Adaptive Trials

Raphaël Porcher

CRESS, Inserm UMR-S 1153, Université Paris Descartes

Modélisation et simulation d'essais cliniques
Toulouse
9–10 avril 2015

Outline

Outline

Adaptive design for a clinical trial[1]

- Uses data accumulated during the trial to possibly modify some aspects of the study
- Without undermining its validity and integrity

[1] Draglin V. Adaptive designs: classification and taxonomy. Adaptive Designs Workshop, 2006

Validity and Integrity?

Definition (Validity)

- Correct statistical inference (test and estimation)
- Consistency between the different trial stages
- Minimizing operational bias

Definition (Integrity)

- Results acceptable for the scientific community
- Preplanning of adaptations as much as possible
- Maintaining confidentiality of data

Main adaptive designs

Type of design	Adaptation
Group sequential trial	Early stopping
Sample size reassessment (blinded–variance, other nuisance parameters)	Increase sample size
Phase 1 dose finding CRM (*Continual Reassment Method*)	Choice of next dose
Combined phase 1–2	Choice of next dose
Phase 2 adaptive dose ranging	Modify the allocation ratio
Sample size reassessment (unblinded - using observed efficacy)	Increase sample size
Population enrichment	Modify inclusion criteria, analysis population $\rightarrow$ subgroup
Combined phases 2–3 (ex-seamless)	Select dose, . . .

Stage of drug development

- Confirmatory trials
  - Goal = market authorization
  - Strict control of type I error rate required
- Exploratory trials
- Regulatory constraints less strong than for confirmatory trials

Exploratory / confirmatory

Type of design	Adaptation
Group sequential trial	Early stopping
Sample size reassessment (blinded–variance, other nuisance parameters)	Increase sample size
Phase 1 dose finding CRM (*Continual Reassment Method*)	Choice of next dose
Combined phase 1–2	Choice of next dose
Phase 2 adaptive dose ranging	Modify the allocation ratio
Sample size reassessment (unblinded - using observed efficacy)	Increase sample size
Population enrichment	Modify inclusion criteria, analysis population $\rightarrow$ subgroup
Combined phases 2–3 (ex-seamless)	Select dose, . . .

Perceived methodology

- By regulatory agencies
  - Well understood methods
  - Less well understood methods
  - Evolved in the last 5-6 years
- By pharmaceutical companies
  - Method accepted by the regulatory
  - Benefit/risk ratio for the trial, for the entire drug development

(Less) Well understood

Type of design	Adaptation
Group sequential trial	Early stopping
Sample size reassessment (blinded–variance, other nuisance parameters)	Increase sample size
Phase 1 dose finding CRM (*Continual Reassment Method*)	Choice of next dose
Combined phase 1–2	Choice of next dose
Phase 2 adaptive dose ranging	Modify the allocation ratio
Sample size reassessment (unblinded - using observed efficacy)	Increase sample size
Population enrichment	Modify inclusion criteria, analysis population $\rightarrow$ subgroup
Combined phases 2–3 (ex-seamless)	Select dose, . . .

Outline

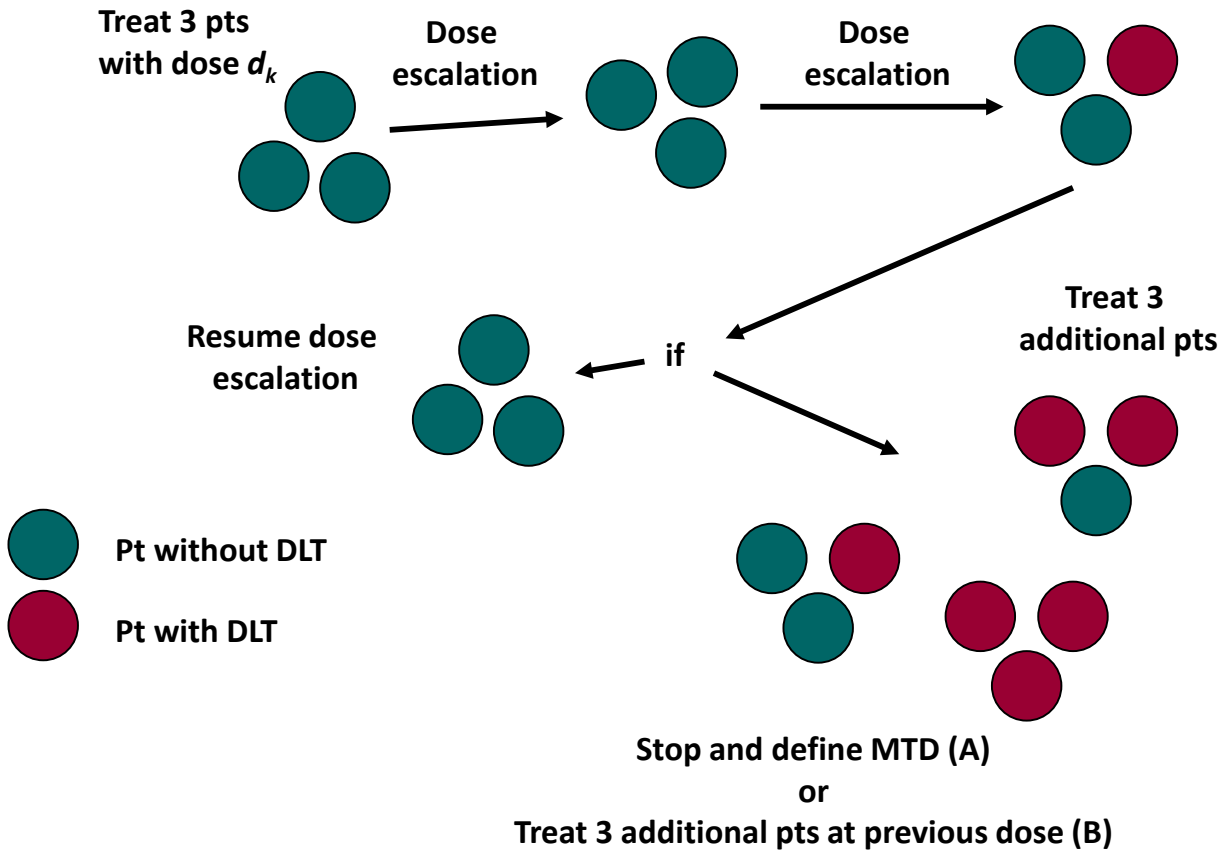
Phase 1: (modified) CRM

- Paradigm of oncology phase 1 trials
- Dose-finding: we search the MTD
  - Dose level associated with an "acceptable" level of toxicity
  - Percentile of the dose–inacceptable (dose-limiting) toxicity relationship
- Underlying paradigm: *more is better (efficacy)*

Standard design: '3+3' dose escalation

- k dose levels administered to cohorts of 3 to 6 patients
- Lowest dose depends on preclinical studies
- Predefined dose levels $d_1 < \ldots < d_k$

3+3 design (2)


Treat 3 pts
with dose d_k
Dose
escalation
Dose
escalation
Treat 3
additional pts
if
Resume dose
escalation
Pt without DLT
Pt with DLT
Stop and define MTD (A)
or
Treat 3 additional pts at previous dose (B)

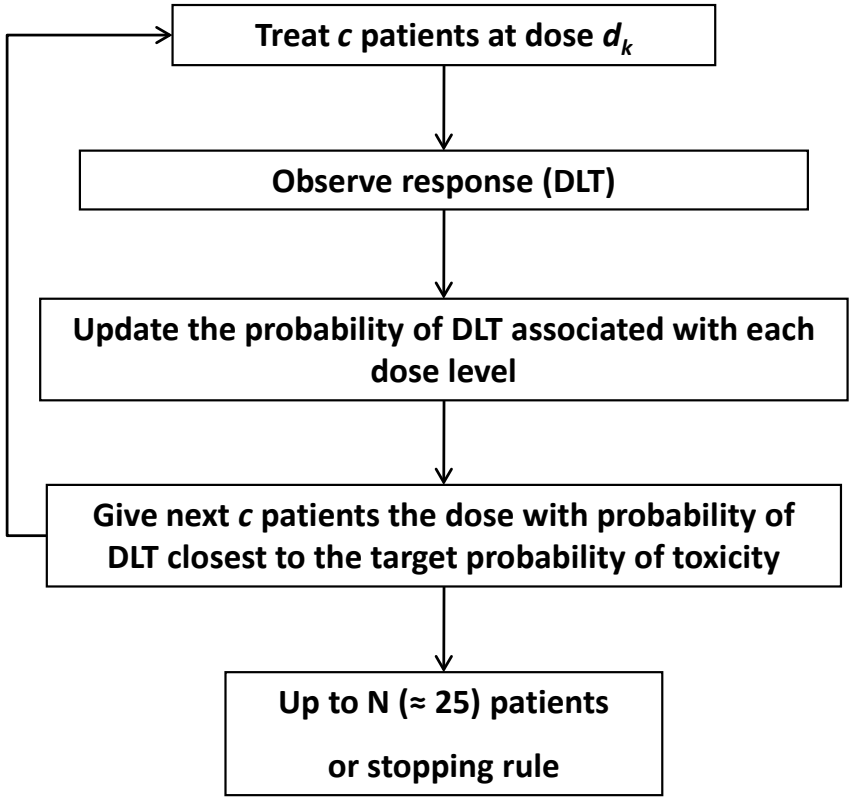
Limits of standard design

- Statistics: lack of precision the the toxicity rate
  - with 3 doses: 0 TDL/3, 1/6, 2/6
  - 90%CI: 0–0.54, 0.01–0.58, 0.06–0.73
  - Targeted probability between 0.17 and 0.33 $\rightarrow \simeq$ 0.25, likely with all three doses!
- Ethics: high probability of dose escalation at the MTD (30 to 80%)
  - Do not undertreat too many patients
  - Do not overtreat too many patients

CRM

- Sequential and adaptive design:
  - Dose for next cohort determined all previous observations (process memory)
- And parametric (model for the dose-effect relationship)
- Inference (parameter estimation)
  - "Frequentist" (likelihood)
  - Bayesian (parameter = random variable)

CRM: schematic representation of the process



Combined phase 1–2

- Guide dose finding on both toxicity and efficacy
- Methodology quite similar to the CRM
- For instance with Bayesian inference
- Observed outcome = (Toxicity,Efficacy)

Bayesian dose finding using efficacy–toxicity trade-offs[2]

- Estimate $\pi_E(d) = \Pr(\text{Efficacy}|d)$ and $\pi_C(d) = \Pr(\text{Toxicity}|d)$
- Acceptability criteria: $\pi_E(d) \geq l_E$ and $\pi_C(d) \leq u_C$
- Several optimality criteria in terms of $\pi_E(d)$ and $\pi_C(d)$
- An several methods of estimation (we won't go into the details)

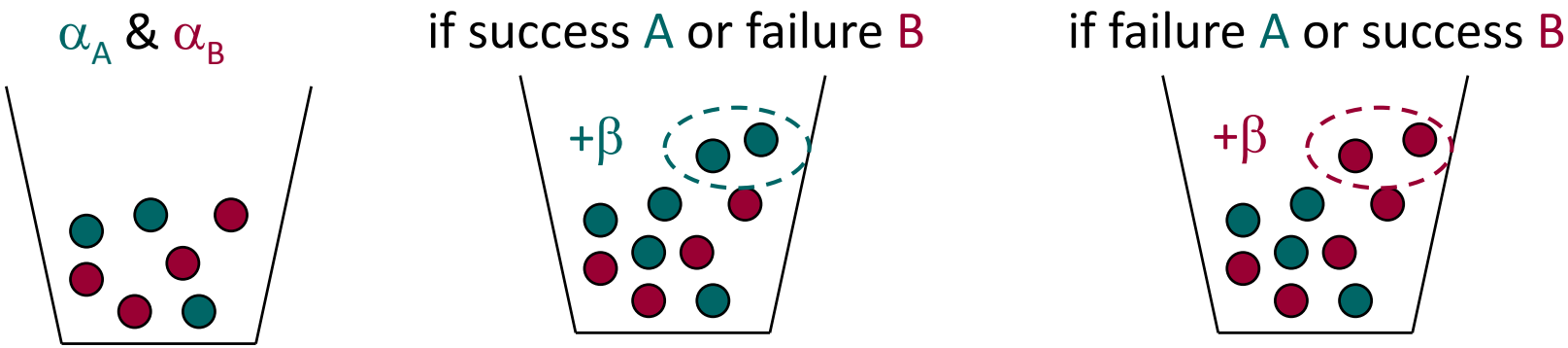
[2] Thall, Russell, 1998; Thall, Cook, 2004; ...

Phase 2: Adaptive dose ranging

- Phase 2: exploratory trial of drug's efficacy
- Search for the right dose to be administred
- As opposed to *dose finding* (previous slides)
- Adaptation: allocate more patients to the doses that seem more effective

Reevaluation of allocation ratio

- One possible method: randomized play-the-winner
- Sequential reevaluation of the probability to receive each treatment (dose) at random allocation



Sample size reassessment

- Two paradigms
  - Blinded (to efficacy results)
  - Unblided to efficacy results
- Different objectives
  - First case: reassess nuisance parameters
  - Second case: a bit more complex . . .

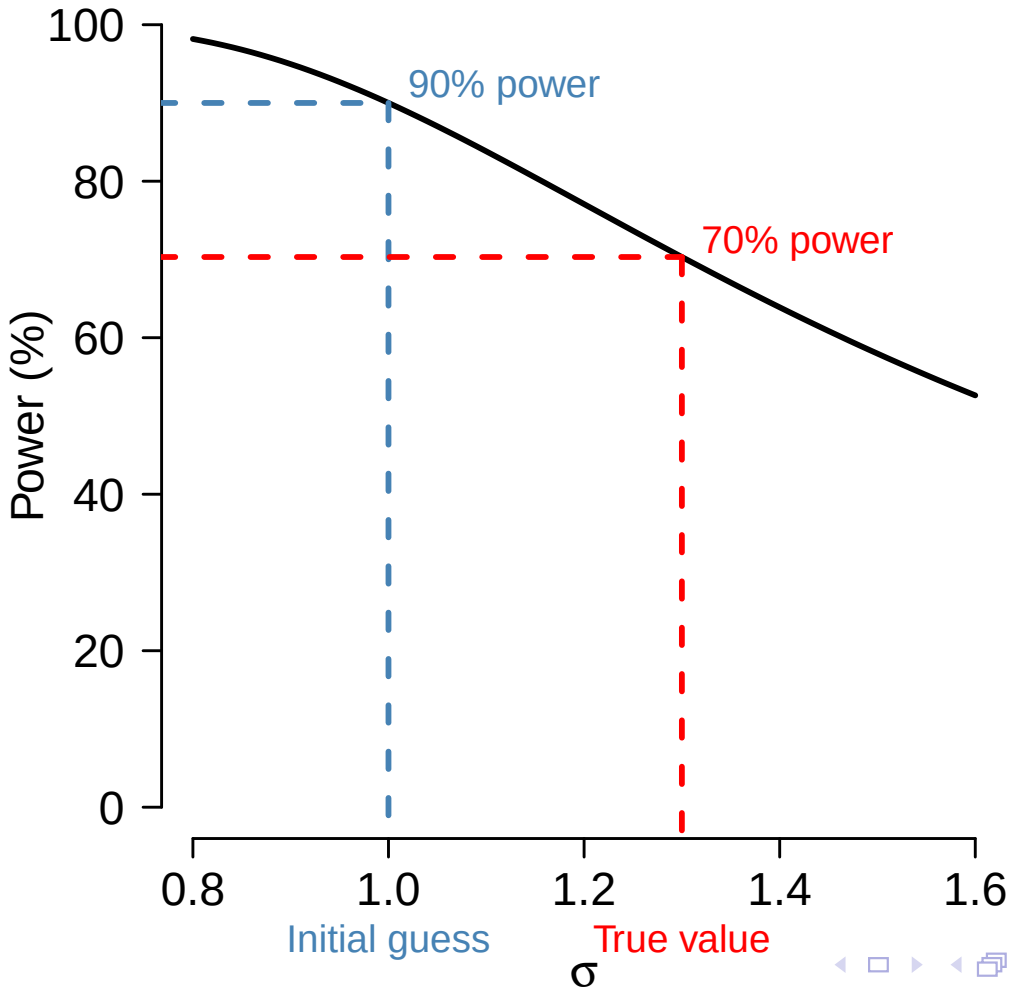
Blinded SSR

- The sampel size depends on
  - Type I et II error rates: α and β (1-power)
  - Difference to be detected: Δ (in a general sense: MD, RD, HR . . .)
  - Variance of the outcome
- Simple case, continuous outcome

$$n = \frac{2\left(z_\alpha + z_\beta\right)^2 \sigma^2}{\Delta^2}$$

- α et β are quite "standard"
- If we make an error on $\sigma \rightarrow$ loss of power

Influence of an error on σ



Example[3]

- Multicenter randomized double-blind trial evaluating lumiracoxib vs ibuprofen on the blood pressure in patients with osteoarthritis and controlled hypertension
- Primary outcome: 24-h mean systolic blood pressure at 4 weeks
- Planning $\alpha = 0.025$ (1-sided), power 80%, meaningful difference $\Delta = 2$ mmHg
- SD $\sigma =$??? mmHg

[3] MacDonald et al. J Hypertension 2008;26:1695–1702. Thanks to Karine Lheritier, Marianne Notter, and Tim Friede

Example: σ and influence on N

- Other studies
  - White et al. (2002): 9 mmHg observed (slightly different population)
  - Sowers et al. (2005): trial planned with 7.5 mmHg, but observed SD 12 mmHg (at 6 w)
  - Other studies with the same outcome but different populations: up to 14 mmHg

σ	7.5	9	12	14
N	442	636	1130	1538

Example (cont'd)

- Fixed trial size: 1020 patients
- Planned blinded SSR after 600 patients
- Blinded estimation of SD: 8.33 mmHg
- Revised sample size : 550
- 787 patients already recruited
- Decision to stop recruitment
- Final analysis showed a significant effect
- Post-hoc power 91% (vs 80% initially planned)
- No increase of type I error rate
- No other impact on the conduct of the trial and blinding

Unblinded SSR

- Uncertainty on Δ
  - Over-optimistic: risk of missing an interesting effect
  - More pessimistic: too large a N to achieve the trial
- Solution: take quite an optimistic Δ, with a clause to extend the trial if exults are promising
  - Prespecify in the protocol the upper limit of same size
  - IDMC will give instructions to the sponsor, who remains blinder to the study results.
- Alternative +++: Group sequential design
  - Planned with a larger sample size from the beginning
  - With the possibility for early stopping

Promising zone design[4]

- Example of an oncology trial
- Median survival with control: 8 months
- HR 0.70 under the alternative plausible +++
- But HRs up to 0.80 would be interesting anyway
- $\alpha = 5\%$, power 90%

HR	No. events	No. subjects	Duration (months)
0.70	330	430	42
0.72	390	510–430	42–68
0.74	464	...	...
0.76	558	...	...
0.78	680	...	...
0.80	844	1100–930–?	42–68–?

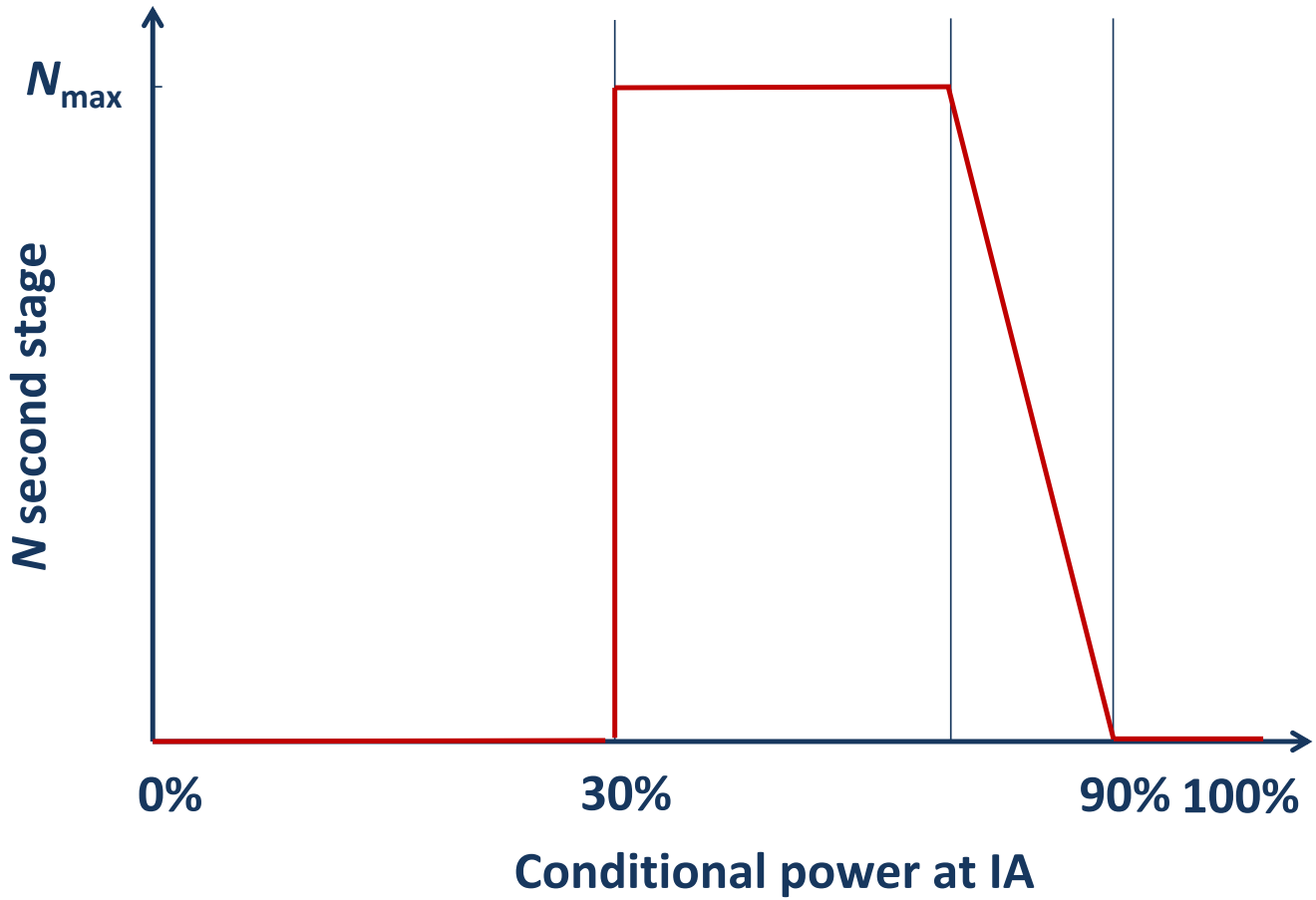
[4] Mehta and Pocock, 2011

Promising zone design (2)

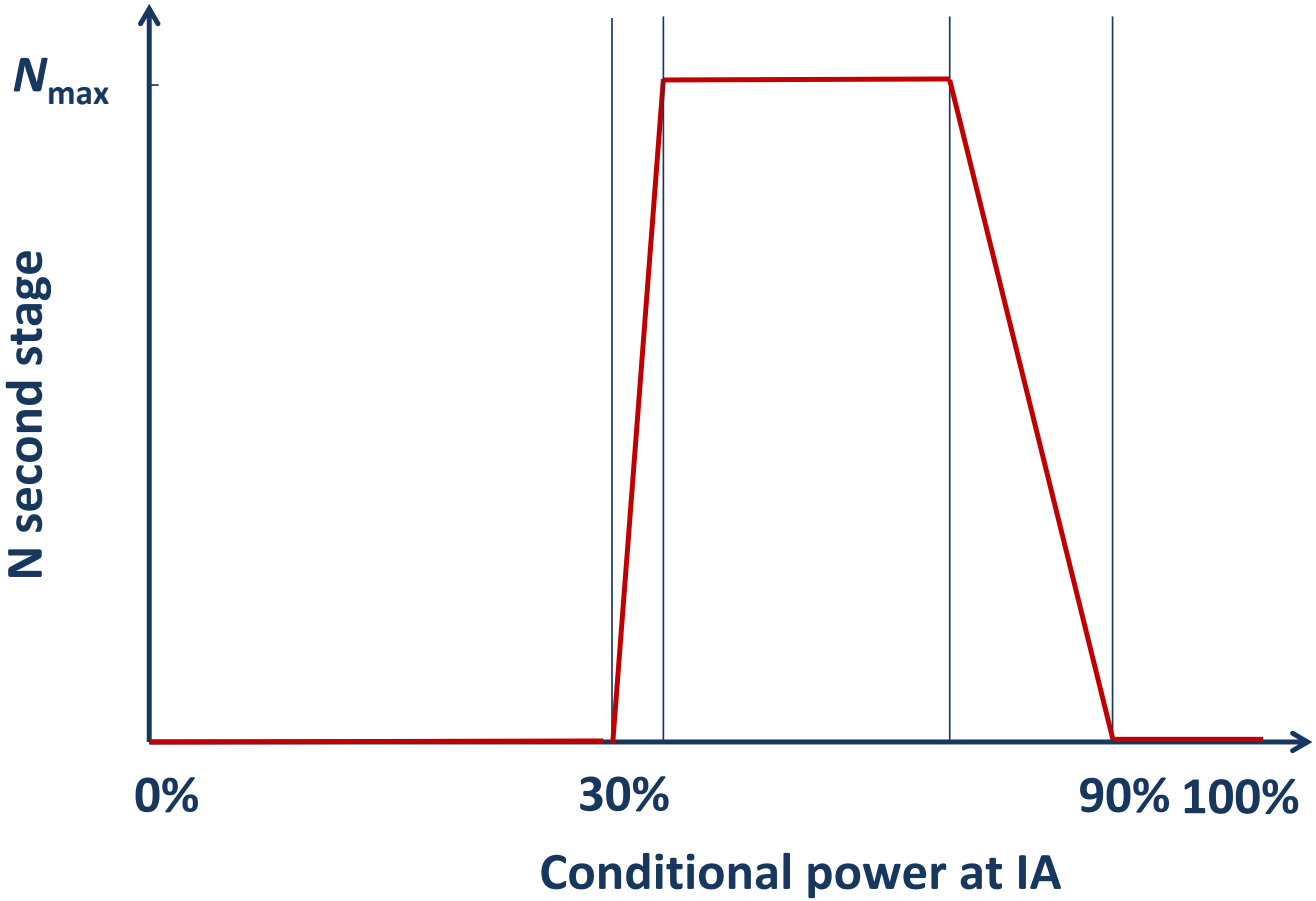
- Plan with HR$=$ 0.70
- Interim analysis with conditional power calculation

Conditional power	Zone	Decision
$> c_{\text{eff}}$	Efficacy	Stop
90%–c_{eff}	Favorable	Continue with no change
30%–90%	Promising	Reassess N
c_{fut}–30%	Unfavorable	Continue with no change
$< c_{\text{fut}}$	Futility	Stop

Reevaluation of N



Or rather . . .



Properties[5]

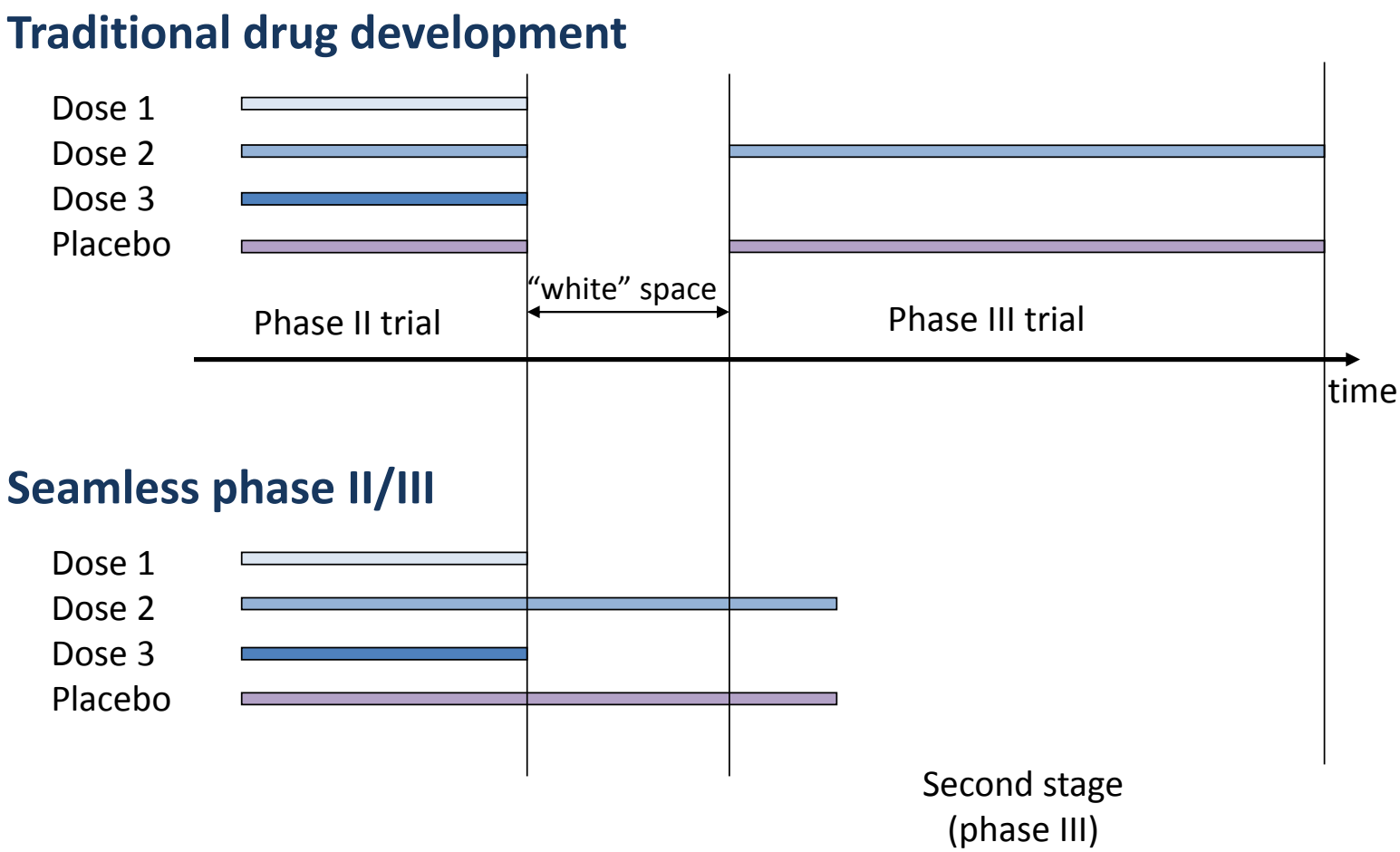
			Classical trial		PZD	
True HR	Zone	Pr(zone)	Power	Evts	Power	Evts
0.76	Unfavorable	20%	42%	423	42%	423
0.76	Promising	24%	75%	423	93%	656
0.76	Favorable	57%	95%	423	95%	423
0.78	Unfavorable	25%	34%	423	34%	423
0.78	Promising	25%	68%	423	88%	658
0.78	Favorable	50%	93%	423	93%	423
0.80	Unfavorable	31%	28%	423	28%	423
0.80	Promising	26%	62%	423	84%	668
0.80	Favorable	43%	93%	423	93%	423

[5] Thanks to Y. Jemiai, Cytel Inc.

Combined phase 2(b)–3

- One trial, two "traditional" phases
  - Stage 1: phase 2 (e.g. dose ranging)
  - Stage 2: phase 3
- Confirmatory trial
- Distinguish
  - Trials that are operationally *seamless*
  - Trials that are inferentially *seamless*
- In the latter case, the final analysis uses all included patients

Schematic representation



General methodology

- Null hypothesis for stage 1 H_{01} (e.g. no difference on early endpoint)
- Other null hypothesis H_{02} (e.g. no difference on clinical endpoint)
- Global null hypothesis $H = H_{01} \cap H_{02}$
- Goal: to combine results from the two stages to control α under H

First stage

- Test $H_1 = H_{01}$
- Recruit n_1 patients $\rightarrow Z_1 \rightarrow p_1$
- If $p_1 \leq \alpha_1$: Reject H_{01} and continue to test H_{02}
- If $\alpha_1 < p_1 \leq \alpha_0$: Do not reject H_{01} (yet) but continue to test$H_{01} \cap H_{02}$, H_{01} and H_{01}
- If $p_1 > \alpha_0$: Stop for futility

Second stage

- Test $H_2 = H_{02}$ or $\{H_{01} \cap H_{02}, H_{01}, H_{01}\}$
- Recruit n_2 additional patients
- $Z_2 \rightarrow$ p-value $p_2(Z_1, Z_2)$
- Reject H_2 and thus H if $p_2 \leq C(z_1)$
  ($C(.) =$ conditional error function)

Phase 3: population enrichment

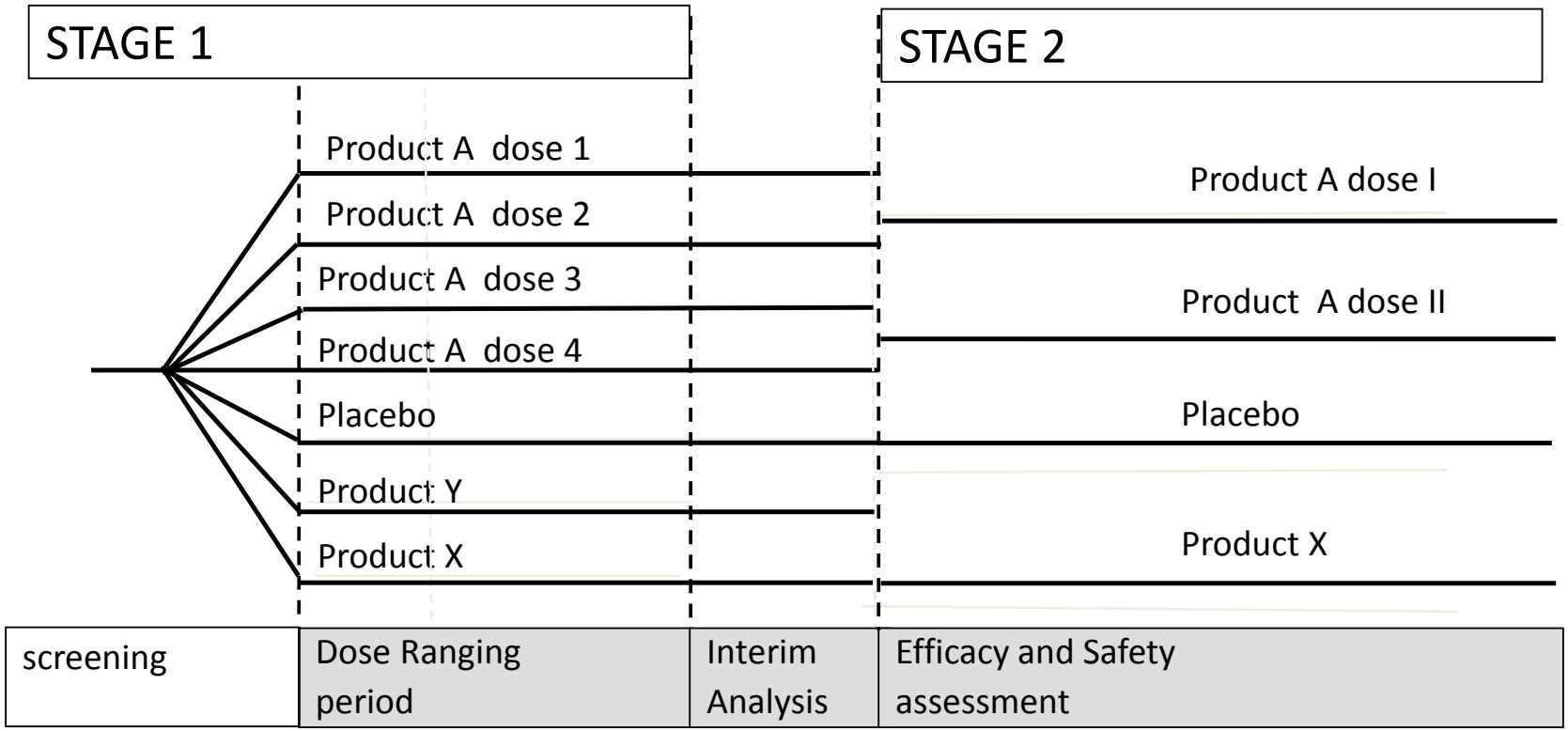
- Trial that begins with a "wide" population
- And possibly continues in a targeted subpopulation if efficace is shown in the subgroup
- Recognized methodology when
  - Subgroups are defined in advance
  - The trial is planned that way from the beginning
- Methods to control the type I error rate α

Post-hoc enrichment

- Analysis that was not pre specified
- Or trial that was not planned with an adaptive design
- Cases where such trials were conducted with a "clean" rationale: e.g. new marker discovered outside the trial
- Other rationales more debated . . .

Outline

(True) phase 2–3 trial



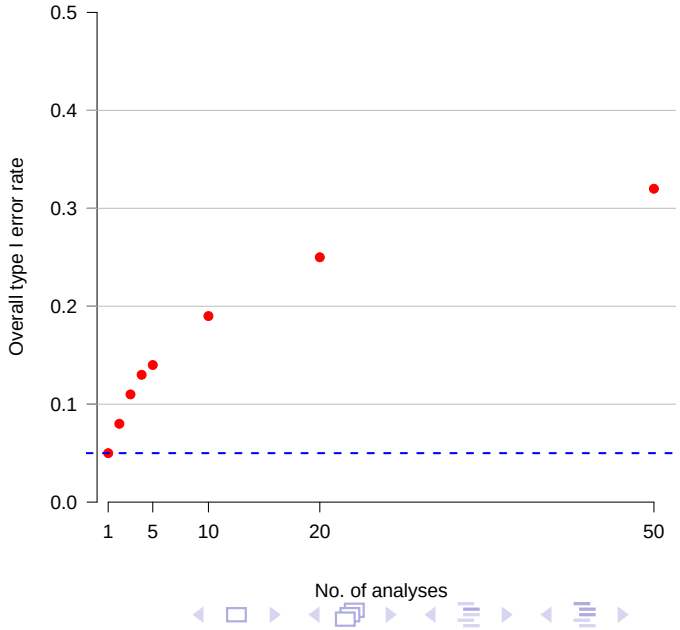
What statistical issues should be accounted for?

Control α for group sequential analyses

- Interim analyses
  - First analysis with n_1/arm
  - Second analysis with $(n_1 + n_2)$/arm

$\rightarrow$ Increase of global α

No. tests at 5% level	False positive rate
1	5%
2	8%
3	11%
5	14%
10	19%
20	25%
50	32%



Control α for multiplicity

- Multiple hypotheses

No. hypotheses	False positive rate
1	5%
2	10%
3	14%
4	19%
5	23%
8	34%
10	40%

$\rightarrow$ Increase of global α

Control of α

- Adapted statistical methods
- Interim analyses
  - Rejection boundaries for group sequential trials
  - O'Brien & Fleming, Pocock, Wang & Tsiatis ...
- Multiplicity
  - Correction of *p*-values / local α
  - Bonferroni, Holm, Hochberg, Sidak, ...

Other issues

- How to combine the two stages?
- How to dimension the second stage to control the power
- Which power (Conditional? For what difference?)
- How to analyze/report the results

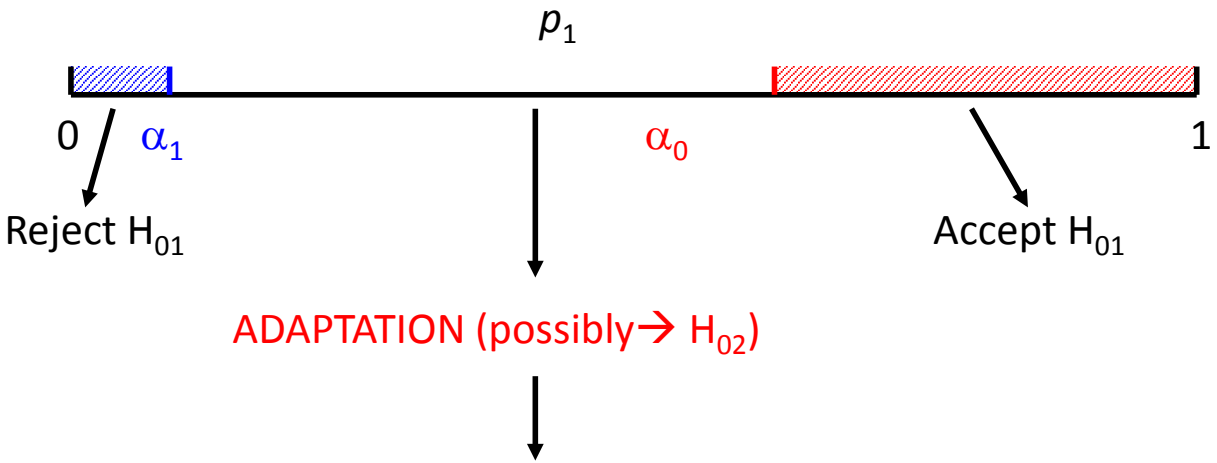
Conditional error and invariance principle

- Conditional error
  - Probability of a type I error at final analysis given what is observed at the IA
- Invariance principle
  - Any modification preserving the conditional error preserves the global type I error
- Methodology of adaptive designs
  - Replace the sequel of a trial by a design which, conditional on what has been observed, preserves the initial conditional type I error

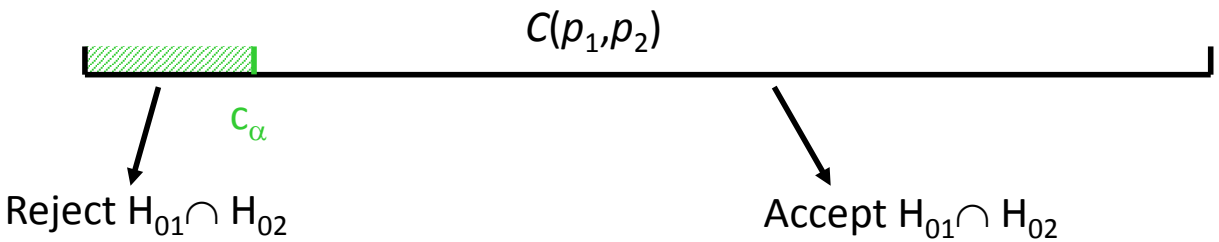
Combining different stages

Stage 1: null hypothesis H_{01}

n_1 patients → p-value p_1



Stage 2: n_2 patients, p-value p_2



Heuristics: from a sequential to an adaptive design

- Test $H_0 : \mu \leq 0$ vs. $H_1 : \mu > 0$
- Working model:
  - μ = mean of a Gaussian variable
  - Variance σ^2 known, equal to 1

Stage 1

n_1 observations, $z_1 = \sqrt{n_1}\bar{x}_1$

- Reject H_0 if $z_1 \geq z_{\alpha_1}$
- Stop for futility si $z_1 < z_{\alpha_0}$

Stage 2

n_2 observations, mean of the $(n_1 + n_2)$, $\bar{x}$

Reject H_0 if $z = \sqrt{n_1 + n_2}\bar{x} \geq z_{\alpha_2}$

$\Leftrightarrow w_1 z_1 + w_2 z_2 \geq z_{\alpha_2}$,

with $w_i = \sqrt{\frac{n_i}{n_1+n_2}}$ and $z_2 = \sqrt{n_2}\bar{x}_2$

- With
  - n_1 and n_2 prespecified
  - α_0, α_1, α_2 determined to control the global type I error rate

From a sequential to an adaptive design (cont'd)

- Interim analysis: adapt $n_2 \rightarrow \tilde{n}_2$
- If we decide to reject H_0 if

$$\tilde{z} = w_1 z_1 + w_2 \tilde{z}_2 \geq z_{\alpha_2}$$

$$\text{with } \tilde{z}_2 = \sqrt{\tilde{n}_2}\bar{x}_2$$

- Then the global level of the test is α provided weights w_i are those defined at the beginning
- i.e. with the "original" n_1 and n_2
- Combination test: tests statistics were combined with prespecified rule

Combination test

- Combine the results of the different stages
- Combine the test statistics (previous slide)
- Or combine *p*-values
- Many combination functions possible
  - Fisher's product test: $C(p_1, p_2) = p_1 \times p_2$
  - Weighted inverse normal combination: $C(p_1, p_2) = 1 - \Phi[w_1 \Phi^{-1}(1 - p_1) + w_2 \Phi^{-1}(1 - p_2)]$, with $0 < w_i < 1$ et $w_1^2 + w_2^2 = 1$

Conditions

- The combination rule has to be fixed in advance
- *p*-values must be "p-clud"

$$\Pr_{H_0}(p_1 \leq \alpha) \leq \alpha \text{ et } \Pr_{H_0}(p_2 \leq \alpha | p_1) \leq \alpha, \quad \forall \alpha \in [0,1]$$

- If p_1 and p_2 are independent and normally distributed, they are p-clud
- Determine decision boundaries to control α

$$\alpha_1 + \int_{\alpha_1}^{\alpha_0} \int_0^1 \mathbf{1}_{[C(x,y) \leq c_{\alpha_2}]} dxdy = \alpha$$

Conditional error function

- Another equivalent concept
- Reject $C(p_1, p_2) \leq c$
- Or reject if $p_2 \leq A(p_1)$
- Where $A(.)$ is the conditional error function
- Working example: reject if $\tilde{z}_2 \geq \frac{z_{\alpha_2} - w_1 z_1}{w_2} = z_{A(z_1, \alpha_2)}$

Multiple testing

- Previous phase 2–3 trial: several hypotheses tested
- Let's note these null hypotheses $H_1, \ldots, H_k$
- Strict control of α
  - Familywise error rate (FWER)
  - Maximum probability to reject at least one of the true H_i's
- Closed testing procedure to control α

Closed testing procedure

- For a given H_i
  - Define all the sub-hypotheses $H_{\mathcal{S}} = \cap_{\mathcal{S}} H_j$ that include H_i
  - Test each of the $H_{\mathcal{S}}$'s with a test of level α
  - Reject H_i iff all $H_{\mathcal{S}}$'s are rejected
  - Strict control of the global type I error rate
- The tests for the different hypotheses may not be the same, only α matters
- Case of two-stage adaptive designs
  - Combination test for each hypothesis
  - If one dose is dropped, p_2 only uses data for the remaining arms

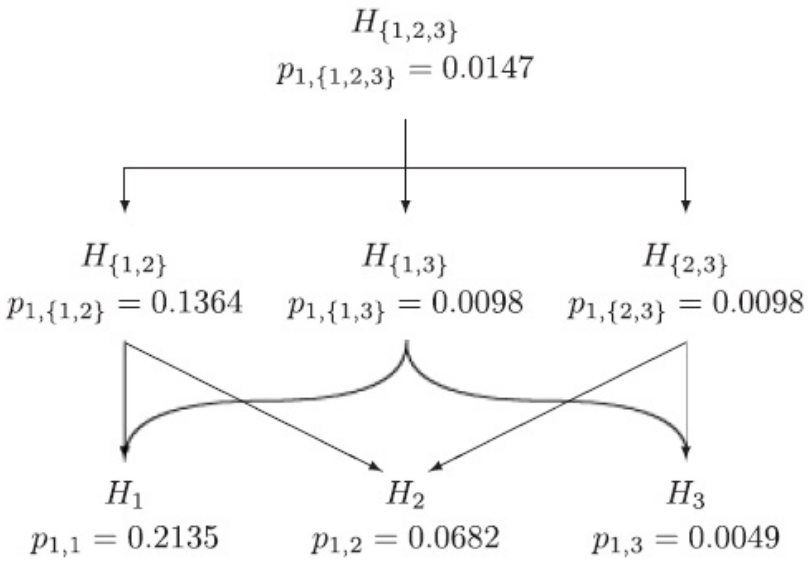
Example: Phase 2–3 trial[6]

- 3 doses and one placebo; 1 dose to be selected for further investigation
- Gaussian outcome with SD $\sigma = 6$
- $n = 142$ / group, IA at $n_1 = 71$
- H_i: $\mu_i \leq \mu_0 \; \forall i = 1, 2, 3$ (μ_0 for placebo)
- Combination test: Weighted inverse normal combination with weights $\sqrt{1/2}$ ($n_1 = n_2$)
- OBF: $\alpha_0 = 0.1$, $\alpha_1 = 0.0054$, $\alpha = 0.025$ and $c = 0.0359$
- Confirmatory trial: first test the global null $H_{\{1,2,3\}}$ with Bonferroni correction

[6] Bretz et al., *Stat Med* 2009

Example: Interim analysis

- Results: $p_{1,1} = 0.2135$, $p_{1,2} = 0.0682$, $p_{1,3} = 0.0049$
- Bonferroni correction: $p_{1,\{i,j\}} = 2\min(p_{1,i}, p_{1,j})$ et $p_{1,\{1,2,3\}} = 3\min(p_{1,1}, p_{1,2}, p_{1,3})$



Interpretation

- $p_{1,\{1,2,3\}} > \alpha_1 \rightarrow$ no early rejection
- $p_{1,\{1,2,3\}} < \alpha_0 \rightarrow$ the trial continues
- $p_{1,\{1,2\}} > \alpha_0 \rightarrow$ accept $H_{\{1,2\}}$, H_1 et H_2
- Only the dose 3 (and placebo) are continued

Example: Final analysis

- We obtain $p_{2,3} = 0.0296$ (other doses stopped)
- $p_{2,3}$ is the second-stage p-value for $H_{\{1,2,3\}}$, $H_{\{1,3\}}$, $H_{\{2,3\}}$, and H_3
- Combination test
  - $C(p_{1,\{1,2,3\}}, p_{2,3}) < c$
  - $C(p_{1,\{1,3\}}, p_{2,3}) < c$
  - $C(p_{1,\{2,3\}}, p_{2,3}) < c$
  - $C(p_{1,3}, p_{2,3}) < c$
- We can thus reject H_3
- We conclude at the superiority of dose 3 over placebo

Power in complex situations

- Up to now sample size to demonstrate one single effect (only one hypothesis)
- If several hypotheses, several choices for the power
  - Probability to reject at least one false H_i ($\mu_i > \mu_0$)
  - Probability to reject all false H_i's
  - Probability to reject the H_i corresponding to the best dose
- But "best" could involve an efficacy–tolerance trade-off ...
- Envisage several definitions and scenarios to power the study $\rightarrow$ simulations

Conditional power

- Like the conditional error, but under H_1
- Probability of rejection at the final analysis given p_1
- Useful for
  - Decision (early stopping, . . .)
  - SSR
  - Other adaptations
- Computing $\tilde{n}_2$: $CP(z_1) = 1 - \Phi\left[(z_{\alpha_2}\sqrt{n_1 + n_2} - z_1\sqrt{n_1})/\sqrt{n_2} - \frac{\Delta\sqrt{\tilde{n}_2}}{\sqrt{2}}\right]$
- What should we take for Δ?
  - $\Delta = d_1$ (*predictive power*) $\rightarrow$ could be inefficient
  - $\Delta = \Delta_0$ (*conditional power*)
  - A combination of both
  - Bayesian predictive power

Issues for inference

- Up to now the methods presented focused on the control of the type I error rate
- Most adaptive designs methods were first targeting testing rather than estimation
- That remains a field for research
- Especially for confidence intervals

Point estimates

- The MLE is typically biased for the mean
- The bias depends on the alternative hypothesis, the stopping rules and the adaptation rules $\rightarrow$ unknown in practice
- Unbiased mean estimators exist but they are generally inefficient
- More efficient unbiased median estimators exist
- Even more sever issues after treatment selection
- UMVCUE can be found
  - Bias(UMVCUE) = 0 < Bias(MLE) but MSE(UMVCUE) > MSE(MLE)
  - Choice on a case-by-case basis

Bayesian methods

- Less (almost never?) used for confirmatory trials
- More frequent in earlier phases trials
  - CRM
  - Phase 2 trials
- Methods also exist for phase 2–3 and phase 3 trials
- Even mixing Bayesian methodology with frequentist testing to show a control of the type I error rate

Outline

Why choose an adaptive design?[7]

- Obtain the same information as with a classical design, but with an increased efficiency
- Increase the probability to attain the trial's objectives
- Improve the knowledge about the treatment
- But also
  - May shorten the drug development
  - Conceptually attractive

[7] Guidance for Industry Adaptive Design Clinical Trials for Drugs and Biologics, FDA, Draft 2010

Constraints to be taken into account

- Regulatory
  - Authorization
  - Maintaining the 'confirmatory' nature (seek formal statistical advice)
- Logistics
  - For all these designs, except phase 1 and 1–2
- Benefit/constraints or benefit/risk balance according to development phase or objectives
- Other constraint
  - Need of an 'expert' statistician

References I


Bretz F, Koenig F, Brannath W, Glimm E, Posch M. Adaptive designs for confirmatory clinical trials. *Statistics in Medicine* 2009; **28**:1181–1217. URL http://dx.doi.org/10.1002/sim.3538.


Bauer P, Bretz F, Dragalin V, König F, Wassmer G. Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. *Statistics in Medicine* 2015; :n/a–n/aURL http://dx.doi.org/10.1002/sim.6472.


O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* 1990; :33–48.


Garrett-Mayer E. The continual reassessment method for dose-finding studies: a tutorial. *Clinical Trials* 2006; **3**:57–71.


Thall PF, Russell KE. A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics* 1998; **54**:251–264.


Thall PF, Cook JD. Dose-finding based on efficacy–toxicity trade-offs. *Biometrics* 2004; **60**:684–693. URL http://dx.doi.org/10.1111/j.0006-341X.2004.00218.x.

References II


Zohar S, Chevret S. Recent developments in adaptive designs for phase I/II dose-finding studies. *Journal of Biopharmaceutical Statistics* 2007; **17**:1071–1083.


Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Statistics in medicine* 2011; **30**:3267–84.


Posch M, Bauer P, Brannath W. Issues in designing flexible trials. *Statistics in Medicine* 2003; **22**:953–969. URL `http://dx.doi.org/10.1002/sim.1455`.


Friede T, Parsons N, Stallard N, Todd S, Valdes Marquez E, Chataway J, Nicholas R. Designing a seamless phase II/III clinical trial using early outcomes for treatment selection: An application in multiple sclerosis. *Statistics in Medicine* 2011; **30**:1528–1540. URL `http://dx.doi.org/10.1002/sim.4202`.


Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, Racine-Poon A. Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Statistics in Medicine* 2009; **28**:1445–1463. URL `http://dx.doi.org/10.1002/sim.3559`.

References III


Friede T, Parsons N, Stallard N. A conditional error function approach for subgroup selection in adaptive clinical trials. *Statistics in Medicine* 2012; **31**:4309–4320. URL http://dx.doi.org/10.1002/sim.5541.


Mehta C, Schäfer H, Daniel H, Irle S. Biomarker driven population enrichment for adaptive oncology trials with time to event endpoints. *Statistics in Medicine* 2014; **33**:4515–4531. URL http://dx.doi.org/10.1002/sim.6272.


Brannath W, König F, Bauer P. Estimation in flexible two stage designs. *Statistics in Medicine* 2006; **25**:3366–3381. URL http://dx.doi.org/10.1002/sim.2258.


Brannath W, Mehta CR, Posch M. Exact confidence bounds following adaptive group sequential tests. *Biometrics* 2009; **65**:539–546. URL http://dx.doi.org/10.1111/j.1541-0420.2008.01101.x.


Bowden J, Glimm E. Conditionally unbiased and near unbiased estimation of the selected treatment mean for multistage drop-the-losers trials. *Biometrical Journal* 2014; **56**:332–349. URL http://dx.doi.org/10.1002/bimj.201200245.