INTRODUCTION TO CAUSAL INFERENCE AND DIRECTED ACYCLIC GRAPHS

ACTELION
A JANSSEN PHARMACEUTICAL COMPANY
OF Johnson & Johnson

EVA-MARIA DIDDEN

JULY 5TH, 2019

OUTLINE

1. CAUSAL INFERENCE

Background

Association versus causation

Key conditions for causal inference

2. DIRECTED ACYCLIC GRAPHS

Background

Paradoxes

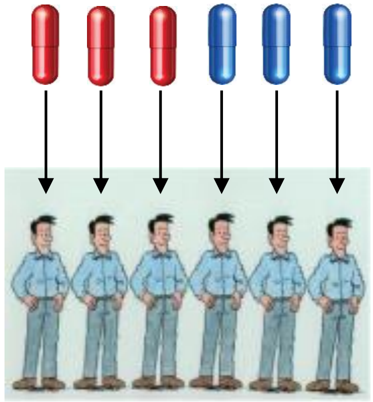
Definitions and illustrations

CAUSAL INFERENCE

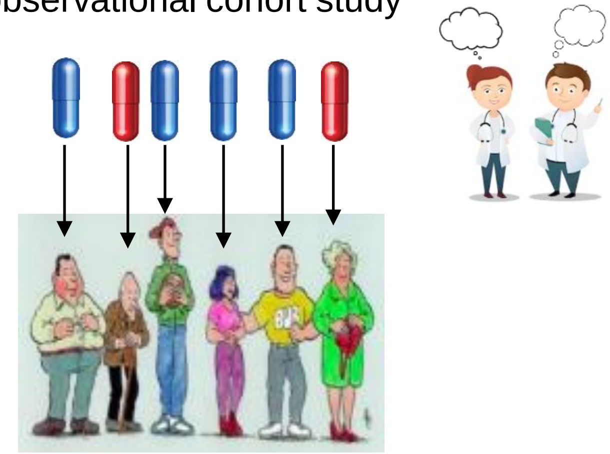
WHY?

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

randomized controlled trial



observational cohort study



NOTATION

Y: outcome (here: binary 0/1)

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

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

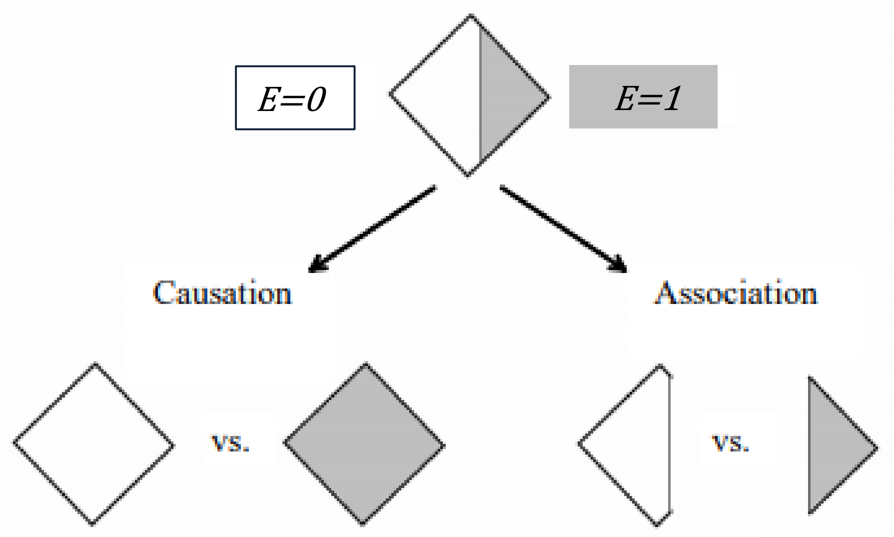
$P(Y=1/E=1)$: probability of $Y=1$ in a population exposed to $E=1$

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

→ $Y^{e=0}, Y^{e=1}$: potential/counterfactual outcomes

ASSOCIATION VERSUS CAUSATION (1/2)

population of interest



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

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

$$P(Y=1|E=1)-P(Y=1|E=0)$$

➔ value of 0 $\triangleq$ Y independent of E

- RISK RATIO

$$\frac{P(Y=1|E=1)}{P(Y=1|E=0)}$$

➔ value of 1 $\triangleq$ Y independent of E

- 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

$$P(Y^{e=1} = 1) - P(Y^{e=0} = 1)$$

➔ value of 0 ≙ no causal effect

- CAUSAL RISK RATIO

$$\frac{P(Y^{e=1} = 1)}{P(Y^{e=0} = 1)}$$

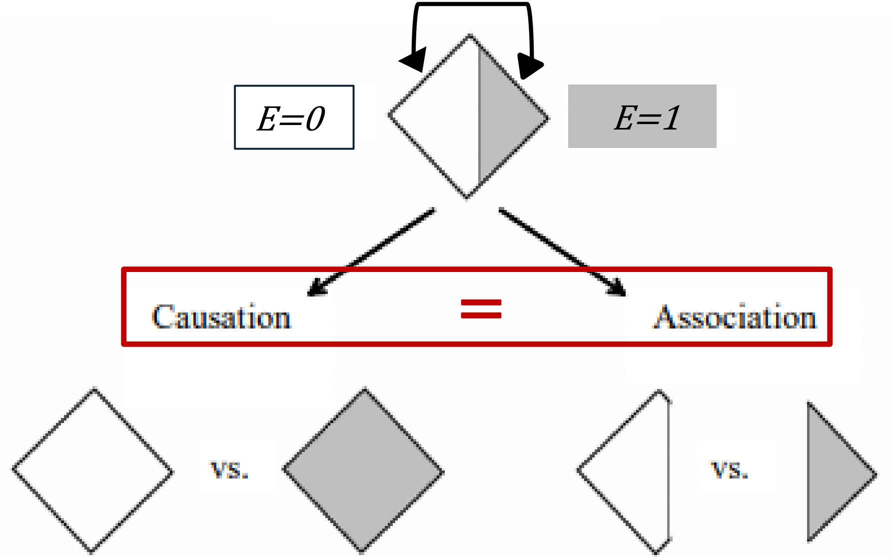
➔ value of 1 ≙ no causal effect

- CAUSAL ODDS RATIO

$$\frac{P(Y^{e=1} = 1)/P(Y^{e=1} = 0)}{P(Y^{e=0} = 1)/P(Y^{e=0} = 0)}$$

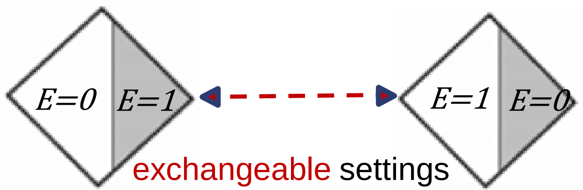
IDEAL RANDOMIZED CONTROLLED TRIAL

2 exchangeable sub-populations



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

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



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)

	Conditionally randomized controlled trial (stratification, e.g. by gender *G*, before randomization)	**Observational cohort study** (confounding due to a set of variables *C*, e.g. gender, co-medication,…, with a causal effect on exposure and outcome)
Conditional exchangeability	Exchangeable exposure groups within each stratum of *G*	Exchangeable exposure groups within each stratum of *C*
Conditional positivity	No empty exposure subgroups across all strata of *G* $P(E=e\|G=g)>0,$ for all e, g	No empty exposure subgroups across all strata of *C* $P(E=e\|C=c)>0,$ for all e, c
Consistency	Well defined interventions (e.g. drug and placebo)	Well defined interventions (e.g. oral and intravenous treatment)

CONDITIONS FOR CAUSAL INFERENCE (2/2)

	Conditionally randomized controlled trial (stratification, e.g. by gender *G*, before randomization)	**Observational cohort study** (confounding due to a set of variables *C*, e.g. gender, co-medication,…, with a causal effect on exposure and outcome)
Conditional exchangeability	Exchangeable exposure groups within each stratum of *G*	Exchangeable exposure groups within each stratum of *C*
Conditional positivity	No empty exposure subgroups across all strata of *G* $P(E=e/G=g)>0$, for all e, g	No empty exposure subgroups across all strata of *C* $P(E=e/C=c)>0$, for all e, c
Consistency	Well defined interventions (e.g. drug and placebo)	Well defined interventions (e.g. oral and intravenous treatment)

CONDITIONS FOR CAUSAL INFERENCE (2/2)

	Conditionally randomized controlled trial (stratification, e.g. by gender *G*, before randomization)	**Observational cohort study** (confounding due to a set of variables *C*, e.g. gender, co-medication,…, with a causal effect on exposure and outcome)
Conditional exchangeability	Exchangeable exposure groups within each stratum of *G*	Exchangeable exposure groups within each stratum of *C*
Conditional positivity	No empty exposure subgroups across all strata of *G* $P(E=e\|G=g)>0,$ for all e, g	No empty exposure subgroups across all strata of *C* $P(E=e\|C=c)>0,$ for all e, c
Consistency	Well defined interventions (e.g. drug and placebo)	Well defined interventions (e.g. oral and intravenous treatment)

DIRECTED ACYCLIC GRAPHS (DAGs)

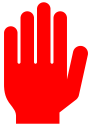
WHY?

- 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!



Allen Wilcox (2001): *On the importance – and the unimportance – of birthweight*, International Journal of Epidemiology, Vol. 30: 1233-1241

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)

E=1: exposed to treatment; E=0: not exposed
Y=1: recovered; *Y=0*: not recovered

- Exposure *E* harmful in female patients

Females	Y=1	Y=0	Total	Recovery rate
E=1	2	8	10	20%
E=0	9	21	20	30%
Total	11	29	40	

- Exposure *E* harmful in male patients

Males	Y=1	Y=0	Total	Recovery rate
E=1	18	12	30	60%
E=0	7	3	10	70%
Total	25	15	40	

- PARADOX FINDING:

  Exposure *E* not harmful in the overall population?

All	Y=1	Y=0	Total	Recovery rate
E=1	20	20	40	50%
E=0	16	24	40	40%
Total	36	24	80	

Edward H. Simpson (1951): *The interpretation of Interaction in Contingency Tables*, Journal of the Royal Statistical Society, Series B, Vol 13 (2), 238-241.

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

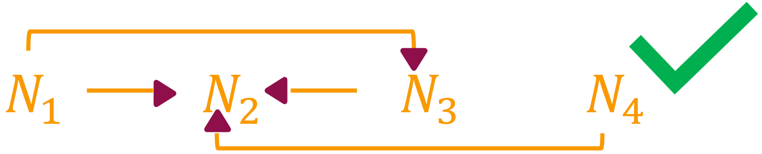
→ Important causal considerations
→ Combined view leading to misinterpretations

CHARACTERISTICS OF A **DAG**

- Graph: nodes/variables N_1 N_2 N_3 N_4

  edges

- Directed Graph:
  (from cause $\xrightarrow{\text{to}}$ outcome)



- Directed Acyclic Graph:

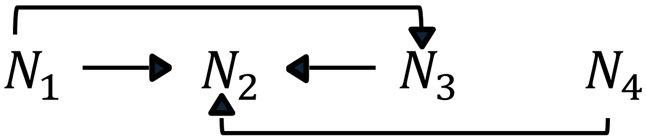
GENERAL NOTE ON INTERPRETATION

NO EDGE ≙ NO DIRECT CAUSAL EFFECT (SHARP NULL ASSUMPTION)

EDGE ≙ EXPECTED CAUSAL EFFECT (OF ANY STRENGTH)

Absence-oriented approach:

- More edges ➔ less causal assumptions



- Less edges ➔ 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):

$$\begin{tikzcd} N_1 \arrow[dr] & \\ & L \\ N_2 \arrow[ur] & \end{tikzcd}$$

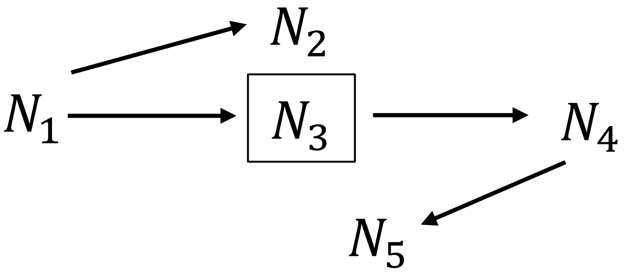
CONDITIONING ON VARIABLES (1/2)

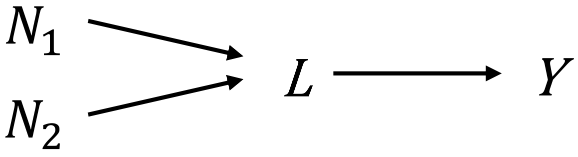
BLOCKED PATH:

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 $\triangleq \square$):





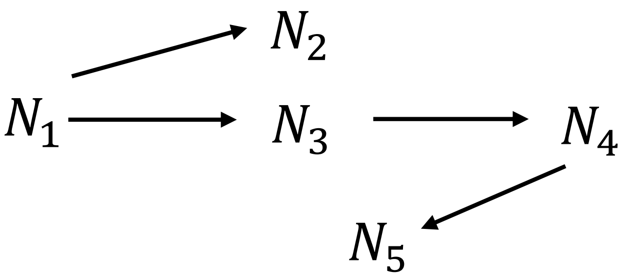
CONDITIONING ON VARIABLES (2/2)

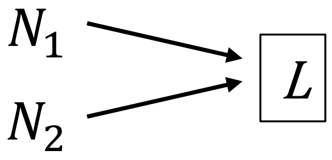
OPEN PATH ≜ UNBLOCKED PATH:

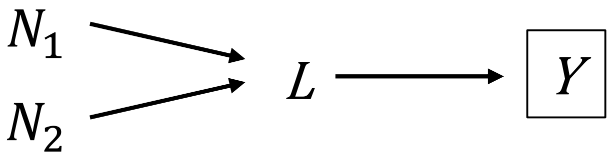
Path with

- 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

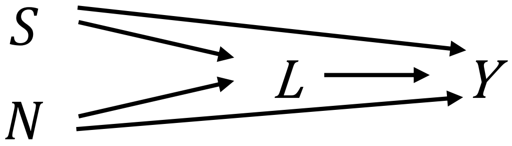
EXAMPLE: Birth Weight Paradox

S: smoking status

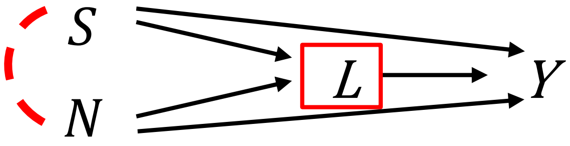
N: smoking-independent medical or genetic factors

L: birthweight

Y: mortality



View on general population

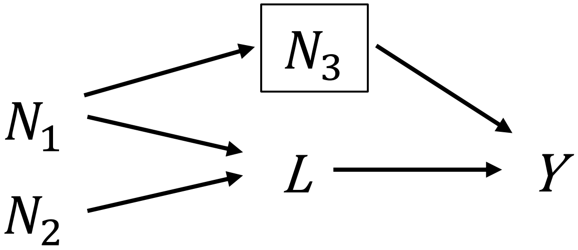


Selection bias

Allen Wilcox (2006): *The Perils of Birth Weight – A Lesson from Directed Acyclic Graphs*, American Journal of Epidemiology, Vol. 164 (11): 1121 - 1123

DIRECTED SEPARATION (D-SEPARATION)

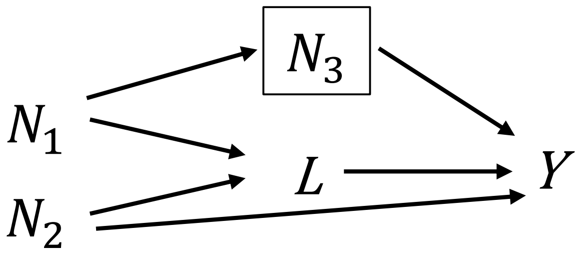
D-SEPARATION BETWEEN TWO VARIABLES $\triangleq$ BLOCKAGES OF ALL PATHS BETWEEN THEM

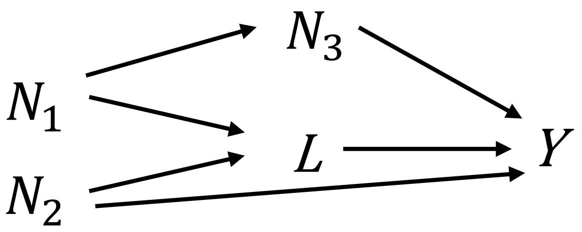


- D-separation between N_1 and Y
- D-separation between N_2 and Y

DIRECTED CONNECTION (D-CONNECTION)

D-CONNECTION OF TWO VARIABLES $\hat{=}$ AT LEAST ONE OPEN PATH BETWEEN THEM





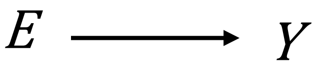
- D-separation between N_1 and Y
- D-connection of N_2 and Y

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

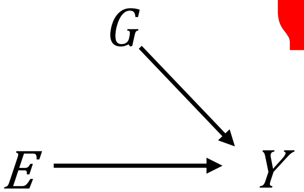
CONFOUNDING

EXAMPLE: Simpson's Paradox:

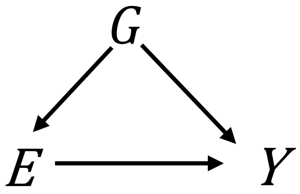
E: exposure Y: recovery G: gender



ignoring G



sharp null assumption between G and E



accounting for G as a common cause of E and Y

➔ ACCOUNTING FOR CONFOUNDING

CAUSAL DAGs FOR CAUSAL INFERENCE

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

- S. Greenland (1990). "Randomization, statistics, and causal inference." Epidemiology: 421-429
- J.M. Robins (1999): "Association, causation, and marginal structural models." Synthese 121.1: 151-179.
- S. Greenland, J. Pearl, and J.M. Robins (1999). "Causal diagrams for epidemiologic research." *Epidemiology* 10: 37-48.
- M.A. Hernán, and J.M. Robins (2006). "Estimating causal effects from epidemiological data." *Journal of epidemiology & community health* 60.7: 578-586.
- J. Pearl (2009). "Causal inference in statistics: An overview." Statistics surveys 3: 96-146.
- G.W. Imbens, and D.B. Rubin (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.

THANK YOU.

BACK-UP SLIDES.

WHICH VARIABLES ARE D-SEPARATED/CONNECTED?