$u^b$

# Data Monitoring Committees in Clinical Trials
## Part II: Statistical Issues

**Lukas Bütikofer, CTU Bern**
DCR-CTU Lecture, 15.02.2023

CTU Bern, University of Bern

# Contents

1. Introduction: adaptive designs and interim analysis

2. Operational characteristics of a clinical trial

3. Adaptive designs in detail:

    – Sample size readjustment

    – Formal safety interim analyses

    – Stopping for futility based on conditional power

    – Group sequential designs and stopping for efficacy

    – Optimal zone design

CTU Bern, University of Bern

# Adaptive design

- 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 with **full control of the type I error** (EMA reflection paper on adaptive designs, 2007).

- An adaptive design is defined as a clinical trial design that allows for **prospectively planned modifications** to one or more aspects of the design based on accumulating data from subjects in the trial. (FDA guidance on adaptive designs, 2019)

CTU Bern, University of Bern

# Why adaptive designs?

$u^b$

- **Statistical efficiency**: increased power, smaller expected sample size

- **Ethical**: stop earlier if intervention is not safe or ineffective, allow broader application if intervention is effective

- **Economical**: save resources if intervention is ineffective or enough evidence for effectiveness has already accumulated

- **Administrative**: ensure the trial is conducted as planned (correct population, eligibility criteria)

CTU Bern, University of Bern

# Interim analysis

$u^b$

- **Any examination of data** obtained from subjects in a trial **while that trial is ongoing** and is not restricted to cases in which there are formal between-group comparisons (FDA guidance on adaptive designs, 2019)

- The evaluation of the current data from an ongoing trial, which has the **potential for modifying the conduct of the study** (Whitehead,1999).

CTU Bern, University of Bern

# Modifications at the interim analysis

$u^b$

- Stop for safety

- Readjust sample size (non-comparative vs comparative)

- Stop for futility

- Stop for efficacy (group sequential design)

- *Adapt randomization (adaptive randomization design)*

- *Drop an arm (e.g. drop-the-loser design)*

- *Switch treatment (adaptive treatment-switching design)*

- *Escalate dose (adaptive dose finding design)*

- *Adapt hypothesis (adaptive-hypotheses design)*

- *…*

CTU Bern, University of Bern

# Planning of an interim analysis

$u^b$

- All interim analyses should be carefully **planned in advance** and described in the protocol.

- The **stopping guidelines and their properties** should be clearly described in the protocol or amendments. The potential effects of early stopping on the analysis of other important variables should also be considered.

- This material should be written or approved by the Data Monitoring Committee when the trial has one.

- Deviations from the planned procedure always bear the potential of **invalidating the trial results**.

- The procedures selected should always ensure that **the overall probability of type I error is controlled**.

CTU Bern, University of Bern

# Operational characteristics of a clinical trial

$\boldsymbol{u}^{b}$

- Type I error probability

- Power or type II error probability

- Sample size (expected, maximal)

- Proportion of trials stopped (e.g. for futility or efficacy)

- Duration (needs assumptions about recruitment)

- …

CTU Bern, University of Bern

# Type I and II errors

$u^b$



Null hypothesis: what we would like to reject (e.g. no effect)

Alternative hypothesis: what we would like to find (e.g. a decrease in mortality by 20%)

Type II error probability: Pr(not reject H0 | H1 is true)

Power (1- type II error): Pr(reject H0 | H1 is true)

|  |  | TRUTH | |
|---|---|---|---|
|  |  | H0 | H1 |
| TRIAL | H0 not rejected (null finding) | Ok | Type II error |
|  | H0 rejected (success) | Type I error | Great |

Based on observed data and statistical test(s)

Type I error probability: Pr(reject H0 | H0 is true)

# Power is controlled via the sample size

$u^b$

- Binary endpoint with $p_{control}$ of 50%

- Chi-squared test with an alpha of 0.05

- H0 is well defined and type I error is controlled via test

- Power controlled via sample size but depends on an (arbitrary) effect (H1)

- It is helpful to look at different alternatives

CTU Bern, University of Bern
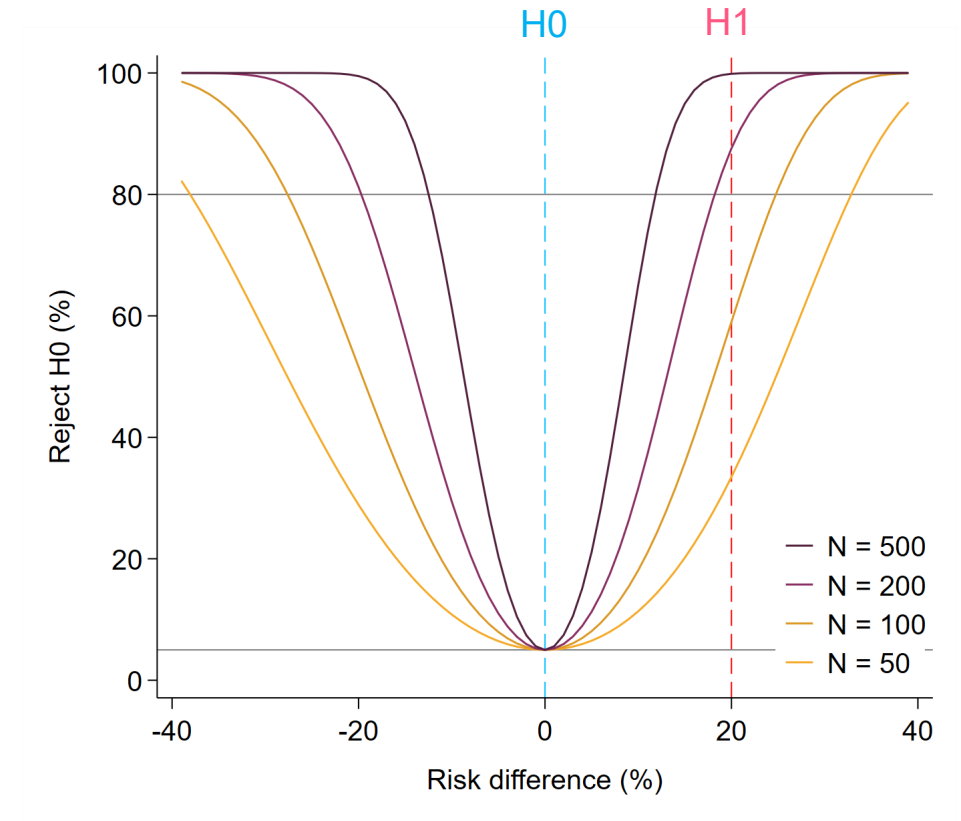
# $u^b$  Multiplicity issue

If there is more than one test, the overall type I error probability is larger than the alpha of the tests

Assuming H0 is true:

Test 1
$\alpha$    $1-\alpha$

Reject H0     Test 2

$\alpha$    $1-\alpha$

Reject H0     Test 3

$\alpha$    $1-\alpha$

Reject H0     …

Family-wise type I error probability:
Pr(reject H0 at any test | H0 is true) = $1 - (1-\alpha)^{\#tests}$

CTU Bern, University of Bern
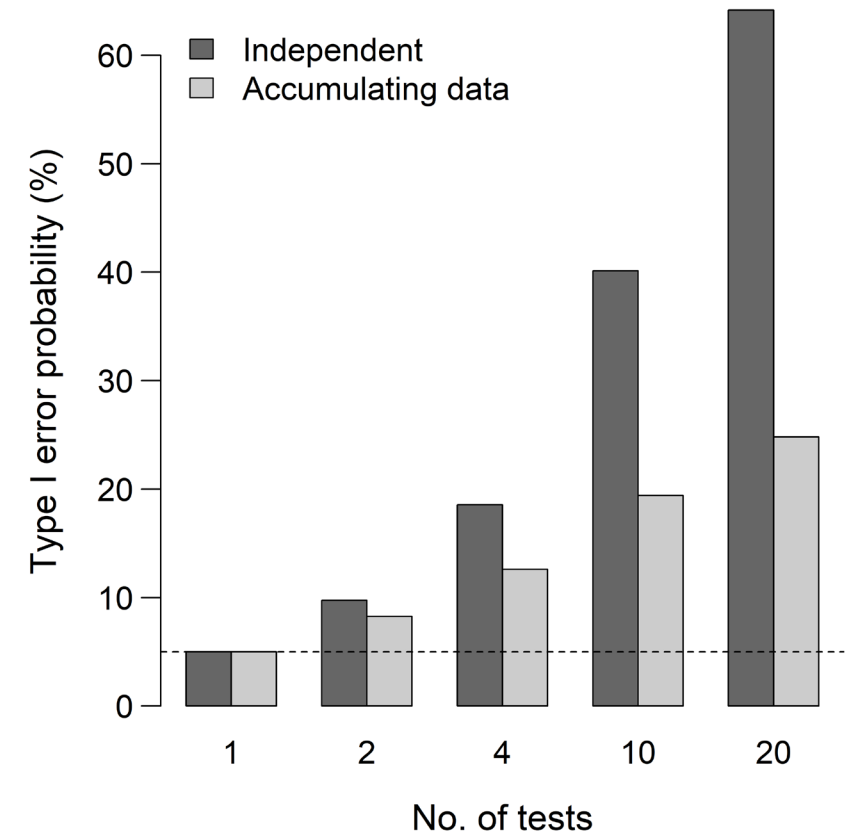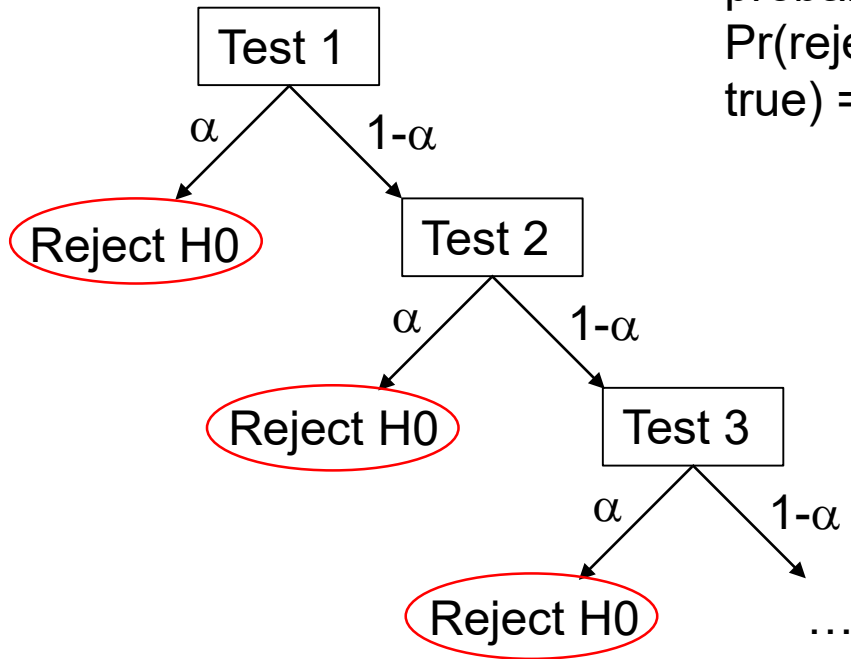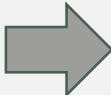
# $u^b$ Expected sample size

- If there are interim analysis, the sample size may depend on the decision taken at interim.

- The expected sample size is then the average (if the trial would be repeated). It is usually assessed under H0 and H1.

- If there is e.g. a stop at interim at N/2 which happens with a probability of 10%, the expected sample size is 0.9*N + 0.1*N/2 = 0.95*N

- The expected may be minimized but the maximal has also to be kept in mind (it must be feasible as it can happen in the actual trial)

CTU Bern, University of Bern

# Effect of interim analyses on operating characteristics

| | Type I error probability | Power | Expected sample size | DMC needed | Binding rules |
|---|---|---|---|---|---|
| Non-comparative sample size readjustment | → | → ↑* | → ↑* | No | Yes |
| Stopping for safety | ↓ | ↓ | ↓ | Yes | No |
| Stopping for futility | ↓ | ↓ | ↓ | Yes | No |
| Stopping for efficacy | ↑ | ↑ | ↓ | No | Yes |
| Comparative sample size readjustment | ↑ | ↑ | ↑ | Yes | Yes |

\* if only increased

CTU Bern, University of Bern

# Non-comparative sample size readjustment

Re-calculate sample size based on non-comparative interim data and adjust if necessary

- When: Uncertainty about nuisance parameters used for sample size calculation (e.g. standard deviation, control proportion or correlation)

➕ Ensure that the desired power is maintained

➖ Logistically difficult as samples size might increase

⚠ Pre-define limits within the sample size is increased

- Adaptations generally do not inflate the type I error probability (Friede and Kieser, 2003)

CTU Bern, University of Bern

# Formal stopping for safety

Stop the trial for safety based on formal rules (binding or non-binding)

- When: there might be some risk associated with the intervention

➕ Support and facilitate decision making of the DMC

➖ Potentially less flexible (if binding rules)

⚠ Consequence of type I and II errors are shifted, i.e. it is typically worse not to stop if there is a safety issue (type II error) than to stop if there is none (type I error)

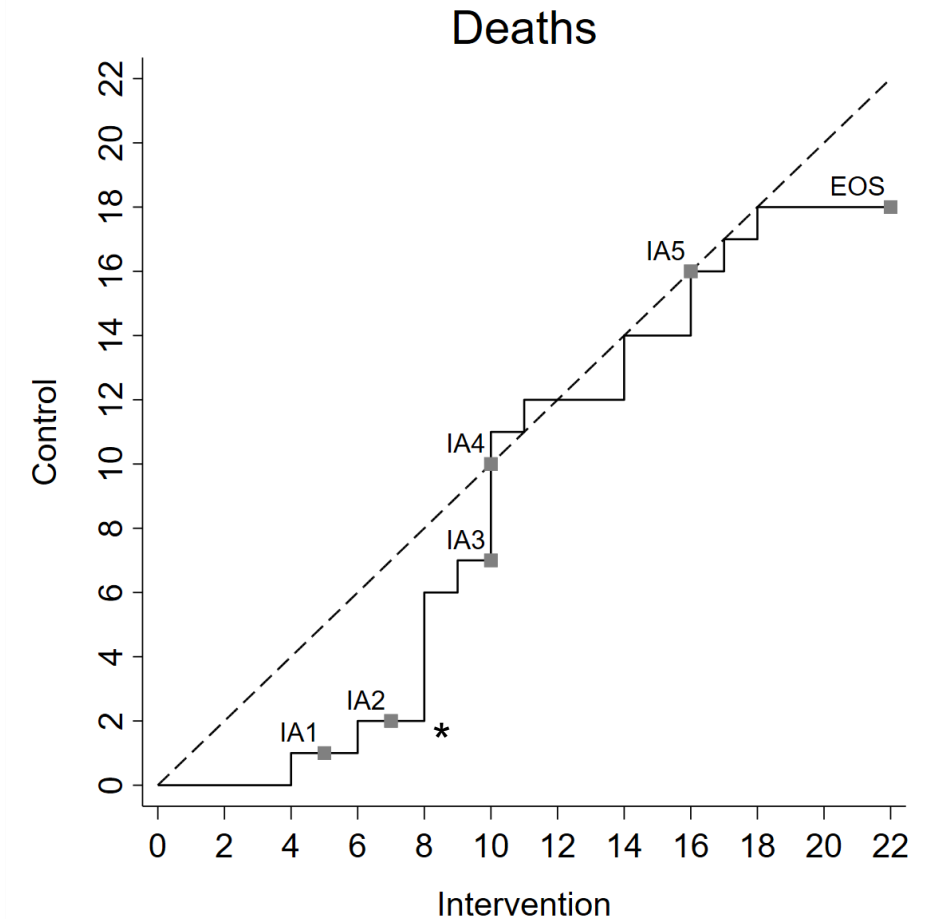→ Strict type I error control may not make sense

CTU Bern, University of Bern

# Example: SWIFT-DIRECT (Fischer, 2023)

Five safety interim analyses:

"If the risk [of the safety outcome] is significantly larger in the intervention group (at a two-sided alpha of 0.05), the DMC will notify the Sponsor that a safety monitoring boundary has been breached, review all available data, and make a recommendation regarding continuation of the study."

| IA | Control | Intervention | P-value |
|----|---------|--------------|---------|
| 1 | 1/25 (4%) | 5/25 (20%) | 0.08 |
| 2 | 2/50 (4%) | 7/50 (14%) | 0.08 |
| * | *2/53 (4%)* | *8/51 (16%)* | *0.039* |
| 3 | 7/77 (9%) | 10/73 (14%) | 0.37 |
| 4 | 10/105 (10%) | 10/97 (10%) | 0.85 |
| 5 | 16/175 (9%) | 16/173 (9%) | 0.84 |



Deaths

CTU Bern, University of Bern

# Simulations to construct rules or guidelines
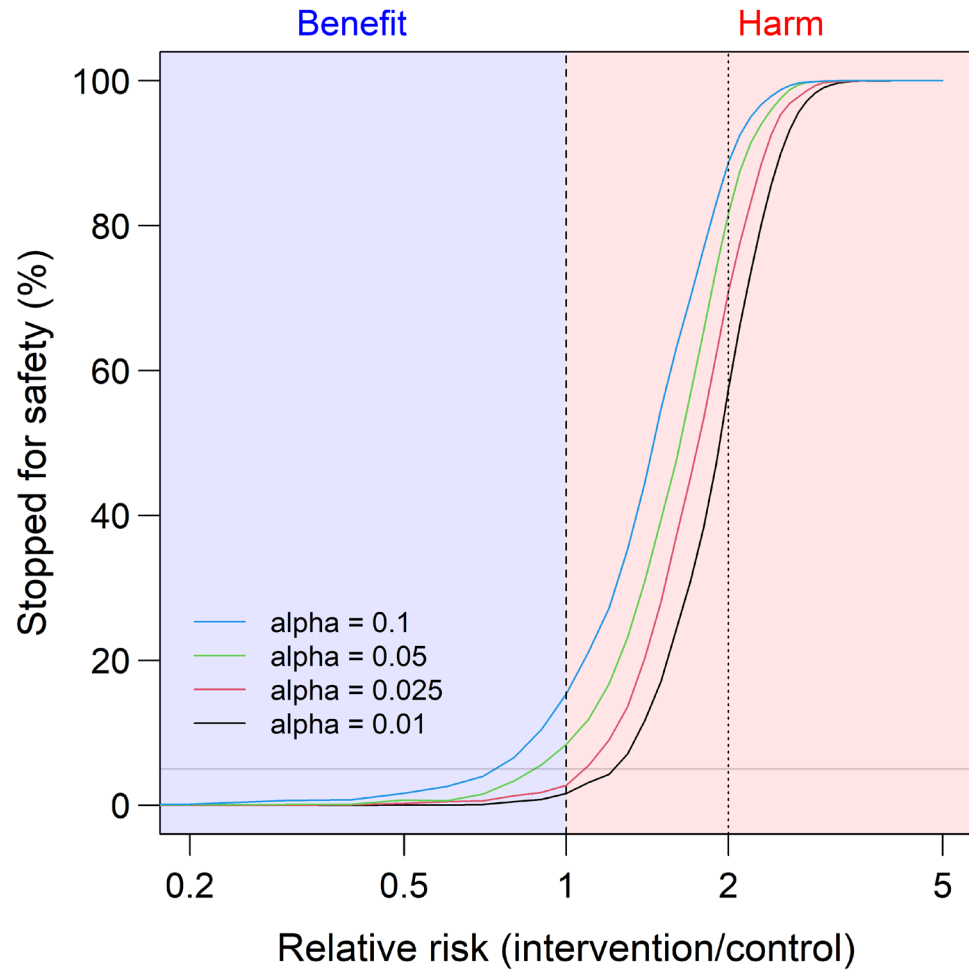


- Strict type I error control may lead to an unacceptable probability of type II error

- The alpha per test can be modified so that enough trials are stopped

- The steepness of the curve depends on the number of events

CTU Bern, University of Bern

# Example: TECNO (in progress, NCT05499832)

- RCT to study the effect of a drug on reperfusion in stroke patients

- Safety interim analyses at 40 and 80 patients, only using the experimental arm

- We will recommend to do DMC to stop the trial if the probability that the prevalence of sICH is >12% in the experimental arm exceeds **80%**

- Calculation will be based on Bayesian methodology with a prior probability of sICH of 8.0% and a weight of 10 patients

CTU Bern, University of Bern
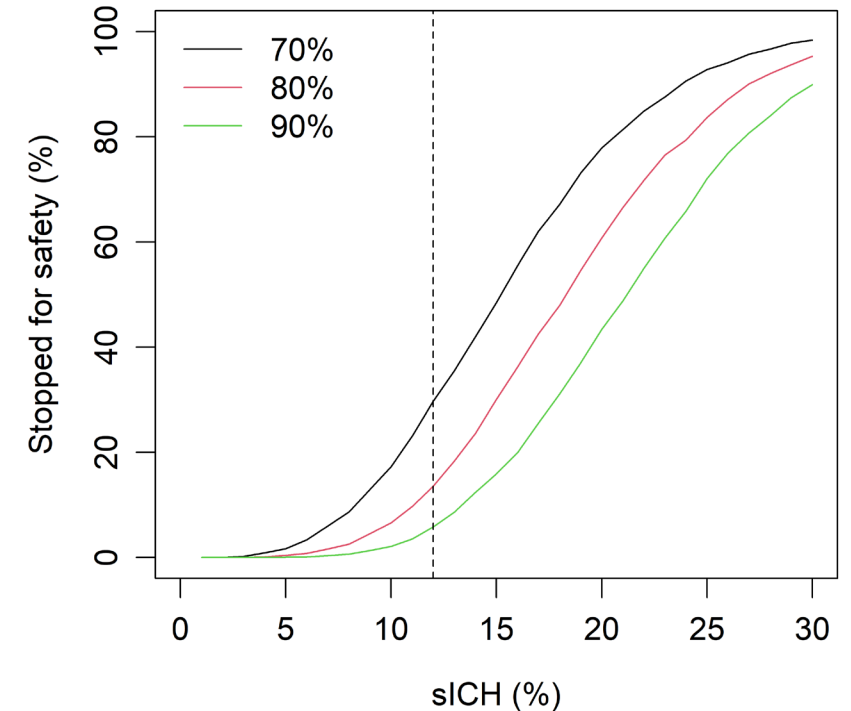
# Stopping for futility

Stop the trial if it is unlikely to reject H0 at the end

- When: large and/or slow recruiting trials with uncertain prior evidence for the intervention

➕ Not exposing patients to an ineffective treatment, save resources

➖ Less data on secondary outcomes or subgroups, reduced power

CTU Bern, University of Bern

# Simon's two stage design (Simon, 1989)

- One-arm phase-II trial with an assessment for futility

- Completely predefined

- Simple but not very flexible

- $N_1$ patients are recruited

- If there are $\leq r_1$ success, the trial will be stopped for futility

- Otherwise it continuous until N patients are recruited

CTU Bern, University of Bern

# Simon's two-stage design (cont'd)

Assumptions:

- Uninteresting $p_{H0}$: 30%

- Good $p_{H1}$ : 60%

- Type I error: 5%

- Power: 80%

Design and operational characteristics:

- Minimax design: 2/10, 8/17

- Expected sample size: 14.3, 16.9

- Probability of stop: 38.3%, 1.2%

- Probability of type 1 error:  4%

- Probability of type 2 error: 19.9%

Sample size

CTU Bern, University of Bern

# Conditional power

$u^b$
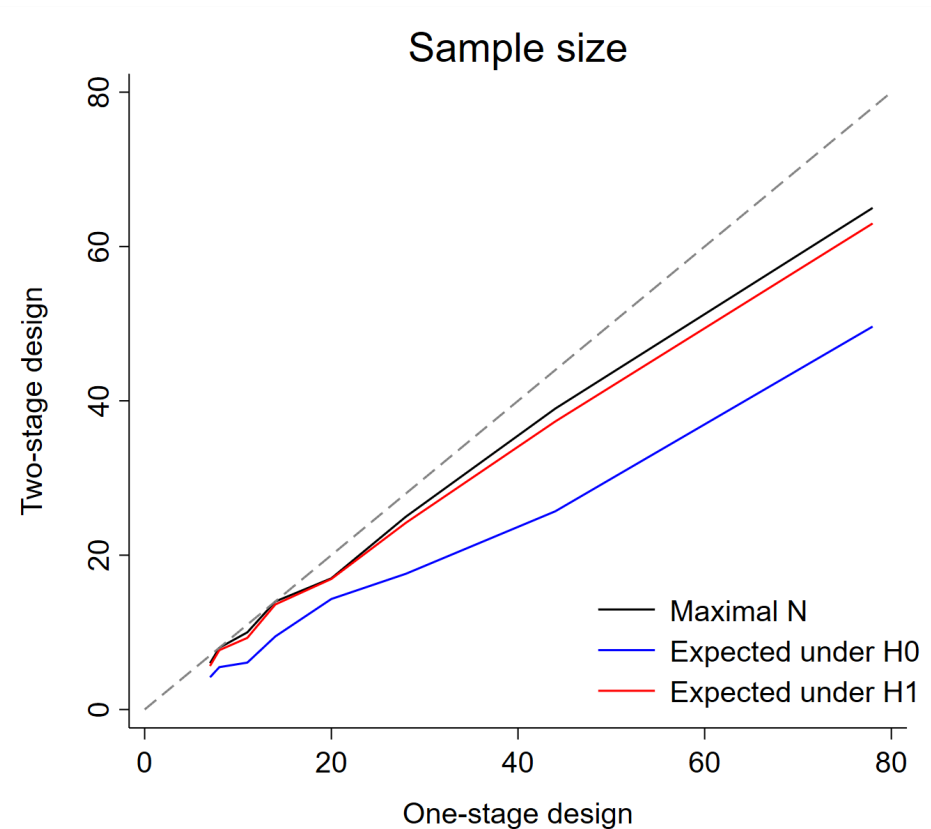
- Power to reject H0 at the end of the trial given the data at interim an some assumption about the future effect

- The future effect usually is the one from sample size calculation (H1) or the one observed at interim

- If it is low (e.g. <20%), the trial might be stopped

- Conservative regarding type I error (is re-used in some designs)

- Non-binding rules (unless re-used)

- Loss of power, simulations might be necessary

CTU Bern, University of Bern

# Example

- Binary endpoint with $p_{control}$: 40%

- H1: risk difference of 20% (i.e. $p_{exp}$: 60%)

- N = 200, power $\approx$ 80%

- Interim analysis after 100 patients, observed $p_{control}$ = 40%

CTU Bern, University of Bern
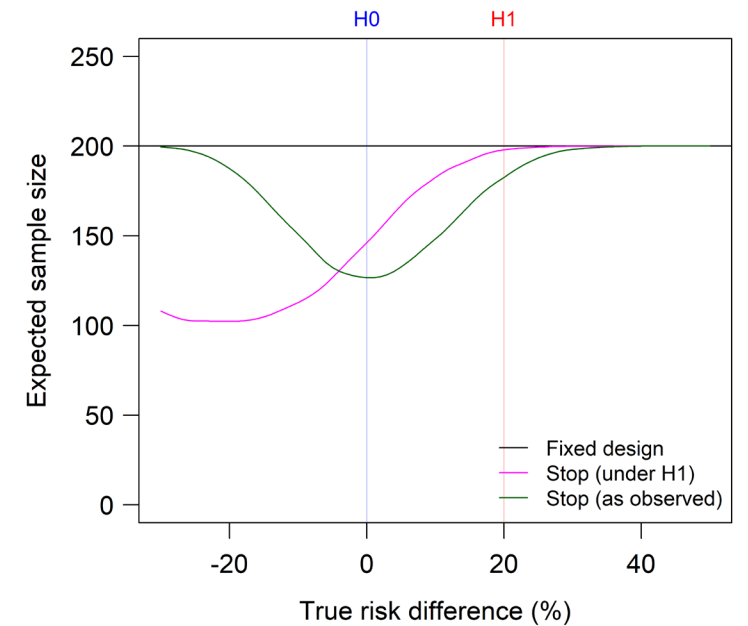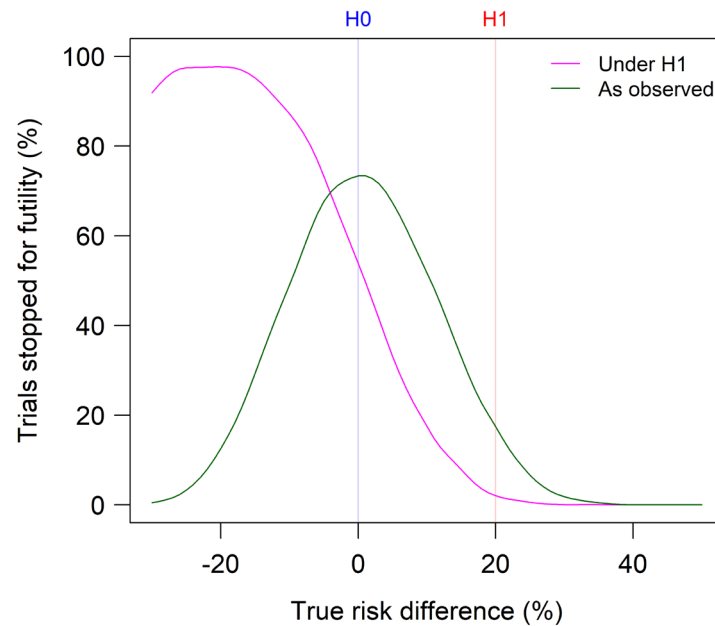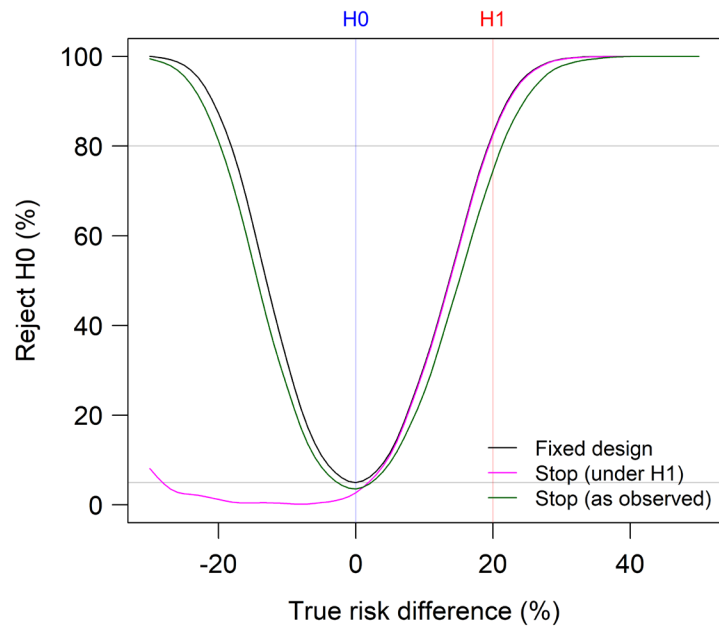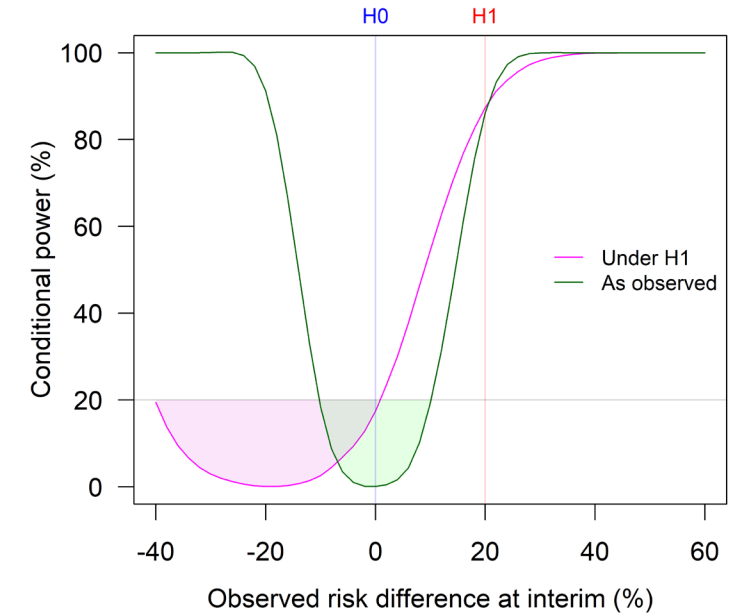
# Stopping for efficacy

Stop the trial if there is enough evidence that the treatment is effective at one or more prospectively planned interim analyses of comparative data

- When: Slow recruiting and/or large trials with good prior evidence for the intervention

➕ Reduce expected sample size and duration, accelerate the approval of new treatments

➖ Not enough power for secondary or safety outcomes, final analysis at a reduced nominal alpha, more complicated design

# Group sequential designs

*$u^b$*

- Type I error inflation due to multiple testing: adjustment using group sequential designs

- E.g. Trial with 300 patients and 3 interim analyses:
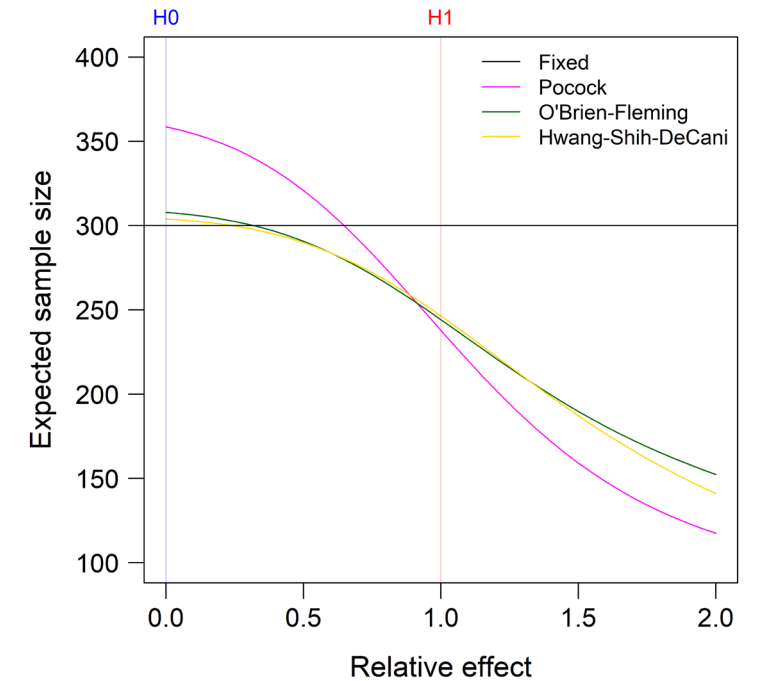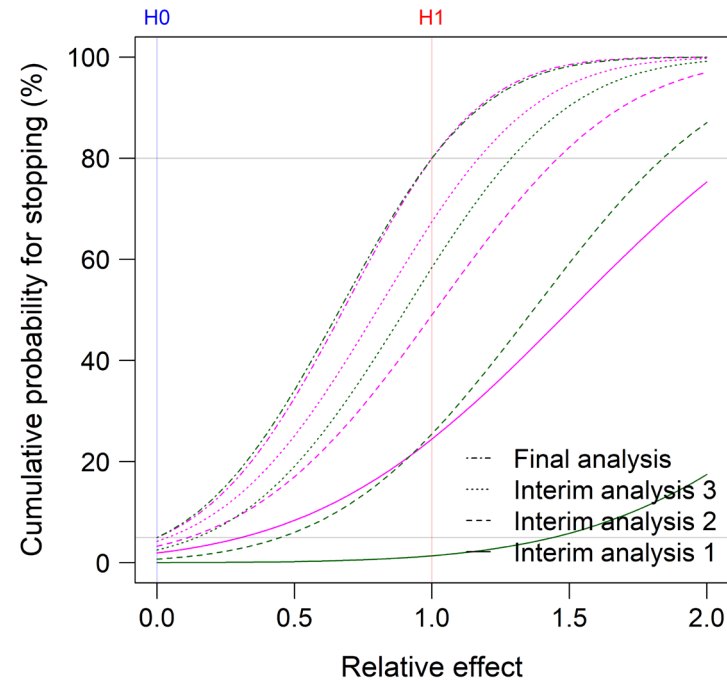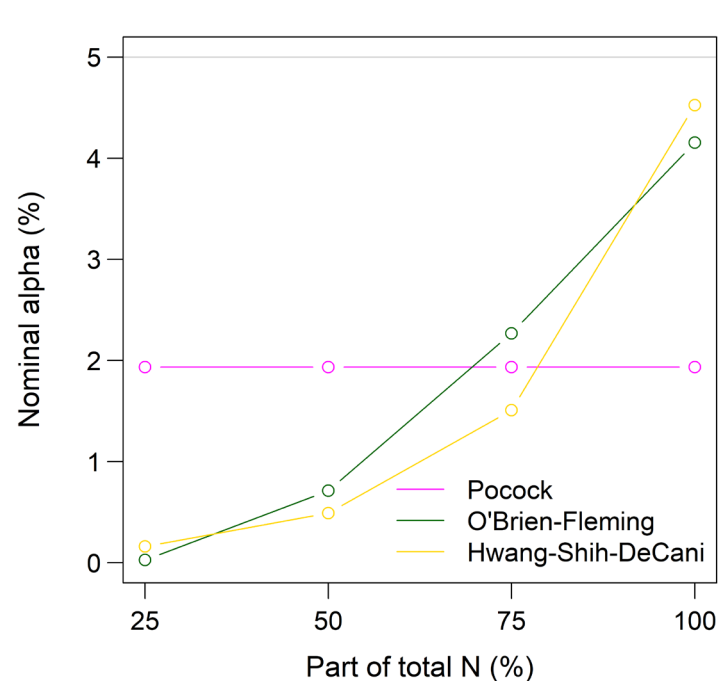
CTU Bern, University of Bern

O'Brien and Fleming, 1979; Pocock 1977

# Comparative sample size readjustment

$u^b$

Increase sample size based on the observed effect

- When: Slow recruiting and/or large trials with uncertainty about the expected (or relevant) effect size

➕ Adapt sample size selectively, i.e. increase in promisingtrials

➖ Design more complicated, simulation are necessary, logistically difficult as samples size might increase

$u^b$ # Control of type I error inflation

- When sample size is increased in a data-dependent manner, the type I error is increased (Cui 1999, Proschan 1995).

- Correction possible (CHW statistics) but suboptimal properties

- No inflation if sample size is only increased when interim results are promising based on conditional power (Chen 2004)

- Promising zone design (Mehta, 2011)

- Interim analysis after N/2
- Recalculate sample size based on the observed effect
- Increase if less than 10%, 50% or 100% additional samples are needed

CTU Bern, University of Bern
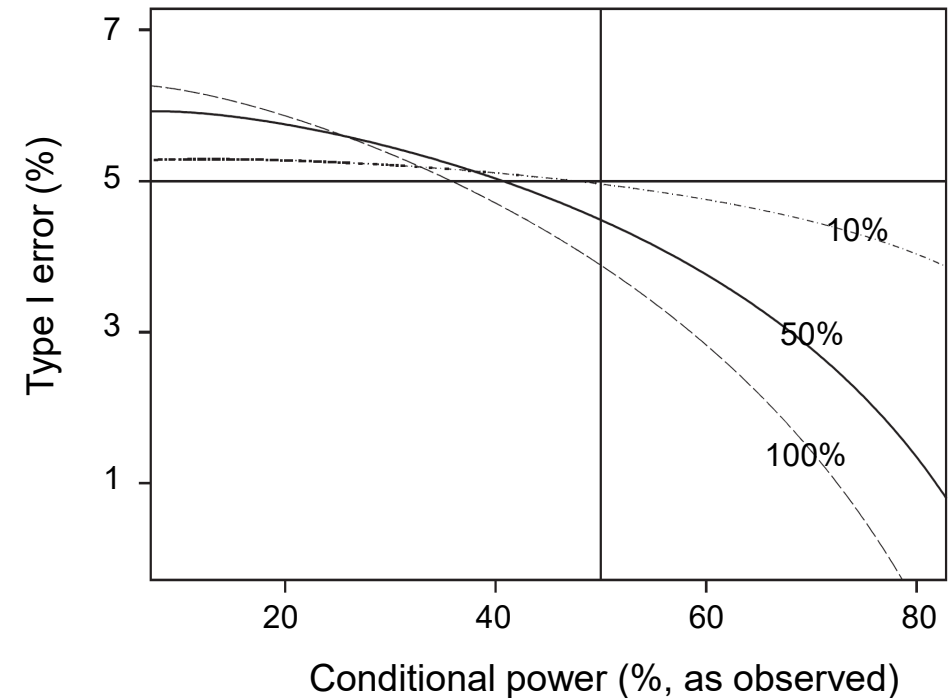
# Example: OTTAWA (in planning, submitted to IICT call)

- RCT with a time-to-event primary outcome

- Initial sample size calculation:

  - $p_{control}$: 27% after two years

  - Relative risk reduction (RRR): 40% after two years

  - Log-rank-test with an alpha of 5%

  - 99 events corresponding to 482 patients assuming uniform recruitment over 4 years and 0.5 year of additional follow-up

- Adaptations: Increase sample size based on the observed effect (to reach a conditional power of 80% or a maximum of 800) and stop for futility

CTU Bern, University of Bern

$u^b$
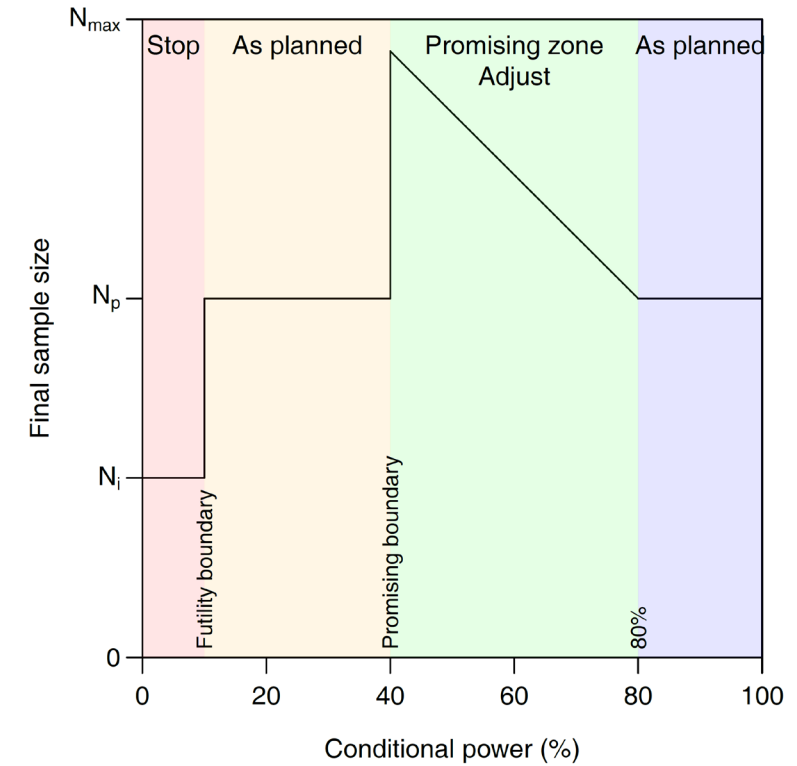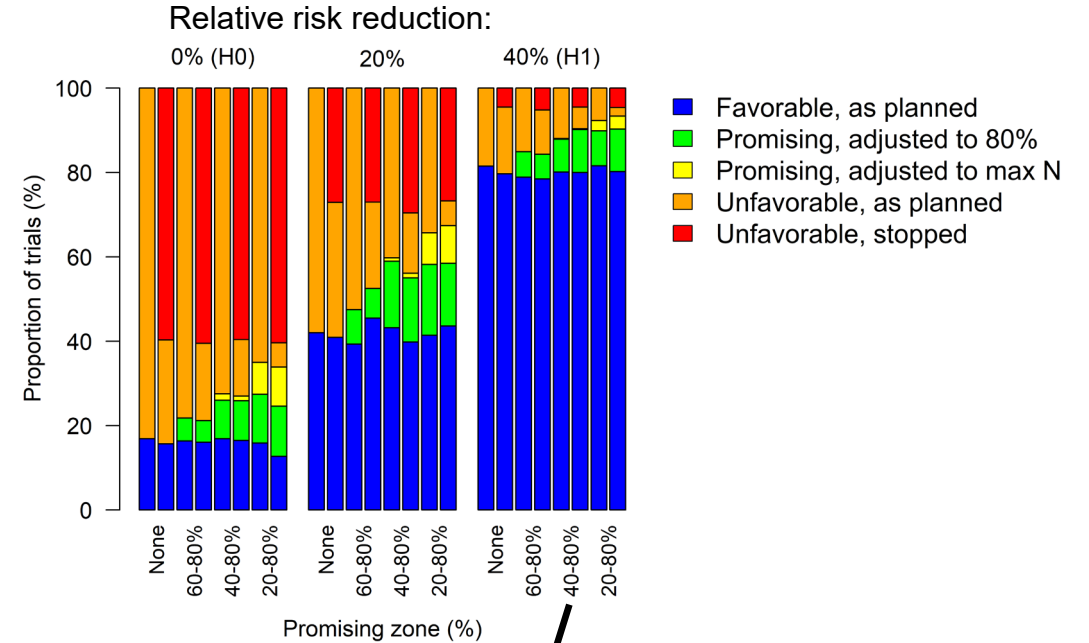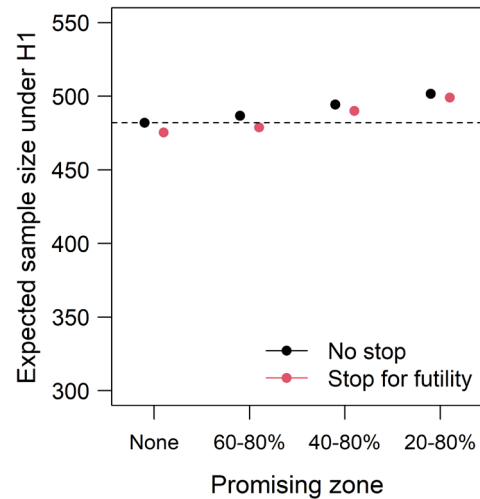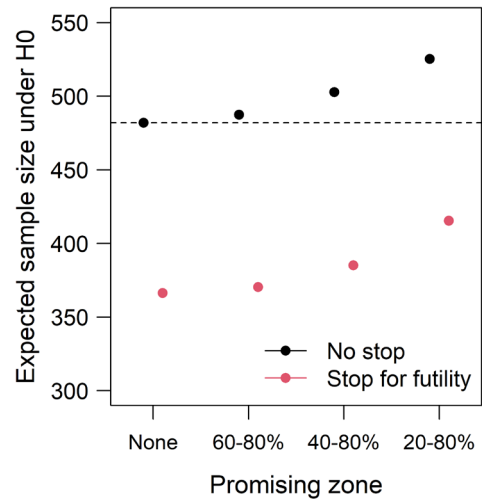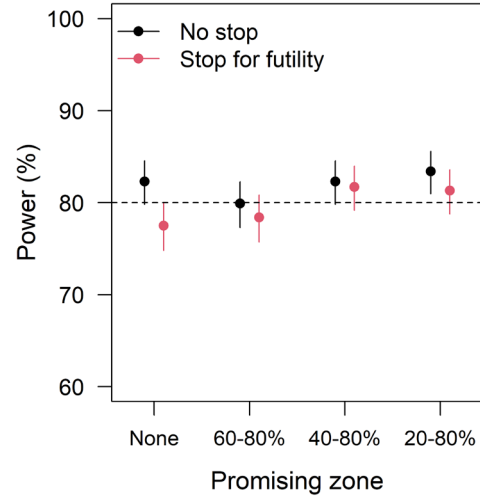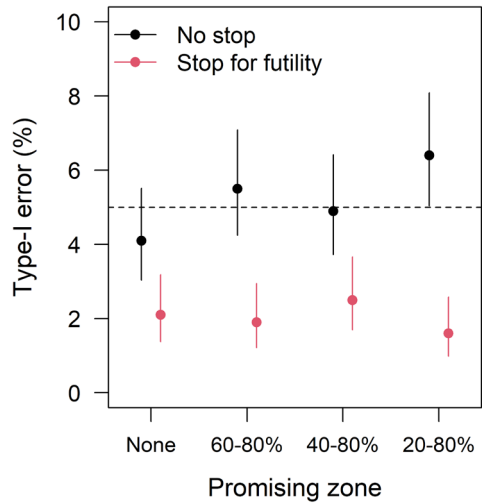


| Interim result | Probability of interim result (%) | Power (%) Adaptive | Fixed | Expected N Adaptive | Fixed |
|---|---|---|---|---|---|
| Favorable | 76 | 90 | 91 | 482 | 482 |
| **Promising** | **12** | **81** | **60** | **617** | **482** |
| Unfavorable | 7 | 33 | 40 | 482 | 482 |
| Stopped | 5 | 0 | 16 | 344 | 482 |

→ Selective power increase for promising trials

CTU Bern, University of Bern

# References

- Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned with an Adaptive Design, EMA, 2007

- E9 Statistical principles for clinical trials, ICH, 1998

- Adaptive Design Clinical Trials for Drugs and Biologics, FDA, 2019

- Chen YH, DeMets DL, Lan KK. Increasing the sample size when the unblinded interim result is promising. *Statistics in Medicine* 2004; **23**:1023--1038.

- Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics* 1999; **55**:853-857

- Fischer, Urs, et al. "Thrombectomy alone versus intravenous alteplase plus thrombectomy in patients with stroke: an open-label, blinded-outcome, randomised non-inferiority trial." The Lancet 400.10346 (2022): 104-115.

- Friede, T, and M Kieser, 2003, Blinded sample size reassessment in non-inferiority and equivalence trials, Stat Med, 22(6):995-1007.

- Lan, KG and DL DeMets, 1983, Discrete Sequential Boundaries for Clinical Trials, Biometrika, 70(3):659–663

- Mehta, Cyrus R., and Stuart J. Pocock. "Adaptive increase in sample size when interim results are promising: a practical guide with examples." *Statistics in medicine* 30.28 (2011): 3267-3284.

- O'Brien, PC and TR Fleming. A Multiple Testing Procedure for Clinical Trials, 1979, Biometrics, 35(3):549–556.

- Pocock, SJ, 1977, Group Sequential Methods in the Design and Analysis of Clinical Trials, Biometrika, 64(2):191–199.

- Proschan MA, Hunsberger SA. Designed extension of studies based on conditional power. *Biometrics* 1995; **51**:1315-1324

- Simon, Richard. "Optimal two-stage designs for phase II clinical trials." Controlled clinical trials 10.1 (1989): 1-10.

- Whitehead, John, et al. "Interim analyses in clinical trials." *British Journal of Clinical Pharmacology* 51.5 (2001): 393.

CTU Bern, University of Bern