Causal questions

Kim Luijken

SIKS May 31st 2023
This lecture

1. Articulating causal questions

2. Exercise: the importance of articulating a causal question

   Two analyses both suitable to estimate a causal effect, yet answering a different causal question

3. Examples of causal questions in a longitudinal setting
Causal questions

What would happen to outcome Y had exposure A been different from what was observed?
3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: What if I had done ... ? Why?
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache?
Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: What if I do ... ? How?
(What would Y be if I do X? How can I make Y happen?)

EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: What if I see ... ?
(How are the variables related? How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?
Articulating causal questions

• Algorithms with causal aim are intended to inform future decisions

• It is therefore of utmost importance that their outputs are interpreted correctly

• Formulating the causal question addressed in an analysis is quite the challenge → let’s practice!
What is the causal question that is answered by a quantitative analysis?
Example on influenza vaccination

• People can receive an invitation for vaccination against influenza through general practitioner in the Netherlands (Oct – Nov)

• Want to know whether the influenza vaccine is effective in reducing mortality risk in people who receive this invitation

• Observational data are available on people invited for vaccination (general practitioner records: vaccination status, mortality, and relevant covariates)
Articulating causal questions

Typical formulation of a causal analysis question would be:

What is the effect of influenza vaccination compared to no vaccination on 3-month mortality risk in adults invited for vaccination?
Do it yourself!
Do it yourself! Exercise 1

Write down the causal questions underlying the two analyses in the practical (around 20 minutes).

- No need to understand all steps of the analysis
- Look at similarities and differences between the two and write down some thoughts how this might affect interpretation of findings

https://github.com/KLuijken/SIKS_2023
Exercise 1 – Discussion

What is the difference between analysis 1 and 2?

Which causal questions are underlying?
Exercise 1 – Discussion

• Analysis 1: average treatment effect
• Analysis 2: average treatment effect on the treated
Exercise 1 – Discussion

Typical formulation of a causal analysis question would be:

What is the effect of influenza vaccination on 3-month mortality risk in adults ≥60 years of age compared to not being vaccinated?

However, this would allow for either analysis, while the interpretation differs!
Exercise 1 – Discussion

<table>
<thead>
<tr>
<th>Causal question</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average treatment effect (ATE)</strong></td>
<td></td>
</tr>
<tr>
<td>[ \Pr[Y^{a=1} = 1] - \Pr[Y^{a=0} = 1] ]</td>
<td>What would be the difference in average 3-month mortality risk if all adults who were invited to receive the influenza vaccination had taken it, compared to if they had not taken it?</td>
</tr>
<tr>
<td></td>
<td>-0.34 (95% CI, -0.36 to -0.33)</td>
</tr>
<tr>
<td><strong>Average treatment effect on the treated (ATT)</strong></td>
<td></td>
</tr>
<tr>
<td>[ \Pr[Y^{a=1} = 1</td>
<td>A = 1] - \Pr[Y^{a=0} = 1</td>
</tr>
<tr>
<td></td>
<td>-0.50 (95% CI, -0.52 to -0.48)</td>
</tr>
</tbody>
</table>
### Exercise 1 – Discussion

<table>
<thead>
<tr>
<th>Causal question</th>
<th>Medical decision to be informed by causal question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average treatment effect (ATE) [\Pr[Y^{a=1} = 1] - \Pr[Y^{a=0} = 1]]</td>
<td>Implementing a population-based influenza vaccination policy, where this study provides information on potential maximal mortality reduction in the population due to the vaccine</td>
</tr>
<tr>
<td>What would be the difference in average 3-month mortality risk if all adults who were invited to receive the influenza vaccination had taken it, compared to if they had not taken it?</td>
<td></td>
</tr>
<tr>
<td>Average treatment effect on the treated (ATT) [\Pr[Y^{a=1} = 1</td>
<td>A = 1] - \Pr[Y^{a=0} = 1</td>
</tr>
<tr>
<td>What would be the difference in average 3-month mortality risk if all adults who took the influenza vaccination had instead not taken it?</td>
<td></td>
</tr>
</tbody>
</table>
In this exercise, we did some reverse engineering! We determined the causal question based on the performed analysis.

The backwards process of the statistical analysis implicitly defining an otherwise unspecified causal research question is not acceptable

(paraphrased from Ratitch, 2020 TIRS)
Exercise 1 – Lesson learned

Formulating a clear causal question:

• Prevents misinterpretation of results
• Informs the choice of data collection and quantitative analysis
Elements of a causal question

Population: Who and at what time

Contrasted treatments: What, when, and how

Endpoint: What, when, and how

Population-level summary measure
### Elements of a causal question

<table>
<thead>
<tr>
<th>Population</th>
<th>All individuals registered at a general practice invited for vaccination through a National Influenza Prevention Program in the period October and November</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrasted treatments</td>
<td>Taking an intramuscular influenza vaccination versus not taking an influenza vaccination</td>
</tr>
<tr>
<td>Endpoint</td>
<td>3-Months risk of all-cause mortality</td>
</tr>
<tr>
<td>Population-level summary measure</td>
<td>Marginal risk difference</td>
</tr>
</tbody>
</table>
## Another causal question

<table>
<thead>
<tr>
<th>Causal question</th>
<th>Medical decision to be informed by causal question</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average treatment effect (ATE)</strong></td>
<td>Implementing a population-based influenza vaccination policy, where this study provides information on potential maximal mortality reduction in the population due to the vaccine.</td>
</tr>
<tr>
<td>What would be the difference in average 3-month mortality risk if all adults who were invited to receive the influenza vaccination had taken it, compared to if they had not taken it?</td>
<td></td>
</tr>
<tr>
<td><strong>Average treatment effect on the treated (ATT)</strong></td>
<td>Discontinuing an already implemented influenza vaccination policy because of insufficient effectiveness.</td>
</tr>
<tr>
<td>What would be the difference in average 3-month mortality risk if all adults who took the influenza vaccination had instead not taken it?</td>
<td></td>
</tr>
<tr>
<td><strong>Average treatment effect on the untreated (ATU)</strong></td>
<td>Stimulating uptake of an implemented vaccination policy among individuals who do not take the invitation for vaccination.</td>
</tr>
<tr>
<td>What would be the difference in average 3-month mortality risk if all who did not take the influenza vaccination had instead taken it?</td>
<td></td>
</tr>
</tbody>
</table>
Exercise 1 – Summary

Formulating a clear causal question:

- Prevents misinterpretation of results
- Informs the choice of data collection and quantitative analysis
Who wants to estimate causal effects from observational data?

Who wants to define the target estimand and effect estimator?

Who wants to use propensity score matching?
Know how to interpret a causal analysis

- Each quantitative analysis has a specific result
- Understanding what the purpose of the analysis implies in words
- Alignment between results and subsequent acts
Tell me what you want, what you really really want: Estimands in observational pharmacoepidemiologic comparative effectiveness and safety studies

Kim Luijken¹ | Rik van Eekelen² | Helga Gardarsdottir³,⁴,⁵ | Rolf H. H. Groenwold⁶,⁷ | Nan van Geloven⁶

¹Department of Epidemiology, Utrecht University Medical Center, University Utrecht, Utrecht, The Netherlands
²Centre for Reproductive Medicine, Amsterdam University Medical Center, Amsterdam, The Netherlands
³Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands
⁴Department of Clinical Pharmacy, Division Laboratories, Pharmacy and Biomedical Genetics, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
⁵Faculty of Pharmaceutical Sciences, University of Iceland, Reykjavik, Iceland
⁶Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands
⁷Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

Abstract

Purpose: Ideally, the objectives of a pharmacoepidemiologic comparative effectiveness or safety study should dictate its design and data analysis. This paper discusses how defining an estimand is instrumental to this process.

Methods: We applied the ICH-E9 (Statistical Principles for Clinical Trials) R1 addendum on estimands – which originally focused on randomized trials – to three examples of observational pharmacoepidemiologic comparative effectiveness and safety studies. Five key elements specify the estimand: the population, contrasted treatments, endpoint, intercurrent events, and population-level summary measure.

Results: Different estimands were defined for case studies representing three types of pharmacological treatments: (1) single-dose treatments using a case study about the effect of influenza vaccination versus no vaccination on mortality risk in an adult population of ≤60 years of age; (2) sustained-treatments using a case study about the effect of dipeptidyl peptidase 4 inhibitor versus glucagon-like peptide-1 agonist on hypoglycemia risk in treatment of uncontrolled diabetes; and (3) as needed treatments using a case study on the effect of nitroglycerin spray as-needed versus...
Longitudinal questions

- Exercise focused on point exposure to treatment and differences in target population
- What about sustained exposure to treatment?
Longitudinal setting

\[ L_0 \rightarrow L_1 \rightarrow L_2 \rightarrow Y_3 \]

\[ A_0 \rightarrow A_1 \rightarrow A_2 \]
Elements of a causal question

Population
- Who and at what time

Contrasted treatments
- What, when, and how

Endpoint
- What, when, and how

Population-level summary measure
Exercise

Come up with two causal questions that differ in contrasted treatments (around 10 minutes).

Setting:
- Individuals with uncontrolled diabetes
- Diabetes medication A versus B (DPP-4 versus GLP1)
- Outcome of interest is blood sugar (HbA1c level, continuous)
Examples

1. What would be the difference in average 1-year HbA$_1c$ level if all adults with uncontrolled diabetes had initiated a DPP-4 inhibitor, compared to if they had initiated a GLP1 agonist?

2. What would be the difference in average 1-year HbA$_1c$ level if all adults with uncontrolled diabetes had initiated and compliantly used a DPP-4 inhibitor, compared to if they had initiated and compliantly used a GLP1 agonist?
Example 1

\[ L_0 \rightarrow L_1 \rightarrow L_2 \rightarrow Y_3 \]

\[ A_0 \rightarrow A_1 \rightarrow A_2 \]
Example 2

\[ L_0 \rightarrow L_1 \rightarrow L_2 \rightarrow Y_3 \]

\[ A_0 \rightarrow A_1 \rightarrow A_2 \]
Examples

1. What would be the difference in average 1-year HbA$_1$C level if all adults with uncontrolled diabetes had initiated a DPP-4 inhibitor, compared to if they had initiated a GLP1 agonist?

2. What would be the difference in average 1-year HbA$_1$C level if all adults with uncontrolled diabetes had initiated and compliantly used a DPP-4 inhibitor, compared to if they had initiated and compliantly used a GLP1 agonist?
Examples

1. What would be the difference in average 1-year HbA$_1c$ level if all adults with uncontrolled diabetes had initiated a DPP-4 inhibitor, if they had initiated a GLP1 agonist?
   Advising on treatment initiation in the population of adults with uncontrolled diabetes mellitus type 2 in a population with similar treatment compliance and add-on treatments to the study sample

2. What would be the difference in average 1-year HbA$_1c$ level if all adults with uncontrolled diabetes had initiated and compliantly used a DPP-4 inhibitor, compared to if they had initiated and compliantly used a GLP1 agonist?
   Making a medical decision about sustained treatment with DPP-4 inhibitor and GLP1 agonist under perfect adherence for the population of adults with uncontrolled diabetes mellitus type 2
We focused on causal questions which compare pre-defined exposure contrasts
• These are also referred to as “static exposures”

Alternatively, one could be interested in the effect of exposure based on a treatment rule
• These are also referred to as “dynamic exposures”
Basic principle dynamic exposure

Example:
Up the dose of GLP1 from 7mg to 14mg daily if HbA1c ≥ 54 mmol/mol

This is based on patient history on HbA1c
Basic principle dynamic exposure
Basic principle dynamic exposure

- “Modern” analysis techniques: finding optimal treatment rule
- What causal question would fit here?
Estimands

1. Prepare Chocolate Cake Batter
   Preheat oven to 350 degrees, and prepare Willy's Ultimate Chocolate Cake batter. Prepare your pans with parchment. Pour 2 1/2 lbs into each 7" round pan, 1 1/2 lbs into your 6" round pan, and divide the remaining batter evenly between your 5" round pans.

2. Bake Cakes
   Bake your 7" round cakes for 50 minutes, your 6" round cake for 40 minutes, and your 5" round cakes for 30 minutes, or until a toothpick comes out clean. Set aside to cool completely in their pans on a wire rack.

3. Prepare Fillings & Simple Syrup
   Prepare your dark chocolate ganache, Italian meringue buttercream, and simple syrup. Set aside until you're ready to decorate.

4. Level Cakes
   Remove your cooled cakes from their pans and level them with a ruler and serrated knife.

5. Simple Syrup
   Give all of your cakes a simple syrup shower with Sir Squeeze, and allow to fully soak in before moving on to the next step.
Estimands

1. Prepare Chocolate Cake Batter
Preheat oven to 350 degrees, and prepare your Ultimate Chocolate Cake batter. Prepare your pans with parchment. Pour 2 1/3 lbs into each 7" round pan, 1 1/3 lbs into your 9" round pan, and divide the remaining batter evenly between your 8" round pans.

2. Bake Cakes
Bake your 7" round cakes for 50 minutes, your 6" round cake for 40 minutes, and your 8" round cakes for 30 minutes, or until a toothpick comes out clean. Set aside to cool completely in their pans on a wire rack.

3. Prepare Fillings & Simple Syrup
Prepare your dark chocolate ganache, Italian meringue buttercream, and simple syrup. Set aside until you’re ready to decorate.

4. Level Cakes
Remove your cooled cakes from their pans and level them with a ruler and serrated knife.

5. Simple Syrup
Dunk all of your cakes in a simple syrup shower with Sir Squeeze, and allow to fully soak in before moving on to the next step.

Estimand

Estimator

Estimate

Credits to Peter Tennant & Oisín Ryan
ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials

Step 5
Overview

Each setting requires formulation of a specific causal question

Formulating a clear causal question:

• Prevents misinterpretation of results
• Informs the choice of data collection and quantitative analysis
Questions or further discussion?

k.luijken@umcutrecht.nl
Further reading

- Ratitch, Bell, ... , Lipkovich (2020). Choosing estimands in clinical trials: putting the ICH E9 (R1) into practice. *Therapeutic innovation & regulatory science*, 54, 324-341.