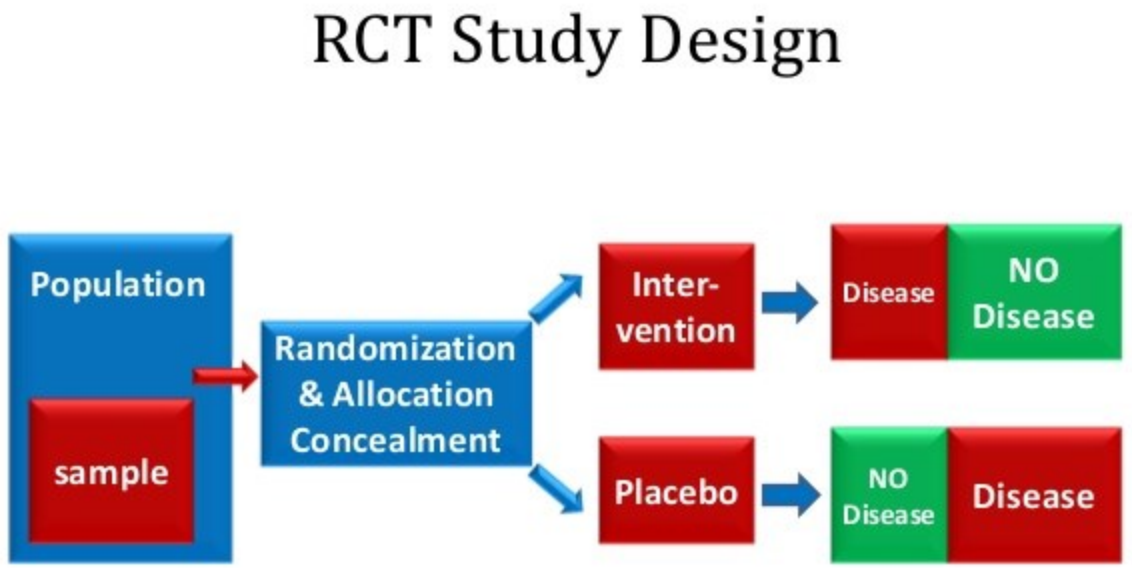
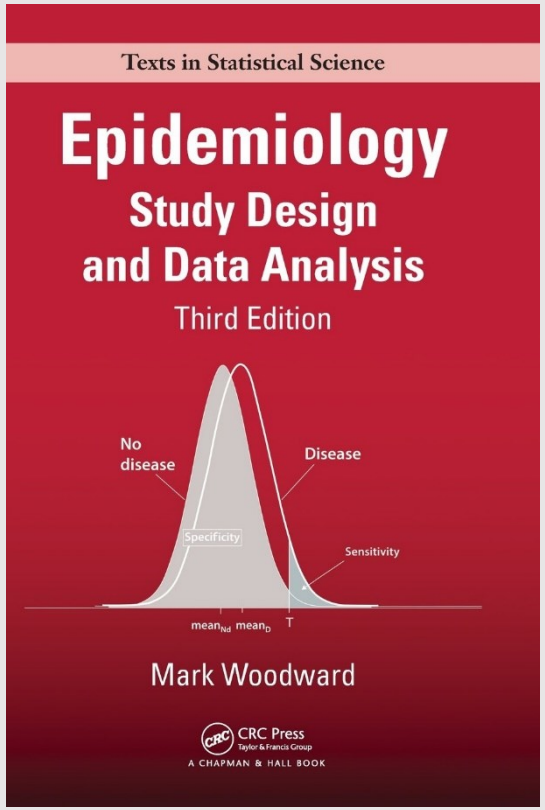
Study Designs in Epidemiology

- General Introduction
- Randomized Clinical Trials : RCTs



Science is built of facts the way a house is built of bricks: but an accumulation of facts is no more science than a pile of bricks is a house.

Henry Poincaré
(1854 – 1912)



To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of.

Ronald A. Fisher
(1890-1962)

5-stage representation of the statistical method applied to scientific investigation:

Problem -> **Plan** -> Data -> Analysis -> Conclusion

- **Which** groups do you want to describe or compare ?
- **What** parameters do you want to measure ?
- **When** do you want to measure them ?
- What parameters do you need **to "account for"** to correctly evaluate the effect of interest ?
- ...

4 pillars of the research project *:

1. Relevance & Rationale:

Research questions should be **highly topical** and **meaningful** from a clinical, policy, or research methodology perspective.

[not only at time of study conception but, perhaps more importantly, at the anticipated time of submission for publication or presentation to the relevant audience].

2.Specificity:

Research questions should be **concise yet unambiguous**, should clearly state:

- **intervention** and **outcome of interest**
- should identify the **patient population**
- should focus on **one primary end point**
- existing data sources must be **adequate** to provide valid answers

*Berger ML, Mamdani M, Atkins D, Johnson ML. Good Research Practices for Comparative Effectiveness Research, Value Health 2009

3. Novelty:

Proposals should clearly identify **what a new study can add to existing knowledge**:

- **absence** of literature thereby making the proposed research question **novel**.
- previous findings may have been **inconclusive, conflicting or questioned** because of study limitations.
- even when some research exists, **there may be a need to validate** findings

4. Feasibility

The proposal should be **feasible** with respect to:

- **power of the study** to answer a question
- **time and resources**, ability to link necessary data sources
- **adequate numbers of patients** and **events** to yield sufficient power for the analysis
- **timing**: answers may no longer be relevant if it takes years to collect and analyze data

RESEARCH OBJECTIVES

Describe - associations/patterns.
Moderately simple .
Report of the descriptive statistical analyses
("**IDA**": initial/exploratory data analysis).

Inference - generalize from sample to population
Moderately difficult .
Requires an analysis plan before starting.

Predict/counterfactual - what *will happen* to the subjects (*if*…).
Descriptive + Inference + (causal) Pre / post report analysis plan.

STATISTICAL TOOLS

Descriptive Statistics
[+***unsupervised*** *approaches*]

Inferential Statistics
(univariable/stratified methods)

Regression Models*
(linear/generalized/survival)
Explanatory/Prognostic/Predictive

Causal effects *Absolute risk* *Individualized risk*

*__supervised__ machine learning techniques

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

TERMINOLOGY:

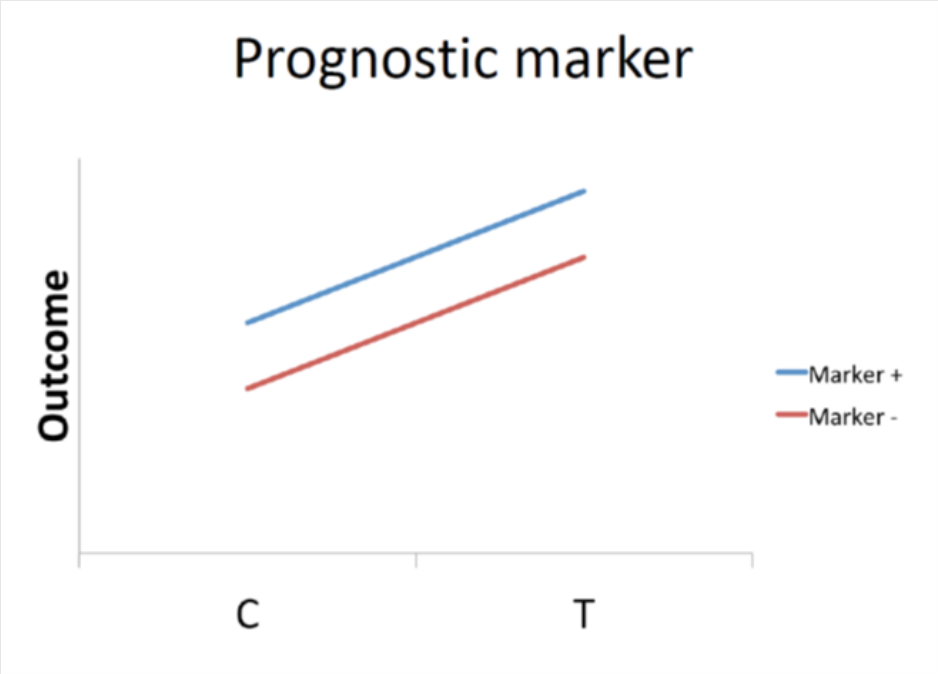
Outcome: result of clinical interest

- time to an event (death / recurrence)
- parameter of physiological functionality
- Binary factor (yes vs no)death/alive

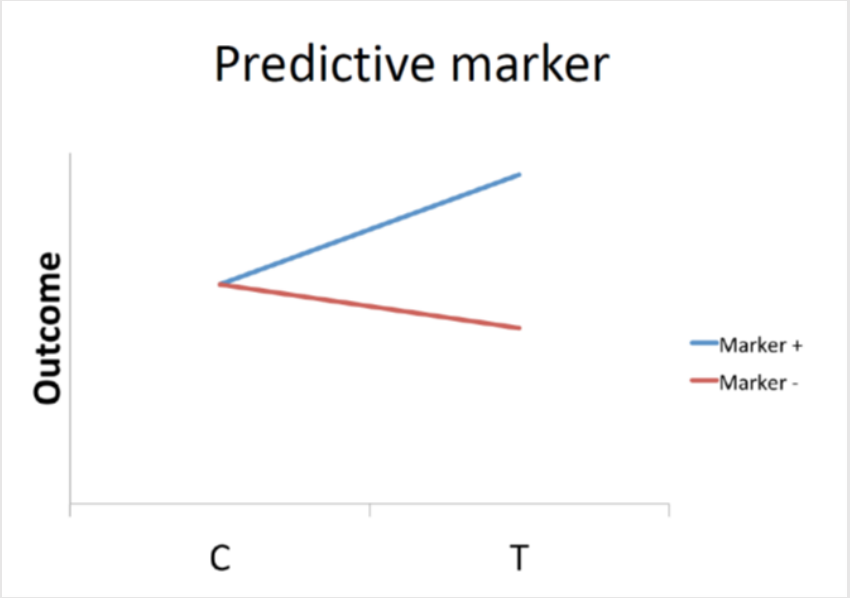
Predictor (s): factors associated with outcome

- genotype (s)
- Biomarker (continuous scale)
- Drug (yes vs no, dosages…)
- Type of chirurgical intervention…(nominal scale)

Prognostic factor is associated with clinical outcome in the absence of therapy or with the application of a standard therapy that patients are likely to receive. It can be thought of as a measure of the *natural history* of the disease.



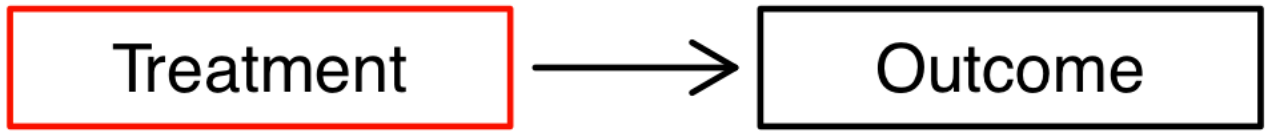
Predictive factor is associated with response or lack of response to a particular therapy.



A predictive factor implies a **differential benefit** from the therapy that depends on the status of the predictive biomarker.
In statistical terms, this constitutes an **interaction (*effect modifier*)** between treatment benefit and biomarker status

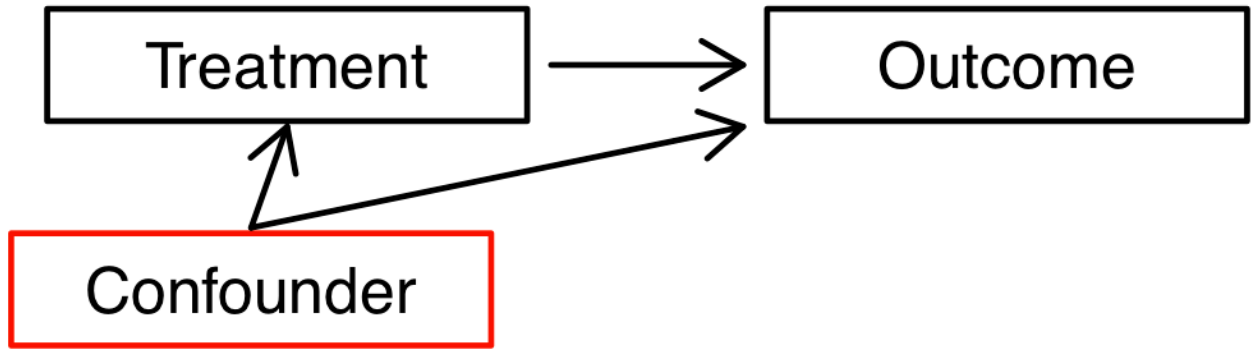
UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

(1)



Increasing level of "complexity" of the study design from (1) to (3)*

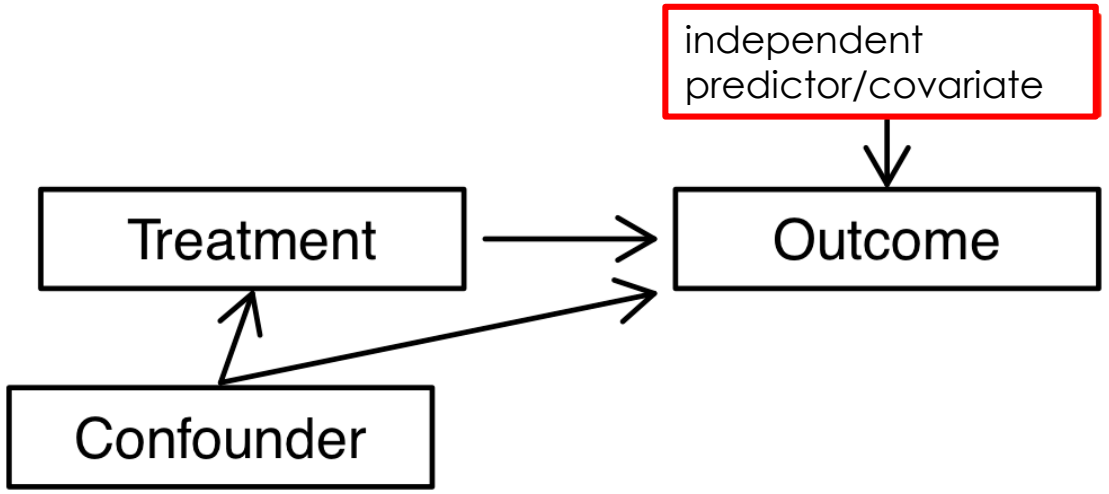
(2)



Associated with outcome (but not with the treatment)
ex: age / gender ...

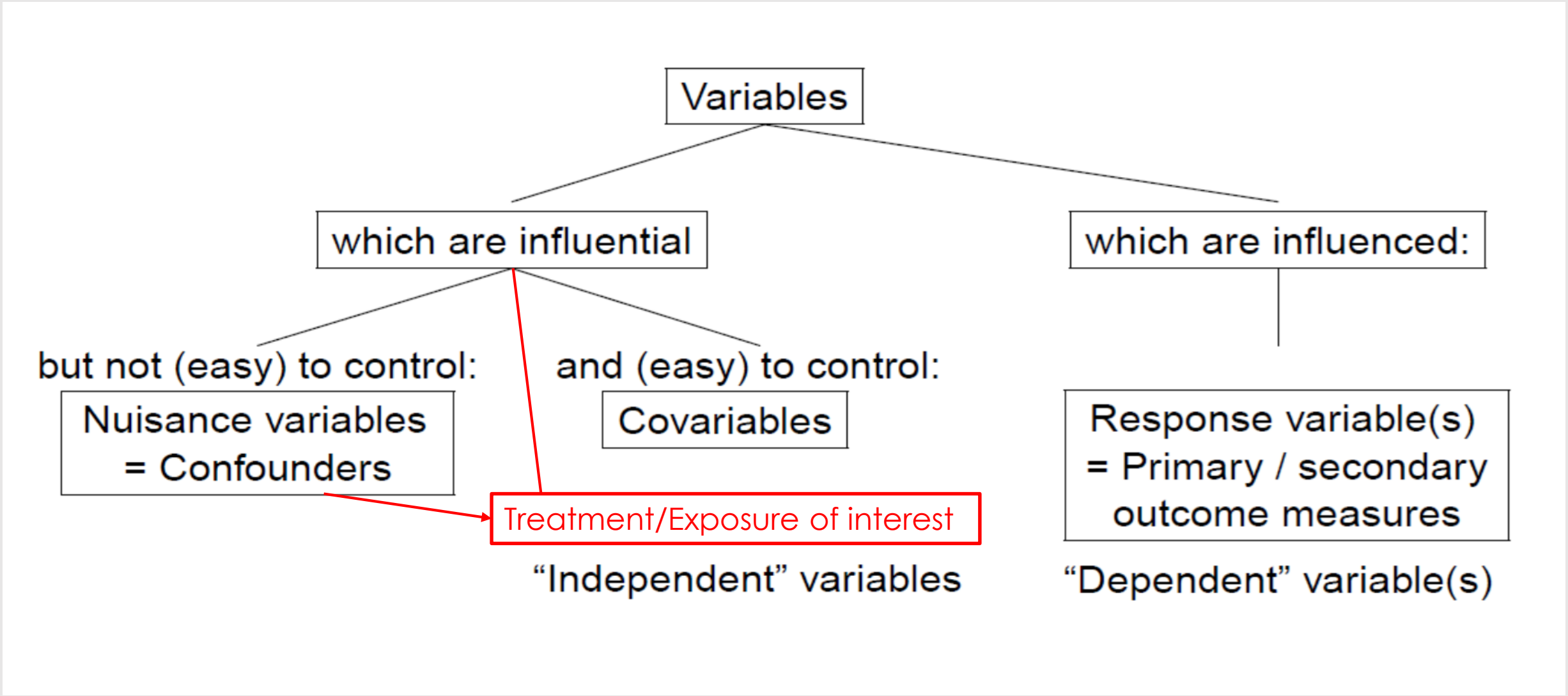
* There are also **mediators/colliders** ... something more in Block 3

(3)



associated with the outcome **and** the treatment:

- clinical indications for the treatment
- Alchool intake w.r.t. exposition to smoke, on oral cancers


Variables
which are influential
which are influenced:
but not (easy) to control:
and (easy) to control:
Nuisance variables
= Confounders
Covariables
Response variable(s)
= Primary / secondary
outcome measures
Treatment/Exposure of interest
“Independent” variables
“Dependent” variable(s)

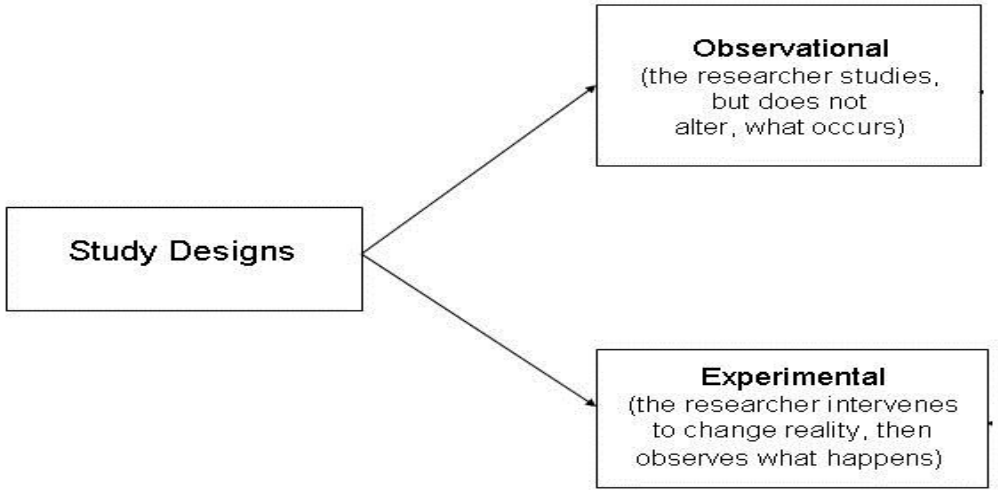
Relevance & Rationale : systematic literature review

- **Background** on the state of art

- **Efficacy/Effectiveness** of the treatment / drug / intervention from previous studies

  (help for quantitative estimates of its effect [sample size])

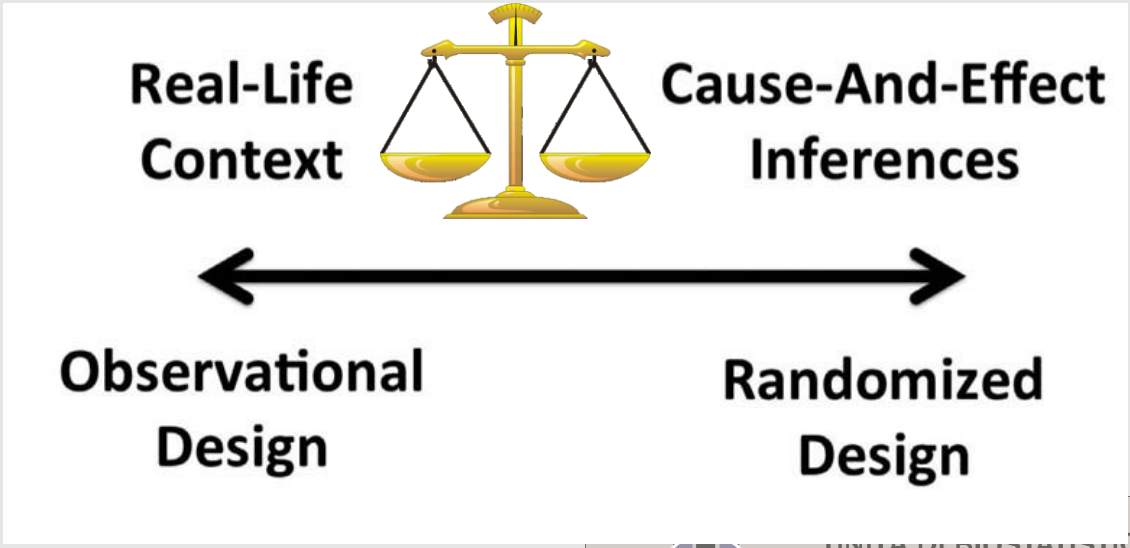
What "type" of study : experiment or observation?



The principle of science, the definition, almost, is the following: ***the test of all knowledge is experiment***. Experiment is the *sole judge* of scientific 'truth'.
(R. Feynman,1963)

"Experiment, observation […] have a **crucial role** for modern therapeutics. ***Arguments about the relative importance of each are an unnecessary distraction***."

Sir Michael Rawlins, National Institute for Health and Clinical Effectiveness (NICE) ***Lancet* 2008; 372: 2152–61**

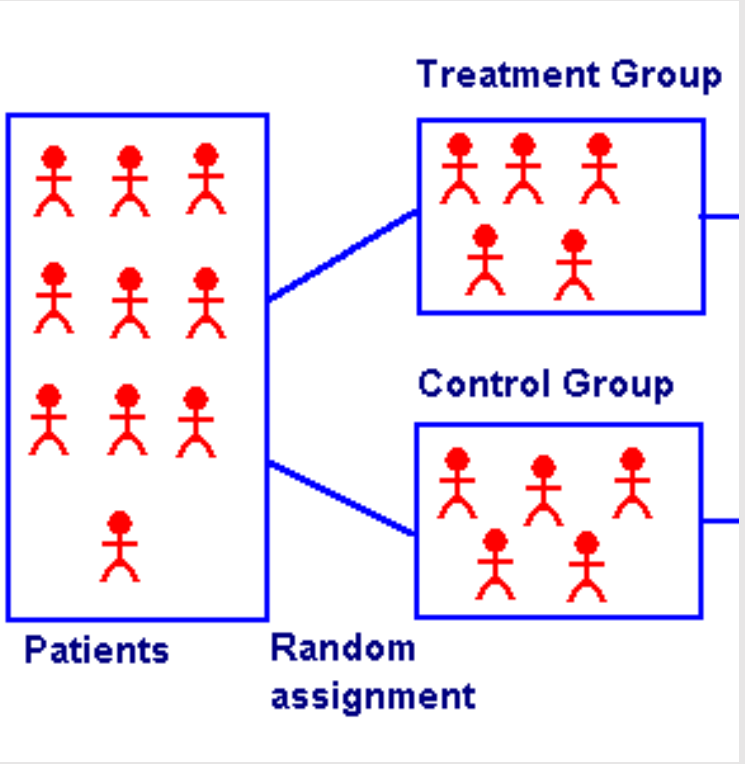


Randomized Clinical Trials

RCTs are the **gold standard** in clinical research to compare interventions in humans. The selected population is usually homogeneous with respect to inclusion / exclusion criteria.

To implement an RCT, a number of rules must be respected:

- scientific- methodological
- statistical
- ethical
- legal



The basic **ethical** principle is that of the **clinical equipoise**: uncertainty in the scientific community on the comparative efficacy of the treatments studied in the trial.

RCTs designed to observe outcomes in humans,
under certain experimental conditions created **ad hoc** by the researcher having **randomized*** subjects to the different treatments.

...*note that the general RCT starting sample is not really *random* (criteria...volunteer bias)

Potential Outcomes Framework

(Rubin-Neyman Causal Model)

Each unit (individual) has **two** potential outcomes:

$Y_0(i)$ is the potential outcome had the unit *i* **not** been treated: **control** outcome

$Y_1(i)$ is the potential outcome had the unit *i* been treated: **treated** outcome

Individual treatment effect for subject *i*:

$$ITE_i = Y_1(i) - Y_0(i)$$

Average Treatment Effect*:

$$ATE = E[Y_1 - Y_0] = E[ITE_i]$$

Average Treatment Effect in the Treated*:

$$ATT = E[Y_1 - Y_0 | T = 1]$$

*RCTs are the **ideal** study design in which estimate these quantities

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Potential Outcomes Framework

(Rubin-Neyman Causal Model)

$$T_i = \begin{cases} 0 & \text{Untreated} \\ 1 & \text{Treated} \end{cases}$$

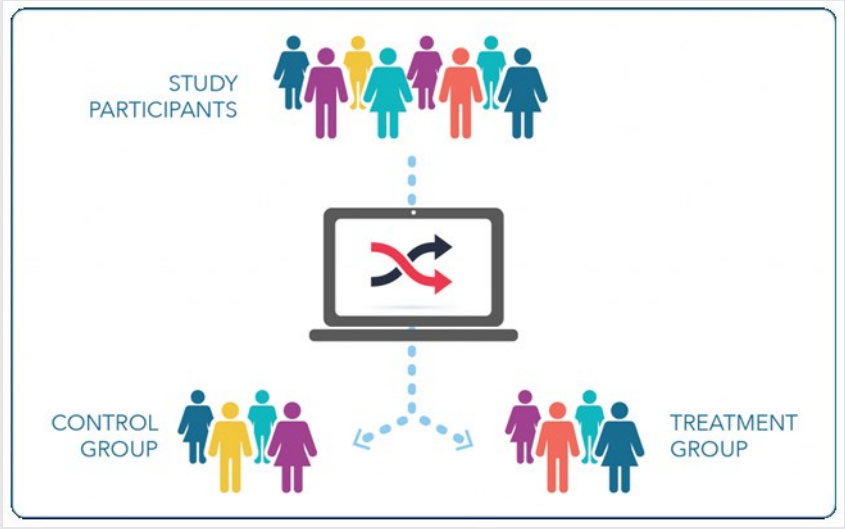
Observed **factual** outcome:

$$y_i = t_i Y_1(i) + (1 - t_i) Y_0(i)$$

Unobserved **counterfactual** outcome:

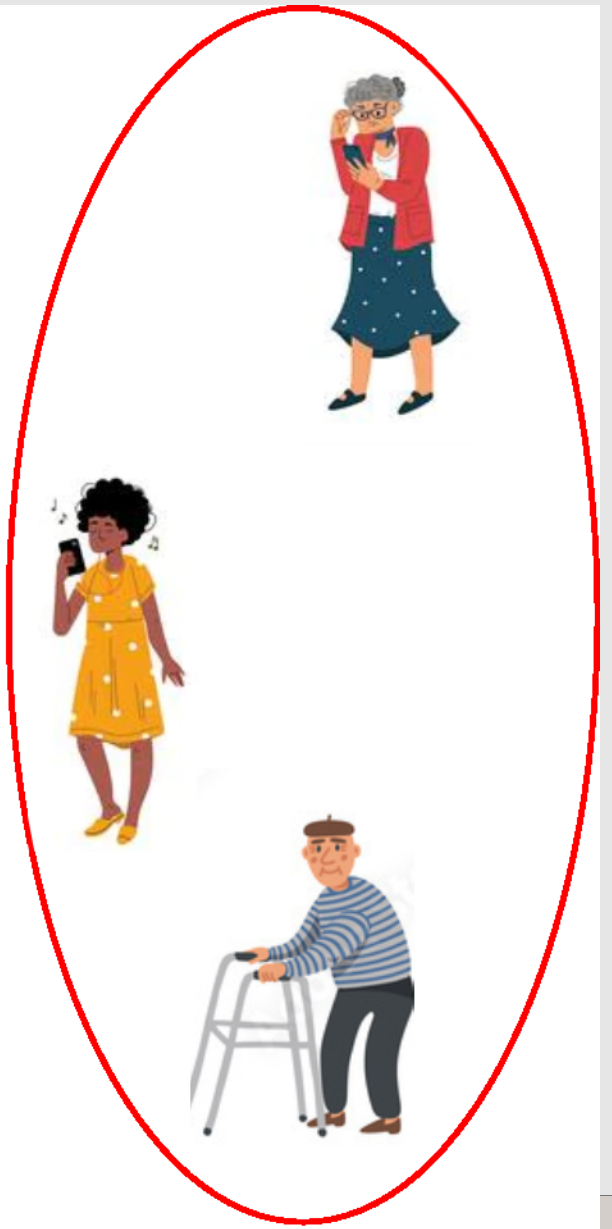
$$y_i^{CF} = (1 - t_i) Y_1(i) + t_i Y_0(i)$$

one can never **directly** observe **individual** causal effects, because we can never observe **both** potential outcomes for any subject. We need to compare *potential* outcomes, but we only have *observed* outcomes.



Control arm

Experimental arm



Randomization is **fair** with respect to the *potential outcomes* that (on average) **are independent** from the assignment to the treatment (**unbiased** estimator **by design**).

Moreover, in a randomized experiment, the treatment and control group tend to be **similar** in terms of their *observed* and *unobserved* covariates (**exchangeability**).

Randomization guarantees: $Y_1, Y_0 \perp T$ + **exchangeability**

$$E(Y_1|T=1) = E(Y|T=1)$$

$$E(Y_0|T=0) = E(Y|T=0)$$

Distributions (and means) of potential outcomes are **independent** from treatment ***assignment***

Y=observed outcome

$$ATE = E[Y_1 - Y_0]$$

We can identify this causal estimand from the *observed* RCT data

Linearity of Expectations

Never observed

$$E(Y_1 - Y_0|T=1) = E(Y_1|T=1) - E(Y_0|T=1)$$

$$= E(Y_1|T=1) - E(Y_0|T=0)$$ → T and potential outcomes are *independent*

$$= E(Y|T=1) - E(Y|T=0)$$ → *exchangeability*: we can use the *observed* Y

$$E(Y_1 - Y_0|T=1) = E(Y_1 - Y_0)$$ → subjects randomized to the treatment are *representative* of the entire population (ATT=ATE)

RCT at the heart of the implementation of new treatments:

PHASE I: Safety of a pharmacological principle / treatment. Small samples (20-80 pts). MTD = Maximum Tolerated Dose on healthy / sick volunteers; informed consent is crucial.

PHASE II: Therapeutic exploration : sample of volunteers with the disease of interest (100-300 pts). Pharmacokinetic / dynamic assessments, optimal dose, frequency of doses, administration protocols, evaluation of end-points of interest.

PHASE III: Comparative efficacy demonstrate / confirm the efficacy of a treatment and identify any adverse effects in clinical practice (large sample: 500-3000+ pts).

PHASE IV: Post-marketing surveillance: **observational** studies to identify less frequent adverse reactions (longer period); cost-effectiveness in the "real world" population.

about 20% of new drugs are "integrated" by side effects in phase IV studies; about 4% are withdrawn

Lasser KE et al., Timing of new black box warnings and withdrawals for prescription medications. JAMA, 2002.

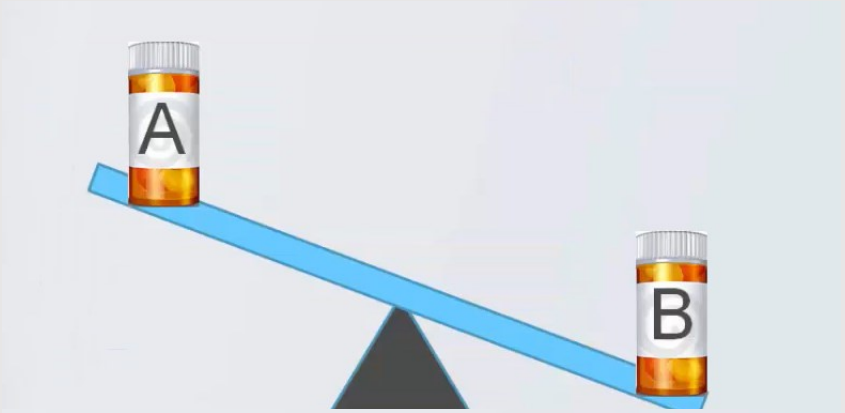
RCT «Hypothesis Types» (Phase III)*

3 general designs of RCTs:

*basic concepts about statistical tests…?

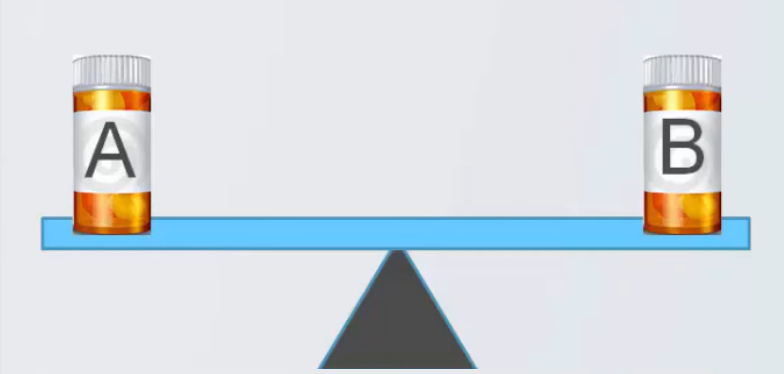
Superiority

$>$



Equivalence

$=$



Non-inferiority

$\not<$

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

The primary objective is to determine the magnitude of increased benefit of the experimental intervention over standard therapy for effectiveness outcomes

Superiority → To show that the treatment is more effective compared to control group

→ Used for establishing new standard of care

- H_0: Treatment *x* is NOT more effective than treatment *y* for given condition
- H_1: Treatment *x* is more effective than treatment *y* for given condition

$$H_0: T_X - T_Y \leq \delta$$

$$H_1: T_X - T_Y > \delta$$

δ is the **minimal clinically relevant «effect size»**

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

2

Equivalence → To show that the treatment is not more effective and is not worse compared to control group

→ Used for establishing generic drugs that are equivalent to their brand formulations

- H_0: Treatment *x* is either worse or better than treatment *y* for given condition by greater than Δ
- H_1: Treatment *x* is NEITHER worse NOR better than treatment *y* for given condition by greater than ±Δ, when Δ is the equivalency margin

$$H_0: |T_X - T_Y| \geq \delta$$

$$H_1: |T_X - T_Y| < \delta$$

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

The researchers are unconcerned if the experimental treatment is better as long as it is not much worse.

Non-inferiority

- To show that the treatment is not worse compared to control group
- Used for establishing alternative care

- H_0: Treatment *x* is <u>worse</u> than treatment *y* for given condition by <u>greater than Δ</u>
- H_1: Treatment *x* is <u>not worse</u> than treatment *y* for given condition by greater than Δ, when Δ is the non-inferiority margin

$$H_0: T_X - T_Y \leq -\delta$$

$$H_1: T_X - T_Y > -\delta$$

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

The magic of RANDOMIZATION

eliminate systematic confounders , **differences** between groups , allows causal estimates from observed data

- **Simple**: randomly allocates each subject to a treatment independent of previous assignments (head or cross): simple but can create some *imbalances*
- **Block**: force the number of subjects to be allocated to groups of ***k*** in each arm

  blocks of size 4 (with two arms) means that in a sequence of 4 subjects 2 subjects are assigned in one arm and 2 in the other [alternate random, in all possible orderings]
  variable size blocks can be used (4, then 8, then 6 ...)

- **Stratification** (**+ blocks**): stratifies the population with respect to specific covariates to be represented (gender, age ...) and then block randomization from each stratum.

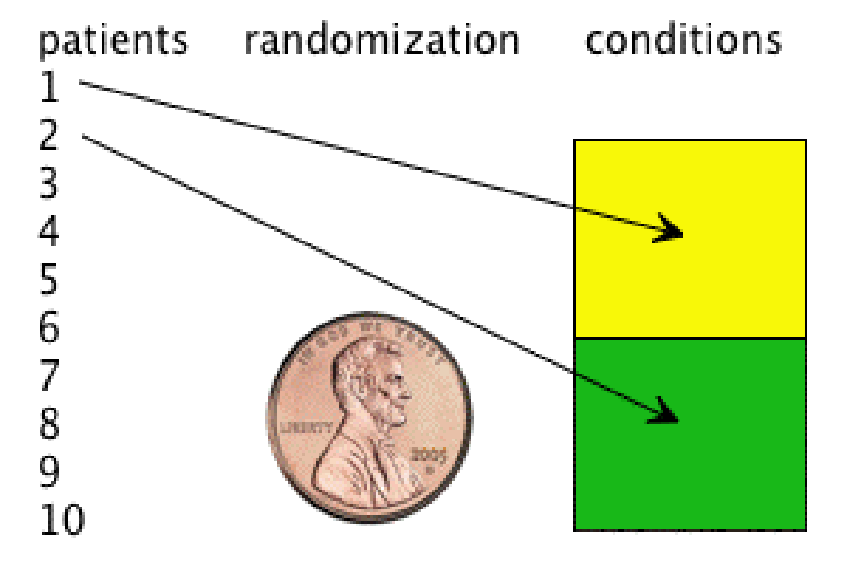
Adaptive randomization: chances of inclusion during the trial **change**

Concealment : hiding of enrollment, those who recruit **do not have to know** the random sequence of assignments

Deciphering the allocation concealment scheme

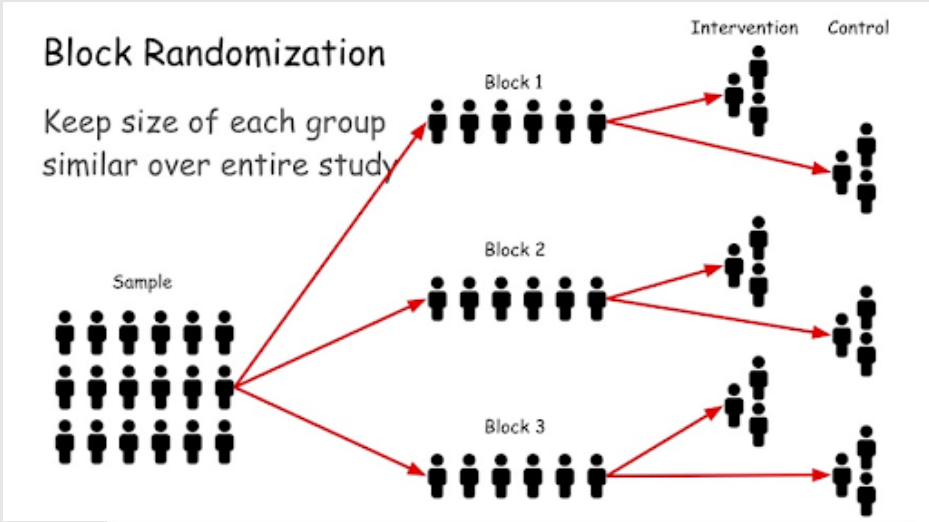
Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet. 2002.

Block 2.1

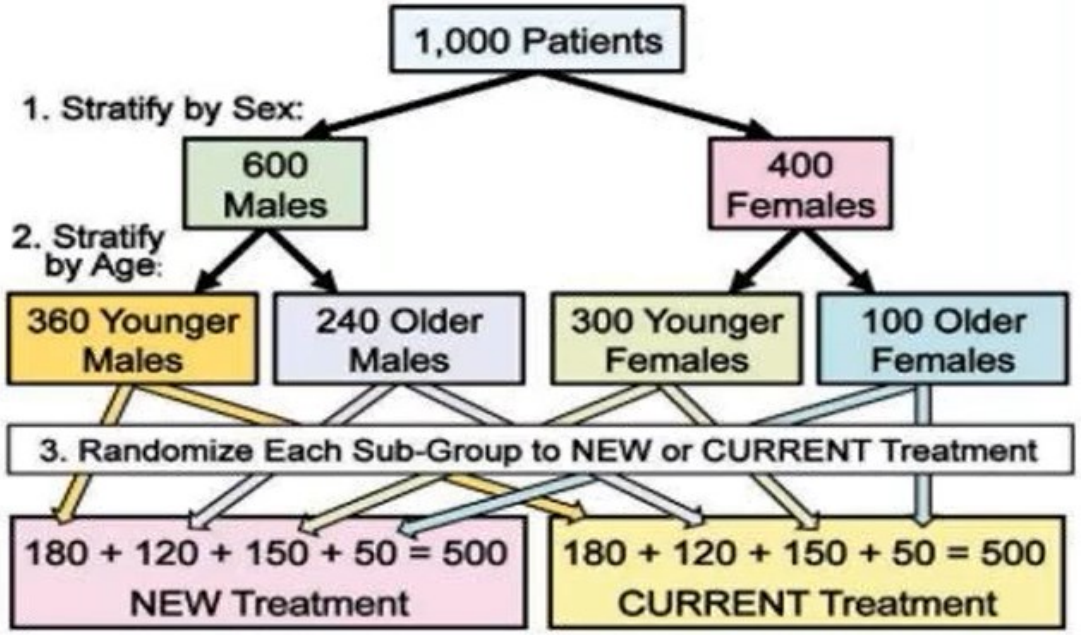


SIMPLE

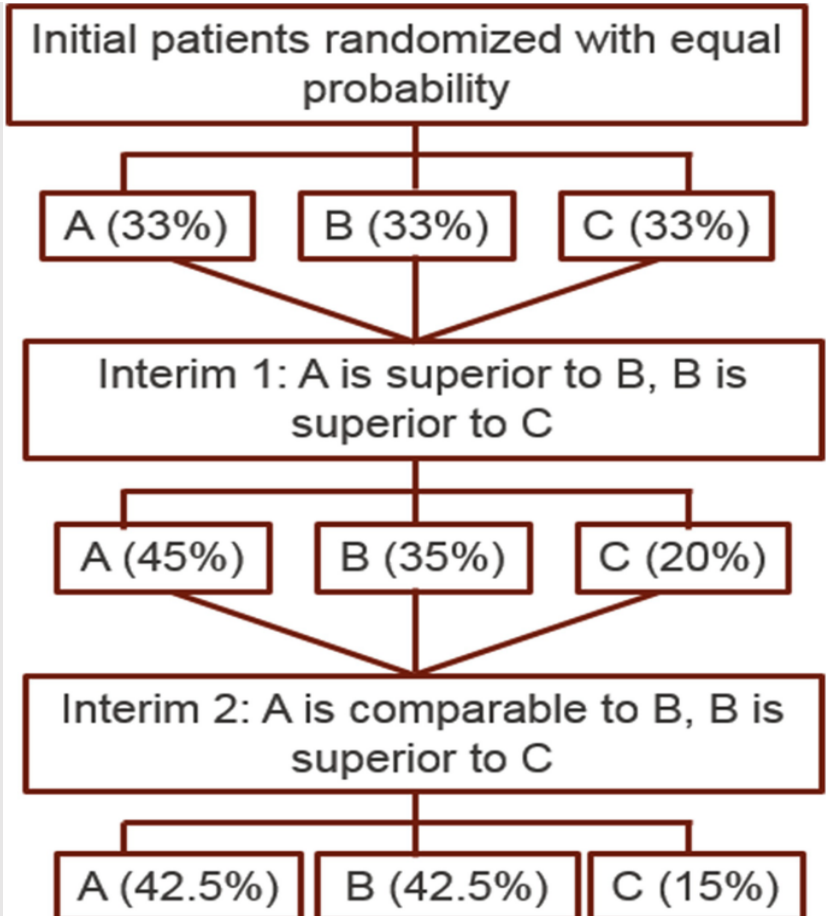
BLOCK



STRATIFIED



ADAPTIVE



Clinical Trials: BLINDING (MASKING)

To minimize the subjectivity of those who evaluate the pts or analyze the data.

SINGLE BLIND: study subjects do not know which treatment they receive
DOUBLE BLIND: subjects and researchers (doctors / biologists) don't know
TRIPLE BLIND: subjects, the researchers, the statisticians don't know

Figure 1: **The authors: double blinded versus single blinded**

Schulz & Grimes. Lancet 2002

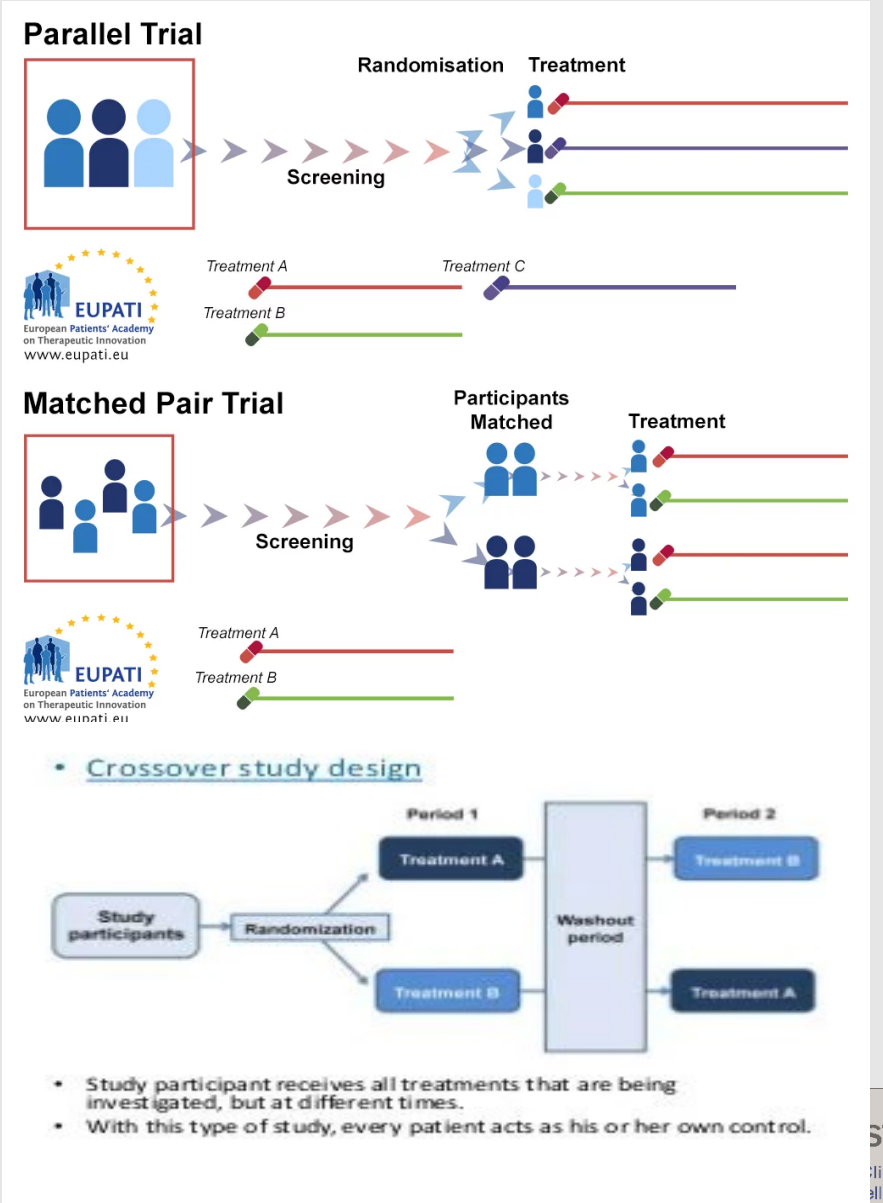
If not done: **OPEN TRIAL**

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

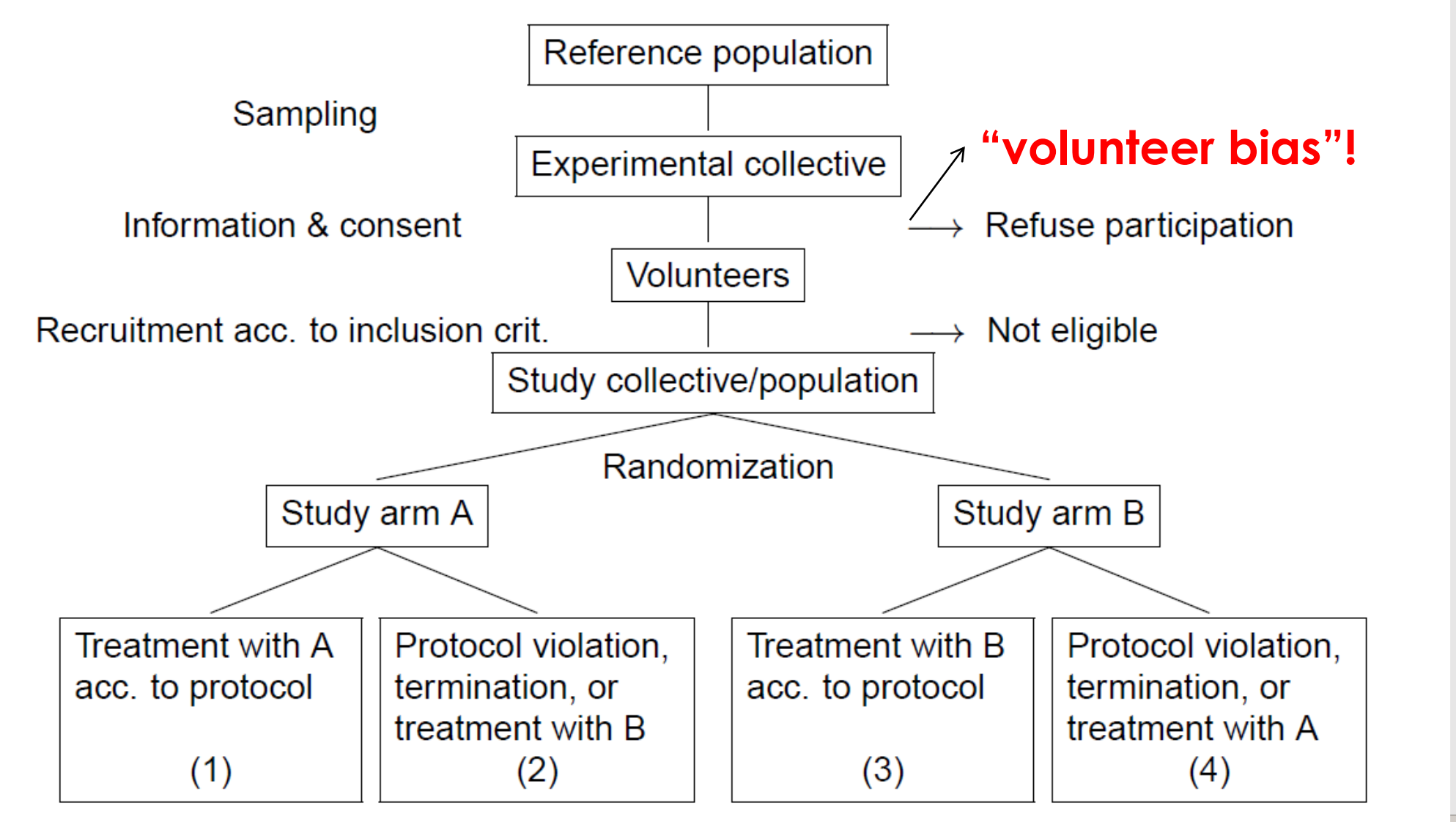
Types of randomised controlled trials

- Parallel
- Matched parallel
- Cross-over
- Sequential
- Cluster
- Umbrella & Basket…

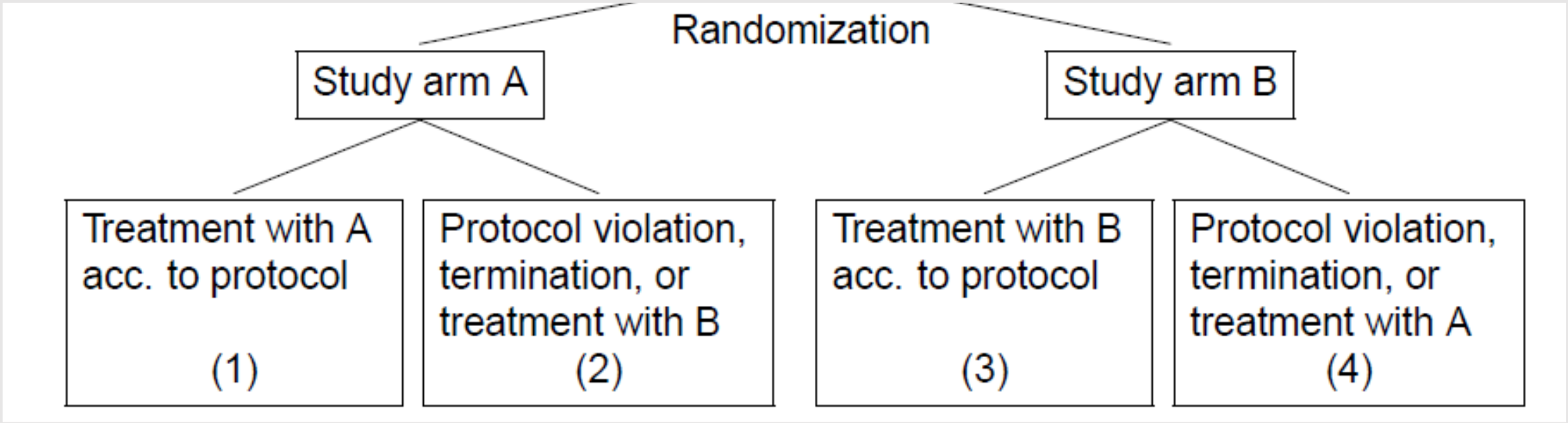
…this is not an exhaustive list!



RCT Flow Chart



RCT data analysis (I)



ITT: intention to treat
(1)+(2) vs (3)+(4)

PP: **per protocol** (1) vs (3)

AT: **as treated** (1)+(4) vs (2)+(3)
(but excluding protocol violations or early drop out)

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

RCT data analysis (II)

1. **Summary trial** results for clinical decision making:

ATE causal effects

Measure	Definition
Absolute risk difference	CER – TER
Relative risk reduction	1 – (TER/CER)
Odds ratio	[TER/(1 – TER)]/[CER/(1 – CER)]

CER = control event rate; TER = treatment event rate.

2. **Subgroup** analyses: divide the trial population into groups and examine effects

Potentially useful for exploring hypotheses about factors that **modify a treatment effect**, but these "*1-variable-at-a-time*" analyses have important limitations. Briefly: low statistical power, multiplicity, and weak prior theory on relative effect modifiers

CATE causal effects

Table 3. Vaccine Efficacy Overall and by Subgroup in Participants without Evidence of Infection before 7 Days after Dose 2.

Efficacy End-Point Subgroup	BNT162b2 (N=18,198)		Placebo (N=18,325)		Vaccine Efficacy, % (95% CI)†
	No. of Cases	Surveillance Time (No. at Risk)*	No. of Cases	Surveillance Time (No. at Risk)*	
Overall	8	2.214 (17,411)	162	2.222 (17,511)	95.0 (90.0–97.9)
Age group					
16 to 55 yr	5	1.234 (9,897)	114	1.239 (9,955)	95.6 (89.4–98.6)
>55 yr	3	0.980 (7,500)	48	0.983 (7,543)	93.7 (80.6–98.8)
≥65 yr	1	0.508 (3,848)	19	0.511 (3,880)	94.7 (66.7–99.9)
≥75 yr	0	0.102 (774)	5	0.106 (785)	100.0 (–13.1–100.0)
Sex					
Male	3	1.124 (8,875)	81	1.108 (8,762)	96.4 (88.9–99.3)
Female	5	1.090 (8,536)	81	1.114 (8,749)	93.7 (84.7–98.0)
Race or ethnic group‡					
White	7	1.889 (14,504)	146	1.903 (14,670)	95.2 (89.8–98.1)
Black or African American	0	0.165 (1,502)	7	0.164 (1,486)	100.0 (31.2–100.0)
All others	1	0.160 (1,405)	9	0.155 (1,355)	89.3 (22.6–99.8)
Hispanic or Latinx	3	0.605 (4,764)	53	0.600 (4,746)	94.4 (82.7–98.9)
Non-Hispanic, non-Latinx	5	1.596 (12,548)	109	1.608 (12,661)	95.4 (88.9–98.5)
Country					
Argentina	1	0.351 (2,545)	35	0.346 (2,521)	97.2 (83.3–99.9)
Brazil	1	0.119 (1,129)	8	0.117 (1,121)	87.7 (8.1–99.7)
United States	6	1.732 (13,359)	119	1.747 (13,506)	94.9 (88.6–98.2)

Pfizer Trial

RCT data analysis (III)

3. **Predictive** approaches to **HTE** (Heterogeneity of Treatment Effect) analysis (***regression-based estimates***)

$$P(outcome) = f(\alpha + \beta_1 x_1 + \cdots + \beta_p x_p)$$

$$lp = \beta_1 x_1 + \cdots + \beta_p x_p$$

$$P(outcome) = f(\alpha + \boldsymbol{\beta_{tx}} \boldsymbol{tx} + \beta_{lp} lp + \boldsymbol{\delta_{int}} \boldsymbol{tx} * \boldsymbol{lp})$$

$$P(outcome) = f(\alpha + \boldsymbol{\beta_{tx}} \boldsymbol{tx} + \beta_1 x_1 + \cdots + \beta_p x_p)$$

Risk oriented

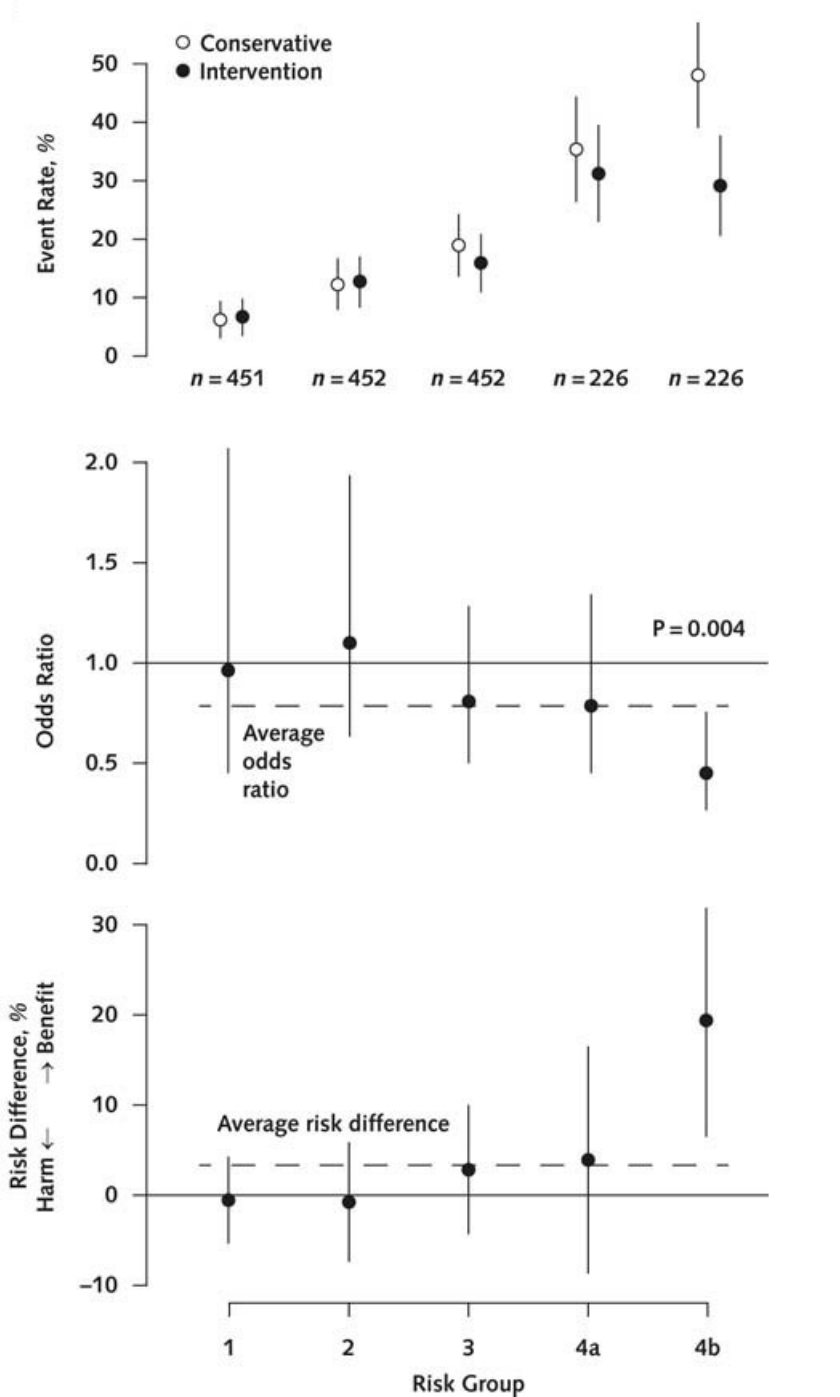
Effect oriented

$$P(outcome) = f(\alpha + \boldsymbol{\beta_{tx}} \boldsymbol{tx} + \beta_1 x_1 + \cdots + \beta_p x_p + \boldsymbol{\beta_{1int}} \boldsymbol{x_1 tx} + \cdots + \boldsymbol{\beta_{pint}} \boldsymbol{x_p tx})$$

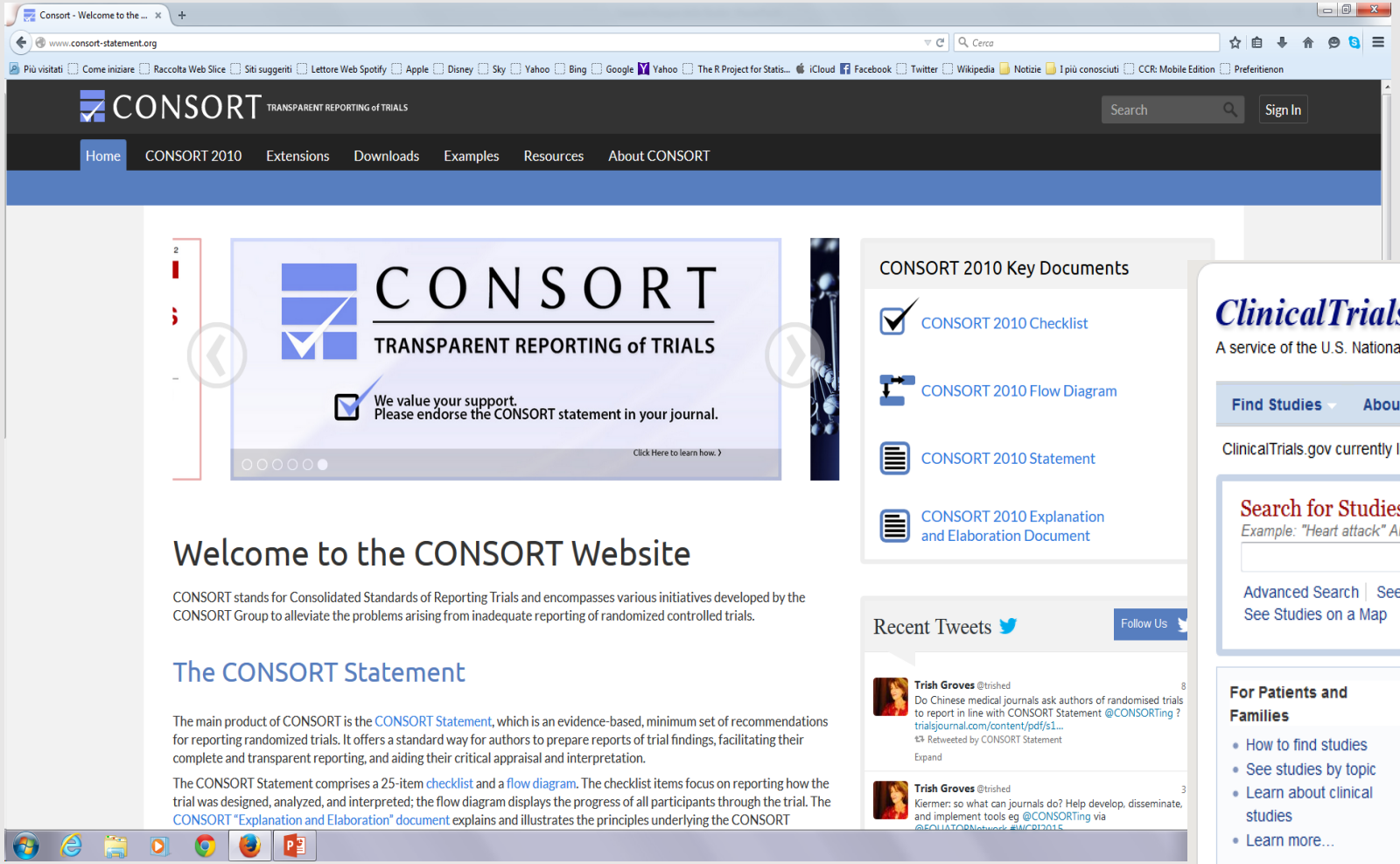
If trial population has substantial **variation in baseline outcome risk** or **interactions**

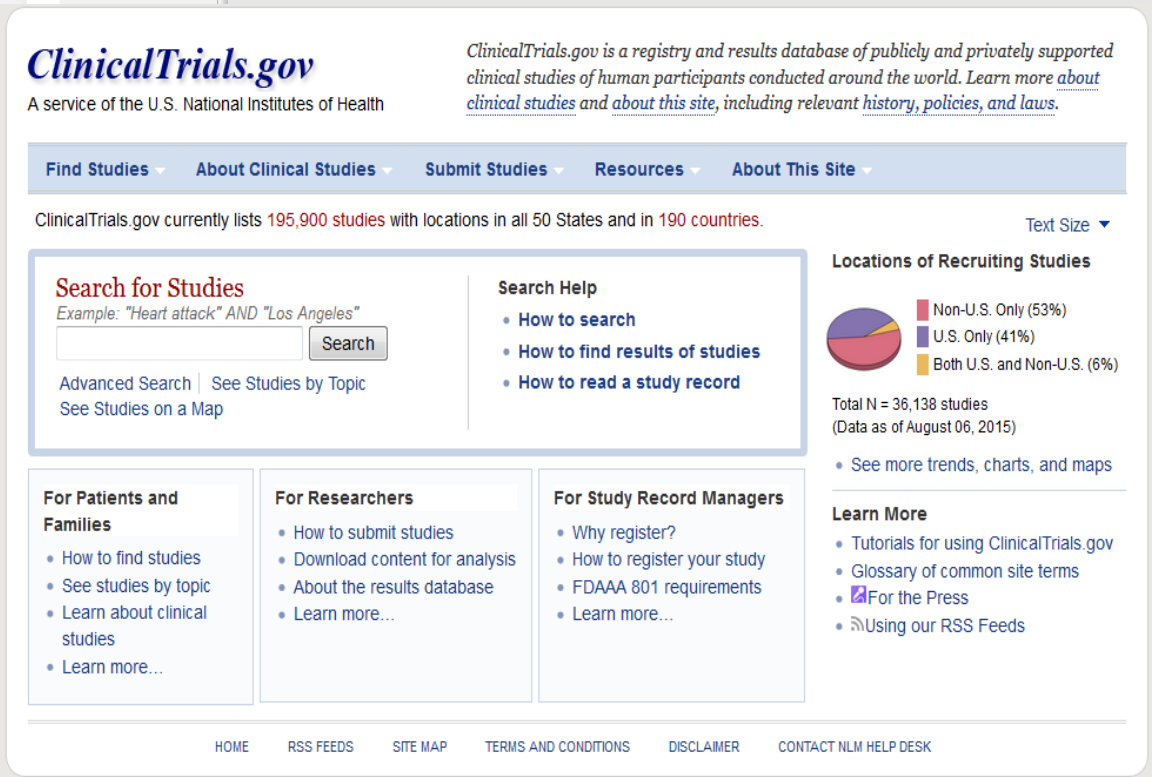
Patient-centered estimates of treatment effects: patients have many attributes that simultaneously affect the outcome of interest and the benefits of treatment.

https://www.acpjournals.org/doi/full/10.7326/M18-3667



Guidelines for carrying out an RCT: it also includes a **25**-point checklist to be filled in when reporting the results





RCTs: limitations (I)

- RCTs are performed in **selected populations** of pts usually for **short periods** of time.

- In clinical practice, intervention / treatment is generally applied in a **heterogeneous** population of pts - often with multiple comorbidities - and usually for **longer** periods.

- RCTs can describe the most common and early adverse reactions*, but they could be unable to identify those that are less common or have longer latency.

* 500-3000 subjects: 6/1000 - 1/1000 adverse events (95% CI)

	Potential problems
Patients	
Age	Effectiveness in younger or older patients
Sex	Effectiveness generally
Severity of the disease	Effectiveness in mild or severe forms of the condition
Risk factors	Effectiveness in patients with risk factors for the condition (eg, smokers)
Comorbidities	Influence of other conditions on effectiveness
Ethnicity	Effectiveness in other ethnic groups
Socioeconomic status	Effectiveness in disadvantaged patients
Treatment	
Dose	Too high a dose used in RCTs
Timing of administration	Influence on adherence (compliance) to treatment regimens
Duration of therapy	Effectiveness during long-term use
Comedication	Adverse interactions
Comparative effectiveness	Effectiveness in comparison with other products used for the same indication
Setting	
Quality of care	Prescription and monitoring by less specialist (expert) healthcare providers

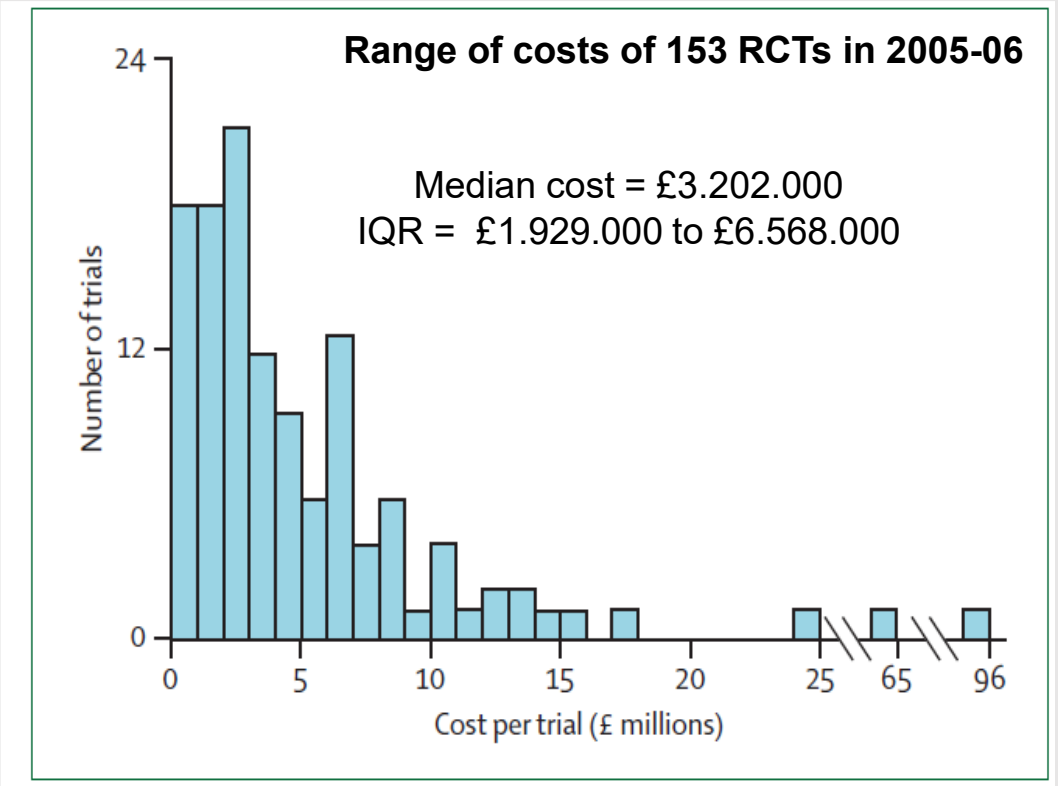
RCTs: limitations (II)

Califf RM. Clinical trials bureaucracy: unintended consequences of well-intentioned policy. *Clin Trials* 2006.

- RCT **costs** are high and show a growing trend

- **Volunteer bias**: study eligibility, compliance, geographical proximity, socio-economic status, health status ...

Homogeneous population selected:
- internal validity (= low variability)
- limit for external validation of results



EFFICACY ≠ EFFECTIVENESS

Success of a treatment in a "artificial" context

Success of a treatment in a "Real world" context

Whilst randomization remains a key design feature for **confirmatory trials aimed at new drug approval**, many innovative non-randomized study designs have been proposed and used in recent years.

The availability of **large databases** of **real world evidence (RWE)** has led to questioning the necessity to carry out expensive and lengthy randomized clinical trials, if **sophisticated statistical techniques** (e.g., **based on causal inference**) could inform the choice between therapeutic options…