Statistical Considerations In Building External Controls

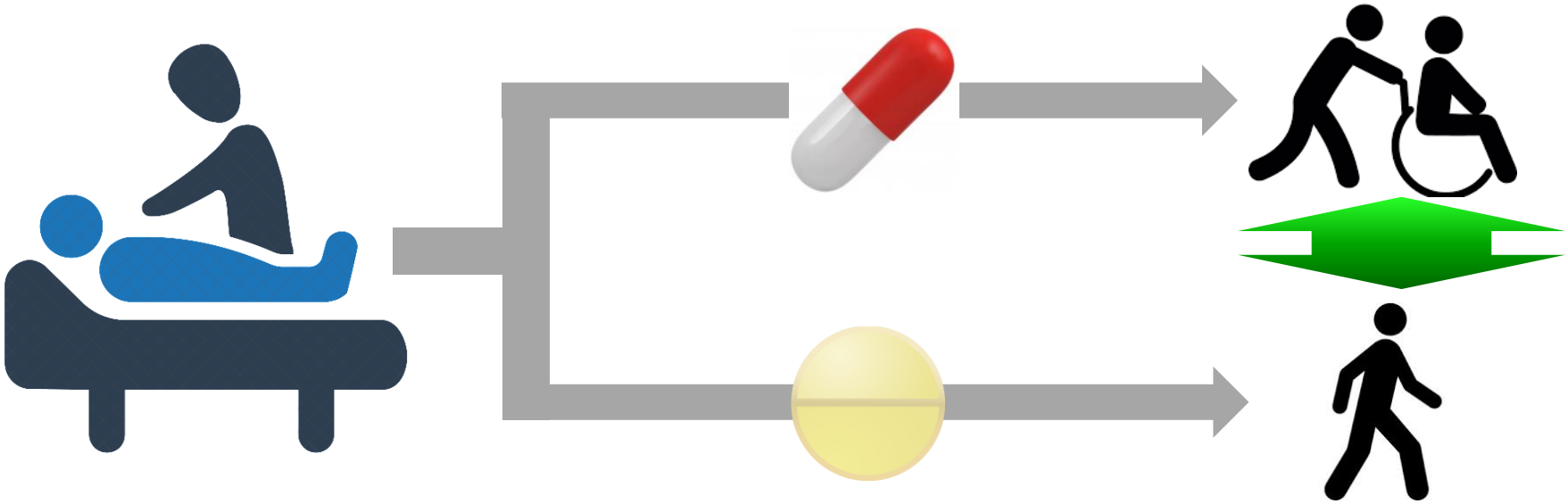
CDDF Multi-stakeholder Workshop / November 2019

Chris Harbron
Expert Statistical Scientist, Roche

Contents

Goal : Understanding Treatment Effects

Understand the difference in outcomes for a patient receiving different treatment regimes



Note : This encompasses the totality of the patient experience : tolerability, adverse events, rescue medications, quality of life as well as the primary and secondary endpoints

Even with complete data, comparisons may be non-trivial

A Thought Experiment of an “Ideal” Study

Roche

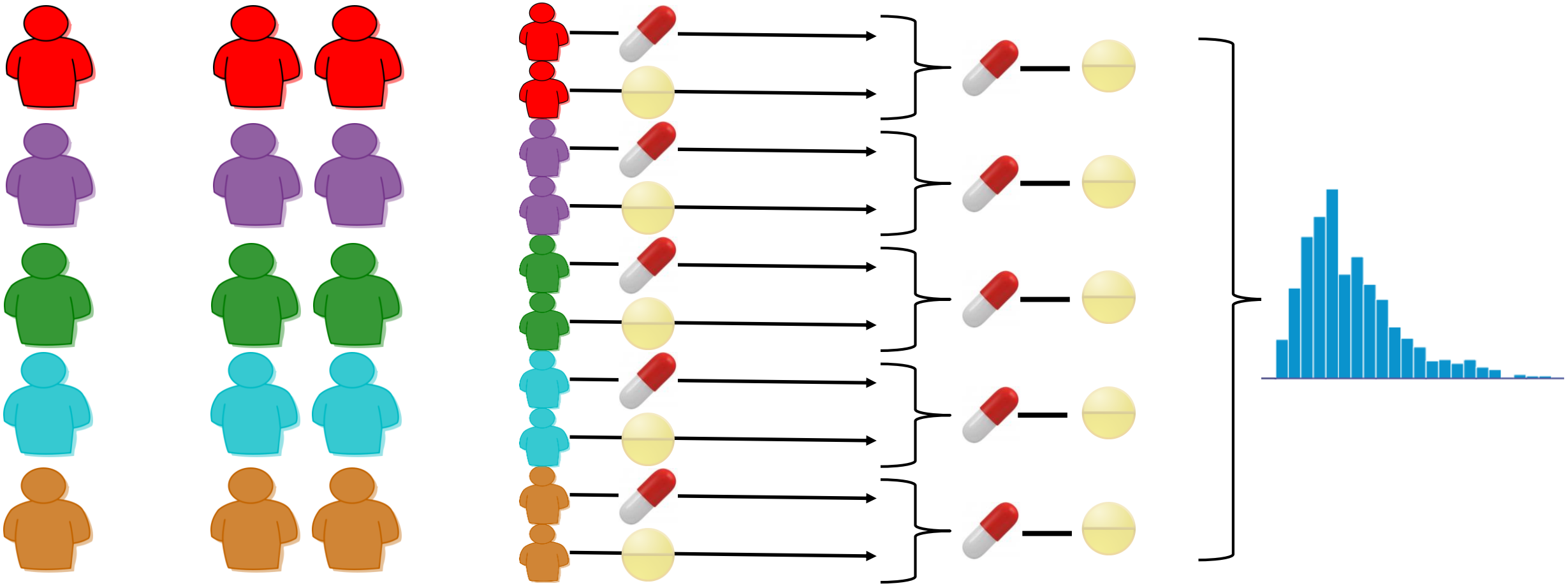
Patients

Clone Each Patient

Give each Clone a Different Treatment

Calculate Differences Between Pairs of Clones

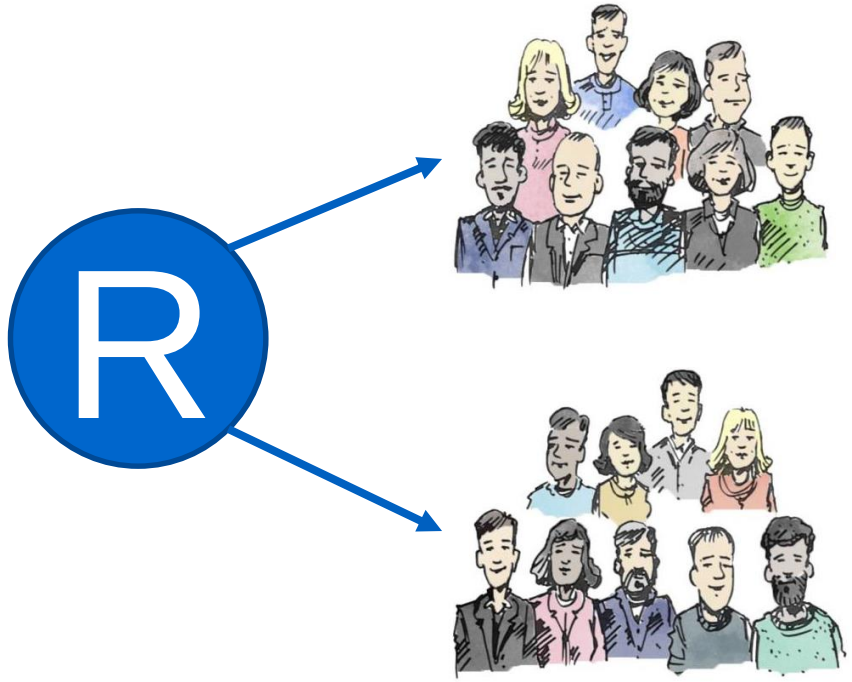
Distribution Of Differences Over Population



Randomisation As A Gold Standard

- We want to estimate the treatment effect :
  - The difference in outcomes from a patient receiving one treatment compared to the other
  - But in practice we can’t observe both outcomes in one patients
    - Cross-over designs approximate this with some assumptions around the stability of a disease – not generally relevant for oncology

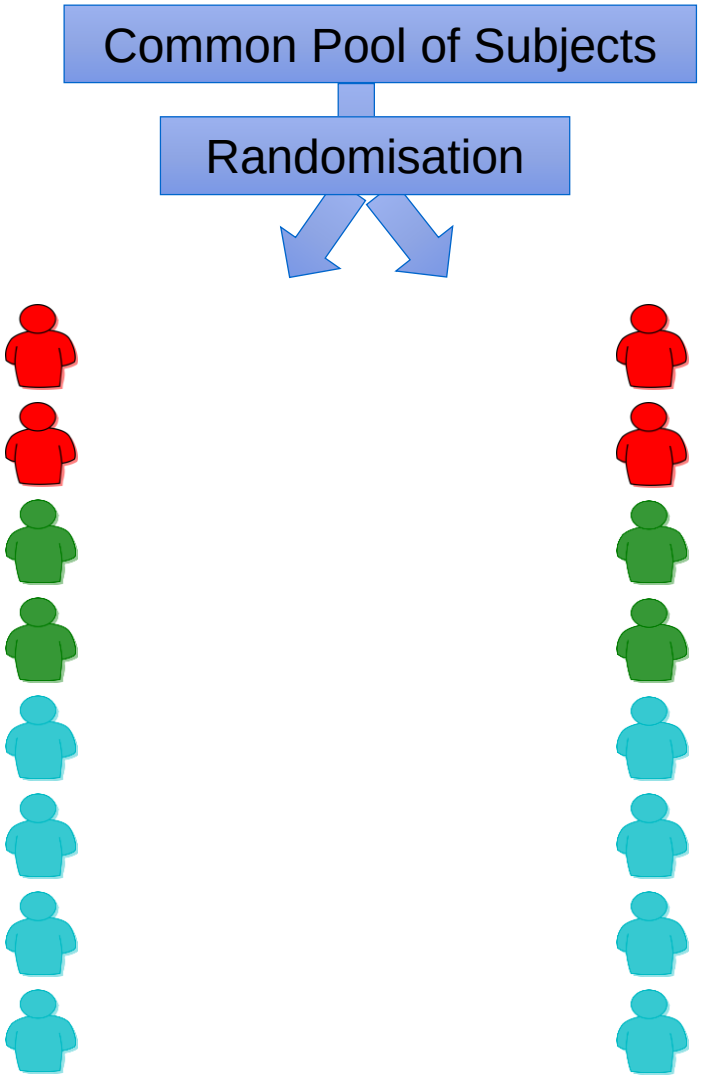
- Randomized controlled trials (RCTs) established as gold-standard for evidence generation in clinical drug development
  - Randomization protects against bias at baseline
  - Allows proper causality assessments and generally lead to unbiased treatment effect estimates



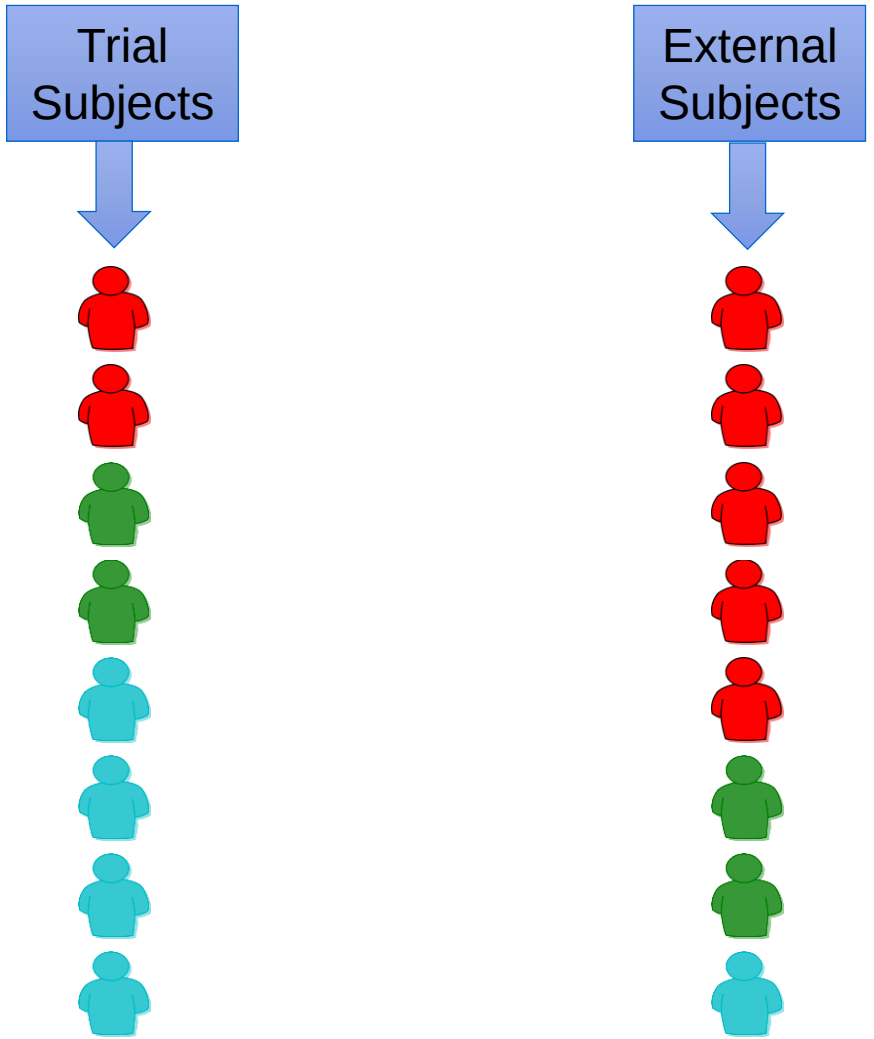
Patient characteristics balanced between the different treatment arms

What's Different with an External Control?

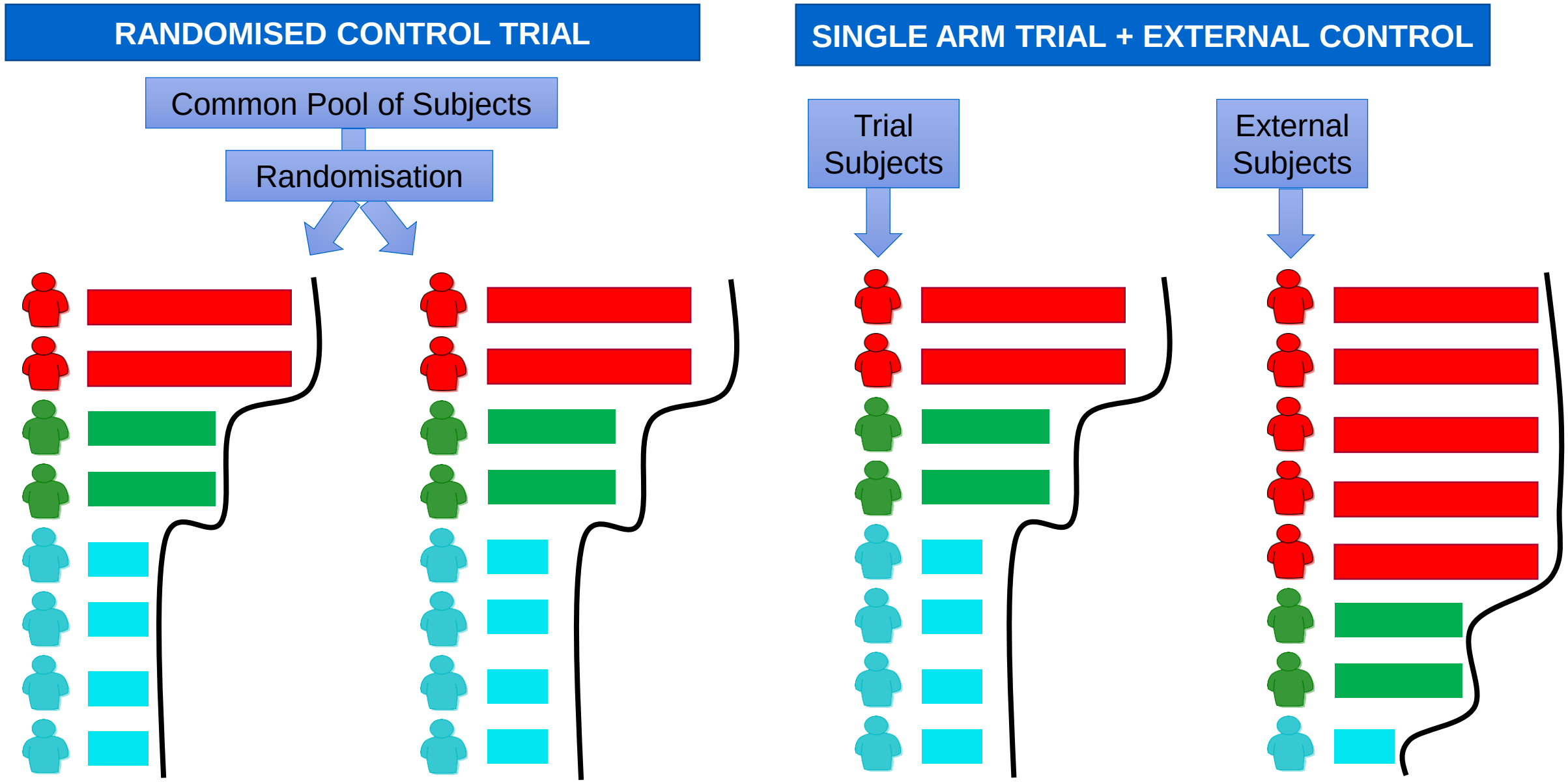
RANDOMISED CONTROL TRIAL



SINGLE ARM TRIAL + EXTERNAL CONTROL



What's Different with an External Control?



Why Might a Randomised and a RWD Control Arm Differ?

STUDY SPECIFIC BIASES

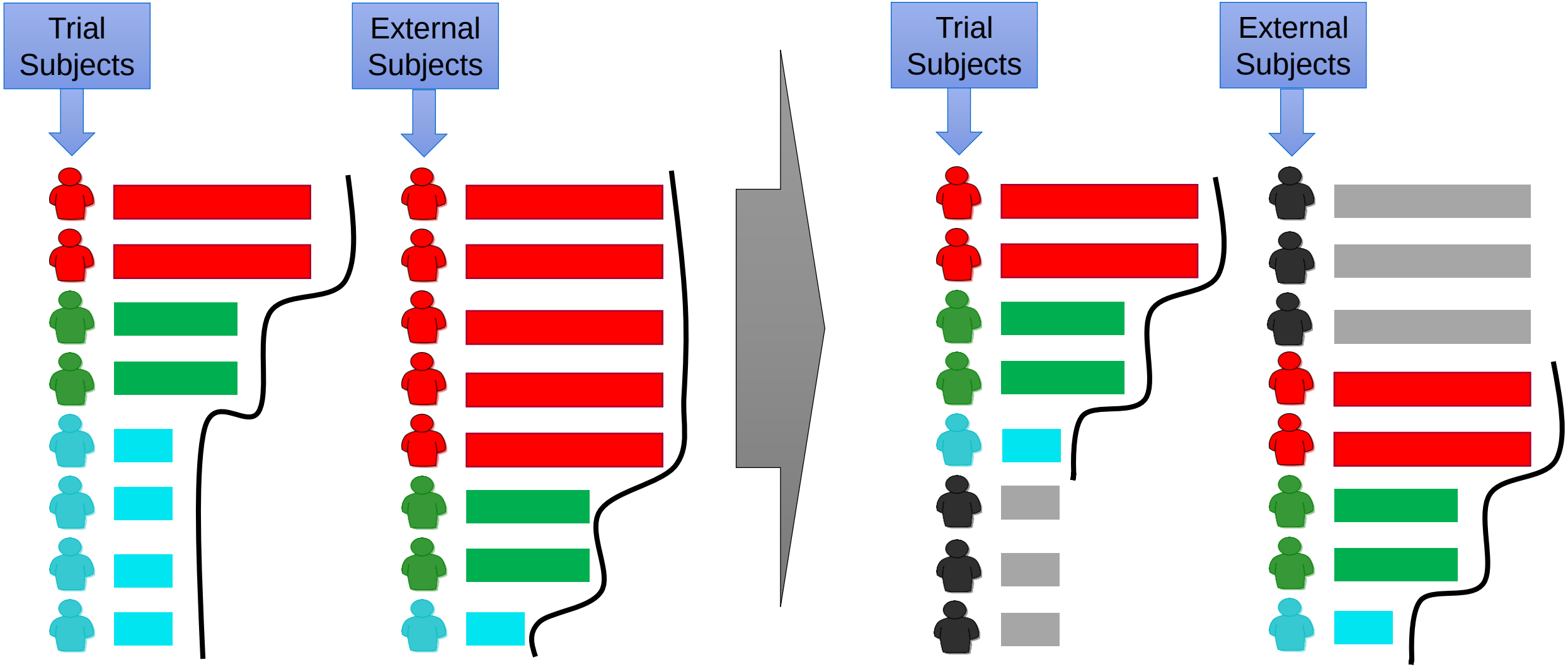
SYSTEMATIC BIASES

- **The process of being in a clinical trial**
  - Patient selection
  - Site selection
  - Higher levels of attention
- **Measurement biases**
  - More events and data captured in clinical trials
  - Variables measured in different ways
  - Some variables (e.g. ECOG) captured in clinical trials but not in clinical practice
  - Clinical trial data may be collected on a more regular basis
- **Unmeasured confounders – Unknown unknowns**
  - We see that absolute results in clinical trials often vary more than we may expect
  - An external control may also have unexpected differences
  - No absolute bound on the size of these differences

Towards Comparable Populations
Inclusion / Exclusion Criteria

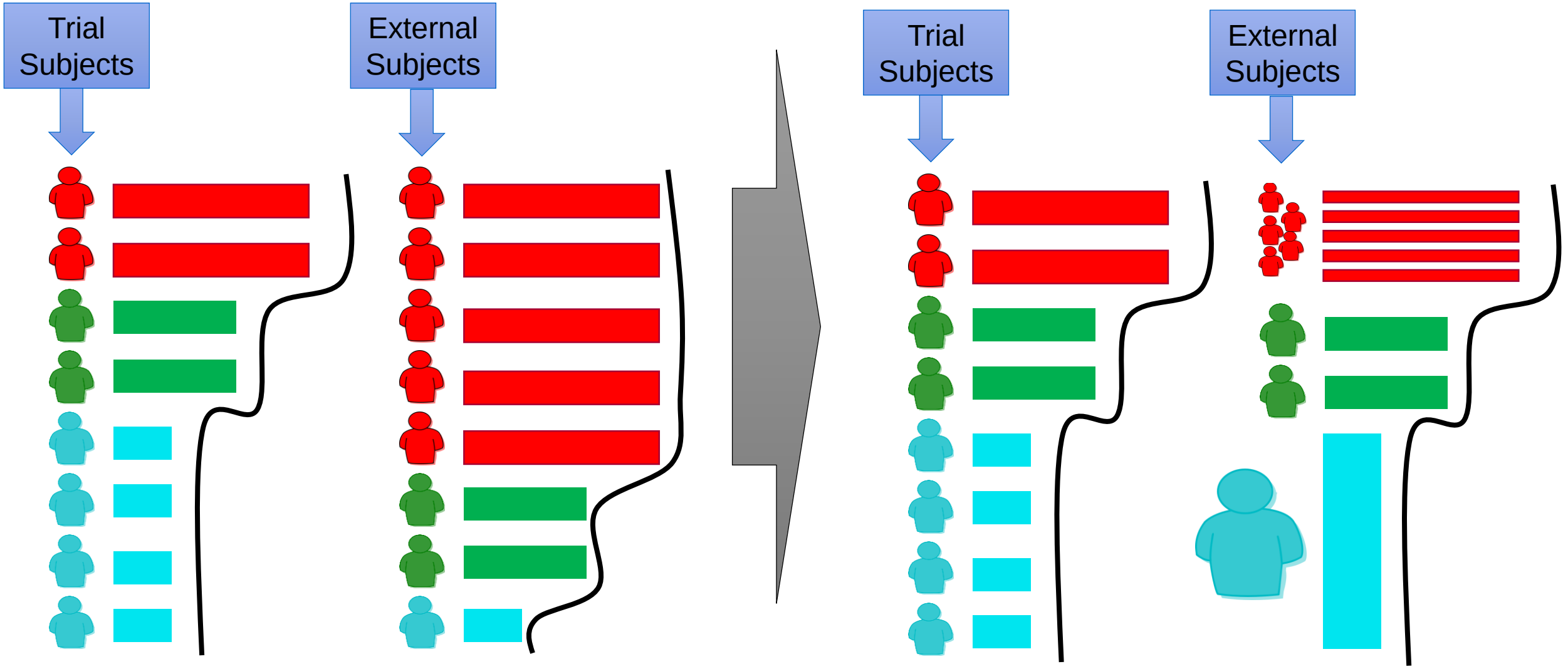
- The **same inclusion and external criteria** need to be applied to the external control subjects as were used in the clinical trial
- This can be challenging if some of the criteria are not included within the external control data or are frequently missing
  - These factors should be **considered** at the **design stage**
- This can often lead to a **reduction in the size of the external control** from repeated shaving of patients from the external control cohort by different criteria
  - This should also be given **consideration** at the **design stage**

Towards Comparable Populations : Matching



Matching may be performed in different ways e.g. selecting the best-match external subject(s) for each trial subject

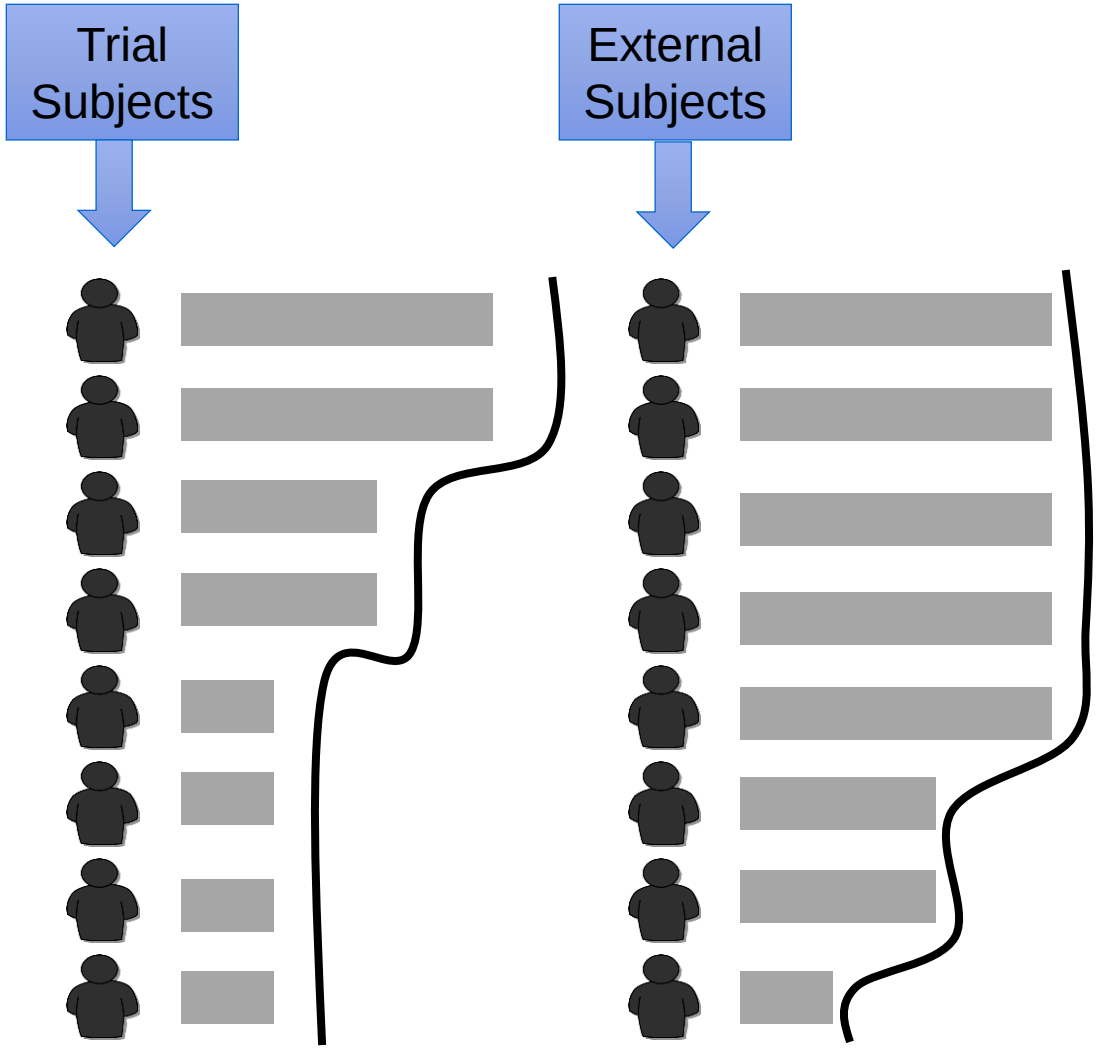
Towards Comparable Populations : Propensity Scoring



Using known prognostic factors to weight external control patients to create a “study-like” population
Note this reweighting can change the statistical behaviour of summary statistics and tests

Towards Comparable Populations : Propensity Scoring

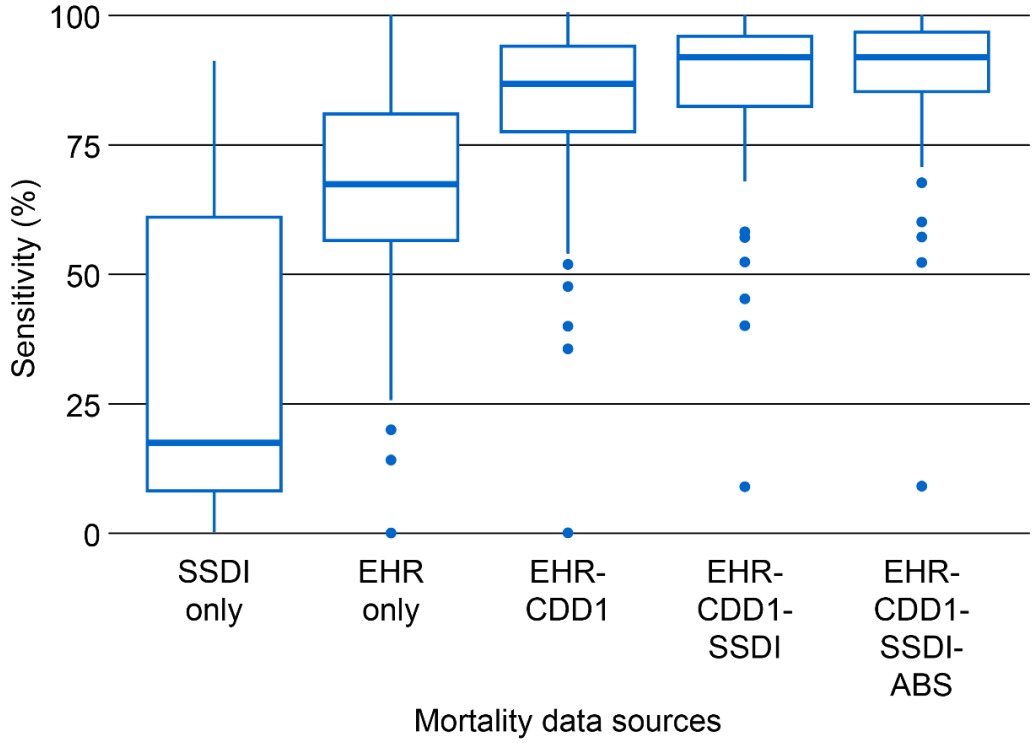
Limitation : Assumes we know and have measured all the relevant factors
May be some unmeasured confounders



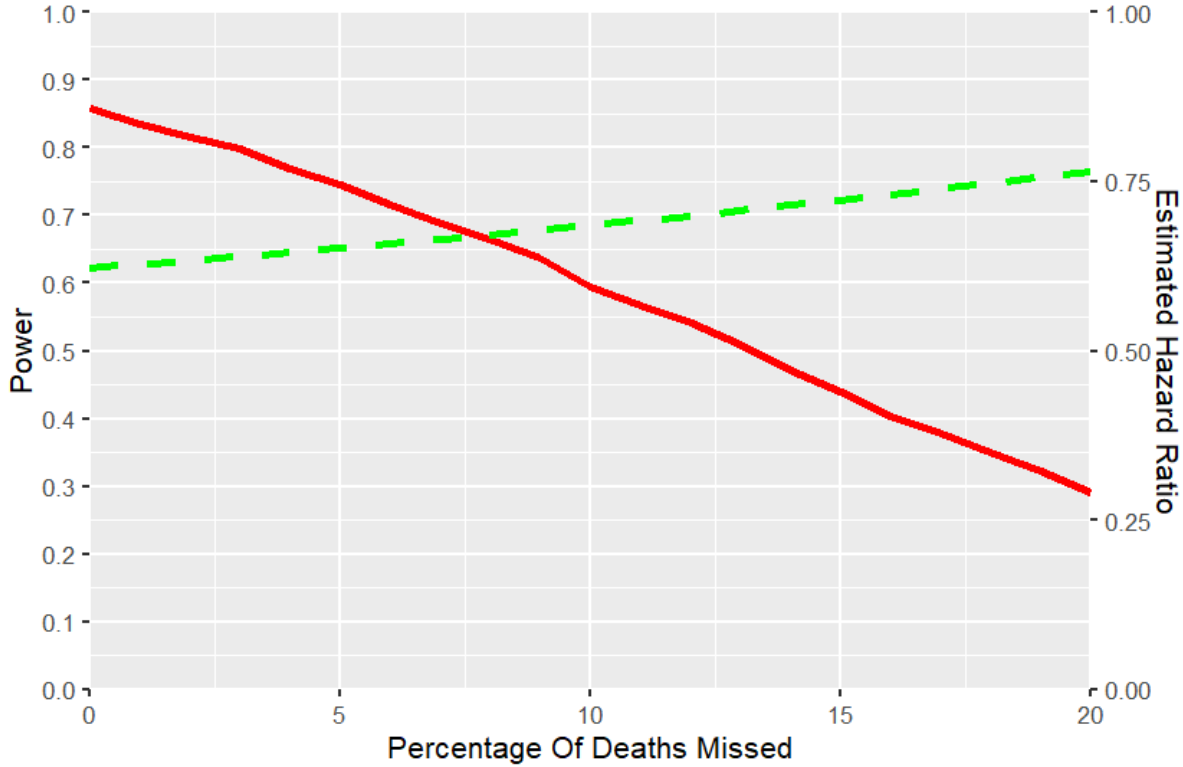
- Unable to apply propensity scoring as we do not know or have the data to weight by
- Most situations we will know, have measured and be able to model some key prognostic variables
- There will inevitably be some degree of unmeasured confounders
- Not easy to say there are no unmeasured confounders or how large any potential bias could be in a situation

Towards Comparable Measurements : Data Quality

Sensitivity of mortality data for advanced non-small-cell lung cancer increased with additional data sources[1]



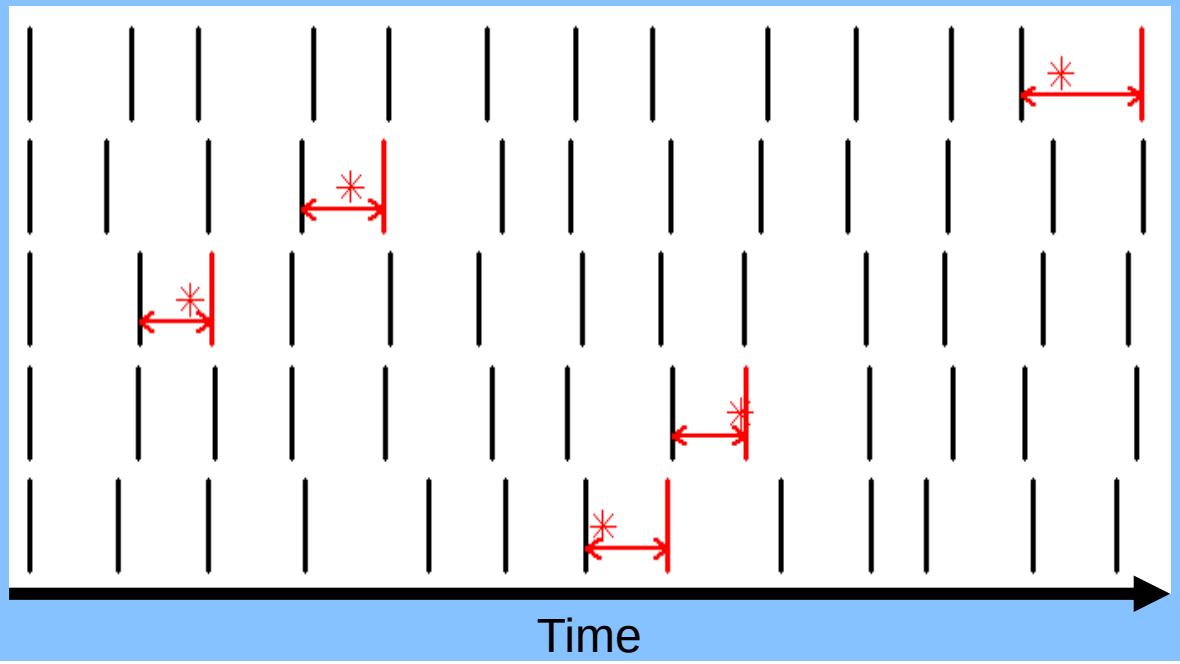
Simulation showing impact of sensitivity of mortality data on power and estimated hazard ratio



ABS, abstracted; CDD1, commercial death data; EHR, electronic health records; IQR, interquartile range; NSCLC, non-small-cell lung cancer; SSDI, social security death index. 1. Curtis MD *et al. Health Serv Res* 2018;53:4460–76

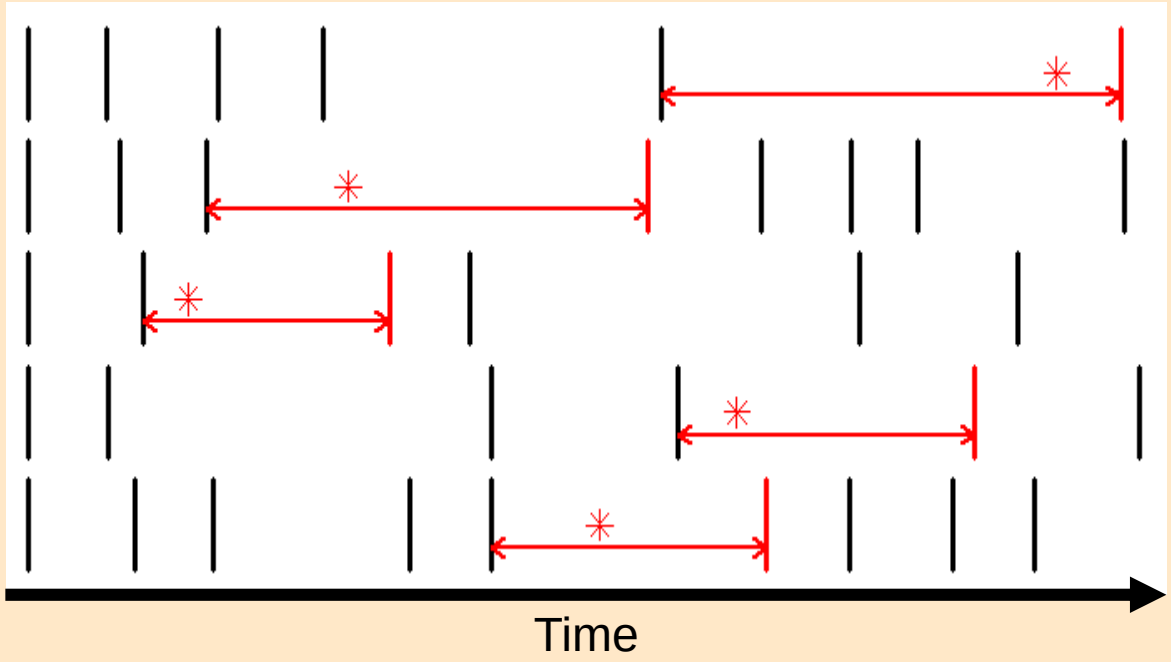
Towards Comparable Measurements :
Measurement Schedules / Data Generating Mechanisms

RCT



Regular visits
Events are interval censored, but effect is comparable over subjects

External Control



Often, irregular and less frequent visits
Impact of interval censoring, may differ between subjects

Towards Comparable Measurements : Baseline Times

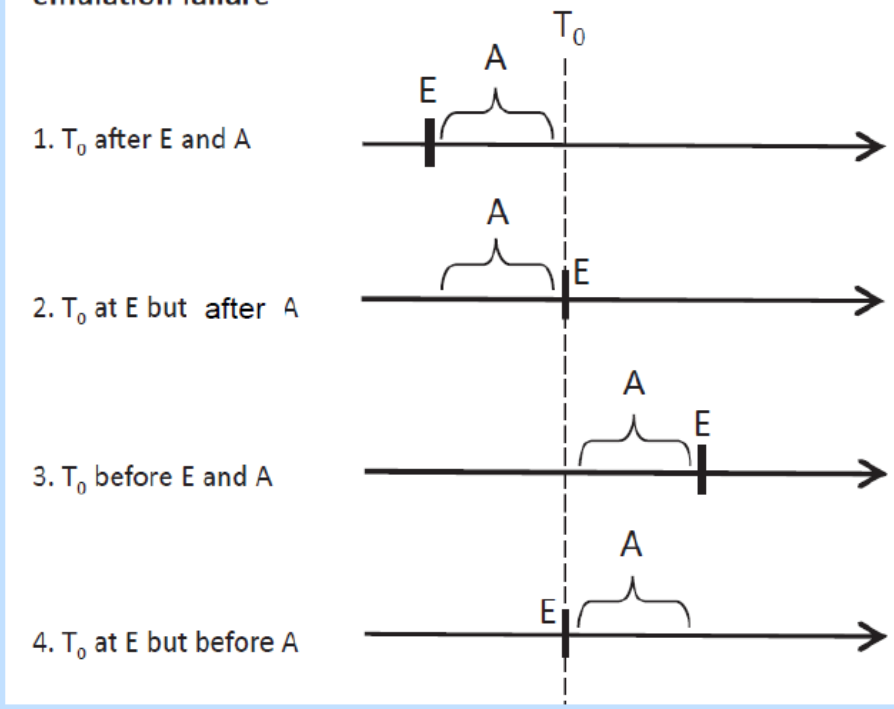
Hernán's criteria for Time Zero :

- Eligibility criteria met
- Treatment strategies assigned
- Study outcomes begin to be counted

- Straightforward in some settings, e.g. a treatment given immediately upon cancer diagnosis

- May be less well defined in other scenarios

- Misalignment can lead to selection bias and immortal time bias

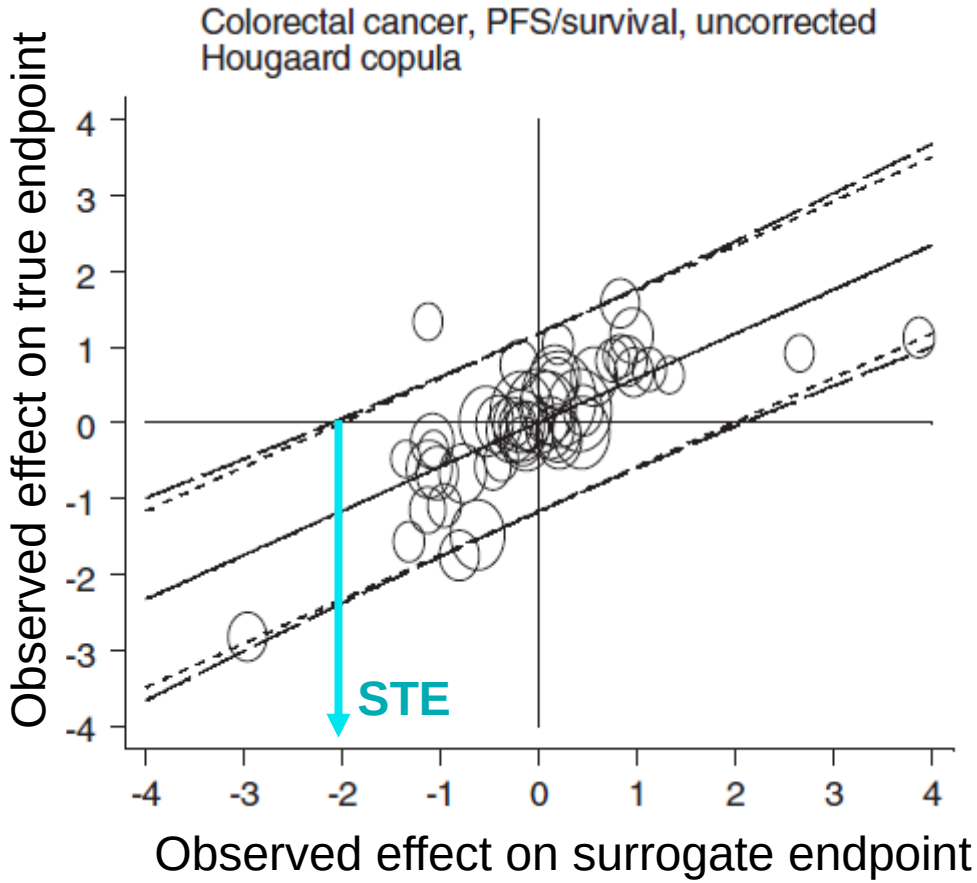


Type of emulation failure	Selection of...	Immortal time
1. T_0 after E and A	eligible individuals who initiate a treatment strategy and remain under follow-up through reset T_0	No
2. T_0 at E but after A	individuals who initiated a treatment strategy before, and remained under follow-up until, eligibility (specified at T_0)	No
3. T_0 before E and A	individuals who initiated a treatment strategy before, and remained under follow-up until, eligibility (specified after T_0)	Yes
4. T_0 at E but before A	eligible individuals at T_0 who remained under follow-up until completing a treatment strategy	Yes

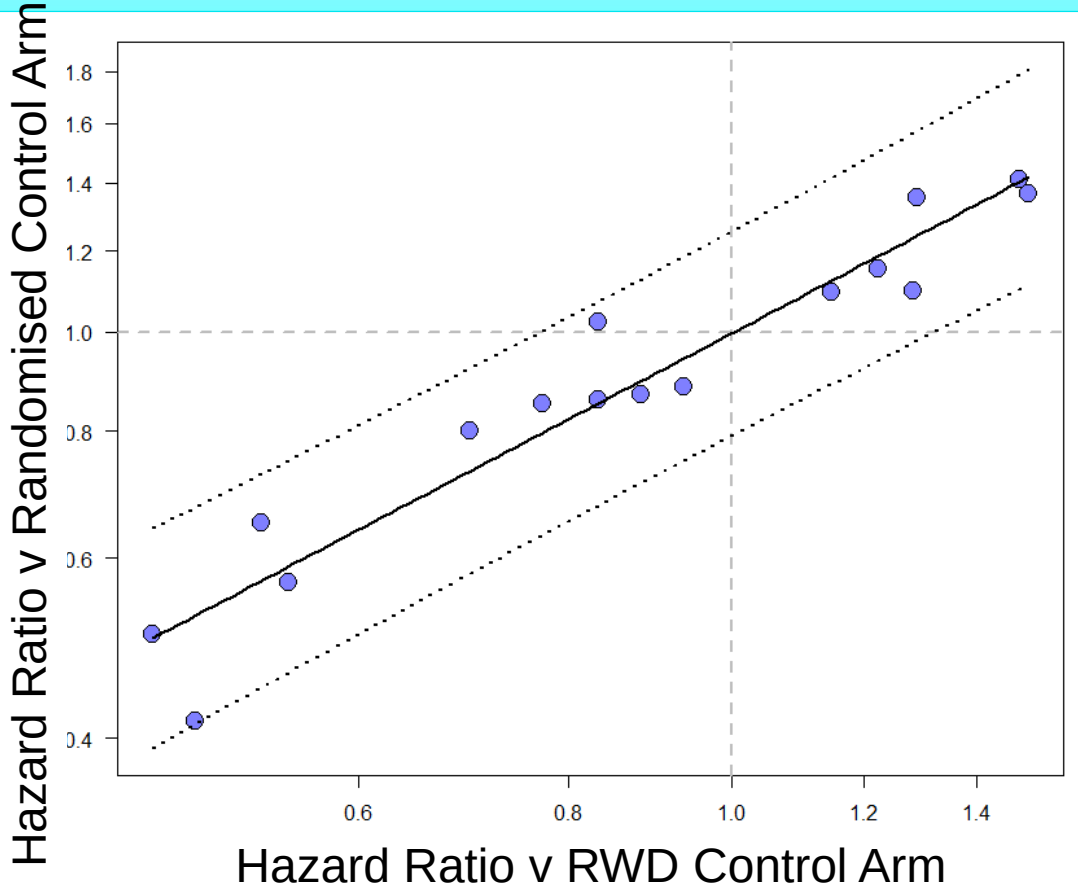
Hernán *et al*. J Clin Epi 2016 79:70-75. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses

Minimising False Positives Idea
Trial Level Surrogacy & Surrogate Threshold Effect

Compare effects in two **endpoints**, e.g. PFS & OS
Objective : to use PFS for **decision making**



Compare effects using two **control arms**, e.g. RCT & RWD
Objective : to use RWD Control for **decision making**



Surrogate threshold effect : An alternative measure for meta-analytic surrogate endpoint validation, Burzykowki & Buyes, Pharmaceutical Statistics 2006;5;173-186

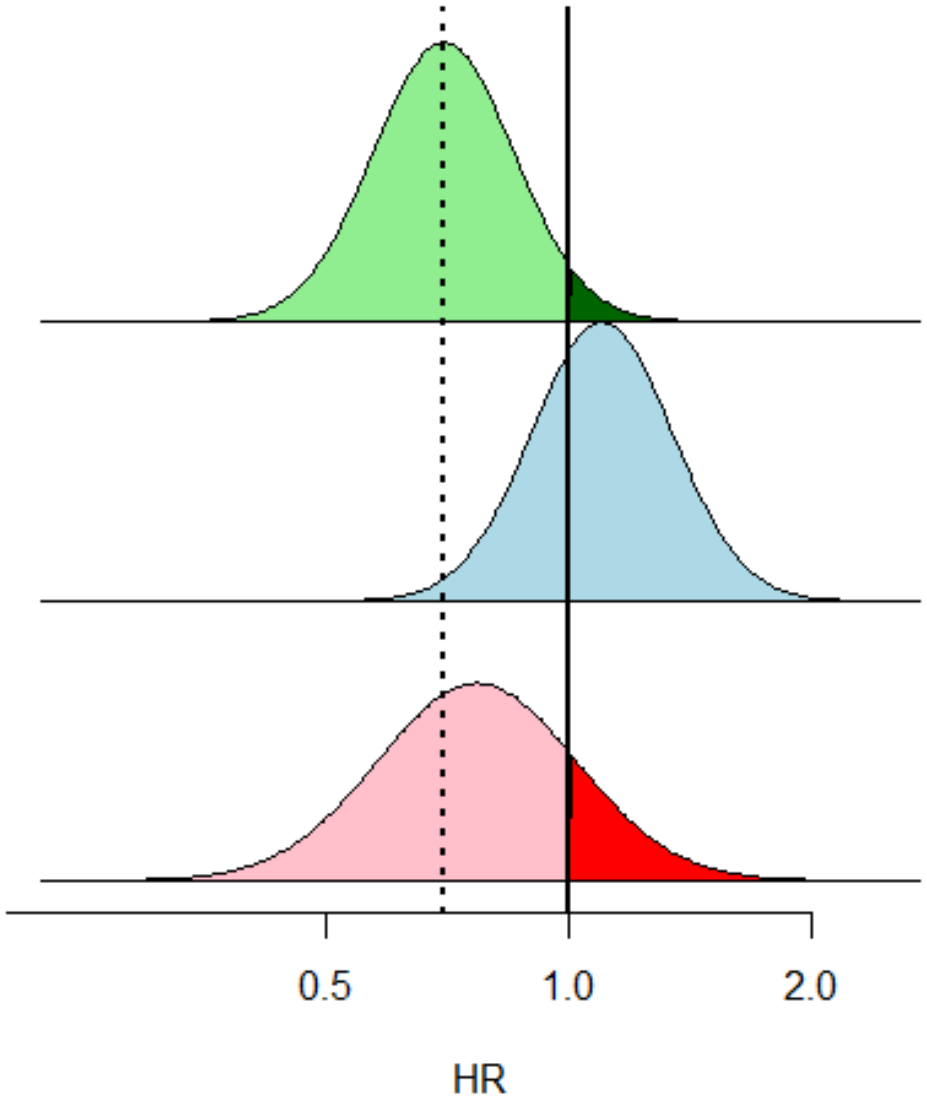
Note : RWD data shown here is artificial data

Minimising False Positives

Treatment vs RWD Control Comparison

Additional Variability Associated With Non-Randomised Comparison

Overall Assessment Of Treatment Effect



- **Increase width of confidence intervals** of estimate of treatment effect representing increased uncertainty from using an external control compared to a randomised control
- **Possibly shift the treatment effect estimate** if a consistent bias had been observed previously
- -> Require a *larger treatment effect* in order to **claim a statistically significant** effect
- -> More **realistic representation** of the **uncertainty** of estimating a **non-randomised treatment effect**

Potential Biases & Mitigations in Using External Controls

Potential Bias	Risk	Potential Mitigation
Selection bias	Different **patient populations** enrolled in clinical trial than in external control.	**Adjust for known confounding** due to differences in patient population
Calendar time bias	Patients treated in the **past** have **worse outcomes** than those treated **today** due to improvements in standard of care over time	Use data from **concurrently treated** external controls
Regional bias	Patient outcomes may **vary between regions** reflecting different healthcare practices between regions.	Use control patients from **same region**
Assessment bias	**Knowledge of therapy can influence** the outcome assessment.	Use a **robust, objective endpoint**
Different endpoint bias	Certain endpoints (e.g. ORR, progression) are **measured differently** in clinical trials than in routine clinical practice (e.g. using standards such as RECIST).	Use an endpoint that is **assessed in the same way** in the clinical trial as the external control
Immortal Bias	**Study start** for every patient **difficult to define** in external data. Differences compared to study can lead to bias.	Use data sources where we can **reliably define study start**, use the equivalent time zero.
Retrospective selection bias	**Retrospective selection** of external data and key analysis features.	**Prospective planning** and **transparent documentation** of all analyses
Study bias	Patients in clinical trials have **different outcomes than in clinical practice.**	An alternative source of external data may be another **clinical trial**

When might external controls be an alternative to randomisation?

- When an RCT may be **unethical** or **impractical** due to practical constraints
  - No accepted Standard Of Care
  - Rare diseases
  - Pediatric indications

- **Natural course of disease is robustly predictable** and outcome is clearly outside of any measurement error
  - Spontaneous shrinkage of a tumour under placebo is highly unlikely

- **Earlier phase studies** – used to inform internal decision making with lower regulatory risk

- Important to consider the **overall objectives and picture**
  - Ethics of clinical trials consider the value to society as a whole
  - Little benefit in running a quicker study which then gets delayed in regulatory/payer approval resulting in a net delay of getting the new treatment to patients

Challenges :
Ensuring robust decision making from the analysis of RWD

- **In which settings could using RWD be most valuable?**
  - When in the development pipeline?
  - Which diseases?

- **How can we avoid biases?**
  - Population biases
  - Measurement biases

- **When biases are unavoidable, or at least the lack of bias cannot be guaranteed, how can we mitigate or minimise them?**

- **How do obtain a realistic estimate of the uncertainty of any conclusions? (avoid inflating type 1 error)?**

- **How do we communicate our levels of confidence to internal and external stakeholders?**

- **What validation experiments could we run to increase confidence?**