OUTLINE

1. CAUSAL INFERENCE
   Background
   Association versus causation
   Key conditions for causal inference

2. DIRECTED ACYCLIC GRAPHS
   Background
   Paradoxes
   Definitions and illustrations
WHY?

TO BE ABLE TO ESTIMATE THE CAUSAL EFFECT OF A VARIABLE (E.G. AN EXPOSURE) ON AN OUTCOME IN SPECIFIC STUDY SETTINGS
NOTATION

$Y$: outcome (here: binary 0/1)

$E$: observed exposure (here: binary 0/1)

$e$: hypothetical exposure (here: binary 0/1)

$P(Y=1|E=1)$: probability of $Y=1$ in a population exposed to $E=1$

$P(Y^e=1 = 1)$: probability of outcome $y=1$, would exposure $e=1$ be chosen

$\rightarrow Y^e=0, Y^e=1$: potential/counterfactual outcomes
ASSOCIATION VERSUS CAUSATION (1/2)

Graph shown in different publications by Miguel A. Hernán and James M. Robins, Harvard T. H. Chan School of Public Health
ASSOCIATION VERSUS CAUSATION (2/2)

ASSOCIATION:

\[ P(Y=1|E=1) \neq P(Y=1|E=0) \]

for two disjoint exposure subgroups

CAUSATION:

\[ P(Y^{e=1} = 1) \neq P(Y^{e=0} = 1) \]

based on a counterfactual view on the entire population

SHARP CAUSAL NULL HYPOTHESIS:

\[ P(Y^{e=1} = 1) = P(Y^{e=0} = 1) \]
MEASURES OF ASSOCIATION

- **RISK DIFFERENCE**

  \[
  P(Y = 1|E = 1) - P(Y = 1|E = 0) \quad \Rightarrow \text{value of 0} \n  \]
  \[\Rightarrow Y \text{ independent of } E\n  \]

- **RISK RATIO**

  \[
  \frac{P(Y = 1|E = 1)}{P(Y = 1|E = 0)} \quad \Rightarrow \text{value of 1} \n  \]
  \[\Rightarrow Y \text{ independent of } E\n  \]

- **ODDS RATIO**

  \[
  \frac{P(Y = 1|E = 1)/P(Y = 0|E = 1)}{P(Y = 1|E = 0)/P(Y = 0|E = 0)} \]
MEASURES OF CAUSAL EFFECTS

• CAUSAL RISK DIFFERENCE

\[ P(Y_e = 1) - P(Y_e = 0) \]

\( \Rightarrow \) value of 0 \( \neq \) no causal effect

• CAUSAL RISK RATIO

\[ \frac{P(Y_e = 1)}{P(Y_e = 0)} \]

\( \Rightarrow \) value of 1 \( \neq \) no causal effect

• CAUSAL ODDS RATIO

\[ \frac{P(Y_e = 1)/P(Y_e = 0)}{P(Y_e = 1)/P(Y_e = 0)} \]
IDEAL RANDOMIZED CONTROLLED TRIAL

2 exchangeable sub-populations

Exchangeability:
Probability of $Y|E$ independent of exposure assignment

$P(Y^e=0 = 1) \quad P(Y^e=1 = 1)$

$P(Y = 1|E = 0) = P(Y^e=0 = 1)$

$P(Y = 1|E = 1) = P(Y^e=1 = 1)$
OBSERVATIONAL COHORT STUDIES

Typically: Association ≠ Causation

Reason: exposure not random, but dependent on other variables $C$ (e.g. age, medical history)

- Absence of exchangeability between exposure subgroups
- Presence of confounding
- Complex causal pathways between variables (incl. exposure) and outcome
CONDITIONS FOR CAUSAL INFERENCE (1/2)

• **EXCHANGEABILITY**
  Outcome $Y|E$ independent of exposure assignment to population subgroups

• **POSITIVITY**
  $$P(E=e)>0, \text{ for all } e$$

• **CONSISTENCY**
  Well-defined controllable types of exposure

  ➔ Fulfilled in “ideal” marginally randomized controlled trials
### CONDITIONS FOR CAUSAL INFERENCE (2/2)

<table>
<thead>
<tr>
<th>Conditionally randomized controlled trial</th>
<th>Observational cohort study (confounding due to a set of variables $C$, e.g. gender, co-medication, ..., with a causal effect on exposure and outcome)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conditional exchangeability</strong></td>
<td>Exchangeable exposure groups within each stratum of $G$</td>
</tr>
</tbody>
</table>
| **Conditional positivity**               | No empty exposure subgroups across all strata of $G$  
$P(E=e|G=g)>0$, for all $e, g$                                         |
| **Consistency**                          | Well defined interventions (e.g. drug and placebo)                                                                            |
|                                           | Exchangeable exposure groups within each stratum of $C$                                                                        |
|                                           | No empty exposure subgroups across all strata of $C$  
$P(E=e|C=c)>0$, for all $e, c$                                         |
|                                           | Well defined interventions (e.g. oral and intravenous treatment)                                                             |
## CONDITIONS FOR CAUSAL INFERENCE (2/2)

<table>
<thead>
<tr>
<th>Conditionally randomized controlled trial (stratification, e.g. by gender $G$, before randomization)</th>
<th>Observational cohort study (confounding due to a set of variables $C$, e.g. gender, co-medication,…, with a causal effect on exposure and outcome)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conditional exchangeability</strong></td>
<td>Exchangeable exposure groups within each stratum of $G$</td>
</tr>
<tr>
<td><strong>Conditional positivity</strong></td>
<td>No empty exposure subgroups across all strata of $G$ $P(E=e</td>
</tr>
<tr>
<td><strong>Consistency</strong></td>
<td>Well defined interventions (e.g. drug and placebo)</td>
</tr>
</tbody>
</table>
### CONDITIONS FOR CAUSAL INFERENCE (2/2)

<table>
<thead>
<tr>
<th></th>
<th>Conditionally randomized controlled trial (stratification, e.g. by gender $G$, before randomization)</th>
<th>Observational cohort study (confounding due to a set of variables $C$, e.g. gender, co-medication, …, with a causal effect on exposure and outcome)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conditional exchangeability</strong></td>
<td>Exchangeable exposure groups within each stratum of $G$</td>
<td>Exchangeable exposure groups within each stratum of $C$</td>
</tr>
<tr>
<td><strong>Conditional positivity</strong></td>
<td>No empty exposure subgroups across all strata of $G$: $P(E=e</td>
<td>G=g)&gt;0$, for all $e, g$</td>
</tr>
<tr>
<td><strong>Consistency</strong></td>
<td>Well defined interventions (e.g. drug and placebo)</td>
<td>Well defined interventions (e.g. oral and intravenous treatment)</td>
</tr>
</tbody>
</table>
DIRECTED ACYCLIC GRAPHS (DAGs)
WHY?

- CONCISE GRAPHICAL VISUALIZATION OF (COMPLEX) CAUSAL ASSUMPTIONS IN OBSERVATIONAL STUDIES
- VISUAL COMPARISON BETWEEN DIFFERENT CAUSAL APPROACHES TO THE SAME PROBLEM
- SUPPORTING TOOL FOR IDENTIFICATION OF POTENTIAL SOURCES OF CONFOUNDING AND BIAS
- SUPPORTING TOOL FOR METHODS CHOICE AND RESULTS INTERPRETATION

Not a pre-requisite, but often very helpful for causal inference
BIRTH WEIGHT PARADOX (1/2)

• In the general population: low birthweight → higher infant mortality

• Paradox finding: lower mortality of babies with low birthweight among smoking mothers than among non-smoking mothers

• Does smoking have a beneficial effect on child mortality?

• Of course not!
BIRTH WEIGHT PARADOX (2/2)

CLARIFICATION:
Rate of babies with low birthweight higher among smoking than among non-smoking mothers ➔ in general higher mortality in babies of smoking mothers

EXPLANATION OF THE PARADOX FINDING:
• Equal “baseline” risk of low birthweight in both groups of mothers
• BUT: birth weight distribution among babies of smoking mothers shifted toward the lower end ➔ low birthweight in some of the otherwise healthy babies ➔ lower mortality among the otherwise healthy babies than among babies with smoking-independent severe medical conditions or unfavorable genetic disposition
SIMPSON’S PARADOX (1/2)

- Exposure $E$ harmful in female patients
- Exposure $E$ harmful in male patients
- PARADOX FINDING:
  Exposure $E$ not harmful in the overall population?

SIMPSON’S PARADOX (2/2)

EXPLANATION OF THE PARADOX FINDING:

- Male and female populations of equal size, BUT
- Higher exposure rate among males than among females
- In general, higher recovery rate in males than in females

→ Important causal considerations
→ Combined view leading to misinterpretations
CHARACTERISTICS OF A **DAG**

• **Graph:** nodes/variables

  \[
  N_1 \quad N_2 \quad N_3 \quad N_4
  \]

  edges

• **Directed Graph:**
  (from cause to outcome)

  \[
  N_1 \rightarrow N_2 \leftarrow N_3 \rightarrow N_4
  \]

• **Directed Acyclic Graph:**

  \[
  N_1 \leftarrow N_2 \rightarrow N_3 \leftarrow N_4
  \]
GENERAL NOTE ON INTERPRETATION

NO EDGE $\triangleq$ NO DIRECT CAUSAL EFFECT (SHARP NULL ASSUMPTION)
EDGE $\triangleq$ EXPECTED CAUSAL EFFECT (OF ANY STRENGTH)

Absence-oriented approach:

- More edges $\Rightarrow$ less causal assumptions

- Less edges $\Rightarrow$ more (sharp!) causal assumptions
COMPONENTS OF A DAG

PATH: Sequence of edges connecting two nodes

POSSIBLE RELATIONSHIPS BETWEEN NODE $N$ AND OTHER NODES:
- Descendant of $N$: a node directly or indirectly caused by $N$
- Child of $N$: a node directly caused by $N$
- Ancestor of $N$: a node directly or indirectly causing $N$
- Parent of $N$: a node directly causing $N$

COLLIDER (L):

```
N_1 --> L --> N_2
```

$N_1$ and $N_2$ are colliders.
CONDITIONING ON VARIABLES (1/2)

BLOCKED PATH:
Path with
• a non-collider $N_i$ being conditioned on OR
• a collider $L$ not being conditioned on and not having any descendent $Y$ being conditioned on

EXAMPLES OF BLOCKED PATHS (CONDITIONING $\perp \!
\!
\!
\vdash$):

\[ N_1 \rightarrow N_2 \rightarrow N_3 \rightarrow N_4 \rightarrow N_5 \]  
\[ N_1 \rightarrow N_2 \rightarrow L \rightarrow Y \]
CONDITIONING ON VARIABLES (2/2)

OPEN PATH $\triangleq$ UNBLOCKED PATH:
Path with
• no non-collider $N_i$ being conditioned on AND
• a collider $L$ being conditioned on or having any descendent $Y$ being conditioned on

EXAMPLES OF OPEN PATHS:
SELECTION BIAS

INDUCED BY
OPENING A PATH BY CONDITIONING ON A COLLIDER OR ONE OF ITS DESCENDANTS

EXAMPLE: Birth Weight Paradox

\( S \): smoking status

\( L \): birthweight

\( N \): smoking-independent medical or genetic factors

\( Y \): mortality

View on general population

Selection bias

DIRECTED SEPARATION (D-SEPARATION)

D-SEPARATION BETWEEN TWO VARIABLES $\triangleright$ BLOCKAGES OF ALL PATHS BETWEEN THEM

- D-separation between $N_1$ and $Y$
- D-separation between $N_2$ and $Y$
DIRECTED CONNECTION (D-CONNECTION)

D-CONNECTION OF TWO VARIABLES $\triangleleft$ AT LEAST ONE OPEN PATH BETWEEN THEM

- D-separation between $N_1$ and $Y$
- D-connection of $N_2$ and $Y$

- D-connection of $N_1$ and $Y$
- D-connection of $N_2$ and $Y$
CONFOUNDING

EXAMPLE: Simpson’s Paradox:

\[
E: \text{exposure} \quad Y: \text{recovery} \quad G: \text{gender}
\]

\[
E \rightarrow Y \quad \text{ignoring } G
\]

\[
E \rightarrow Y \quad \text{sharp null assumption between } G \text{ and } E
\]

\[\begin{align*}
G \\
E & \quad Y
\end{align*}\]

accounting for \(G\) as a common cause of \(E\) and \(Y\)

⇒ ACCOUNTING FOR CONFOUNDING
CAUSAL DAGs FOR CAUSAL INFERENCEx

ASSUMPTIONS:

• All common causes captured by the graph
• No unmeasured confounding

➔ Very strong and critical assumptions
➔ Prerequisites for accurate and reliable causal inference
SOME REFERENCES


THANK YOU,
BACK-UP SLIDES.
WHICH VARIABLES ARE D-SEPARATED/CONNECTED?