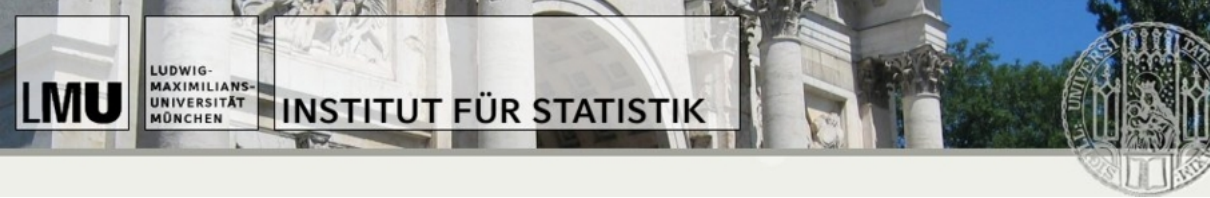


Causality

Chapter 4: Randomized Studies

Summer Semester 2023
Version: May 5, 2023

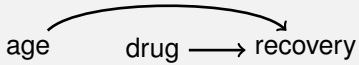
Michael Schomaker
Ludwig-Maximilians-Universität München, Institut für Statistik

What is a randomized study/experiment?

- Randomized studies, and experiments, describe settings where (parts of) the system under study is under control of the investigator.
- In particular, the treatment assignment mechanism is under control of the investigator(s).
- One may assign the treatment randomly, which means that each unit i receives $A = a$ (e.g. $A = 1$ and $A = 0$) based on a prior specified probability $P(A = a) = p_a > 0$.
- A *natural* experiment also has a particular intervention assignment mechanism, which is however not under control of the investigator.

Why does this help us to establish causality?

Example:



There is no arrow that goes into A (i.e., drug): this encodes our *knowledge* that treatment assignment does not depend on any third variables (e.g., age), only on randomness.

It follows immediately from the back-door theorem that the causal effect from A on Y is identified!

Can you see this? What else do we know from the theorem?

Identification assumptions are met by design!

1. Positivity guaranteed due to probabilistic assignment:

$$0 < P(A = a) < 1$$

2. Exchangeability is achieved as well: $Y^a \coprod A \quad \forall a$

   $P(Y^a = 1|A = 1) = P(Y^a = 1|A = 0) = P(Y^a = 1)$

   Even better than conditional exchangeability!

3. It is also in our hands to achieve consistency. Why?

It follows that:

$$E(Y^1) - E(Y^0) \overset{exch.}{=} E(Y^1|A=1) - E(Y^0|A=0)$$
$$\overset{cons.}{=} E(Y|A=1) - E(Y|A=0)$$

We haven't talked about estimation yet, but you can see it will be easy in the case of a *perfect*, idealized, randomized study (experiment).

Causality
Michael Schomaker

Block randomization (conditional randomization)

- Define K blocks of units, b_k, $k = 1, \ldots, K$, through specific (baseline) covariate strata
- Randomize within blocks
- This means that randomization depends on covariate values
- $\rightarrow Y^a \coprod A|L$
- Cluster randomization: randomize *groups* of units
- A combination of block and cluster randomization is possible

What could be the motivation for block randomization?

Example: Ebola

Example: "For this open-label, cluster-randomised ring vaccination trial, suspected cases of Ebola virus disease in Basse-Guinée (Guinea, West Africa) were independently ascertained by Ebola response teams as part of a national surveillance system. [...] clusters of all contacts and contacts of contacts were defined and randomly allocated 1:1 to immediate vaccination or delayed (21 days later) vaccination with rVSV-ZEBOV [...] Adults (age $\geq$18 years) who were not pregnant or breastfeeding were eligible for vaccination. Block randomisation was used, with randomly varying blocks, stratified by location (urban vs rural) and size of rings ($\leq$20 vs $>$20 individuals). [...] The primary analysis was per protocol and compared the incidence of Ebola virus disease in eligible and vaccinated individuals in immediate vaccination clusters with the incidence in eligible individuals in delayed vaccination clusters" [1]

Block Randomization

$$E(Y^1) - E(Y^0)$$

Block Randomization

$$
\begin{aligned}
& E(Y^1) - E(Y^0) \\
= \quad & E_L \left\{ E(Y^1 - Y^0 | L) \right\}
\end{aligned}
$$

Block Randomization

$$
\begin{aligned}
& E(Y^1) - E(Y^0) \\
= \quad & E_L\left\{E(Y^1 - Y^0|L)\right\} \\
= \quad & E_L\left\{E(Y^1|L) - E(Y^0|L)\right\}
\end{aligned}
$$

Block Randomization

$$
\begin{aligned}
& E(Y^1) - E(Y^0) \\
= \quad & E_L\left\{E(Y^1 - Y^0|L)\right\} \\
= \quad & E_L\left\{E(Y^1|L) - E(Y^0|L)\right\} \\
\overset{?}{=} \quad & E_L\left\{E(Y^1|A=1,L) - E(Y^0|A=0,L)\right\}
\end{aligned}
$$

Block Randomization

$$
\begin{aligned}
& E(Y^1) - E(Y^0) \\
= \quad & E_L\left\{E(Y^1 - Y^0|L)\right\} \\
= \quad & E_L\left\{E(Y^1|L) - E(Y^0|L)\right\} \\
\overset{\text{cond. exch.}}{=} \quad & E_L\left\{E(Y^1|A=1,L) - E(Y^0|A=0,L)\right\}
\end{aligned}
$$

Block Randomization

$$
\begin{aligned}
& E(Y^1) - E(Y^0) \\
= \quad & E_L\left\{E(Y^1 - Y^0|L)\right\} \\
= \quad & E_L\left\{E(Y^1|L) - E(Y^0|L)\right\} \\
\overset{\text{cond. exch.}}{=} \quad & E_L\left\{E(Y^1|A=1,L) - E(Y^0|A=0,L)\right\} \\
\overset{?}{=} \quad & E_L\left\{E(Y|A=1,L) - E(Y|A=0,L)\right\}
\end{aligned}
$$

Block Randomization

$$
\begin{aligned}
& E(Y^1) - E(Y^0) \\
= \quad & E_L\left\{E(Y^1 - Y^0|L)\right\} \\
= \quad & E_L\left\{E(Y^1|L) - E(Y^0|L)\right\} \\
\overset{\text{cond. exch.}}{=} \quad & E_L\left\{E(Y^1|A=1,L) - E(Y^0|A=0,L)\right\} \\
\overset{\text{cons.}}{=} \quad & E_L\left\{E(Y|A=1,L) - E(Y|A=0,L)\right\}
\end{aligned}
$$

Block Randomization

$$
\begin{aligned}
& E(Y^1) - E(Y^0) \\
= \quad & E_L\left\{E(Y^1 - Y^0|L)\right\} \\
= \quad & E_L\left\{E(Y^1|L) - E(Y^0|L)\right\} \\
\overset{\text{cond. exch.}}{=} \quad & E_L\left\{E(Y^1|A=1,L) - E(Y^0|A=0,L)\right\} \\
\overset{\text{cons.}}{=} \quad & E_L\left\{E(Y|A=1,L) - E(Y|A=0,L)\right\}
\end{aligned}
$$

If L is binary:

$$
\begin{aligned}
= \quad & (E(Y|A=1,L=1) - E(Y|A=0,L=1)) \times P(L=1) + \\
& (E(Y|A=1,L=0) - E(Y|A=0,L=0)) \times P(L=0)
\end{aligned}
$$

$\rightarrow$ Effect is identified

What could be an estimation technique?

Social randomization

Social interventions are sometimes difficult to randomize or implement

1. Effect of intervention may only be seen after years/decades

   Example: The effect of education programs during childhood may only be seen in the long term.

2. Ethical considerations

   Example: How can you randomize “poverty”?

   Example: Treatment A is known to be superior to B, but you may want to compare C to both A and B. Evaluating B again may be unethical.

3. Considerations of interference and consistency

   Example: “Effect of poverty” may not be well-defined and vary with respect to own social network.

Social randomization

- Option 1: Use observational data and estimation methods from chapters 5-10

  Problem: unmeasured confounding, difficulty to develop a causal model

  Example: Does relocation to a “bad neighbourhood” make you sick; or is it that people with poor health have reduced income due to illness and thus limited options on where to live?

- Option 2: randomize with unambiguous intervention

  Example: Intervention: Voucher to rent in area with $< 10\%$ “poverty”. Eligibility criteria: must live in “poor” area ($> 40\%$ poverty”) and have kids. Outcome: asthma attack (among others). [2]

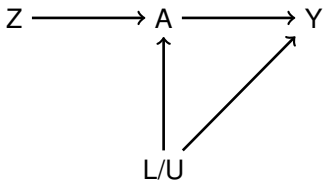
Non-adherence and non-compliance

We may have situations where the offered treatment is not taken:

$$
\begin{aligned}
Z &= \text{randomized treatment (intervention offered)} \\
A &= \text{actual treatment taken} \\
Y &= \text{outcome} \\
L &= \text{measured covariate} \\
U &= \text{unmeasured covariate}
\end{aligned}
$$

Example: Patients may not adhere to the treatment (plan) that was offered to them. In the above example, people may not accept the vouchers that were given to them.

What is the situation?



Non-adherence and non-compliance: 4 groups

Compliers: those who follow the assigned treatment

$(A^{Z=1} = 1; A^{Z=0} = 0)$

Always takers: those who will take the treatment irrespective of assignment

$(A^{Z=1} = 1; A^{Z=0} = 1)$

Never takers: those who will not take the treatment irrespective of assignment

$(A^{Z=1} = 0; A^{Z=0} = 0)$

Defiers: those who will always do the opposite of what they were told

$(A^{Z=1} = 0; A^{Z=0} = 1)$

The last three groups are the “non-compliers”

Implications

- If we *observe* that an individual was assigned to $Z = 1$ and took treatment $A = 1$, we do not know whether she is a complier or an always-taker.

- If we *observe* that an individual was assigned to $Z = 1$ and took treatment $A = 0$, we do not know whether he is a defier or a never-taker.

Non-compliance: example from Kaufman et al. [2]

	All	Complier ($C = 1$)	Never taker ($C = 0$)
Voucher offered ($Z = 1$)	Voucher group $P(Y = 1\|Z = 1)$ observed	$P(Y = 1\|Z = 1, C = 1)$ observed vocher offered and accepted ($A = 1$)	$P(Y = 1\|Z = 1, C = 0)$ observed vocher offered & declined ($A = 0$)
Voucher not offered ($Z = 0$)	control group $P(Y = 1\|Z = 0)$ observed	$P(Y = 1\|Z = 0, C = 1)$ not observed voucher not offered, did not move ($A = 0$); but would have, if possible	$P(Y = 1\|Z = 0, C = 0)$ not observed voucher not offered, did not move ($A = 0$); and would not have, even if possible

No always-takers in this example. Why?; no defiers assumed.

Effect measures under non-compliance

causal?	effect measure	often estimated through
✓	Intention-to-treat (ITT)	$P(Y = 1\|Z = 1) - P(Y = 1\|Z = 0)$
?	As treated	$P(Y = 1\|A = 1) - P(Y = 1\|A = 0)$
?	Per protocol	$P(Y = 1\|A = 1, Z = 1) - P(Y = 1\|A = 0, Z = 0)$
✓	Complier average causal effect (CACE, LACE)	$\{P(Y = 1\|Z = 1) - P(Y = 1\|Z = 0)\}/$ $P(A = 1\|Z = 1) - P(A = 1\|Z = 0)$

Intention-to-treat effect

- Effect identified, easy-to-estimate and causal. Why?
- May be relevant for social interventions or policy changes:

  Example: What would happen if we implemented the voucher policy for everyone; that is, what is the maximum potential of this intervention?

- ITT = “effectiveness of treatment”
- But remember: not the effect of A on Y!
- May not be useful for evaluating the safety of drugs. Why?

As treated

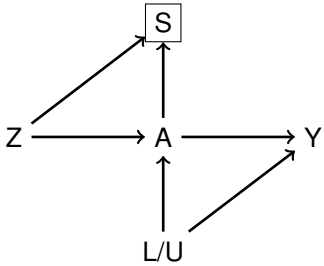
1. Randomisation is not valid anymore if there are common causes of A and Y.
2. Thus, we essentially have observational data again, which means that:
   - Estimation based on the expression $P(Y = 1|A = 1) - P(Y = 1|A = 0)$ is *not* valid to estimate the effect of A on Y.
   - We need to adjust for those common common causes, using the methods from chapter 5 onwards – to obtain a valid causal estimate[1].
   - If we have unmeasured common causes, the naive as-treated effect is definitely non-causal.

Example: The effect of actual housing on asthma may not be identified if there are unmeasured common causes of both; for example, factors related to the socio-economic context.

[1] but consider the reflections of chapter 7 on the implied estimand, depending on the estimation method used

Per protocol

- A naive per-protocol analysis only includes people who followed the protocol, i.e. for which $A = Z$ holds
- This may create selection bias:



Per Protocol

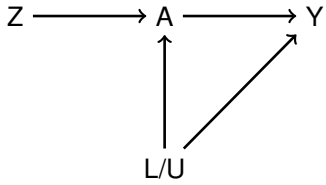
Example: Randomization of colonoscopy (Z). Assumption: actual colonoscopy ($A = 1$) does not affect colon cancer (Y). Among those assigned to $Z = 1$, men with a family history of colon cancer $L = 1$ are more liekly to adhere to the protocol and have a colonoscopy. Despite an actual null effect, we will find a higher risk of death among those with $Z = 1$ because this group has men with a family history of colon cancer overrepresented! [3]

Complier Average Causal Effect (CACE)

As the name implies, the complier average causal effect is defined as:

$$E(Y^{A=1} - Y^{A=0} | A^{Z=1} = 1, A^{Z=0} = 0)$$

This effect can be identified and estimated under our setup, and other assumptions we are going to discuss:



Complier Average Causal Effect (II)

First, consider the ITT and its decomposition with respect to the 4 groups:

$$
\begin{aligned}
E(Y^{Z=1} - Y^{Z=0}) \quad &= \quad E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 1, A^{Z=0} = 1) \times P(A^{Z=1} = 1, A^{Z=0} = 1) \\
&+ \quad E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 0, A^{Z=0} = 0) \times P(A^{Z=1} = 0, A^{Z=0} = 0) \\
&+ \quad E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 0, A^{Z=0} = 1) \times P(A^{Z=1} = 0, A^{Z=0} = 1) \\
&+ \quad E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 1, A^{Z=0} = 0) \times P(A^{Z=1} = 1, A^{Z=0} = 0)
\end{aligned}
$$

First, consider the ITT and its decomposition with respect to the 4 groups:

$$
\begin{aligned}
E(Y^{Z=1} - Y^{Z=0}) &= E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 1, A^{Z=0} = 1) \times P(A^{Z=1} = 1, A^{Z=0} = 1) \\
&+ E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 0, A^{Z=0} = 0) \times P(A^{Z=1} = 0, A^{Z=0} = 0) \\
&+ E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 0, A^{Z=0} = 1) \times P(A^{Z=1} = 0, A^{Z=0} = 1) \\
&+ E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 1, A^{Z=0} = 0) \times P(A^{Z=1} = 1, A^{Z=0} = 0)
\end{aligned}
$$

Second, the intention-to-treat effect in both the always-takers and the never-takers is zero, because

1. Z does *not* affect A in these two strata
2. there is no direct effect of Z on Y

First, consider the ITT and its decomposition with respect to the 4 groups:

$$
\begin{aligned}
E(Y^{Z=1} - Y^{Z=0}) &= \cancel{E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 1, A^{Z=0} = 1) \times P(A^{Z=1} = 1, A^{Z=0} = 1)} \\
&+ \cancel{E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 0, A^{Z=0} = 0) \times P(A^{Z=1} = 0, A^{Z=0} = 0)} \\
&+ E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 0, A^{Z=0} = 1) \times P(A^{Z=1} = 0, A^{Z=0} = 1) \\
&+ E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 1, A^{Z=0} = 0) \times P(A^{Z=1} = 1, A^{Z=0} = 0)
\end{aligned}
$$

Second, the intention-to-treat effect in both the always-takers and the never-takers is zero, because

1. Z does *not* affect A in these two strata
2. there is no direct effect of Z on Y

Complier Average Causal Effect (III)

Third, if we *additionally assume* that there are no defiers, we have:

$$E(Y^{Z=1} - Y^{Z=0}) \quad = \quad E(Y^{Z=1} - Y^{Z=0} \mid A^{Z=1} = 1, A^{Z=0} = 0) \times P(A^{Z=1} = 1, A^{Z=0} = 0)$$

In a randomized study, this may be a reasonable assumption because we may not expect that some individuals will provide consent for participation in a trial with the intention to do exactly the opposite of what they are asked to do.

Now, we are left with the effect in the compliers. Since $A = Z$ in this group, we get:

$$E(Y^{Z=1} - Y^{Z=0}) \quad = \quad E(Y^{A=1} - Y^{A=0} \mid A^{Z=1} = 1, A^{Z=0} = 0) \times P(A^{Z=1} = 1, A^{Z=0} = 0)\,,$$

and thus

$$E(Y^{A=1} - Y^{A=0} \mid A^{Z=1} = 1, A^{Z=0} = 0) \quad = \quad \frac{E(Y^{Z=1} - Y^{Z=0})}{P(A^{Z=1} = 1, A^{Z=0} = 0)}\,.$$

Complier Average Causal Effect (IV)

Now, if we

1. use consistency
2. observe, that the proportion of compliers (denominator) is 1 minus non-compliers (which excludes defiers):

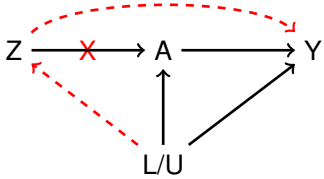
$$\begin{aligned} & 1 - P(A = 0 \mid Z = 1) - P(A = 1 \mid Z = 0) \\ = \; & 1 - P(A = 0 \mid Z = 1) - [1 - P(A = 1 \mid Z = 1)] \\ = \; & P(A = 1 \mid Z = 1) - P(A = 1 \mid Z = 0)\,, \end{aligned}$$

then we get as an expression for the CACE:

$$\frac{E(Y \mid Z = 1) - E(Y \mid Z = 0)}{P(A = 1|Z = 1) - P(A = 1 \mid Z = 0)} = \frac{\text{ITT}}{\text{proportion compliers}}$$

Complier Average Causal Effect (V)

- under perfect compliance: ITT = CACE; otherwise CACE $>$ ITT
- CACE may not always be useful, e.g. if the proportion of compliers is small
- The assumption of “no defiers” is equivalent to *monotinicity*
- If DAG is different, the results for CACE are invalid:



Why is the DAG likely correct in randomized studies?

Reasons for missing data

- Data not captured (administrative)
- Failed experiments
- Drop-out of (longitudinal) study:
  - Problems with treatment
  - No interest, not in the mood
  - Too sick
- No answer in suveys
  - “don’t know”
- ....

Notation

Let $D = \{\mathbf{L}, A, Y\}$ denote the stacked data vector.

C is a vector of the same size with $C_i = 1$ if D_i is observed and $C_i = 0$ otherwise.

The vector C partitions D into the two subsets of observed and unobserved data: D^{obs} with obs $= \{i : c_i = 1\}$ and D^{mis} with mis $= \{i : c_i = 0\}$.

Missing completely at random, MCAR:

$$P(C = c|D^{\text{obs}} = d^{\text{obs}}, D^{\text{mis}} = d^{\text{mis}}; \xi) = P(C = c|\xi) \qquad \forall \xi, c, d^{\text{obs}}, d^{\text{mis}}$$

Probability of missingness depends neither on observed nor unobserved data.

Example: 10% of BMI values are missing because someone forgot to capture them: p(C)=10%.

Missing at random, MAR:

$$P(C = c | D^{\text{obs}} = d^{\text{obs}}, D^{\text{mis}} = d^{\text{mis}}; \xi) = P(C = c | D^{\text{obs}} = d^{\text{obs}}; \xi) \qquad \forall \xi, c, d^{\text{obs}}, d^{\text{mis}}$$

Probability of missingness depends on observed , but not unobserved data.

Example: The probability to be screened for some disease depends on age, which is captured and hence p(C) = f(age).

Missing not random, MNAR:

$$P(C = c | D^{\text{obs}} = d^{\text{obs}}, D^{\text{mis}} = d^{\text{mis}}; \xi) \neq \quad P(C = c | D^{\text{obs}} = d^{\text{obs}}, D^{\text{mis}} = d^{\text{mis}}_*; \xi)$$

$$\text{for some} \quad \xi, c, d^{\text{obs}} \quad \text{and} \quad d^{\text{mis}} \neq d^{\text{mis}}_*$$

Probability of missingness depends on unobserved data.

Example:

- Salary not disclosed if high or low income
- Mortality is higher among those who drop out of a study (too sick, too healthy)
- Unmeasured variable determines missingness probability

Some basic thoughts

- Under MCAR, randomization still valid:
  - Using the complete cases, i.e. those observations for which $C_i = 1$, means that consistent estimates remain consistent
- Under MAR and MNAR we can't assume that randomization still worked and we get causal estimates from the randomized study (without making additional assumptions or using appropriate methods).
- With MAR, often weighting and multiple imputation are viable options, though reflections on what a valid adjustment is requires attention.

Some reflections

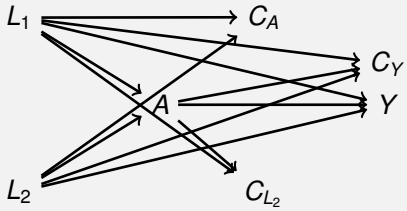
- If multiple variables have missing values simultaneously, it is difficult to practically assess the plausibility of the MAR, MCAR, and MNAR assumptions.
- Whether a desired (causal) target parameter can be estimated, using the complete cases or differently, can not necessarily be determined with the traditional M(N)AR framework alone.
- *Why* are values missing? $\rightarrow$ causal concept
- Bad news: newest research shows that recoverability of a target parameter needs to be ascertained mathematically on a case-by-case basis.

Missingness DAGs

Definition

A missingness DAG (m-DAG) is a DAG that includes all missingness indicator variables $\mathbf{C} = \{C_1, C_2, \ldots\}$. The m-DAG describes the assumptions about the data-generating process and the assumed causes of missingness.

Example:



Definition

Let $\mathbf{L} = \{\mathbf{L}_1, \mathbf{L}_2\}$, where $\mathbf{L}_1$ contains the fully observed variables and $\mathbf{L}_2$ those that are partially observed and contain missing data. Data $(\mathbf{L}, A, Y, \mathbf{C}, U = \emptyset)$ are said to be G-MCAR if $\mathbf{L}, A, Y \coprod \mathbf{C}$. G-MAR is fulfilled if the condition $\mathbf{L}_1, A, Y \coprod \mathbf{C} | \mathbf{L}_2$ is met. The data are defined to be G-MNAR if they are not G-MCAR or G-MAR.

If there is no arrow between $\{\mathbf{L}, A, Y\}$ and $\mathbf{C}$, the data are G-MCAR.

If there is i) no arrow between any variable in $\mathbf{C}$ and any variable in $\mathbf{L}_2$ and ii) no path like $\mathbf{C} \leftarrow U \rightarrow \mathbf{L}_1$ the data are G-MAR.

Note: G-MAR is a stronger requirement than MAR.

Missingness DAGs (IV)

We have graphical rules to derive missingness mechanisms.

The application of missingness DAGs is challenging, but just keep in mind that the traditional definitions can sometimes lead us to conclusions which are not ideal.

Example: Suppose we have a block randomized study (randomization based on L_1). We want to estimate the effect of A on Y. Some values of Y may be missing *because of* unmeasured factors. We may be immediately frightened that we are in a MNAR situation, as missingness depends on unobserved variables. However, problems may only occur once we have arrows from Y, A, L_1 to C_1; which we may solve, or not be able to solve, depending on the specific mechanism. The graphical MCAR definition already helps us to see that there is no problem.

We are not going into more details of this complex topic. See Moreno-Betancur et al. [4] for more detailed practical examples.

Comparison: RCT versus observational study

Randomized Study	Observational Study
high internal validity, low external validity	often better external validity
no confounding by design	confounding due to *measured* variables possible
may be expensive	often less expensive
randomization yields causality, but is not always possible: ethically or practically	no randomization
time: follow-up time limited	time: often longer follow-up possible
intervention: includes implicitly trial condition, which may sometimes limit power to detect events[2]	

[2] Example: PREDICT trial [5]; hardly any events happened under any treatment strategy, likely because every kid got a lot of attention due to very regular trial visits.

Defining the estimand in randomized trials

1. eligibility criteria
2. treatment strategies
3. treatment assignment procedure
4. follow-up time / time zero
5. outcome definition
6. causal contrast
7. statistical analysis

How to make estimands from RCTs and observational data comparable

Effect estimates from randomized trials and observational studies might not be directly comparable because of differences in study design, other than randomization, and in data analysis.

Solution: target trial concept, example from Lodi et al. [6]:

		Harmonization		
	Original randomized trial (START) →	**Actual trial (START)**	**Emulated trial (HIV-CAUSAL)**	← **Original observational analysis (HIV-CAUSAL)**
N	**4685**	4676	14595	**17,612**
Eligibility criteria	**AIDS-free, ART-naïve**	AIDS-free, ART-naïve		**AIDS-free, ART-naïve**
	Age ≥18	Age ≥18		**Age ≥18**
	Two CD4 count>500 cells/mm^3 >14 days apart within 60 days of each other	Two CD4 counts>500 cells/mm^3 >14 days apart and one HIV-RNA measurement		**One CD4 count>500 cells/mm^3 and one HIV-RNA measurement within 3 months of each other**
		within 60 days of each other	within 90 days of each others	
	HIV care in high-, mid- or low-income countries	HIV care in high-, mid- or low-income countries	HIV care in high-income countries and Brazil	**HIV care in high-income countries**
	In 2009-2013	In 2009-2013	In 2005-2013	**In 2000-2013**

Example (II)

Treatment strategies	**1. Immediate ART initiation (grace period not specified) after treatment assignment** **2. Deferred ART initiation at second CD4 count<350 cells/mm^3 or AIDS (grace period not specified)**	1. Immediate ART initiation within 1 month of treatment assignment 2. Deferred ART initiation within 1 month of second CD4 count<350 cells/mm^3 or AIDS		**1. Immediate ART initiation within 6 months of treatment assignment** **2. Deferred ART initiation within 6 months of first CD4 count<350 cells/mm^3 or AIDS**
Randomized assignment	**Physical randomization**	Physical randomization	Randomization assumed within levels of covariates	**Randomization assumed within levels of covariates**

We could formalize the intervention definition (complex here) [3]

In observational data we can not randomize; but we make assumptions what variables we need to have measured to achieve some sort of conditional randomization

[3] We have 1 static and 1 dynamic intervention:

$$\bar{d}_t^1 = \{a_t = 1 \quad \text{for} \quad \forall t$$

$$\bar{d}_t^2 = \begin{cases} a_t = 1 & \text{if CD4 count}_t^{\bar{d}_t} < 350 \text{ (and any CD4 count}_{t^*}^{\bar{d}_{t^*}} < 350 \quad \text{with } t^* < t) \quad \text{or} \quad \text{AIDS}_t^{\bar{d}_t} = 1 \\ & \quad \text{or} \quad a_{t-1} = 1 \\ a_t = 0 & \text{otherwise} \end{cases}$$

Example (III)

Follow-up	**Time zero: strategy assignment**	Time zero: strategy assignment	**Time zero: strategy assignment**
	End of follow-up: outcome, loss to follow-up, 5 years after time zero	End of follow-up: outcome, loss to follow-up, 5 years after time zero	**End of follow-up: outcome, loss to follow-up, 7 years after time zero**
	Loss to follow-up: last date known alive	Loss to follow-up: 12 months without a CD4 cell count or HIV-RNA measurement	**Loss to follow-up: 12 months without a CD4 cell count or HIV-RNA measurement**
Outcome	**Serious AIDS event, serious non-AIDS event, or death**	AIDS event or death	**AIDS event or death**
	AIDS definition: including Hodgkin lymphoma and herpes zoster and excluding candidiasis and herpes simplex virus	AIDS definition: European definition excluding Hodgkin lymphoma and herpes zoster and including candidiasis and herpes simplex virus	**AIDS definition: European definition excluding Hodgkin lymphoma and herpes zoster and including candidiasis and herpes simplex virus**
	Confirmed and probable events after adjudication	All events	**All events**

Example (IV)

Causal contrast	**Intention-to-treat effect**	Per-protocol effect	**Per-protocol effect**
Statistical analysis	**Cox model**	Parametric g-formula	**Parametric g-formula**
	Main effect measure: hazard ratio	Main effect measure: 5-year risk difference	**Main effect measure: 7-year risk difference**

Per-protocol effect means here: the effect that would have been observed under perfect adherence to the trial protocol, i.e. the non-naive per-protcol effect obtained through hypothetically intervening on adherence

Estimation method? Next chapter!

Generalizability: the task of generalizing empirical findings (causal effects) to new environments, settings or populations (often called “external validity”)

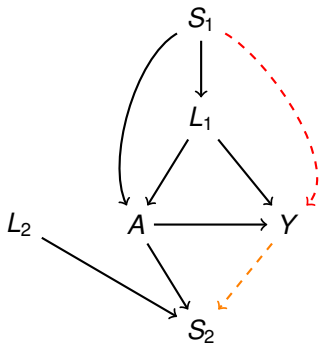
Transportability: the specific generalization task of *transporting* results from a randomized study to a target population, of which we have some data measured through an observational study.

Recoverability: means generally whether a target quantity can be estimated consistently from the available data. Sometimes used specifically for recoverability from *selection bias*, which is a special form of generalizability from the selected population to the full population.

Data Fusion: the most general form of evaluating identification when combining multiple data sources, including randomized studies, observational data, data with selection bias and others.

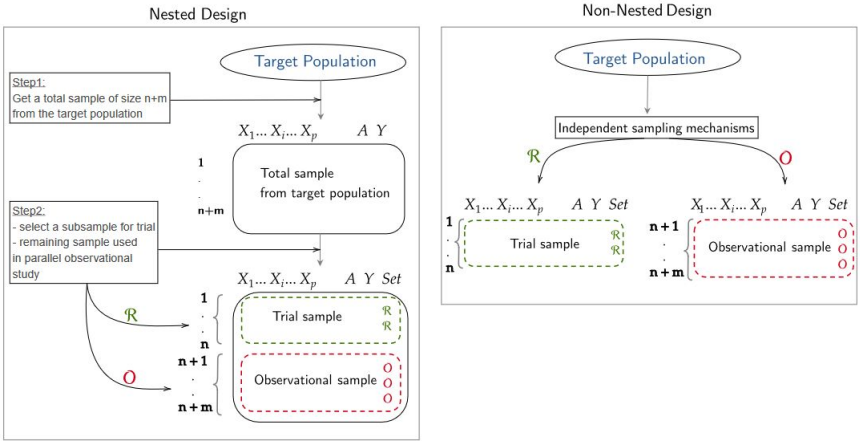
Transportability and Selection Diagrams

- For *transportability* we add selection indicators (S_1) to the DAG. $S_1 = 1$ denotes belonging to population 1 (i.e. the RCT population).
- Transportability selection indicators only have *outgoing* arrows. They describe with respect to which variables (e.g., L_1) the two populations differ.
- In contrast, selection through preferential selection of units for data analysis is described by indicators with *incoming* arrows (S_2). Think also of RCT inclusion and exclusion criteria.
- If we can deal with selection bias based on sample selection requires the so-called *selection back-door criterion* [7], which we do not discuss[a].



[a] obviously, though, the orange arrow is a problem as you might imagine...

Transportability: Designs and Setup[4]



[4] from Colnet et al. [8]

Assumptions for identification (non-nested design)

First, we need *internal validity* in the randomized study:

1. Consistency in the trial population ($S = 1$)
2. (Conditional) exchangeability: $Y^a \coprod A \mid S = 1, L \ \forall a$
3. Positivity of trial participation (in addition to standard positivity): $P(S = 1|L = l) > 0 \ \forall l$

Second, we look at the following assumption:

4. Ignorability on trial participation: $Y^a \coprod S \mid L \ \forall a$

where L are all shifted pre-intervention covariates for which the two populations — trial and target — do not follow the same distribution (and are predictive of Y).

Theorem: Transport Formula

Consider the non-nested setup described above. Under Assumptions (1)–(4) above, it holds that the causal effect of A on Y in the target population can be identified through:

$$P(Y^{A=a}) = \sum_{l} \underbrace{P(Y|A=a, L=l, S=1)}_{\text{RCT}} \times \underbrace{P(L=l)}_{\text{Target Pop.}} \quad (1)$$

Note: the above formula is not necessarily valid if L contains post-treatment variables. In this case, one either has to derive other formulae (using the concept of *S-admissibility*), or transporting the effect may not be possible.

Note: The target population may refer to either all trial eligible people or only those from the observational data ($S = 0$).

Proof

$$E(Y^a)$$

Proof

$$
\begin{aligned}
& E(Y^a) \\
= \quad & E(E(Y^a \mid L))
\end{aligned}
$$

Proof

$$
\begin{aligned}
& E(Y^a) \\
= \quad & E(E(Y^a \mid L)) \\
\overset{?}{=} \quad & E(E(Y^a \mid L, S = 1))
\end{aligned}
$$

Proof

$$
\begin{aligned}
& E(Y^a) \\
= \quad & E(E(Y^a \mid L)) \\
\overset{\text{trial ign.}}{=} \quad & E(E(Y^a \mid L, S = 1))
\end{aligned}
$$

Proof

$$
\begin{aligned}
& E(Y^a) \\
= \quad & E(E(Y^a \mid L)) \\
\overset{\text{trial ign.}}{=} \quad & E(E(Y^a \mid L, S = 1)) \\
\overset{?}{=} \quad & E(E(Y^a \mid L, S = 1, A = a))
\end{aligned}
$$

Proof

$$
\begin{aligned}
& E(Y^a) \\
= \quad & E(E(Y^a \mid L)) \\
\overset{\text{trial ign.}}{=} \quad & E(E(Y^a \mid L, S = 1)) \\
\overset{\text{cond. exch.}}{=} \quad & E(E(Y^a \mid L, S = 1, A = a))
\end{aligned}
$$

Proof

$$
\begin{aligned}
& E(Y^a) \\
= \quad & E(E(Y^a \mid L)) \\
\overset{\text{trial ign.}}{=} \quad & E(E(Y^a \mid L, S = 1)) \\
\overset{\text{cond. exch.}}{=} \quad & E(E(Y^a \mid L, S = 1, A = a)) \\
\overset{?}{=} \quad & E(E(Y \mid L, S = 1, A = a))
\end{aligned}
$$

Proof

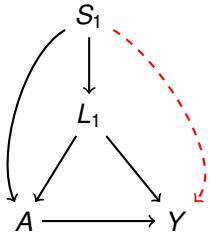
$$
\begin{aligned}
& E(Y^a) \\
= \quad & E(E(Y^a \mid L)) \\
\overset{\text{trial ign.}}{=} \quad & E(E(Y^a \mid L, S = 1)) \\
\overset{\text{cond. exch.}}{=} \quad & E(E(Y^a \mid L, S = 1, A = a)) \\
\overset{\text{cons.}}{=} \quad & E(E(Y \mid L, S = 1, A = a))
\end{aligned}
$$

Proof

$$
\begin{aligned}
& E(Y^a) \\
= \quad & E(E(Y^a \mid L)) \\
\overset{\text{trial ign.}}{=} \quad & E(E(Y^a \mid L, S = 1)) \\
\overset{\text{cond. exch.}}{=} \quad & E(E(Y^a \mid L, S = 1, A = a)) \\
\overset{\text{cons.}}{=} \quad & E(E(Y \mid L, S = 1, A = a)) \\
= \quad & \sum_l E(Y \mid L = l, S = 1, A = a) \times P(L = l)
\end{aligned}
$$

Example

- Under all black arrows the effect of A on Y is transportable
- With the red arrow, the effect is not transportable because ignorability on trial participation is not achieved. Can you see?
- If we had an intermediate variable L_2 with $A \to L_2 \to Y$, and $S_1 \to L_2$, then we have a problem: L_2 is a post-intervention variable which we need for transportation, but we can't estimate the effect in the trial as L_2 is a mediator; thus the transport formula needs modification (or cannot be established in other examples).



Summary

- Randomization helps us to achieve causality by design!
- However, in some situations implementating randomization may not be possible: practically or ethically
- Practical problems may (partially) destroy randomization benefits:
  - non-compliance
  - missing data

  $\rightarrow$ a careful analytical approach is thus needed
- Also: the population eligible for a trial may sometimes not be identical to the target population
  - harmonize observational and trial data analysis for *valid* comparisons
  - reflect on transportability (and interpretations, as discussed in next chapter)

Bibliography

[1] A. M. Henao-Restrepo, I. M. Longini, M. Egger, N. E. Dean, W. J. Edmunds, et al.
Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial.
Lancet, 386(9996):857–66, 2015.

[2] J. S. Kaufman, S. Kaufman, and C. Poole.
Causal inference from randomized trials in social epidemiology.
Social Science & Medicine, 57(12):2397–409, 2003.

[3] M. A. Hernan and S. Hernandez-Diaz.
Beyond the intention-to-treat in comparative effectiveness research.
Clinical Trials, 9(1):48–55, 2012.

[4] Margarita Moreno-Betancur, Katherine J Lee, Finbarr P Leacy, Ian R White, Julie A Simpson, and John B Carlin.
Canonical Causal Diagrams to Guide the Treatment of Missing Data in Epidemiologic Studies.
American Journal of Epidemiology, 187(12):2705–2715, 2018.

[5] T. Puthanakit, V. Saphonn, J. Ananworanich, P. Kosalaraksa, R. Hansudewechakul, U. Vibol, S. J. Kerr, S. Kanjanavanit, C. Ngampiyaskul, J. Wongsawat, W. Luesomboon, N. Ngo-Giang-Huong, K. Chettra, T. Cheunyam, T. Suwarnlerk, S. Ubolyam, W. T. Shearer, R. Paul, L. M. Mofenson, L. Fox, M. G. Law, D. A. Cooper, P. Phanuphak, M. C. Vun, K. Ruxrungtham, and Predict Study Group.
Early versus deferred antiretroviral therapy for children older than 1 year infected with hiv (predict): a multicentre, randomised, open-label trial.
Lancet Infect Dis, 12(12):933–41, 2012.

[6] S. Lodi, A. Phillips, J. Lundgren, R. Logan, S. Sharma, S. R. Cole, A. Babiker, M. Law, H. Chu, D. Byrne, A. Horban, J. A. C. Sterne, K. Porter, C. Sabin, D. Costagliola, S. Abgrall, J. Gill, G. Touloumi, A. G. Pacheco, A. van Sighem, P. Reiss, H. C. Bucher, A. Montoliu Gimenez, I. Jarrin, L. Wittkop, L. Meyer, S. Perez-Hoyos, A. Justice, J. D. Neaton, and M. A. Hernan.
Effect estimates in randomized trials and observational studies: comparing apples with apples.
American Journal of Epidemiology, 188:1569.

[7] Elias Bareinboim, Jin Tian, and Judea Pearl.
Recovering from selection bias in causal and statistical inference.
Proceedings of the AAAI Conference on Artificial Intelligence, 28(1), 2014.

[8] Bénédicte Colnet, Imke Mayer, Guanhua Chen, Awa Dieng, Ruohong Li, Gaël Varoquaux, Jean-Philippe Vert, Julie Josse, and Shu Yang.
Causal inference methods for combining randomized trials and observational studies: a review.
https://arxiv.org/pdf/2011.08047.pdf, 2023.