# Sample Size Estimation

## why do we need it and what do we need for it?

Marco Cattaneo

Department of Clinical Research
University of Basel

29 October 2019

# outline

# what does it mean?

- *sample size*: number of patients (or participants) to recruit for the study

# what does it mean?

- *sample size*: number of patients (or participants) to recruit for the study
- *estimation* / *calculation* / *determination*: misleading terminology (no correct value to determine, no statistical noise, maybe *evaluation* / *planning* / *discussion* better)

# what does it mean?

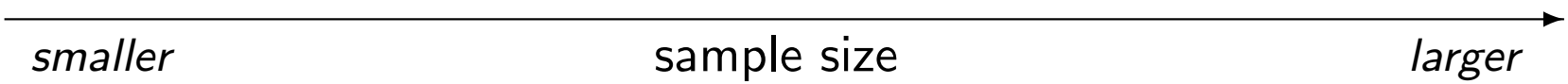
- *sample size*: number of patients (or participants) to recruit for the study
- *estimation* / *calculation* / *determination*: misleading terminology (no correct value to determine, no statistical noise, maybe *evaluation* / *planning* / *discussion* better)
- SSE is about finding an equilibrium:

**more potential value:**

- less precision
- higher risk of inconclusive results
- *higher risk of giving patients suboptimal treatment for nothing*

**more actual costs:**

- more money
- more time
- *more patients receiving suboptimal treatment*



# why do we need it?

- to allow **feasibility** evaluation

# why do we need it?

- to allow **feasibility** evaluation
- to help time/budget planning

# why do we need it?

- to allow **feasibility** evaluation
- to help time/budget planning
- to help clarifying ethical soundness

# why do we need it?

- to allow **feasibility** evaluation
- to help time/budget planning
- to help clarifying ethical soundness
- to promote contact with **statistician** at right point in time

  *To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.* (Fisher, 1938)

# why do we need it?

- to allow **feasibility** evaluation
- to help time/budget planning
- to help clarifying ethical soundness
- to promote contact with **statistician** at right point in time

  *To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.* (Fisher, 1938)
- to induce clarification and planning of primary outcomes and their statistical analyses (rather than data dredging)

  *If you torture the data enough, nature will always confess.* (Coase, 1982)

# my experience so far

- in my first year at CTU Basel, I performed 10 SSE

# my experience so far

- in my first year at CTU Basel, I performed 10 SSE
- studies:
  - 9x RCT (goal: significance test):
  - 1x survey (goal: CI length)

# my experience so far

- in my first year at CTU Basel, I performed 10 SSE
- studies:
  - 9x RCT (goal: significance test):
    - 8x superiority trial
    - 1x noninferiority trial
  - 1x survey (goal: CI length)

# my experience so far

- in my first year at CTU Basel, I performed 10 SSE
- studies:
  - 9x RCT (goal: significance test):
    - 8x superiority trial
    - 1x noninferiority trial
  - 1x survey (goal: CI length)
- endpoints:
  - 6x continuous
  - 4x binary (including time-to-event)

# my experience so far

- in my first year at CTU Basel, I performed 10 SSE
- studies:
  - 9x RCT (goal: significance test):
    - 8x superiority trial
    - 1x noninferiority trial
  - 1x survey (goal: CI length)
- endpoints:
  - 6x continuous
  - 4x binary (including time-to-event)
- range of sample sizes: 20 – 50 000

# what do we need for it?

- **primary endpoint** (a precisely defined measurement)

# what do we need for it?

- **primary endpoint** (a precisely defined measurement)
- **statistical model** (often simplified version of the analysis one)

# what do we need for it?

- **primary endpoint** (a precisely defined measurement)
- **statistical model** (often simplified version of the analysis one)
- **significance level** (probability of false positives): lower level $\rightsquigarrow$ larger SS

# what do we need for it?

- **primary endpoint** (a precisely defined measurement)
- **statistical model** (often simplified version of the analysis one)
- **significance level** (probability of false positives): lower level $\rightsquigarrow$ larger SS
- possibly also **noninferiority/equivalence margins** (clinically irrelevant differences): margins nearer to 0 $\rightsquigarrow$ larger SS

# what do we need for it?

- **primary endpoint** (a precisely defined measurement)
- **statistical model** (often simplified version of the analysis one)
- **significance level** (probability of false positives): lower level $\rightsquigarrow$ larger SS
- possibly also **noninferiority/equivalence margins** (clinically irrelevant differences): margins nearer to 0 $\rightsquigarrow$ larger SS

---

- information about (multivariate) **distribution** of endpoint (and covariates), such as variability: higher variability $\rightsquigarrow$ larger SS

# what do we need for it?

- **primary endpoint** (a precisely defined measurement)
- **statistical model** (often simplified version of the analysis one)
- **significance level** (probability of false positives): lower level $\rightsquigarrow$ larger SS
- possibly also **noninferiority/equivalence margins** (clinically irrelevant differences): margins nearer to 0 $\rightsquigarrow$ larger SS

---

- information about (multivariate) **distribution** of endpoint (and covariates), such as variability: higher variability $\rightsquigarrow$ larger SS
- **power** (probability of true positives): higher power $\rightsquigarrow$ larger SS

# what do we need for it?

- **primary endpoint** (a precisely defined measurement)
- **statistical model** (often simplified version of the analysis one)
- **significance level** (probability of false positives): lower level $\rightsquigarrow$ larger SS
- possibly also **noninferiority/equivalence margins** (clinically irrelevant differences): margins nearer to 0 $\rightsquigarrow$ larger SS

---

- information about (multivariate) **distribution** of endpoint (and covariates), such as variability: higher variability $\rightsquigarrow$ larger SS
- **power** (probability of true positives): higher power $\rightsquigarrow$ larger SS
- hypothetical **treatment effect** (or clinically relevant difference): effect nearer to 0 $\rightsquigarrow$ larger SS

# what do we need for it?

- **primary endpoint** (a precisely defined measurement)
- **statistical model** (often simplified version of the analysis one)
- **significance level** (probability of false positives): lower level $\rightsquigarrow$ larger SS
- possibly also **noninferiority/equivalence margins** (clinically irrelevant differences): margins nearer to 0 $\rightsquigarrow$ larger SS

---

- information about (multivariate) **distribution** of endpoint (and covariates), such as variability: higher variability $\rightsquigarrow$ larger SS
- **power** (probability of true positives): higher power $\rightsquigarrow$ larger SS
- hypothetical **treatment effect** (or clinically relevant difference): effect nearer to 0 $\rightsquigarrow$ larger SS
- **dropout/nonresponse rate** (proportion of patients with incomplete results): higher rate $\rightsquigarrow$ larger SS

# example: superiority trial, continuous endpoint

- **primary endpoint**: continuous measurement at follow-up (larger is better)

# example: superiority trial, continuous endpoint

- **primary endpoint**: continuous measurement at follow-up (larger is better)
- **statistical model**: baseline-adjusted ANCOVA

# example: superiority trial, continuous endpoint

- **primary endpoint**: continuous measurement at follow-up (larger is better)
- **statistical model**: baseline-adjusted ANCOVA
- **significance level**: 5%

# example: superiority trial, continuous endpoint

- **primary endpoint**: continuous measurement at follow-up (larger is better)
- **statistical model**: baseline-adjusted ANCOVA
- **significance level**: 5%
- **noninferiority/equivalence margins**: not needed

# example: superiority trial, continuous endpoint

- **primary endpoint**: continuous measurement at follow-up (larger is better)
- **statistical model**: baseline-adjusted ANCOVA
- **significance level**: 5%
- **noninferiority/equivalence margins**: not needed

---

- **distribution**: bivariate distribution of baseline and follow-up measurements

# example: superiority trial, continuous endpoint

- **primary endpoint**: continuous measurement at follow-up (larger is better)
- **statistical model**: baseline-adjusted ANCOVA
- **significance level**: 5%
- **noninferiority/equivalence margins**: not needed

---

- **distribution**: bivariate distribution of baseline and follow-up measurements
- **power**: 80%

# example: superiority trial, continuous endpoint

- **primary endpoint**: continuous measurement at follow-up (larger is better)
- **statistical model**: baseline-adjusted ANCOVA
- **significance level**: 5%
- **noninferiority/equivalence margins**: not needed

---

- **distribution**: bivariate distribution of baseline and follow-up measurements
- **power**: 80%
- **treatment effect**: additional effect of treatment vs control is 40

# example: superiority trial, continuous endpoint

- **primary endpoint**: continuous measurement at follow-up (larger is better)
- **statistical model**: baseline-adjusted ANCOVA
- **significance level**: 5%
- **noninferiority/equivalence margins**: not needed

---

- **distribution**: bivariate distribution of baseline and follow-up measurements
- **power**: 80%
- **treatment effect**: additional effect of treatment vs control is 40
- **dropout/nonresponse rate**: not considered (divide SS by 1 − rate)

# example: noninferiority trial, continuous endpoint

- **primary endpoint**: continuous measurement at follow-up (larger is better)
- **statistical model**: baseline-adjusted ANCOVA
- **significance level**: 5%
- **noninferiority/equivalence margins**: treatment noninferior to control if effect $> -20$

---

- **distribution**: bivariate distribution of baseline and follow-up measurements
- **power**: 80%
- **treatment effect**: additional effect of treatment vs control is 40
- **dropout/nonresponse rate**: not considered (divide SS by 1 − rate)

# example: superiority trial, binary endpoint

- **primary endpoint**: binary version of previous continuous endpoint
- **statistical model**: baseline-adjusted logistic regression (baseline is also binary)
- **significance level**: 5%
- **noninferiority/equivalence margins**: not needed

---

- **distribution**: bivariate distribution of baseline and follow-up measurements
- **power**: 80%
- **treatment effect**: additional effect of treatment vs control is 40
- **dropout/nonresponse rate**: not considered (divide SS by 1 − rate)

# what do we need for it? (CI version)

- **parameter of interest** (for which CI should be constructed)
- **statistical method** (to construct CI)
- **confidence level** (coverage probability): higher level $\rightsquigarrow$ larger SS

---

- **desired interval length** (maximal): shorter intervals $\rightsquigarrow$ larger SS
- information about (multivariate) **distribution** of involved measurements, such as variability: higher variability $\rightsquigarrow$ larger SS
- **probability of desired length** (or shorter): higher probability $\rightsquigarrow$ larger SS
- **dropout/nonresponse rate** (proportion of patients with incomplete results): higher rate $\rightsquigarrow$ larger SS

# example: CI length, continuous endpoint

- **parameter of interest**: baseline-adjusted treatment effect on previous continuous endpoint
- **statistical method**: ANCOVA CI
- **confidence level**: 95%

---

- **desired interval length**: 100
- **distribution**: bivariate distribution of baseline and follow-up measurements (with arbitrary treatment effect)
- **probability of desired length**: 90%
- **dropout/nonresponse rate**: not considered (divide SS by 1 − rate)