Inferring Causality in Microbial Communities

Goals and Challenges

Niklas Pfister

May 9, 2023

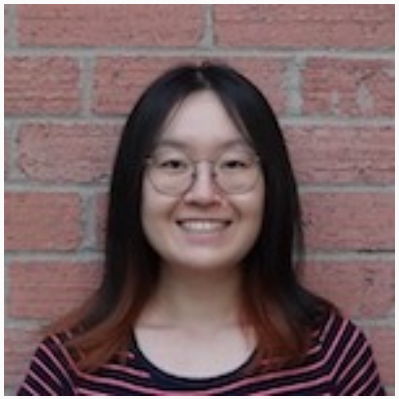
Two-day meeting - DSTS

novo nordisk fonden
Benefiting people and society

UNIVERSITY OF COPENHAGEN

Collaborators

At the University of Copenhagen



Shimeng Huang

Anton Rask Lundborg

At Helmholtz AI in Munich

Elisabeth Ailer

Niki Kilbertus

Microbiome

What is the microbiome and why is it important?

The microbiome is the collection of all mircoorganisms (bacteria, viruses, fungi,...) residing within/on a host.

The microbiome is the collection of all mircoorganisms (bacteria, viruses, fungi,...) residing within/on a host.

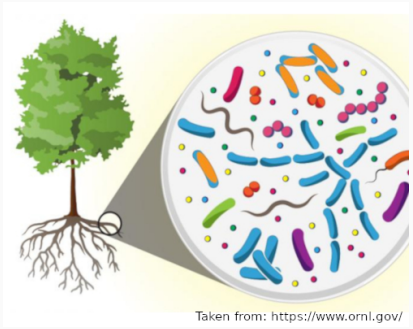
Human

- *habitats:* gut, skin, oral and nasal cavities
- *involved in* diet, immune response and diseases

The microbiome is the collection of all mircoorganisms (bacteria, viruses, fungi,...) residing within/on a host.

Human

- *habitats:* gut, skin, oral and nasal cavities
- *involved in* diet, immune response and diseases



Plant

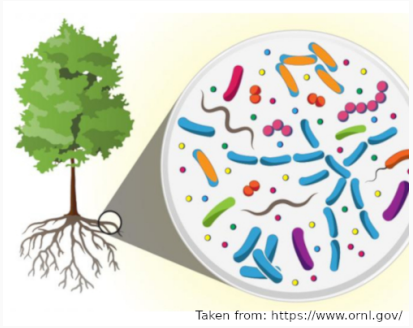
- *habitats:* roots, soil, leaf surface, intercellular
- *involved in* absorption of nutrients, protection against diseases

The microbiome is the collection of all mircoorganisms (bacteria, viruses, fungi,...) residing within/on a host.


Taken from: https://drannwellness.com/

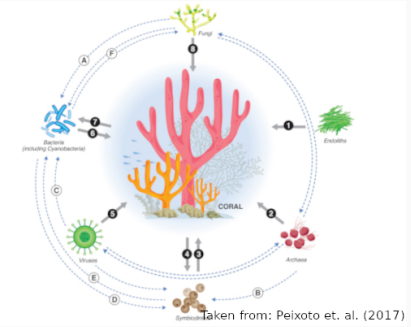
Human

- *habitats:* gut, skin, oral and nasal cavities
- *involved in* diet, immune response and diseases


Taken from: https://www.ornl.gov/

Plant

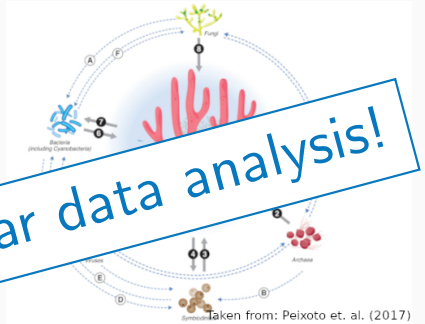
- *habitats:* roots, soil, leaf surface, intercellular
- *involved in* absorption of nutrients, protection against diseases


Taken from: Peixoto et. al. (2017)

Marine

- *habitats:* animals, algae, corals, sponges
- *involved in* biodiversity, water quality, diseases in fish

The microbiome is the collection of all mircoorganisms (bacteria, viruses, fungi,...) residing within/on a host.



Wide range of applications, but very similar data analysis!

H

- *habitats:* gut, skin, oral and nasal cavities
- *involved in* diet, immune response and diseases

Plant

- *habitats:* roots, soil, leaf surface, intercellular
- *involved in* absorption of nutrients, protection against diseases

Marine

- *habitats:* animals, algae, corals, sponges
- *involved in* biodiversity, water quality, diseases in fish

Measuring the microbiome

(1) Extract sample and preprocess

(2) Perform sequencing:

- marker gene sequencing (amplify specific DNA target)
- metagenome sequencing (directly sequence DNA fragments)

(3) Match each DNA read to a specific microbe
(substantially more complex in reality)

Measuring the microbiome

(1) Extract sample and preprocess

(2) Perform sequencing:

- marker gene sequencing (amplify specific DNA target)
- metagenome sequencing (directly sequence DNA fragments)

(3) Match each DNA read to a specific microbe
(substantially more complex in reality)

Problems:

- compositional (more recently, absolute measures are also possible)
- multiple sources of bias (host DNA, amplification bias, ...)
- zero-inflated and high-dimensional
- varies substantially over time

What are interesting research questions?

What are interesting research questions?

- [Plant] Which microbes in the soil increase the nutrition uptake?
  Increase crop yield by modifying soil.

What are interesting research questions?

- [Plant] Which microbes in the soil increase the nutrition uptake?
  Increase crop yield by modifying soil.
- [Human] How does the gut microbiome interact with a drug?
  Avoid side effects and increase efficacy of drugs.

What are interesting research questions?

- [Plant] Which microbes in the soil increase the nutrition uptake?
  Increase crop yield by modifying soil.
- [Human] How does the gut microbiome interact with a drug?
  Avoid side effects and increase efficacy of drugs.
- [Human] How does the diet affect the microbiome and vice versa?
  Create personalized diets and help people loose weight.

What are interesting research questions?

- [Plant] Which microbes in the soil increase the nutrition uptake?
  Increase crop yield by modifying soil.
- [Human] How does the gut microbiome interact with a drug?
  Avoid side effects and increase efficacy of drugs.
- [Human] How does the diet affect the microbiome and vice versa?
  Create personalized diets and help people loose weight.
- [Food] How do microbial communities interact to create taste?
  Produce better and more efficient food (e.g., cheese).

What are interesting research questions?

- [Plant] Which microbes in the soil increase the nutrition uptake?
  Increase crop yield by modifying soil.
- [Human] How does the gut microbiome interact with a drug?
  Avoid side effects and increase efficacy of drugs.
- [Human] How does the diet affect the microbiome and vice versa?
  Create personalized diets and help people loose weight.
- [Food] How do microbial communities interact to create taste?
  Produce better and more efficient food (e.g., cheese).

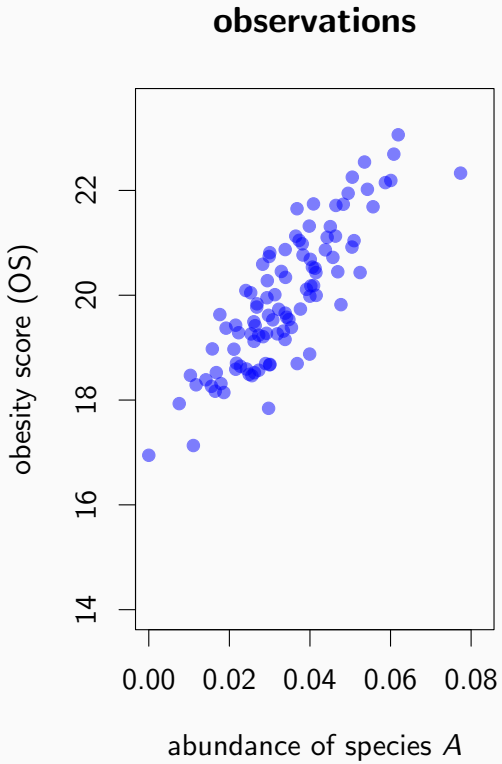
What are interesting research questions?

- [Plant] Which microbes in the soil increase the nutrition uptake?
  Increase crop yield by modifying soil.
- [Human] How does the gut microbiome interact with a drug?
  Avoid side effects and increase efficacy of drugs.
- [Human] How does the diet affect the microbiome and vice versa?
  Create personalized diets and help people loose weight.
- [Food] How do microbial communities interact to create taste?
  Produce better and more efficient food (e.g., cheese).

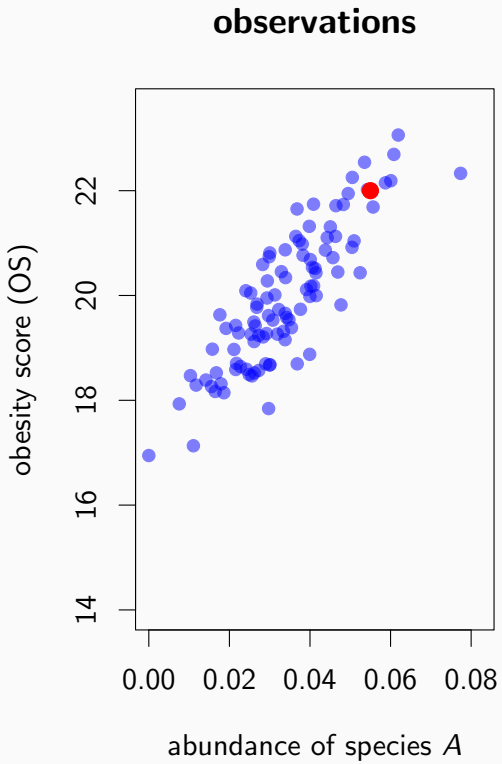
Shared goal: Infer underlying causal mechanism!

Causality

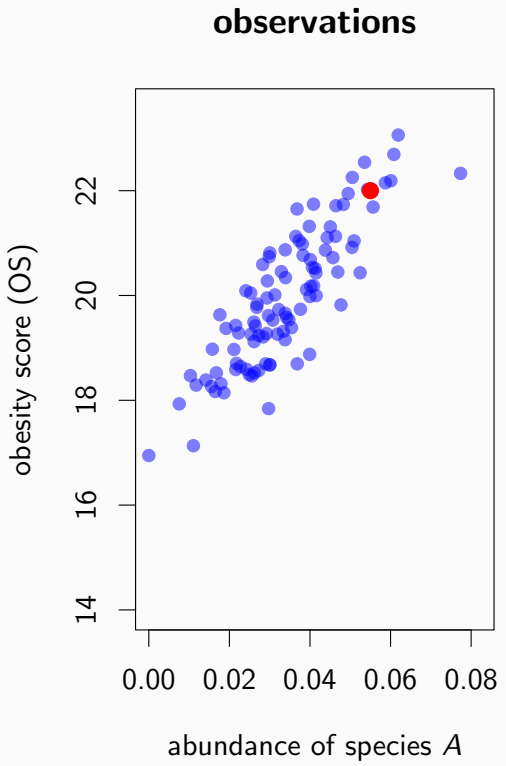
What is a causal model and how is it different from an observational statistical model?

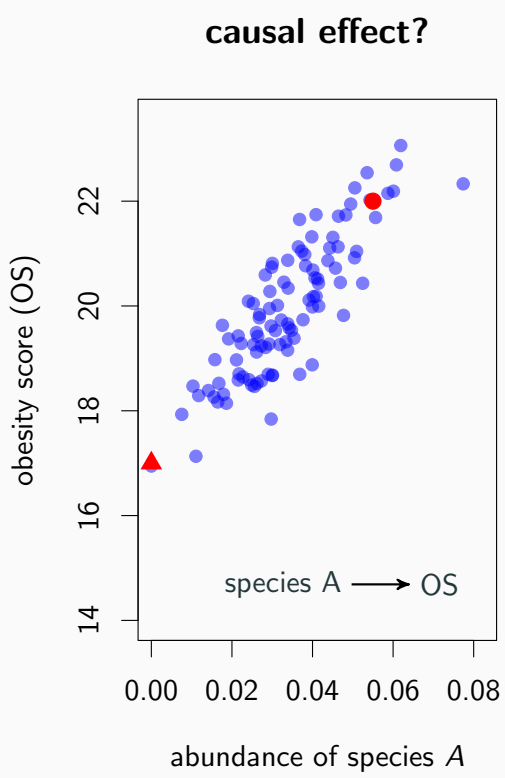

observations
obesity score (OS)
14
16
18
20
22
0.00
0.02
0.04
0.06
0.08
abundance of species A

Question: What if species A is removed for a specific mouse?

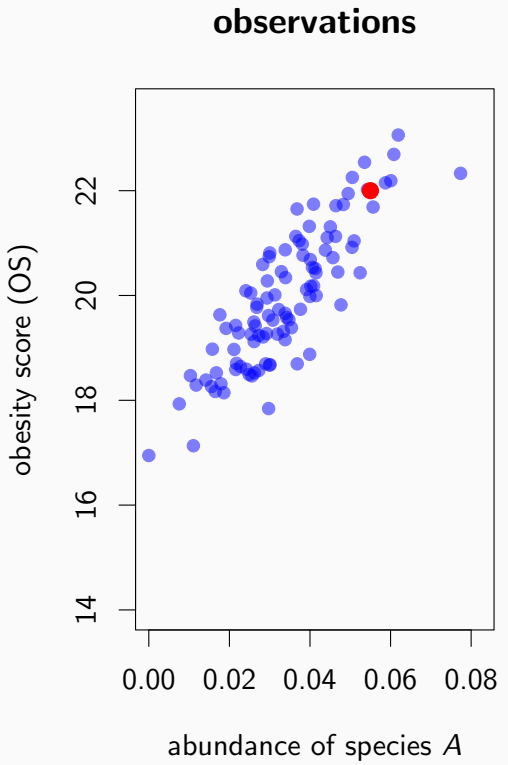


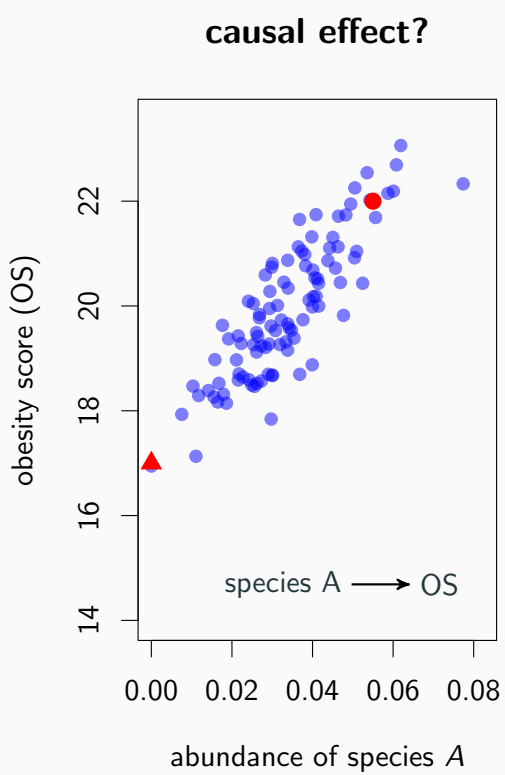
Question: What if species A is removed for a specific mouse?

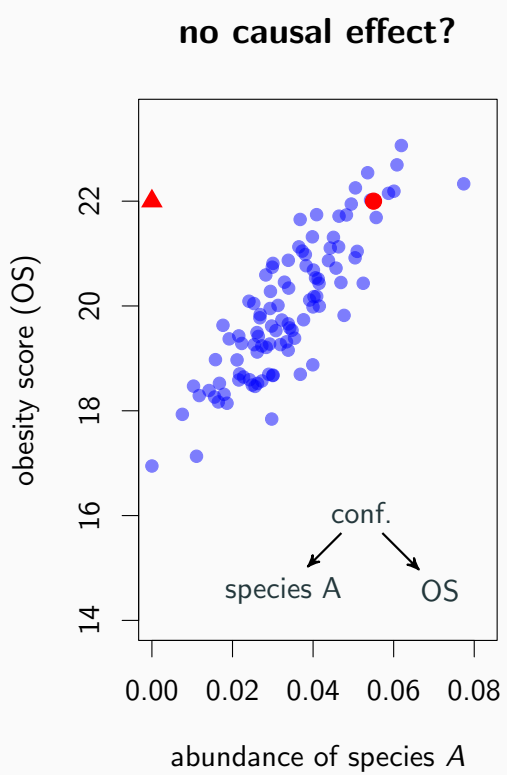


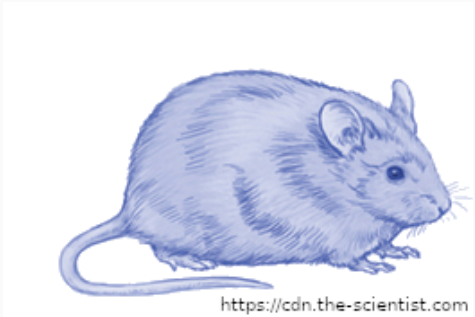


Question: What if species A is removed for a specific mouse?

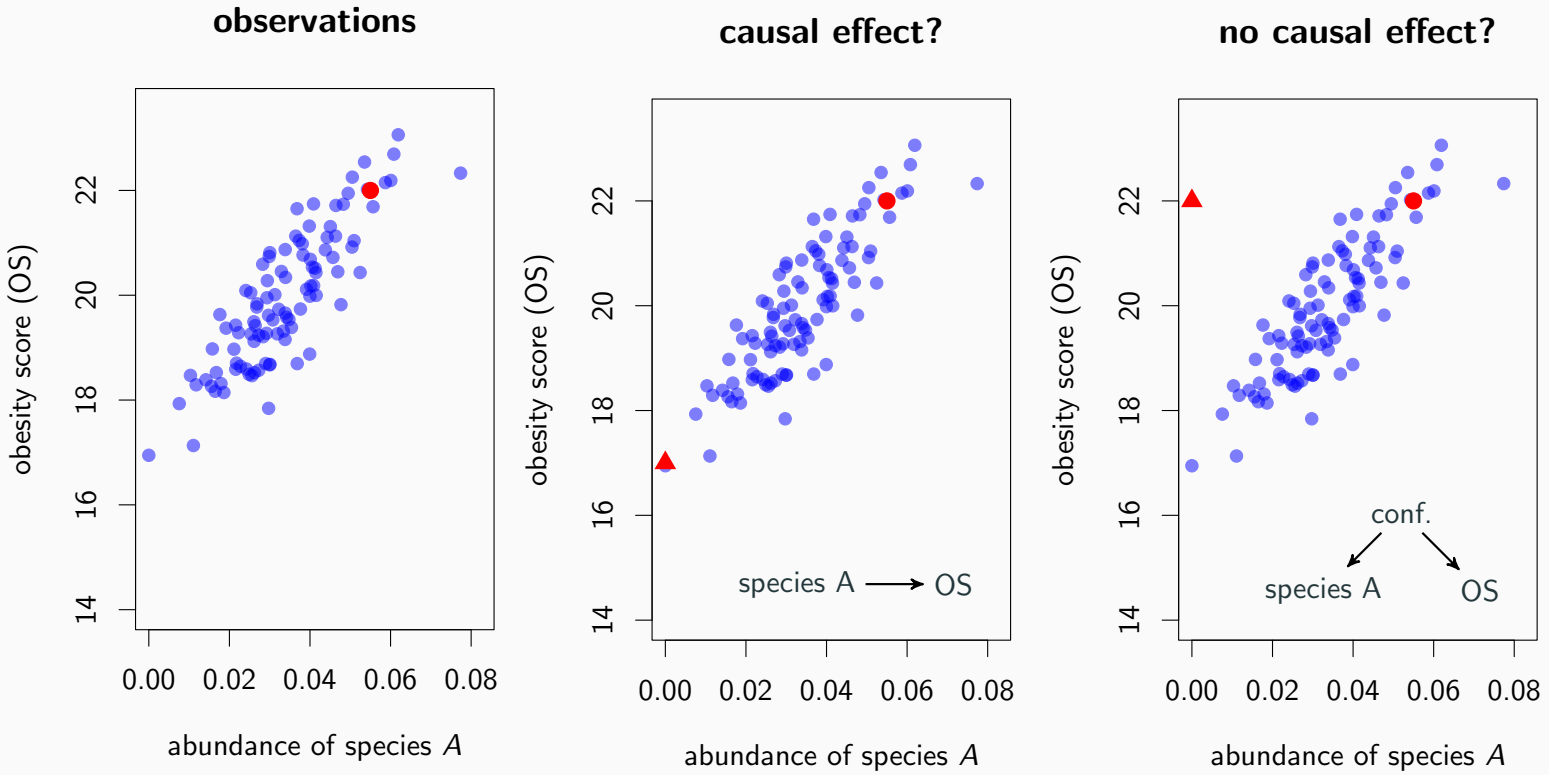








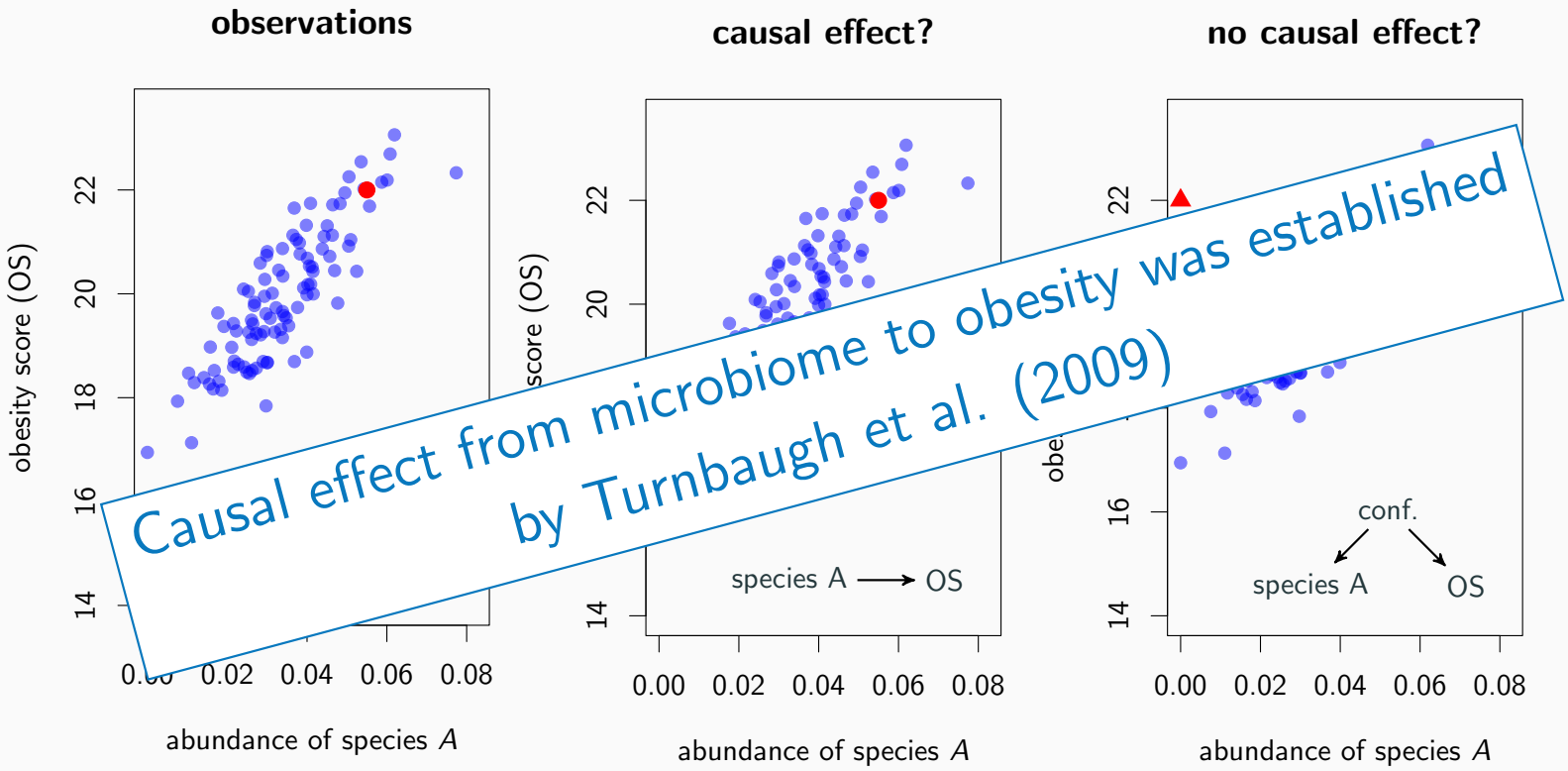
Question: What if species A is removed for a specific mouse?



Only causal models can answer these types of questions!

correlation $\neq$ causation

Question: What if species A is removed for a specific mouse?



Only causal models can answer these types of questions!

correlation $\neq$ causation

Causal models describe both:

- the observation distribution

  What can we say about new observations from the same system?

- the relevant intervention distributions

  What happens if we intervene on (change) the system?

Causal models describe both:

- the observation distribution

  What can we say about new observations from the same system?
- the relevant intervention distributions

  What happens if we intervene on (change) the system?

Two main types of models exist:

- Potential outcome models (Imbens and Rubin (2015))
  - starts from a set of interventions
- Structural causal models (Pearl (2009))
  - starts from the causal mechanism

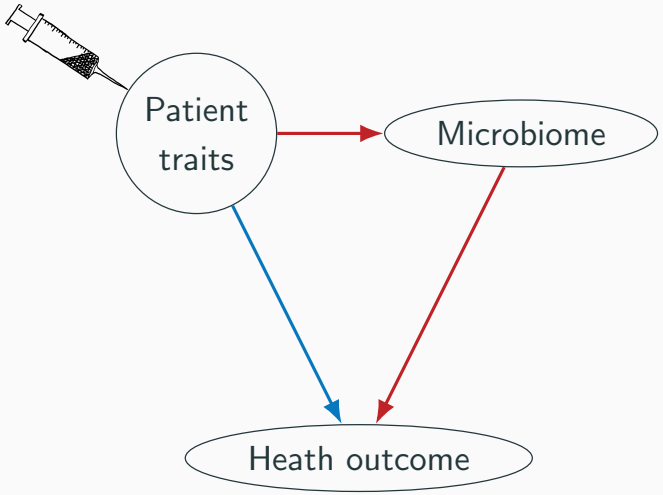
$\rightarrow$ Equivalent, but provide different perspective/focus.

Two challenges with microbiome data

Model specification and confounding

Challenge 1: Model specification

Example: Microbiome-mediated effect of drugs. How can we decompose the causal effects of a drug?



Challenge 1: Model specification

At what level should we model causal mechanisms?

Micobe level Composition level Proxy level

Challenge 1: Model specification

At what level should we model causal mechanisms?

Micobe level

- individual microbes are modeled
- Problems:
  - highly complex
  - hard to collect data
- Advantages:
  - fully captures the mechanism

Composition level

Proxy level

Challenge 1: Model specification

At what level should we model causal mechanisms?

Micobe level

- individual microbes are modeled
- Problems:
  - highly complex
  - hard to collect data
- Advantages:
  - fully captures the mechanism

Composition level

- relative abundances of microbes are modeled
- Problems:
  - not all effects are captured
  - tricky to analyze
- Advantages:
  - reduced complexity
  - corresponds to measurements

Proxy level

Challenge 1: Model specification

At what level should we model causal mechanisms?

Micobe level

- individual microbes are modeled
- Problems:
  - highly complex
  - hard to collect data
- Advantages:
  - fully captures the mechanism

Composition level

- relative abundances of microbes are modeled
- Problems:
  - not all effects are captured
  - tricky to analyze
- Advantages:
  - reduced complexity
  - corresponds to measurements

Proxy level

- proxies of microbiome are modeled
- Problems:
  - requires meaningful proxies
  - no interactions between microbes
- Advantages:
  - easy to integrate in causal analysis

Challenge 1: Model specification

At what level should we model causal mechanisms?

Micobe level

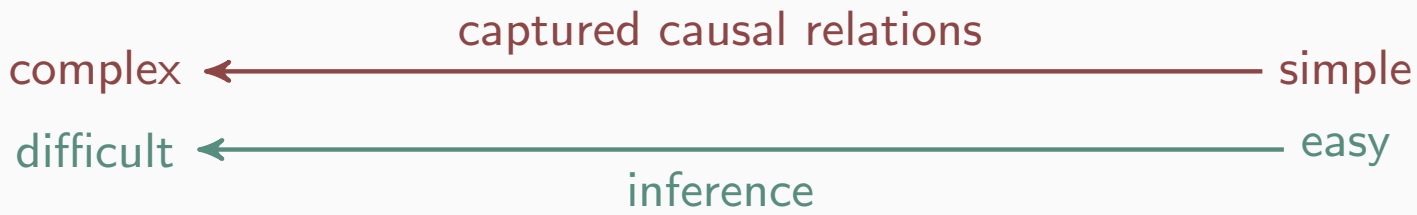
- individual microbes are modeled
- Problems:
  - highly complex
  - hard to collect data
- Advantages:
  - fully captures the mechanism

Composition level

- relative abundances of microbes are modeled
- Problems:
  - not all effects are captured
  - tricky to analyze
- Advantages:
  - reduced complexity
  - corresponds to measurements

Proxy level

- proxies of microbiome are modeled
- Problems:
  - requires meaningful proxies
  - no interactions between microbes
- Advantages:
  - easy to integrate in causal analysis



Compositional level

Simple mechanistic model

- $Y \in \mathbb{R}$ response of interest
- $X = (X^1, \ldots, X^n) \in \mathbb{R}^d$ predictor variables
- Structural causal model encodes the relation of X on Y,

$$Y = f(X, \epsilon) \quad \text{with} \quad X \perp\!\!\!\perp \epsilon$$

Compositional level

Simple mechanistic model

- $Y \in \mathbb{R}$ response of interest
- $X = (X^1, \ldots, X^n) \in \mathbb{R}^d$ predictor variables
- Structural causal model encodes the relation of X on Y,

$$Y = f(X, \epsilon) \quad \text{with} \quad X \perp\!\!\!\perp \epsilon$$

Goal: Learn causal effects corresponding to interventions that modularly increase individual coordinates, i.e.,

$$I^j(x_0) := \mathbb{E}[\tfrac{\partial}{\partial x^j} f(x_0, \epsilon)]$$

$\rightarrow$ Similarly, averaged effect $I^j := \mathbb{E}[\tfrac{\partial}{\partial x^j} f(X, \epsilon)]$.

Compositional level

Simple mechanistic model

- $Y \in \mathbb{R}$ response of interest
- $X = (X^1, \ldots, X^n) \in \mathbb{R}^d$ predictor variables
- Structural causal model encodes the relation of X on Y,

$$Y = f(X, \epsilon) \quad \text{with} \quad X \perp\!\!\!\perp \epsilon$$

Goal: Learn causal effects corresponding to interventions that modularly increase individual coordinates, i.e.,

$$I^j(x_0) := \mathbb{E}[\tfrac{\partial}{\partial x^j} f(x_0, \epsilon)]$$

$\rightarrow$ Similarly, averaged effect $I^j := \mathbb{E}[\tfrac{\partial}{\partial x^j} f(X, \epsilon)]$.

Intuition: Captures infinitesimal intervention on variable X^j.

Compositional level

Simple mechanistic model

- $Y \in \mathbb{R}$ response of interest
- $X = (X^1, \ldots, X^n) \in \mathbb{R}^d$ predictor variables
- Structural causal model encodes the relation of X on Y

Problem: Individually changing a single coordinate in compositional vector is meaningless.

Goal: Learn causal effects corresponding to interventions that modularly increase individual coordinates, i.e.,

$$I^j(x_0) := \mathbb{E}[\frac{\partial}{\partial x^j} f(x_0, \epsilon)]$$

$\rightarrow$ Similarly, averaged effect $I^j := \mathbb{E}[\frac{\partial}{\partial x^j} f(X, \epsilon)]$.

Intuition: Captures infinitesimal intervention on variable X^j.

Compositional level

Simple compositional mechanistic model

- $Y \in \mathbb{R}$ response of interest
- $Z = (Z^1, \ldots, Z^n) \in \Delta^d$ predictor variables, where $\Delta^d = \{z \in [0,1]^d \mid \sum_{j=1}^d z^j = 1\}$ is the simplex.
- Structural causal model encodes the relation of Z on Y,

$$Y = f(Z, \epsilon) \quad \text{with} \quad Z \perp\!\!\!\perp \epsilon$$

Goal: Learn causal effects corresponding to interventions that modularly increase individual coordinates relative to the others, i.e.,

$$I^j(z_0) := \mathbb{E}[\tfrac{\partial}{\partial c} f(\phi^j(z_0, c), \epsilon)\big|_{c=1}] \quad \text{(CFI)}$$

with $\phi^j(z, c) = (s_c z^1, \ldots, s_c z^{j-1}, c^j, s_c z^{j+1}, \ldots, s_c z^d)$.

Further considerations for model specification

- Complex data structure

  high-dimensional, sparse (many zeros), compositional

  - What types of statistical procedures apply?

- Integration of multiple data types

  prior knowledge (e.g., phylogenetic structure), host traits

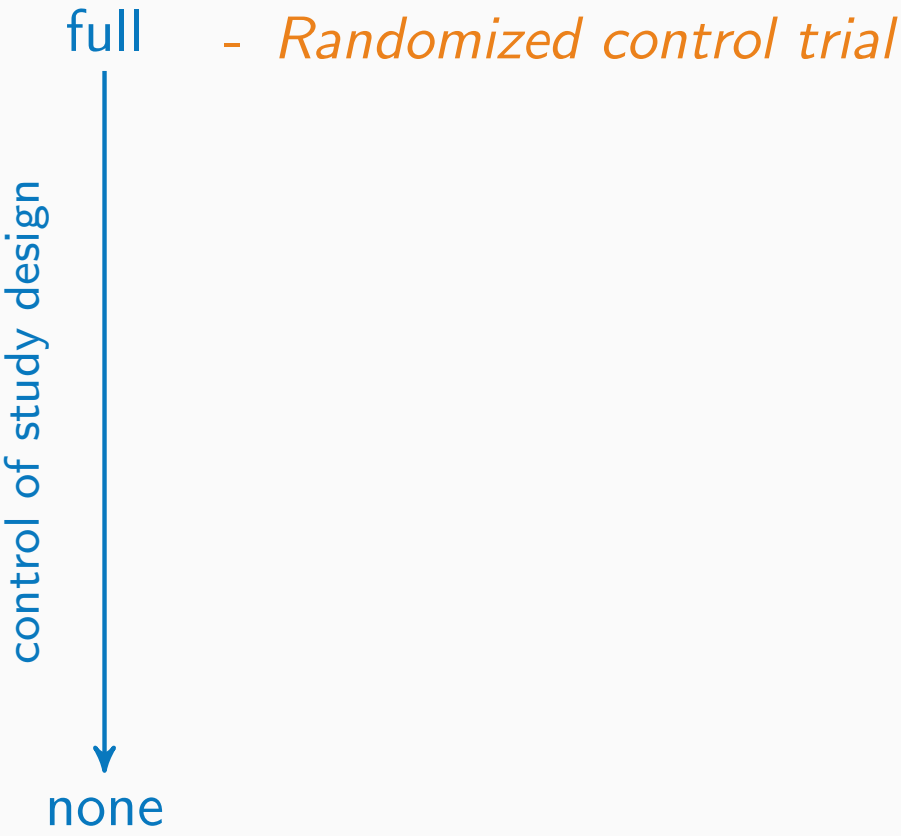
  - How should this be included in causal analysis?

- Volatile over time

  microbiome may oscillate over time

  - When do we need to account for time and when not?

Challenge 2: Confounding



Challenges: requires targeted interventions, expensive, difficult outside of the lab

Challenge 2: Confounding

- *Randomized control trial*

- *Procedures with no prior causal knowledge*

  (e.g., score- or constraint-based)

Challenges: model entire system, often under-identified, difficult to achieve uncertainty quantification

Challenge 2: Confounding

- *Randomized control trial*

- *Adjustment/inverse probability weighting*
  (e.g., many semiparametric approaches)

- *Procedures with no prior causal knowledge*
  (e.g., score- or constraint-based)

Challenges: lots of prior knowledge about causal structure, no unobserved confounding (ignorability)

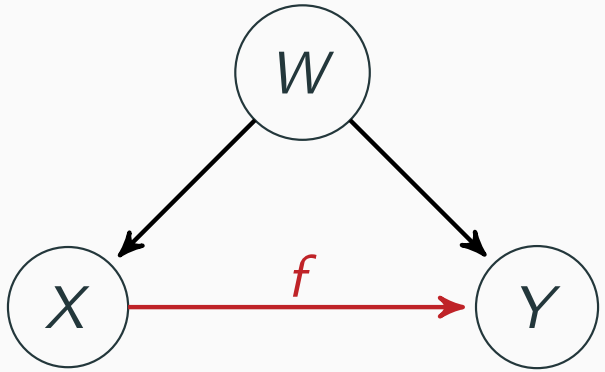
Challenge 2: Confounding

- *Randomized control trial*
- *Invariance based methods*
  (e.g., instrumental variables/Mendelian randomization)
- *Adjustment/inverse probability weighting*
  (e.g., many semiparametric approaches)
- *Procedures with no prior causal knowledge*
  (e.g., score- or constraint-based)

Challenges: some causal structure is assumed, requires heterogeneous study design

Causal inference with observed confounding

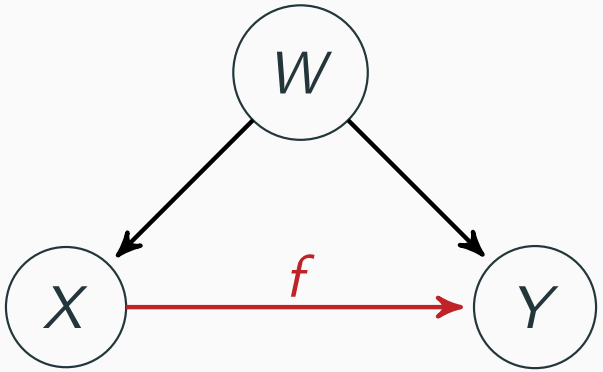
- response Y, e.g. health outcome
- treatment variable X, e.g., microbiome or individual microbe
- covariates W, e.g., age, diet



assumed causal structure

Causal inference with observed confounding

- response Y, e.g. health outcome
- treatment variable X, e.g., microbiome or individual microbe
- covariates W, e.g., age, diet



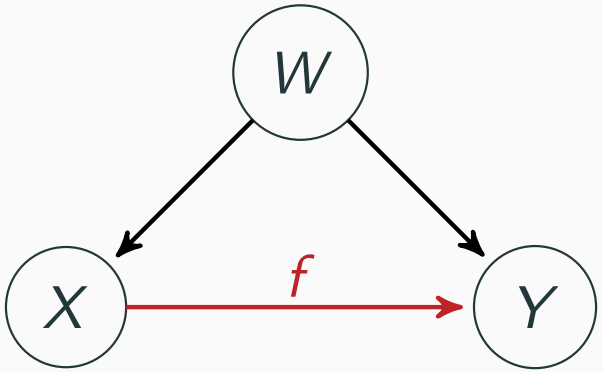
assumed causal structure

Goal: Learn the function f.

- Adjustment: Learn f by regressing Y on X and W jointly and marginalizing out W (outcome model)
- Inverse probability weighting: Learn $X|W$ and use it to reweight data so that edge $W \to X$ is removed (propensity model)

Causal inference with observed confounding

- response Y, e.g. health outcome
- treatment variable X, e.g., microbiome or individual microbe
- covariates W, e.g., age, diet



assumed causal structure

Goal: Learn the function f.

- Adjustment: Learn f by regressing Y on X and W jointly and marginalizing out W (outcome model)
- Inverse probability weighting: Learn $X|W$ and use it to reweight data so that edge $W \to X$ is removed (propensity model)

Next talk: Combines both to achieve parametric rates!

Conclusions

- Microbial communities are important parts of biological systems.
- Research questions in microbiome sciences are causal in nature.
- Two challenges when learning causal effects:
  (1) Model specification: compositional (+ other structure)
  (2) Confounding: Requires causal methods

Conclusions

- Microbial communities are important parts of biological systems.
- Research questions in microbiome sciences are causal in nature.
- Two challenges when learning causal effects:
  (1) Model specification: compositional (+ other structure)
  (2) Confounding: Requires causal methods

- **Solutions?** Coming in next talk...

Conclusions

- Microbial communities are important parts of biological systems.
- Research questions in microbiome sciences are causal in nature.
- Two challenges when learning causal effects:
  (1) Model specification: compositional (+ other structure)
  (2) Confounding: Requires causal methods

- **Solutions?** Coming in next talk...

Thank you!

- Pearl, J. (2009). *Causality.* Cambridge university press.
- Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.
- Turnbaugh, P. J., Bäckhed, F., Fulton, L. and Gordon, J. I. (2008). *Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome.* Cell host and microbe, 3(4), 213-223.
- Huang, S., Ailer, E., Kilbertus, N and Pfister, N. (2022). *Supervised Learning and Model Analysis with Compositional Data.* Preprint: https://arxiv.org/abs/2205.07271.