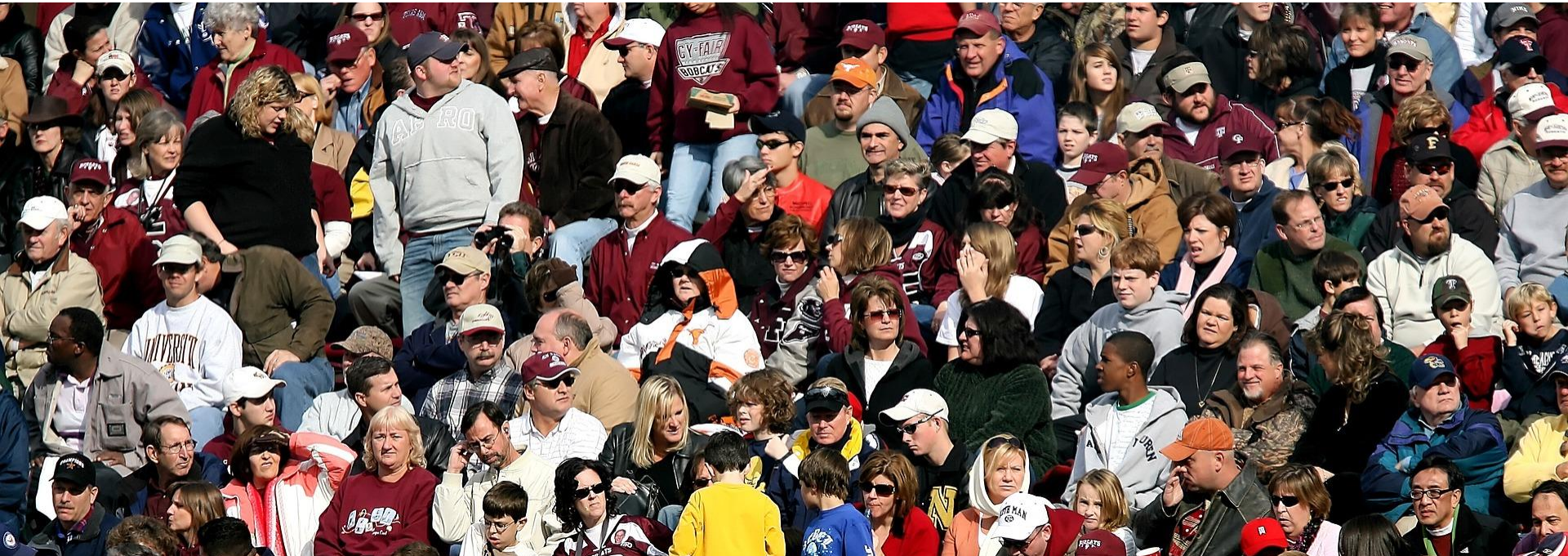
50 oder 500?

Prinzipien der Sample Size Berechnung erklärt

Sven Trelle, CTU Bern

Credits to Irina Irincheeva (CTU Bern), Stephen Senn and other references



KEK Bern feedback

Or any other ethics committee

B. Primärer Endpunkt: Der gewählte ‚primary outcome', ‚safety', (Protokoll 11.1) ist mit max 25 Patienten per Definition nicht mit genügend engem Vertrauensintervall erfassbar. (Dose-limiting toxicity wäre z.B. eine erfassbare Alternative).

- Eine Begründung für die Patientenzahl muss vorliegen.

- Seite 21, Kapitel 8.5:es fehlt die Fallzahlberechnung aufgrund des als relevant zu betrachtenden Unterschieds beim Primärparameter zwischen den beiden Behandlungsgruppen, womit die Anzahl Patienten begründet wird.

Framework of this lecture

Frequentist

– Null hypothesis significance testing
→ CTU Lecture 06/2018

The p-value – what it is (not)!

Sven Trelle

CTU Bern

Framework of this lecture

Frequentist

- Null hypothesis significance testing
  → CTU Lecture 06/2018
- Parameter of interest e.g. treatment effect is unknown but fixed (no uncertainty)
- A parameter estimate e.g. treatment effect estimate from a clinical trial has a distribution (uncertainty)
  - The given experiment is one realization of an infinite number (repetition) of the same experiment

Null hypothesis testing

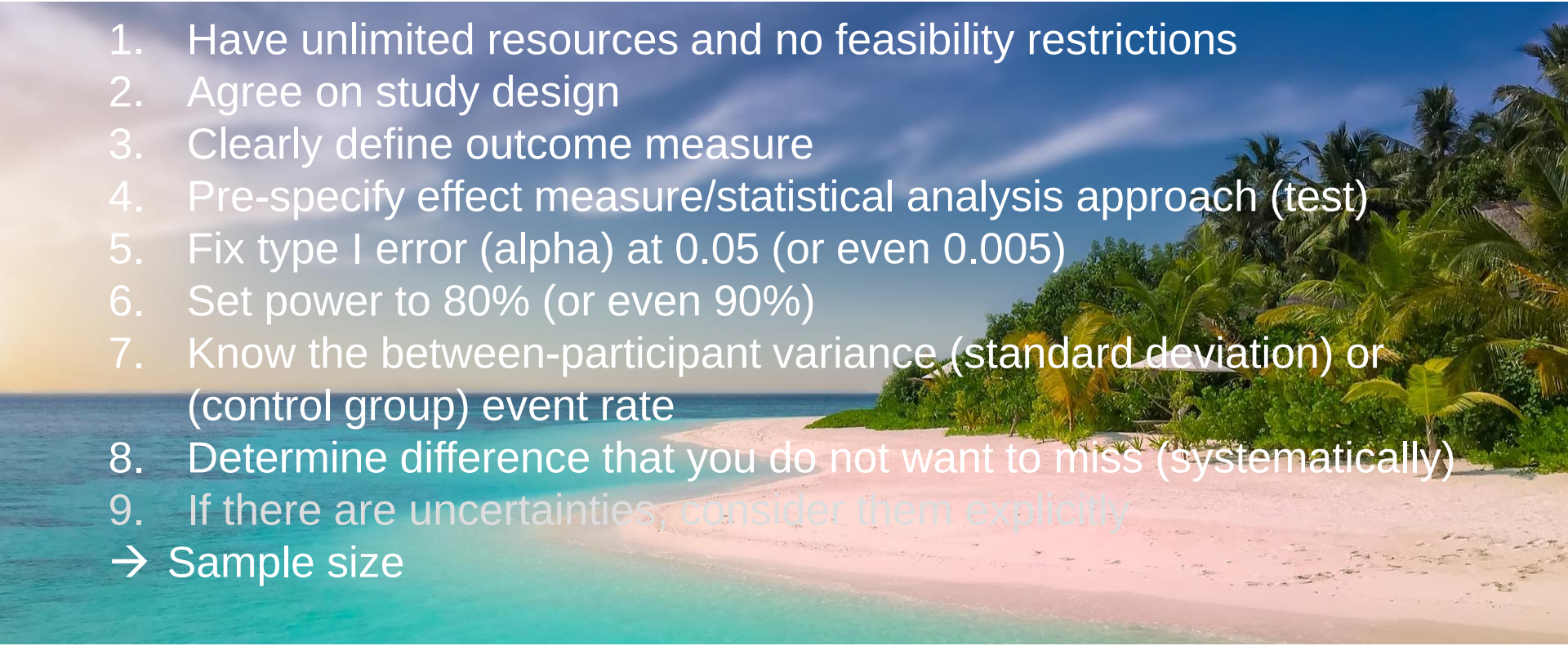
UNIVERSITÄT BERN

"Truth (known)"

Test outcome	Null hypothesis is true	Alternative hypothesis is true
Reject the null ($p < \alpha$)	Type I error, false positive	Correct conclusion = **Power**
Do not reject the null ($p >/= \alpha$)	Correct conclusion	Type II error, false negative

Sample size calculation of a controlled trial

In a perfect world

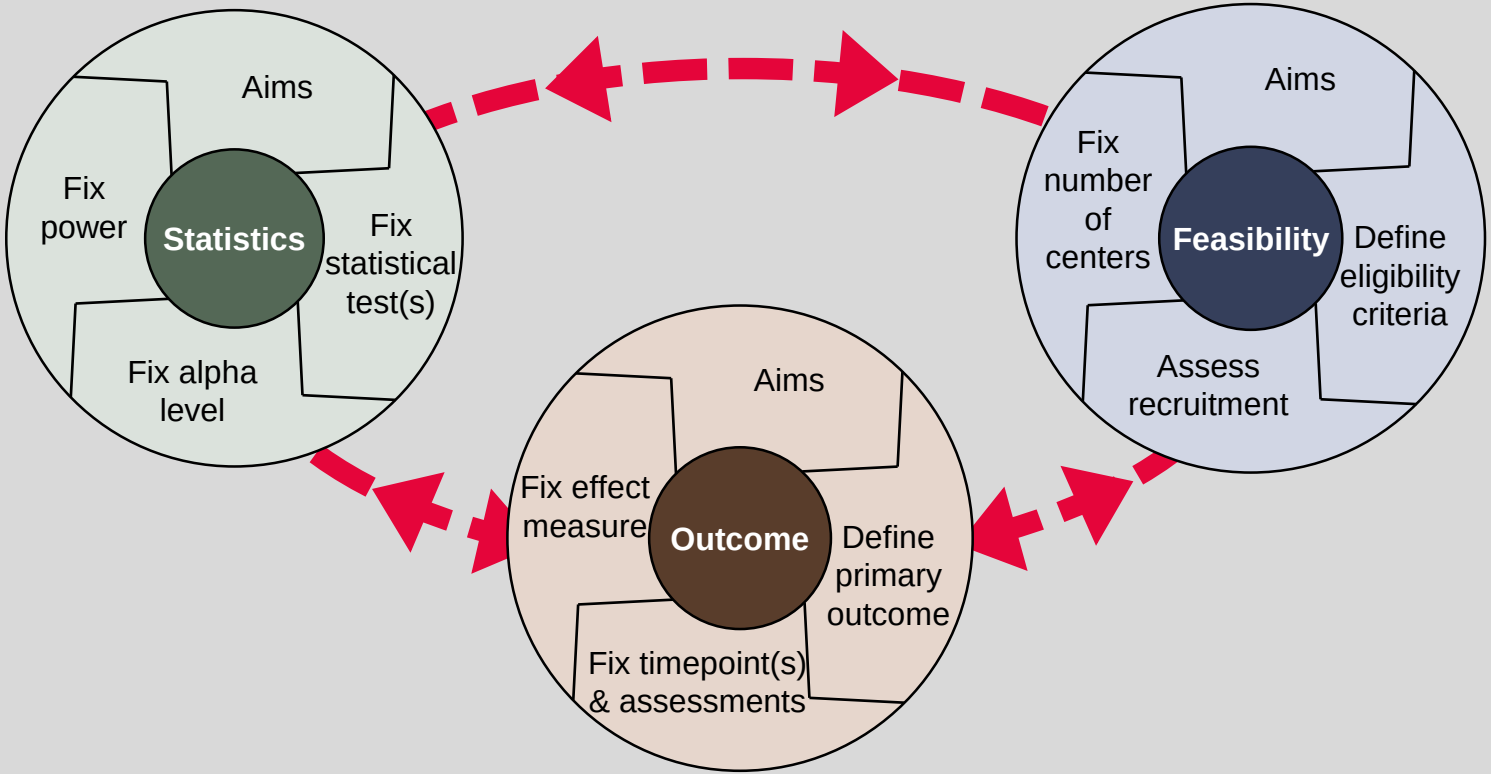


1. Have unlimited resources and no feasibility restrictions
2. Agree on study design
3. Clearly define outcome measure
4. Pre-specify effect measure/statistical analysis approach (test)
5. Fix type I error (alpha) at 0.05 (or even 0.005)
6. Set power to 80% (or even 90%)
7. Know the between-participant variance (standard deviation) or (control group) event rate
8. Determine difference that you do not want to miss (systematically)
9. If there are uncertainties, consider them explicitly

→ Sample size

Sample size calculation

An interative process with different aspects



What does power mean?

UNIVERSITÄT BERN

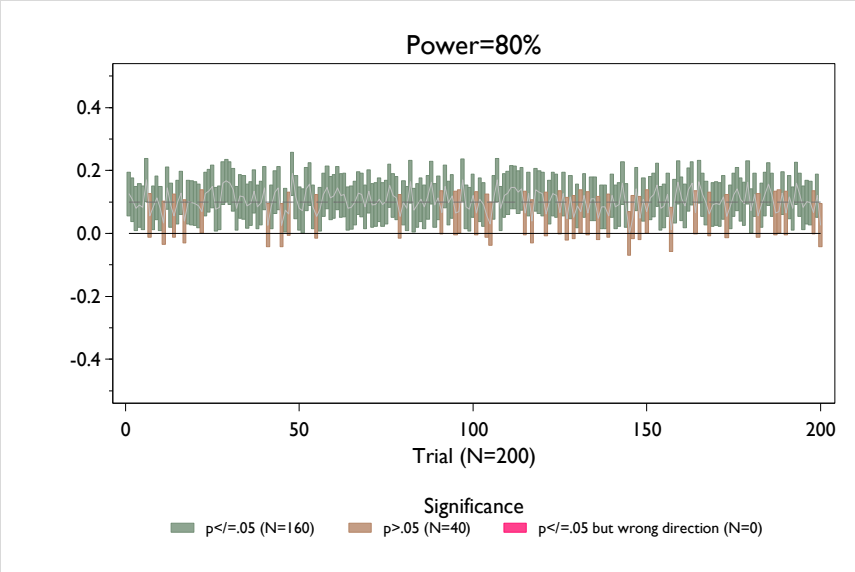
- Given that alternative hypothesis is true e.g. no difference
- Probability of rejecting null hypothesis
- Repetition

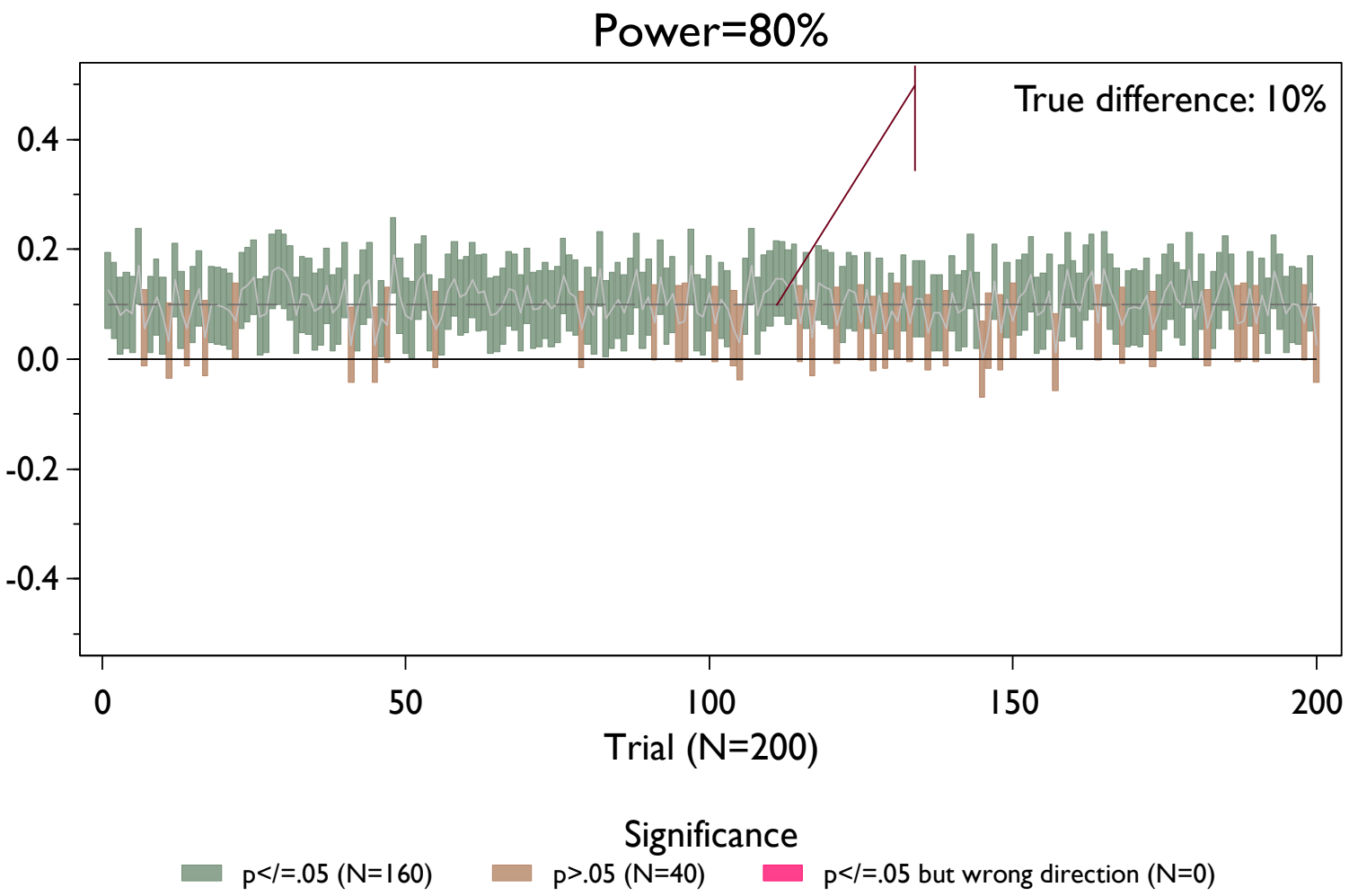
Example using repetition

Simulation

UNIVERSITÄT BERN

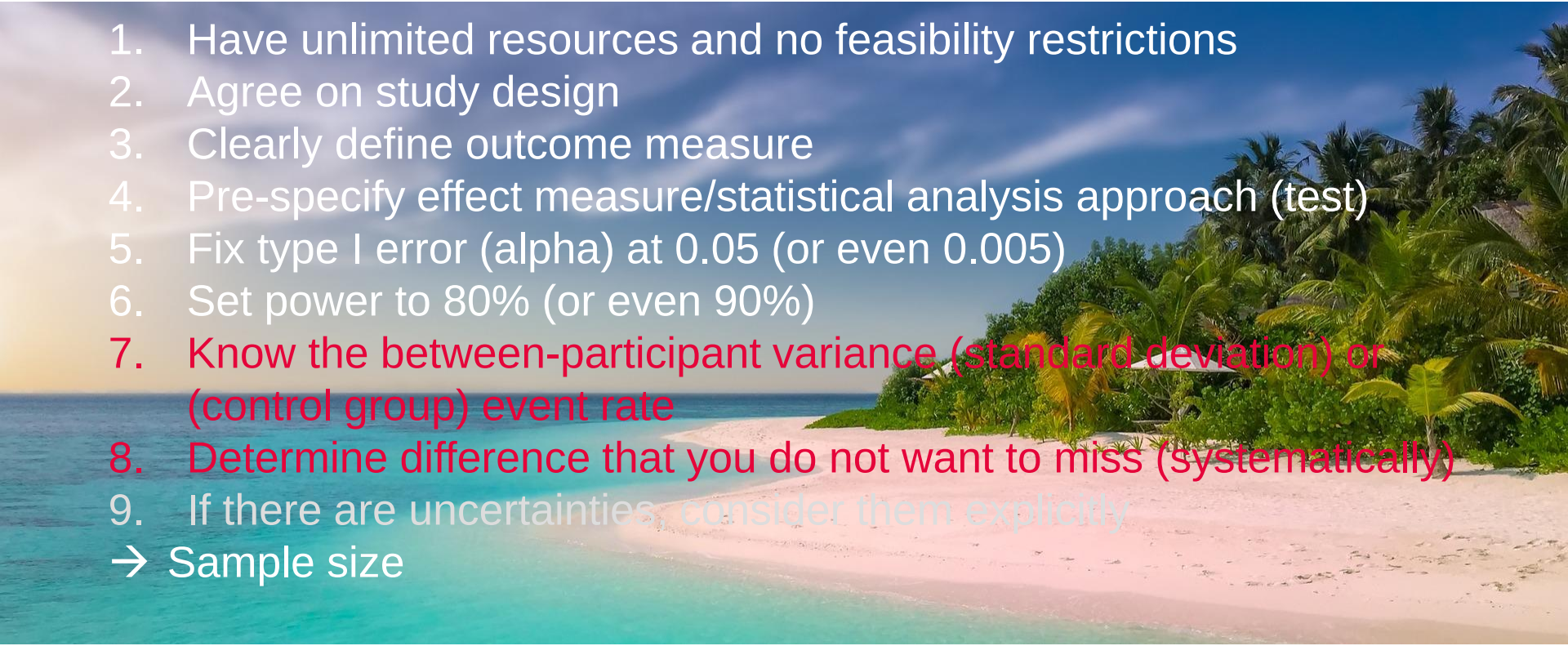
– Power: 80%; placebo response rate: 40%; effect: 10% improvement




Power=80%
True difference: 10%
0.4
0.2
0.0
-0.2
-0.4
0
50
100
150
200
Trial (N=200)
Significance
p</=.05 (N=160)
p>.05 (N=40)
p</=.05 but wrong direction (N=0)

Sample size calculation of a controlled trial

In a perfect world



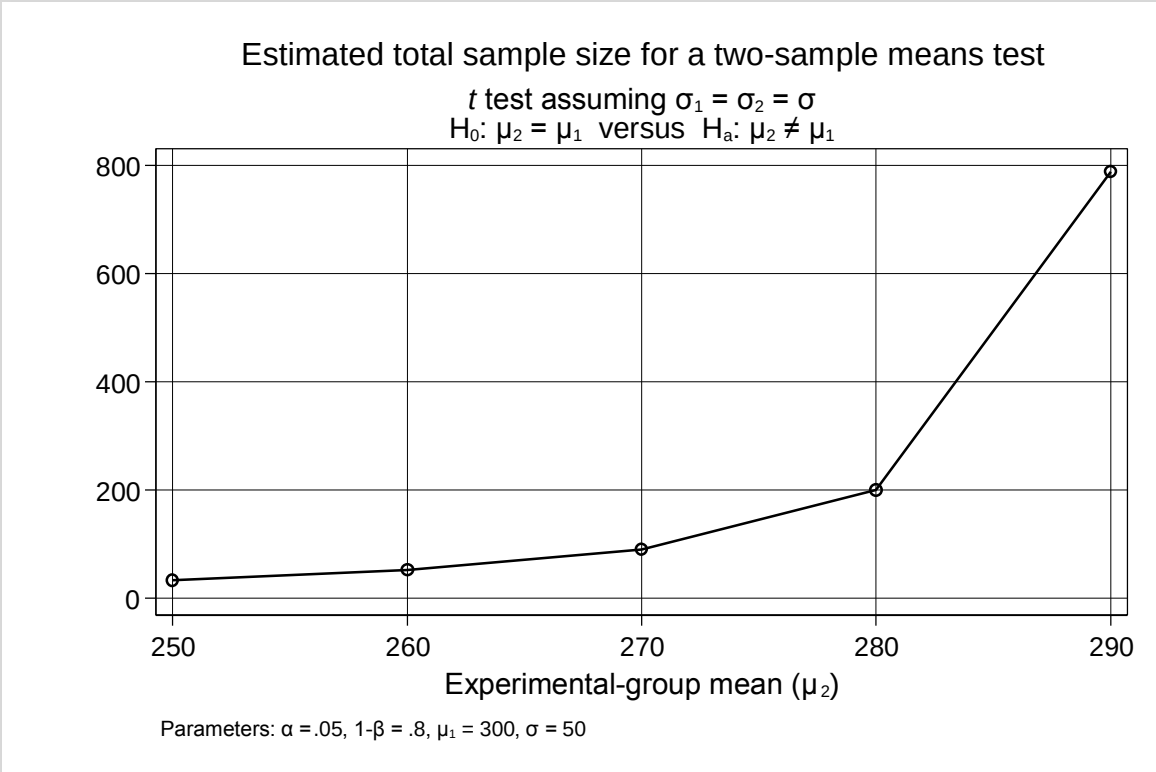
1. Have unlimited resources and no feasibility restrictions
2. Agree on study design
3. Clearly define outcome measure
4. Pre-specify effect measure/statistical analysis approach (test)
5. Fix type I error (alpha) at 0.05 (or even 0.005)
6. Set power to 80% (or even 90%)
7. Know the between-participant variance (standard deviation) or (control group) event rate
8. Determine difference that you do not want to miss (systematically)
9. If there are uncertainties, consider them explicitly

→ Sample size

Continuous outcomes

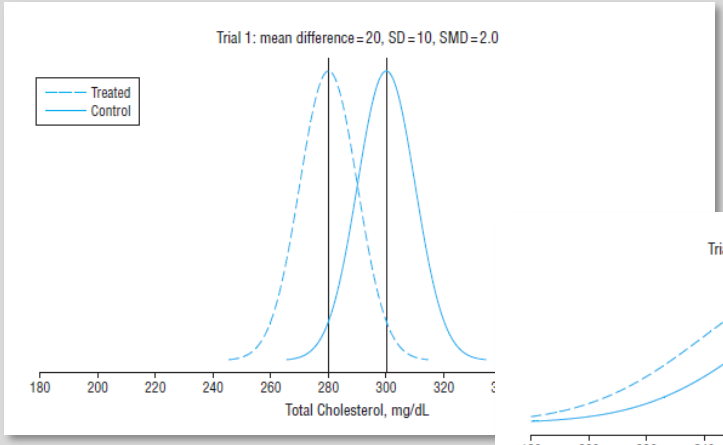
Influence of difference

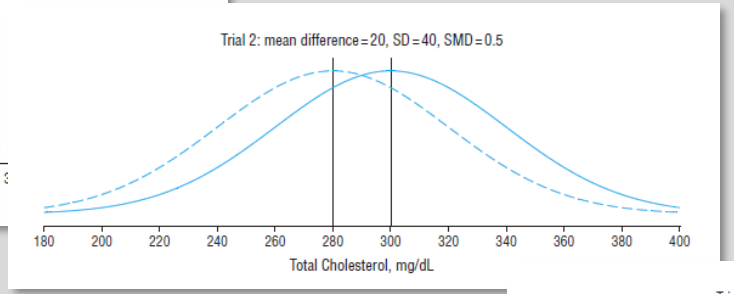
Continuous outcomes (variance/SD fixed)

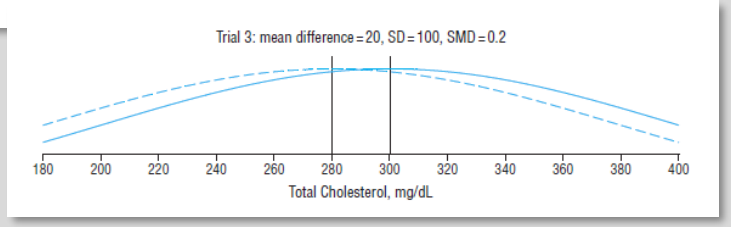


Influence of between-participant variance

Continuous outcomes (difference fixed)

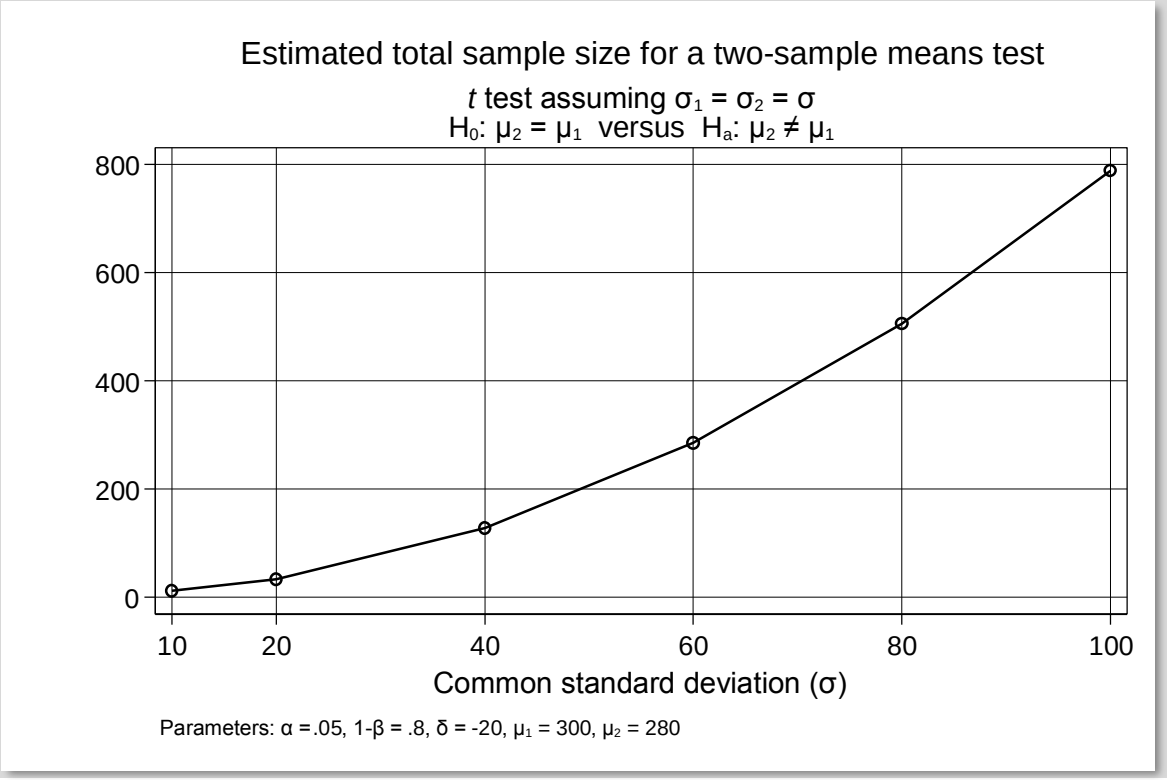






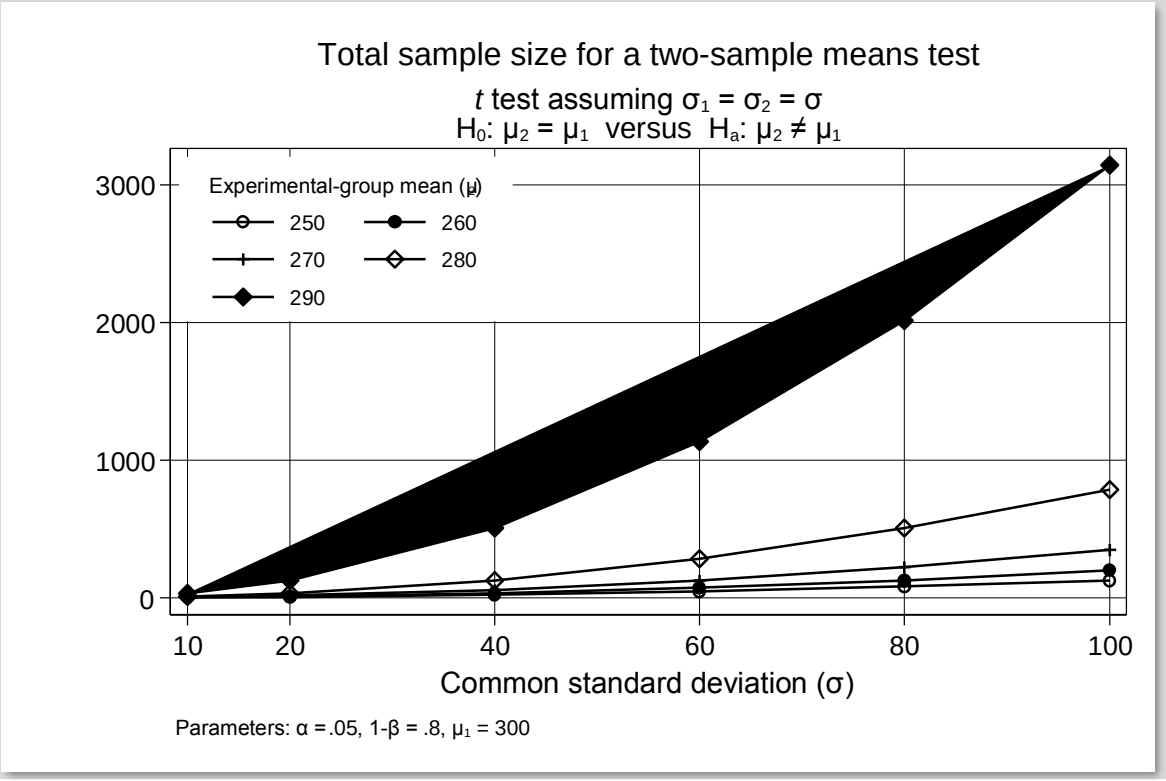
Influence of between-participant variance

Continuous outcomes (difference fixed)

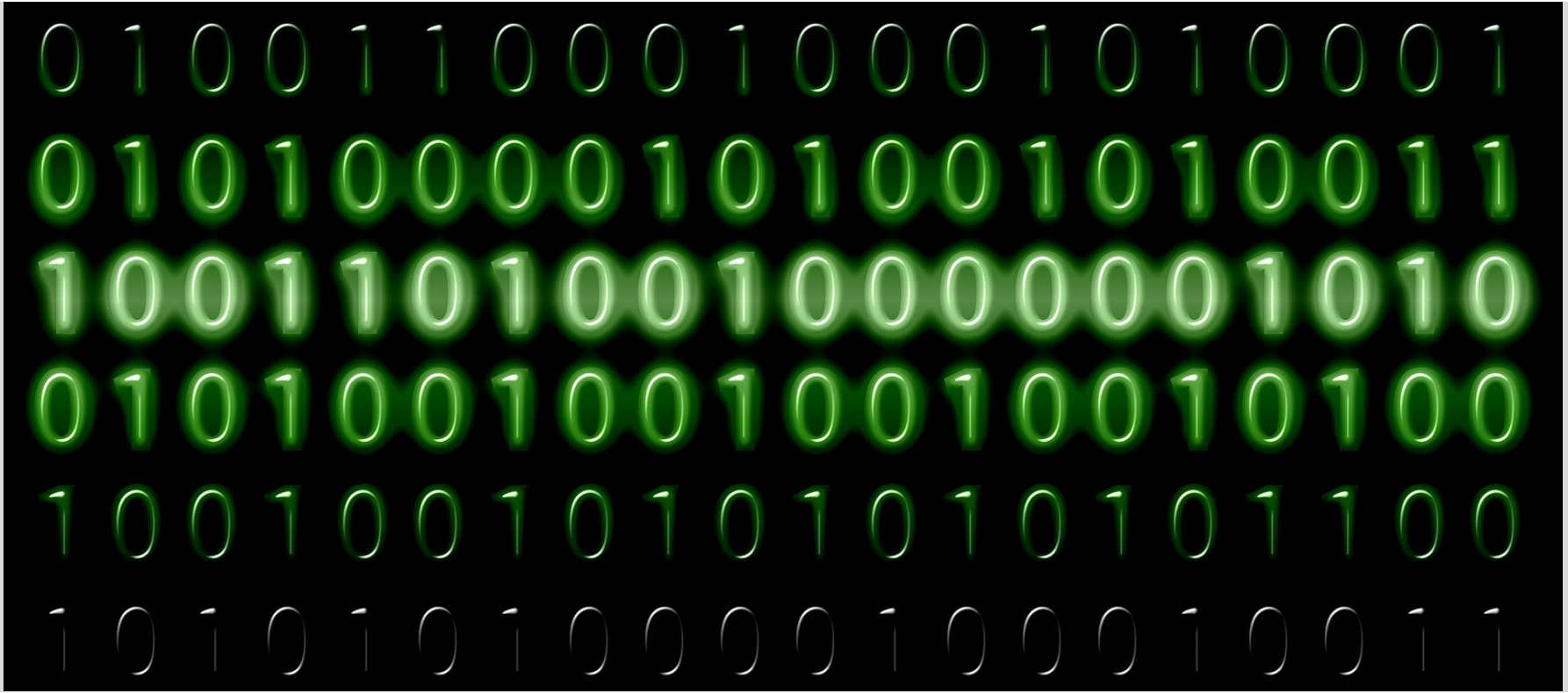


Difference & variance

Continuous outcomes

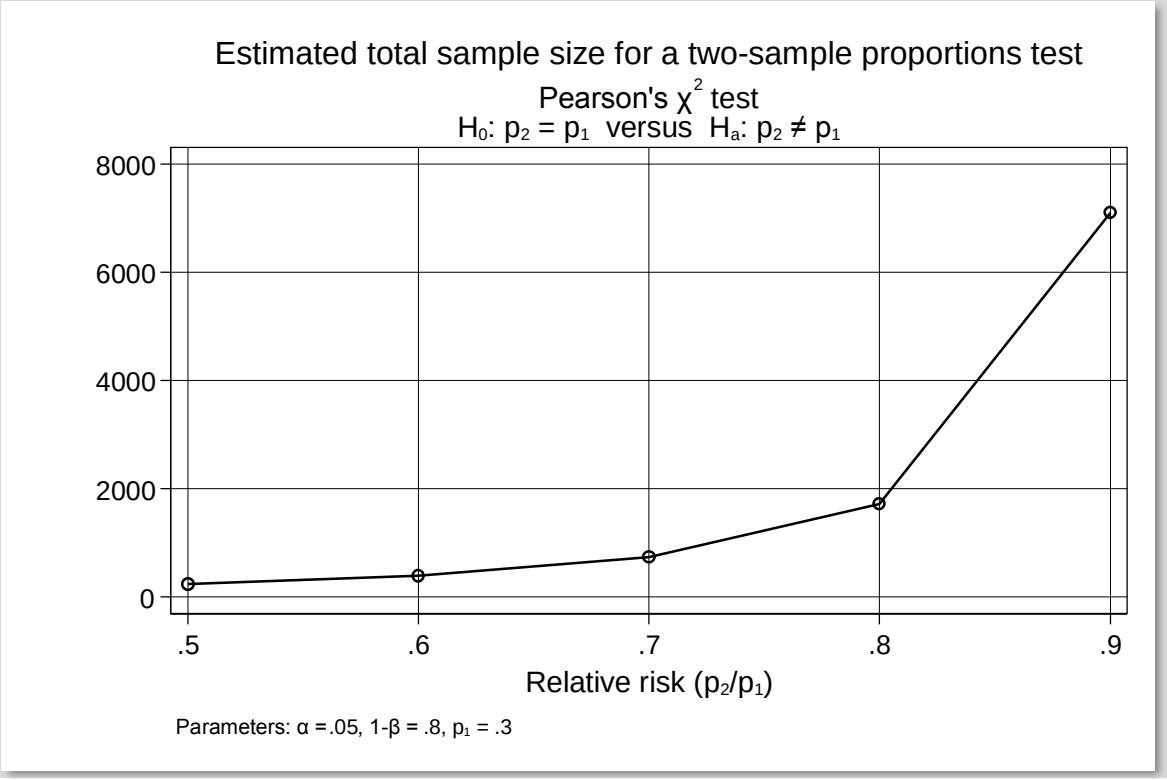


Binary outcomes



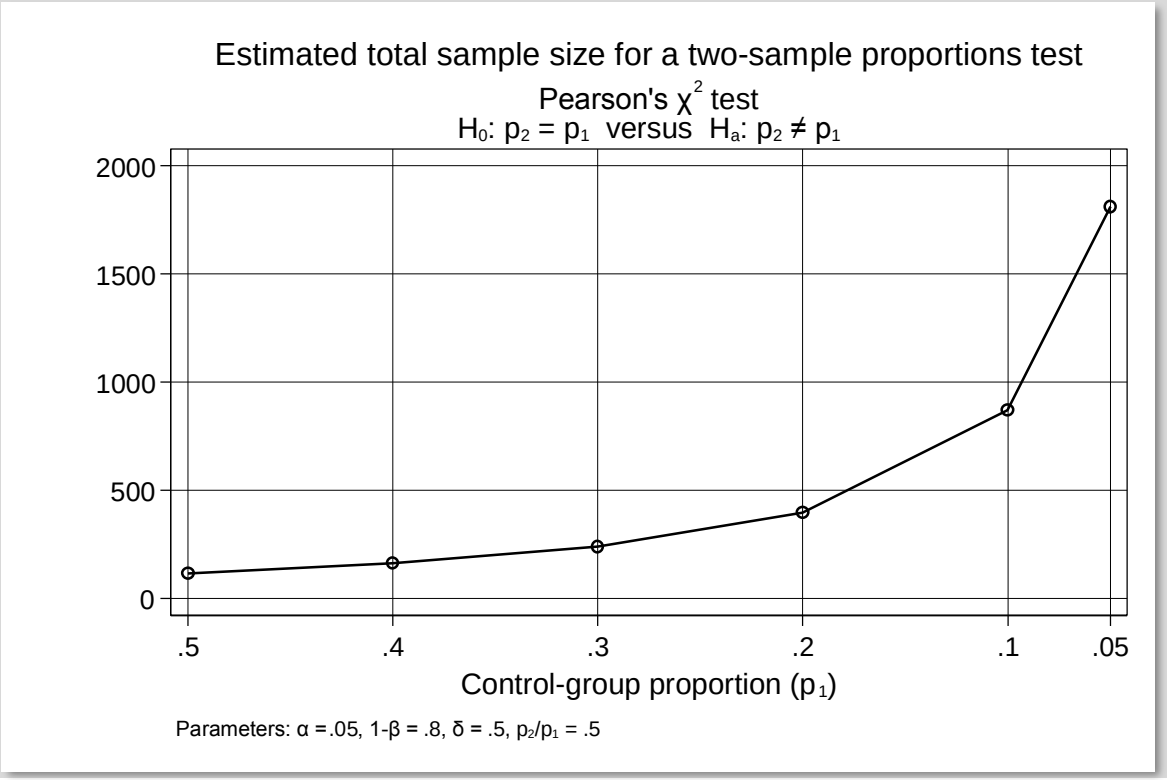
Influence of effect (relative)

Binary outcome ((control) event rate fixed)



Influence of event rate (relative)

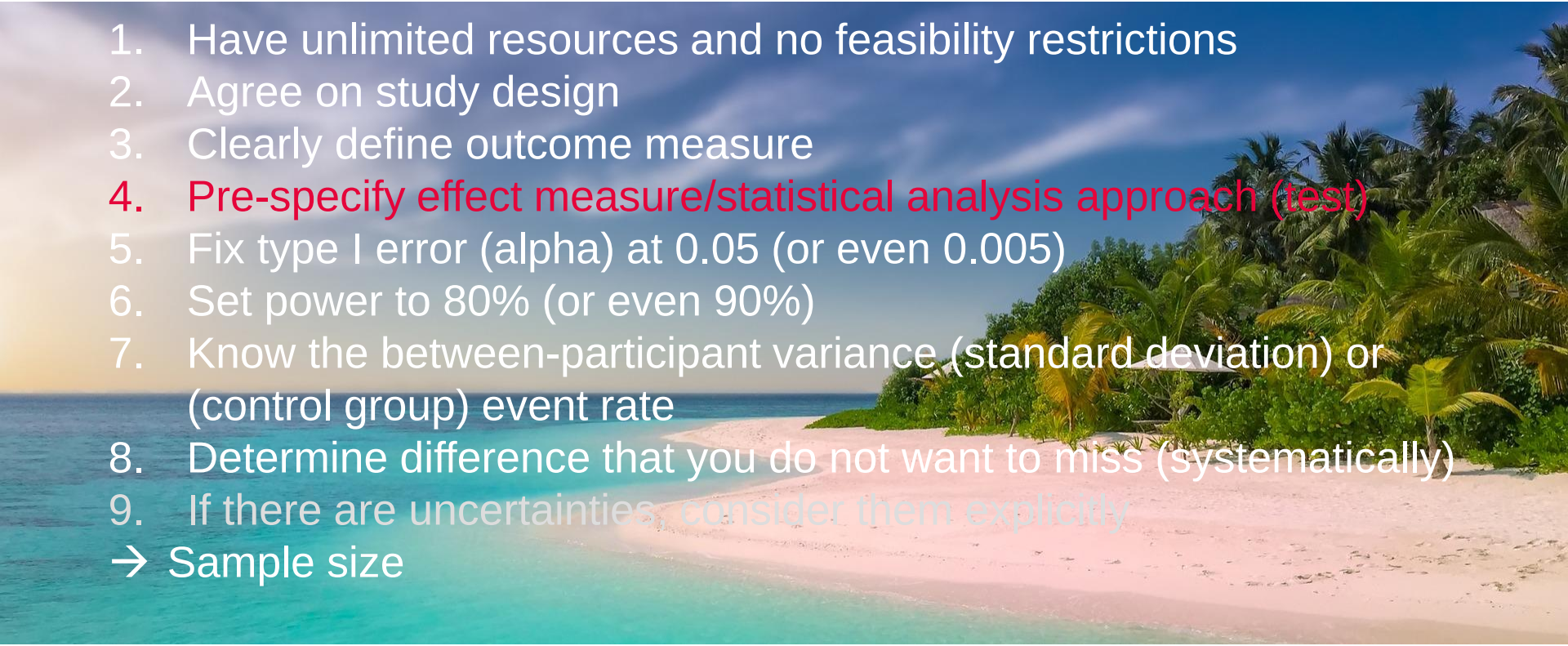
Binary outcome (**Relative Risk** fixed)



Sample size calculation of a controlled trial

UNIVERSITÄT BERN

In a perfect world

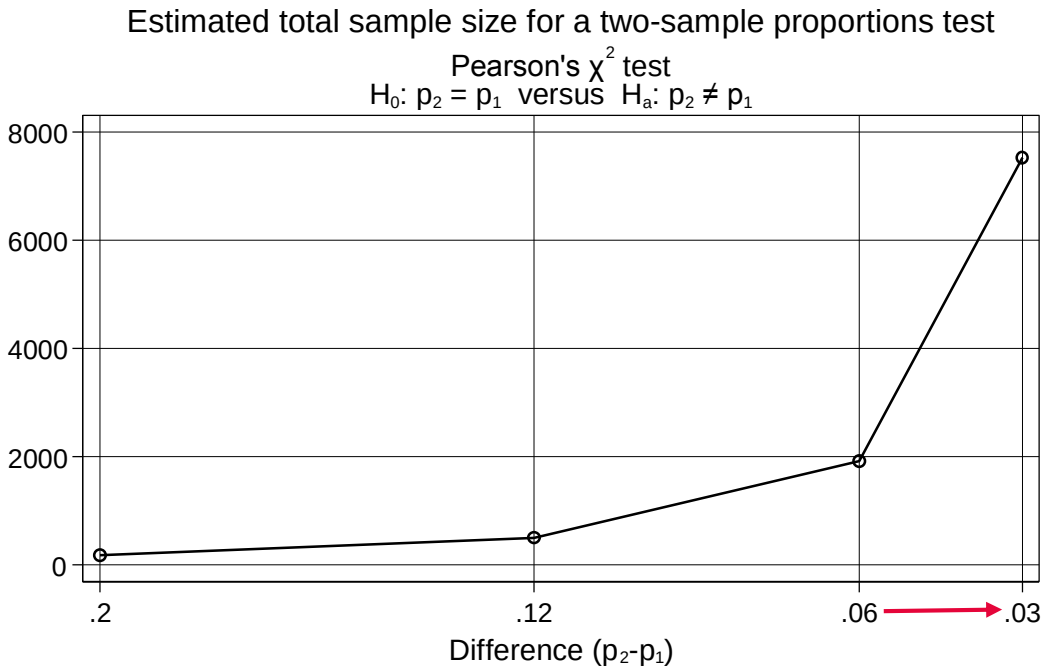


1. Have unlimited resources and no feasibility restrictions
2. Agree on study design
3. Clearly define outcome measure
4. Pre-specify effect measure/statistical analysis approach (test)
5. Fix type I error (alpha) at 0.05 (or even 0.005)
6. Set power to 80% (or even 90%)
7. Know the between-participant variance (standard deviation) or (control group) event rate
8. Determine difference that you do not want to miss (systematically)
9. If there are uncertainties, consider them explicitly

→ Sample size

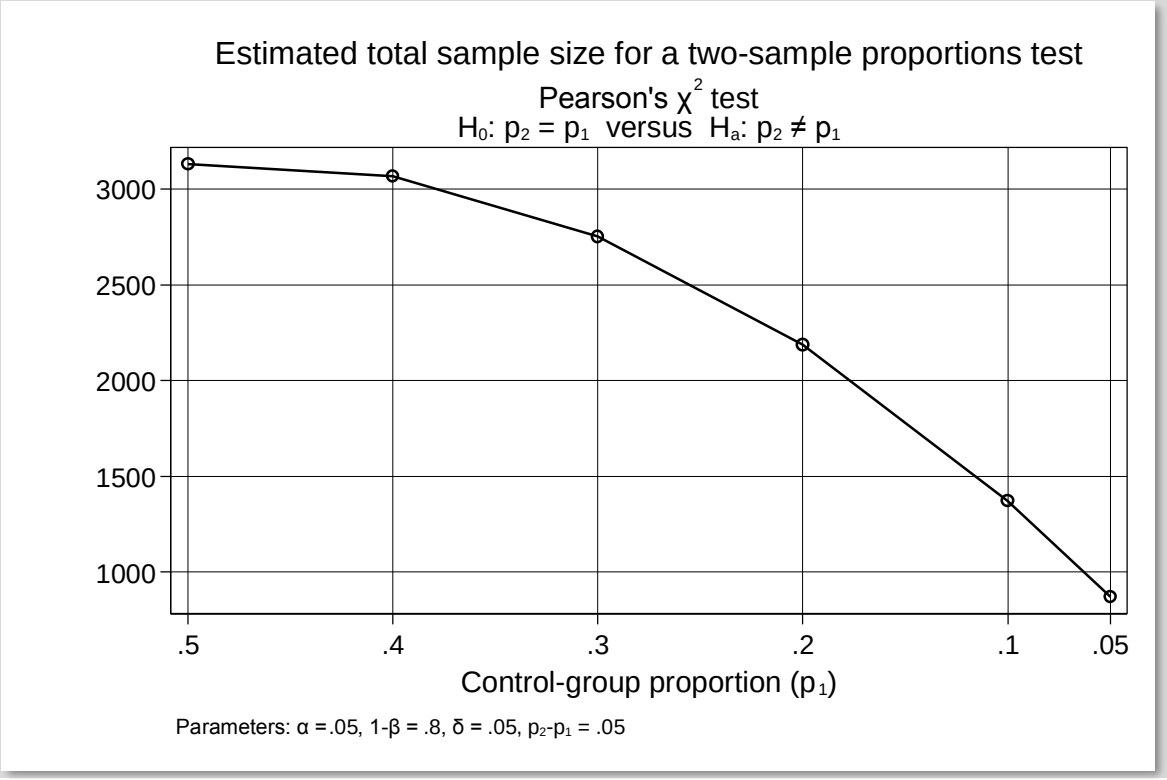
Influence of effect measure (absolute)

Binary outcome ((control) event rate fixed)



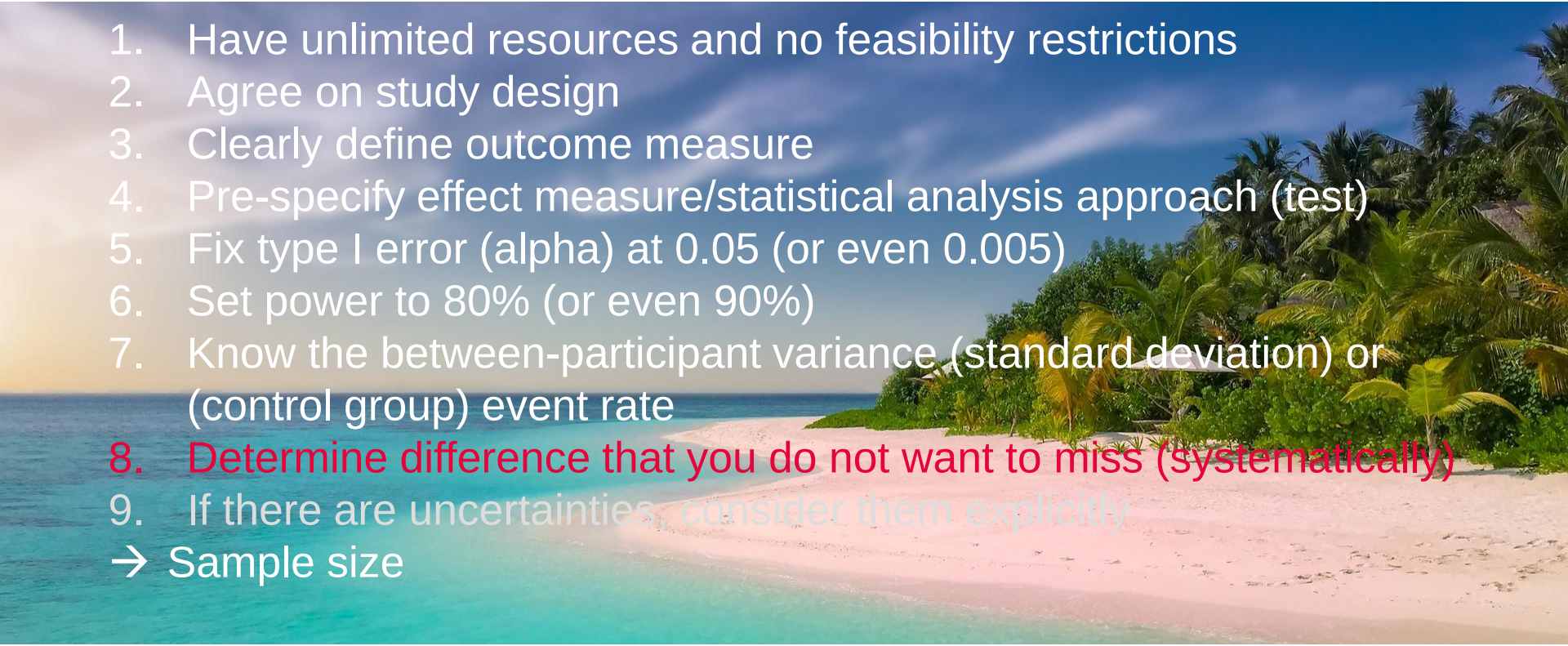
Influence of effect measure (absolute)

Binary outcome (Difference fixed)



Sample size calculation of a controlled trial

In a perfect world



1. Have unlimited resources and no feasibility restrictions
2. Agree on study design
3. Clearly define outcome measure
4. Pre-specify effect measure/statistical analysis approach (test)
5. Fix type I error (alpha) at 0.05 (or even 0.005)
6. Set power to 80% (or even 90%)
7. Know the between-participant variance (standard deviation) or (control group) event rate
8. Determine difference that you do not want to miss (systematically)
9. If there are uncertainties, consider them explicitly

→ Sample size

What is this difference?

Delta (δ)

Interpreting δ

"Minimally clinically relevant difference"

UNIVERSITÄT BERN

- "Minimally clinically relevant difference" taken literally
  → Requires test against δ not 0 (1) i.e. lower confidence interval > δ
- The believed true difference
  → Historical information biased towards larger effects (Gelman & Carlin 2014)
  → Given limited resources, money will need to be spent on interventions with smaller effects (Senn 2014)
- The difference we would like to observe
  → Will result in actual power of ca. 50% not 80%
- **The difference we would not want to miss**

Interpreting δ

"Minimally clinically relevant difference"

UNIVERSITÄT BERN

- "Minimally clinically relevant difference" taken literally
  → Requires test against δ not 0 (1) i.e. lower confidence interval > δ

→ **Important to be clear about null and alternative hypothesis**

Post-hoc power

"Observed power"

UNIVERSITÄT BERN

- Power is a pre-study concept, there is no "observed power"
- Almost always shows a low power (< 50%) for non-significant differences
- Non-significant differences plus high power does not necessarily imply support for the null over (relevant) alternative hypotheses (Greenland 2012)
- Gives overconfidence with respect to significant differences

Other considerations

UNIVERSITÄT BERN

- Time-to-event outcomes → use binary as a start
  - Consider varying length of follow-up, drop-outs, …
- More complex analysis (slide 5 #4) → use simulations
- 'Non-standard' designs
  - Adaptive designs, interim analysis → adjust alpha (α); often size↑
  - Cluster trials → adjust for correlation within clusters (ICC), size↑
  - Dose-finding trials → Modelling, often size not absolutely fixed
  - Historically controlled → multi-stage e.g. Simon's two stage design
- Observational studies → power considerations?

Alternatives

UNIVERSITÄT BERN

- Aim at precision of confidence interval (and do not test hypotheses)
- Bayesian statistics

Thank you

for your attention!

Sven Trelle, CTU Bern

UNIVERSITÄT BERN

References

- Gelman A, Carlin JB 2014. Perspectives on Psychological Science; 9: 641-51.
- Goodman SN, Berlin JA 1994. Ann Intern Med; 121: 200-6.
- Greenland 2012. Ann Epidemiol; 22: 364-8.
- Senn S 2014. https://errorstatistics.com/2014/03/17/stephen-senn-on-how-to-interpret-discrepancies-against-which-a-test-has-high-power-guest-post/

Pictures from

- Pixabay

CAS Clinical Research

https://www.cas-clinicalresearch.ch/

UNIVERSITÄT BERN

Next CTU Lecture

http://www.ctu.unibe.ch/training_courses/ctu_lectures/index_eng.html

UNIVERSITÄT BERN

- Do 14. März 2019
- ZMK
  André Schröder
  Auditorium