Design klinischer Studien mit Schwerpunkt frühe Phase II Studien - Dos and Don'ts

Dr. Dirk Hasenclever

Institute for Medical Informatics, Statistics & Epidemiology (IMISE),
and
Clinical Trial Centre (ZKS),
University Leipzig

dirk.hasenclever@imise.uni-leipzig.de

SaxoCell Clinics Workshop - Klinische Studien mit ATMPs

2023-03-16 Leipzig

What is the role of Phase II studies?

Public funding?

Roadmap for the development of a new standard therapy

- **Phase 0**: Experiments with cell cultures and animals
- **Phase 1**: First in men
  - Tolerability
  - PK
- **Phase 2:**
  - Finetuning the intervention
  - Preliminary evidence for short term efficacy
- **Phase 3:** Large definite trial demonstrating superiority over standard treatment in clinically relevant endpoint
- **New standard Therapy**

Public funding for early trials in Germany BMBF

Bundesministerium für Bildung und Forschung

Richtlinie zur Förderung von frühen klinischen Studien

Ziel dieser Fördermaßnahme ist es, bestehende Barrieren in der Translationskette zwischen präklinischer und klinischer Forschung in Deutschland zu überwinden und die Erfahrung und Expertise in diesem Forschungsbereich in Deutschland auszubauen. Vielversprechende Forschungsansätze und neue Wirkstoffanwendungen sollen, in größerem Maße als bislang, frühzeitig identifiziert und konsequent weiterverfolgt werden. Die Ziele der Fördermaßnahme werden erreicht, wenn durch die Erhöhung der Zahl an frühen klinischen Studien bis Phase II vermehrt vielversprechende, hoch qualitative Erkenntnisse aus der präklinischen Forschung in spätere Phasen der klinischen Forschung überführt werden.

https://www.gesundheitsforschung-bmbf.de/de/12905.php

Public funding for early trials in Germany
DFG

https://www.dfg.de/formulare/17_01/index.jsp

Deutsche Forschungsgemeinschaft
Kennedyallee 40 · 53175 Bonn, Germany · Postal address: 53170 Bonn, Germany
Tel.: + 49 228 885-1 · Fax: + 49 228 885-2777 · postmaster@dfg.de · www.dfg.de

DFG

The aim of a feasibility study is to provide initial evidence of the efficacy of a method as well as to examine the feasibility of a subsequent interventional trial. The method can be therapeutic, diagnostic or prognostic in nature. In terms of feasibility, possible reasons for applying for funding include the validation of the intervention (e.g. dose finding), the operationalisation of the endpoints, the assessment of the effect size and the sample size, the assessment of the practicability of the randomisation procedure, and the investigation of the planned study design. The proposal must clearly show how the feasibility study will serve as the lead-up to a subsequent interventional trial. Feasibility studies must be prospective and include a control group, a randomised allocation of patients, and sample size calculations. Monocentric studies are possible.

Reviewer's check list for Phase I/II studies

- **Plausibly argued biological-medical rationale?**

- **Clearly stated study question?**
- **Appropriate primary endpoint(s)?**
  - Multiplicity issues?
- **Adequate design** controlling **sources of bias**?
- **Sufficient sample size** to achieve a **meaningful evidencial signal**?

- **Right study question on a roadmap of steps** to a definite Phase III study?

Wide spectrum of questions

~~**Schaun'mer 'mal?**~~ is **not a scientific question!**

What is the question?
Types of questions and corresponding endpoints

- **Don't be lazy: Disentangle the bundle of your questions!**
  - Safety?
  - Confirmation of biological rationale?
  - Finetune the intervention?
  - Clinical feasibility?
  - Preliminary clinical efficacy?

  - Other planning information for next step?
    - E.g. endpoint validation

Write it up and get it right!

- All your questions should be addressed and detailed in the study protocol.
- Make sure your study sufficiently answers your questions – vague information is not enough!
- Prioritise if necessary! First steps first!
- Your study is early phase and has a limited time horizon.

  → Focus on those questions that can be answered within this setting

Safety

- Your intervention has
  - Known potential side effects and
  - Potential unknown side effects
- For **known potential side effects**
  - **define AEs of special interest**
  - make sure these are observed and documented
  - Consider what level/grade would still be acceptable
  - Consider defining a respective alarm trigger/stopping rule
- For **potential unknown side effects**, watch out and document general AEs.

Confirmation of biological rationale

- **Reviewer/Sceptic**: "**Biopoetry?**" - **Or sound rationale?**

- **Consider adding biological endpoints** to confirm your biological model of the intervention in a clinical setting.
  - Can you measure **what is important for your theory**?
  - Don't measure what you can measure

    because you can measure it,

    if it is not what you need
- Regulatory problem: If you use fancy new methods, you will be asked for validation. -> experimental endpoint(s)

Finetuning the intervention

- Is the **intervention already well defined**?

- If NO
  - it is a early Phase 2a study
  - Showing efficacy is not the priority

- **Issues**
  - **Dose finding**: maximum tolerable dose
  - **Heterogeneity** in PK?
  - **Intervention modification rules**

Finetuning the intervention
Dose finding

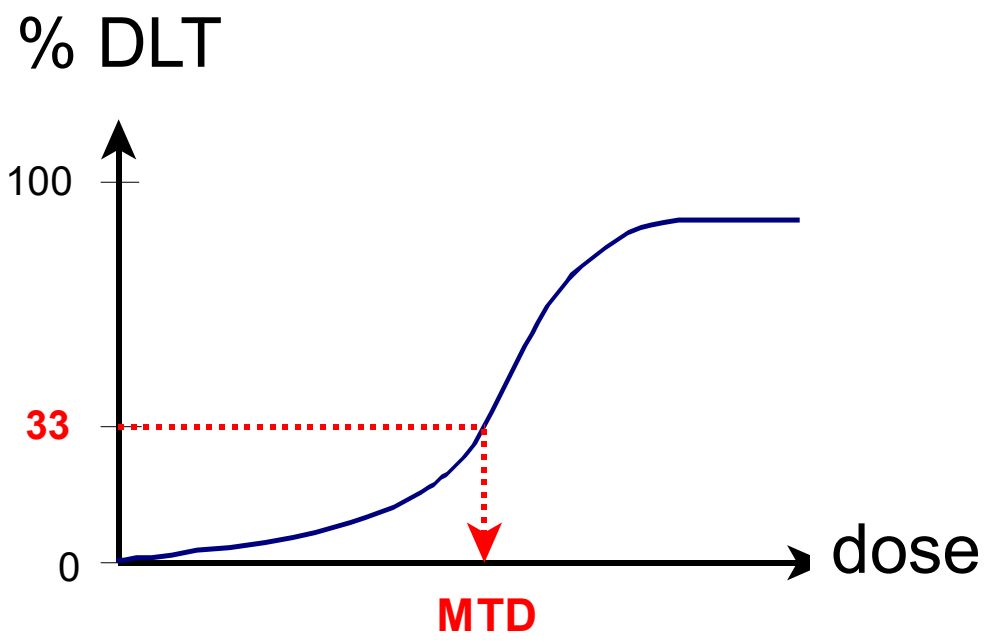
- Dose finding: maximum tolerable dose
  - If toxicity profile unclear:
    - only by “clinical judgement”
    - maximum tolerable dose is ill defined
    - Question ill defined – stop?
  - If dose limiting toxicities are known
    - Define specific AEs of special interest
    - Define what grade of expected toxicity just acceptable in a certain proportion of patients

- Various study designs
    - 3+3, up and down, Continuous reassessment

Dose finding designs

Steps to specify a dose finding question

- Define the maximum tolerable dose (MTD)
  - Acceptable grades in ~ 33% patients
    - E.g. severe neutropenia for more than 4 days
- If multiple cycle intervention:
  - Short-term reversable tox: e.g. blood counts
  - Cumulative toxicity: e.g. neurotoxicity
- Time horizon: first cycle only or whole intervention



Steps to specify a dose finding question II

- Define the dose levels
  - Often equidistant on logarithmic scale (or Fibonacci)
- Choose a starting dose
  - Low for safety
  - Not too low for efficacy
- Dose allocation algorithm
  - Choosing the dose for the next patients
    based on toxicity results up to inclusion
- Estimate the MTD with sufficient precision

3+3 design
often used – pretty much voodoo…

# patients	# DLTs observed	Action
3	2+	STOP
3	1	three more!
3	0	next level up
6	2+	STOP
6	1	next level up

Up and down design
rarely used – better, but more patients

# patients	# DLTs observed	Action for next 3 patients
3	0	three more on level +1!
3	1	three more on same level!
3	2	three more on level -1!

Even better: **Continous Reassessment Method (O‘ Quigley, 1990)**

Dealing with patient heterogeneity

Finetuning the intervention Heterogeneity

- Heterogeneity in PK?
  - Example: Carboplatin dosing depends on renal function to achieve a uniform AUC.

  - If substantial heterogeneity seen in PK / toxicity,
    - **Understand!**
    - **Predict! and**
    - **adjust dosing!**

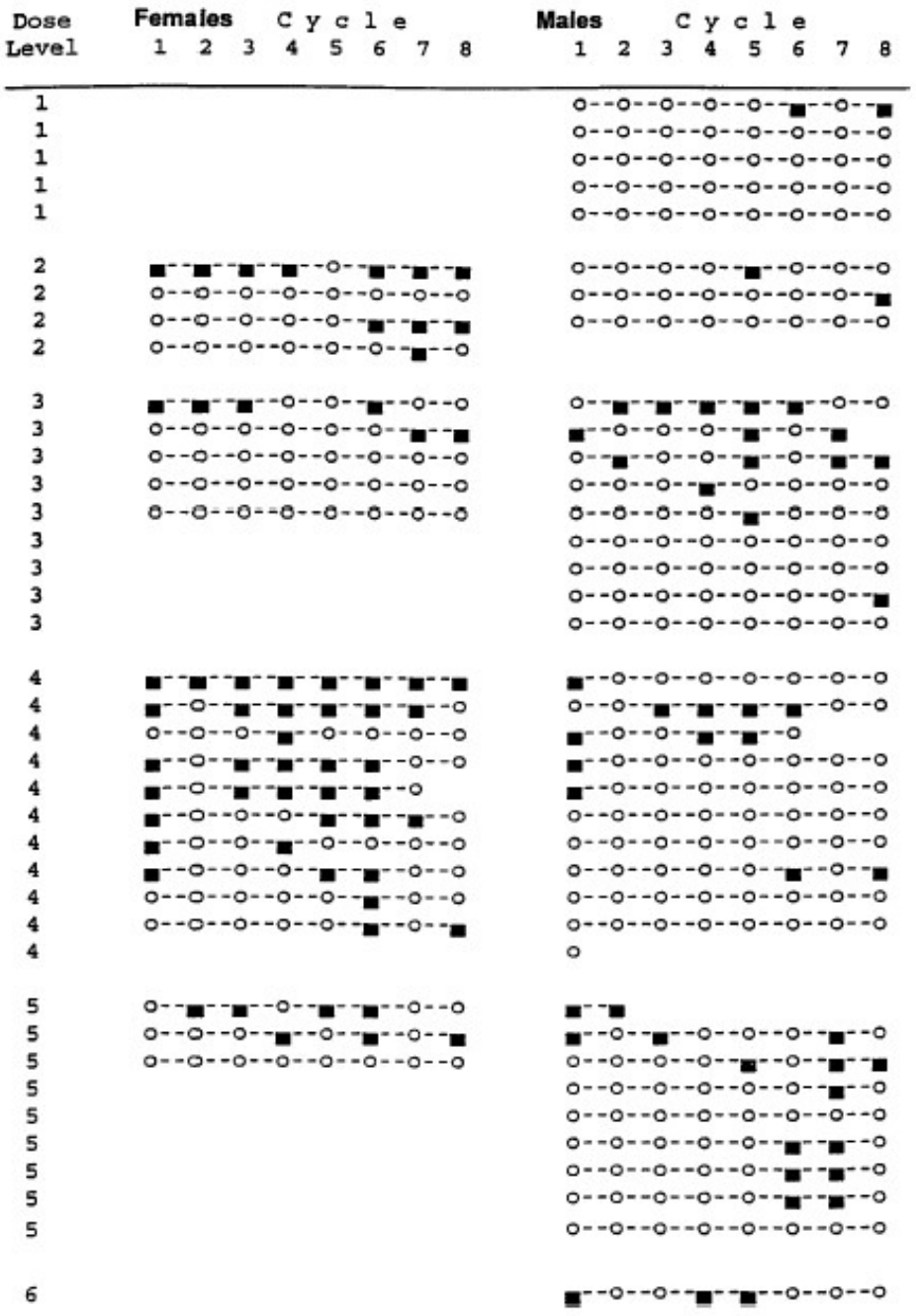


Fig 3. Occurence of toxic events in individual patients treated at different dose levels. (■) Toxic response; (○) no toxicity.

Tesch H, 1998 Blood

Unexplained Heterogeneity in toxicity disposition

between

Patients

by sex

Encountered in a dose escalation trial for

an 8 cycle polychemotherapy

Finetuning the intervention
Intervention modification rules

- **Intervention modification rules for multi cycle therapies**
  - Intra-individual, toxicity dependent
  - dose reduction or escalation
  - Lengthening of time between cycles etc..
- Depends on your control of PK/Tox heterogeneity
  - No tox – underdosed?
  - Make sure each patient gets effective dose
- Consider general de-escalation or escalation strategy

- CAVE: phrase dose modification in protocol as guidance not as algorithmic rule whenever feasible to avoid a tsunami of monitoring findings…

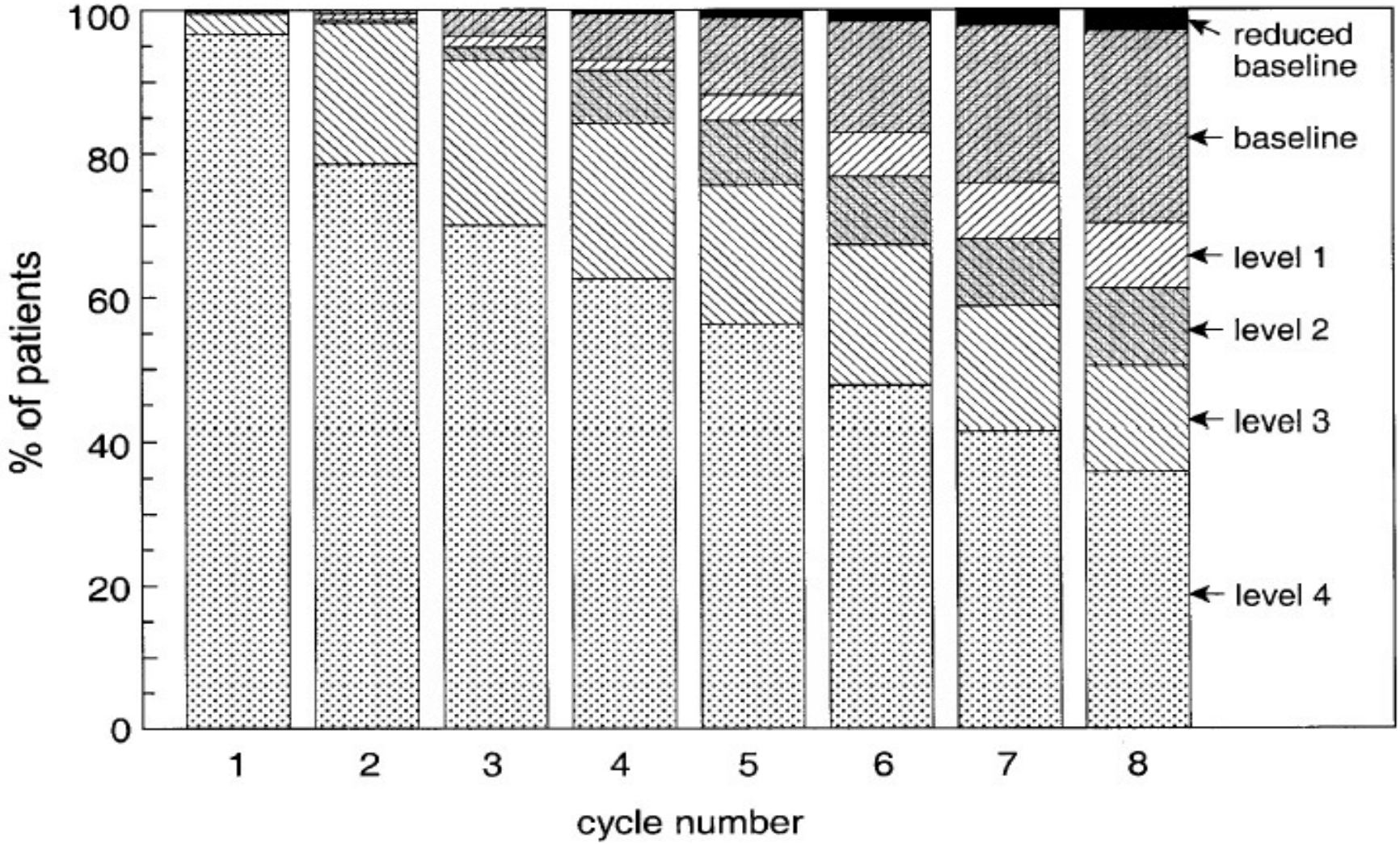


Figure 2. Deescalation of BEACOPP-escalated over the successive therapy cycles. Stacked bars show the percentages of patients being treated at the specified dose levels.

Engel C. 2000, Ann. Oncol.

Clinical feasibility

- Is the treatment program feasible?
  - Overall summary endpoint: e.g.
    - Rate treatment fully delivered
    - Time on treatment etc.
  - Are there particular risks
    - critical processes,
    - logistics,
    - time lines?
  - Define specific feasibility endpoints
    - E.g. Waiting time for treatment relevant analyses
    - E.g. Rate ex-vivo produced specific product ready before disease progression

Surrogate efficacy endpoints

Clinical efficacy

- **Most phase II endpoints are surrogates!**
- A **surrogate-endpoint** is
  - a lab parameter,
  - a biological marker, or
  - a histopathological or a
  - clinical assessment = “response”

  used as **short-term substitute** for a **clinically relevant long-term endpoint**.

Valid surrogate endpoints

- **Two statistical criteria**

  - High correlation between surrogate and relevant endpoint

  AND

  - Treatment effects in surrogate correlate with treatment effects on relevant endpoint

Example: DFS is a valid surrogate for OS in metastatic colon cancer

Disease-Free Survival Versus Overall Survival As a Primary End Point for Adjuvant Colon Cancer Studies: Individual Patient Data From 20,898 Patients on 18 Randomized Trials

J Clin Oncol 23:8664-8670.

Daniel J. Sargent, Harry S. Wieand, Daniel G. Haller, Richard Gray, Jacqueline K. Benedetti, Marc Buyse, Roberto Labianca, Jean Francois Seitz, Christopher J. O'Callaghan, Guido Francini, Axel Grothey, Michael O'Connell, Paul J. Catalano, Charles D. Blanke, David Kerr, Erin Green, Norman Wolmark, Thierry Andre, Richard M. Goldberg, and Aimery De Gramont

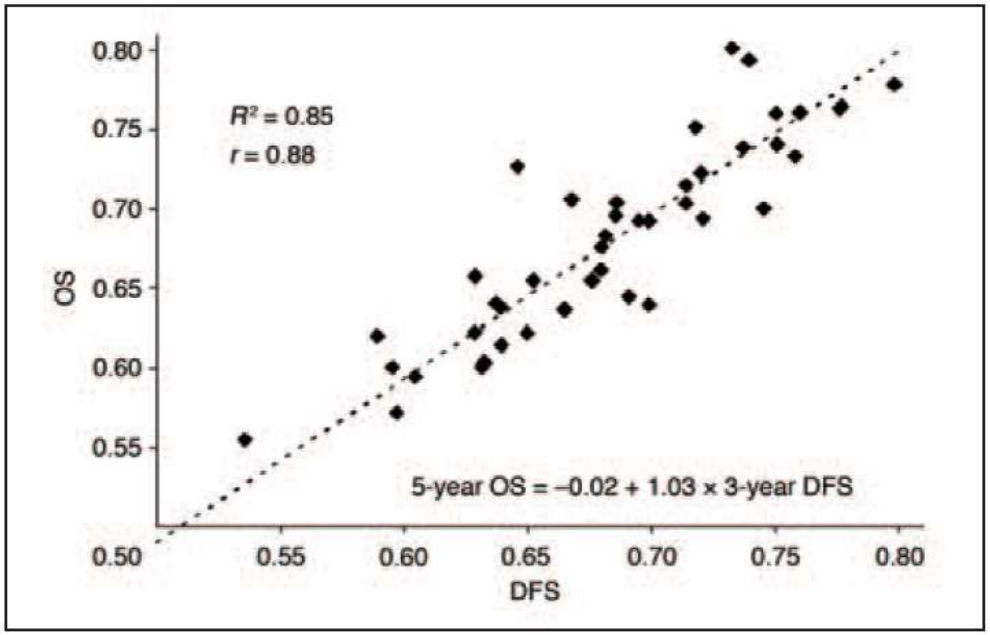


Fig 1. Three-year disease-free survival (DFS) versus 5-year overall survival (OS) by study arm.

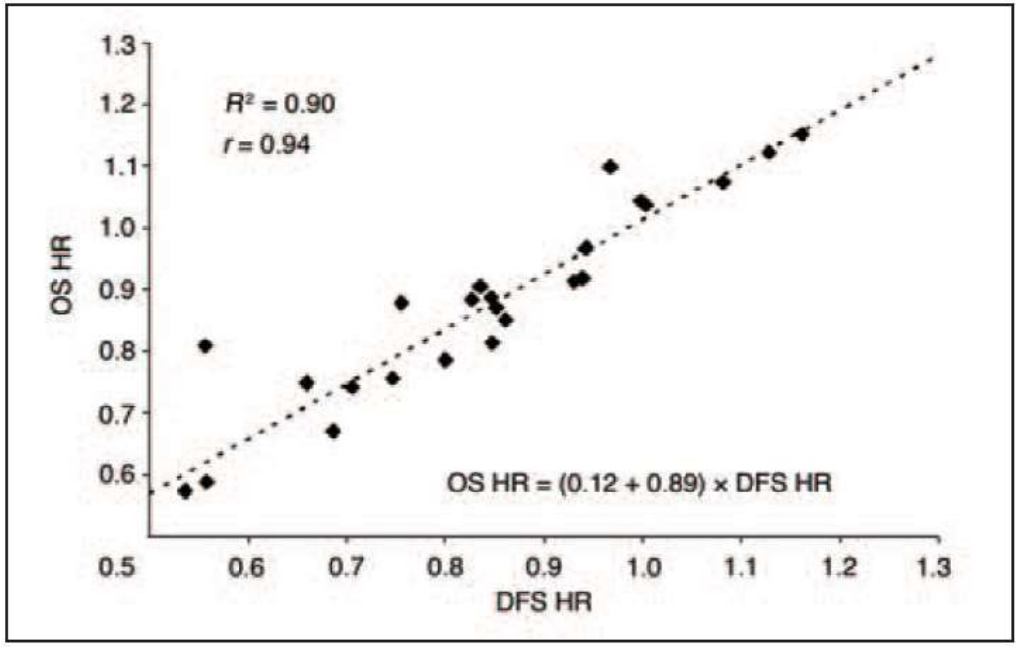
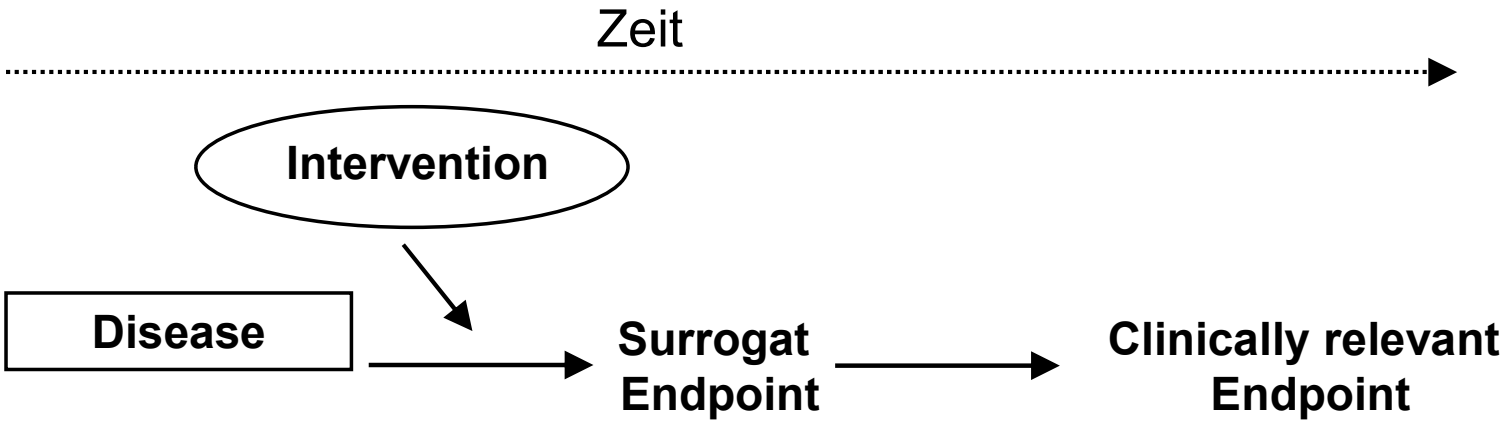


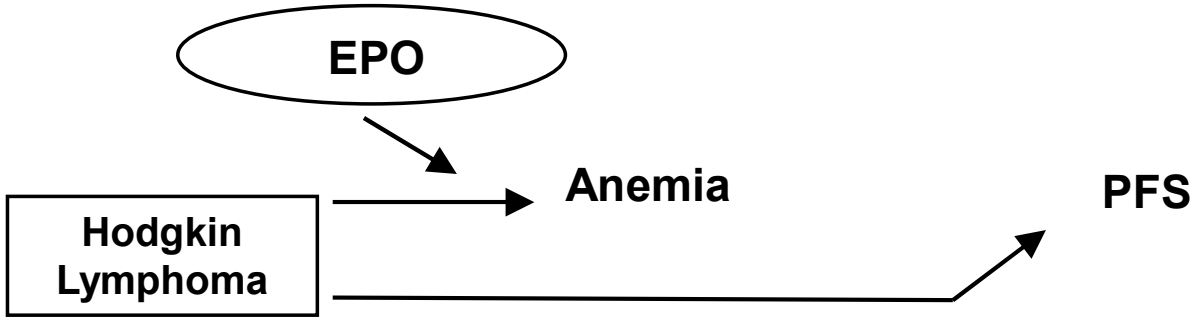
Fig 2. Disease-free survival (DFS) versus overall survival (OS) hazard ratios (HR) by trial.

Think in causal models!

- **Validity** given if surrogate endpoint is a mediator.



- **Nonsense** if e.g. surrogate endpoint is just a symptom.

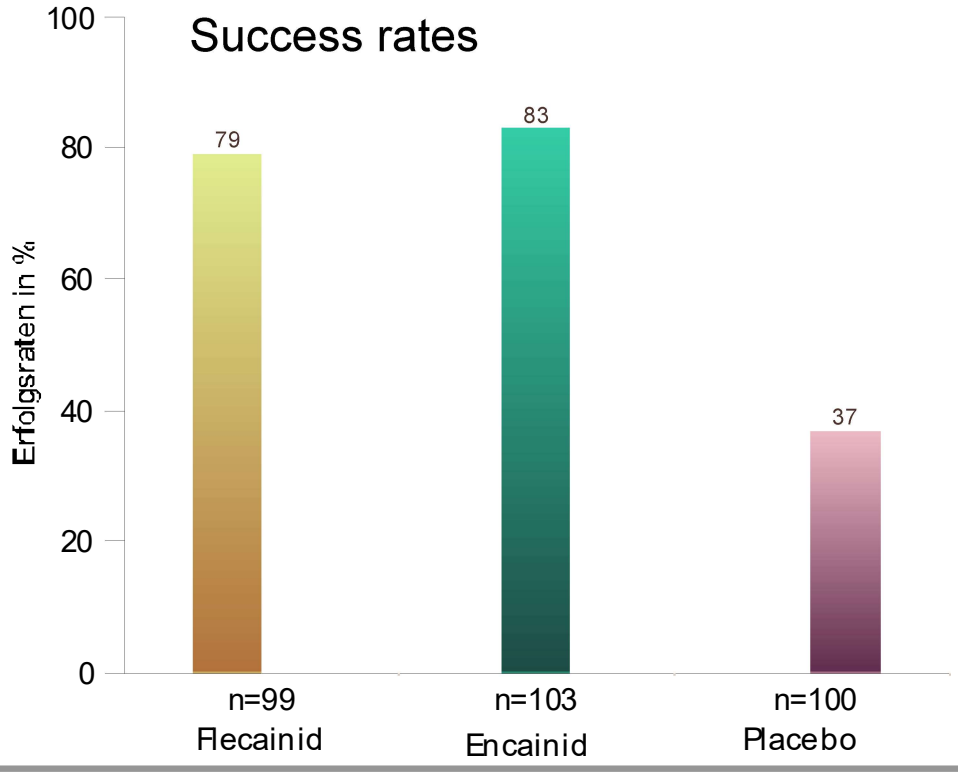


Disaster with surrogate endpoints I

Drug treatment of **arrhythmia after myocardial infarction.**

Pilot study CAPS:

=> Can drugs suppress **arrhythmia ?**



VES: ventricular extrasystole
VT: ventricular tachycardia

Erfolg: VES ⇓ 70% und VT ⇓ 90%

Am J Cardiol 1988; 61: 501-509

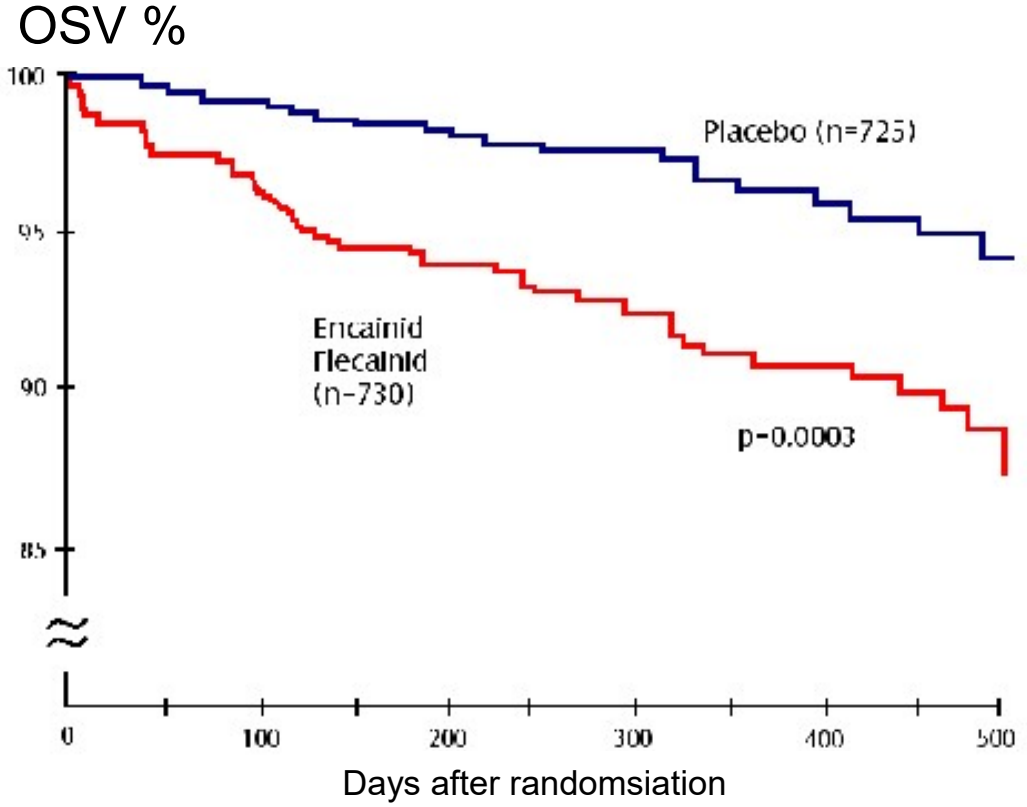
➔ licenced for large market

Disaster with surrogate endpoints II

Drug treatment of **arrhythmia after myocardial infarction.**

CAST-Studie:

=> **Do anti-arrhythmic drugs prolong survival?**



Planned:

4.400 patients in 3 years

➔ **Study was stopped at the first interim analysis after 1.5 years.**

N Engl J Med 1989; 321: 406-412

To randomise or not to randomise…

Specify the type of evidence you want to obtain

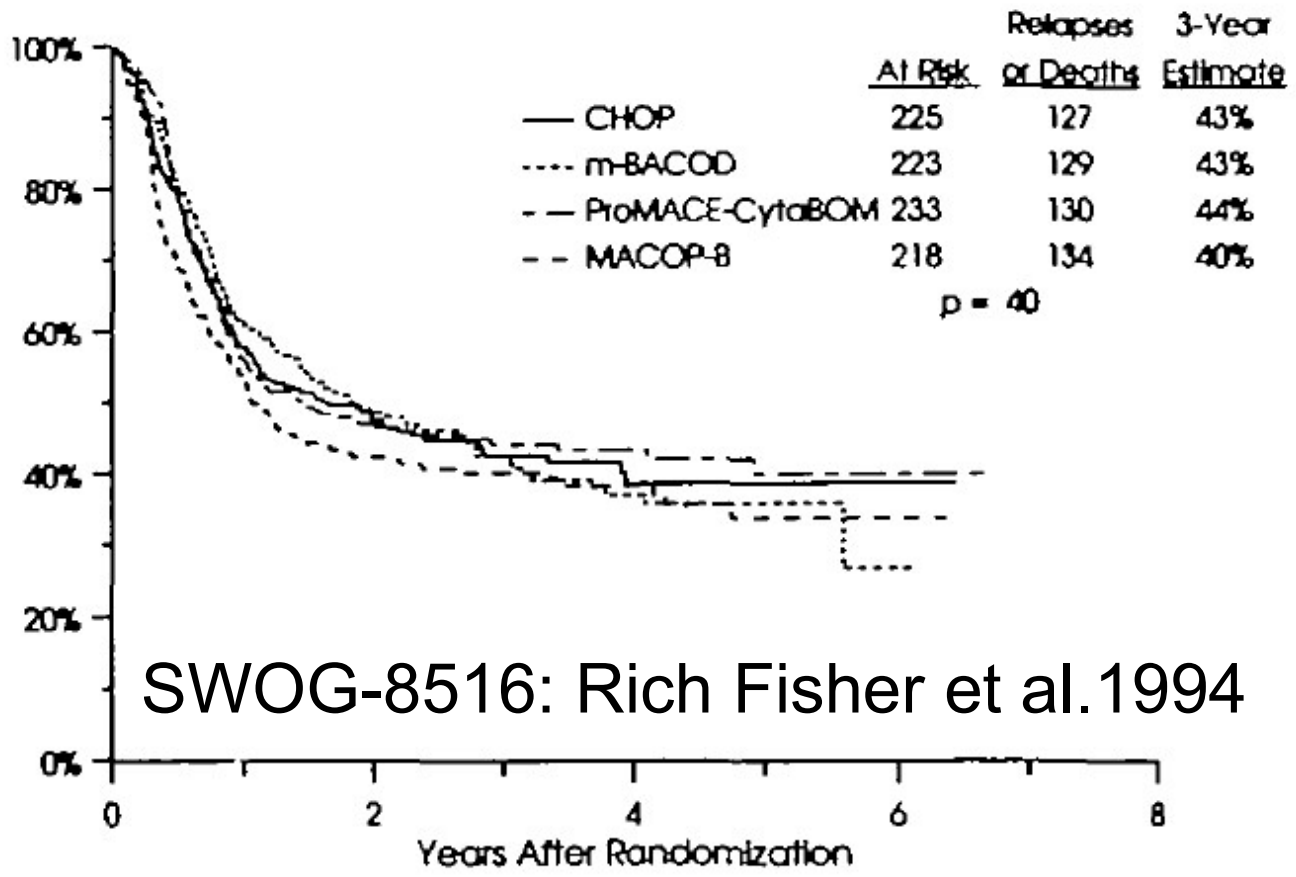
Phase I/II studies are steps to a definitive Phase III trial

- Only in very rare cases Phase II studies already **deservedly change the clinical standard**
- Exceptionally if they are
  - Randomised **and**
  - Show a large effect in a validated surrogate endpoint and/or a relevant clinical endpoint

- In all other cases: **Resist the temptation** to claim so.
  - Deterrent examples in this talk.

Promising Uncontrolled Phase II trials may badly fail in Phase III

- DLBCL CHOP ~ 40%
- Several uncontrolled Phase II trials with very intensive chemotherapy regimen: ~60% ?
- Germany: Confirmative Phase III trial deemed unethical.
- USA: National Priority Study: Identical results with higher toxicity.



SWOG-8516: Rich Fisher et al.1994

Fig. 1. Time-to-treatment-failure curve for all eligible patients according to the randomized treatment.

Controlled or uncontrolled?

- Uncontrolled Phase II trials are OK as proof of principle studies where **a response in a potential control arm can be practically excluded a priori**.
- Uncontrolled trials are OK **if a control arm clearly does not add information**: e.g.
  - Tolerability of new drug (background AEs in control known to be limited)
  - Feasibility of new regimen
- Otherwise promising results from Phase II trials may be misleading due to **selection of a favourable study population**.
  - **If in doubt, randomise!**

Phase II trials require a clearly stated evidential objective

- **Phase II trials involve compromises** because they can only
  - look at short term surrogate endpoints and
  - sample size is generally limited.

- In Phase III, we require strong evidence.
- Phase II trials are rarely powered for strong evidence.
  - **Make clear what you can realistically expect from your trial.**

Example: Uncontrolled Phase II

Early clinical endpoint:

The trial is designed to test the
null hypothesis that the **response probability is ≤ 5%** against
the alternative hypothesis that the **response probability is ≥20%.**

Using a one-sided significance level of 2.5% and
requiring a power of 85% leads to a sample size of N=41
with the exact binomial test.

Biological endpoint:

With 41 patients, we will have > 85% power to detect
even a moderate effect size of 0.5 in the mean change in a biological marker
(paired t-test on the log scale) at a one-sided significance level of 2.5%.

External benchmark design
when standard therapy shows response?
Discouraged!

- **Experimental** : Standard therapy + **new drug**
- **Benchmark control:** Standard therapy

- **Objective** demonstrate a response rate higher than the assumed benchmark response rate.

- Marked biological heterogeneity -> **Substantial uncertainty on outcome with standard therapy:**
  - Company: Benchmark 50%
  - Registry data: perhaps 65-70%

- If say 65% response is observed, is drug **promising** or **useless**?
- **High risk of uninterpretable result**.
- Beware of selection effects. **Randomise**!

Example: Sample Size Discussion in controlled trial: Consider using strong and moderate evidential signal

From the **randomized** XYZ study we expect a response rate of 36% in the control arm.

For planning purposes we define **two alternative scenarios**:

- **Strong effect size** scenario: an improvement in response rate by an odds ratio of 0.316 corresponding to an **increase of 28%** from 36% to 64%.
- **Moderate effect size** scenario: an improvement in response rate by an odds ratio of 0.5 corresponding to an **increase of 17%** from 36% to 53%.

Example: Sample Size Discussion in controlled trial: Consider using strong and moderate evidential signal II

The study will randomize N=102 patients, such that accounting for about 5% drop-outs N=98 patients will be informative.
With this sample size

- Assuming a strong effect size, we will have 80% power to obtain a significant **one-sided test at a 2.5% significance level**.

- Assuming a moderate effect size, we will have 80% power to obtain a a significant **one-sided test at a 20% significance level**, corresponding to excluding equality from a one-sided 80% confidence interval for the difference.

A misleading phase II study…

Do not delude yourself… and others! (1)

N ENGL J MED 379;10 NEJM.ORG SEPTEMBER 6, 2018

Rituximab plus Lenalidomide in Advanced Untreated Follicular Lymphoma

F. Morschhauser, N.H. Fowler, P. Feugier, R. Bouabdallah, H. Tilly, M.L. Palomba, C. Fruchart, E.N. Libby, R.-O. Casasnovas, I.W. Flinn, C. Haioun, H. Maisonneuve, L. Ysebaert, N.L. Bartlett, K. Bouabdallah, P. Brice, V. Ribrag, N. Daguindau, S. Le Gouill, G.M. Pica, A. Martin Garcia-Sancho, A. López-Guillermo, J.-F. Larouche, K. Ando, M. Gomes da Silva, M. André, P. Zachée, L.H. Sehn, K. Tobinai, G. Cartron, D. Liu, J. Wang, L. Xerri, and G.A. Salles, for the RELEVANCE Trial Investigators*

Phase III

Objective: "90% power to detect a between-group **difference of 12 percentage points in the rate of confirmed Or unconfirmed complete response at 120 weeks** (72% in the rituximab–lenalidomide group vs. 60% in the rituximab–chemotherapy group), at a two-sided alpha level of 0.05."

Do not delude yourself… and others! (2)

120 week response

Variable	Rituximab–Lenalidomide Group (N=513)	Rituximab–Chemotherapy Group (N=517)	Hazard Ratio (95% CI)	P Value
Response status at 120 weeks, as assessed by independent review committee				
Overall response — no. (% [95% CI])	312 (61 [56–65])	336 (65 [61–69])		
Confirmed or unconfirmed complete response — no. (% [95% CI])	247 (48 [44–53])	274 (53 [49–57])		0.13

Oops, slightly worse!

IDMC who had to discuss stopping the trial because of
clear inferiority in 70 weeks response at an early interim analysis.

Phase III was based on a Phase II trial...
based on which L+R had already obtained FDA approval!

Do not delude yourself… and others! (3)

Lancet Oncol 2014; 15: 1311–18

Safety and activity of lenalidomide and rituximab in untreated indolent lymphoma: an open-label, phase 2 trial

Nathan H Fowler, R Eric Davis, Seema Rawal, Loretta Nastoupil, Fredrick B Hagemeister, Peter McLaughlin, Larry W Kwak, Jorge E Romaguera, Michelle A Fanale, Luis E Fayad, Jason R Westin, Jatin Shah, Robert Z Orlowski, Michael Wang, Francesco Turturro, Yasuhiro Oki, Linda C Claret, Lei Feng, Veerabhadran Baladandayuthapani, Tariq Muzzafar, Kenneth Y Tsai, Felipe Samaniego, Sattva S Neelapu**

Chemotherapy free treatment of indolent lymphoma!
Spectacular response rates!

(1) **Inflated response rate reported** based on evaluable patients (imaging for response assessment)
excluding patients with prior progression / withdrawal from the denominator. This borders at scientific fraud.

(2) **Favourably selected** patient population!

Phase II

FLIPI score	
0–1	11 (22%)
2	25 (50%)
3–5	14 (28%)

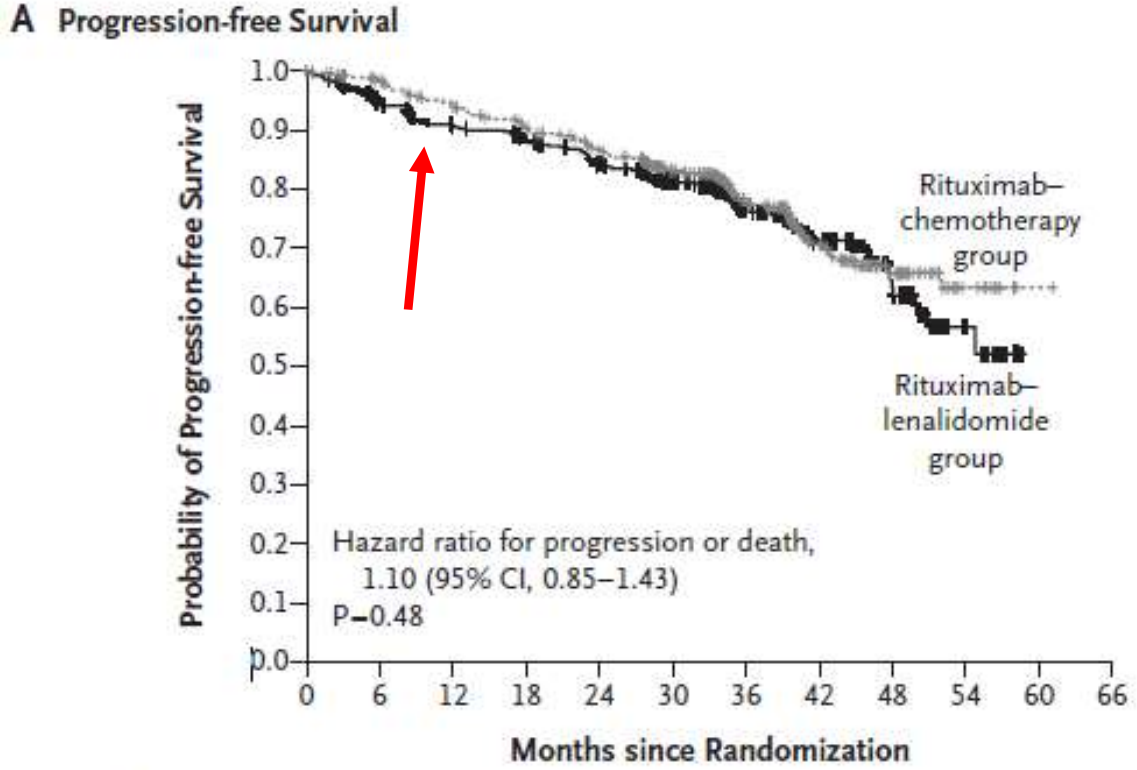
FLIPI score — no. (%)	
0 or 1	153 (15)
2	374 (36)
3 to 5	503 (49)

Phase III

Do not delude yourself… and others! (4)

The IDMC let the trial continue based on the indolent nature of the disease
– against the protocol.

The Phase III trial should have been designed to show non-inferiority in PFS.



Sample size considerations

Sample sizes

Basic facts of life

- You will die
- You have to pay taxes
- **You will hate the statistically required sample size**

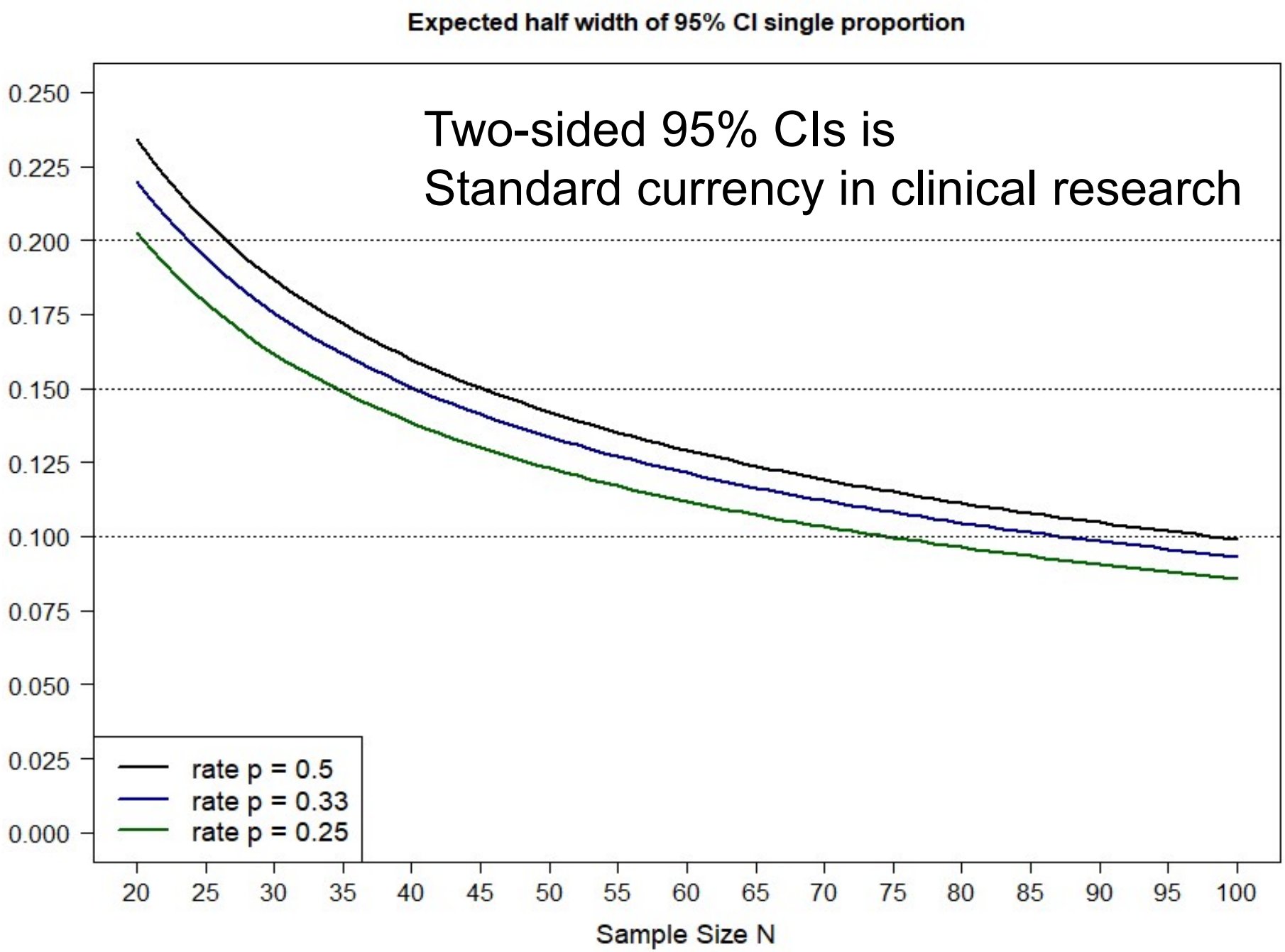
Do not shoot the biometrician!

Estimating a single probability

- Specify the **desired precision of the estimate**

- Precision ~ Expected half width
  of the two-sided confidence interval
  - ±20%, ±15%, ±10%?
- Evidence level: Confidence level
  - 95%, 90%, 80%?

- **N = 45-50 is often a reasonable order of magnitude**


Expected half width of 95% CI single proportion
Two-sided 95% CIs is
Standard currency in clinical research
0.250
0.225
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
rate p = 0.5
rate p = 0.33
rate p = 0.25
20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Sample Size N

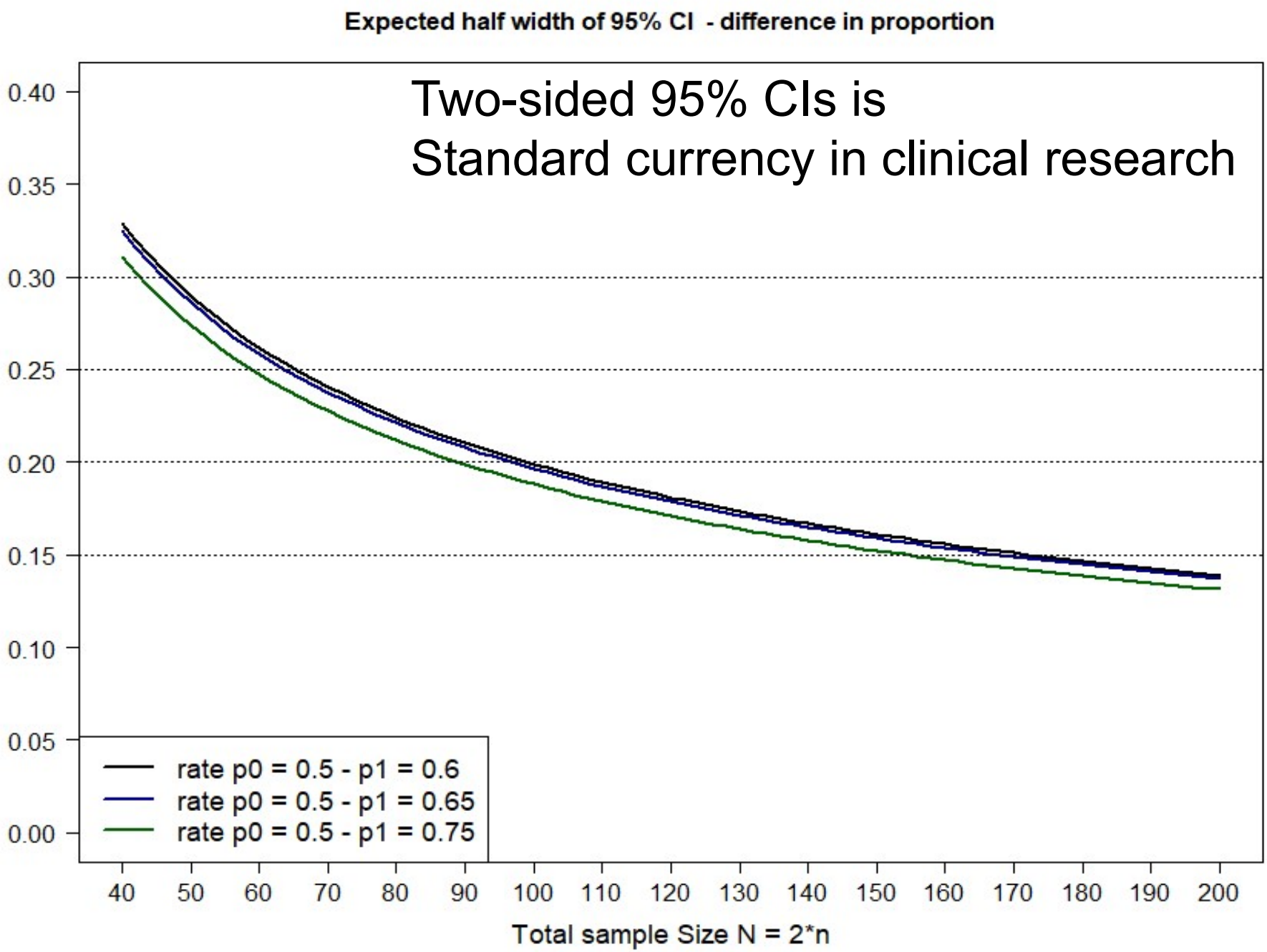

Expected half width of (1-alpha) - CI single proportion
Do you accept
weakening your standards and
your evidential requirement?
0.250
0.225
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
two-sided confidence = 0.95
two-sided confidence = 0.9
two-sided confidence = 0.8
20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Sample Size N

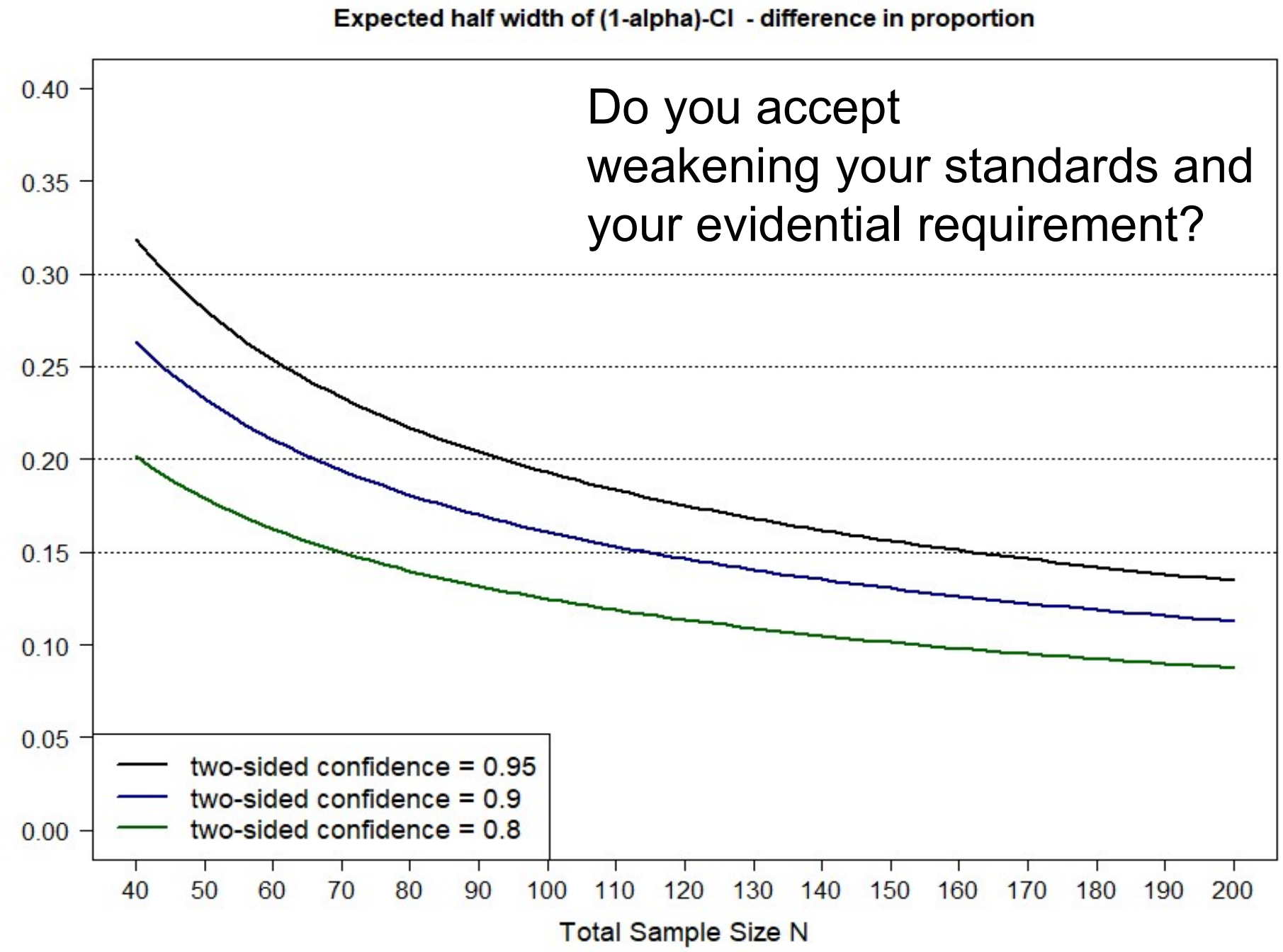
Estimating a treatment response rate difference - randomised two arm trial

- Specify the **desired precision of the estimate**

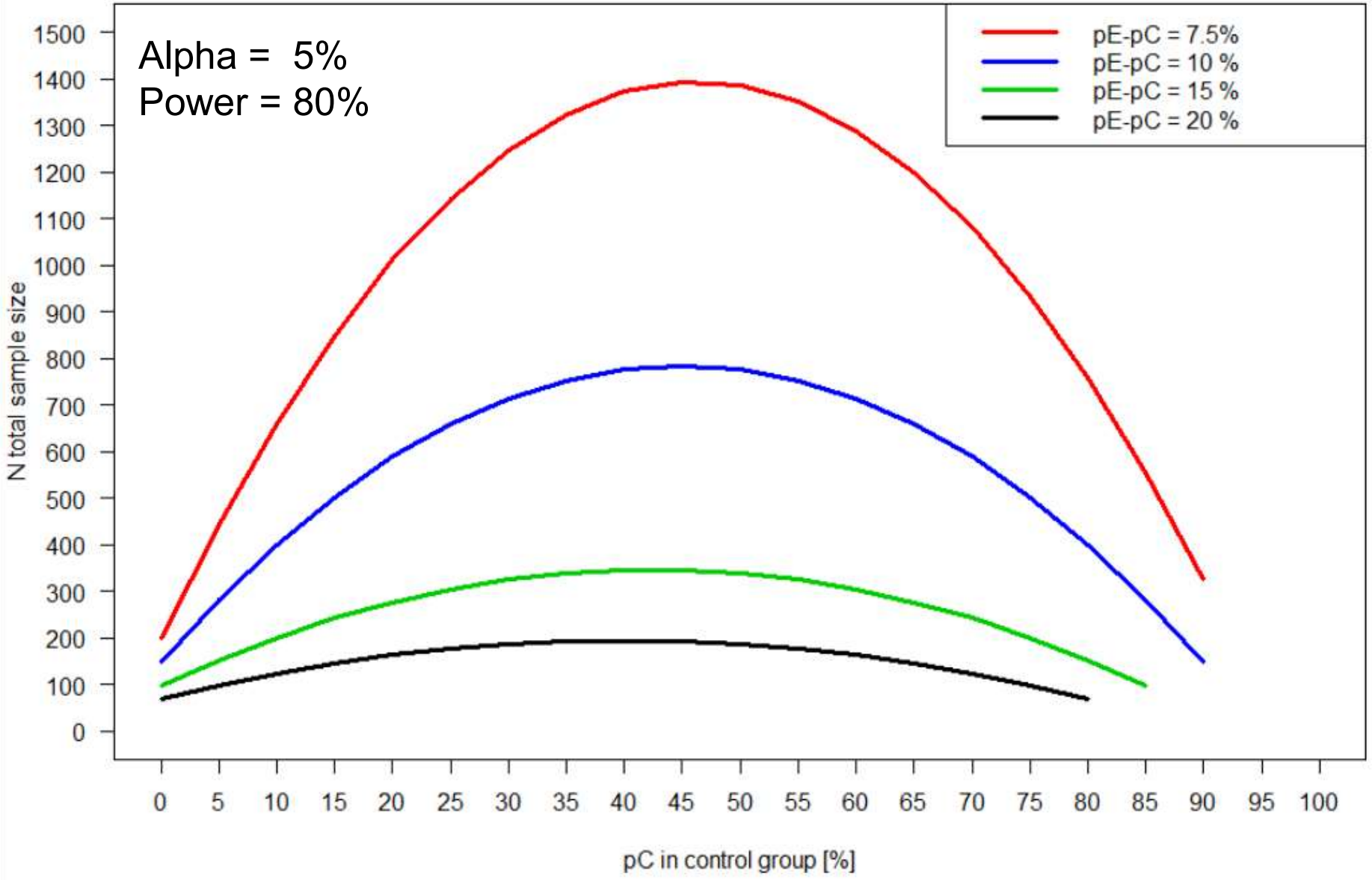
- Precision ~ Expected half width
  of the two-sided confidence interval
  - ±25%, ±20%, ±15%?
- Evidence level: Confidence level
  - 95%, 90%, 80%?
- Expected Probabilities: p_control, p_experimental

- **N = 100-200 is often a reasonable order of magnitude for a Phase II study**


Expected half width of 95% CI - difference in proportion
Two-sided 95% CIs is
Standard currency in clinical research
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
rate p0 = 0.5 - p1 = 0.6
rate p0 = 0.5 - p1 = 0.65
rate p0 = 0.5 - p1 = 0.75
40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
Total sample Size N = 2*n


Expected half width of (1-alpha)-CI - difference in proportion
Do you accept
weakening your standards and
your evidential requirement?
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
two-sided confidence = 0.95
two-sided confidence = 0.9
two-sided confidence = 0.8
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
Total Sample Size N

Total sample size Chi2-test



Interim analysis?

Types of interim analyses

- **Specify precisely in the protocol**
  what decisions the interim analysis may trigger

- **Early stopping for early success:**
  - Already clear that the trial objective is met.
  - Speed up development and go to Phase III
    - Rare…

- **Early stopping for futility:**
  - Already clear that the trial objective cannot be met.
  - Do not expose further patients to risks of unpromising therapy
  - Spare resources and study patients

Think twice about interim analyses

- **Only possible with short-term response endpoints**
- **Statistically more complicated**
  - Multiplicity issues
  - Difficult to explain in publication
- **Logistically more complicated**
- **Contracts more complicated**
  - Predetermined breaking point

Interim analysis in an uncontrolled Phase II study

Simon two-step design

Controlled Clinical Trials

Volume 10, Issue 1, March 1989, Pages 1-10

Optimal two-stage designs for phase II clinical trials

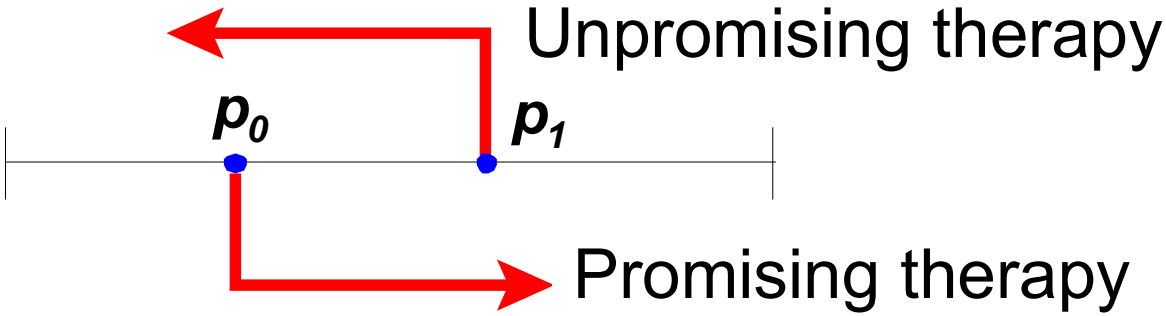
Richard Simon PhD

Show more

https://doi.org/10.1016/0197-2456(89)90015-9

Two step design (Simon, 1989)

Chose two clinically relevant p_0 and p_1 for the response rate



1. Hypothesis $p \leq p_0$ rejected, if enough responses observed
   ⟶ Therapy promising

2. Hypothesis $p \geq p_1$ rejected, if not enough responses observed
   ⟶ Therapy not promising

Example: LUCAS

- Response rate to be rejected: $\pi_0 = 0.05$
  (Nullhypothesis H0: $\pi \leq \pi_0$)

- Response rate not to be missed: $\pi_A = 0.20$
  (Specific alternative hypothesis HA: $\pi \geq \pi_A$)

- Control relevant error rates:
  - $\alpha = 0.05$ probability to declare an inactive drug (response probability π_0) promising
  - $\beta = 0.1$ probability to dismiss an active drug that has response probability π_A

Example: LUCAS

- Enroll N=42 patients in two stages, with the option to **stop early for futility** after the first 21 patients have their response documented:

- **Stage 1:**
  - **Dismiss Drug X** - if less than two responses are documented in the first 21 patients.

- **Stage 2:** (**only if 2 or more responses from stage 1**):
  - **Dismiss Drug X** - if response less than five responses are documented in all 42 patients
  - Otherwise reject null-hypothesis: **DrugX promissing**

Example: LUCAS

- This design has the following properties:
  - Expected sample size EN (π0 = 0.05) = 26.7
  - Probability of early stopping if Nullhypothesis π0 ≤ 0.05 is true: ≥ 72%
- Note: One-sided test.
-
- Note: Stop study – waiting for 21st response?
  - When 21 patients have been included and it is not yet clear that the study goes to the second stage, **we will / will not stop accrual** in order not to prolong the study duration.

Take home messages

- Clearly elaborate the biological-medical rationale.
- Clearly state the **primary study questions**.
- Carefully consider the primary clinical and primary biological endpoint.Beware of **misleading surrogates**.
- If in doubt, **randomise** in phase II to avoid selection bias.
- **Do not cheat or delude yourself**!
- Detail clearly **what evidence signal you realistically expect from the trial**!
- Use **appropriate sample sizes**!

- Make sure that you have **enough information to plan the subsequent definitive phase III trial** afterwards.