Sense and sensibility of estimands for health technology assessment (HTA)

Jason Wang, Celgene

Christine Fletcher, Amgen

Neil Hawkins, Glasgow University

Disclaimer

The views expressed are those of the presenters and should not be understood or quoted as being made on behalf of their affiliations.

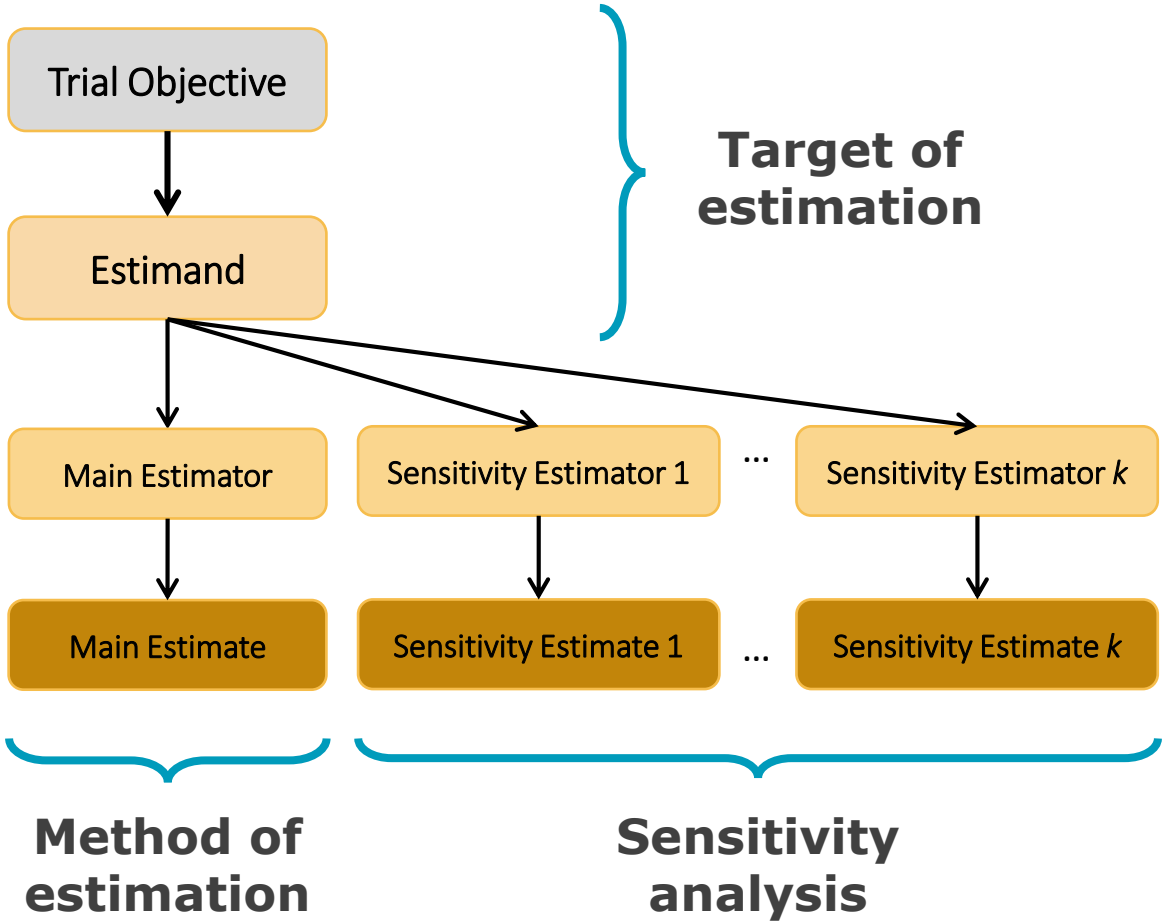
Outlines

- Recent discussion on estimands includes some considerations in HTA aspects, but has not fully covered the wide range of HTA aspects.
- Why considering estimands is important for HTA and what are different from regulatory aspects?
- Determining estimands that make sense for HTA and their estimators.
- What can we learn from HTA about treatment/policy evaluation without randomization or randomization is broken?

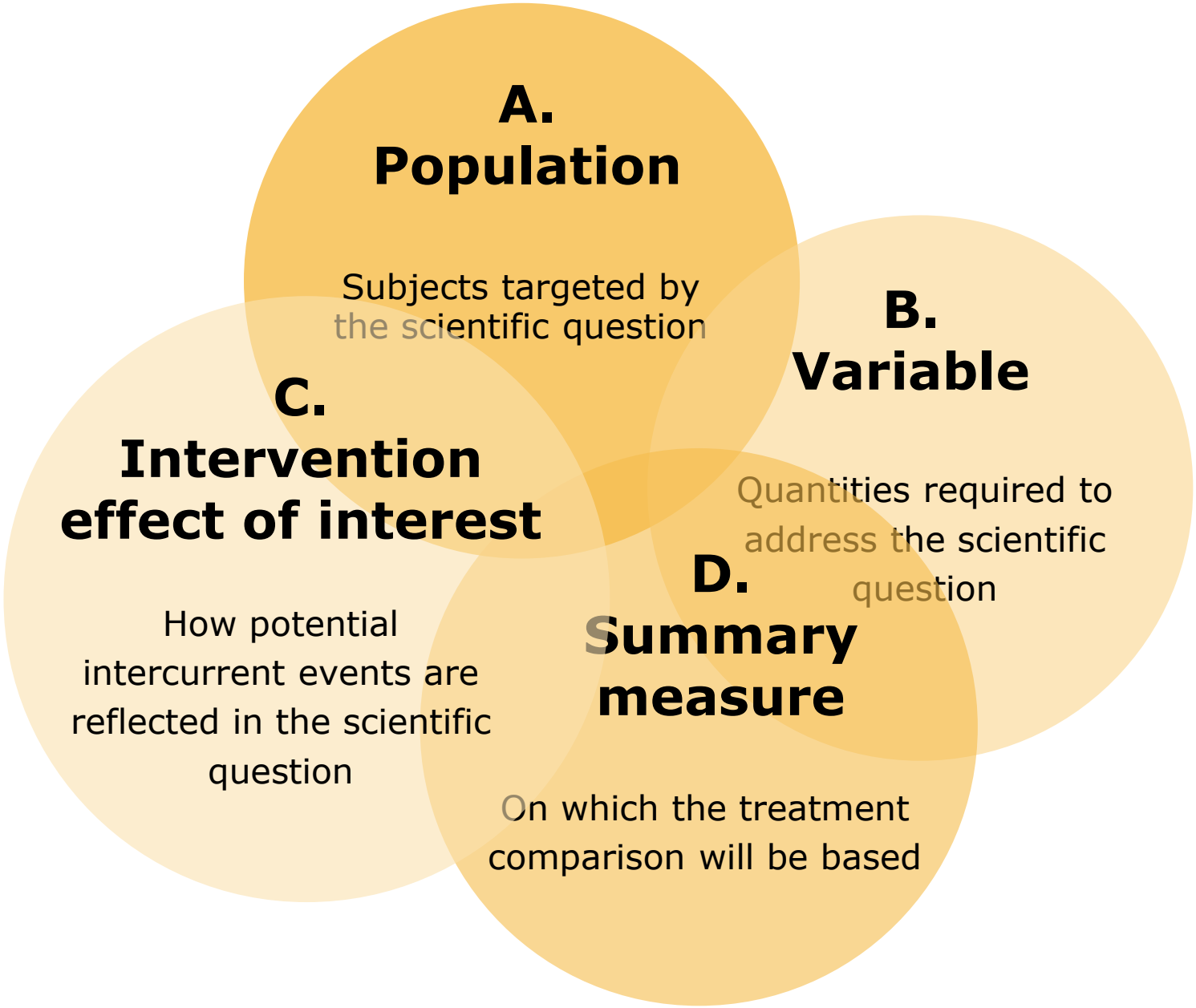
Health technology assessment

- Health technology assessment (HTA) evaluates treatment effectiveness
  - in terms of its clinical, social and economical outcomes,
  - when the test treatment is used in clinical practices
  - in the target patient population.

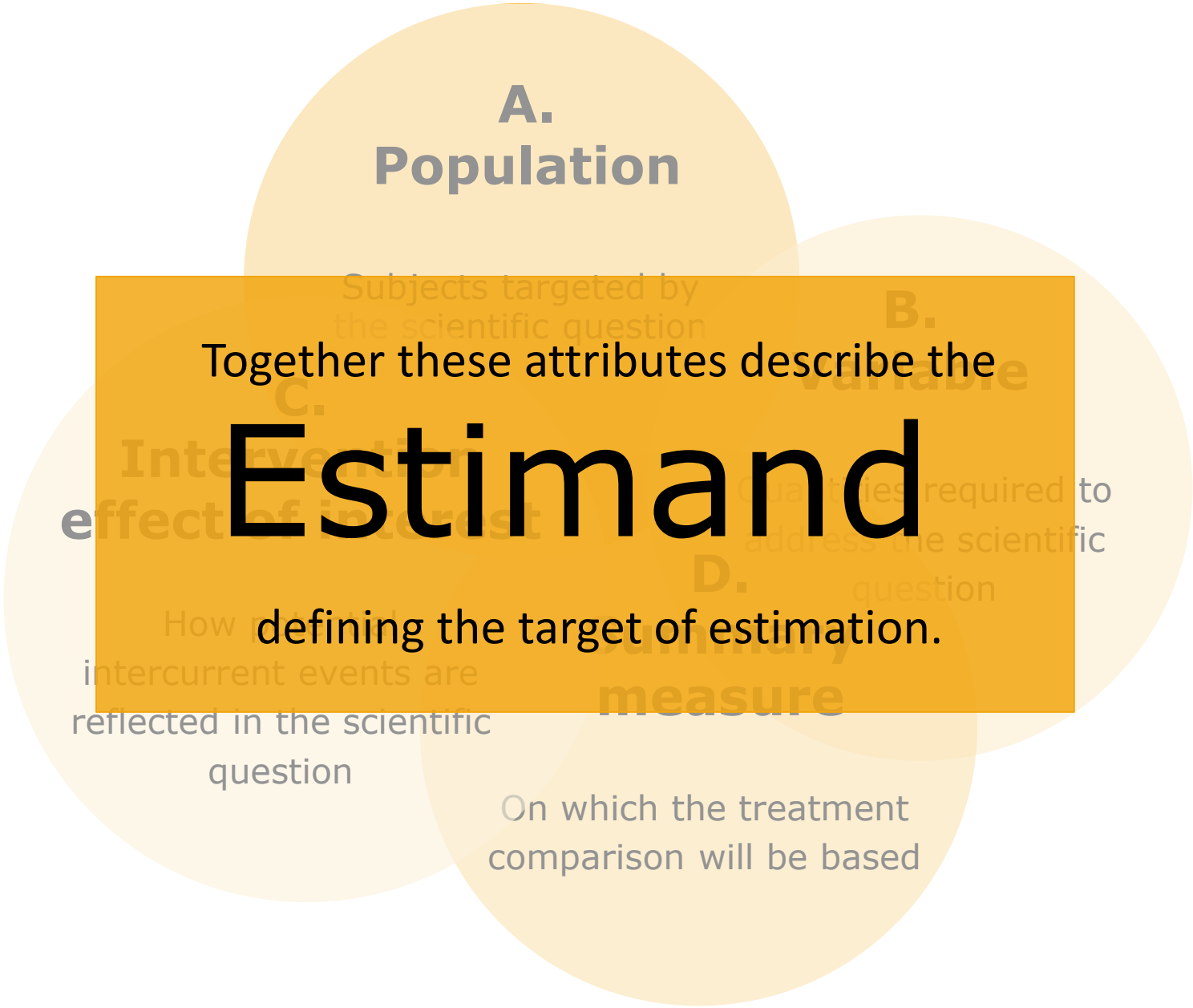
Estimands: a new framework



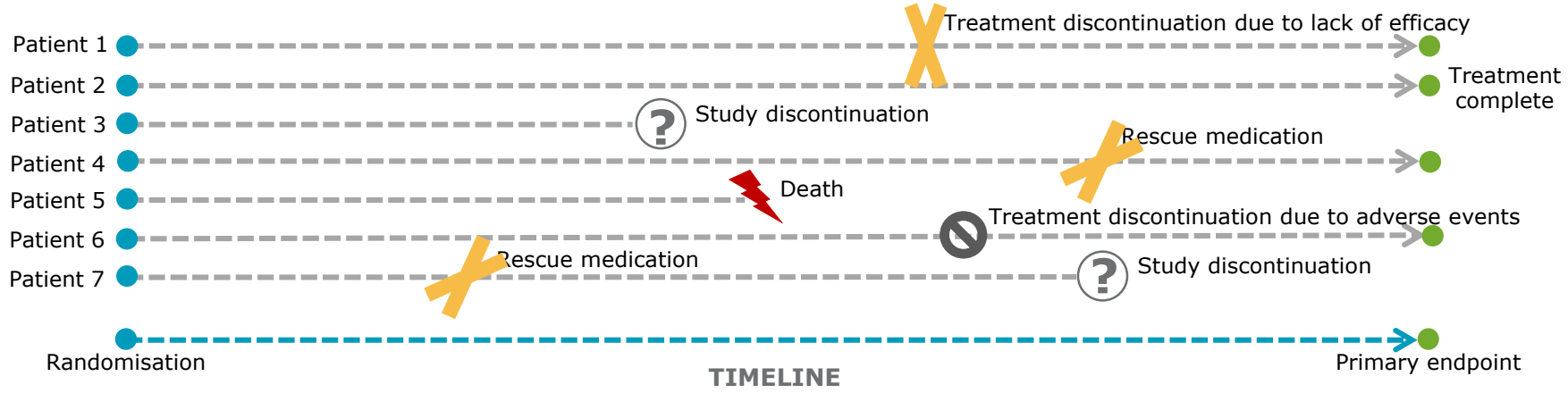
Description of an estimand



Description of an estimand



Intercurrent events



- Such events may include: death, treatment discontinuation due to adverse events or lack of efficacy, use of other medicines affecting the outcome, whether specified or prohibited by the protocol.
- Some reflect clinical practice and some do not.
- Intercurrent events may make the relevance, the definition, or even the existence of the primary variable questionable.

The benefits of considering estimands for HTA:

- It helps to clarify the need for HTA at trial design stages or earlier, along with the development of value dossier.
- The coordination of using estimands for multiple purposes can facilitate using planned trial analyses for HTA.
- It helps to identify statistical problems in estimation, e.g., confounding bias, at early stages.
- It is important for regulatory and HTA parallel consultation, e.g., to confirm estimands for long term treatment effects.

How RCT evidence is used for HTA?

- HTA approaches to using evidences from randomized controlled trials (RCT) vary, but some common features are:
  - Using long term effects to evaluate/predict clinical, social and economical outcomes.
  - Evaluating outcomes when the test treatment is used in current clinical practice, likely as a part of a treatment policy.
  - Using HTA-specific outcomes (e.g., resource data).
  - Using quantitative evaluation, rather than mainly hypothesis testing results.

RCT evidence for HTA: examples

- Oncology trials
  - Some registration trials use progression free survival (PFS) and some use overall survival (OS) as the primary endpoint. Switching from the control to the test treatment upon disease progression is often allowed.
  - OS is the key endpoint for HTA, but PFS may also be used as a surrogate of OS (e.g., the ASCO value framework).
  - Cost-effectiveness analysis (CEA) typically needs both PFS and OS.
  - This approach is often preferred to modeling time-to-progression and time-to-death, as it avoids dealing with competing risks.
  - For CEA, switching to the test treatment may need adjustment, as it does not reflect the clinical practice if the test treatment is no available, but switching from the test treatment to feasible options needs no adjustment.

RCT evidence for HTA: examples

- Type 2 Diabetes:
  - Although HbA1c is often used as the primary endpoint in RCTs for registration, HTA needs to assess clinical outcomes such as diabetic events.
  - CEAs are often based on two sets of models:
    - Models for long term HbA1c changes, using treatment effect from short term RCTs.
    - Models for the relationship between diabetic events and HbA1c.
  - Typical outcomes in RCT, mean HbA1c change at 12/24 weeks
  - Measure at 24 weeks is a better surrogate for long term effect, but treatment changes, dropouts are more likely to happen before the time.

Strategies dealing with intercurrent events

Strategy	Examples	Treatment	Outcome	Population	Randomization based estimator
Treatment policy	Mean difference between randomized groups (ITT estimand)	Varying	Original	Whole	Yes
Composite endpoint	Response rate at n weeks, counting treatment changes before that as failure	As randomized	Modified	Whole	Yes
Hypothetical	The effect if everyone compliances.	As randomized	Original	Whole	Varying
Principal stratum	The effects within compliers	As randomized	Original	Varying	Varying
While on treatment	Last HbA1c measure while on randomized treatment.	As randomized	Modified	Whole	Yes

Strategies dealing with intercurrent events

Strategy	Pros	Cons
Treatment policy	• HTA typically evaluates effectiveness of a policy rather than a treatment.	• Hard/impossible to replicate feasible policies for multiple payers in a single trial. • Effects of test treatment may be diluted.
Composite endpoint	• A good composite endpoint (e.g., QALY) may fit to HTA well • Meaningful for BRA and HTA if it reflects benefit-risk balancing	• Only work well for some events (e.g., death) • May have to categorize continuous measures, which may not be good for HTA.
Hypothetical	• Comparisons between randomized treatments	• The hypothetical scenario is meaningful only under some situations
Principal stratum	• Randomization may still be useful • Extensive researches have been done	• Principal strata may not be identifiable • They may not be of interest.
While on treatment	• May construct meaningful estimands, e.g., together with the composite approach	• May only measure short-term effects for some patients.

Treatment or policy estimand?

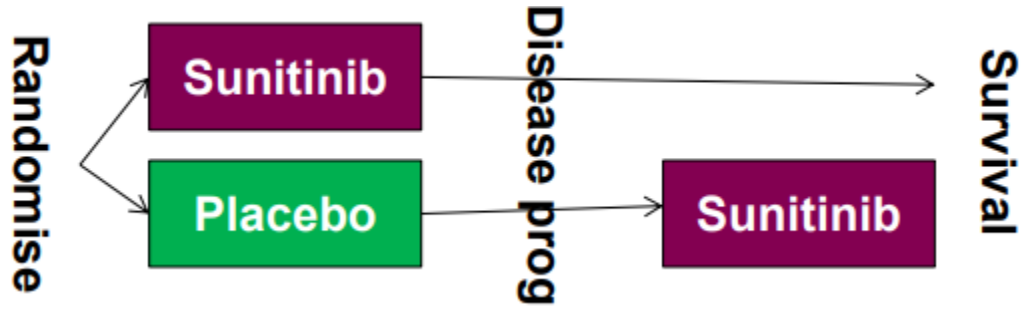
- Changes to subsequent treatments, including switching between randomized treatments, are common.
- Adjust them or not, and if yes, to what?
  - In general we should always adjust artificial changes (or no change) to that of common practice to construct two treatment policies : one with and another without the test treatment.
  - If patients on the test treatment A may be switched (following clinical practices) to either B or C, while only B is licensed in the target population, then those switched to C should be adjusted to as if they had switched to B.

Example: Treatment switching and progression free survival (PFS) as estimand

- Disease progression is an intercurrent event.
- Two strategies lead to PFS
  - While on treatment: No switching before event/censoring happens.
  - Composite endpoint: Combining death and progression.
- Pros: Randomization based, hence estimation is easy.
- Cons:
  - PFS may not be a good OS predictor. E.g., the ASCO framework counts PFS evidence as 70% of OS in the scoring algorithm.
  - CEA will need both PFS and OS.

Treatment switching and overall survival (OS) as estimand

- Example: The Sunitinib trial
  - Patients on pcb can switch to sunitinib.
  - Patients on Sunitinib can't switch.



- Approaches to deal with treatment switching

1. Treatment policy approach: Sunitinib before or after placebo?
2. Hypothetical approach: adjusting to no switching:
   - Randomization based estimator: g-estimation based on rank preserving AFT models, assuming no heterogeneity.
   - Non-randomization based estimator: inverse probability weighting.

Estimands of average effect for HTA, health economics etc.

- Several estimands have been used for, e.g., policy evaluation, and have been extensively investigated.
  - Average treatment effects (ATE): average effect in the whole population.
  - Local average treatment effect (LATE): average effect among compliers. (also known as complier's average causal effect(CACE).
  - Average treatment effect on treated: average effect among treated patients.
  - Average treatment effect on untreated: average effect among untreated, had they been treated.
  - When they can be identified and how to estimate them?

Principal strata

Treated/Randomized	R=0	R=1
T=0	Complier/never-taker	Never-take/defier
T=1	Always-take/defier	Complier/always taker

A typical principal stratum: the compliers stratum

- Compliers: those who take active treatment (T=1) if given 1, and take control (T=0) if given 0.
- Strata are not fully identifiable.
- They may not be of interest: we are interested in those comply to the active treatment (treated).
- But in some situations, **compliers = treated**.

Hypothetical example: switching 1 -> 0 due to safety events.

- Assuming T=1 has additional toxicity, hence if one can tolerate T=1, he will tolerate T=0.
- 100 patients are randomized to R=1 and 50 stay on T=1 and 50 switch to T=0.
- 100 patients are randomized to R=0 and all stay on T=0.
- Can we estimate the effect of T=1 among those staying on 1?
- Can we estimate ATE?

	0 (n=100)	1 (n=100)
0	100	50
1	0	50

Principal stratification and IV estimator

Example: (continued)

- Assuming T=1 has 1 unit effect and T=0 has 0 unit effect, so the mean treatment effects in the R=1 and R=0 groups are 0.5 and 0.
- Intuitively since among R=1 the exposure rate to the active is 50% so the effect per exposed is 0.5 / 50% =1; we recover the true treatment effect.
- This is the instrumental variable (IV) estimator (Wald estimator) for LATE:
- $$\frac{\text{mean(eff.}|\text{R=1)} - \text{mean (eff.}|\text{R=0)}}{(\text{P(T=1}|\text{R=1)} - \text{P(T=0}|\text{R=0)}} = \frac{0.5 - 0}{0.5 - 0} = 1.$$

- When the estimator is valid?
  - Constant treatment effect.
  - Assuming (deterministic) monotonicity: for every subject T(R=1)>=T(R=0), valid even when T depends on heterogeneity (essential heterogeneity).
  - Can we replace it with stochastic monotonicity (P(T=1|R=1)>=P(T=1|R=0) in a population)?
- No good estimator for ATE (except when it is the same as LATE).

Discussion

- Estimands play an equally important role in HTA, just as for regulatory purposes
- Early consideration of what should be estimated for HTA in trial design stage brings multiple benefits.
- Estimation of some estimands is challenging, but extensive approaches have been done in relevant areas
- How far randomization based estimation can go?
- The role of alternatives (modeling, covariates) to the randomization based estimator.
- Challenges and opportunities for statistician to contribute in HTA.