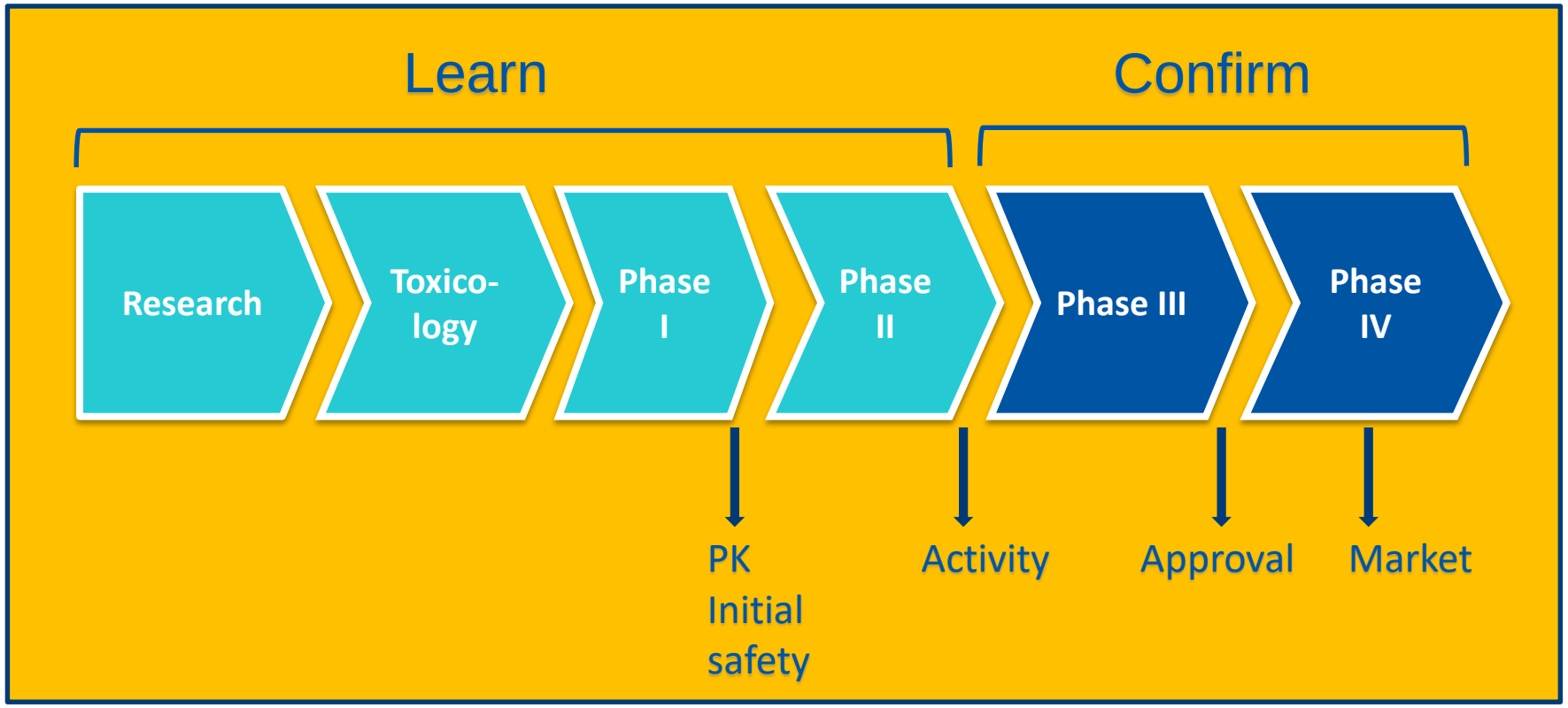
Adaptive trial designs

Laurence Collette, PhD

Head of the Statistics Department

EORTC Headquarters, Brussels, Belgium

Traditional Paradigm of Drug Research and Development



Traditional paradigm

- Inefficient use of patient and financial resources
- Slow and not flexible:
  - does not allow for real-time learning during the course of a trial
- Need to allow modifications during course of trial to increase the chance of success of the drug development strategy

The drug development process needs to be

Adaptive designs

What is an adaptive design? (EMA)

"… A study design is called 'adaptive' if statistical methodology allows the modification of a design element (e.g. sample-size, randomisation ratio, number of treatment arms) at an interim analysis..."

European Medicines Agency (2007). *C*
HMP/EWP/2459/02 Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003616.pdf

Blinded vs Unblinded Adaptions

Based on interim non-comparative analyses

- Study endpoint data in control arm
- Discontinuation rates
- Baseline characteristics

- **No specific statistical concerns**

Based on interim comparative analyses of study endpoints or on outcomes potentially correlated with these endpoints

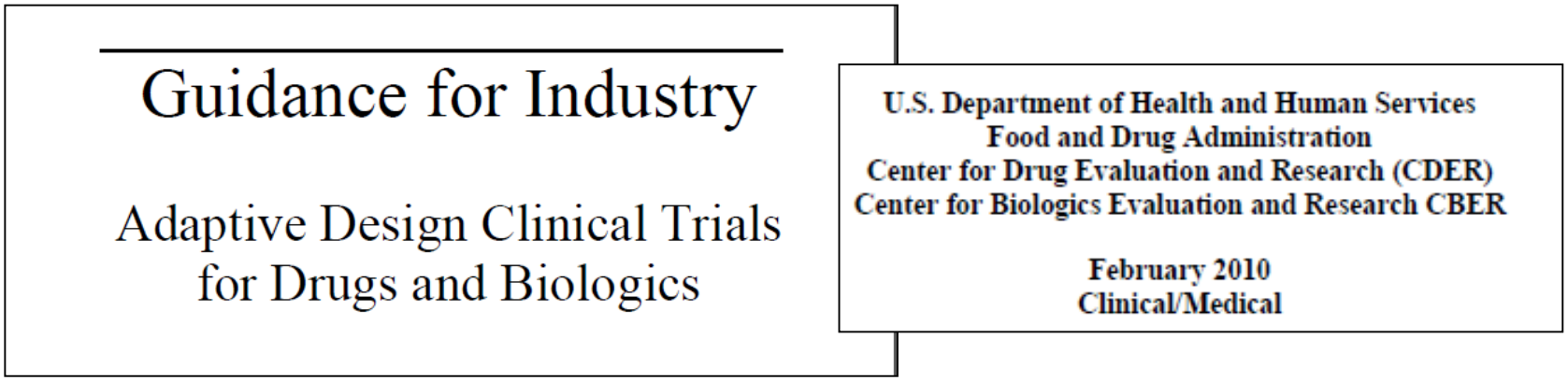
Statistically more challenging

Risk of bias and of type I inflation

What is an adaptive design? (FDA)

"… 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. "….

"…without undermining the validity and integrity of the trial".


Guidance for Industry

Adaptive Design Clinical Trials for Drugs and Biologics

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research CBER

February 2010
Clinical/Medical

Prospectively planned means

- that the adaptation was planned (and details specified & documented) before data were examined in an un-blinded manner by any personnel involved in planning the revision

Auditable by FDA

Validity means

- providing correct statistical inference (such as adjusted p-values to control risk of false positive findings, unbiased estimates and adjusted confidence intervals, etc…)
- assuring consistency between different stages of the study
- minimizing operational bias

Requires specific firewalls & processes

Integrity means

- maintaining confidentiality of data
- providing convincing results to a broader scientific community
- preplanning, as much as possible, based on intended adaptations

Requires simulations

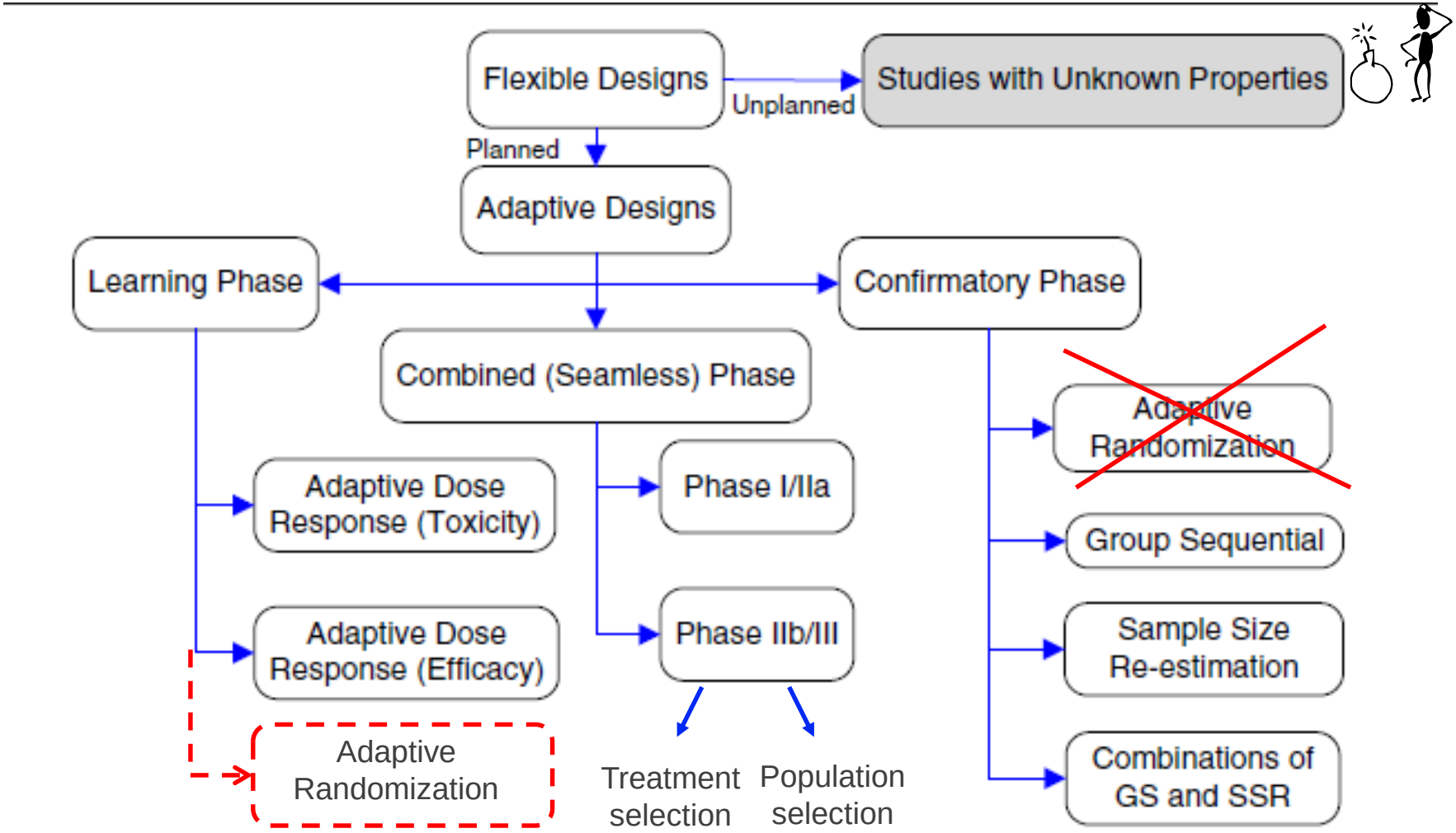
Regulatory Aspects of Adaptive Clinical Trial Designs

Both the FDA & EMEA are

- more **open** when the trial objective is to explore, or "learn".
- much more **cautious** when the trial objective is to "confirm"**.**

➢ **Risk benefit of using an adaptive design versus a more classic "well understood" design must be properly assessed**

Adaptive designs: the very many



Kairalla et al. Trials 2012

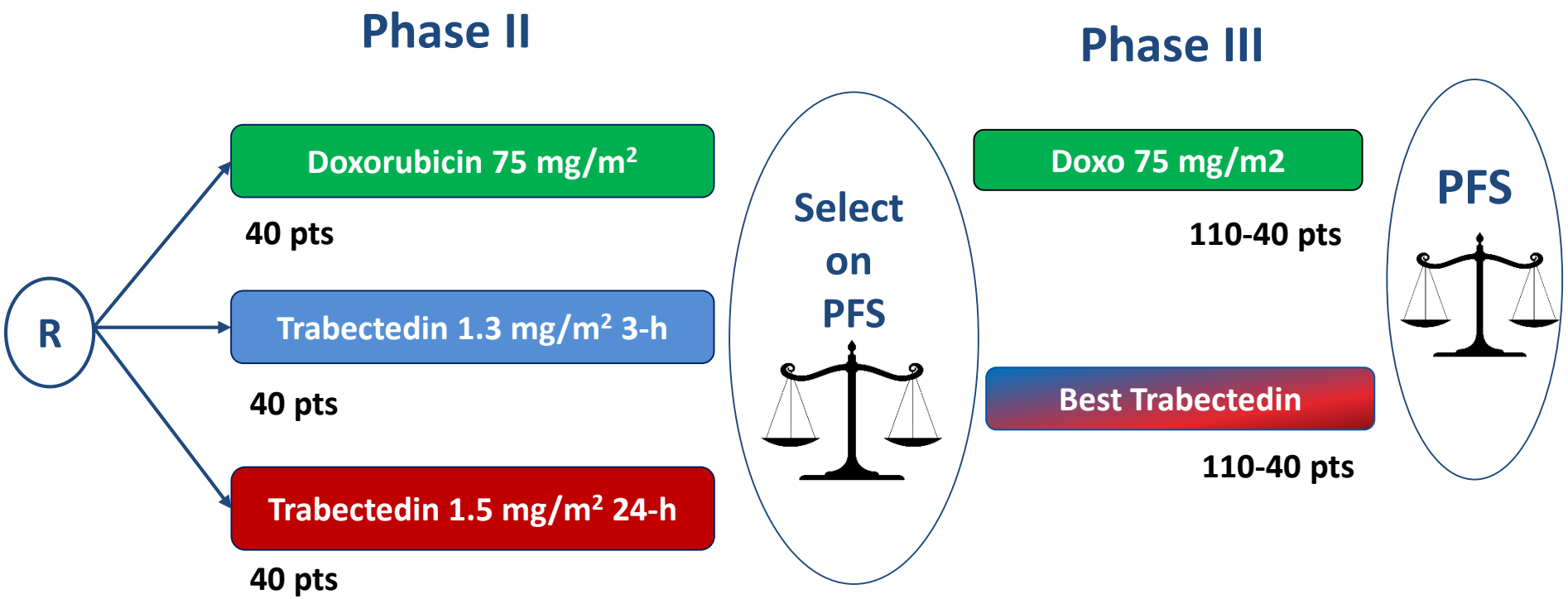
Adaptive designs at EORTC

- Standard to our studies: "well understood" adaptions (FDA 2010)
  - All adaptions based on blinded intermediate results (incl. SS re-estimation)
  - Group sequential designs (early stopping rules)
  - Phase I Continual Reassessment methods
  - Pre-planned switch from superiority to non-inferiority
- We also apply the following adaptions:
  - (Seamless) phase II-III designs with treatment selection
  - Population selection during the study (biomarker enrichment)
- We consider the following:
  - Sample-size re-estimation based on unblinded trt effect estimates
  - Response-adaptive randomization designs

SEAMLESS PHASE II/III WITH SELECTION

- To select the dose regimen within the confirmatory trial (multi-arms trials)
- To select the population within the confirmatory trial (biomarker driven trials)

EORTC 62091 in advanced or metastatic STS

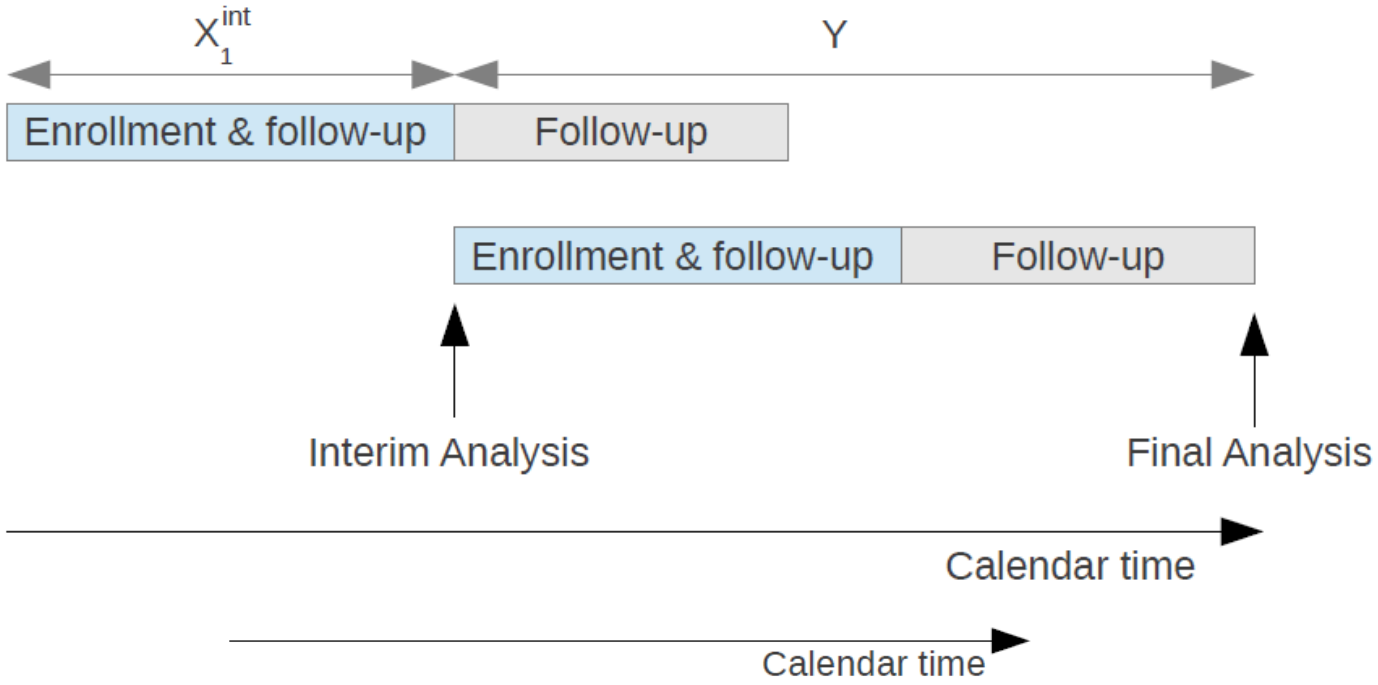


Challenges:

to control type I error at the final test

to control type II error at initial selection

Combination test approach

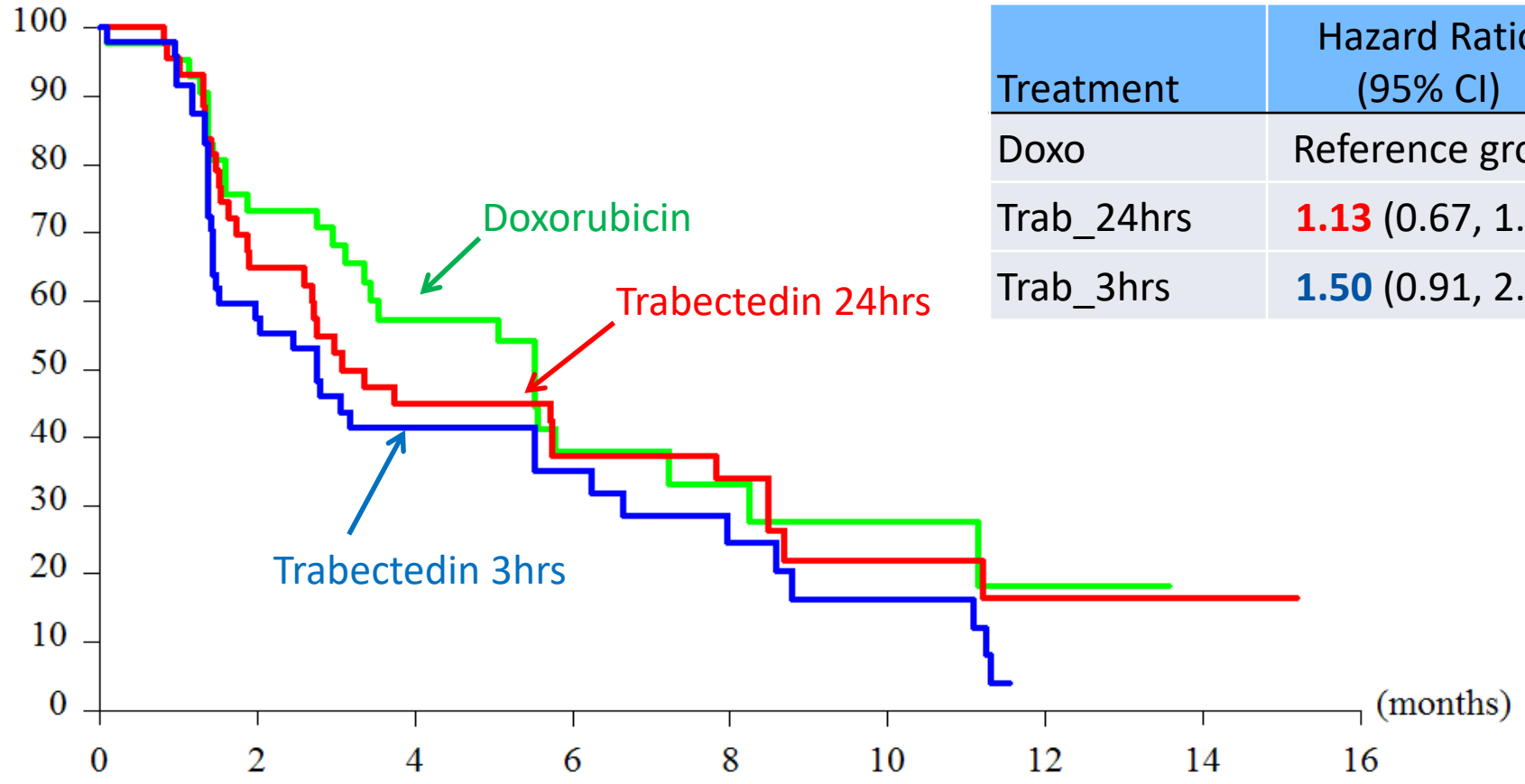


E.g., under H_0,

$$\frac{1}{\sqrt{2}}\Phi^{-1}\left\{1-p_1(X_1^{\text{int}})\right\}+\frac{1}{\sqrt{2}}\Phi^{-1}\left\{1-p_2(Y)\right\} \overset{?}{\sim} \mathcal{N}(0,1)$$

Magirr D et al. PLOS one 2016

Progression free survival



Treatment	Hazard Ratio (95% CI)
Doxo	Reference group
Trab_24hrs	**1.13** (0.67, 1.90)
Trab_3hrs	**1.50** (0.91, 2.48)

O	N	Number of patients at risk :						
26	43	30	20	10	6	5	1	0
31	43	27	18	13	10	5	3	1
37	47	26	16	11	6	4	0	0

Treatment
Doxo
Trab_24hrs
Trab_3hrs

EORTC

Bui, CTOS 2013 abstract

EORTC 62091 in advanced or metastatic STS

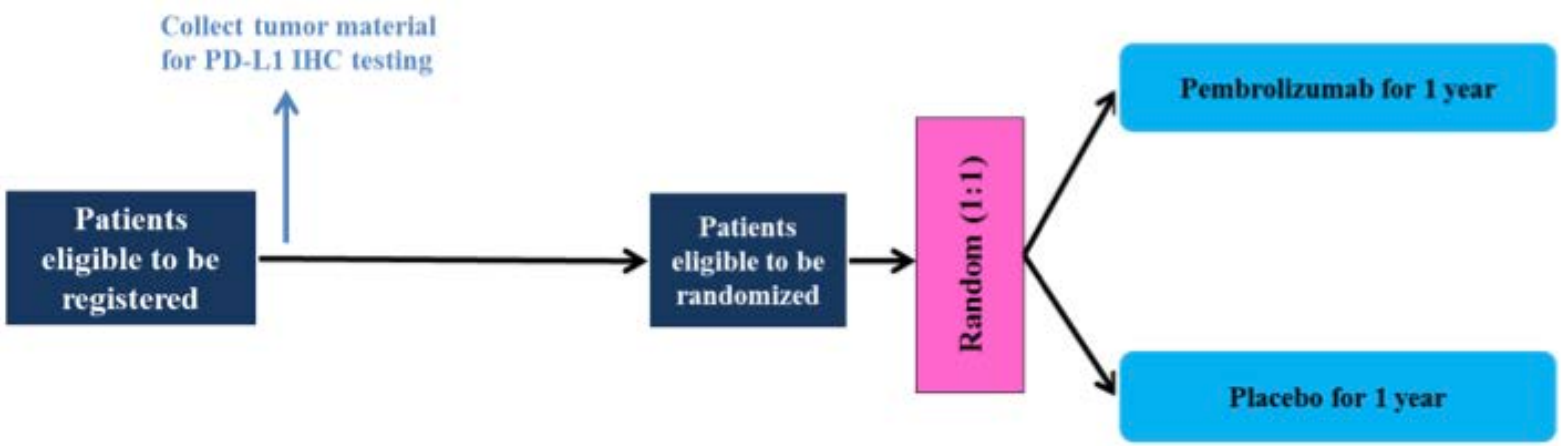
- Both steps are conducted independently and the results of both steps are combined in the end in an overall test result
- Shortens time and patient exposure
- Relatively flexible
- Efficient use of patient resources

- Complex design: statistics are difficult to explain
- Gap in accrual between phase II and phase III if accrual fast (to limit overrunning)
- Logistically challenging
- Difficult in studies with long-term endpoints
  - Unless in combination with a short-term endpoint for the phase II part … another long and complex story on type I error and correlation…

INTERIM BIOMARKER BASED POPULATION SELECTION

PEARLS (EORTC 1416):

A randomized, phase 3 trial with anti-PD-1 monoclonal antibody pembrolizumab (MK-3475) versus placebo for patients with early stage NSCLC after resection and completion of standard adjuvant therapy



- Stage (IB vs II vs IIIA);
- Adjuvant CT (no adjuvant CT versus adjuvant CT);
- PD-L1 status: negative (PS=0) versus weak positive (PS = 1-49%) - versus strong positive (PS≥50%);
- Region (Western EU vs Eastern EU vs the Rest of the world vs Asia)

PEARLS (EORTC 1416)
- Co-primary endpoints

- **DFS in the overall population**
  - an improvement of 13.5 months in median DFS or equivalent to HR = 0.78 is aimed for the whole population.
- **DFS in the PD-L1 strong positive sub-group**
  - *It is assumed* that this subgroup represents 55% of the sample
  - *It is assumed* that around 15% of DFS events at the final analysis will be in the PD-L1 strong positive population (based on the available limited epidemiology data).
  - An improvement of 39.3 months in median DFS or equivalent to HR=0.55 is the effect targeted in this subgroup.
- 1380 randomized patients are required

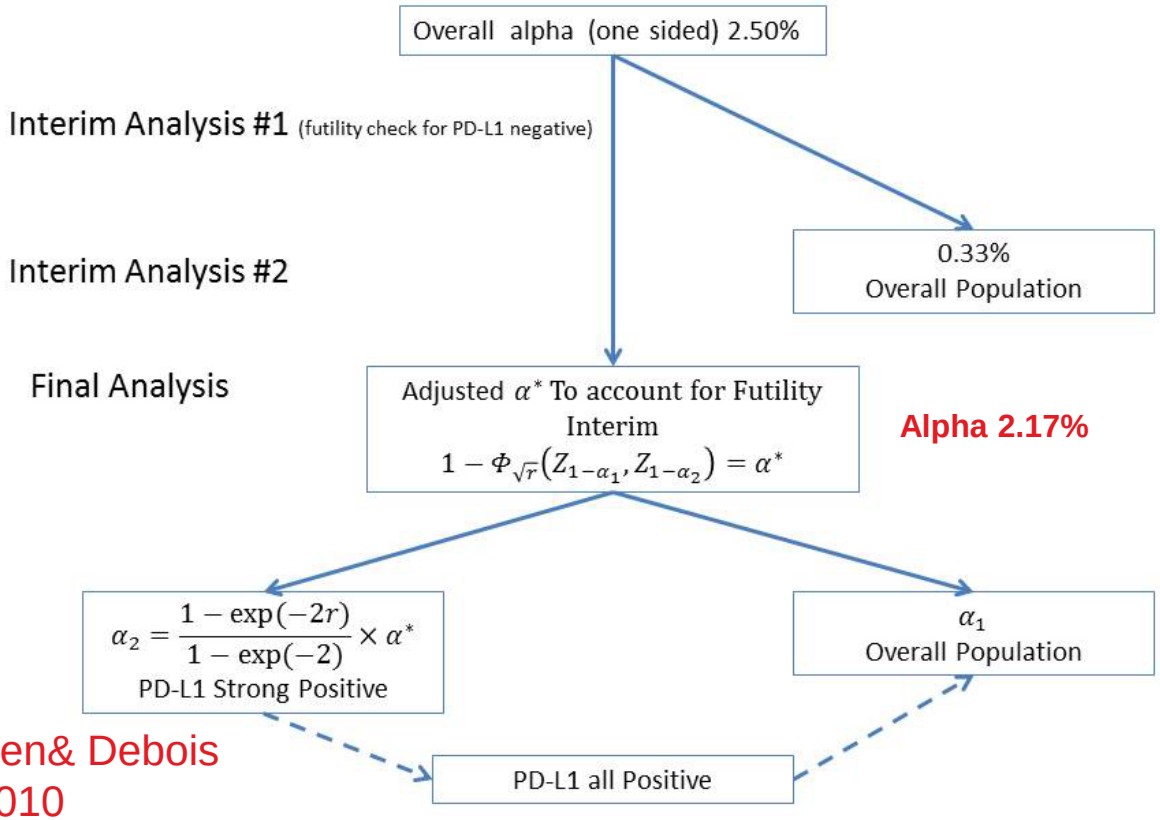
Multiple testing strategy to ensure that if either of the tests is significant, then the study can be declared successful in their respective population/sub-population

PEARLS (EORTC 1416) – Further adaptions during the study

- Interim look to test for futility in the PDL1- population
  → may result in selecting out the PDL1- group and continuing with only the PDL1+ subgroup "POPULATION ENRICHMENT"
  → Adaptions of the final test in this case is covered in the protocol

- Interim look to test for superiority in the full group (group sequential testing)
  → may terminate the trial early if early evidence of overall benefit
- Monitoring of the assumptions regarding rate of PDL1+ and strong PDL1+ patients, sample size may be adjusted if the rate departed strongly from assumptions to ensure sufficient strong PDL1+ in the study

PEARLS (EORTC 1416) – Multiple testing strategy



Bonferroni

Alpha 2.17%

Spiessen& Debois CCT 2010

*If event ratio from PD-L1 Strong Positive vs. Overall population is r

Graphical sequential method to recycle alpha between tests over major secondary endpoint OS and between subgroups.

Bretz et al. SIM 2009

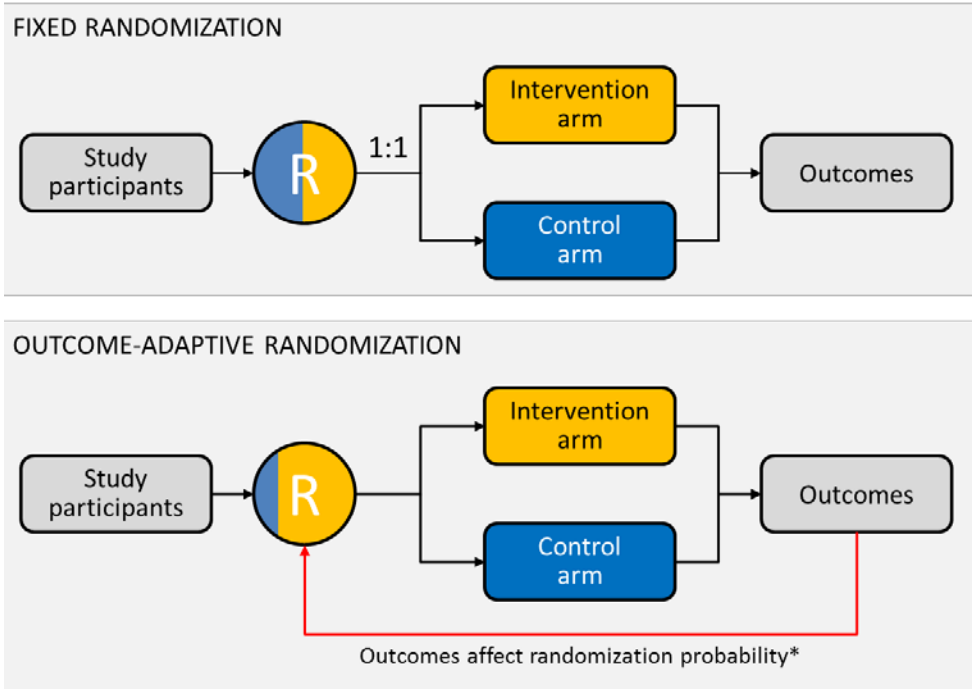
(R Package appended to protocol)

PEARLS (EORTC 1416): Current status

- Study is recruiting

- One interim look at the distribution of PDL1 indicated better than anticipated distribution (greater proportion of PDL1+ than anticipated)

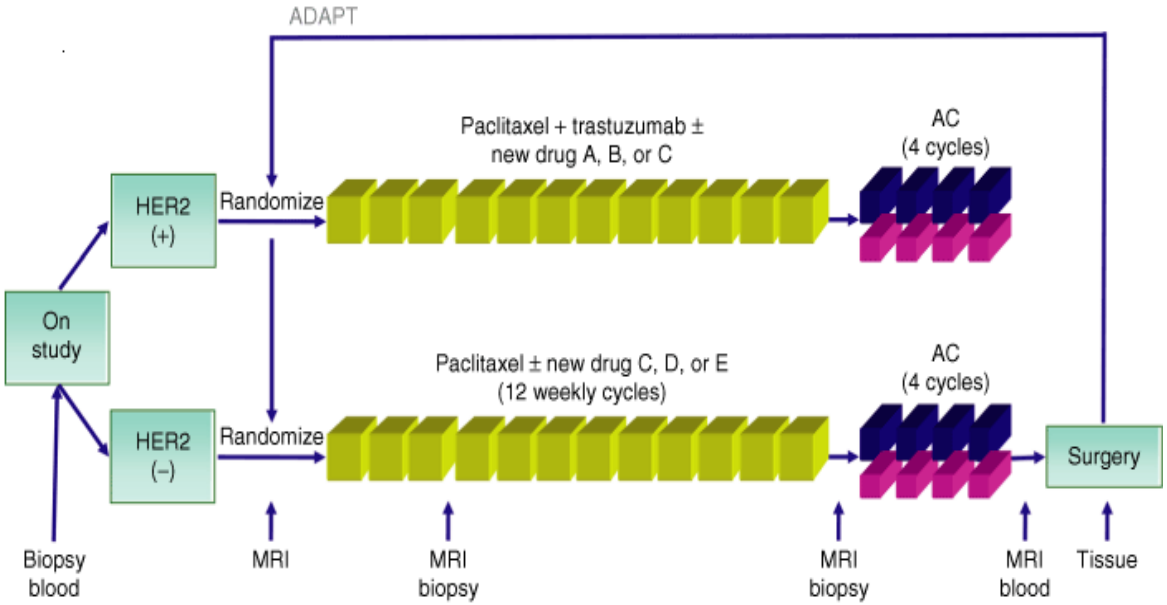
RESPONSE-ADAPTIVE RANDOMIZATION



Mathematical model

"Mm. Fornier, I have two possible treatments for your cancer, A and B, but I do not know which is better. So I would like to enroll you in a clinical trial aimed at comparing these treatments to each other. If you agree to enter the trial, your treatment will be chosen *randomly by a computer, based on the data that we have so far on how well these two treatments have done with previous patients in the trial.*

I-SPY2 Adaptive randomization



10 subgroups investigated

6 “graduations”

Multiple drugs tested

Regimens that show a **high** *(>85%)* Bayesian predictive probability* of being more effective than standard therapy will graduate from the trial with their corresponding biomarker signature.

** (in an equivalent 1:1 randomized phase III trial in the biomarker group)*

Barker et al. Clin Pharmacol Ther. 2009; Berry et al. Molecular oncology 2015

Extremely marginal gain for 2 arms

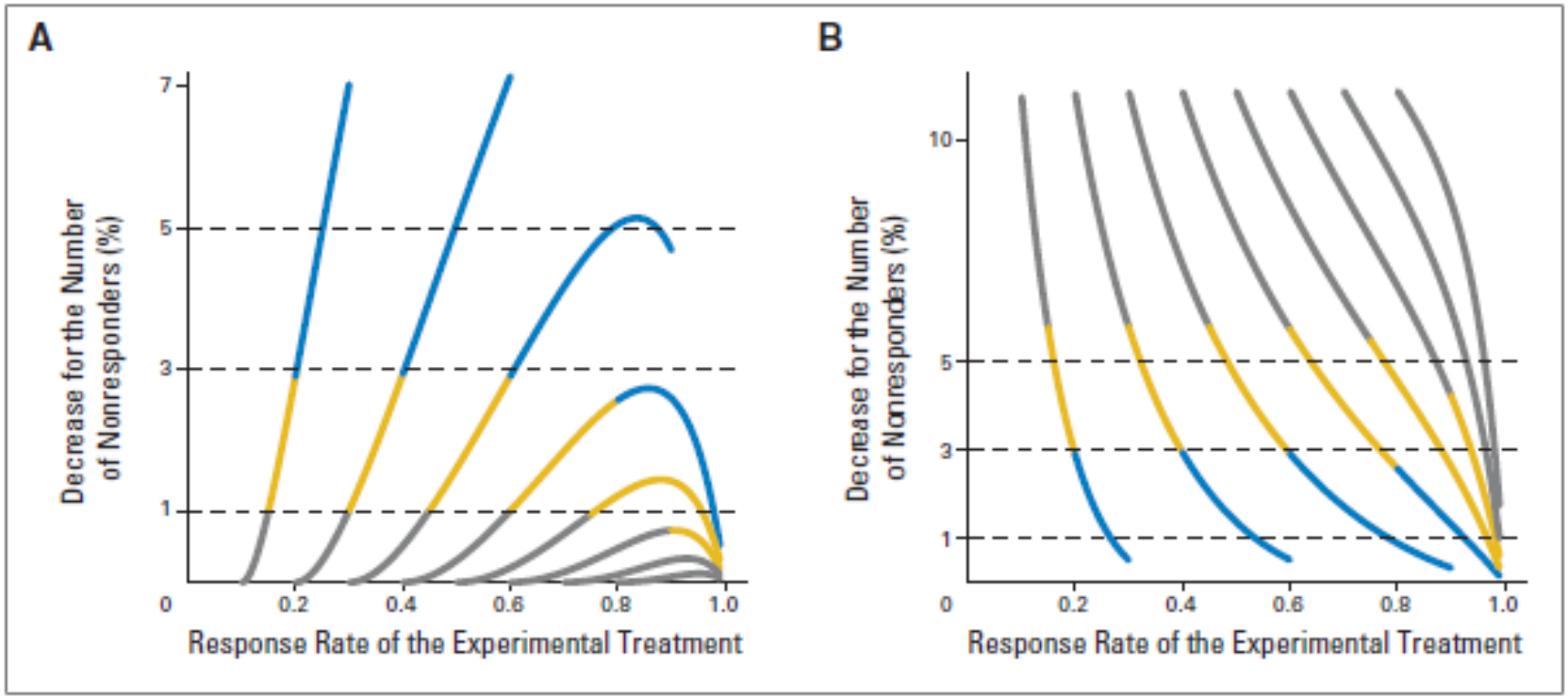


Fig 1. The maximum percentage of reduction in the number of nonresponders for outcome-adaptive randomization with respect to that for (A) 1:1 fixed-ratio randomization and (B) 2:1 fixed-ratio randomization. With respect to the curves from left to right, the response rates of the standard treatment are 0.1, 0.2, ..., 0.8, respectively, and the gray, yellow, and blue segments of each curve indicate that the response rate of the experimental treatment is 0% to 50%, 50% to 100%, and 100% to 200% higher than that of the standard treatment, respectively.

Yuan and Yin JCO 2011

Risk of taking wrong path

More patients go on worst treatment (A)

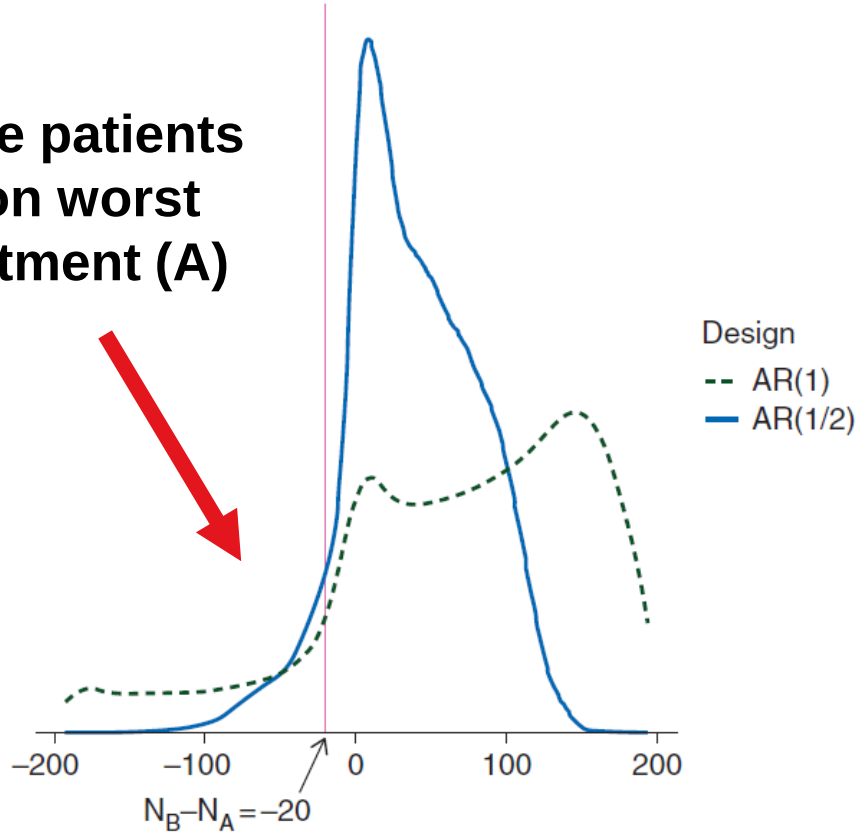


Figure 2. Distributions of the achieved sample size difference $N_B - N_A$ for a 200-patient trial conducted using either AR(1) or AR(1/2), when $p_A = 0.25$ and $p_B = 0.35$.

Randomized 2-arm 200 pt trial
Pa=0.25, Pb=35%

AR(1): adaptive randomization with probability 1 to attribute next pt to current best arm, 0 to current worse arm

AR(1/2): probability 0.67 to current best arm and 0.33 to current best arm

Thall et al. Ann oncol 2015

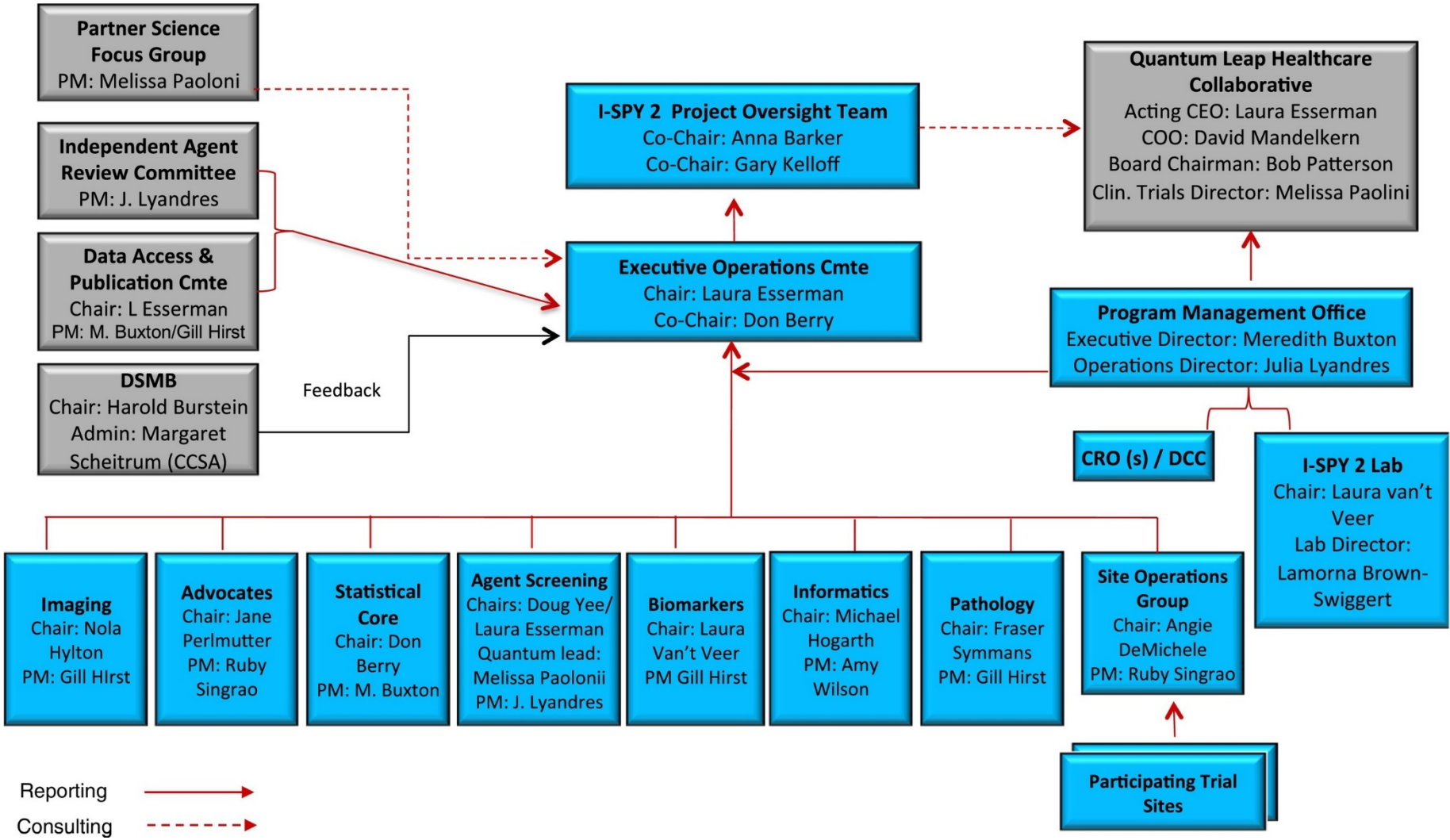
Adaptive randomization

- Improvement in efficiency and individual ethics **only IF >2 arms** and if at least **1 treatment is MUCH better** than all others
- Concerns about **risk of inflation of type I error**
- Difficult Interpretation of results? **Bias in estimation of treatment effect**?
- On average distribution of patients on "winner" arms is better but variance of the distribution is very wide → **risk of exposing more patients to less effective treatment is not negligible**
  Overall advantage over balanced randomization with early stopping rules modest → SIMULATIONS NEEDED
- **Sensitivity to time-drift** (improved prognosis of patients over time)
- Practicability in oncology?: **short term "partial surrogate" needed**

EORTC

Korn and Freidlin JCO 2011; Yuan and Yin JCO 2011 ; Lee et al., CCR 2012; Thall Ann Oncol 2015; Bowden et Trippa Stat Meth Med Res 2015

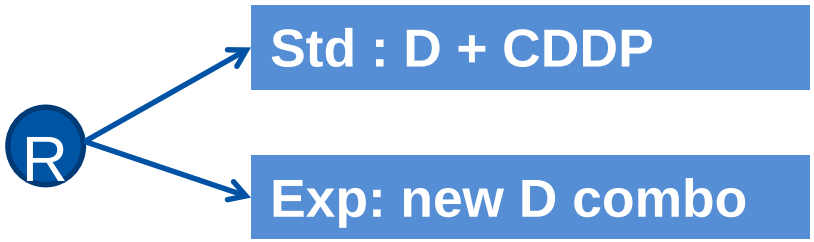
I-SPY 2 Governance structure



Sample Size Adaptation Versus Group Sequential Design

- Group sequential: Start big, stop early if - sign
  - Trial designed to detect a treatment difference Δ
  - Stopping rule at interim for efficacy or futility ("safety-belt")
  - Operating characteristics well known (i.e. control of Type I and Type II error through error spending functions)
  - Conventional analysis

- Sample Size Adaptation: Start small, increase if + sign
  - Trial initially designed to detect an optimistic treatment difference Δ^*
  - Increase sample size based on interim treatment effect
  - Operating characteristics to be simulated in order to understand impact on Type I and Type II error
  - Non conventional analysis (weighted statistic)
  - FDA only allows sample size increase not decrease



- Primary endpoint : overall survival
- Aiming at an increase in median OS from 8 to 11.4 months (HR=0.7)
- …

0.7 too optimistic?

0.77 clinically relevant?

	Wbk1:Des1	Wbk1:Des2
Mnemonic	SU-2S-LRSD	SU-2S-LRSD
Test Parameters		
Design Type	Superiority	Superiority
No. of Looks	2	2
Test Type	1-Sided	1-Sided
Specified α	0.025	0.025
Power	0.9	0.9
Model Parameters		
Hazard Ratio (Alt.)	0.701	0.77
Var (Log HR)	Null	Null
Allocation Ratio (nt/nc)	1	1
Boundary Parameters		
Futility Boundary	Gm (-5) (B)	Gm (-5) (B)
Spacing of Looks	Equal	Equal
Efficacy Boundary	LD (OF)	LD (OF)
Accrual & Dropout Parameters		
Subjects are Followed	Until End of Study	Until End of Study
No. of Accrual Periods	1	1
No. of Dropout Pieces	0	0
Sample Size		
Maximum	483	876
Expected Under H0	421.84	769.186
Expected Under H1	453.819	822.402
Events		
Maximum	334	620
Expected Under H0	258.805	480.406
Expected Under H1	290.35	538.891
Accrual Duration		
Maximum	24	24
Expected Under H0	20.961	21.074
Expected Under H1	22.55	22.532
Study Duration		
Maximum	30	30
Expected Under H0	22.931	23.386
Expected Under H1	26.966	26.959

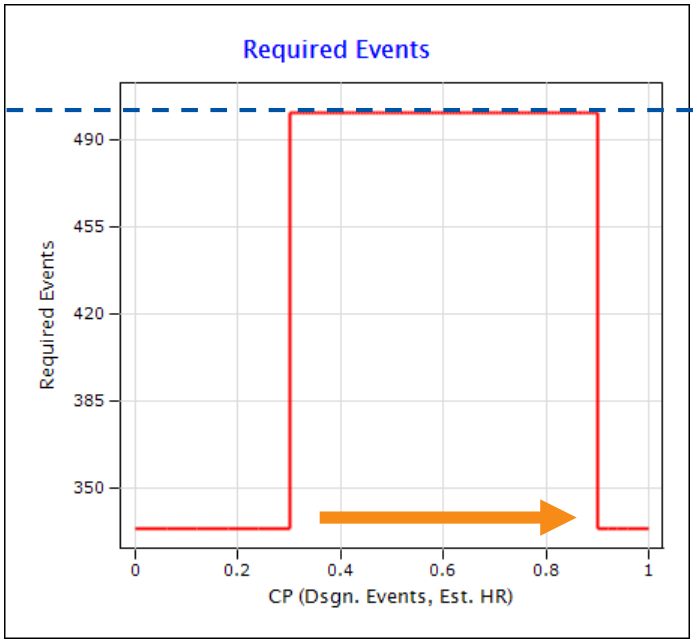
Sample Size Re-Estimation (Adaption)

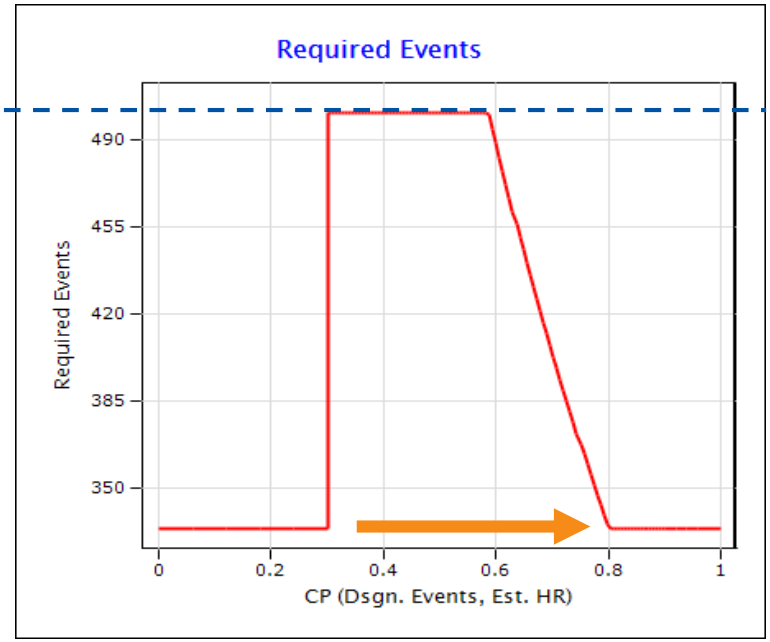
- The study is started with a design based on an optimistic treatment effect (here HR=0.70)
- The possible interim outcomes are split into 3 zones, defined in terms of conditional power (probability of success given current data))

Unfavorable / Futility CP<? (eg. 30%)	• No change in initial design May stop
Promising ?<CP<? (eg. 30-90%)	• **Increase sample size** • (e.g. by max 50%)
Favorable / Efficacy CP>? (eg. 90%)	• No change in initial design Continues to the end

Pattern of sample size increase after interim

Max 501 events





Promising
30%<CP<90%

Promising
30%<CP<90%

- **Increase sample size**
- **Max increase 50% in #events (max=501 events)**
- **#patients increased 50% (max=724 pts)**

Adaptive Sample Size Reestimation

- Very attractive to sponsors: small upfront commitment, additional resources only if “promising” results
- Targeted effect can be adjusted as information comes
- May reduce total sample size compared to pessimistic target HR

- Interim estimates of treatment effect can be misleading (do not apply too early)
- Firewall needed to prevent leakage of information about adaptive rules or decisions
  - Double blind / auditable SOPs & DMC charter / show no change of baseline charactreristics before-after adaption
- Decision to increase or not not in hands of sponsor
- Requires meticulous upfront planning (simulations+++) to demonstrate operational properties and benefit over standard sequential design
- Non conventional analyses (weighted test with prespecified weights)
- Logistically heavy!!

Bhatt and Metha, NEJM 2016

Conclusions

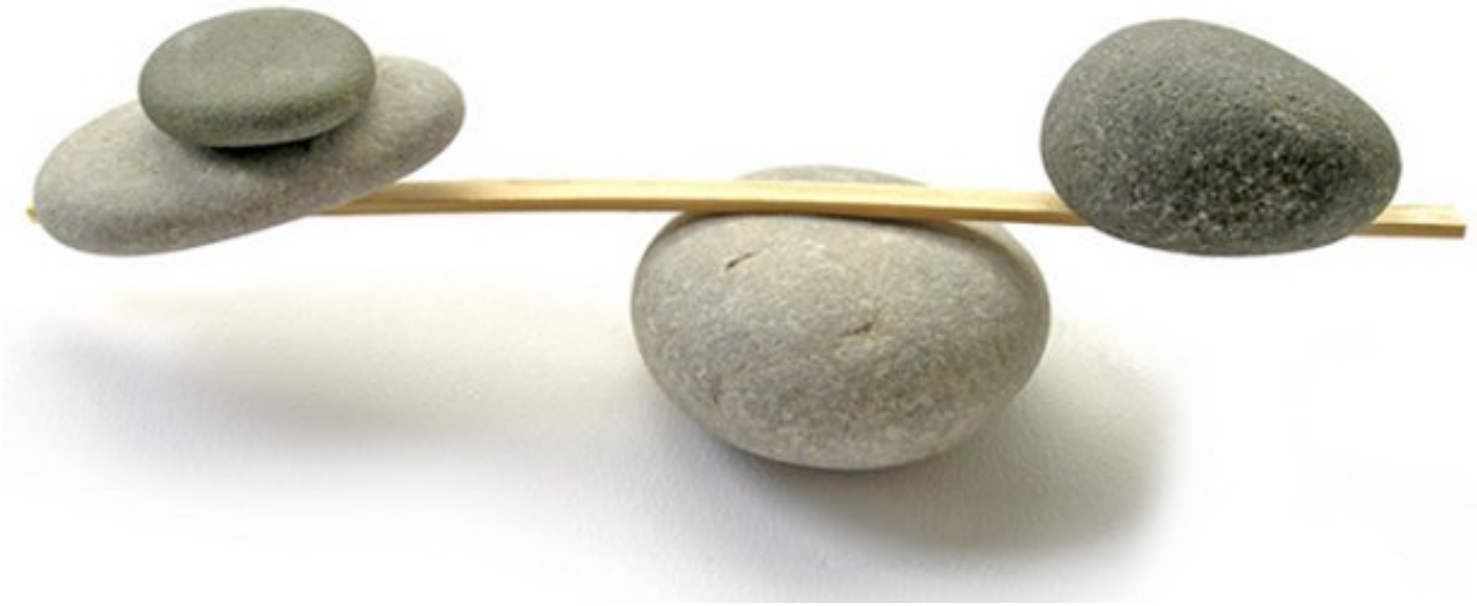
Adaptive study designs have advantages

Flexible design strategies that incorporate new knowledge

Shorter total development process

But this comes at the cost of greater complexity in order to preserve study validity and integrity of the study

Mind operational bias & statistical risks



- This presentation was prepared with the help of my colleagues

  - Saskia Litière, ScD – Associate head of department
  - Murielle Mauer, ScD – Lead statistician
  - Catherine Fortpied, MSc – Lead statistician

References

- Bauer P, Bretz F, Dragalin V et al. Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. Stat med 2016; 35: 325-47. (with commentaries)
- Berry DA. The brave new world The Brave New World of clinical cancer research: Adaptive biomarker-driven trials integrating clinical practice with clinical research. Mol Oncol 2015; 9: 951-9
- Bhatt DL, Metha C. Adaptive designs for clinical trias. NEJM 2016; 375 (1): 65-75
- Bowden J, Trippa L Bias adjusted estimation for binary outcomes in clinical trials using adaptive randomisation. Stat Meth Med Res 2015; 26(5) 2377-86
- Das S, Lo W. Re-inventing drug development: A case study of the I-SPY 2 breast cancer clinical trials program. CCT 2017; 62: 168-74.
- European Medicines Agency. CHMP/EWP/2459/02 Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design (2007)
- Jennison C and Thurnbull BW. Adatpive sample size modification in clinical trials. Start small then ask for more? Stat Med 2015; 34: 3793-3810.
- Keiralla JA, Coffey JS, Thomann MA et al. Adaptive trial designs: a review of barriers and opportunities. Trials 2012; 13:145
- Korn EL and Freidlin B. Outcome-Adaptive Randomization: Is It Useful? J Clin Oncol 2011; 29: 771-6.
- Lee JJ, Chen N and Yin G. Worth adapting? Revisiting the usefulness of outcome adaptive designs CCR 2012; 18(17):4498-507
- Thall P, Fox P and Whathen J: statistical controversies in clinical research:scientific and ethical problems with adaptive randomization in comparative clinical trials. Ann Oncol 2015; 26: 1621-28
- US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics. Draft Guidance February 2010.
- Yuan Y and Yin G. On the usefulness of outcome adaptive randomization. J Clin Oncol 2011; 13: e390-92