INSTITUT
JULES BORDET
INSTITUUT

Clinical trials methodology

Marianne Paesmans – PGMO 15/05/2021

ULB iris

Clinical trials categorization

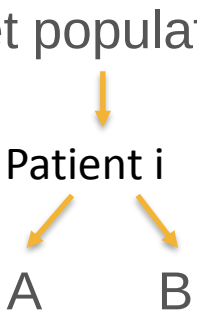
- Reminder : clinical trial = interventional study with investigational medicinal product
- Main objective : to get information about the IMP(s)

Category	Objective	N patients	Statistical design
Phase I	Safety	Small	No true inference
Phase I/II	Safety / Preliminary efficacy	Small	Accuracy of estimation / hypothesis testing
Phase II	Early efficacy	Medium	Hypothesis testing
Phase III	Efficacy compared to standard	Large to very large	Hypothesis testing
Phase IV	Pharmacovigilance	Large	Estimation

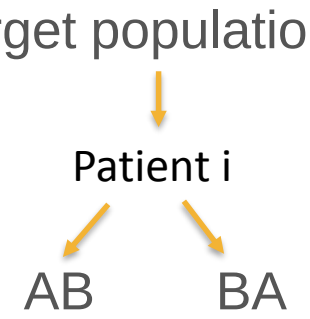
Importance of randomization <-> Personalized medicine

Bias control

- Bias = systematic error
- Randomization : makes the patients « groups » comparable and different only by the intervention (known and unknown confounders, valid p values)
- Parallel design



Cross-over design



- Cross-over design : each patient = own control
- Factorial design : two by two randomisation

Randomization

- Does not imply that groups will be formally compared
- Ratio 1 to 1 :
  -statistically most efficient for comparative trials
  -equipoise principle
- Ratio 2 to 1 might be considered when
  -more data needed on a specific endpoint in one arm
  -non comparative purpose
  -not as an incentive for patients ?

Trial objectives

- Defining the question(s)
- Defining the intervention(s)
- Defining the target patients population (eligibility criteria)

Trial objectives

- Primary versus secondary, related to trial phase
- Primary
  - most often 1 (or 2)
  - most clinically relevant, with assessment not subject to bias
  - will drive sample size :
    - estimation accuracy -> confidence interval
    - hypothesis testing : with control of type I error (α) and type II error (β)
- Secondary
  - control of α only, allow more complete evaluation with risk-benefit balance
- Exploratory

Primary objective(s)

Translated into an hypothesis test :

H_0 : null hypothesis versus H_1 : alternative hypothesis

Exemples :
H_0 : pCR < 20% versus H_1 : pCR ≥ 20% (one-sided alternative)
H_0 : $S_E(t)=S_C(t)$ versus H_1 : $S_E(t)\neq S_C(t)$ (two-sided alternative)

	H_0 rejected	Fail to reject H_0
H_0 false	Correct	Type II error
H_0 true	Type I error	correct

Alpha (α) = Prob (Type I error)

Beta (β) = Prob (Type II error)

Power = $1-\beta$

Sample size
(control of random errors)
driven by :
- α
- β
- detectable difference

If more than one 1ary :
adjustment for multiplicity will be needed

A priori sample size calculation, timing of analysis and planning of interim analyses

INSTITUT JULES BORDET INSTITUUT

ULB iris

Example

- Hypothesis : immune checkpoint inhibitor added to neoadjuvant chemotheray will increase pCR in patients operable bladder cancer
- CT alone : expected 20% pCR
- H_0 : pCR ≤ 20% versus H_1 pCR > 20%
- Phase II design (randomized or not) :

Detectable pCR	1-tailed α	β	n
30%	5%	10%	156
35%	5%	10%	72
40%	5%	10%	42
40%	10%	10%	33
40%	5%	20%	29

INSTITUT JULES BORDET INSTITUUT

ULB iris

Example

- Phase III superiority design
- $H_0 : p_E CR = p_C CR$ versus $H_1 : p_E CR <> p_C CR$
- Expected : $p_C CR = 20\%$

True pCR with exp	2-tailed α	β	n
30%	5%	10%	2*392
35%	5%	10%	2*185
40%	5%	10%	2*109
40%	10%	10%	2*89
40%	5%	20%	2*82

- Phase III non inferiority design
- $H_0 : p_E CR - \varepsilon \leq p_C CR$ versus $H_1 : p_E CR - \varepsilon > p_C CR$
- Expected : $p_C CR = 30\%$, $\varepsilon = 10\%$

True pCR with exp	2-tailed α	β	n
30%	5%	10%	2*442

INSTITUT JULES BORDET INSTITUUT

ULB iris

Comparative trials : hypothesis testing

- Superiority : to show experimental arm better than control
- Equivalence : to show experimental arm sufficiently close than control
- Non inferiority : to show experimental is not worse than control by a small amount
- In equivalence / non inferiority trials : detectable difference should be small -> large sample size
- Adaptive designs : prospectively planned to change design or hyotheses based on interim data

Outcomes / endpoints

- Should match the objectives : efficacy, safety, costs, PROs, prediction, compliance, …
- Measured on **each patient** included in a trial
- Objectives : reached or unreached by data aggregation on endpoints
- Types of endpoints :
  - Binary
  - Categorical
  - Continuous
  - Time-to-event

Primary endpoint

- Clinically relevant
- Accurate and reliable measurement
- High probability of being assessable in all patients
- Assessment not linked to treatment arm
- Subjectivity in assessment <-> need of blinding
- **Bias control**
- Improving objectivity :
  - Well defined criteria for assessment, validated measures
  - Training of assessors
  - Independent (blinded) assessment

Analysis

- Statistical methods in the protocol
- Detailed statistical analysis plan : patients populations, methods, contents including subgroups analyses, interim analyses –early efficacy, futility or both-, hierarchical testing if applicable, adjustment for multiplicity
- ITT principle : analysis of all randomized patients (preserving randomization)
  - pragmatic trials versus explanatory trials
  - exception for non inferiority trials
- Primary analyses versus sensitivity analyses
- **Bias control**

Cochrane risk of bias tool (2019)

5 domains to be assessed

1) Risk of bias arising from the randomization process :
   - Random allocation / concealment / baseline differences
   - **Y / PY** / **PN / N** / NI -> low / high risk of bias / some concerns
   - Assessment of the bias direction
2) Risk of bias arising from deviations to the interventions
   - Blinding / impact of deviations and of patients exclusion on outcomes
3) Risk of bias arising from missing outcome data
   - Amount of missingness and potential for bias / missingness at random or not (association with outome)
4) Risk of bias in measurement of the outcome
   - Adequacy of method, association with arm, blinding of assessors
5) Risk of bias arising from selection in the reported results
   - Consistency with protocol, multiple analyses of data, subgroups analyses only, …

Reporting and interpreting results

Consort statement (2010)

Study flow chart

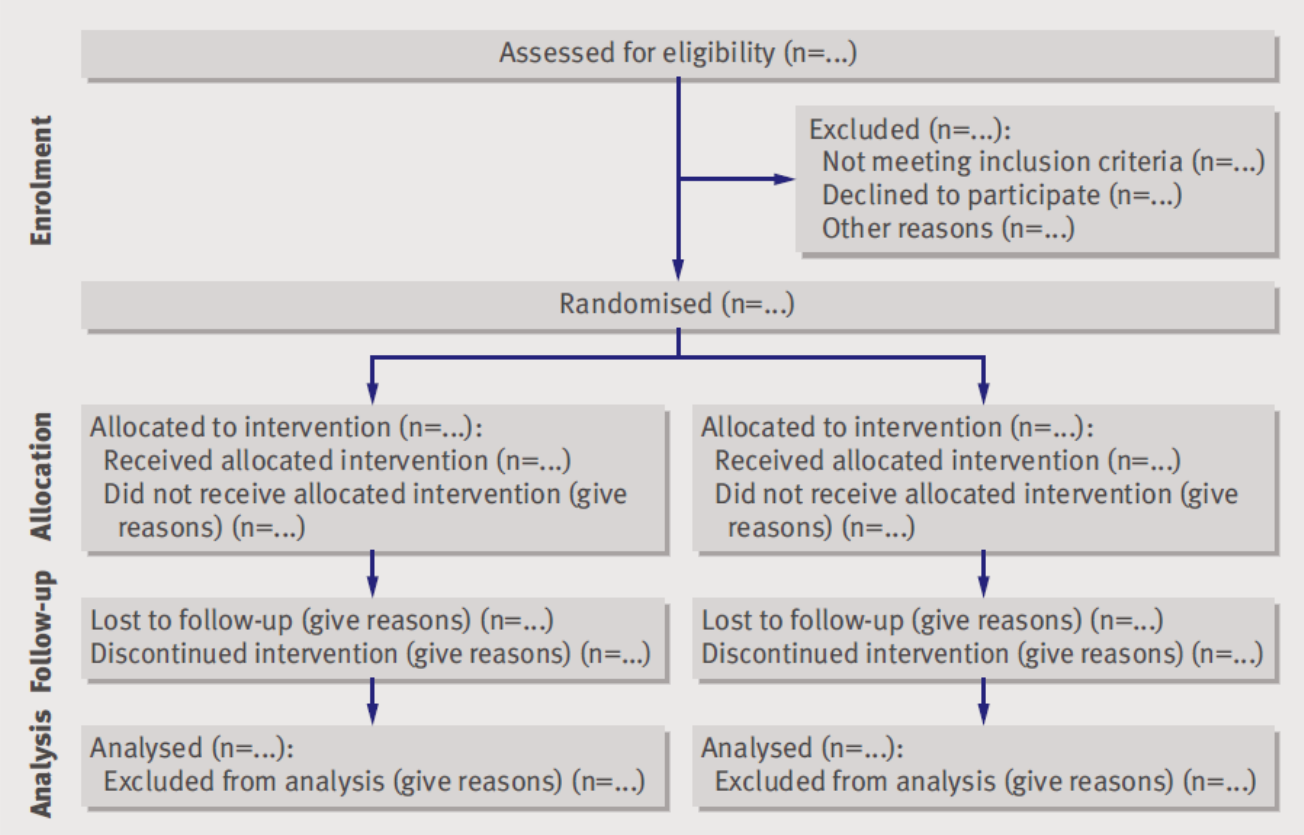


Fig 1 | Flow diagram of the progress through the phases of a parallel randomised trial of two groups (that is, enrolment, intervention allocation, follow-up, and data analysis)[52-54]

INSTITUT JULES BORDET INSTITUUT

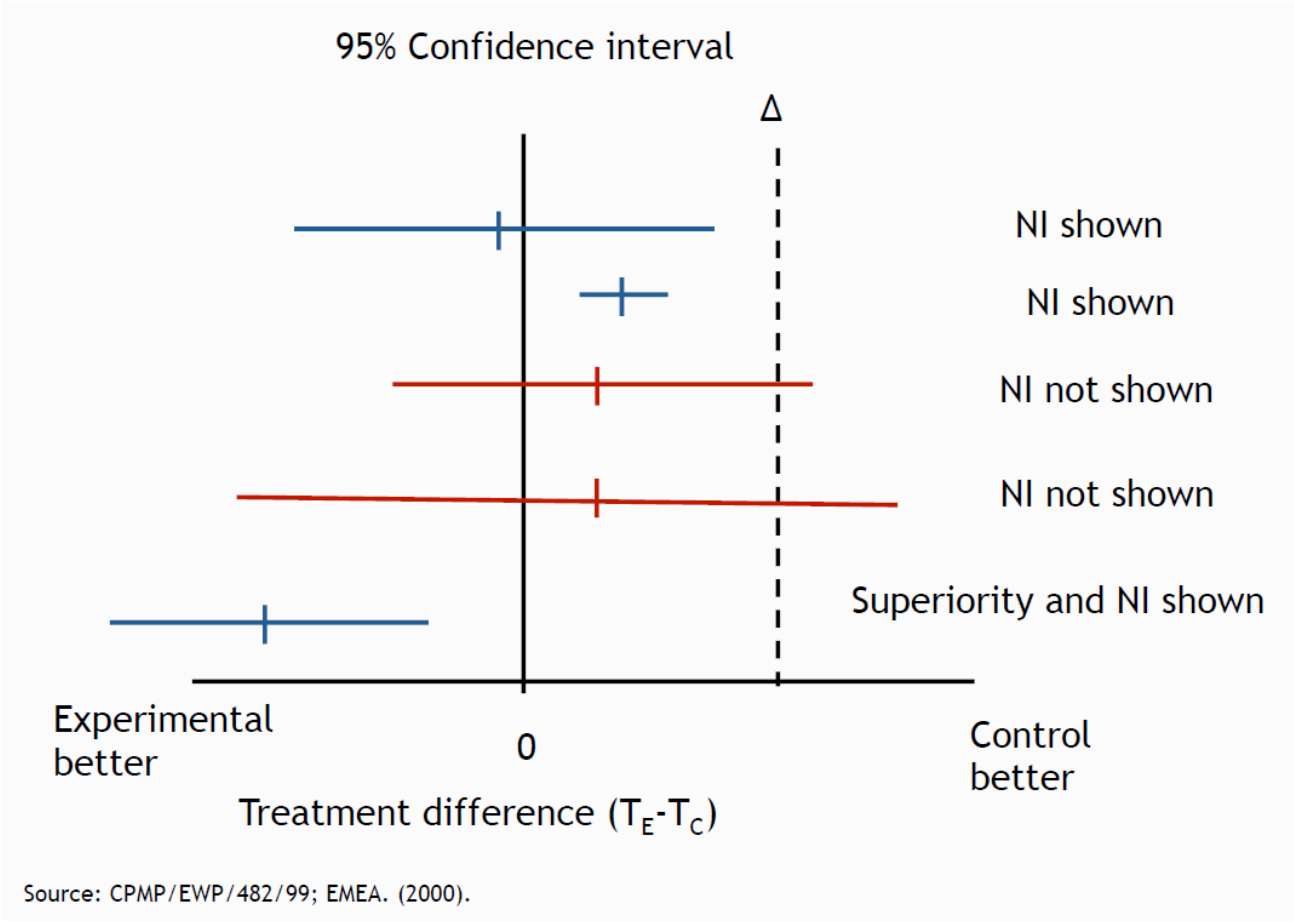
ULB iris

Reporting and interpreting results

- Patients flow
- Accrual period and follow-up period
- Reason for stopping the trial
- Baseline data
- Outcomes : estimation per group and confidence intervals for intervention effect (all outcomes); both relative and absolute effects
- Other planned and unplanned analyses

Confidence intervals and p values

P value : probability of rejecting null while null is true
Confidence interval : contains the true treatment effect with high confidence
P value : combination of magnitude of effect and sample size
Confidence interval : allows to interpret the magnitude of effect



Source: CPMP/EWP/482/99; EMEA. (2000).

INSTITUT JULES BORDET INSTITUUT

ULB iris

Subgroups analyses

- Planned versus unplanned or posthoc (hypothesis behind the analysis versus fishing expedition or data driven analysis)

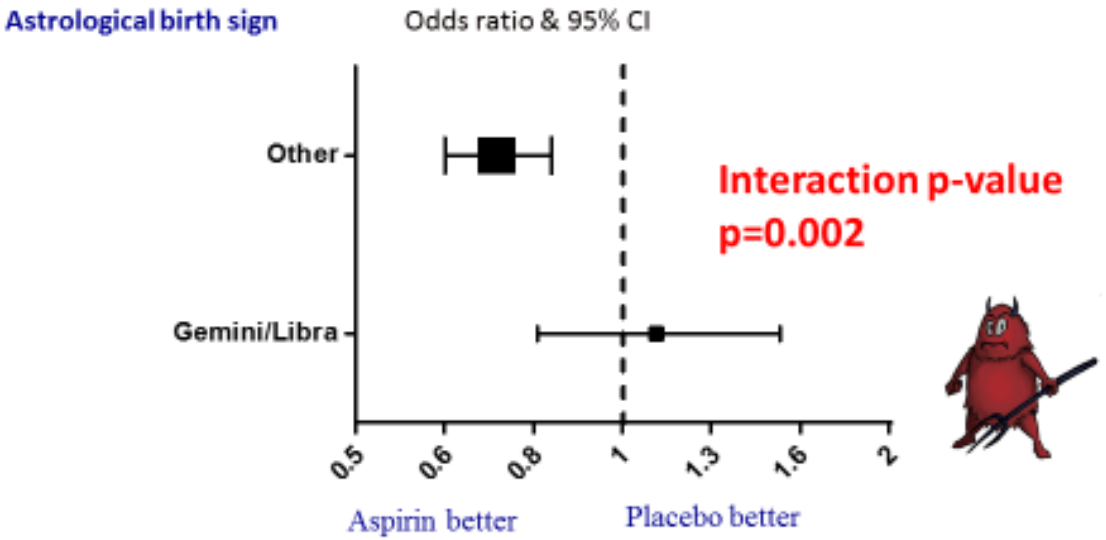
- Multiplicity : 10 covariables -> 10 subgroups analyses -> false positive result in 40% of trials ; difficult to assess cf reporting bias

What you can get

MOST FAMOUS SUBGROUP?

Empruntée à S. Michiels

ISIS-2: aspirin vs control - effects on vascular death in 17,187 patients with acute myocardial infarction (Peto et al, Lancet 1988)



Likely, an interaction test on the 12 signs would not have been significant

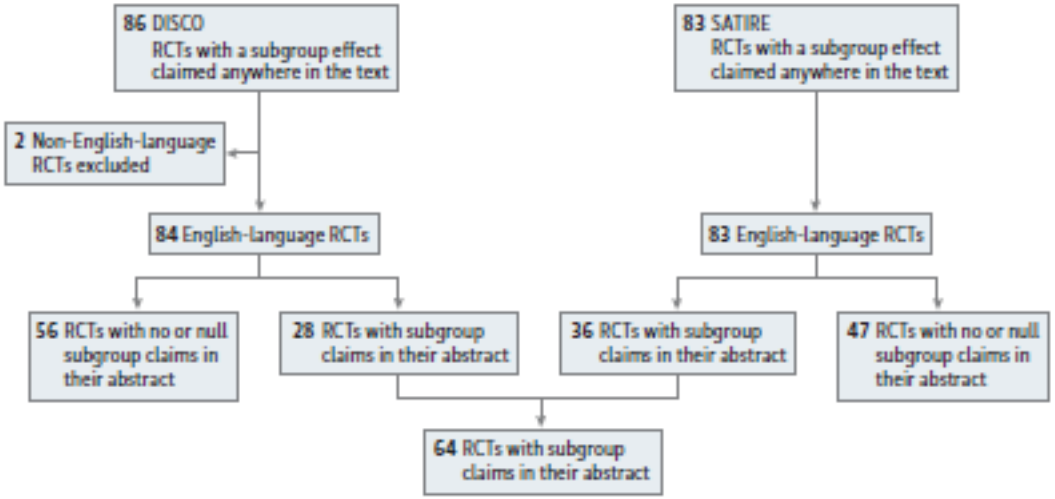
This is not only theory …

Evaluation of Evidence of Statistical Support and Corroboration of Subgroup Claims in Randomized Clinical Trials

Joshua D. Wallach, MS, PhD; Patrick G. Sullivan, MD, MS; John F. Trepanowski, PhD; Kristin L. Sainani, MS, PhD; Ewout W. Steyerberg, MSc, PhD; John P. A. Ioannidis, MD, DSc

JAMA 2017

Figure. Flow Diagram of DISCO and SATIRE Articles Review Process



117 subgroups analyses

39% considered statistically valid