# The role of causal inference in clinical trials: an introduction

Kelly Van Lancker

UNIVERSITEIT GENT

FACULTY OF SCIENCES

# Acknowledgements


- Stijn Vansteelandt and Oliver Dukes (Ghent University)
- Hege Michiels (argenx)
- Mouna Akacha, Tianmeng Lyu and Jiarui Lu (Novartis)

# Outline

1 Introduction

2 Counterfactuals and Causal Estimands

3 Identification Assumptions
- The consistency assumption
- Exchangeability
- Positivity

4 Examples in RCTs
- Treatment-Policy Estimand
- A Hypothetical Estimand

# Chocolate Consumption and Nobel Prizes

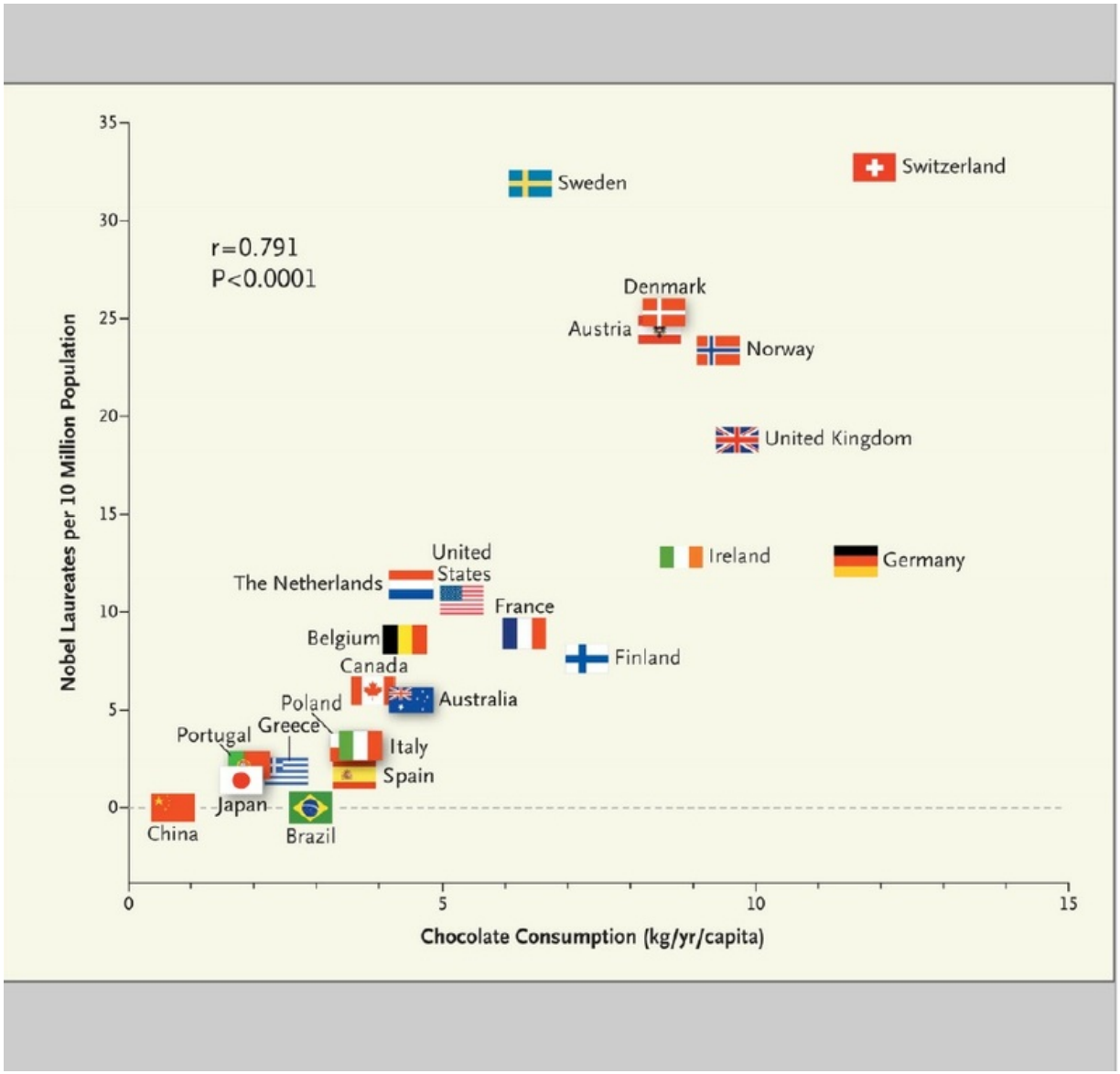
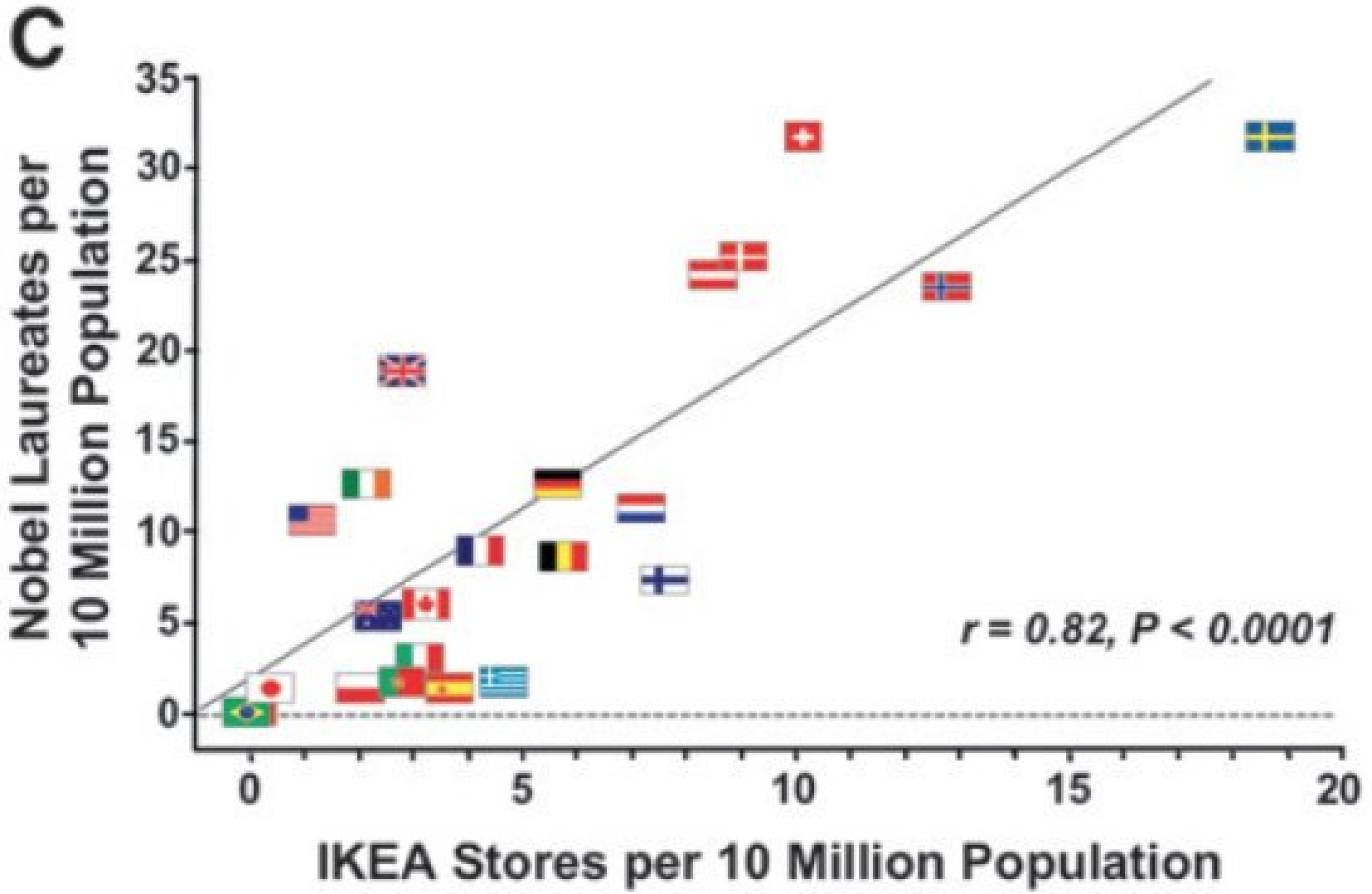



**Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

# What about visiting IKEA?



# Causal Inference

- Do they sound ridiculous? How about:
  "I took an aspirin and my headache went away - the drug worked!"

# Causal Inference

- Do they sound ridiculous? How about:
  "I took an aspirin and my headache went away - the drug worked!"
- In this talk, we will develop insight by explicitly **distinguishing association from causation**.
  - What is the effect of consuming more chocolate?

# Causal Inference

- Do they sound ridiculous? How about:
  "I took an aspirin and my headache went away - the drug worked!"
- In this talk, we will develop insight by explicitly **distinguishing association from causation**.
  - What is the effect of consuming more chocolate?
- So the question is: what are we actually **trying to estimate**, and when does association imply causation?

# Estimands first!

- In this talk, we will introduce measures of causal effect, so-called **causal estimands**.
- This is the **first step in a causal analysis**.

# Estimands first!

- In this talk, we will introduce measures of causal effect, so-called **causal estimands**.
- This is the **first step in a causal analysis**.
- This may sound obvious, but it is not.
  - For a statistician / data scientist, the first step is often formulating a model / algorithm.

# A wrong first step

- ICH E9 (FDA and EMA, 1998) and EMA (2015) guidelines are written with the understanding that the **target treatment effect is a model parameter**; e.g.,

$$g\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X$$

where $g(\cdot)$ is a pre-specified link function.

# A wrong first step

- ICH E9 (FDA and EMA, 1998) and EMA (2015) guidelines are written with the understanding that the **target treatment effect is a model parameter**; e.g.,

$$g\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X$$

  where $g(\cdot)$ is a pre-specified link function.

- This model implies no interaction between $Z$ and $X$:
  - A **statistical modelling assumption**, **not implied by randomization**.
  - When the model is misspecified, the standard likelihood-based estimators of $\beta_1$ **may not generally target a causal effect**.

# Why estimands first?

# Why estimands first?

- Causal estimands **translate the scientific question into a quantity that we can (hopefully) communicate well** to clinicians/investigators/....
- Models and algorithms are only tools to learn an estimand, but should never be the primary aim of a causal analysis.

# Why estimands first?

- Causal estimands **translate the scientific question into a quantity that we can (hopefully) communicate well** to clinicians/investigators/....
- Models and algorithms are only tools to learn an estimand, but should never be the primary aim of a causal analysis.
- In this talk, we will introduce popular causal estimands and study how to **identify** them from data.

# Causal inference road map

- Road map for inferring causal effects
  1. Defining the **estimand**
  2. Stating the **identification assumptions**
  3. **Estimation** method(s) along with statistical assumptions

# Causal inference road map

- Road map for inferring causal effects
  1. Defining the **estimand**
  2. Stating the **identification assumptions**
  3. **Estimation** method(s) along with statistical assumptions
- Despite the many positive steps, statisticians often tend to go straight to Step 3.
- In my opinion, we should strive to fully follow this road map to achieve the most benefits.

# Outline


1 Introduction

2 Counterfactuals and Causal Estimands

3 Identification Assumptions
- The consistency assumption
- Exchangeability
- Positivity

4 Examples in RCTs
- Treatment-Policy Estimand
- A Hypothetical Estimand

# A simple example

- 18 subjects each suffer a **headache**.
- Some take a **potion**; others don't.
- One hour later, we ask each of the 18 whether or not his/her headache has **disappeared**.

# The observed data (1)

| | $Z$ (potion taken?) | $Y$ (headache disappeared?) |
|---|---|---|
| Fay | 0 | 1 |
| George | 0 | 1 |
| Tom | 0 | 1 |
| Mary | 0 | 1 |
| Chris | 0 | 0 |
| Anna | 0 | 0 |
| Rose | 1 | 1 |
| Jack | 1 | 1 |
| Lee | 1 | 1 |
| Adam | 0 | 1 |
| John | 0 | 0 |
| Ian | 0 | 0 |
| Betsy | 1 | 1 |
| Claus | 1 | 1 |
| Sara | 1 | 1 |
| Lisa | 1 | 1 |
| Peter | 1 | 0 |
| Sue | 1 | 0 |

# The observed data (2)

| | $Z$ (potion taken?) | $Y$ (headache disappeared?) |
|---|---|---|
| Fay | 0 | 1 |
| George | 0 | 1 |
| Tom | 0 | 1 |
| Mary | 0 | 1 |
| Chris | 0 | 0 |
| Anna | 0 | 0 |
| Rose | 1 | 1 |
| Jack | 1 | 1 |
| Lee | 1 | 1 |
| Adam | 0 | 1 |
| John | 0 | 0 |
| Ian | 0 | 0 |
| Betsy | 1 | 1 |
| Claus | 1 | 1 |
| Sara | 1 | 1 |
| Lisa | 1 | 1 |
| Peter | 1 | 0 |
| Sue | 1 | 0 |

- Sara took the potion, and her headache disappeared.
- Did the potion **cause** her headache to disappear?

# The observed data (2)

| | $Z$ (potion taken?) | $Y$ (headache disappeared?) |
|---|---|---|
| Fay | 0 | 1 |
| George | 0 | 1 |
| Tom | 0 | 1 |
| Mary | 0 | 1 |
| Chris | 0 | 0 |
| Anna | 0 | 0 |
| Rose | 1 | 1 |
| Jack | 1 | 1 |
| Lee | 1 | 1 |
| Adam | 0 | 1 |
| John | 0 | 0 |
| Ian | 0 | 0 |
| Betsy | 1 | 1 |
| Claus | 1 | 1 |
| Sara | 1 | 1 |
| Lisa | 1 | 1 |
| Peter | 1 | 0 |
| Sue | 1 | 0 |

- Sara took the potion, and her headache disappeared.
- Did the potion **cause** her headache to disappear?
- We don't know.

# The observed data (2)

| | $Z$ (potion taken?) | $Y$ (headache disappeared?) |
|---|---|---|
| Fay | 0 | 1 |
| George | 0 | 1 |
| Tom | 0 | 1 |
| Mary | 0 | 1 |
| Chris | 0 | 0 |
| Anna | 0 | 0 |
| Rose | 1 | 1 |
| Jack | 1 | 1 |
| Lee | 1 | 1 |
| Adam | 0 | 1 |
| John | 0 | 0 |
| Ian | 0 | 0 |
| Betsy | 1 | 1 |
| Claus | 1 | 1 |
| Sara | 1 | 1 |
| Lisa | 1 | 1 |
| Peter | 1 | 0 |
| Sue | 1 | 0 |

- Sara took the potion, and her headache disappeared.
- Did the potion **cause** her headache to disappear?
- We don't know.
- To answer this, we need to know what **would** have happened **had she not** taken the potion.

# Counterfactuals and potential outcomes

- Write $Y^0$ and $Y^1$ to represent the **potential outcomes** under both treatments.[1]
  - $Y^0$ is the outcome which would have been seen **had the potion NOT been taken**.
  - $Y^1$ is the outcome which would have been seen **had the potion been taken**.
- One of these is observed: if $Z = 0$, $Y^0$ is observed; if $Z = 1$, $Y^1$ is observed.
- The other is **counterfactual**.
- Suppose that we can observe the unobservable...

[1] Some use $Y(0)$ and $Y(1)$

# The ideal data

| | $Y^1$ | $Y^0$ |
|---|---|---|
| Fay | 1 | 1 |
| George | 1 | 1 |
| Tom | 1 | 1 |
| Mary | 1 | 1 |
| Chris | 1 | 0 |
| Anna | 1 | 0 |
| Rose | 1 | 1 |
| Jack | 1 | 1 |
| Lee | 1 | 0 |
| Adam | 1 | 1 |
| John | 1 | 0 |
| Ian | 0 | 0 |
| Betsy | 1 | 1 |
| Claus | 1 | 1 |
| Sara | 1 | 0 |
| Lisa | 1 | 0 |
| Peter | 0 | 0 |
| Sue | 0 | 0 |

- For Sara, the potion **did** have a causal effect.
- She did take it, and her headache disappeared;
  but **had she not taken it**,
  her headache **would not** have disappeared.
- Thus the potion had a causal effect on her headache.
- What about Fay?
- and Chris?
- and Ian?

# The fundamental problem of causal inference

Back to reality...

| | $Z$ | $Y$ | $Y^1$ | $Y^0$ |
|---|---|---|---|---|
| Fay | 0 | 1 | ? | 1 |
| George | 0 | 1 | ? | 1 |
| Tom | 0 | 1 | ? | 1 |
| Mary | 0 | 1 | ? | 1 |
| Chris | 0 | 0 | ? | 0 |
| Anna | 0 | 0 | ? | 0 |
| Rose | 1 | 1 | 1 | ? |
| Jack | 1 | 1 | 1 | ? |
| Lee | 1 | 1 | 1 | ? |
| Adam | 0 | 1 | ? | 1 |
| John | 0 | 0 | ? | 0 |
| Ian | 0 | 0 | ? | 0 |
| Betsy | 1 | 1 | 1 | ? |
| Claus | 1 | 1 | 1 | ? |
| Sara | 1 | 1 | 1 | ? |
| Lisa | 1 | 1 | 1 | ? |
| Peter | 1 | 0 | 0 | ? |
| Sue | 1 | 0 | 0 | ? |

- In reality, we **never** observe **both** $Y^0$ and $Y^1$ on the same individual.

# The fundamental problem of causal inference

Back to reality...

| | $Z$ | $Y$ | $Y^1$ | $Y^0$ |
|---|---|---|---|---|
| Fay | 0 | 1 | ? | 1 |
| George | 0 | 1 | ? | 1 |
| Tom | 0 | 1 | ? | 1 |
| Mary | 0 | 1 | ? | 1 |
| Chris | 0 | 0 | ? | 0 |
| Anna | 0 | 0 | ? | 0 |
| Rose | 1 | 1 | 1 | ? |
| Jack | 1 | 1 | 1 | ? |
| Lee | 1 | 1 | 1 | ? |
| Adam | 0 | 1 | ? | 1 |
| John | 0 | 0 | ? | 0 |
| Ian | 0 | 0 | ? | 0 |
| Betsy | 1 | 1 | 1 | ? |
| Claus | 1 | 1 | 1 | ? |
| Sara | 1 | 1 | 1 | ? |
| Lisa | 1 | 1 | 1 | ? |
| Peter | 1 | 0 | 0 | ? |
| Sue | 1 | 0 | 0 | ? |

- In reality, we **never** observe **both** $Y^0$ and $Y^1$ on the same individual.
- Sometimes called the **fundamental problem of causal inference**.

# The fundamental problem of causal inference

Back to reality...

| | $Z$ | $Y$ | $Y^1$ | $Y^0$ |
|---|---|---|---|---|
| Fay | 0 | 1 | ? | 1 |
| George | 0 | 1 | ? | 1 |
| Tom | 0 | 1 | ? | 1 |
| Mary | 0 | 1 | ? | 1 |
| Chris | 0 | 0 | ? | 0 |
| Anna | 0 | 0 | ? | 0 |
| Rose | 1 | 1 | 1 | ? |
| Jack | 1 | 1 | 1 | ? |
| Lee | 1 | 1 | 1 | ? |
| Adam | 0 | 1 | ? | 1 |
| John | 0 | 0 | ? | 0 |
| Ian | 0 | 0 | ? | 0 |
| Betsy | 1 | 1 | 1 | ? |
| Claus | 1 | 1 | 1 | ? |
| Sara | 1 | 1 | 1 | ? |
| Lisa | 1 | 1 | 1 | ? |
| Peter | 1 | 0 | 0 | ? |
| Sue | 1 | 0 | 0 | ? |

- In reality, we **never** observe **both** $Y^0$ and $Y^1$ on the same individual.
- Sometimes called the **fundamental problem of causal inference**.
- It is therefore over-ambitious to infer anything about individual-level causal effects.

# Population-level causal effects

- A less ambitious goal is to focus on the **population-level** or **average** causal effect:

$$E\left(Y^1 - Y^0\right) \quad \text{or} \quad \frac{E\left(Y^1\right)}{E\left(Y^0\right)}.$$

- We can also define causal effects in a **subpopulation**, e.g. the treated:

$$E\left(Y^1 - Y^0 \,|Z = 1\right)$$

or, for **precision medicine**, in strata defined by pre-treatment characteristics $X$:

$$E\left(Y^1 - Y^0 \,|X\right)$$

# Summary so far...

- We now have notation
  to **distinguish causation**

$$E\left(Y^1 - Y^0\right) \quad \text{or} \quad \frac{E\left(Y^1\right)}{E\left(Y^0\right)}$$

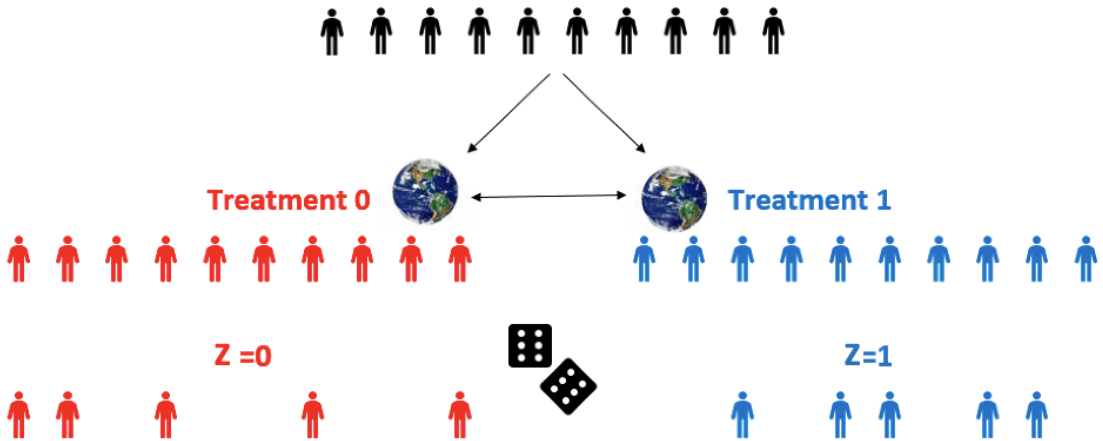
  **from association**:

$$E\left(Y|Z=1\right) - E\left(Y|Z=0\right) \quad \text{or} \quad \frac{E\left(Y|Z=1\right)}{E\left(Y|Z=0\right)}.$$

- Historically, this has been key
  to the development of methods for causal inference.

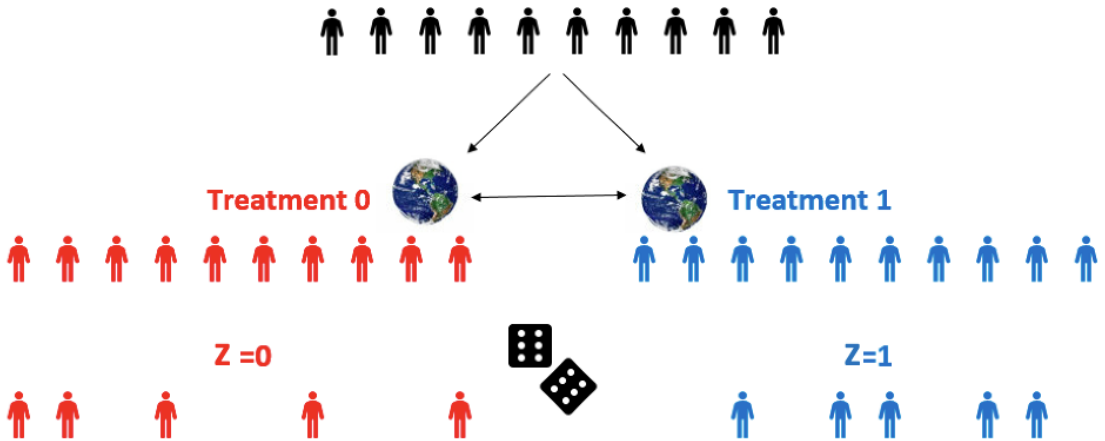
# causation $\neq$ association? (1)

Consider a randomized trial



# causation $\neq$ association? (1)

Consider a randomized trial




- In real life, patients are randomized to only one group.
- Randomization ensures that causal contrasts correspond to statistical contrasts:
  - $E\left(Y^1\right) - E\left(Y^0\right) = E\left(Y|Z=1\right) - E\left(Y|Z=0\right)$.

# causation $\neq$ association? (2)

- So, do we need to care about the question whether causation $\neq$ association in a randomized trial?

# causation $\neq$ association? (2)

- So, do we need to care about the question whether causation $\neq$ association in a randomized trial?
- Yes!

# causation $\neq$ association? (2)

- So, do we need to care about the question whether causation $\neq$ association in a randomized trial?
- Yes!
- Randomization can be broken
  - due to intercurrent events,
  - missing data, or
  - when interest lies in generalizing trial results.

# causation $\neq$ association? (3)

In that case, we are in a similar situation as an observational study...

| Id | $X$ | $Z$ | $Y$ | $Y^1$ | $Y^0$ |
|---|---|---|---|---|---|
| Fay | 0 | 0 | 1 | 1 | 1 |
| George | 0 | 0 | 1 | 1 | 1 |
| Tom | 0 | 0 | 1 | 1 | 1 |
| Mary | 0 | 0 | 1 | 1 | 1 |
| Chris | 0 | 0 | 0 | 1 | 0 |
| Anna | 0 | 0 | 0 | 1 | 0 |
| Rose | 0 | 1 | 1 | 1 | 1 |
| Jack | 0 | 1 | 1 | 1 | 1 |
| Lee | 0 | 1 | 1 | 1 | 0 |
| Adam | 1 | 0 | 1 | 1 | 1 |
| John | 1 | 0 | 0 | 1 | 0 |
| Ian | 1 | 0 | 0 | 0 | 0 |
| Betsy | 1 | 1 | 1 | 1 | 1 |
| Claus | 1 | 1 | 1 | 1 | 1 |
| Sara | 1 | 1 | 1 | 1 | 0 |
| Lisa | 1 | 1 | 1 | 1 | 0 |
| Peter | 1 | 1 | 0 | 0 | 0 |
| Sue | 1 | 1 | 0 | 0 | 0 |

$$\frac{P(Y^1 = 1)}{P(Y^0 = 1)} = \frac{15/18}{9/18} = \frac{5}{3}$$

$$\frac{P(Y = 1|Z = 1)}{P(Y = 1|Z = 0)} = \frac{7/9}{5/9} = \frac{7}{5}$$

# Fundamental problem of causal inference

- Since we don't know $Y^1$ for every subject, we can't easily estimate $E\left(Y^1\right)$ as the proportion of all subjects with $Y^1 = 1$.
- Likewise, we can't simply calculate $E\left(Y^0\right)$ as the proportion of all subjects with $Y^0 = 1$.
- Our task is therefore to choose quantities from the observed data (i.e. involving $Z$, $Y$ and other observed variables) that represent **reasonable substitutes** for hypothetical quantities such as $E\left(Y^1 - Y^0\right)$.[2]

[2] Note that we know how to do this when randomization is not broken in an RCT.

# Outline


1 Introduction

2 Counterfactuals and Causal Estimands

3 Identification Assumptions
- The consistency assumption
- Exchangeability
- Positivity

4 Examples in RCTs
- Treatment-Policy Estimand
- A Hypothetical Estimand

# Outline

1 Introduction

2 Counterfactuals and Causal Estimands

3 Identification Assumptions
- The consistency assumption
- Exchangeability
- Positivity

4 Examples in RCTs
- Treatment-Policy Estimand
- A Hypothetical Estimand

# The consistency assumption

- So far, we have implicitly used that:

$$Z = z \Rightarrow Y^z = Y$$

  in order to link counterfactuals to the observed data.

- This may appear logical, but is nonetheless called an assumption:
  the **consistency assumption**.

- The reason is that we define **'causal effects'**
  as expressing what would happen under **hypothetical interventions**,
  but no interventions may have been considered in the study.

# Example: the effect of weight loss

*Nutrition Research Reviews* (2009), **22**, 93–108
doi:10.1017/S0954422409990035


**A review and meta-analysis of the effect of weight loss on all-cause mortality risk**

Mary Harrington[1], Sigrid Gibson[2] and Richard C. Cottrell[3]*

[1]*The Sugar Bureau, London WC2B 5JJ, UK*
[2]*Sig-Nurture Ltd, Guildford, Surrey GU1 2TF, UK*
[3]*World Sugar Research Organisation, London SW1V 3LX, UK*

*What is meant by the effect of weight loss on mortality?*

*If* $Z = 1$ *means weight loss, then does* $Y^1 = Y$ *for those with* $Z = 1$*?*

# What does weight loss mean? (1)

There are many **different versions** of weight loss:

- Does losing 10 kg of weight prolong life?

# What does weight loss mean? (1)

There are many **different versions** of weight loss:

- Does losing 10 kg of weight prolong life?
- Does losing 10 kg of weight by age 40,
  and maintaining that weight loss between ages 40 and 50
  prolong life?

# What does weight loss mean? (1)

There are many **different versions** of weight loss:

- Does losing 10 kg of weight prolong life?
- Does losing 10 kg of weight by age 40,
  and maintaining that weight loss between ages 40 and 50
  prolong life?
- Does losing 10 kg of weight by age 40,
  and maintaining that weight loss between ages 40 and 50 via
  restricted caloric intake
  prolong life?

# What does weight loss mean? (1)

There are many **different versions** of weight loss:

- Does losing 10 kg of weight prolong life?
- Does losing 10 kg of weight by age 40,
  and maintaining that weight loss between ages 40 and 50
  prolong life?
- Does losing 10 kg of weight by age 40,
  and maintaining that weight loss between ages 40 and 50 via restricted caloric intake
  prolong life?
- Does losing 10 kg of weight by age 40,
  and maintaining that weight loss between ages 40 and 50 via restricted caloric intake
  and physical exercise prolong life?

# What does weight loss mean? (1)

There are many **different versions** of weight loss:

- Does losing 10 kg of weight prolong life?
- Does losing 10 kg of weight by age 40,
  and maintaining that weight loss between ages 40 and 50
  prolong life?
- Does losing 10 kg of weight by age 40,
  and maintaining that weight loss between ages 40 and 50 via restricted caloric intake
  prolong life?
- Does losing 10 kg of weight by age 40,
  and maintaining that weight loss between ages 40 and 50 via restricted caloric intake
  and physical exercise prolong life?
- ...

# What does weight loss mean? (2)

- We could go on forever... and will never be satisfied.
- Quantitative statements such as

  *'Intentional weight loss had a small benefit for individuals with obesity-related risk factors (RR 0.87 (95% CI 0.77, 0.99); P = 0.028) ...'*

  are therefore very **difficult to understand**:
  unclear precisely what intervention on weight is considered.

# What does weight loss mean? (2)

- We could go on forever... and will never be satisfied.
- Quantitative statements such as

  *'Intentional weight loss had a small benefit for individuals with obesity-related risk factors (RR 0.87 (95% CI 0.77, 0.99); P = 0.028) ...'*

  are therefore very **difficult to understand**:
  unclear precisely what intervention on weight is considered.
- If our interest was in quantifying the effect of weight loss via physical exercise,
  the study will not help when participants lost weight via gastric bypass.

# What does weight loss mean? (2)

- We could go on forever... and will never be satisfied.
- Quantitative statements such as

  *'Intentional weight loss had a small benefit for individuals with obesity-related risk factors (RR 0.87 (95% CI 0.77, 0.99); P = 0.028) ...'*

  are therefore very **difficult to understand**:
  unclear precisely what intervention on weight is considered.
- If our interest was in quantifying the effect of weight loss via physical exercise,
  the study will not help when participants lost weight via gastric bypass.
- The **consistency assumption**:

$$Z = z \Rightarrow Y^z = Y$$

  then fails.

# Outline


1 Introduction

2 Counterfactuals and Causal Estimands

3 Identification Assumptions
- The consistency assumption
- Exchangeability
- Positivity

4 Examples in RCTs
- Treatment-Policy Estimand
- A Hypothetical Estimand

# Exchangeability (1)

- What might be a good substitute for $E\left(Y^1\right)$?
- *What about* $\boldsymbol{E}\left(\boldsymbol{Y} \mid \boldsymbol{Z}=\mathbf{1}\right)$*?*
- This is the proportion whose headache disappeared among those who actually took the potion.
- *Is this the same as* $E\left(Y^1\right)$*?*

# Exchangeability (1)

- What might be a good substitute for $E\left(Y^1\right)$?
- *What about* $\boldsymbol{E\left(Y \mid Z=1\right)}$*?*
- This is the proportion whose headache disappeared among those who actually took the potion.
- *Is this the same as* $E\left(Y^1\right)$*?*
- Only if those who took the potion are **exchangeable** with those who didn't.
  Mathematically, $Z \perp\!\!\!\perp Y^0$ and $Z \perp\!\!\!\perp Y^1$.
- This would be the case if the choice to take the potion was made **at random**.
- This is why **ideal randomised experiments** are the **gold standard** for inferring causal effects.

# Exchangeability (2)

**With exchangeability, analyses of randomised experiments return causal effects**

$$E(Y^1 - Y^0) \quad = \quad E(Y^1) - E(Y^0)$$

# Exchangeability (2)

**With exchangeability, analyses of randomised experiments return causal effects**

$$\begin{aligned} E(Y^1 - Y^0) &= E(Y^1) - E(Y^0) \\ &= E(Y^1|Z=1) - E(Y^0|Z=0) \end{aligned}$$

# Exchangeability (2)

**With exchangeability, analyses of randomised experiments return causal effects**

$$
\begin{aligned}
E(Y^1 - Y^0) &= E(Y^1) - E(Y^0) \\
&= E(Y^1|Z=1) - E(Y^0|Z=0) \\
&= E(Y|Z=1) - E(Y|Z=0)
\end{aligned}
$$

The righthand side is obtainable by taking the difference of the mean of the outcomes ($Y$) in each arm ($Z = 1$ and $Z = 0$).

# Conditional exchangeability (1)

- In observational data (or RCT with intercurrent events), exchangeability is usually **implausible**.
- Those with a worse headache
  are probably more likely to take the potion.

# Conditional exchangeability (1)

- In observational data (or RCT with intercurrent events), exchangeability is usually **implausible**.
- Those with a worse headache
  are probably more likely to take the potion.
- Suppose we asked each subject at the beginning of the study:
  "is your headache **severe**?".
- Then, we might propose that,
  **after taking severity into account**,
  the **decision** as to whether or not to take the potion was effectively taken **at random**.

# Conditional exchangeability (2)

- Suppose $X$ denotes severity.
- Then, under this assumption, within strata of $X$, the exposed and unexposed subjects are **exchangeable**.
- This is called **conditional exchangeability** (given $X$). Mathematically, $Z \perp\!\!\!\perp Y^0 | X$ and $Z \perp\!\!\!\perp Y^1 | X$.
- We can't check this from our data; we need to believe it from **a priori knowledge**.

# Conditional exchangeability (3)

**With conditional exchangeability, regression delivers conditional causal effects**

$$E(Y^1 - Y^0|X) \quad = \quad E(Y^1|X) - E(Y^0|X)$$

# Conditional exchangeability (3)

**With conditional exchangeability, regression delivers conditional causal effects**

$$\begin{aligned} E(Y^1 - Y^0|X) &= E(Y^1|X) - E(Y^0|X) \\ &= E(Y^1|Z=1,X) - E(Y^0|Z=0,X) \end{aligned}$$

# Conditional exchangeability (3)

## With conditional exchangeability, regression delivers conditional causal effects

$$\begin{aligned} E(Y^1 - Y^0|X) &= E(Y^1|X) - E(Y^0|X) \\ &= E(Y^1|Z=1,X) - E(Y^0|Z=0,X) \\ &= E(Y|Z=1,X) - E(Y|Z=0,X) \end{aligned}$$

It follows that the **marginal causal effect** equals

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

# Basis for G-computation

**Step 1:** **Model fitting**

Fit a model for $E(Y|Z,X)$

- E.g., $P(Y=1|Z,X) = logit^{-1}(\gamma_0 + \gamma_1 \cdot Z + \gamma_2 \cdot X + \gamma_3 \cdot Z \cdot X)$.

# Basis for G-computation

**Step 1: Model fitting**
Fit a model for $E(Y|Z,X)$

- E.g., $P(Y=1|Z,X) = logit^{-1}(\gamma_0 + \gamma_1 \cdot Z + \gamma_2 \cdot X + \gamma_3 \cdot Z \cdot X)$.

**Step 2: Predicting**
Use this model to impute outcome **under treatment ($Z=1$) and control ($Z=0$)** for **all** patients:

| Id | $X$ | $Z$ | $Y$ | $Y^1$ | $\hat{P}^1$ | $Y^0$ | $\hat{P}^0$ |
|---|---|---|---|---|---|---|---|
| Fay | 0 | 0 | 1 | ? | 0.6 | 1 | 0.5 |
| Rose | 0 | 1 | 1 | 1 | 0.7 | ? | 0.5 |
| Adam | 1 | 0 | 1 | ? | 0.8 | 1 | 0.7 |
| Lisa | 1 | 1 | 1 | 1 | 0.8 | ? | 0.7 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

# Basis for G-computation

**Step 1: Model fitting**

Fit a model for $E(Y|Z,X)$

- E.g., $P(Y=1|Z,X) = logit^{-1}(\gamma_0 + \gamma_1 \cdot Z + \gamma_2 \cdot X + \gamma_3 \cdot Z \cdot X)$.

**Step 2: Predicting**

Use this model to impute outcome **under treatment ($Z=1$) and control ($Z=0$)** for **all** patients:

| Id | $X$ | $Z$ | $Y$ | $Y^1$ | $\hat{P}^1$ | $Y^0$ | $\hat{P}^0$ |
|---|---|---|---|---|---|---|---|
| Fay | 0 | 0 | 1 | ? | 0.6 | 1 | 0.5 |
| Rose | 0 | 1 | 1 | 1 | 0.7 | ? | 0.5 |
| Adam | 1 | 0 | 1 | ? | 0.8 | 1 | 0.7 |
| Lisa | 1 | 1 | 1 | 1 | 0.8 | ? | 0.7 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

**Step 3: Averaging**

Take the average of imputed outcomes, and calculate treatment effect of interest: $\frac{1}{n}\sum_{i=1}^{n} \hat{P}_i^1 - \frac{1}{n}\sum_{i=1}^{n} \hat{P}_i^0$.

# The hazards of hazard ratios

- It is common to measure the effect of a randomized treatment $Z$ on a time-to-event endpoint $T$ in terms of the **hazard ratio**:

$$\frac{P(T = t | T \geq t, Z = 1)}{P(T = t | T \geq t, Z = 0)}$$

# The hazards of hazard ratios

- It is common to measure the effect of a randomized treatment $Z$ on a time-to-event endpoint $T$ in terms of the **hazard ratio**:

$$\frac{P(T = t | T \geq t, Z = 1)}{P(T = t | T \geq t, Z = 0)}$$

- By randomization, we have exchangeability.

# The hazards of hazard ratios

- It is common to measure the effect of a randomized treatment $Z$ on a time-to-event endpoint $T$ in terms of the **hazard ratio**:

$$\frac{P(T = t | T \geq t, Z = 1)}{P(T = t | T \geq t, Z = 0)}$$

- By randomization, we have exchangeability.
- This allows us to re-express the hazard ratio as

$$\frac{P(T^1 = t | T^1 \geq t)}{P(T^0 = t | T^0 \geq t)}$$

# The hazards of hazard ratios

- It is common to measure the effect of a randomized treatment $Z$ on a time-to-event endpoint $T$ in terms of the **hazard ratio**:

$$\frac{P(T = t | T \geq t, Z = 1)}{P(T = t | T \geq t, Z = 0)}$$

- By randomization, we have exchangeability.
- This allows us to re-express the hazard ratio as

$$\frac{P(T^1 = t | T^1 \geq t)}{P(T^0 = t | T^0 \geq t)}$$

- Note that this continues to be an **apple versus orange comparison**,
  except under the null.

# Outline


1 Introduction

2 Counterfactuals and Causal Estimands

3 Identification Assumptions
- The consistency assumption
- Exchangeability
- Positivity

4 Examples in RCTs
- Treatment-Policy Estimand
- A Hypothetical Estimand

# Positivity (1)

## Positivity assumption

Conditional on covariates $X$, there is a **probability greater than zero** of being assigned to each of the treatment levels

$$0 < P(Z = 1|X) < 1 \text{ with probability 1}$$

Here, $X$ is a set of variables that satisfies exchangeability

# Positivity (1)

## Positivity assumption

Conditional on covariates $X$, there is a **probability greater than zero** of being assigned to each of the treatment levels

$$0 < P(Z = 1|X) < 1 \text{ with probability 1}$$

Here, $X$ is a set of variables that satisfies exchangeability

- Important that there is **variability in treatment assignment**

# Positivity (2)

- In **RCTs**, positivity is usually **guaranteed** by design for the randomized treatment of interest.
  - The protocol will specify the assignment mechanism, meaning that $P(Z = 1|X)$ is known.

# Positivity (2)

- In **RCTs**, positivity is usually **guaranteed** by design for the randomized treatment of interest.
  - The protocol will specify the assignment mechanism, meaning that $P(Z = 1|X)$ is known.
- **Not always realistic** in observational studies or RCTs with **intercurrent events** (see later).

# Positivity (2)

- In **RCTs**, positivity is usually **guaranteed** by design for the randomized treatment of interest.
  - The protocol will specify the assignment mechanism, meaning that $P(Z = 1|X)$ is known.
- **Not always realistic** in observational studies or RCTs with **intercurrent events** (see later).
- Even when this assumption holds,
  **unstable estimates** are typically obtained when it is **nearly violated**.

# Outline


1 Introduction

2 Counterfactuals and Causal Estimands

3 Identification Assumptions
- The consistency assumption
- Exchangeability
- Positivity

4 Examples in RCTs
- Treatment-Policy Estimand
- A Hypothetical Estimand

# Trial Setting

- **Population**: Patients with Type 2 diabetes.

# Trial Setting

- **Population**: Patients with Type 2 diabetes.
- **Treatment**: Assignment to active treatment ($Z = 1$) vs placebo ($Z = 0$).

# Trial Setting

- **Population**: Patients with Type 2 diabetes.
- **Treatment**: Assignment to active treatment ($Z = 1$) vs placebo ($Z = 0$).
- **Outcome variable**: Change in HbA1c from baseline to week 36.

# Trial Setting

- **Population**: Patients with Type 2 diabetes.
- **Treatment**: Assignment to active treatment ($Z = 1$) vs placebo ($Z = 0$).
- **Outcome variable**: Change in HbA1c from baseline to week 36.
- **Intercurrent events**: Use of rescue medication.

# Trial Setting

- **Population**: Patients with Type 2 diabetes.
- **Treatment**: Assignment to active treatment ($Z = 1$) vs placebo ($Z = 0$).
- **Outcome variable**: Change in HbA1c from baseline to week 36.
- **Intercurrent events**: Use of rescue medication.
- **Strategy**: Will discuss two settings; treatment policy and hypothetical strategy.

# Trial Setting

- **Population**: Patients with Type 2 diabetes.
- **Treatment**: Assignment to active treatment ($Z = 1$) vs placebo ($Z = 0$).
- **Outcome variable**: Change in HbA1c from baseline to week 36.
- **Intercurrent events**: Use of rescue medication.
- **Strategy**: Will discuss two settings; treatment policy and hypothetical strategy.
- **Population summary**: later

# Trial Setting

- **Population**: Patients with Type 2 diabetes.
- **Treatment**: Assignment to active treatment ($Z = 1$) vs placebo ($Z = 0$).
- **Outcome variable**: Change in HbA1c from baseline to week 36.
- **Intercurrent events**: Use of rescue medication.
- **Strategy**: Will discuss two settings; treatment policy and hypothetical strategy.
- **Population summary**: later
- $Z \in \{0, 1\}$ is randomized with $P(Z = 1) = p_1$
  - $p_1 \in (0, 1)$ is some fixed constant.
  - Non-stratified randomziation between $Z = 0, 1$.

# Outline


1 Introduction

2 Counterfactuals and Causal Estimands

3 Identification Assumptions
- The consistency assumption
- Exchangeability
- Positivity

4 Examples in RCTs
- Treatment-Policy Estimand
- A Hypothetical Estimand

# "Intention-to-treat Analysis"

**Strategy**: Use outcome values regardless of discontinuation or use of rescue medication (i.e., **treatment policy strategy**).

# "Intention-to-treat Analysis"

**Strategy**: Use outcome values regardless of discontinuation or use of rescue medication (i.e., **treatment policy strategy**).

**Population summary**: $E(Y^1 - Y^0)$.

# "Intention-to-treat Analysis"

**Strategy**: Use outcome values regardless of discontinuation or use of rescue medication (i.e., **treatment policy strategy**).

**Population summary**: $E(Y^1 - Y^0)$.

- **Consistency** ($Y_i^z = Y_i$ for every individual with $Z_i = z$):
  - Ensured through a proper definition of treatments (e.g., dose, frequency, route of administration)

# "Intention-to-treat Analysis"

**Strategy**: Use outcome values regardless of discontinuation or use of rescue medication (i.e., **treatment policy strategy**).

**Population summary**: $E(Y^1 - Y^0)$.

- **Consistency** ($Y_i^z = Y_i$ for every individual with $Z_i = z$):
  - Ensured through a proper definition of treatments (e.g., dose, frequency, route of administration)
- **Exchangeability** ($Y^z \perp\!\!\!\perp Z$):
  - Random process for treatment assignment does not depend on any covariates/confounders or on potential outcomes.

# "Intention-to-treat Analysis"

**Strategy**: Use outcome values regardless of discontinuation or use of rescue medication (i.e., **treatment policy strategy**).

**Population summary**: $E(Y^1 - Y^0)$.

- **Consistency** ($Y_i^z = Y_i$ for every individual with $Z_i = z$):
  - Ensured through a proper definition of treatments (e.g., dose, frequency, route of administration)
- **Exchangeability** ($Y^z \perp\!\!\!\perp Z$):
  - Random process for treatment assignment does not depend on any covariates/confounders or on potential outcomes.
- **Positivity**:
  - $P(Z = 1) = p_1 > 0$ and $P(Z = 0) = 1 - p_1 > 0$.

# Causal vs Statistical Assumptions

- Suppose interest lies in estimating the **conditional (causal) contrasts**: $\frac{E(Y^1|X=x)/\{1-E(Y^1|X=x)\}}{E(Y^0|X=x)/\{1-E(Y^0|X=x)\}}$.

# Causal vs Statistical Assumptions

- Suppose interest lies in estimating the **conditional (causal) contrasts**: $\frac{E(Y^1|X=x)/\{1-E(Y^1|X=x)\}}{E(Y^0|X=x)/\{1-E(Y^0|X=x)\}}$.
- For a binary outcome $Y$, it is common to fit

$$logit\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X.$$

# Causal vs Statistical Assumptions

- Suppose interest lies in estimating the **conditional (causal) contrasts**: $\frac{E(Y^1|X=x)/\{1-E(Y^1|X=x)\}}{E(Y^0|X=x)/\{1-E(Y^0|X=x)\}}$.
- For a binary outcome $Y$, it is common to fit

$$logit\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X.$$

- **Statistical modelling assumption**: no interaction between $Z$ and $X$ on the linear scale
  - **Not implied by randomization**.

# Causal vs Statistical Assumptions

- Suppose interest lies in estimating the **conditional (causal) contrasts**: $\frac{E(Y^1|X=x)/\{1-E(Y^1|X=x)\}}{E(Y^0|X=x)/\{1-E(Y^0|X=x)\}}$.
- For a binary outcome $Y$, it is common to fit

$$logit\{E(Y|Z,X)\} = \beta_0 + \beta_1 Z + \beta_2 X.$$

- **Statistical modelling assumption**: no interaction between $Z$ and $X$ on the linear scale
  - **Not implied by randomization**.
- When the model is misspecified, the standard likelihood-based estimators of $\beta_1$ may not generally target either $\frac{E(Y^1|X=x)/\{1-E(Y^1|X=x)\}}{E(Y^0|X=x)/\{1-E(Y^0|X=x)\}}$ or $\frac{E(Y^1)/\{1-E(Y^1)\}}{E(Y^0)/\{1-E(Y^0)\}}$.

# Outline


1 Introduction

2 Counterfactuals and Causal Estimands

3 Identification Assumptions
- The consistency assumption
- Exchangeability
- Positivity

4 Examples in RCTs
- Treatment-Policy Estimand
- A Hypothetical Estimand

# A Hypothetical Estimand (1)

**Strategy**: outcome value of interest is the value that would have been observed if no one would start rescue medication (i.e., **hypothetical strategy**).

# A Hypothetical Estimand (1)

**Strategy**: outcome value of interest is the value that would have been observed if no one would start rescue medication (i.e., **hypothetical strategy**).

- Let $S$ denote switching status: $S = 1$ for switchers, and 0 otherwise.

# A Hypothetical Estimand (1)

**Strategy**: outcome value of interest is the value that would have been observed if no one would start rescue medication (i.e., **hypothetical strategy**).

- Let $S$ denote switching status: $S = 1$ for switchers, and 0 otherwise.

**Population summary**: $E(Y^{Z=1,S=0} - Y^{Z=0,S=0})$.

# A Hypothetical Estimand (1)

**Strategy**: outcome value of interest is the value that would have been observed if no one would start rescue medication (i.e., **hypothetical strategy**).

- Let $S$ denote switching status: $S = 1$ for switchers, and 0 otherwise.

**Population summary**: $E(Y^{Z=1,S=0} - Y^{Z=0,S=0})$.

- **Consistency** ($Y_i^{z,s} = Y_i$ for every individual with $Z_i = z$ and $S_i = s$):
  - Ensured through a proper definition of treatments and (switching to) rescue medication.

# A Hypothetical Estimand (1)

**Strategy**: outcome value of interest is the value that would have been observed if no one would start rescue medication (i.e., **hypothetical strategy**).

- Let $S$ denote switching status: $S = 1$ for switchers, and 0 otherwise.

**Population summary**: $E(Y^{Z=1,S=0} - Y^{Z=0,S=0})$.

- **Consistency** ($Y_i^{z,s} = Y_i$ for every individual with $Z_i = z$ and $S_i = s$):
  - Ensured through a proper definition of treatments and (switching to) rescue medication.
- **Conditional exchangeability** ($Y^{z,s} \perp\!\!\!\perp S|X, Z = z$):
  - Depends on whether all variables which explain $Y^{z,s}$ given $X$ and $Z$ are measured.

# A Hypothetical Estimand (1)

**Strategy**: outcome value of interest is the value that would have been observed if no one would start rescue medication (i.e., **hypothetical strategy**).

- Let $S$ denote switching status: $S = 1$ for switchers, and 0 otherwise.

**Population summary**: $E(Y^{Z=1,S=0} - Y^{Z=0,S=0})$.

- **Consistency** ($Y_i^{z,s} = Y_i$ for every individual with $Z_i = z$ and $S_i = s$):
  - Ensured through a proper definition of treatments and (switching to) rescue medication.
- **Conditional exchangeability** ($Y^{z,s} \perp\!\!\!\perp S|X, Z = z$):
  - Depends on whether all variables which explain $Y^{z,s}$ given $X$ and $Z$ are measured.
- **Positivity**: often not realistic.

# A Hypothetical Estimand (2)

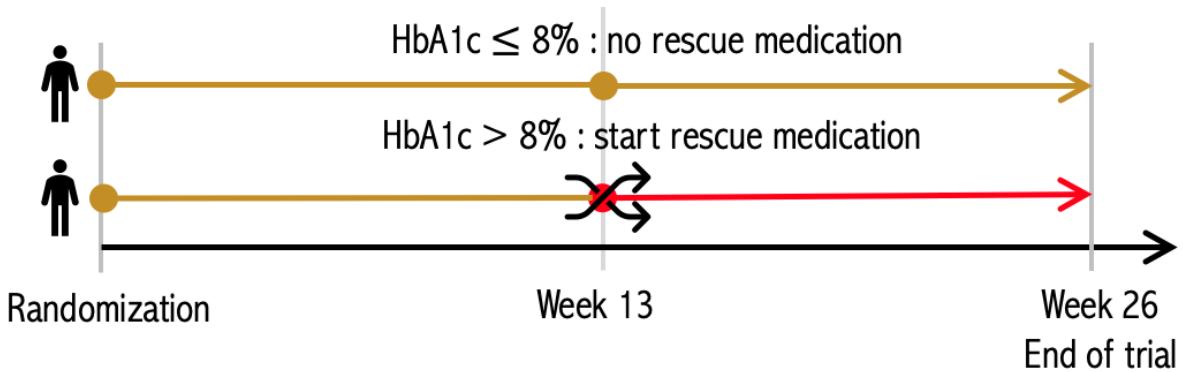
- Often not realistic.

# A Hypothetical Estimand (2)

- Often not realistic.
- No similar patients who did and did not receive rescue medication,
  esp. in case of **deterministic rules** in the protocol

# A Hypothetical Estimand (2)

- Often not realistic.
- No similar patients who did and did not receive rescue medication,
  esp. in case of **deterministic rules** in the protocol




- Most estimands and estimators are developed for settings without deterministic rules,
  but they are used in these settings!

Thank you for your attention!

E-mail: kelly.vanlancker@ugent.be

# References I

EMA (2015). Guideline on adjustment for baseline covariates in clinical trials. Last checked: 2022-05-30.

FDA and EMA (1998). E9 statistical principles for clinical trials. U.S. Food and Drug Administration: CDER/CBER. European Medicines Agency: CPMP/ICH/363/96. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf. Last checked: 2021-02-03.