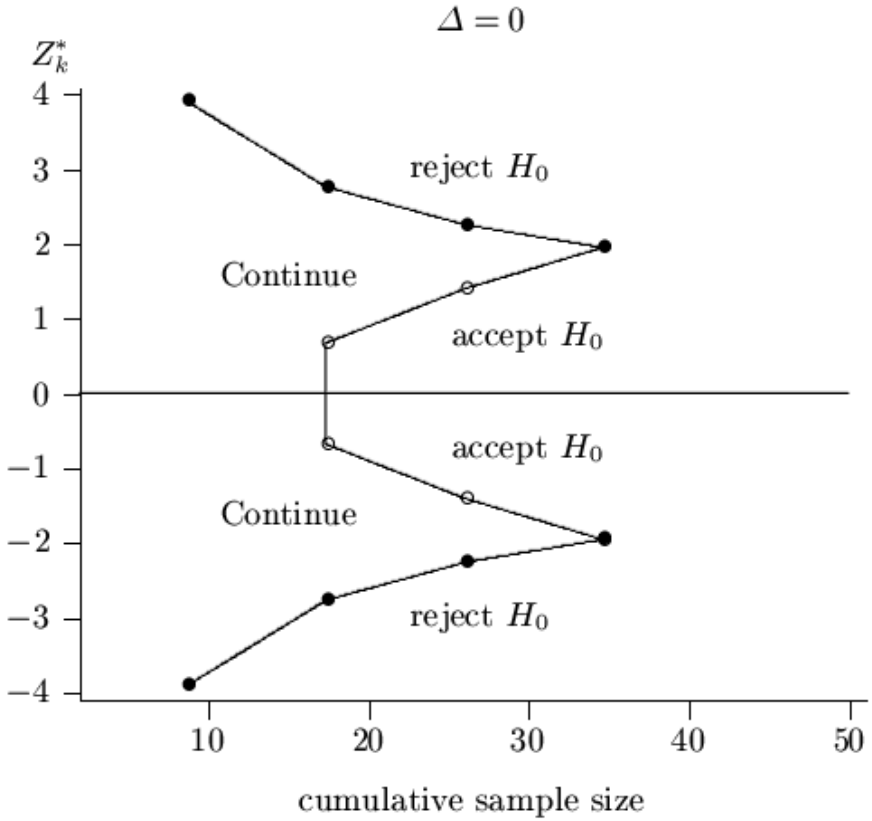
Pampallona & Tsiatis design (one-sided)



Continuation and decision regions for the one-sided design of Pampallona & Tsiatis; $K = 4$, $\alpha = 0.025$, $1 - \beta = 0.80$, $\delta = 0.5$

Decision Boundaries

Pampallona & Tsiatis design (two-sided)



Continuation and decision regions for the two-sided design of Pampallona & Tsiatis; $K = 4$, $\alpha = 0.05$, $1 - \beta = 0.80$, $\delta = 0.5$, $\Delta = 0$

Planning Aspects

Design with fixed sample size (two-sided):

$$n_f = (u_{1-\alpha/2} + u_{1-\beta})^2 \frac{2\sigma^2}{(\mu_1 - \mu_2)^2}$$

Group sequential test :

$$N = N(\alpha, \beta, K) = \vartheta^*(\alpha, \beta, K)^2 \frac{2\sigma^2}{(\mu_1 - \mu_2)^2}$$

Inflation factor

$$I = I(\alpha, \beta, K) = \frac{N}{n_f}$$

Planning Aspects

Average sample size:

$$ASN = n_1 + n_2 P(|Z_1^*| < u_1) + n_3 P(|Z_1^*| < u_1, |Z_2^*| < u_2) + \ldots +$$
$$n_K P(|Z_1^*| < u_1, \ldots, |Z_{K-1}^*| < u_{K-1})$$

Expected reduction in sample size under H_1 :

$$\frac{ASN_{H_1}}{n_f}$$

Planning Aspects

Inflation factor and expected reduction in sample size for the designs of O'Brien & Fleming and Pocock, respectively, for different values of K, significance level α, and power $1-\beta$. The expected reduction in sample size is presented in parentheses.

		$1-\beta = 0.80$		$1-\beta = 0.90$	
	K	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$
O'Brien & Fleming	1	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
	2	1.001 (0.947)	1.008 (0.902)	1.001 (0.912)	1.007 (0.851)
	3	1.007 (0.886)	1.017 (0.856)	1.006 (0.837)	1.016 (0.799)
	4	1.011 (0.862)	1.024 (0.831)	1.010 (0.806)	1.022 (0.767)
	5	1.015 (0.847)	1.028 (0.818)	1.014 (0.789)	1.026 (0.750)
Pocock	1	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
	2	1.092 (0.872)	1.110 (0.853)	1.083 (0.798)	1.100 (0.776)
	3	1.137 (0.841)	1.166 (0.818)	1.125 (0.750)	1.151 (0.721)
	4	1.166 (0.828)	1.202 (0.805)	1.152 (0.728)	1.183 (0.697)
	5	1.187 (0.822)	1.228 (0.799)	1.171 (0.717)	1.206 (0.685)

Planning Aspects

Example

α = 0.05 two-sided, 1-β = 0.80, K = 4, one-sample design

Expected or minimum clinically relevant effect

$$\delta = \mu_1 - \mu_2 = 0.50, \quad \sigma = 1$$

$$n_f = \frac{(1.96 + 0.842)^2}{0.50^2} = 31.4$$

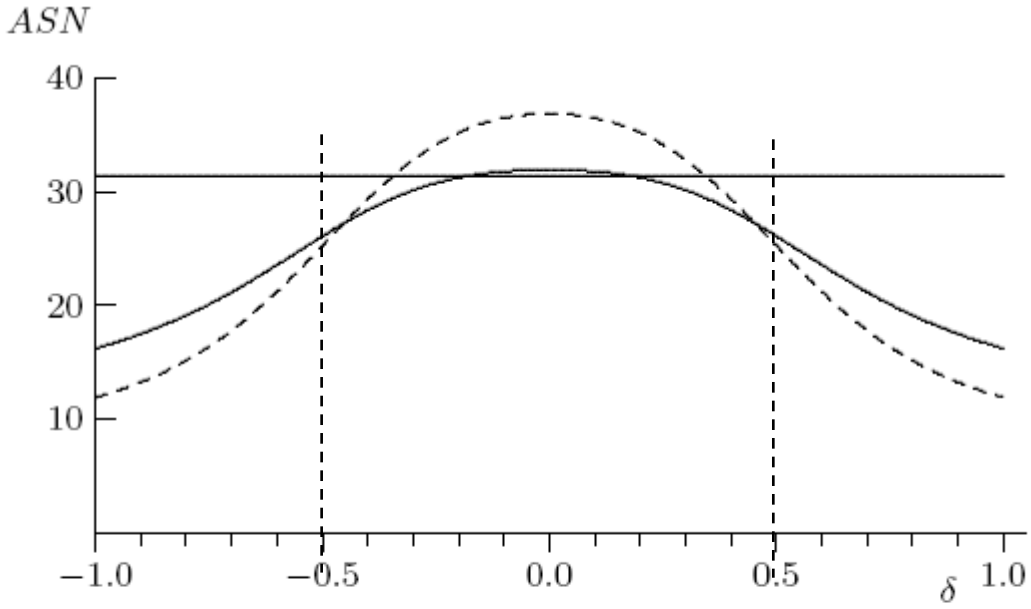
Pocock :

$$N = 1.202 \cdot 31.4 = 37.8 \text{ , } ASN_{H_1} = 0.805 \cdot 31.4 = 25.3$$

O'Brien & Fleming :

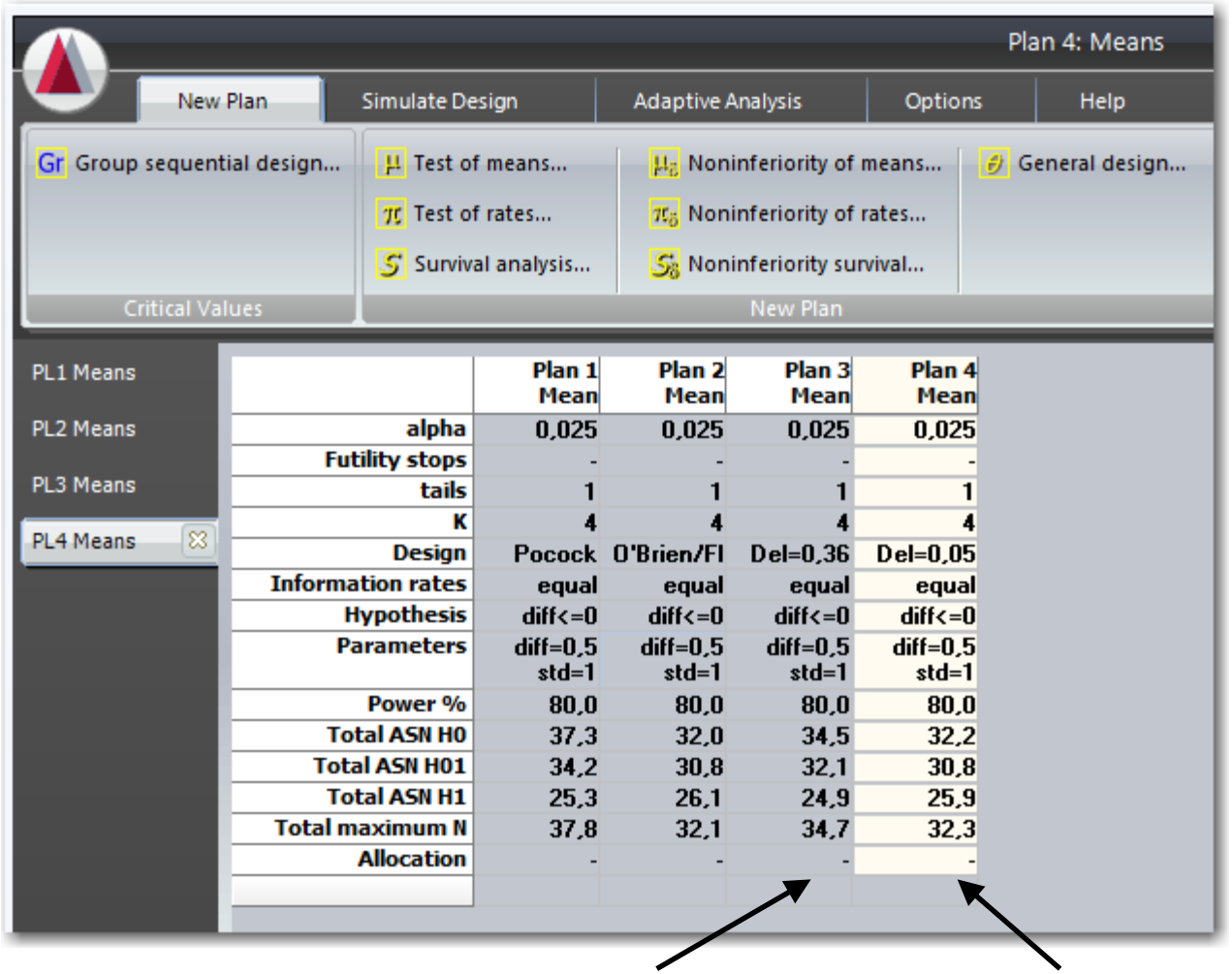
$$N = 1.024 \cdot 31.4 = 32.1, ASN_{H_1} = 0.831 \cdot 31.4 = 26.1$$

Planning Aspects



Average sample size of O'Brien & Fleming's (solid line) and Pocock's (dashed line) test. The sample size was calculated to achieve power $1 - \beta = 0.80$ at $|\delta| = 0.50$ ($\alpha = 0.05$, four-stage design). Horizontal line: sample size for fixed sample design.

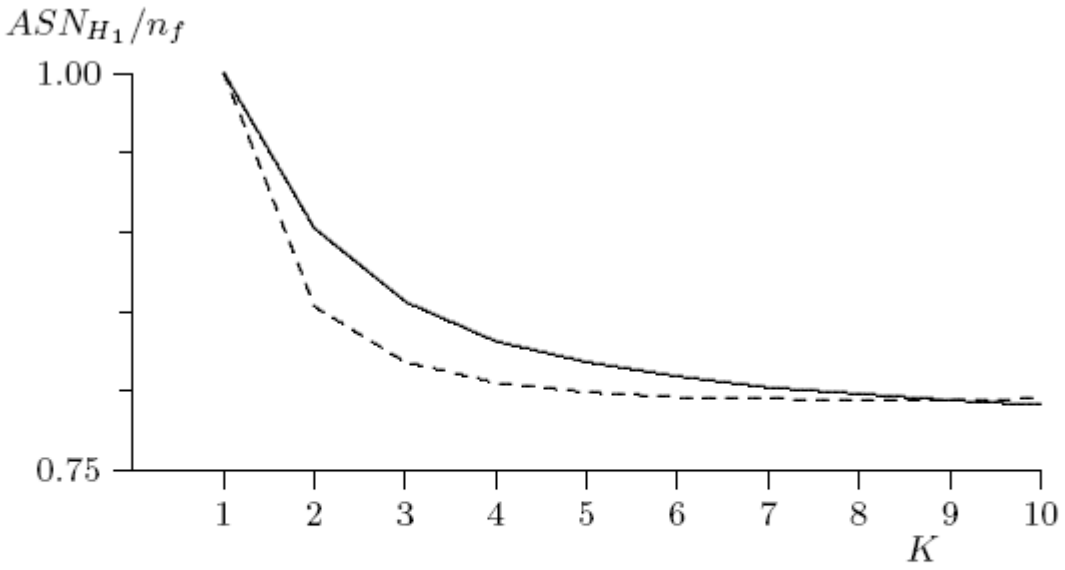
Planning Aspects



	Plan 1 Mean	Plan 2 Mean	Plan 3 Mean	Plan 4 Mean
alpha	0,025	0,025	0,025	0,025
Futility stops	-	-	-	-
tails	1	1	1	1
K	4	4	4	4
Design	Pocock	O'Brien/Fl	Del=0,36	Del=0,05
Information rates	equal	equal	equal	equal
Hypothesis	diff<=0	diff<=0	diff<=0	diff<=0
Parameters	diff=0,5 std=1	diff=0,5 std=1	diff=0,5 std=1	diff=0,5 std=1
Power %	80,0	80,0	80,0	80,0
Total ASN H0	37,3	32,0	34,5	32,2
Total ASN H01	34,2	30,8	32,1	30,8
Total ASN H1	25,3	26,1	24,9	25,9
Total maximum N	37,8	32,1	34,7	32,3
Allocation	-	-	-	-

Optimum designs: minimum ASN_{H_1}, minimum $ASN_{H_0}+ASN_{H_{0/1}}+ASN_{H_1}$

Planning Aspects: Choice of *K*



Expected reduction in sample size under H_1 relative to the sample size in a fixed sample size design, ASN_{H_1}/n_f, for different K of O'Brien & Fleming's (solid line) and Pocock's (dashed line) test; power $1-\beta = 0.80$, significance level $\alpha = 0.05$.

„α-spending“ or „use function“ Approach

Lan & DeMets (1983), Kim & DeMets (1987)

- α-spending function $\alpha^*(t_k)$ with $\alpha^*(0) = 0$ and $\alpha^*(1) = \alpha$,
- Fix N (maximum amount of information)
- Determine critical values $u_1, u_2, \ldots, u_k$ successively

$$P_{H_0}(|Z_1^*| \geq u_1) = \alpha^*(t_1),$$

$$P_{H_0}(|Z_1^*| \leq u_1, \ldots, |Z_{k-1}^*| \leq u_{k-1}, |Z_k^*| \geq u_k) = \alpha^*(t_k) - \alpha^*(t_{k-1})$$

Examples of α-spending Functions

$$\alpha_1^*(t_k) = 4(1-\Phi(u_{1-\alpha/4}/\sqrt{t_k})) \rightarrow \text{O'Brien \& Fleming (two-sided)}$$

$$\alpha_2^*(t_k) = \alpha \ln(1+(e-1)t_k) \rightarrow \text{Pocock}$$

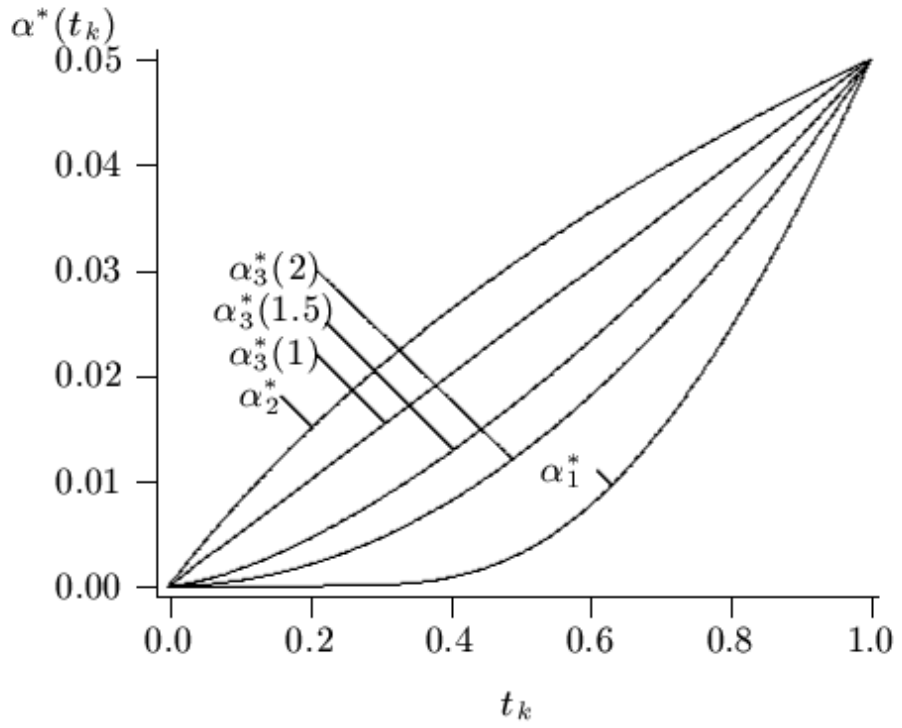
Kim & DeMets (1987):

$$\alpha_3^*(t_k) = \alpha t_k^\rho$$

Hwang et al. (1990):

$$\alpha_4^*(t_k) = \begin{cases} \alpha \dfrac{1-e^{-\gamma t_k}}{1-e^{-\gamma}}, & \gamma \neq 0 \\ \alpha t_k, & \gamma = 0 \end{cases}$$

Examples of α-spending Functions



Examples of α-spending functions. α_1^* and α_2^* approximate O'Brien & Fleming's and Pocock's design, respecitively. $\alpha_3^*(\varrho)$ is plotted for $\varrho = 1.0$, 1.5, and 2.0; $\alpha = 0.05$.

„α-spending“ or „use function“ Approach

- Computation of critical values does not depend on future information rates.
- Accounting for random under- and overrunning is possible.
- Specifically applicative for survival data
- Number of interim analyses need not be fixed in advance.
- Planning is usually based on assuming equidistant information rates but can also be performed for suitably chosen information rates.

For all approaches (including α-spending):

Don´t fix the subsequent sample sizes in a „data driven“ way. This could lead to a serious inflation of the Type I error rate. The effects of not considering this is described in, e.g., Proschan, Follmann & Waclawiw (1992).

Furthermore, you have to fix the designing parameters (e.g., shape of decision boundaries, the test statistic to be used, the hypothesis to be tested) prior to the experiment. These cannot be changed during the course of the trial.

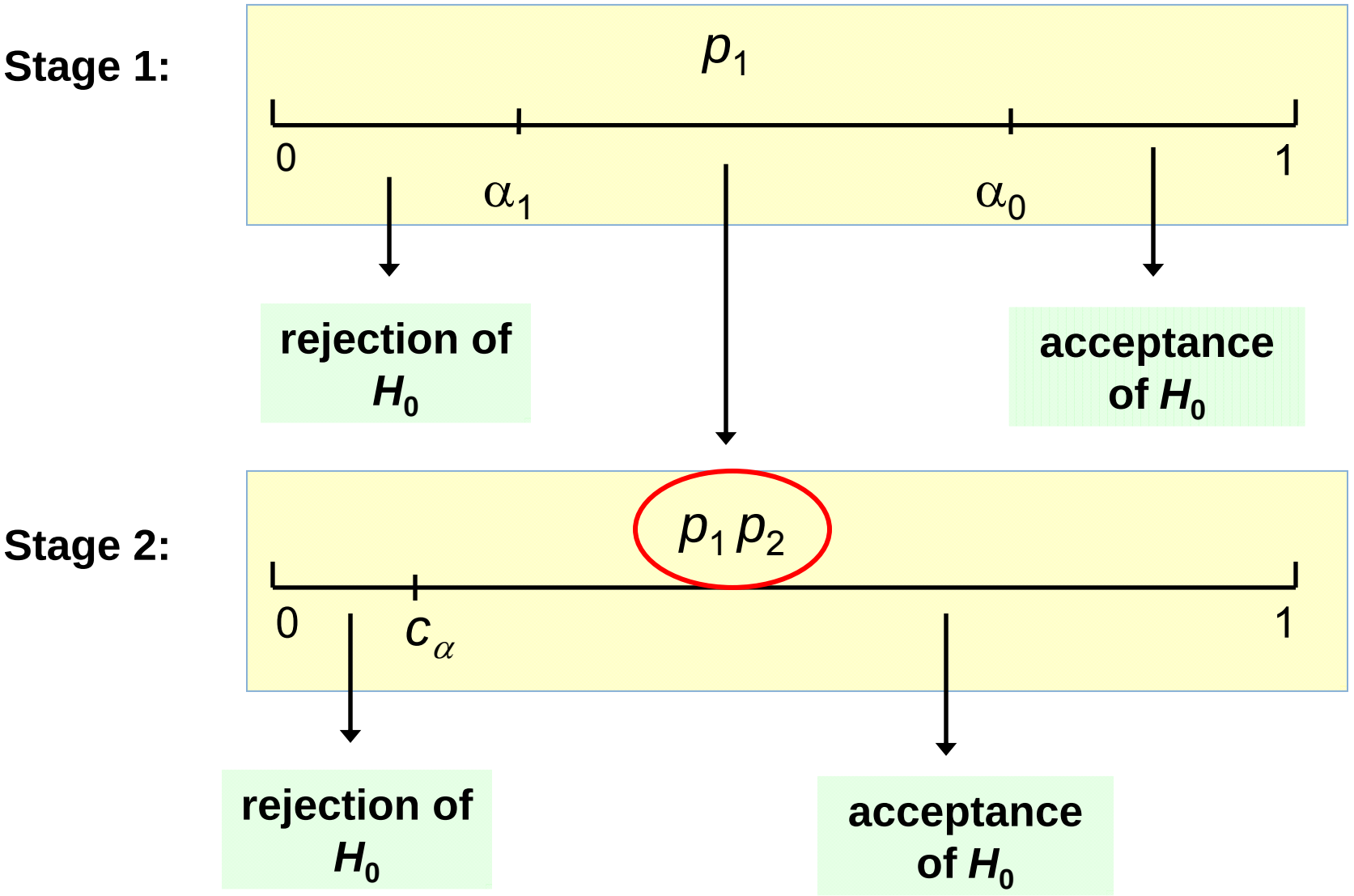
Adaptive Confirmatory Designs

All information available in an interim analysis may be used for planning the subsequent stages of the trial, under control of the prespecified Type I error rate.

Two pioneering proposals:

1. Bauer & Köhne (Biometrics, 1994):
   Combination of *p*-values with a specific combination function (Bauer, 1989)
2. Proschan & Hunsberger (Biometrics, 1995):
   Specification of a conditional error function

Procedure of Bauer & Köhne (1994)



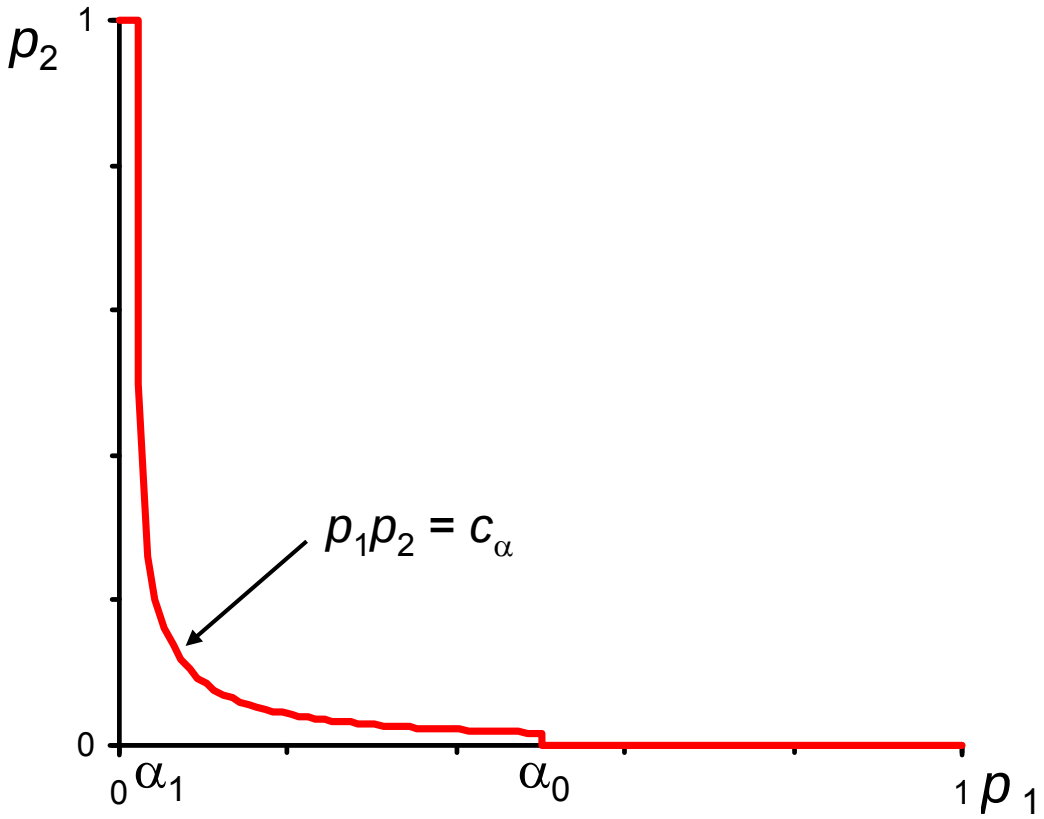
Procedure of Bauer & Köhne (1994)

- Use of Fisher´s combination test to combine the separate stage p-values p_1 and p_2, i.e., $C(p_1,p_2) = p_1 p_2$
- Under H_0, the p-values are stochastically independent, irrespective of the choice of the design for the second stage.
- H_0 is rejected after the second stage if

$$p_1 p_2 \leq c_\alpha = \exp(-1/2\, \chi^2_{4,\alpha})$$

- Other combination functions $C(p_1,p_2)$ and/or more than two stages can also be considered.
- In the two stages, different hypotheses can be considered, the considered global test is a test for $H_0 = H_0^1 \cap H_0^2$

Procedure of Bauer & Köhne (1994)



$$c_\alpha = \exp(-1/2\chi^2_{4,\alpha})$$

Compute α_1 from $\alpha_1 + c_\alpha(\ln\alpha_0 - \ln\alpha_1) = \alpha$

Possible Data-Dependent Changes of Design

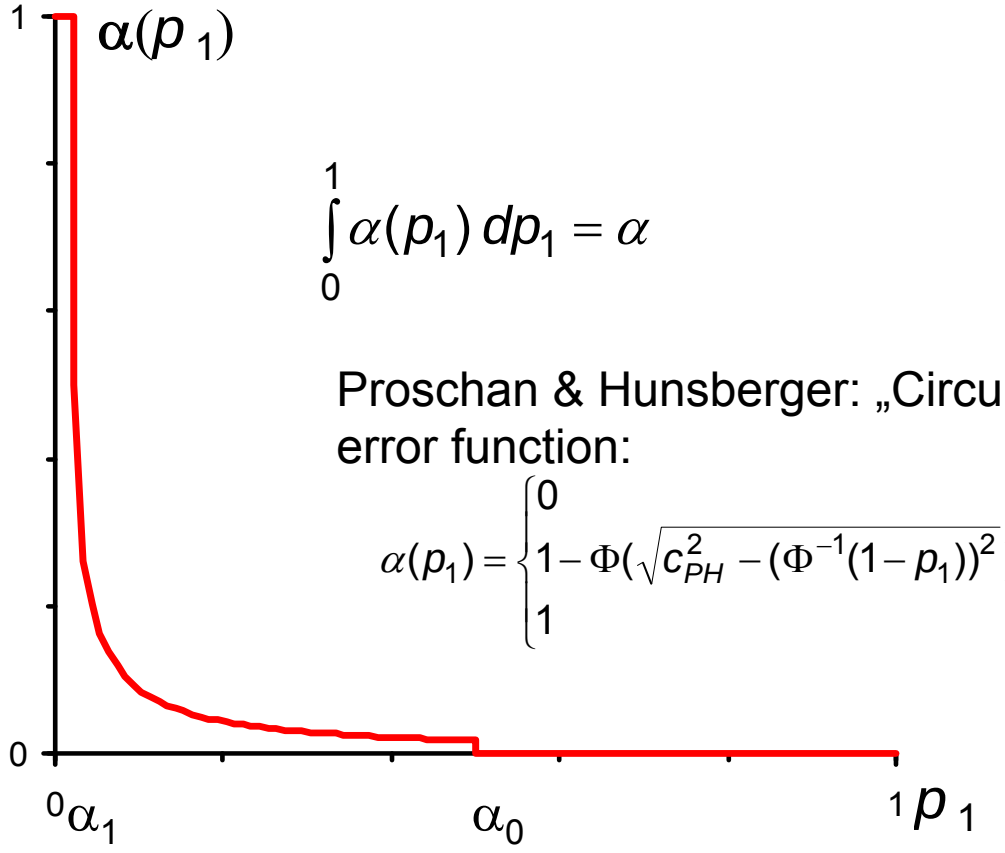
- Reassessment of sample size
- Adaptive choice of test statistic
- Combining Phase II/III studies

  (adaptive seamless phase II/III designs)
- Selection of study population (enrichment designs)
- Selection of endpoints
- Change of target parameter
- Modification of ordering of hypothesis

Alternative (Proschan & Hunsberger, 1995)

Specification of a „conditional error function“ $\alpha(p_1)$

Rejection of H_0 in second stage if $p_2 \leq \alpha(p_1)$



$$\int_0^1 \alpha(p_1)\, dp_1 = \alpha$$

Proschan & Hunsberger: „Circular“ conditional error function:

$$\alpha(p_1) = \begin{cases} 0 & \text{if} \quad p_1 \geq \alpha_0 \\ 1 - \Phi\left(\sqrt{c_{PH}^2 - (\Phi^{-1}(1-p_1))^2}\right) & \text{if} \quad 1\text{-}\Phi(c_{PH}) < p_1 < \alpha_0 \\ 1 & \text{if} \quad p_1 \leq 1\text{-}\Phi(c_{PH}) \end{cases}$$

General Result

The method based on the concept of the conditional error function can be looked at in terms of combination tests and vice versa.

Conditional error function of Bauer & Köhne procedure:

$$\alpha(p_1) = \begin{cases} 0 & \text{if} \quad p_1 \geq \alpha_0 \\ c_\alpha / p_1 & \text{if} \quad \alpha_1 < p_1 < \alpha_0 \\ 1 & \text{if} \quad p_1 \leq \alpha_1 \end{cases}$$

Adaptive Design using the *inverse normal method*

Consider at *k*th stage, $k = 1,2,...,K$:

$$T_k^* = C(p_1,...,p_k) = \frac{\Phi^{-1}(1-p_1) + \Phi^{-1}(1-p_2) + \ldots + \Phi^{-1}(1-p_k)}{\sqrt{k}}$$

$\Phi^{-1}(1-p_k) \sim N(0;1)$ if p_k uniformly distributed on [0; 1]

Under H_0, the same distributional assumption as for the group sequential tests applies and, therefore, the decision regions of the traditional group sequential tests can be used.

Lehmacher & Wassmer, 1999

Properties

- Decision regions of group sequential tests can be used. Generalization to more than two stages and more general designs straightforward.
- Use *unweighted* mean of test statistics from the separate stages also for unequal and arbitrarily (data dependent) fixed sample sizes.
- Effect on power is small unless „dramatic“ changes in sample size were performed.
- Can also be used in testing situations with nuisance parameters.
- If no design changes were performed, the inverse normal technique yields the traditional test.

More general:

Consider at *k*th stage

$$T_k^* = \frac{w_1 \Phi^{-1}(1-p_1) + w_2 \Phi^{-1}(1-p_2) + \ldots + w_k \Phi^{-1}(1-p_k)}{\sqrt{w_1^2 + w_2^2 + \ldots + w_k^2}},$$

where $w_1, w_2, \ldots, w_K$ are weights fixed prior to the trial.

Weighted inverse normal method, proposed for adaptive designs by Lehmacher & Wassmer (Biometrics, 1999).

Another (equivalent) Approach

Cui, Hung & Wang (Biometrics, 1999) independently proposed the same approach by considering the overall test statistic

$$Z_k^* = \frac{w_1 Z_1 + w_2 Z_2 + \ldots + w_k Z_k}{\sqrt{w_1^2 + w_2^2 + \ldots + w_k^2}}$$

However, they only considered the aspect of reassessing the sample size when considering the test statistic for the known variance case. The (weighted) inverse normal method is much more general!

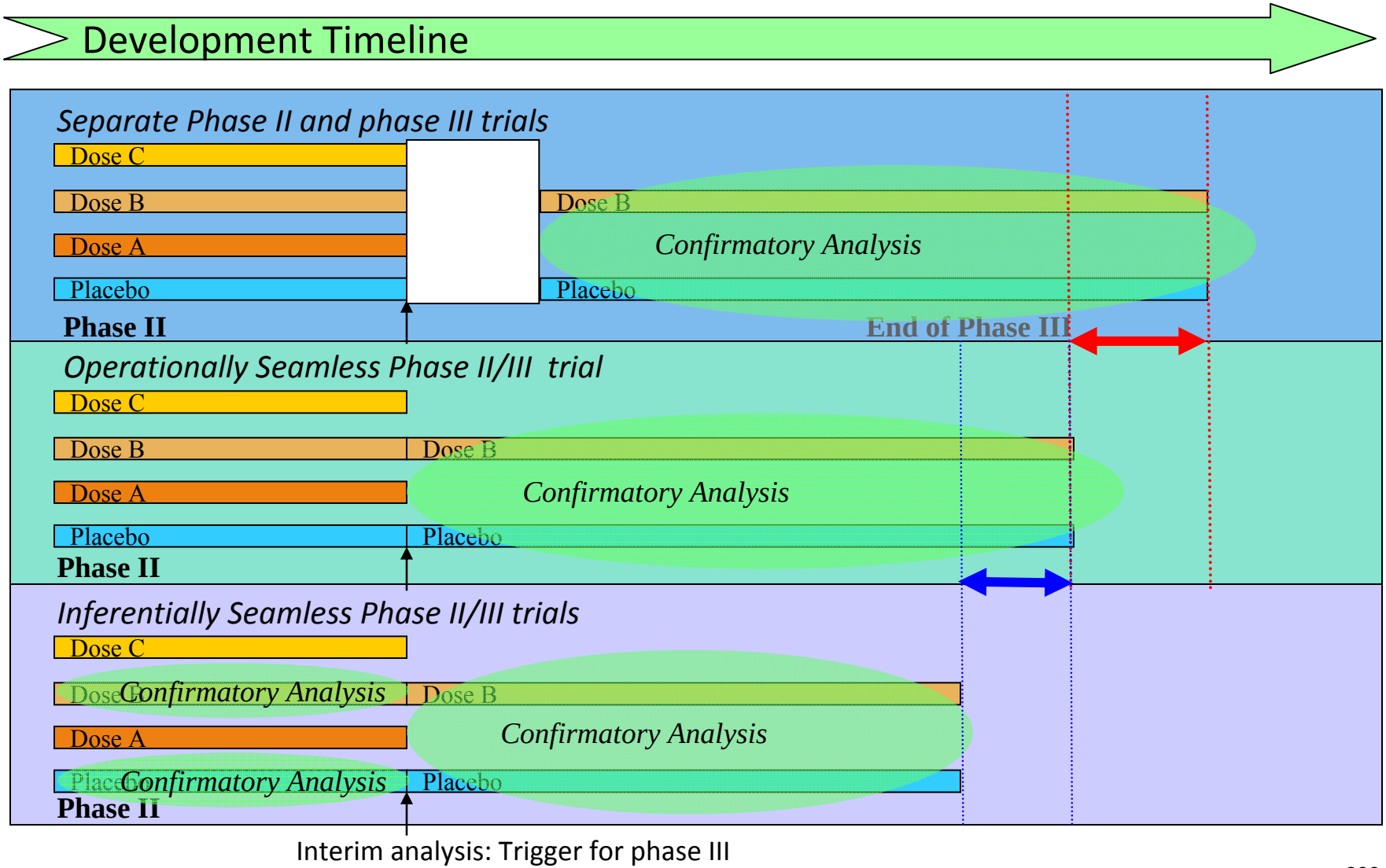
Outline

- Introduction and Taxonomy of Clinical Trial Designs
- Basic Principles of Adaptive designs
  - Allocation Rule
  - Sampling Rule
  - Stopping Rule
  - Decision Rule
- Phases of Development
- Adaptive Designs for the Learn Phase of Drug development
  - First-in Human / MTD
  - Two-Stage Designs
  - Adaptive Dose-Ranging designs
- Adaptive Designs for the Confirmatory Phase of Drug Development
  - Sample Size Re-Assessment
  - Adaptive Group Sequential designs
  - Seamless Phase II/III Designs
  - Population Enrichment Designs
- Practical Aspects of Adaptive Design Implementation
- Discussion

Seamless Designs

- Seamless design
  - A clinical trial design which combines into a single trial objectives which are traditionally addressed in separate trials (*operationally seamless*)
- Adaptive Seamless design
  - A seamless trial in which the final analysis will use data from patients enrolled before and after the adaptation (*inferentially seamless*)
- Primary objective – combine "dose selection" and "confirmation" into a single trial
- Benefits: Efficiency; faster and more informed decision-making
- Challenges: Effective and Efficient Implementation

Efficiency of Adaptive Seamless Phase II/III Designs



Adaptations

- 1st Stage is mainly for “dose selection”
- No intention for early stopping for efficacy (trial can always be stopped for safety considerations or for futility)
- After dose selection, the only change is to new enrollments (patients are generally not re-randomized)
- Patients on terminated treatment groups could be followed
- All data from the 1st and 2nd stages is used in the final analysis. Appropriate statistical methods must be used

Methodology

- Sources for alpha inflation
  - Interim analyses
  - Sample size reassessment
  - Multiple arms
- The proposed adaptive procedure fulfils the regulatory requirements for the analysis of adaptive trials in that it strongly controls the pre-specified Type I error rate (FEW control)
- The procedure is based on the application of the closed test procedure together with combination tests

Major References

Testing and estimation in flexible group sequential designs with adaptive treatment selection

Martin Posch[1,*,†], Franz Koenig[1], Michael Branson[2], Werner Brannath[1], Cornelia Dunger-Baldauf[2] and Peter Bauer[1]

[1]*Section of Medical Statistics, Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Spitalgasse 23, A-1090 Wien, Austria*
[2]*Novartis Pharma AG, Biostatistics and Statistical Reporting, Lichtstrasse 35, CH-4002 Basel, Switzerland*

SUMMARY

Integrating selection and confirmation phases into a single trial can expedite the development of new treatments and allows to use all accumulated data in the decision process. In this paper we review adaptive treatment selection based on combination tests and propose overall adjusted *p*-values and simultaneous confidence intervals. Also point estimation in adaptive trials is considered. The methodology is illustrated in a detailed example based on an actual planned study. Copyright © 2005 John Wiley & Sons, Ltd.

STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:3697–3714

TUTORIAL IN BIOSTATISTICS

Adaptive designs for confirmatory clinical trials

Frank Bretz[1,2,‡], Franz Koenig[3,*,†,‡], Werner Brannath[3], Ekkehard Glimm[1] and Martin Posch[3]

[1]*Novartis Pharma AG, Lichtstrasse 35, 4002 Basel, Switzerland*
[2]*Department of Biometry, Medical University of Hannover, 30623 Hannover, Germany*
[3]*Section of Medical Statistics, Medical University of Vienna, Spitalgasse 23, 1090 Wien, Austria*

SUMMARY

Adaptive designs play an increasingly important role in clinical drug development. Such designs use accumulating data of an ongoing trial to decide how to modify design aspects without undermining the validity and integrity of the trial. Adaptive designs thus allow for a number of possible adaptations at midterm: Early stopping either for futility or success, sample size reassessment, change of population, etc. A particularly appealing application is the use of adaptive designs in combined phase II/III studies with treatment selection at interim. The expectation has arisen that carefully planned and conducted studies based on adaptive designs increase the efficiency of the drug development process by making better use of the observed data, thus leading to a higher information value per patient.

In this paper we focus on adaptive designs for confirmatory clinical trials. We review the adaptive design methodology for a single null hypothesis and how to perform adaptive designs with multiple hypotheses using closed test procedures. We report the results of an extensive simulation study to evaluate the operational characteristics of the various methods. A case study and related numerical examples are used to illustrate the key results. In addition we provide a detailed discussion of current methods to calculate point estimates and confidence intervals for relevant parameters. Copyright © 2009 John Wiley & Sons, Ltd.

STATISTICS IN MEDICINE
Statist. Med. 2009; **28**:1181–1217

Example

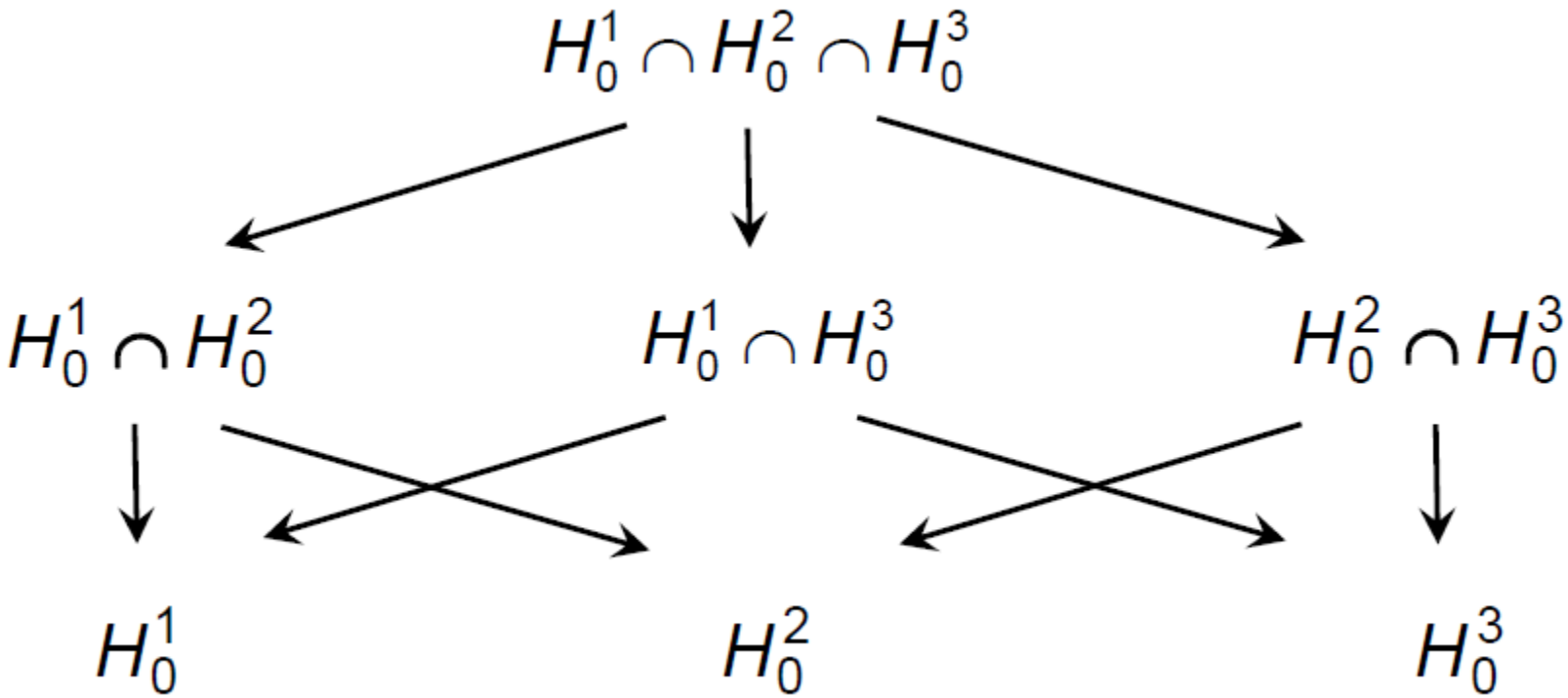
- *G* = *3*, equal sample sizes between the treatment groups

$$H_0^1 : \mu_1 = \mu_0 ,$$
$$H_0^2 : \mu_2 = \mu_0 ,$$
$$H_0^3 : \mu_3 = \mu_0 .$$

- Assume that one treatment arm is to be selected at the first interim stage
- Confirmatory analysis should be possible for the comparison of the selected dose with the control

Closed Testing Principle

- Closed system of hypotheses



Intersection test

- At the first interim analysis, consider a test statistic for

$$H_0^1 \cap H_0^2 \cap H_0^3$$

$$Z_1^* = \max(Z_1^{1*}, Z_1^{2*}, Z_1^{3*})$$

- where Z_1^{i*} denotes the first stage test statistic for $H_0^i, i = 1,2,3$

- That is, compute Dunnett's adjusted *p-value for each intersection* hypothesis

Decision at the 1st Stage

- At the first interim analysis, it is possible to stop the trial while showing significance of one (or more) treatment arms.
- It is also possible to stop the trial due to futility arguments. These are usually based on conditional power calculations.
- It is expected, however, that the first stage is specifically used to select a treatment arm to be considered in the subsequent stages of the trial and to reassess the sample size for the subsequent stages.

Test decision for the 2nd Stage

H_0^S is rejected if

$$\min_{J \ni S} \varphi(p_J, q_S) \geq u_2 ,$$

where p_J is the p-value of the Dunnett test for testing $\bigcap_{i \in J} H_0^i$,

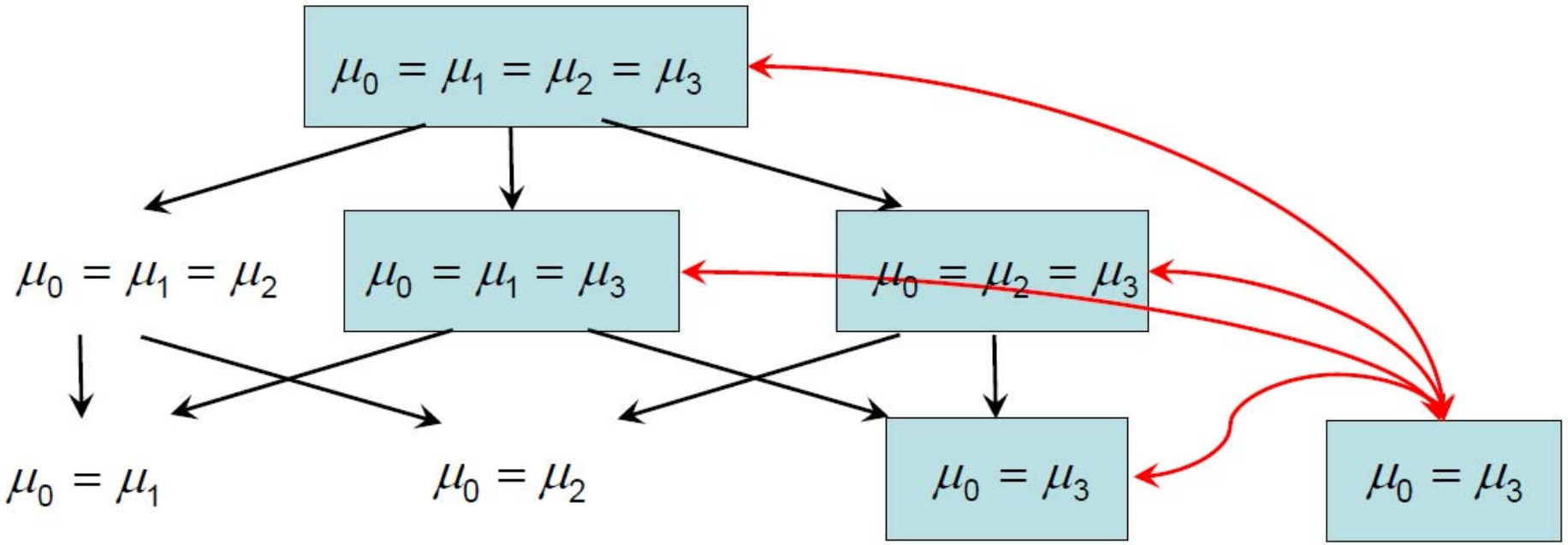
q_S is the second stage test statistic for the selected dose,

and u_2 is the critical value for the second stage.

$$\varphi(p,q) = \frac{w_1 \Phi^{-1}(1-p) + w_2 \Phi^{-1}(1-q)}{\sqrt{w_1^2 + w_2^2}}$$

Test decision for the 2nd Stage

1st STAGE **2nd STAGE**



H_0^3 can be rejected if all combination tests exceed the critical value u_2 .

Designing options in ADDPLAN

- The performance of a multi-armed adaptive designs depends on
  - the test procedure
  - the global (intersection) test
  - the sequential design
  - the effect sizes
  - the selection procedure
  - the sample size determination

Test strategies

- Combination test
  - Inverse normal method
  - Fisher's combination test
- Adaptive Dunnett
  - "Pure" conditional Dunnett: second stage is a conditional Dunnett test
  - "Recursive" conditional Dunnett: second stage is an unconditional Dunnett test
- Separate Phase II/III: Phase II only for selection, Phase III is group sequential
- One stage: only selection at interim

Assumptions

G experimental treatment groups and one control group, many-to-one comparisons

G elementary hypotheses $H_0^g : \theta_0 = \theta_g,\ g = 1, \ldots, G$

Global hypothesis $H_0 = \bigcap_{g=1}^{G} H_0^g : \theta_0 = \cdots = \theta_G$

Let $p_{(1)} \leq \ldots \leq p_{(G)}$ denote the ordered p-values of the G comparisons.

Let p_g denote the p-value for testing $H_0^g : \theta_0 = \theta_g,\ g = 1, \ldots, G$.

Let $J \subset \{1, \ldots, G\}$ and $|J|$ be defined as the number of all indices $g \in J$.

Intersection hypotheses $H_0^J = \bigcap_{g \in J} H_0^g$

Multiple Comparison Tests

Bonferroni Test for Many to One Comparisons

Using the Bonferroni test, the p-value for testing the global hypothesis H_0 is given by

$$p^{\text{BON}} = \min\{Gp_{(1)}, 1\} \; . \qquad \text{(BONFERRONI-GLOBAL)}$$

The adjusted p-value for the hypothesis H_0^J is given by

$$p_J^{\text{BON}} = \min\{|J| \min_{g \in J}\{p_g\}, 1\} \; . \qquad \text{(BONFERRONI)}$$

Šidák Method for Many to One Comparisons

With the Šidák test, the p-value for testing the global hypothesis H_0 is given by

$$p^{\text{SID}} = 1 - (1 - p_{(1)})^G \; . \qquad \text{(SIDAK-GLOBAL)}$$

The adjusted p-value for the hypothesis H_0^J is given by

$$p_J^{\text{SID}} = 1 - (1 - \min_{g \in J}\{p_g\})^{|J|} \; . \qquad \text{(SIDAK)}$$

Simes Method for Many to One Comparisons

With the Simes global test, the p-value for testing the global hypothesis H_0 is given by

$$p^{\text{SIM}} = \min_g \{\frac{G}{g} p_{(g)})\} \; . \qquad \text{(SIMES-GLOBAL)}$$

The adjusted p-value for the hypothesis H_0^J is given by

$$p_J^{\text{SIM}} = \min_{g \in J} \{\frac{|J|}{g} p_{(g_J)})\} \; , \qquad \text{(SIMES)}$$

where $p_{(1_J)} \leq \ldots \leq p_{(|J|_J)}$ denote the ordered p-values from the subset $J \subset G$.

Dunnett Test for Many to One Comparisons

The p-value for testing H_0 is calculated through

$$p^{\text{DUN}} = 1 - F_{\mathbf{\Sigma},\sum n_g-(G+1)}\left(\max_{g\in G}\frac{\bar{x}_g - \bar{x}_0}{s\sqrt{1/n_0 + 1/n_g}}\right), \qquad \text{(DUNNETT-GLOBAL)}$$

where $F_{\mathbf{\Sigma},(G+1)(n-1)}(\cdot)$ denotes the cdf of the multivariate t distribution. The elements of the correlation matrix $\mathbf{\Sigma}$ are $\rho_{gg'} = \varsigma_g\varsigma_{g'}$ for $g \neq g'$ where

$$\varsigma_g = \sqrt{n_g/(n_0 + n_g)}\,, \; g = 1\ldots, G\,.$$

Equivalently,

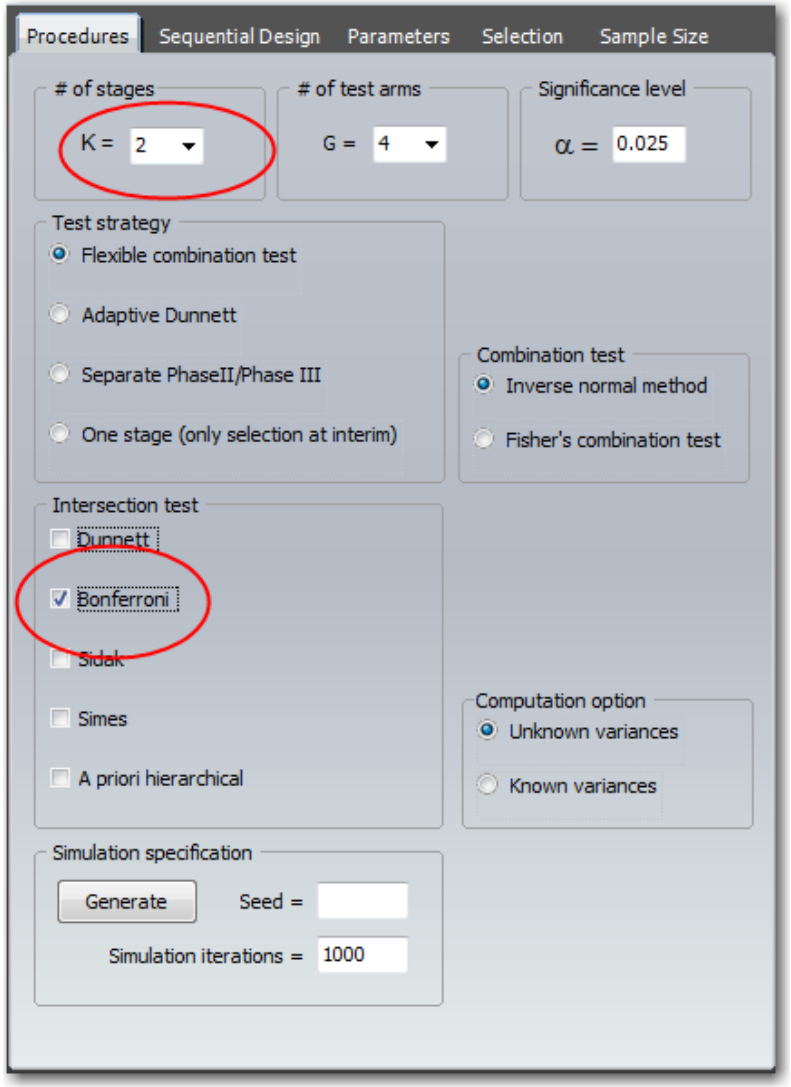
$$p_J^{\text{DUN}} = 1 - F_{\mathbf{\Sigma}_J,\sum_{g\in J} n_g-(|J|+1)}\left(\max_{g\in J}\frac{\bar{x}_g - \bar{x}_0}{s\sqrt{1/n_0 + 1/n_g}}\right), \qquad \text{(DUNNETT)}$$

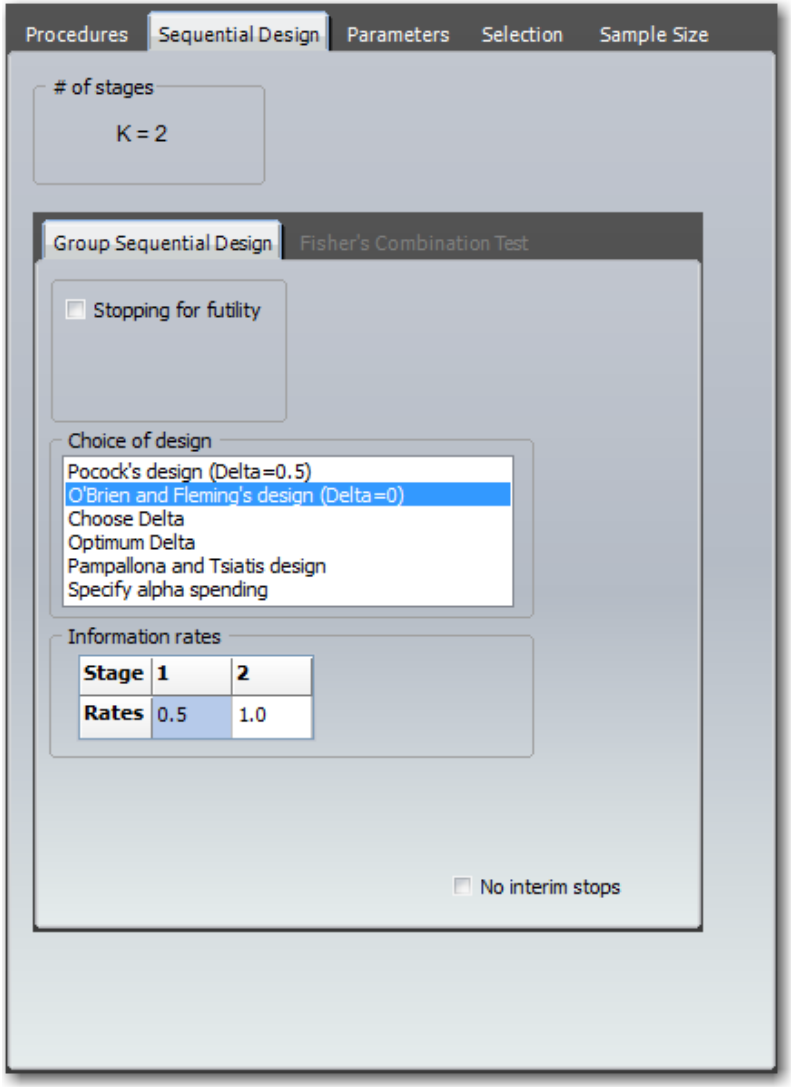
where $\mathbf{\Sigma}_J$ is the corresponding submatrix of the matrix $\mathbf{\Sigma}$.

Sequential Design

- Combination test
  - Inverse normal method
  - Fisher's combination test
- Group sequential design
  - Boundary approach
  - α spending function approach
- Only futility stopping

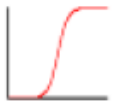
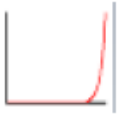
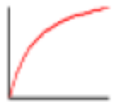
Procedures and Sequential Design



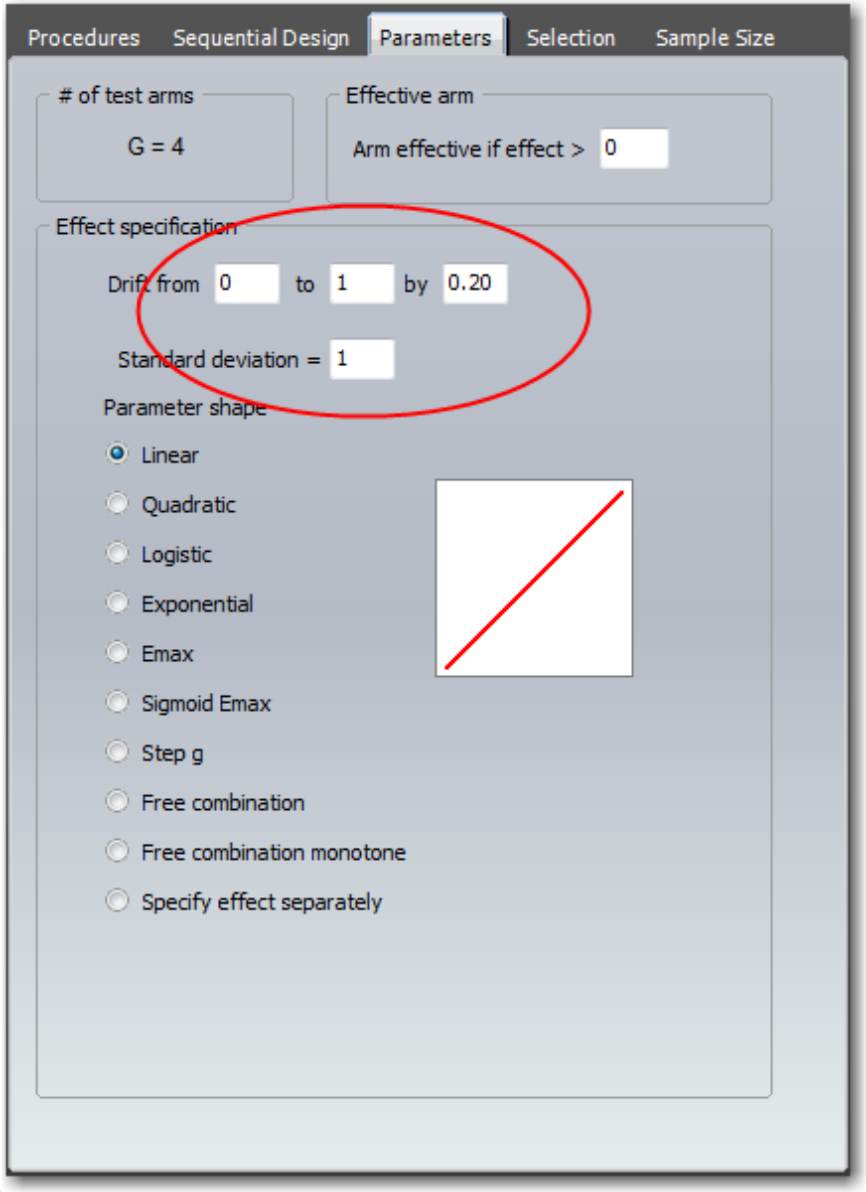


Treatment Effect

Different parameter shapes together with „drift“ parameters can be used to assess the statistical performance of a specified procedure

- Linear  $\delta_j = \frac{\vartheta}{G} j, \quad j = 1,\ldots,G$
- Quadratic  $\delta_j = \frac{2\vartheta}{g^2}(g\,j - \frac{j^2}{2}), \quad j = 1,\ldots,G, \quad \max \text{ at } j = g$
- Logistic  $\delta_j = \frac{\vartheta}{1+\exp(g-j)}, \quad j = 1,\ldots,G, \quad ED50 \text{ at } j = g$
- Exponential  $\delta_j = \frac{\vartheta}{\exp(G)}\exp(j), \quad j = 1,\ldots,G,$
- Emax  $\delta_j = \frac{\vartheta\, j}{g+j}, \quad j = 1,\ldots,G, \quad ED50 \text{ at } j = g$
- Step  Step at g

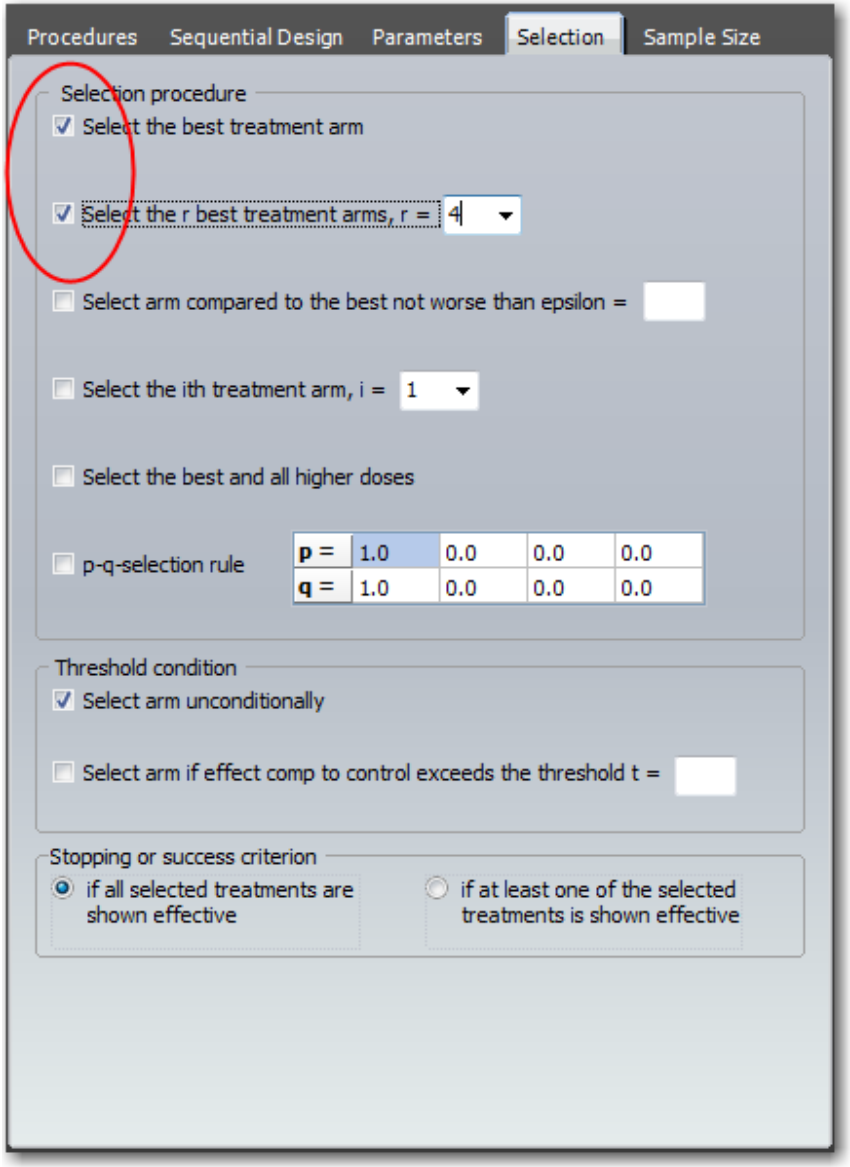
Parameters



Selection Procedure

- Select the best treatment arm
- Select the *r best treatment arms, specify r*
- Select arm compared to the best not worse than ε, specify ε
- Select the *i-th treatment arm, specify i*
- *p-q-selection rule:*
  - p_i = *Pr(select i treatment arms)*
  - q_i = *Pr(start selection at the i-th best treatment arm)*
- Select arm unconditionally
- Select arm if it exceeds a threshold *t, specify t*

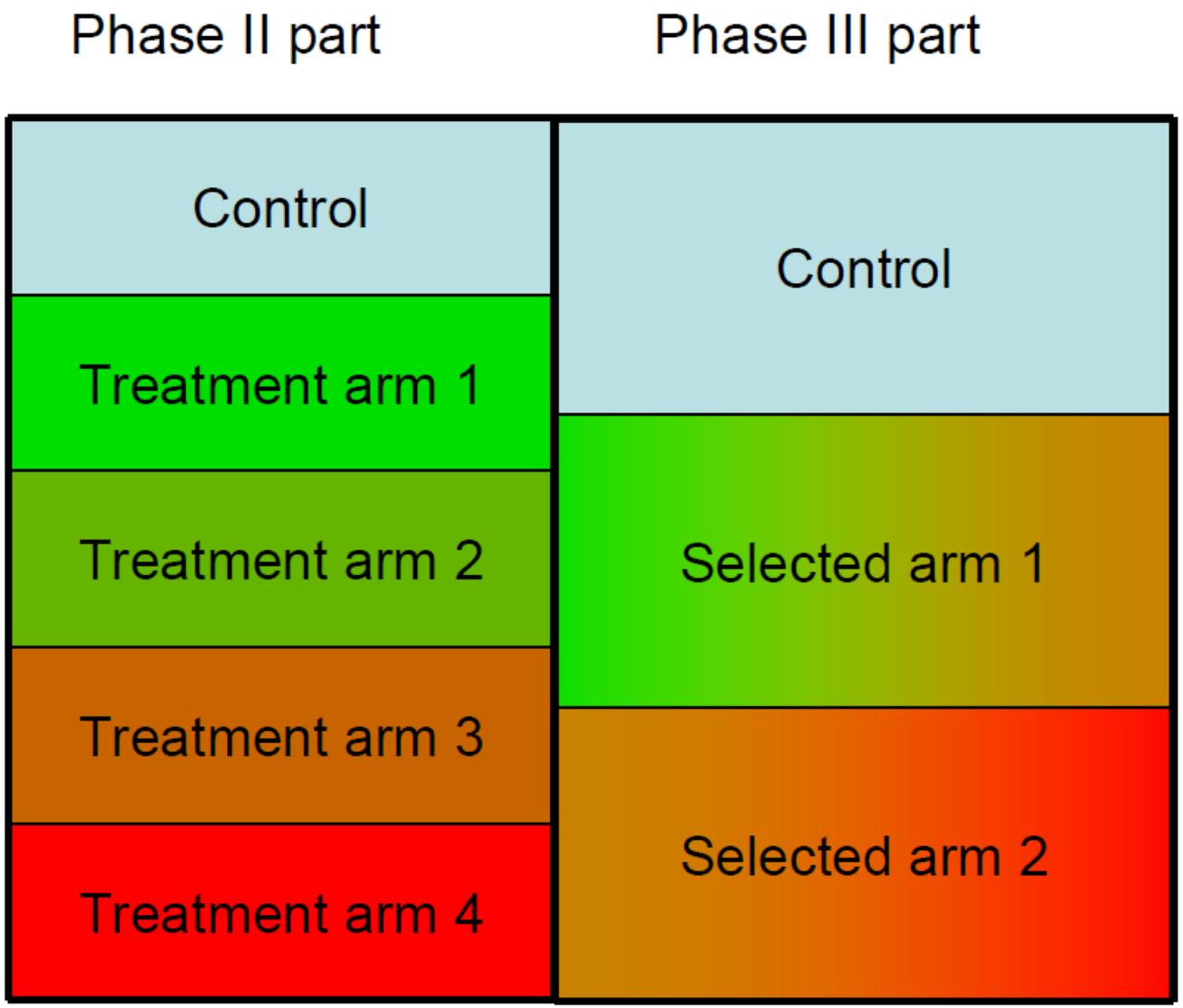
Selection Procedure



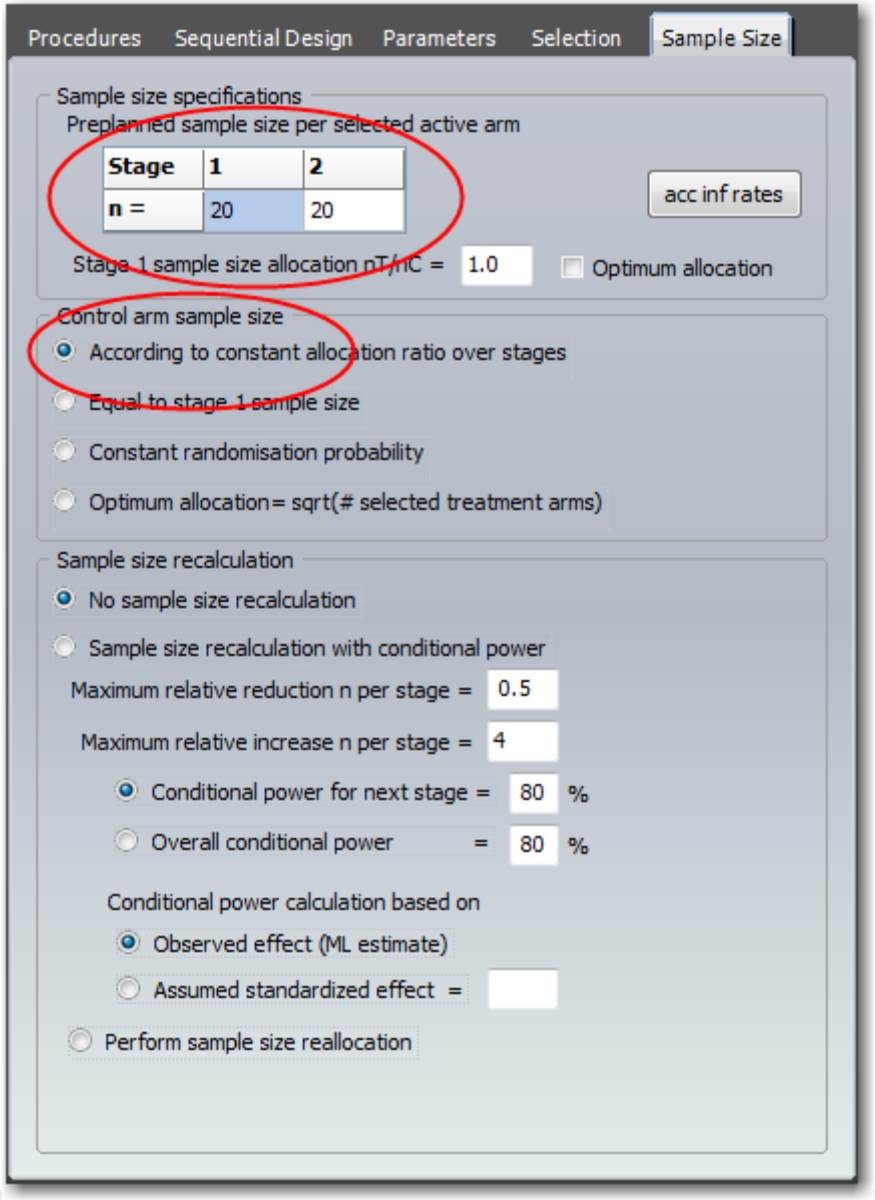
Sample Size

- **Sample size partition (sample size of control arm)**
  - Control arm sample size equal stage I sample size
  - Control arm sample size according to optimal allocation ratio: sqrt(# selected trt arms)
  - Control arm sample size with constant randomization probability
- **Sample size determination**
  - Fixed sample sizes
  - Sample size reassessment based on conditional power
  - Re-allocation of sample size

Re-allocation of Sample Size



Sample Size



Analyzing a Multi-Armed Design

Choose between the following testing strategies

- Combination test
  - Inverse normal method
  - Fisher‘s combination test
- Adaptive Dunnett
  - „Pure“ conditional Dunnett: second stage is a conditional Dunnett test
  - „Recursive“ conditional Dunnett: second stage is a unconditional Dunnett test
- Conventional group sequential designs
- Intersection tests and sequential designs as for the simulation module

Example: Seamless Phase II/III study in an Orphan Condition

- Two-stage group sequential design with O'Brien & Fleming boundaries
- Dunnett intersection test
- Three doses of a drug with pre-specified effect sizes
- Primary endpoint: Short-term response (0 or 1) (7 days)

Assumptions

- One-sided type I error 0.025
- Power 80%
- Placebo rate 45%

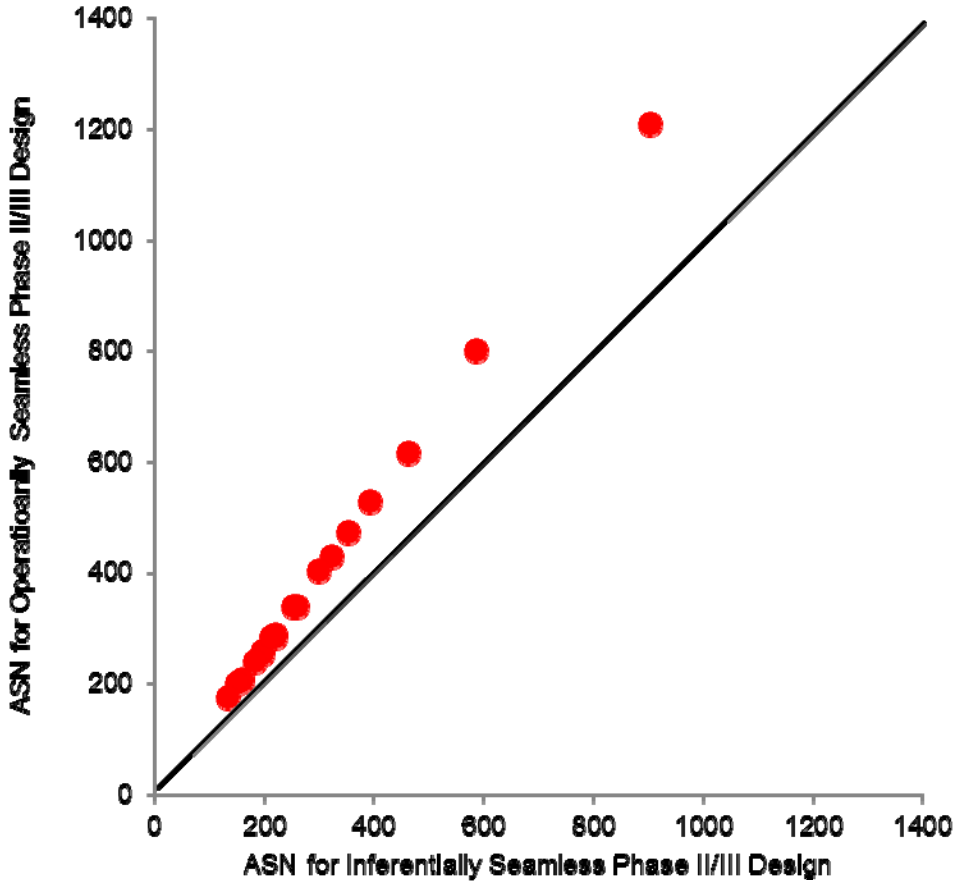
Comparison

- A seamless adaptive Phase II/III design with stopping after Part A only for efficacy using an O'Brien and Fleming boundary. The interim is assumed to be conducted at an information rate of 1/3.
- A non-seamless Phase II study with the same decision rule, and sample size, as the seamless design followed by a phase III study powered at 80%.

ADDPLAN MC Simulation Results

				Phase II/III Seamless OBF				Operationally Seamless (Phase II then Phase III)			
Scenarios (Response Rates)				Sample Size/Arm				Phase III		Overall Study	
Placebo	1	2	3	Phase II	Phase III	Power	ASN	Sample Size	% Increase	ASN	% Increase
0,45	0,35	0,35	0,35	115	230	0.799	902	376	63	1210	34
0,45	0,35	0,35	0,3	75	150	0.799	586	250	67	800	37
0,45	0,35	0,35	0,25	45	90	0.800	352	147	63	474	35
0,45	0,35	0,35	0,2	28	56	0.803	220	86	54	284	29
0,45	0,35	0,3	0,3	59	118	0.802	462	190	61	616	33
0,45	0,35	0,3	0,25	41	82	0.801	321	133	62	430	34
0,45	0,35	0,3	0,2	28	56	0.810	220	88	57	288	31
0,45	0,35	0,25	0,25	33	66	0.806	259	103	56	338	31
0,45	0,35	0,25	0,2	25	50	0.810	197	77	54	254	29
0,45	0,35	0,2	0,2	20	40	0.807	158	62	55	204	29
0,45	0,3	0,3	0,3	50	100	0.800	393	164	64	528	34
0,45	0,3	0,3	0,25	38	76	0.802	298	124	63	404	36
0,45	0,3	0,3	0,2	27	54	0.803	212	88	63	284	34
0,45	0,3	0,25	0,25	34	61	0.797	253	102	67	340	34
0,45	0,3	0,25	0,2	25	50	0.814	196	79	58	258	32
0,45	0,3	0,2	0,2	20	40	0.802	158	64	60	208	32
0,45	0,25	0,25	0,25	27	54	0.803	212	88	63	284	34
0,45	0,25	0,25	0,2	23	46	0.815	181	74	61	240	33
0,45	0,25	0,2	0,2	19	38	0.801	150	62	63	200	33
0,45	0,2	0,2	0,2	17	34	0.814	134	53	56	174	30

Increased ASN for Operationally Seamless Design



Outline

- Introduction and Taxonomy of Clinical Trial Designs
- Basic Principles of Adaptive designs
  - Allocation Rule
  - Sampling Rule
  - Stopping Rule
  - Decision Rule
- Phases of Development
- Adaptive Designs for the Learn Phase of Drug development
  - First-in Human / MTD
  - Two-Stage Designs
  - Adaptive Dose-Ranging designs
- Adaptive Designs for the Confirmatory Phase of Drug Development
  - Sample Size Re-Assessment
  - Adaptive Group Sequential designs
  - Seamless Phase II/III Designs
  - Population Enrichment Designs
- Practical Aspects of Adaptive Design Implementation
- Discussion

Patient Enrichment Designs

- Applicable where studies of unselected patients are unable to detect a drug effect and it seems necessary to "enrich" the study with potential responders (Temple, *Comm Stat Theory Meth 1994*).
- If this is done in an adaptive way (i.e., it is not clear upfront whether to use the selected population) we might use adaptive enrichment designs (Wang et al, 2009).
- Baseline characteristics that are used for patient selection are known as biomarkers, and often genetic.
- Proof of efficacy is done in a confirmatory sense. Hence, we use confirmatory adaptive designs that control prespecified Type I error rate.

340

Key Concepts

- Extension from the conventional single population design objective to an objective that encompasses several possible patient sub-populations
- Allow more informative evaluation in the patients having different degrees of responsiveness to the therapy
- Allow modification to study hypothesis, reallocating the patients and reestimation of the sample size midstream to achieve the pre-planned objective

The Enrichment Test Procedure

- For simplicity, we consider a two-sample comparison case although an extension to the multi-armed case is straightforward.
- Consider prespecified subpopulation(s) $S_1,\ldots,S_G$, which can be nested, and a full population F:

$$S_G \subset \ldots \subset S_1 \subset F$$

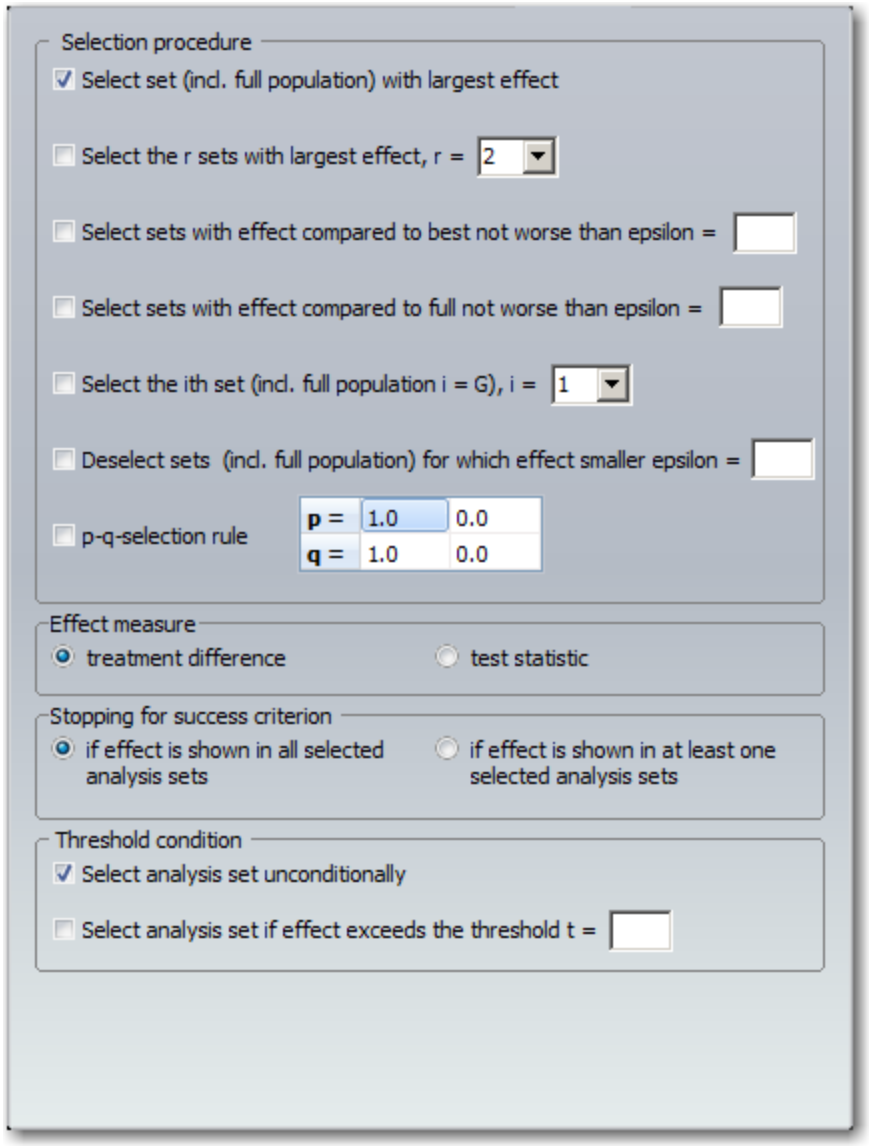
- At an interim stage it is decided which subpopulation is selected for further inference (including all subpopulations, i.e., full population).
- Not only selection procedures, but also other adaptive strategies (e.g., sample size reassessment) can be performed.

342

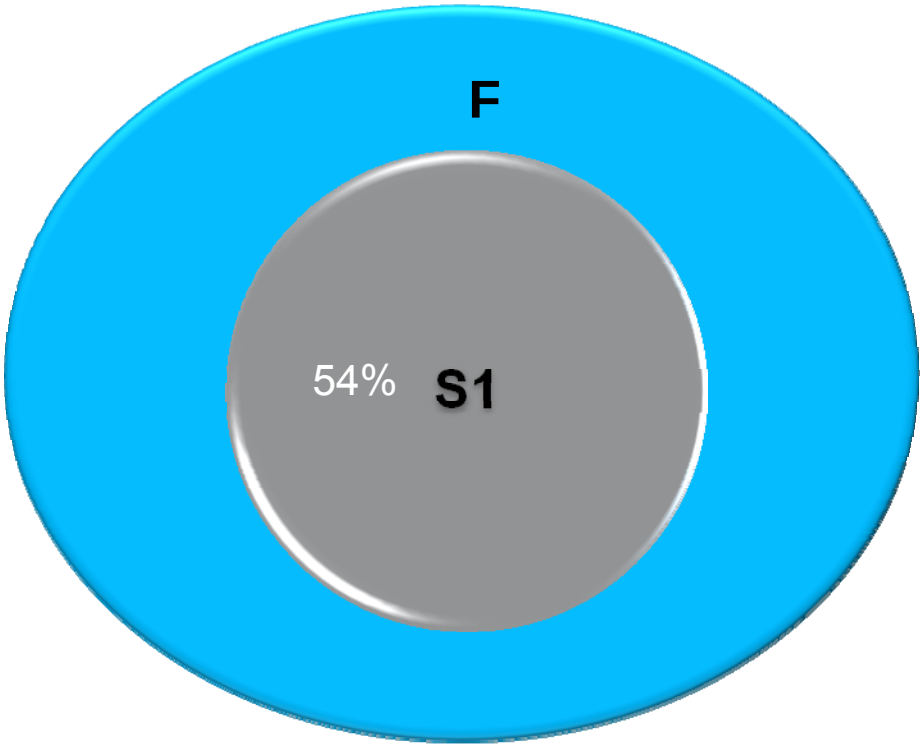
Test strategies

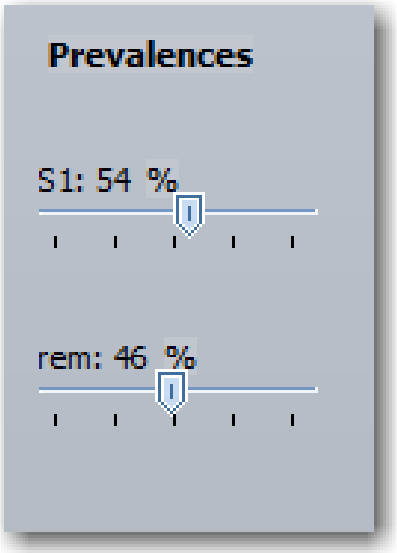
- Combination test:
  - Inverse normal method
  - Fisher's combination test
- Separate Phase II/III:
  - Phase II only for sub-population selection
  - Phase III is group sequential
- Intersection Tests:
  - Dunnett
  - Bonferroni
  - Sidak
  - Simes
  - Hierarchical

Selection Procedure

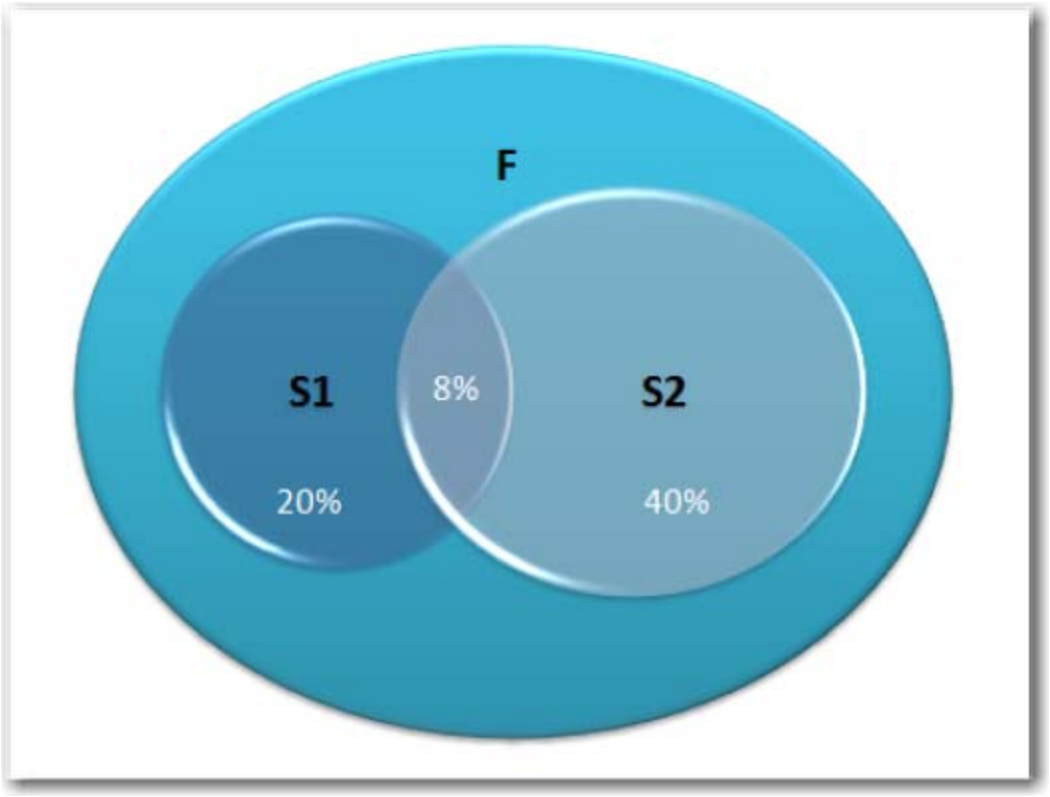


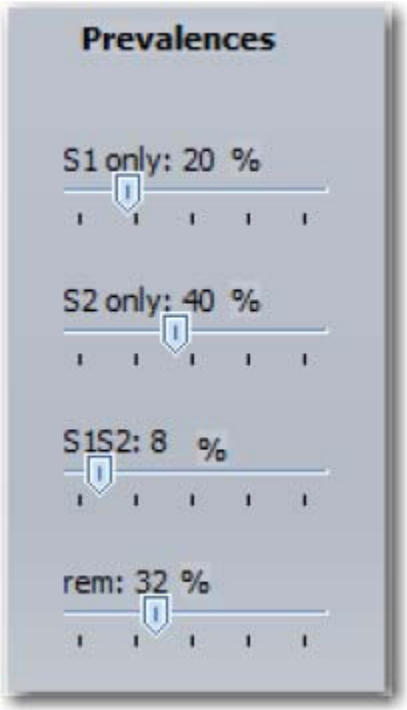
One sub-population





Two sub-populations of interest



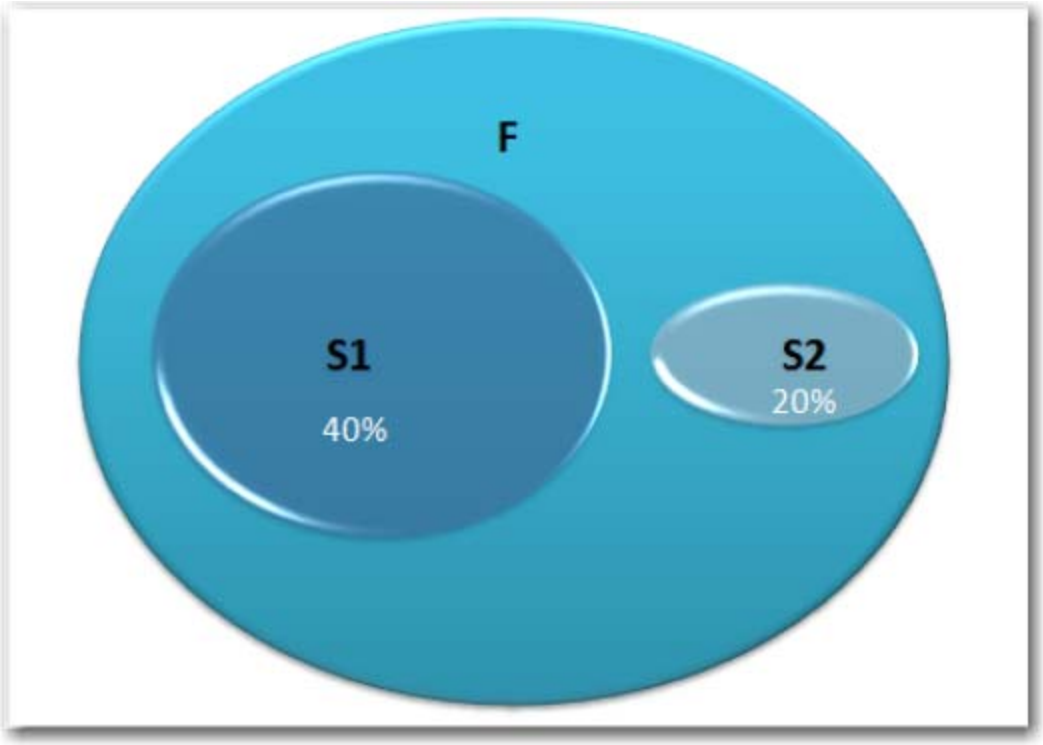


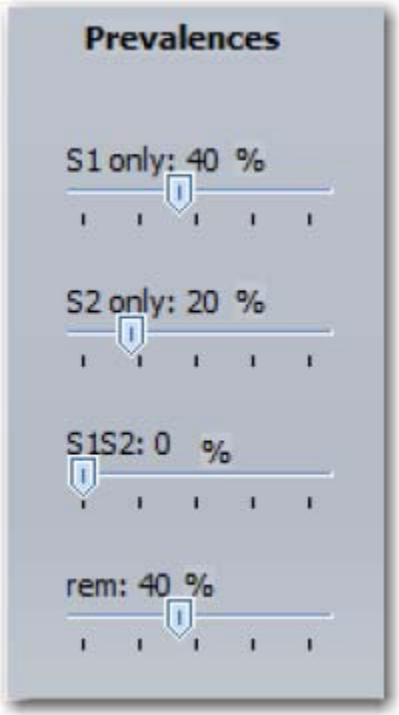
Three sub-populations of interest

Prevalences

S1 only =	20	%
S2 only =	40	%
S3 only =	0	%
S1S2 only =	0	%
S1S3 only =	0	%
S2S3 only =	0	%
S1S2S3 =	8	%
rem =	32	%

Two non-overlapping sub-populations





Sources for alpha Inflation

- Interim analyses
- Multiple hypotheses
- Sample size reassessment

The proposed adaptive procedure fulfils the regulatory requirements for the analysis of adaptive trials in that it strongly controls the prespecified (familywise) Type I error rate.

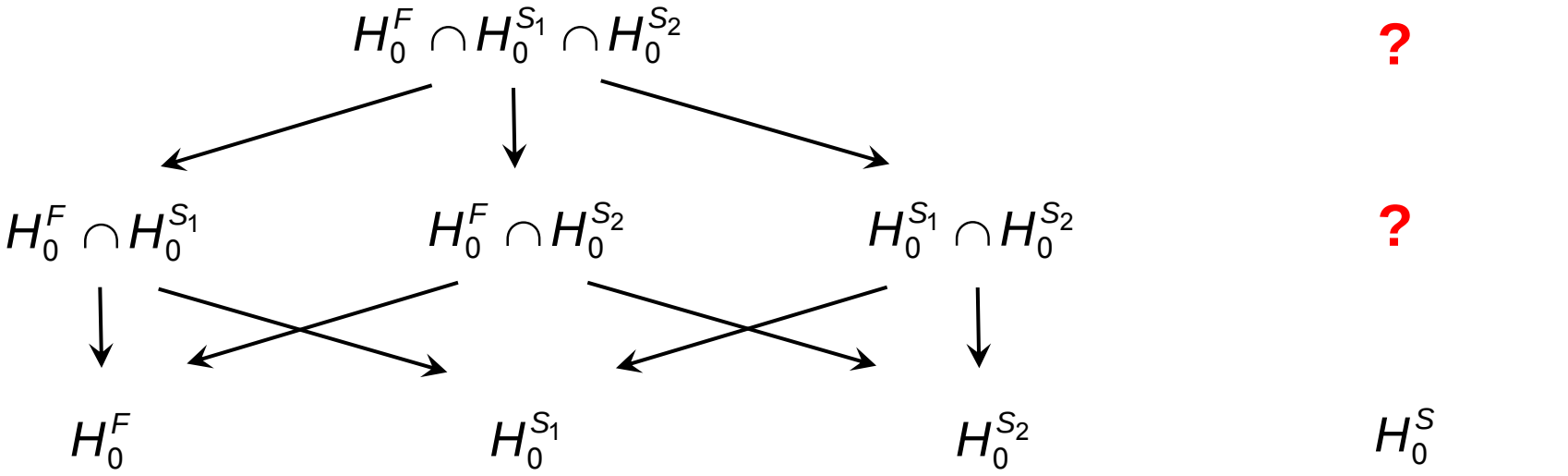
This procedure will be based on the application of the closed test procedure together with combination tests. For multi-armed designs this was proposed be several authors (e.g., Bauer & Kieser, 1999; Posch et al., 2005, Bretz et al., 2009, Wassmer, 2011).

Obviously, the same procedures as for multi-armed designs can be applied.

349

Closed testing procedure

Stage I **Stage II** **…**



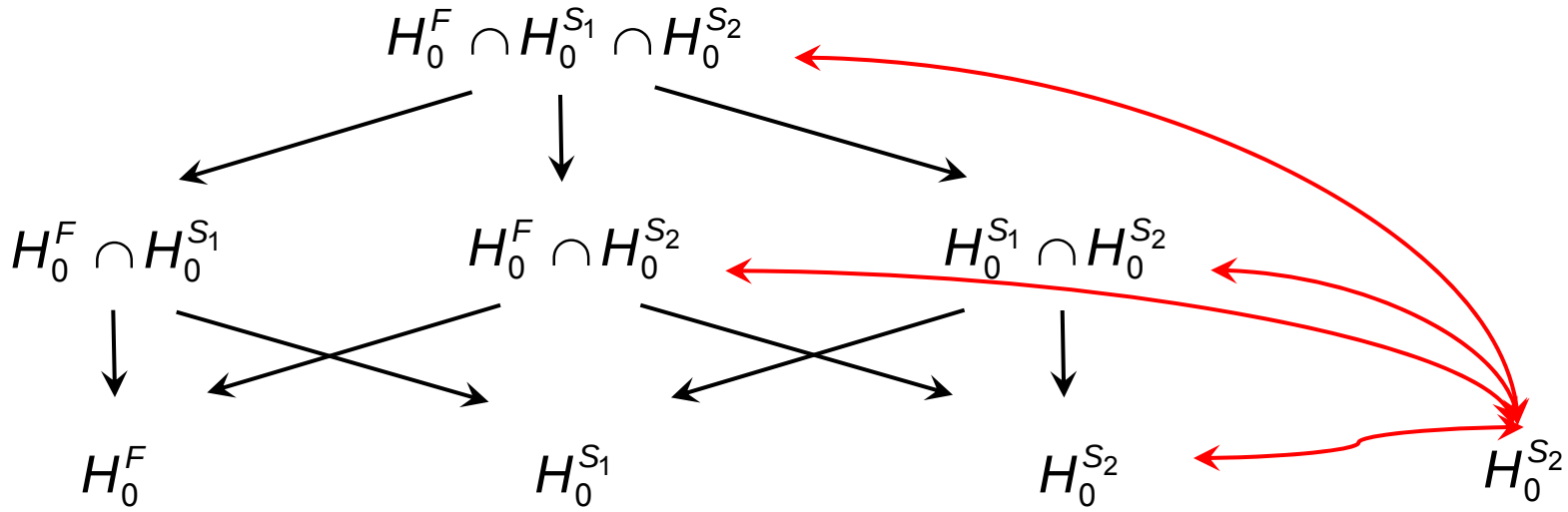
Simple “trick”: Test of intersection hypotheses are formally performed as tests for H_0^S.

H_0^S can be rejected if all combination tests exceed the critical value u_2.

Closed testing procedure

Example $S = S_2$

Stage I **Stage II** **…**



- $H_0^{S_2}$ can be rejected if all combination tests exceed the critical value u_2.
- The choice of tests for intersection hypotheses is free. You might use Bonferroni, Simes or Sidak tests.
- For one subgroup also Dunnett‘s test can be applied

351

Overall *p*-values

- Defined as smallest significance level for which the test results yield rejection of the considered (single) hypothesis
- Overall *p*-value can be calculated at any stage of the trial („Repeated *p*-value“).
- That is,

$$p_k^g \leq \alpha \quad \Leftrightarrow \quad H_0^g \text{ can be rejected at stage } k$$

- *p*-values account for the step-down nature of the closed testing principle and are completely consistent with the test decision.

352

Overall confidence intervals

- 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 confidence intervals based on the single step adjusted overall *p*-values. These can also be applied for the conditional Dunnett test.
- The RCIs are not, in general, consistent with the test decision. It might happen that, e.g., a hypothesis is rejected but the lower bound of the CI is smaller 0.
- They can be provided for each step of the trial.

353

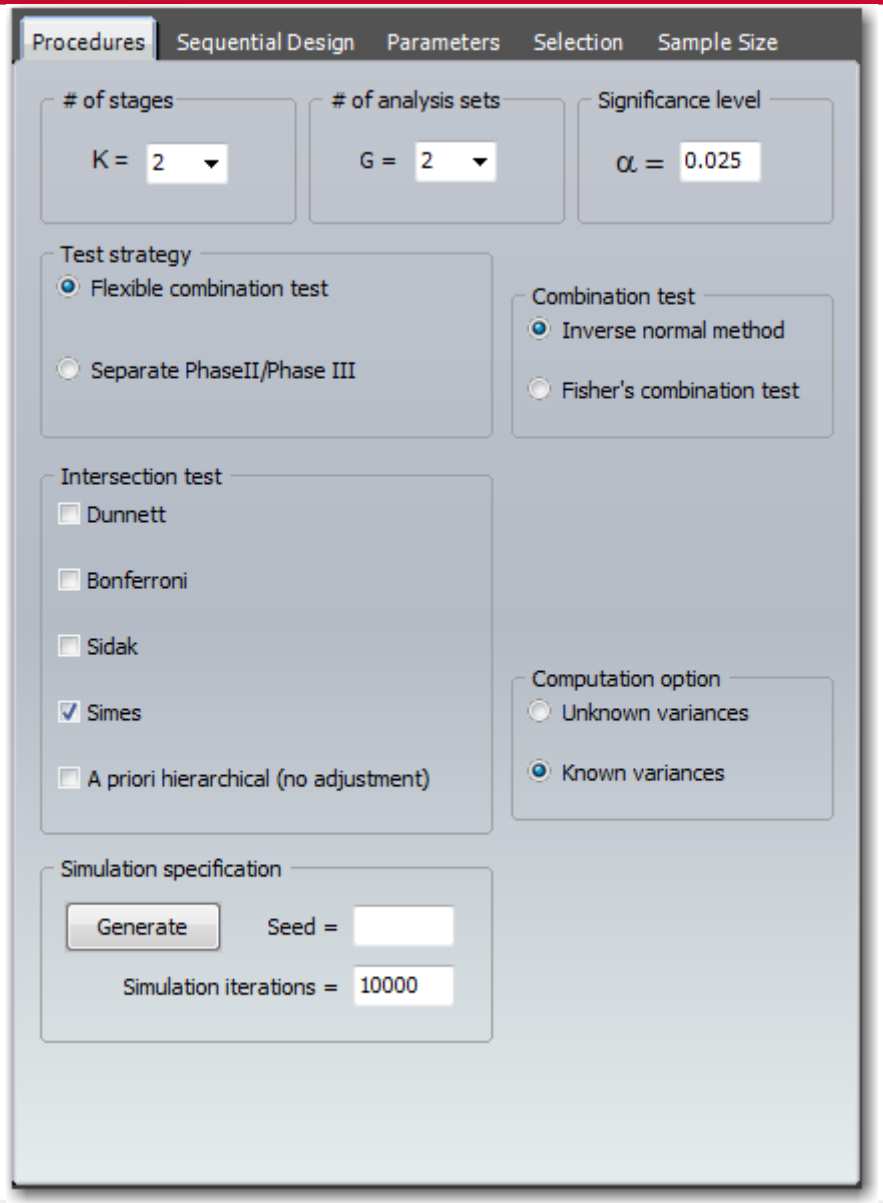
Patient Enrichment Designs: Simulation

Simulation Example

- Two-stage design with no early stopping, one sub-population
- In the biomarker positive population a standardized effect of 0.5 is assumed, biomarker negative population has effect sizes ranging from 0 to 0.5
- Selection rules
  - Select the population with highest effect size
  - Select the population with effect size compared to the better not worse than 0.25 (say)
  - Never select
- Prevalances of biomarker positive population is 5%, 10%, 20%.
- Sample sizes 100 patients per stage
- Simes‘ test is used for testing intersection hypotheses.

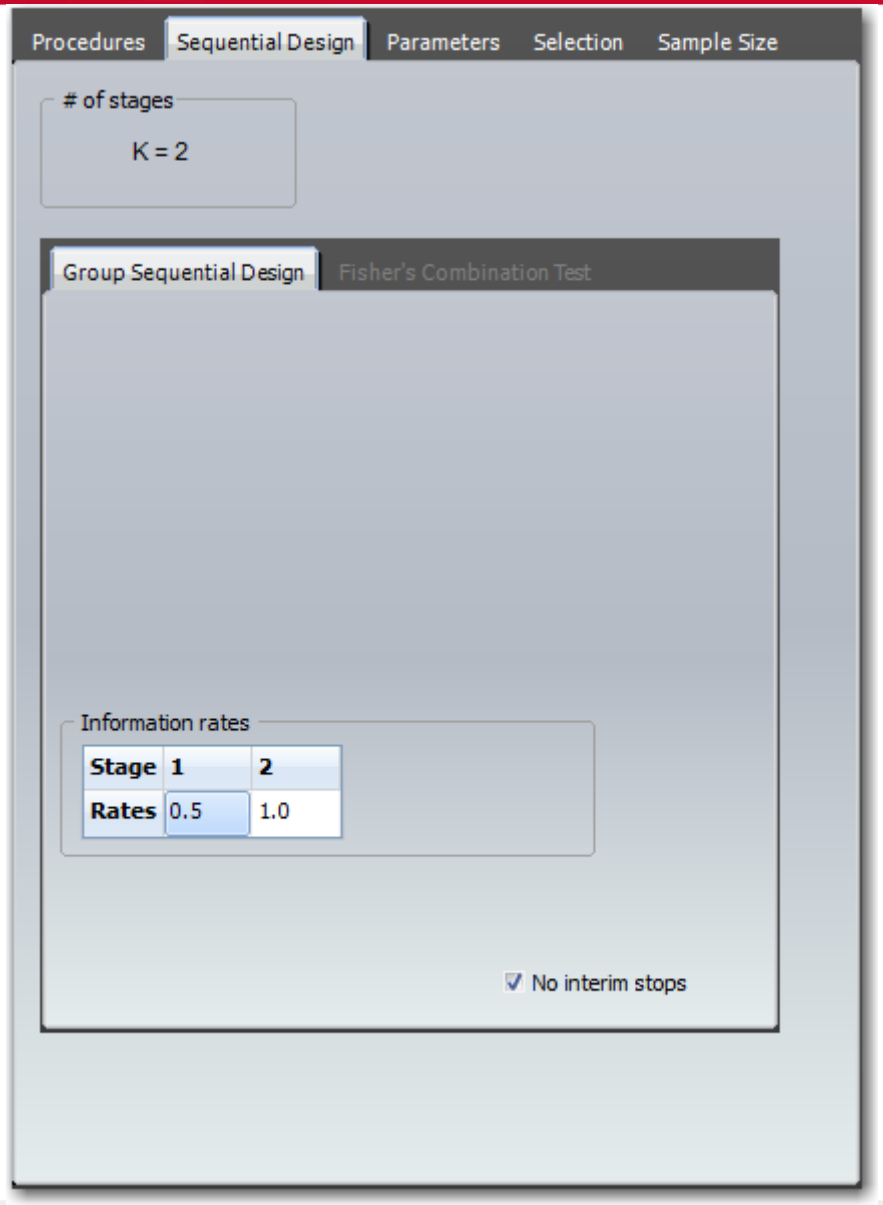
355

Specifications



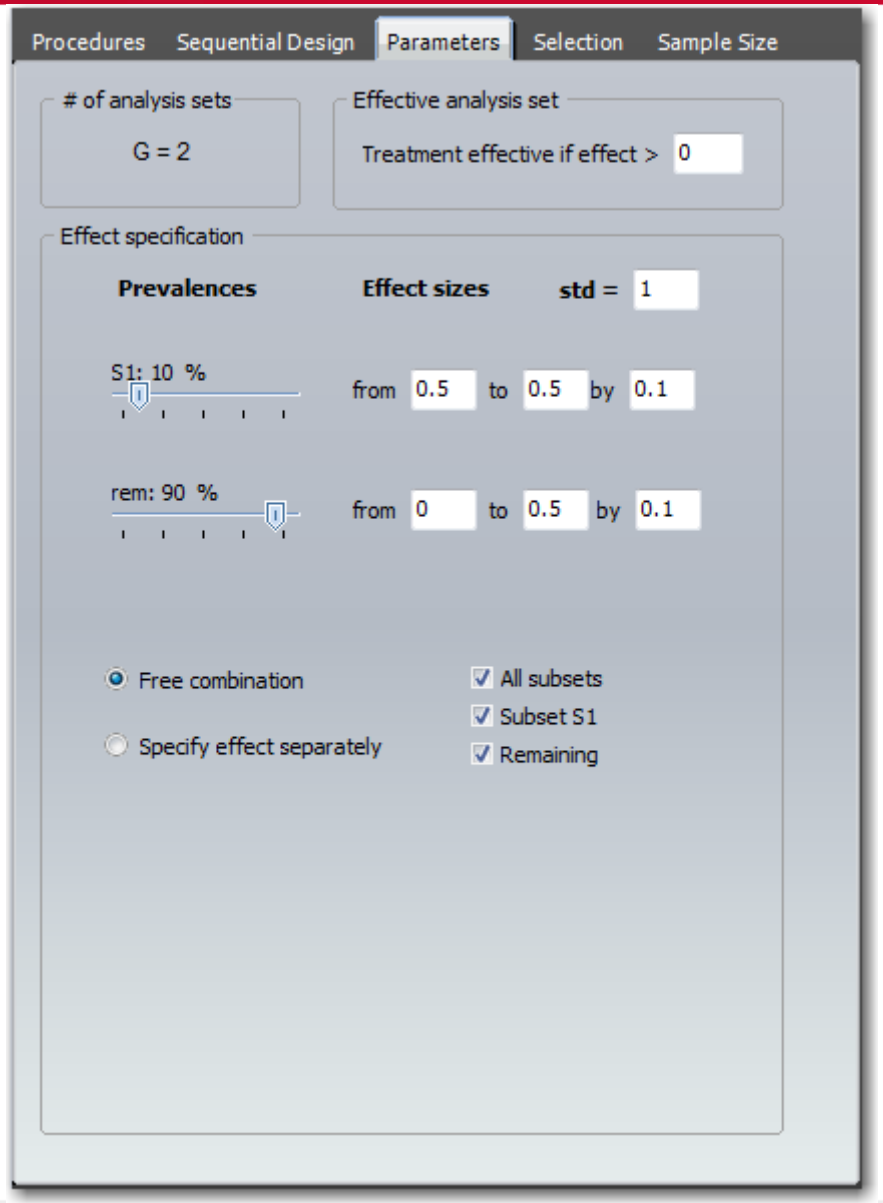
356

Specifications



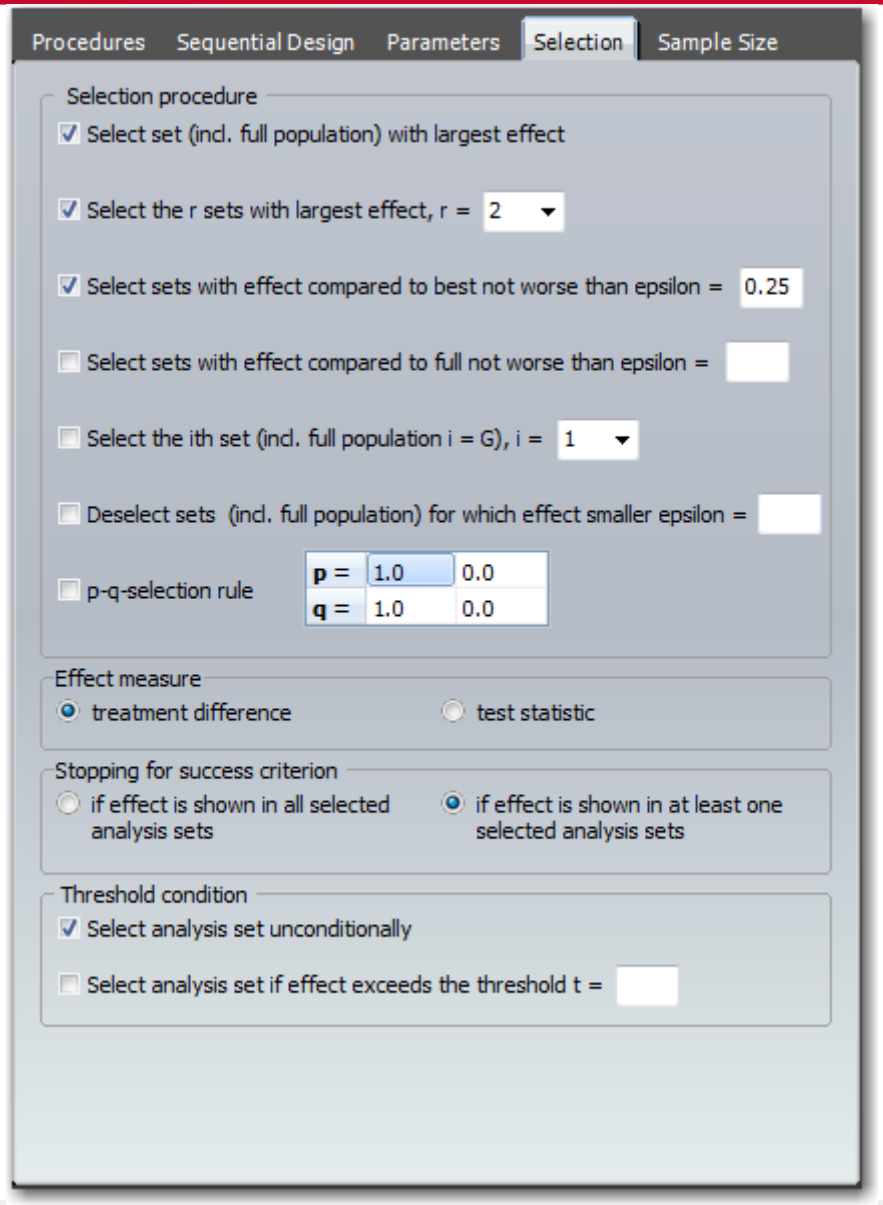
357

Specifications



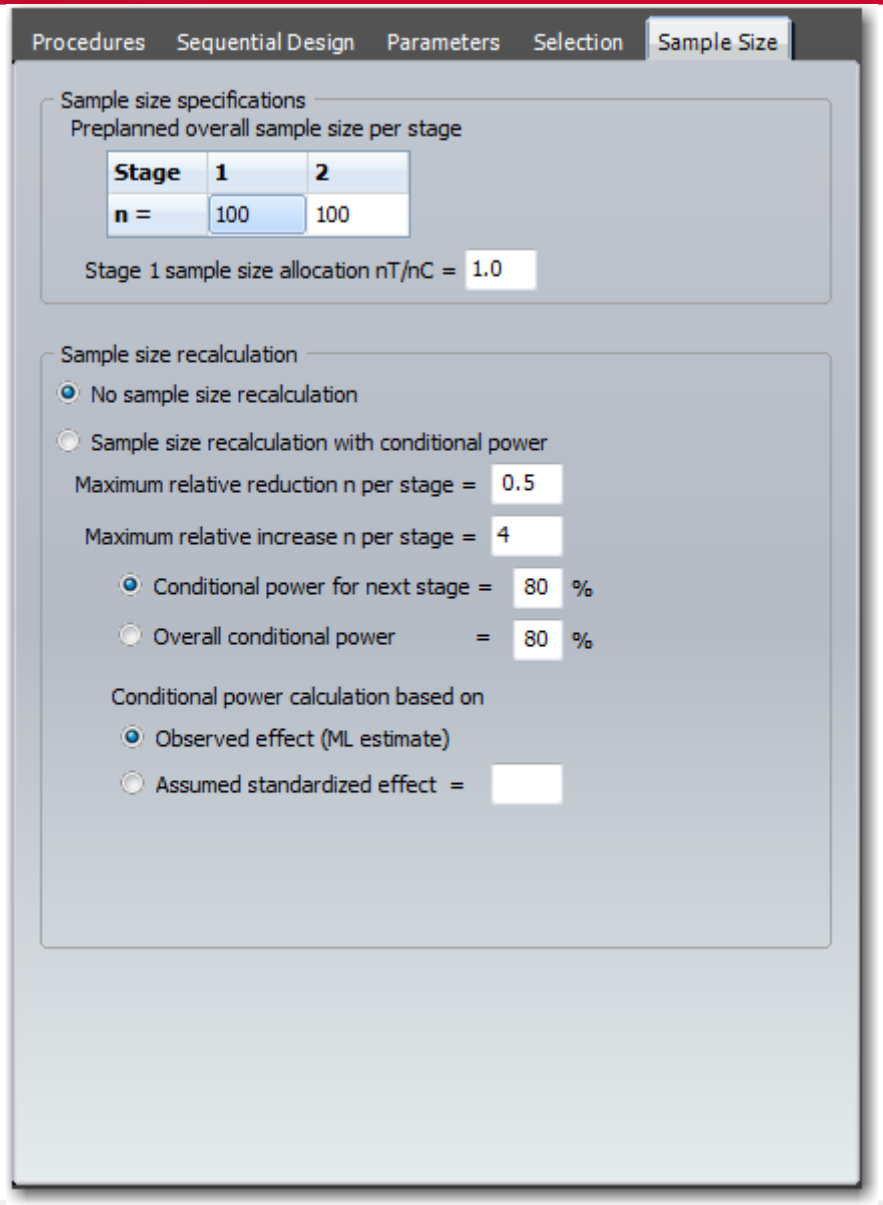
358

Specifications



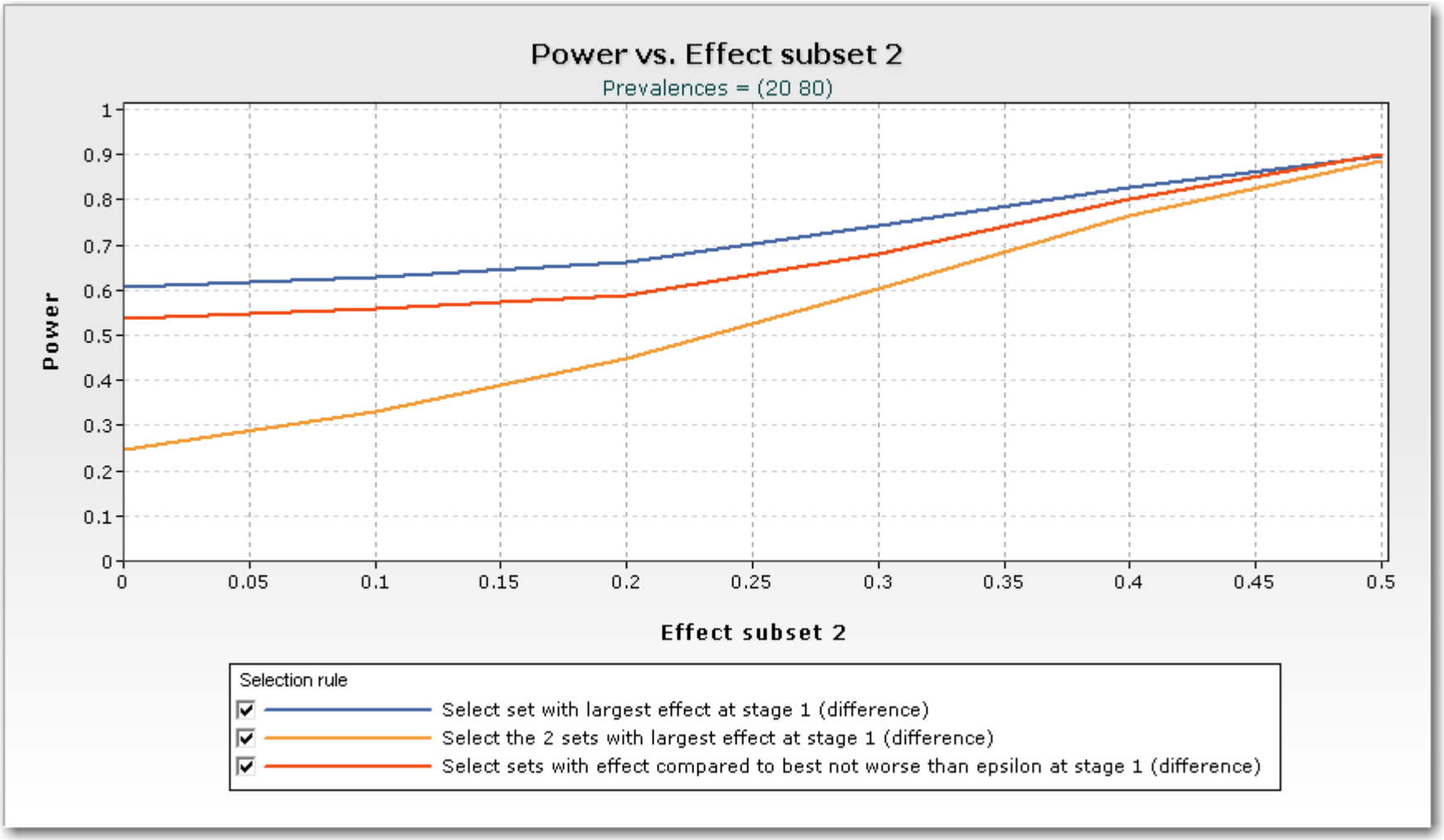
359

Specifications

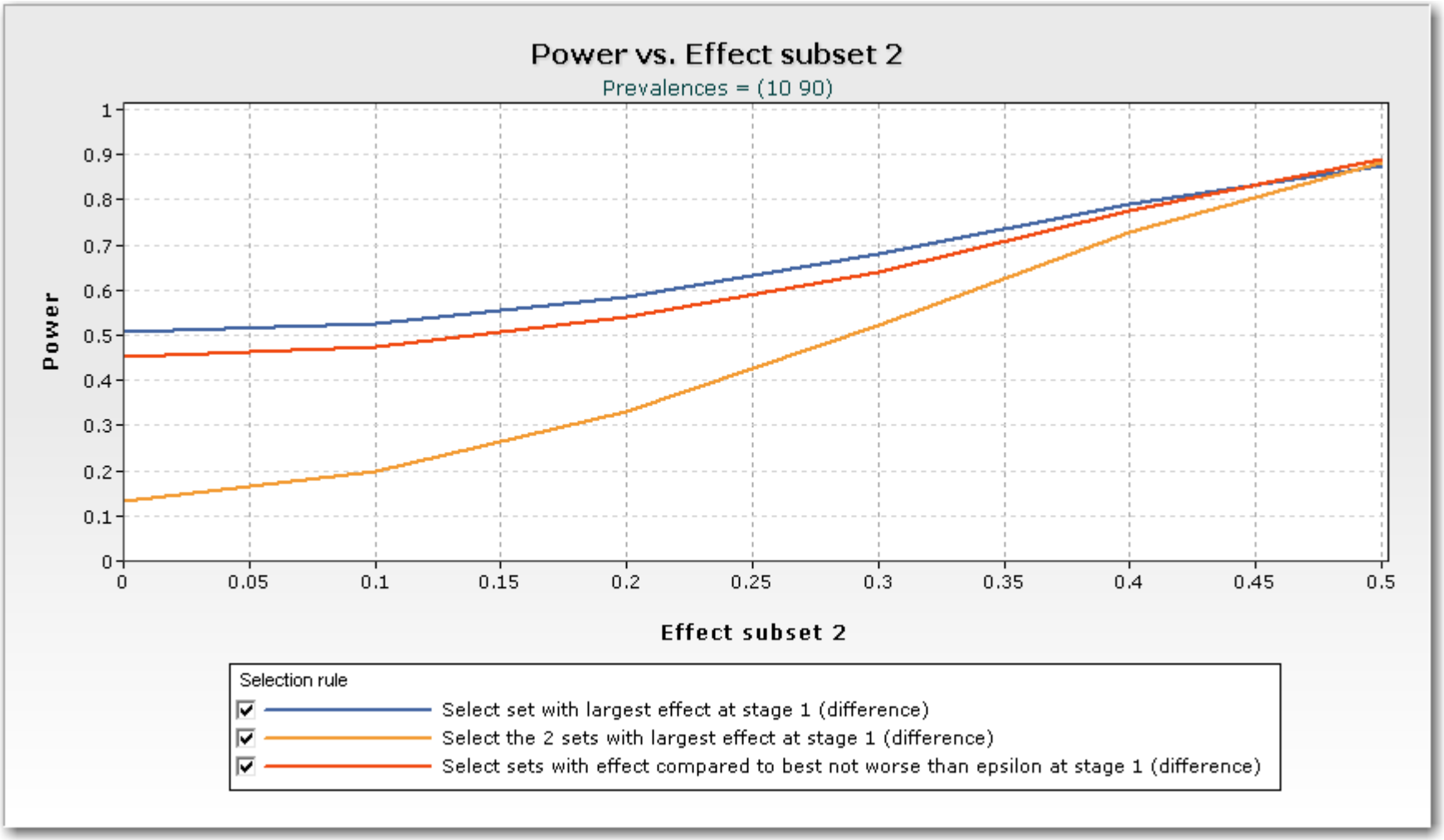


360

Power plot

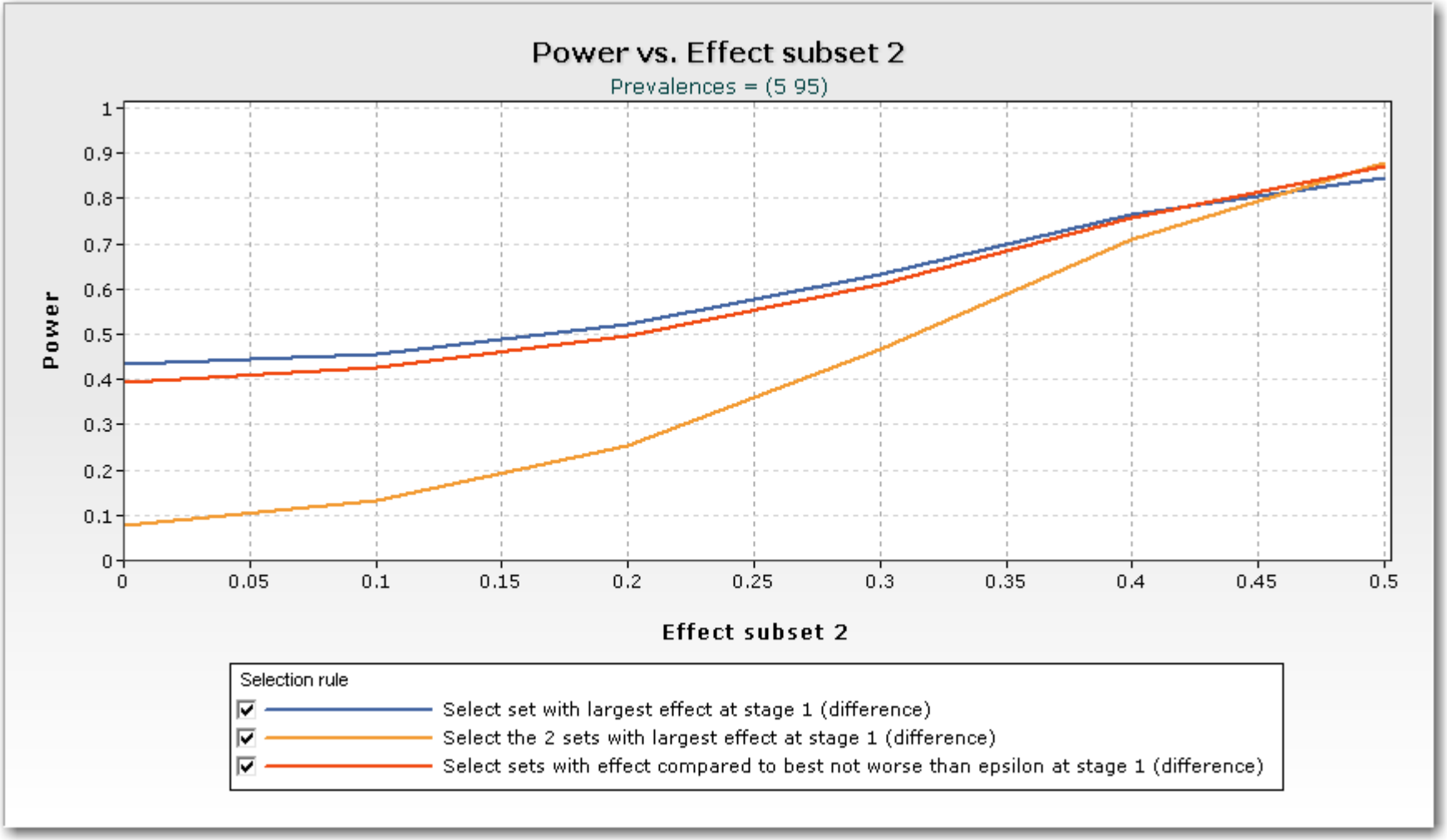


Power plot



362

Power plot



363

Results

- Clear power disadvantage for procedure that never selects a sub-population
- No clear advantage of selecting always (and only) the best population
- For small prevalences, always selecting the best can even provide a small loss in power

364

Conclusion

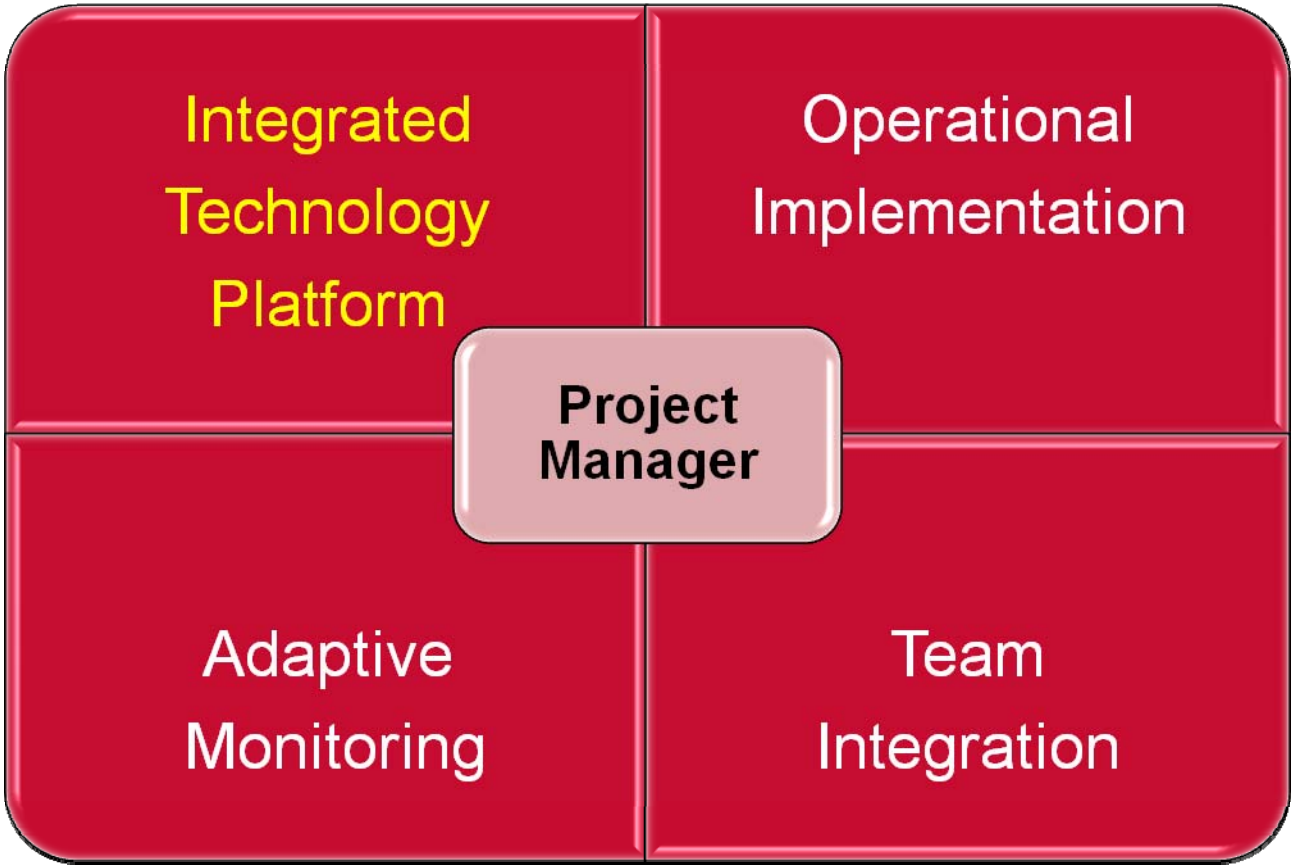
The main reason for enrichment is study efficiency

- increases the chance of success, often with a smaller sample size
- provides major benefits of individualization,
- directs treatment where it will do the most good
- spares potential harm for people who cannot respond

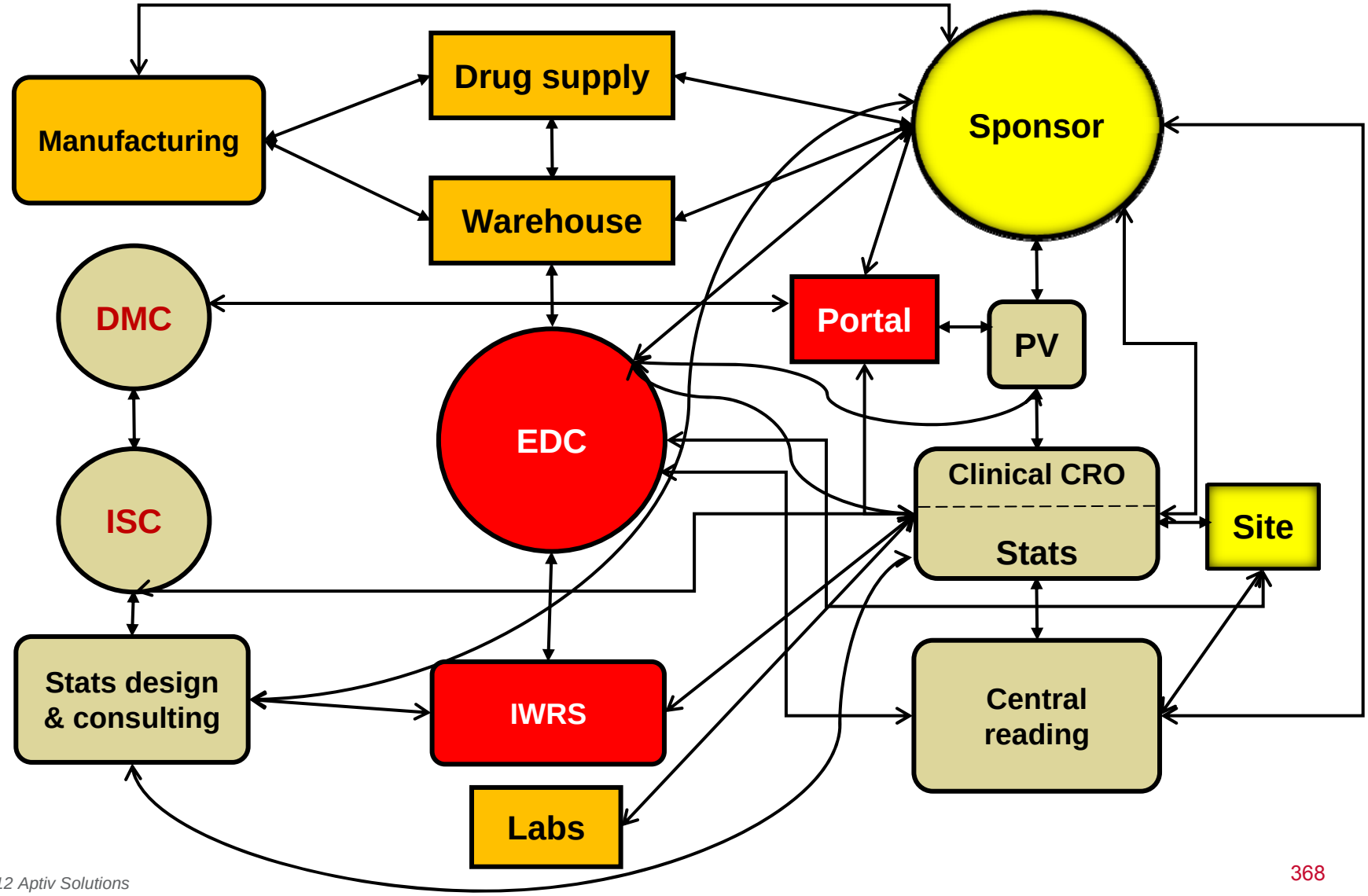
Outline

- Introduction and Taxonomy of Clinical Trial Designs
- Basic Principles of Adaptive designs
  - Allocation Rule
  - Sampling Rule
  - Stopping Rule
  - Decision Rule
- Phases of Development
- Adaptive Designs for the Learn Phase of Drug development
  - First-in Human / MTD
  - Two-Stage Designs
  - Adaptive Dose-Ranging designs
- Adaptive Designs for the Confirmatory Phase of Drug Development
  - Sample Size Re-Assessment
  - Adaptive Group Sequential designs
  - Seamless Phase II/III Designs
  - Population Enrichment Designs
- Practical Aspects of Adaptive Design Implementation
- Discussion

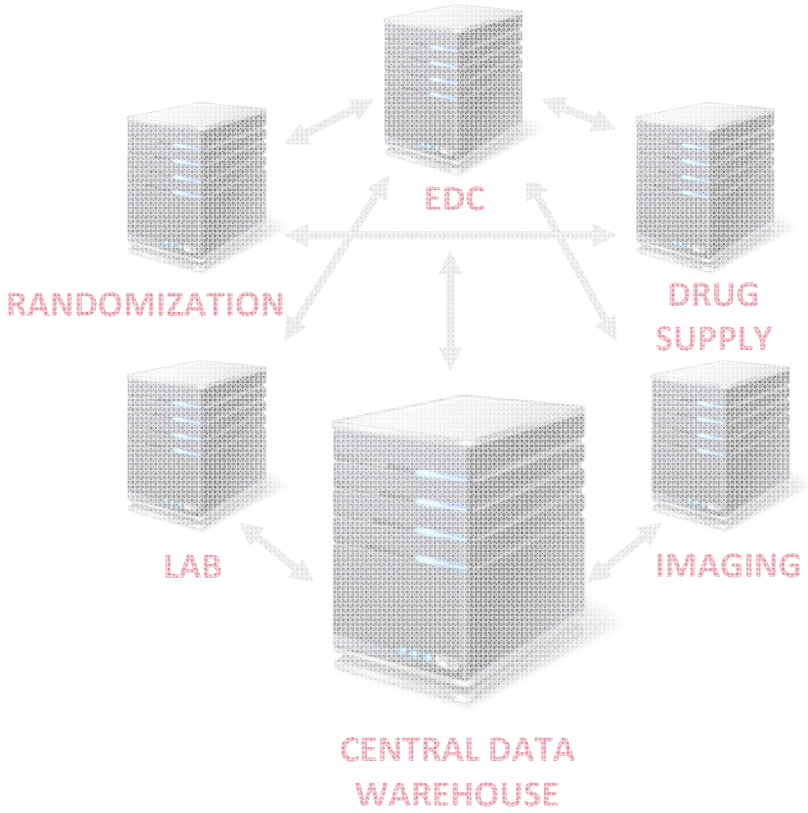
Successful Adaptive Clinical Trial Execution



The Logistics of Implementation are very complex



Integrated Technology Solution

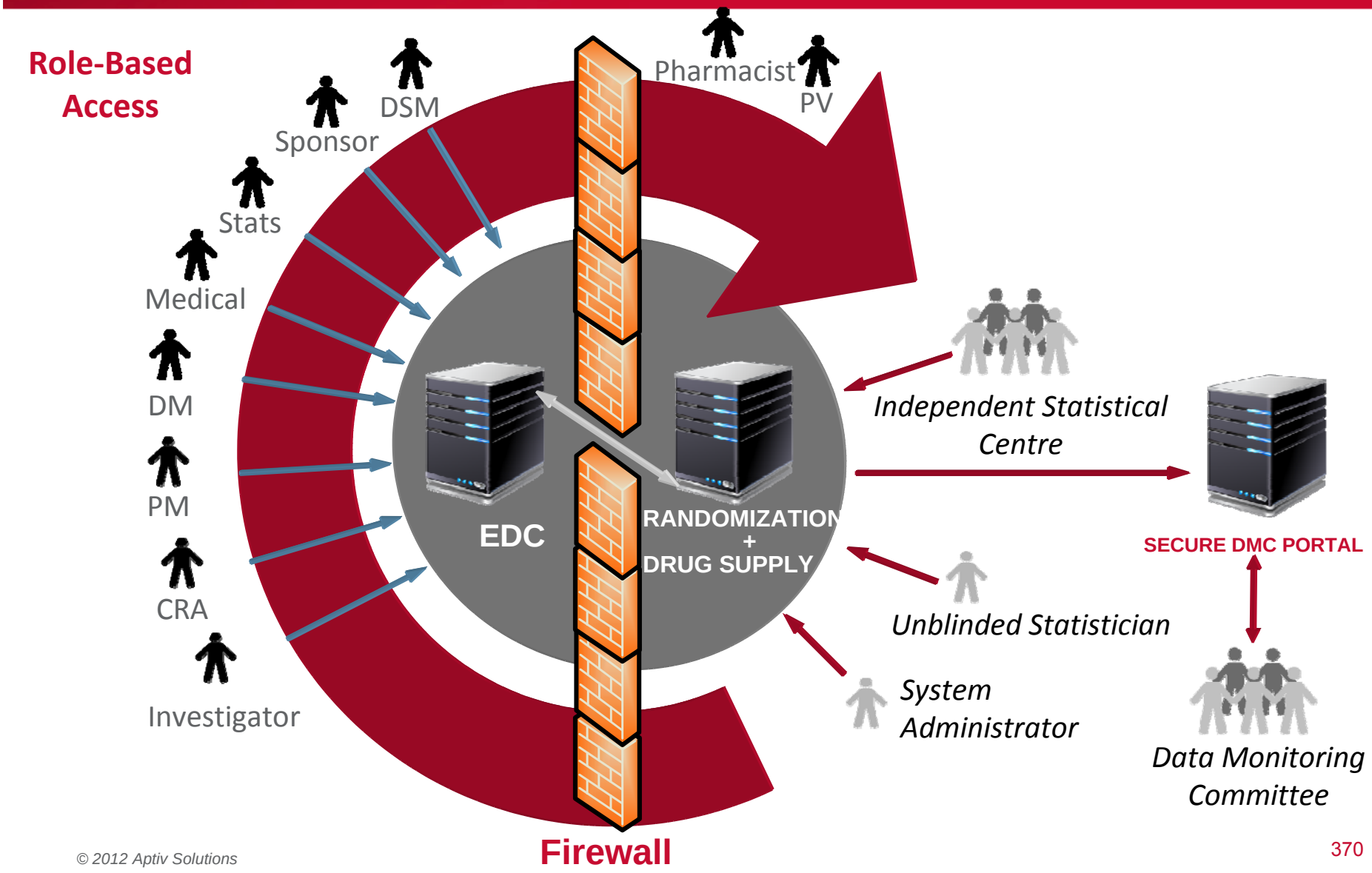


Integrated EDC, Drug Supply Management, Randomization and Design Engine for ACTs

Traditional Clinical Data Warehouse: Customized data export and import between separate entities

Integrated Technology: Minimization of customized or manual data transfer in a changing environment

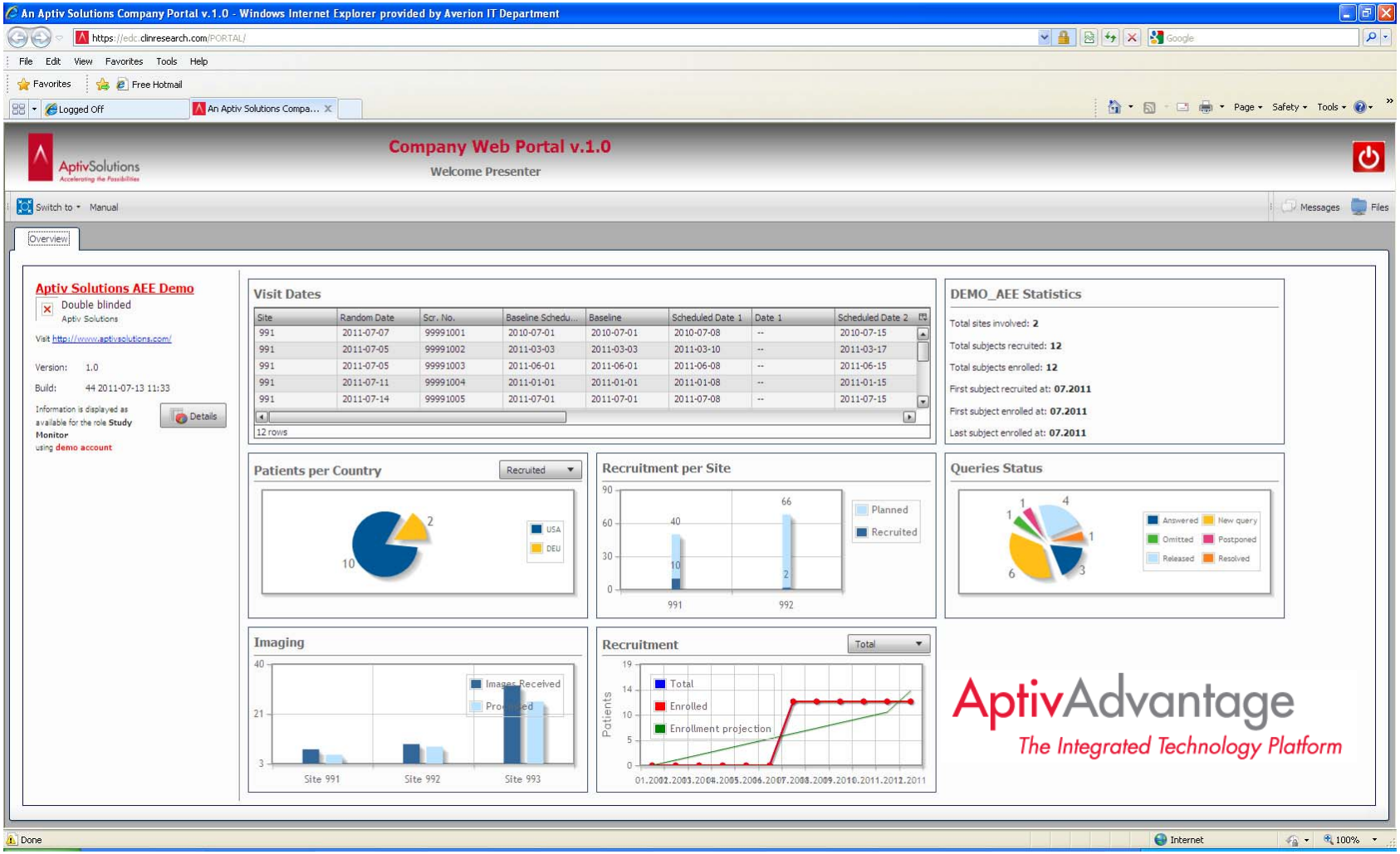
Integrated Technology Platform

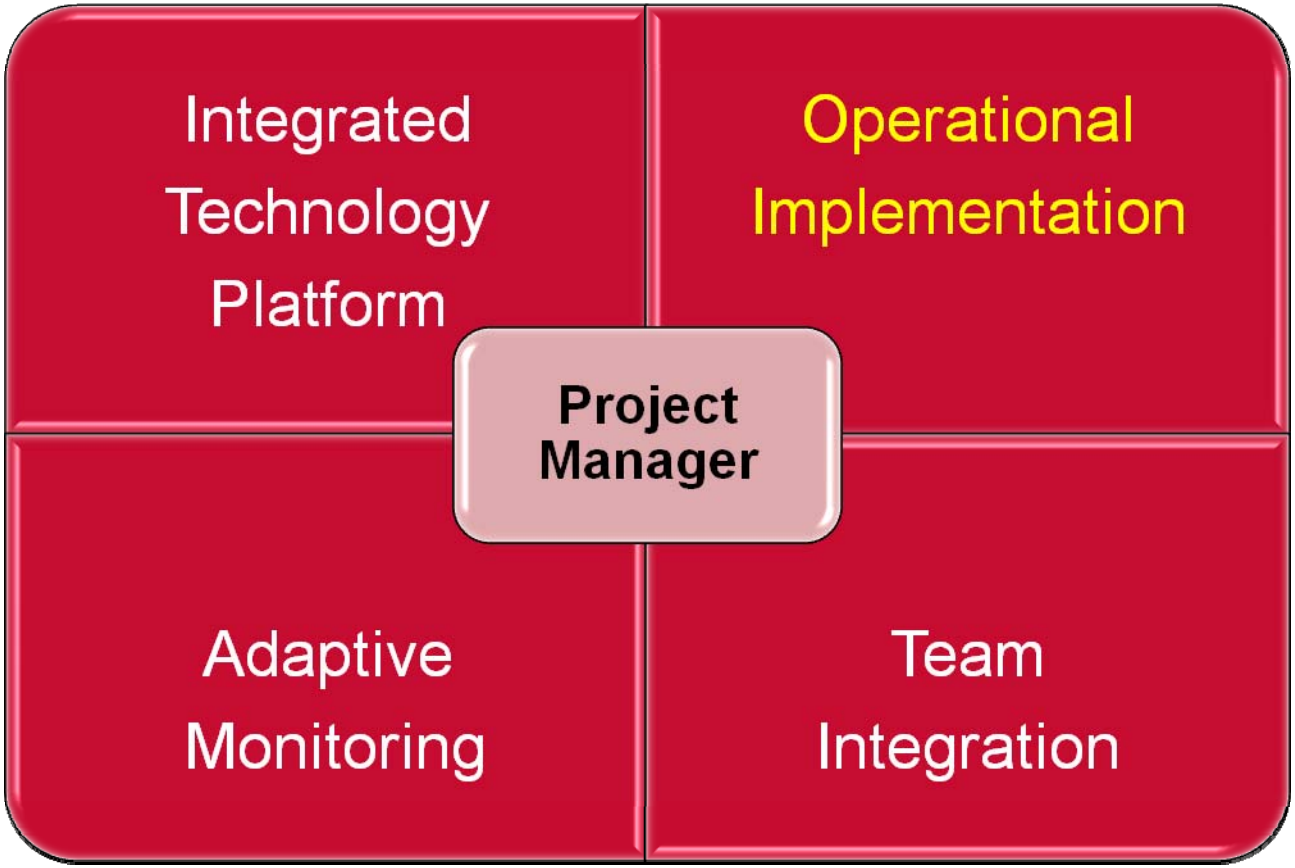


Clinical Data Management System

- In addition to data capture AptivAdvantage™ offers:
  - Online and offline data checks
  - Online reporting and dashboard
  - Direct data access for interim analysis by the stats team
  - Direct access for medical data reviewer during the course of the trial → in-stream data cleaning [interim readiness, medical plausibility, early safety signaling]
    - Eliminates classical sequential batch process by implementation of parallel processing
    - Fulfills needs of adaptive trials by integration of processes and systems
  - Database lock to adaptation decision – < 1 week

AptivAdvantage™ - Workflow Management




Integrated Technology Platform
Operational Implementation
Project Manager
Adaptive Monitoring
Team Integration

Adaptive Trials Require New Operational Processes

- *Adaptive SOPs and working procedures*
- Advanced project management
- New roles & process for data cleaning
- Interim analysis ready at all times
- *Investigator payments linked to data entry*
- *Data-driven monitoring*
- Drug supply management
- *DMC Management*
- Operating procedures to control access to unblinded data
- Firewalls and secure DMC portal

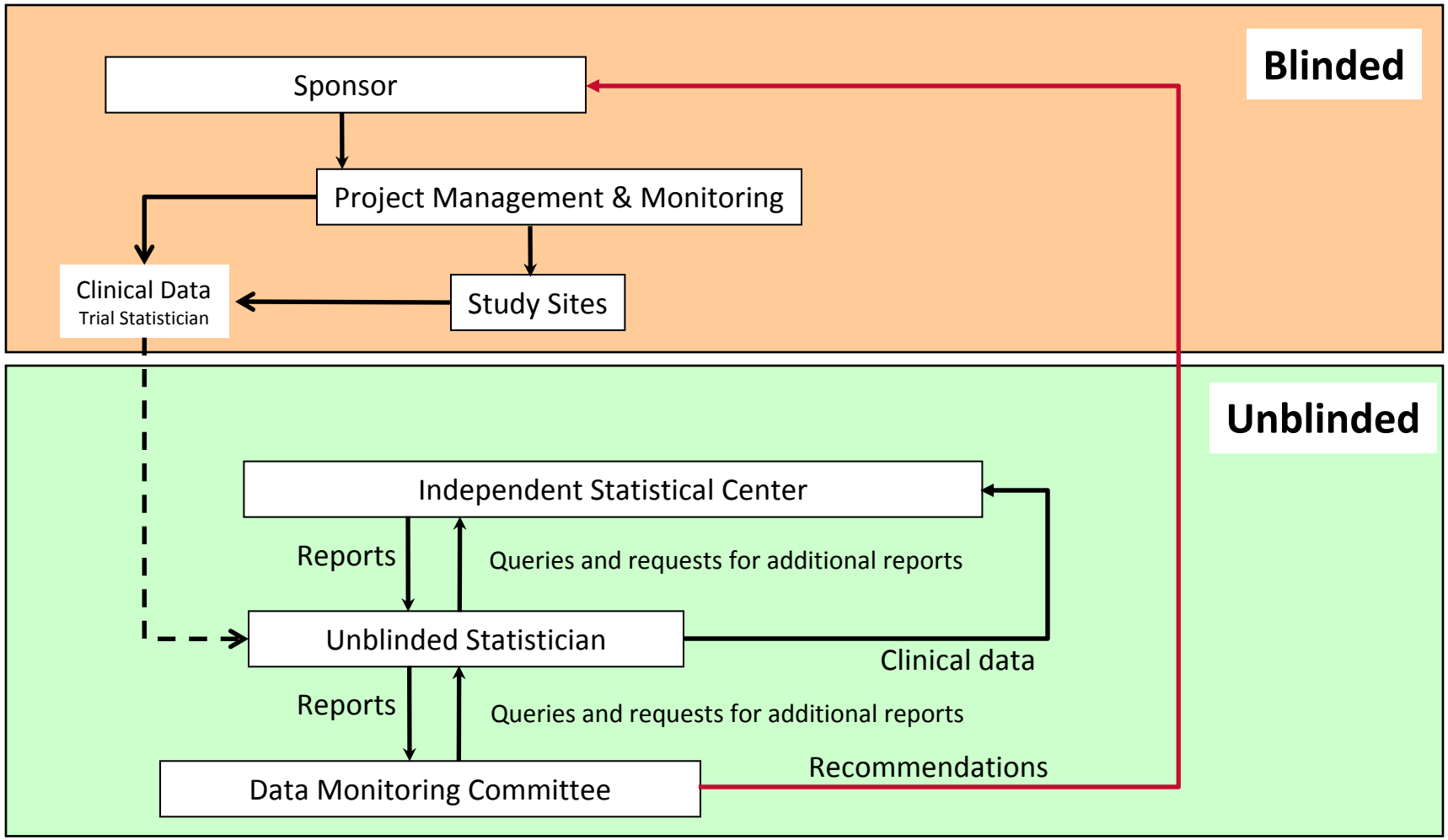
Operational Implementation

- SOPs developed to ensure study integrity
  - All applicable functional areas
  - Planning – details of communication and escalation pathways - “communication firewalls”
  - Monitoring – define and outline parameters for data-driven monitoring and remote monitoring
  - Adaptive execution checklist – serves as guide for implementation
  - SAP/Interim Analysis Plans - integration and timing
  - DMC - firewalls and charter management

Operational Implementation

- Processes and work practices in place supporting adaptive trials
  - Clinical Operations policy
- Oversight Committees
  - Executive Committee comprised of executive management, Innovation Center, implementation/execution experts, and KOLs as needed
  - Study Integrity Teams dependent upon size and scope

Effective Management of the DMC is Critical

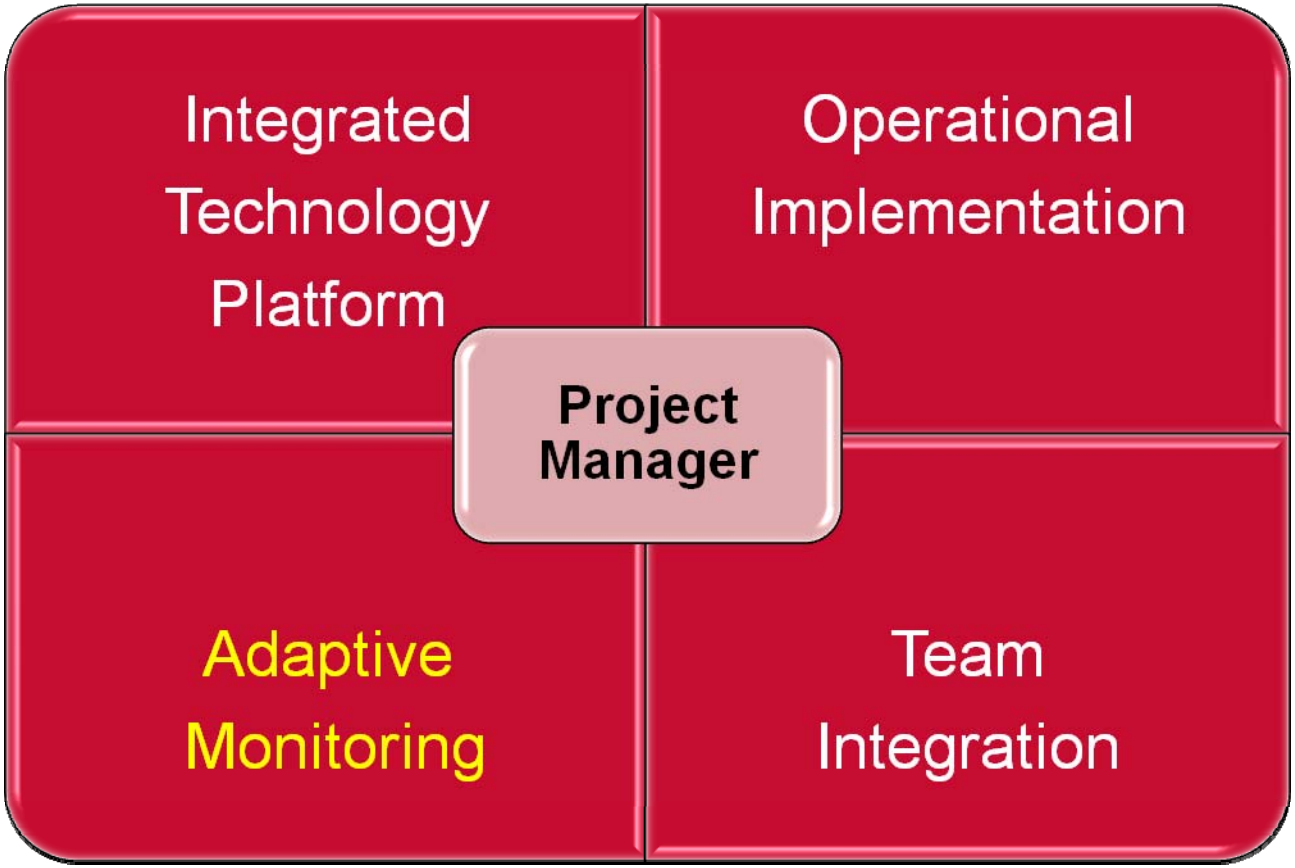


Specific requirements

- While performing adaptive clinical trials **key concern of agencies** is the protection against any issue that could influence the trial (operational bias).
- The key principle to adhere to is **preserving the blind** by:
  - Shield Investigators and other study stakeholders from knowledge of adaptive procedures (interim analyses and the following adaptations)
    - Investigators may wait to include patients to increase the patients' chance for a “better” treatment.
    - Assure consistency between different stages of the study
    - Although there will be an interim analysis, the trial should continue smoothly without any change/stop/break.
  - Documented procedures and reliable technology for managing “who sees what and when” (SOPs)

DMC, Data & Sponsor

- At the completion of each interim analysis the Sponsor will receive recommendations regarding continuation or termination of the trial or modifications to study procedures from the DMC
  - The members of the DMC must be independent of the Sponsor
  - The members of the DMC must not be involved in the study conduct and may not disclose any confidential information to people involved in the conduct of the study
- No unblinded data will be provided
- The Sponsor reviews DMC recommendations and makes a final decision


Integrated Technology Platform
Operational Implementation
Project Manager
Adaptive Monitoring
Team Integration

Adaptive Monitoring

- Adaptive Monitoring replaces
  - **fixed** schedules
  - **rigid** adherence
  - **predefined** plans
- Adaptive Monitoring combines
  - **remote monitoring**
  - **data-driven site visits**
- Adaptive Monitoring optimizes site performance

Adaptive Monitoring (cont'd)

- Adaptive designs = pro-active monitoring
  - Continuous review of real time study data using AptivAdvantage™
  - Adapt frequency and purpose of monitoring visits to the observations
  - Close communication with sites
    - Motivation: Prompt CRF completion/query resolution
    - Good relations: Arrange monitoring visits at short notice
  - CRAs should be unaware of the timing of an interim analyses

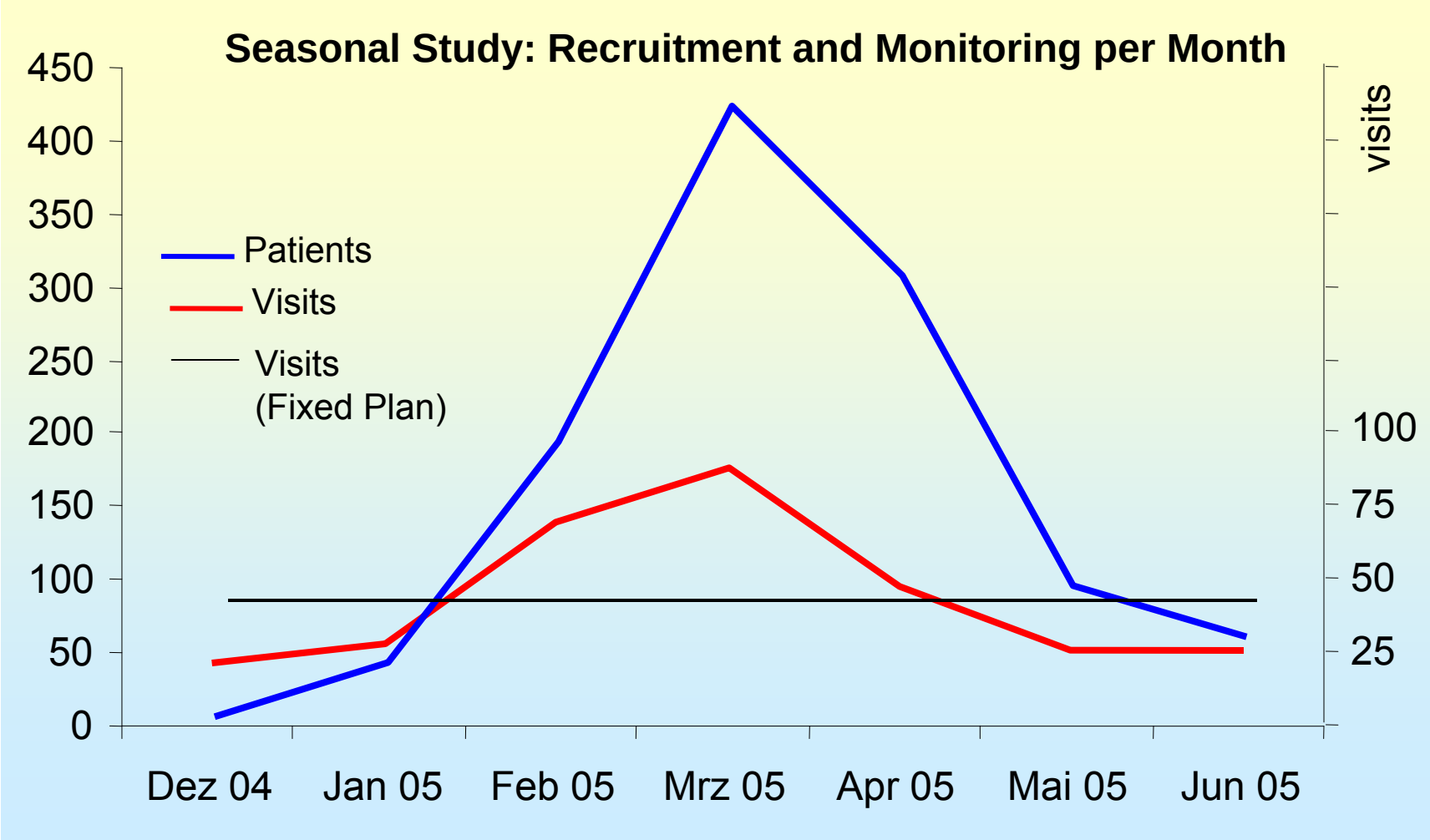
Adaptive Monitoring (cont'd)

- Data-driven monitoring
  - Flexible timing for on-site visits
  - Develop triggered approach driven by site enrollment, data entry and site performance
  - Ensures clean data for all interim analyses
    - Ongoing Query Resolution
    - Continuous Data Cleaning
    - Interim analysis ready at all times

Adaptive Monitoring (cont'd)

- Remote monitoring
  - AptivAdvantage™ to review data trends, medical review feedback, endpoint data, data turn around time, safety trends
  - Site performance and forecasting: monitor recruitment and review drug supply
  - Identify site trends and corrective actions
  - CRAs work close together with DM and in-stream medical reviewers
  - Early intervention in case of systematic errors

Adaptive Monitoring (cont'd)



AptivSolutions℠
Accelerating the Possibilities

US and EU Regulations and Guidance for Adaptive Study Design

US/EU- Changing Regulatory Environment

- US – Critical Path Initiative
- Advancing Innovative Trial Designs
  - Design of Active Controlled Trials
  - Enrichment Designs
  - Use of Prior Experience or Accumulated Information in Trial Design
  - Development of Best Practices for Handling Missing Data
  - Development of Trial Protocols for Specific Therapeutic Areas
  - Analysis of Multiple Endpoints
- Improving Measurement of Patient Responses
- Streamlining the Clinical Trials Process

- EU Innovative Medicines Initiative

- The Roadmap to 2015
- Adopted by EMA in December 2010
- Strategy to drive innovation in the scientific processes through which medical products are developed, evaluated, and manufactured
  - Adaption of existing model for medicines regulation to enable integration of new and emerging science

Regulators Motivations for Accepting Adaptive Design

- Compared to non-adaptive, adaptive design studies may:
  - More efficiently provide the same information
    - Shorter duration
    - Fewer patients
  - Increases the likelihood of success on the study objective
    - More likely to demonstrate effect of drug, if one exists
  - Yields improved understanding of the treatment's effect
    - Broader and better dose-response relationship
    - Sub-group effects
    - Provides foundation for more efficient subsequent studies

Current Regulatory Environment

- Regulatory authorities in EU and US published "Guidance Documents" on use & implementation of adaptive designed trials
  - *EMA Reflection Paper (2007)*

    A study design is called "adaptive" if statistical methodology allows the modification of a design element (e.g., sample size, randomization ratio, number of treatment arms) at an interim analysis with full control of the type 1 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 from subjects in the study" (line 67)

EMA Reflection Paper

- Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design (CHMP/EWP/2459/02-effective date Oct 2007)
- ***Adaptive*** = statistical methodology allows modification of a design element (e.g. sample-size, randomisation ratio, number of treatment arms) at an interim analysis with full control of the type 1 error
- ***Confirmatory*** = adequately controlled with hypotheses stated in advance and evaluated. Confirmatory trials are necessary to provide firm evidence of safety and efficacy.

Summary of EU Guideline

- Purpose of Phase III study is to *confirm findings* from previous studies (CHMP/EWP/2330/99) thus design modification is a contradiction of the confirmatory nature of such studies
  - Adaptive designs best utilised as a tool for planning clinical trials in difficult experimental situations
- For adaptive trials need to describe/plan and justify in protocol
  - Interim Analysis
  - Measures for maintaining confidentiality of interim results
  - Type of anticipated design modification
  - Available statistical methods to control Type I error, correct estimates and confidence intervals for treatment effects
  - Methods for assessment of homogeneity of results from different stages
  - Advantage of adaptive over non-adaptive design

EU Regulatory Environment

- EMA workshops 'Adaptive Design in Confirmatory Clinical Trials' - December 2007 & April 2009
- Workshop Conclusions:
  - Appropriate use of adaptive design to be encouraged e.g. confirmatory studies
  - Adaptive design may not be acceptable for a single pivotal seamless Phase II/III study to support an MAA
  - **Sponsor involvement in interim analysis is discouraged**
  - EMA planning to create a Biostatistics ad hoc group to deal with adaptive design issues and missing data
  - Evolving area, ongoing discussion is recommended
  - EMA strongly advise scientific sdvice meetings to specifically discuss adaptive design studies

EMA – Workshop Conclusions cont.

- Adaptive designs as single pivotal study
  - Where only one study is possible?
  - Where basis for regulatory decision is improved
- Sponsor involvement
  - Strictly limited and controlled involvement has been cautiously accepted, in particular if not single pivotal study
- Change to primary endpoint based on internal information discouraged
- Maintenance of homogenous trial / assessment of heterogeneity is key

EMA – Moving Forward

- Promote use of adaptive designs where appropriate:
  - Difficult experimental situations
  - Confirmatory studies where efficiency gains do not compromise basis for regulatory decision or present unacceptable ‘risk’ to trial / trial programme
- Expanding experience and emerging regulatory preferences to be communicated
  - Recent guideline evaluation of products for bacterial infections
- Further consideration to be given to unresolved issues
- Recent Workshop ‘Subgroup Analysis’ 18 November 2011 – Adaptive trials discussed

FDA Guidance Document On Adaptive Designs

- FDA Guidance from “Adaptive Design Clinical Trials for Drugs and Biologics”, FDA, February 2010 (FDA):
  - “Comprehensive and prospectively written SOPs that define who will implement interim analysis and adaptation plan” (FDA line 1685)
  - “Many CROs do not have long histories of carrying out these responsibilities. Study sponsors should have assurance that the personnel performing these roles have appropriate expertise…” (FDA line 1725)

US Guidance -Adaptive Design Clinical Trials for Drugs and Biologics

- What aspects of adaptive design trials call for special consideration
- When to interact with FDA
- What information to include in adaptive design for FDA review
- Issues to consider in evaluation of a completed adaptive design study
- Discusses clinical, statistical, and regulatory aspects of wide range of adaptive design clinical studies
  - Familiar and unfamiliar approaches
- Not focused on exploratory studies
- Do not rigorously control Type I error rate

Major Categorisation

Generally Well-understood Adaptive Design Approaches

- Well established study designs
- Planned modifications based on
  interim study result analysis
  (one to multiple times in a study)
- Need no statistical correction related to the interim analysis
- No need to properly account for analysis-related multiplicity of choices

Less Well-understood Adaptive Study Designs

- Relatively little regulatory experience
- Primarily intended for where the primary study objective(s) cannot be achieved by other study designs
- All based on unblinded Interim analyses
- Chief concerns
  - controlling Type I error rate
  - Minimizing impact on statistical and operational bias on treatment effects
  - Interpretability of trial results

US Guidance -Adaptive Design Clinical Trials for Drugs and Biologics

- Guidance is specific to adequate and well-controlled (A&WC) trials
- Adequate and well-controlled effectiveness studies intended to provide substantial evidence of effectiveness required by law to support conclusion that drug is effective (21 CFR 314.126)
- Adaptation not based on study design aspects that are revised based on information obtained entirely from sources outside of the specific study
  - e.g. dose response or PK data obtained from another study or new safety or efficacy data obtained from another study

US Guidance –FDA Interaction

- Complex adaptive design studies warrant earlier and more extensive interaction
- Early and mid stage drug development (exploratory)
  - Less formal than late stage (unmet needs get more attention)
- Late stage drug development
  - FDA will not be involved in examining interim data or providing comments on decisions during the study
- Special Protocol Assessments (SPA)
  - Rarely suited for adaptive designs
  - May not be able to comply to 45 day turnaround
  - Encourage EARLY discussions w/FDA before submitting
  - FDA response may have certain limitations due to fact that FDA cannot commit to study design aspects that are not yet determined