Vorbereitende Phase II Studien

Dr. Dirk Hasenclever

Project design considerations: biometry, statistics and more

Dr. Dirk Hasenclever

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

dirk.hasenclever@imise.uni-leipzig.de

Summer School:
CLONE – Clinical Trials in Oncology in the New Era of Omics, Big Data, and Modeling,
24.-28.06.2019 in Halle (Saale)/Germany

Public funding for early trials in Germany

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

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

- Is there a **plausibly argued biological-medical rationale?**
- Is there a **clearly stated study question?**
- Is there an **appropriate primary endpoint?**
- Is the **design adequate** in controlling **relevant sources of bias**?
- Is the **sample size sufficient** to achieve a **clearly stated, meaningful evidence signal**?

- Is it the **right study question on a roadmap** to a definite Phase III study?
- Will the result sufficiently **inform the decision to go to Phase III** and **respective study planning**?

Phase I/II studies mainly prepare 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.
  - Three 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.

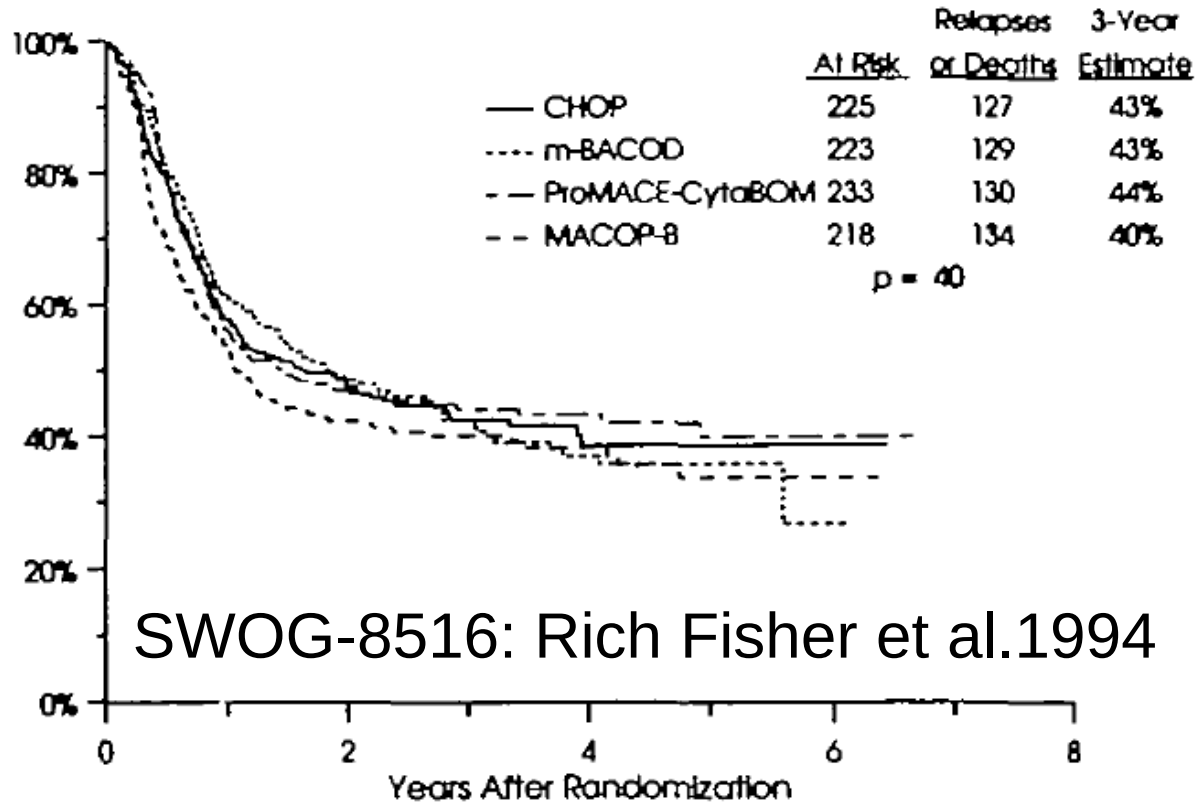


Fig. 1. Time-to-treatment-failure curve for all eligible patients according to the randomized treatment.

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!**

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%.

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*

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

Uups, slightly worse!

I chaired the 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%)

→

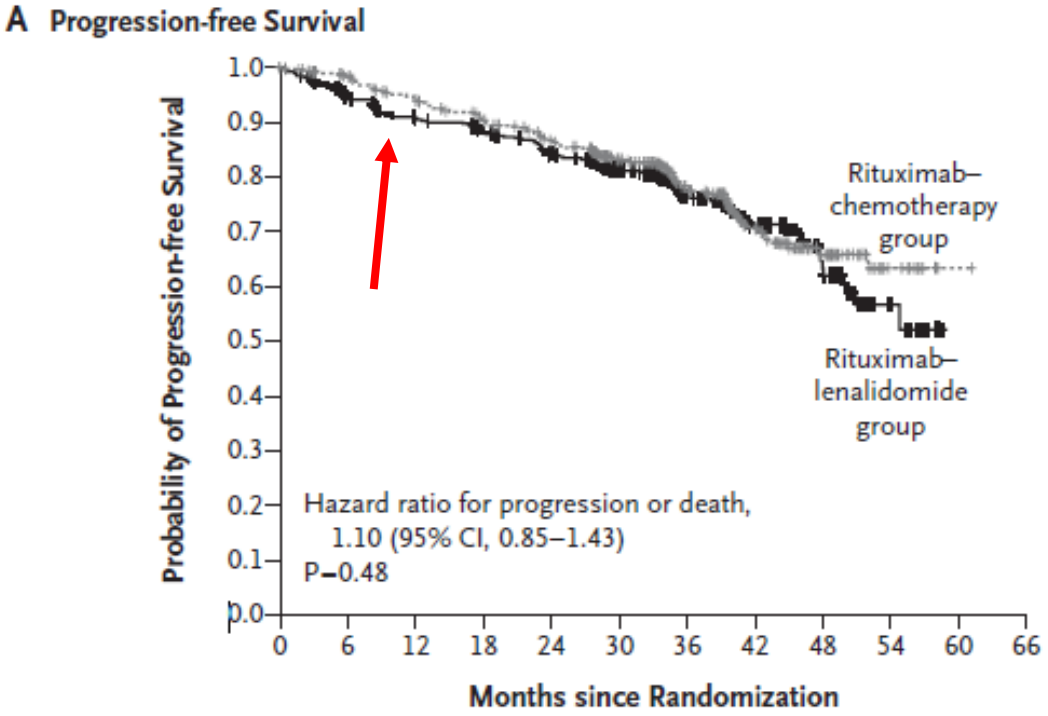
FLIPI score — no. (%)	
0 or 1	153 (15)
2	374 (36)
3 to 5	503 (49)

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.



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.
- Therefore **they tend to need more and not less biometrical consideration**.

- 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: 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%.

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.

Most phase II endpoints are surrogates!

- A **surrogate-endpoint** is
  - a Lab parameter,
  - a biological marker, or
  - a histopathological or a
  - clinical assessment

used as **short-term substitute** for a **clinically relevant 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

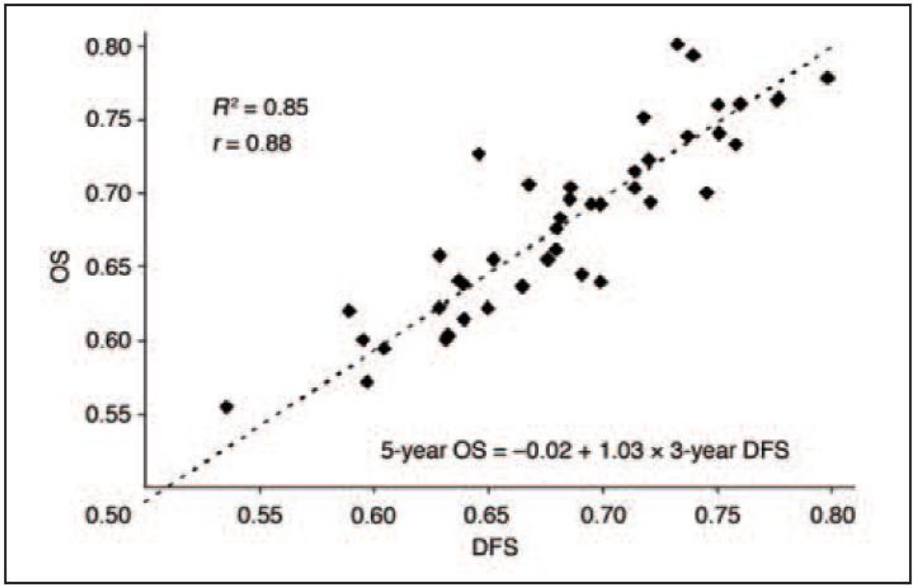


Fig 1. Three-year disease-free survival (DFS) versus 5-year overall survival (OS) by study arm.

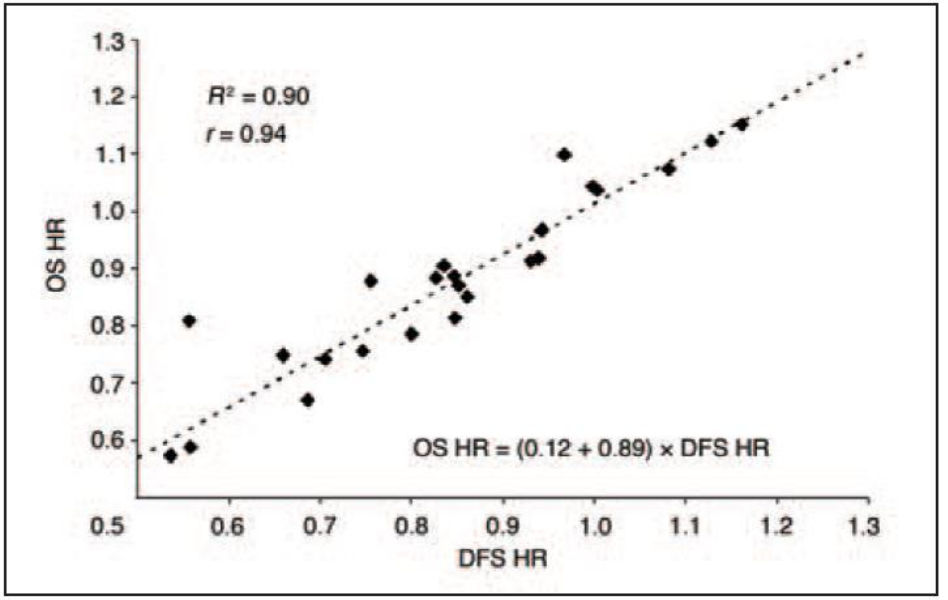
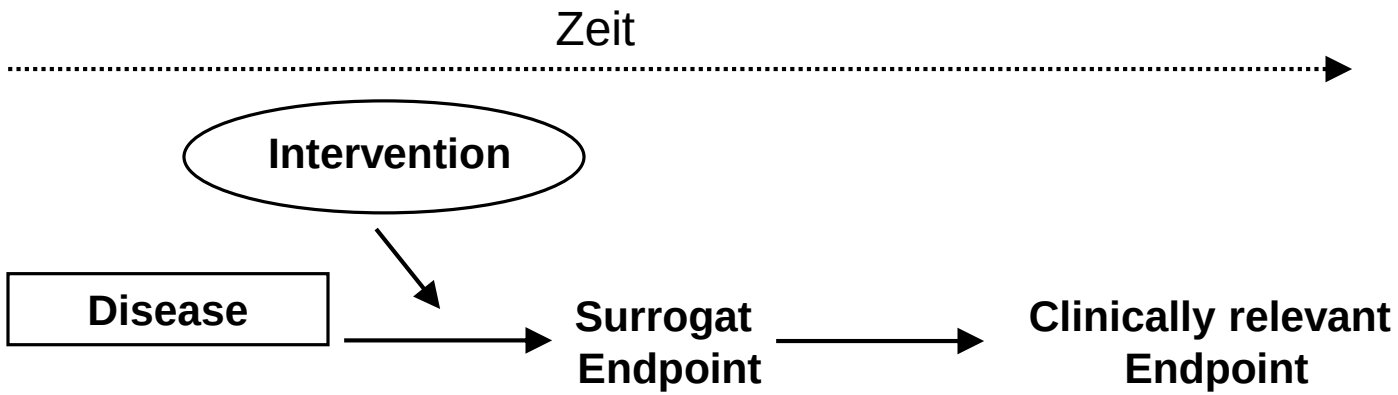


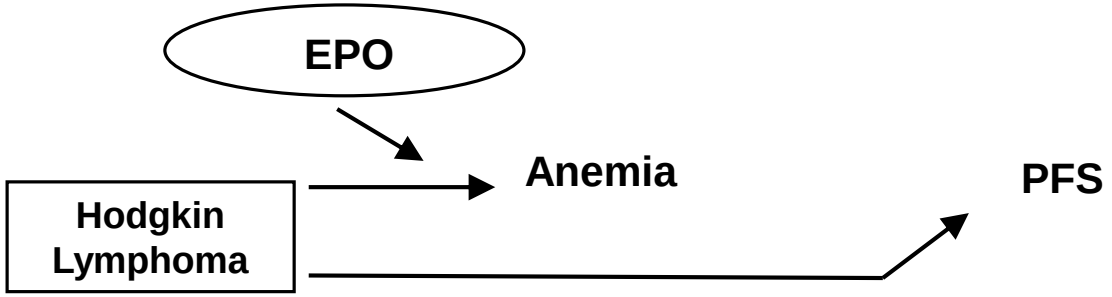
Fig 2. Disease-free survival (DFS) versus overall survival (OS) hazard ratios (HR) by trial.

Think in causal models!

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



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

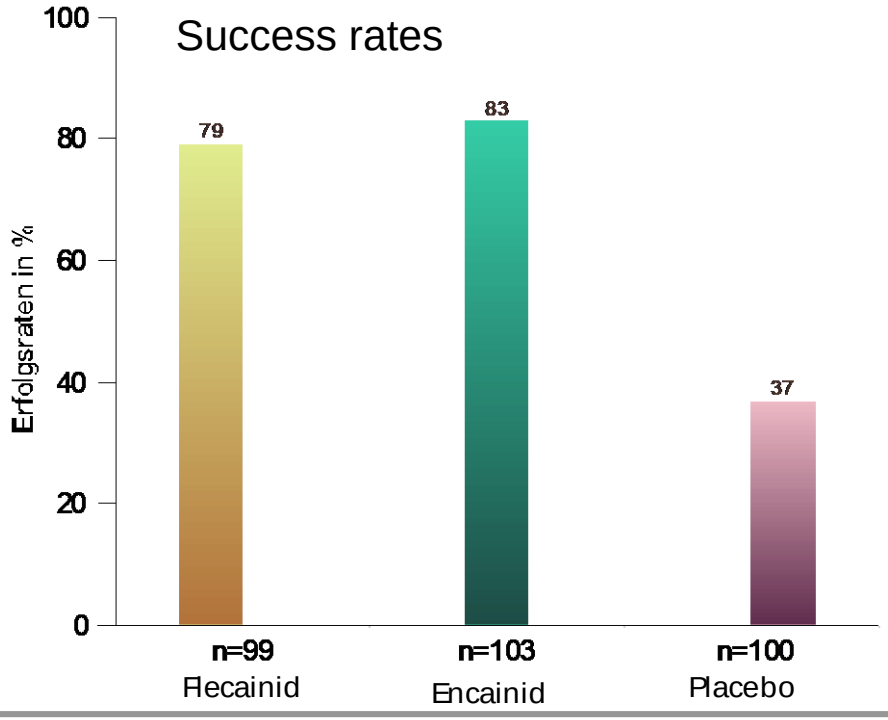


Disaster with surrogate endpoints I

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

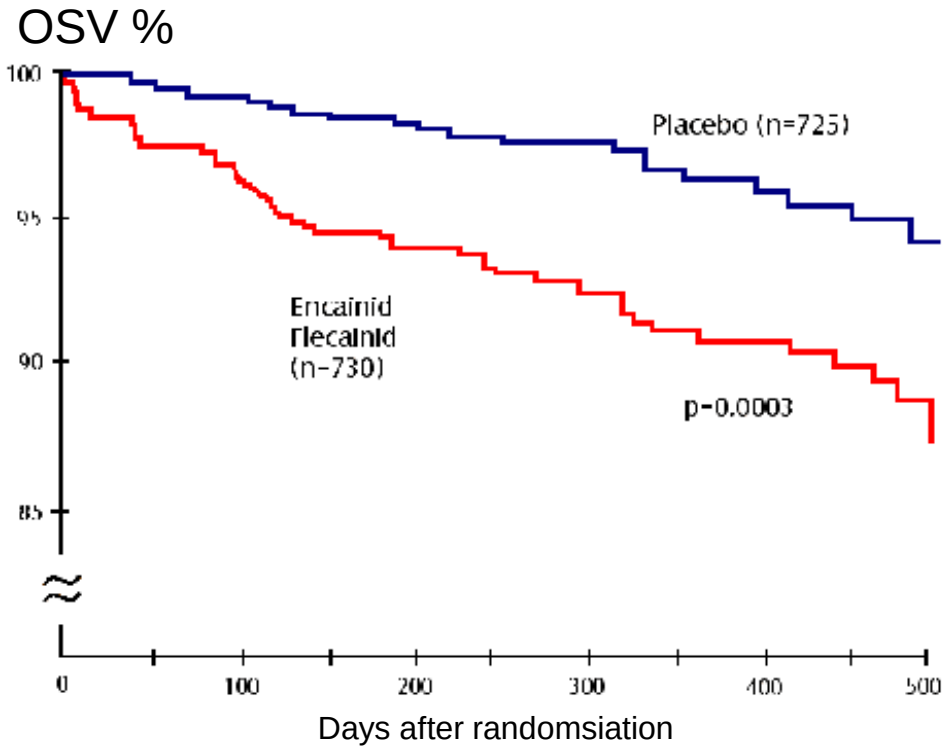
-> licenced for large market

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

Take home messages

- Clearly elaborate the biological-medical rationale.
- Clearly state the primary study question.

- If in doubt, randomise in phase II to avoid selection bias.
- Do not cheat or delude yourself!
- Careful consideration of the primary clinical and primary biological endpoint. Beware of misleading surrogates.
- Detail clearly what evidence signal you realistically expect from the trial! And use appropriate sample sizes!

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

Fallzahlen und Simon Zwei-Stufen Design

Machbarkeitsstudie

Ist eine **neue Therapie** oder Therapiestrategie **geeignet**, gegen eine etablierte **Standardtherapie im Wirksamkeitsvergleich** geprüft zu werden?

Ziele:

- Überprüfung der **Machbarkeit** bezüglich Toxizität und Logistik
- **erster Nachweis** von biologischer/ therapeutischer **Wirkung**
- Priorisierung von Therapien für kontrollierte klinische Therapiestudien
- ggf. Gewinnung zusätzlicher Informationen zur Pharmakologie

Endpunkte:

- **akute Nebenwirkungen** (z.B. WHO-Grade für Infektionen)
- Parameter zur Ermittlung der **Therapieadhärenz** (z.B. Relative Dosisintensität)
- **kurzfristig** zu beobachtende **Wirkungsparameter** (z.B. Rate kompletter Remissionen)

Machbarkeitsstudie

Auswertung:

- **Schätzung** von Toxizitäts- bzw. Erfolgsraten erforderlich
- **Konfidenzintervalle (KI)** beschreiben **Präzision der Schätzung**

Es gilt: Fallzahl ⇔ Präzision der Schätzung

Grobe Regel: $n = \frac{1}{(\text{Breite des 95\%-KI})^2}$

Beispiel:

Präzision = +- 10% ⟶ n=100 Patienten

Präzision = +- 20% ⟶ n= 25 Patienten

Einseitiger Test gegen vorgegebenes p0

```
#           +0.05 +0.1 +0.15 +0.2 +0.25
# p0=0.05    212   67    35   23    16
# p0=0.1     341   98    49   30    21
# p0=0.15    452  124    60   36    24
# p0=0.2     546  146    69   41    27
# p0=0.25    625  164    76   44    29
# p0=0.3     688  178    81   47    30
# p0=0.35    736  188    85   48    31
# p0=0.4     767  194    87   49    31
# p0=0.45    783  196    87   48    30
# p0=0.5     783  194    85   47    29
# p0=0.55    767  188    81   44    27
# p0=0.6     736  178    76   41    24
# p0=0.65    688  164    69   36    21
# p0=0.7     625  146    60   30    16
```

```r
#------------------------------------------------------------------------
###  Sample size to compare a proportion with a fixed p0
###  Desu MM & Raghavarao D. Sample size methodology Academic Press (1990) p 16
#------------------------------------------------------------------------

PowerOneSampleProportion<-function(p0,p1,alpha,beta,SIDED=2){
  ceiling((qnorm(1-alpha/SIDED) + qnorm(1-beta))^2 /
            (2 * (asin(sqrt(p1)) - asin(sqrt(p0))))^2)
}
```

Machbarkeitsstudie

Gewünschter Studienabbruch, falls:

- zu hohe Inzidenz von **Nebenwirkungen**
- untersuchte **Therapie ineffektiv** ist

→ „**early stopping rules**“ erlauben vorzeitigen Studienabbruch wegen **futility**

Interim analysis in an uncontrolled Phase II study



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

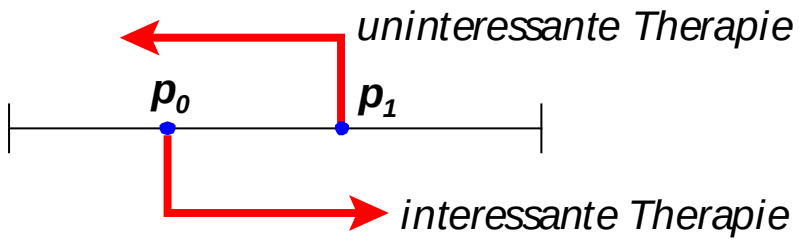
https://doi.org/10.1016/0197-2456(89)90015-9

Get rights and content

Machbarkeitsstudie

2 – Stufendesign (Simon, 1989):

Wahl zweier klinisch relevanter Schwellenwerte p_0 und p_1 für die Erfolgsrate



1. Die Hypothese $p \leq p_0$ wird verworfen, falls viele Erfolge eintreten ⟶ Therapie weiterverfolgen

2. Die Hypothese $p \geq p_1$ wird verworfen, falls zu wenige Erfolge eintreten ⟶ Therapie aufgeben

Machbarkeitsstudie

	Erfolgs-wahrscheinlicheit	Fehler
Beispiel:	$p_0 = 20\%$	$\alpha = 5\%$
	$p_1 = 40\%$	$\beta = 10\%$

Fehler, **Therapie weiter zu verfolgen**, obwohl **wahre Responserate < p_0**

Fehler, **Therapie nicht weiter zu verfolgen**, obwohl **wahre Responserate > p_1**

Optimales 2-Stufen Design: (Auszug)

p_0	p_1	α	β	r_1	n_1	r	n
0.10	0.30	0.05	0.20	1	10	5	29
			0.10	2	18	6	35
0.20	0.40	0.05	0.20	3	13	12	43
			0.10	4	19	15	54
0.30	0.50	0.05	0.20	5	15	18	46
			0.10	8	24	24	63

Machbarkeitsstudie

Studienablauf:

Stufe 1: **19 Patienten** erhalten die Therapie

STOP, falls **≤ 4 Erfolge**

sonst

Stufe 2: **35 weitere Patienten** erhalten die Therapie

Auswertung: betrachte alle 54 Patienten

- Therapie **uninteressant**, falls **≤ 15 Erfolge**
- Therapie **weiterverfolgen,** falls **> 15 Erfolge**