A primer on causal inference

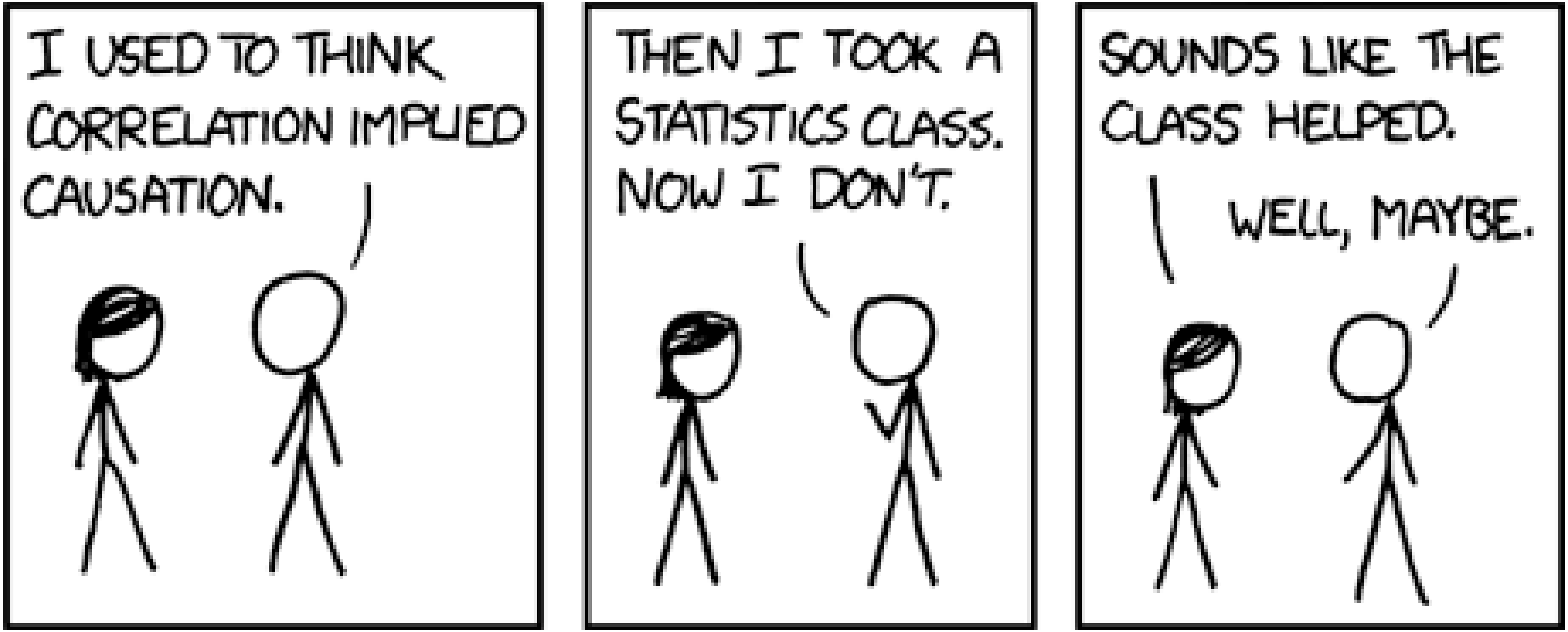
Kristen Hunter
UNSW School of Mathematics and Statistics

18 September 2025

A bit about me

- Lecturer in Statistics and Data Science at UNSW since 2022.
  - Teaching: intro stats, regression, inference, data visualisation.
  - Worked as a data scientist and software engineer.
  - PhD in statistics from Harvard.
- Main research interests: experimental design, causal inference.
  - Applications: environmental policy, education, health.

Causal inference attempts to estimate the causal effect of an intervention on a particular outcome. Causal inference involves the careful design and analysis of experiments and observational studies.



Potential outcomes framework

Image: Flaticon.com

I have a headache, so I take a medication.

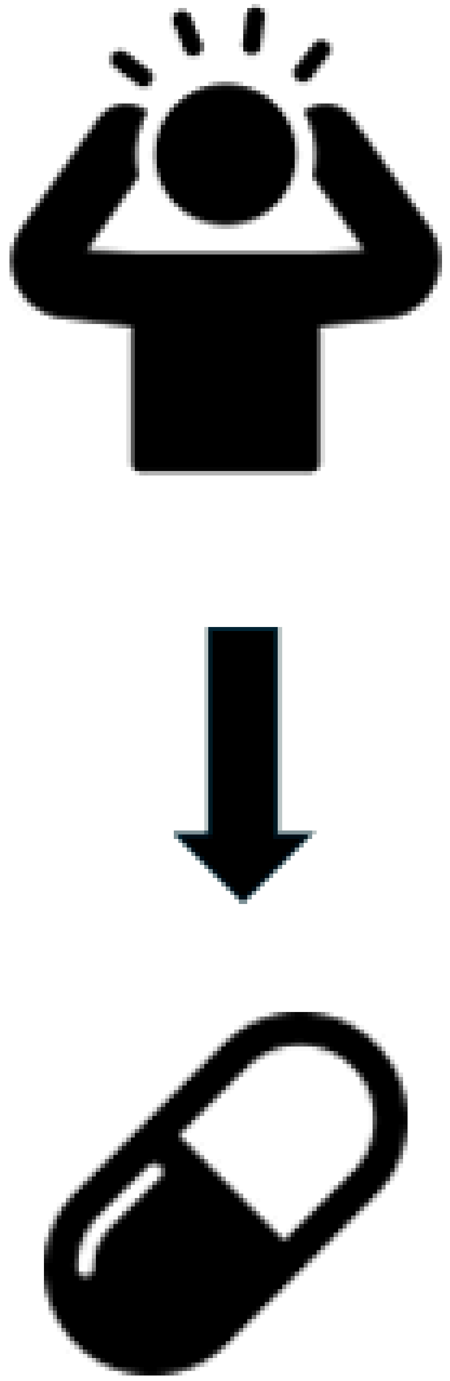


Image: Flaticon.com

An hour later, my headache is gone (so I do some yoga.)

Image: Flaticon.com

Did the medication cause my headache to go away?

This is an ill-defined causal question!

Image: Flaticon.com

In causal inference, we always ask: compared to what?

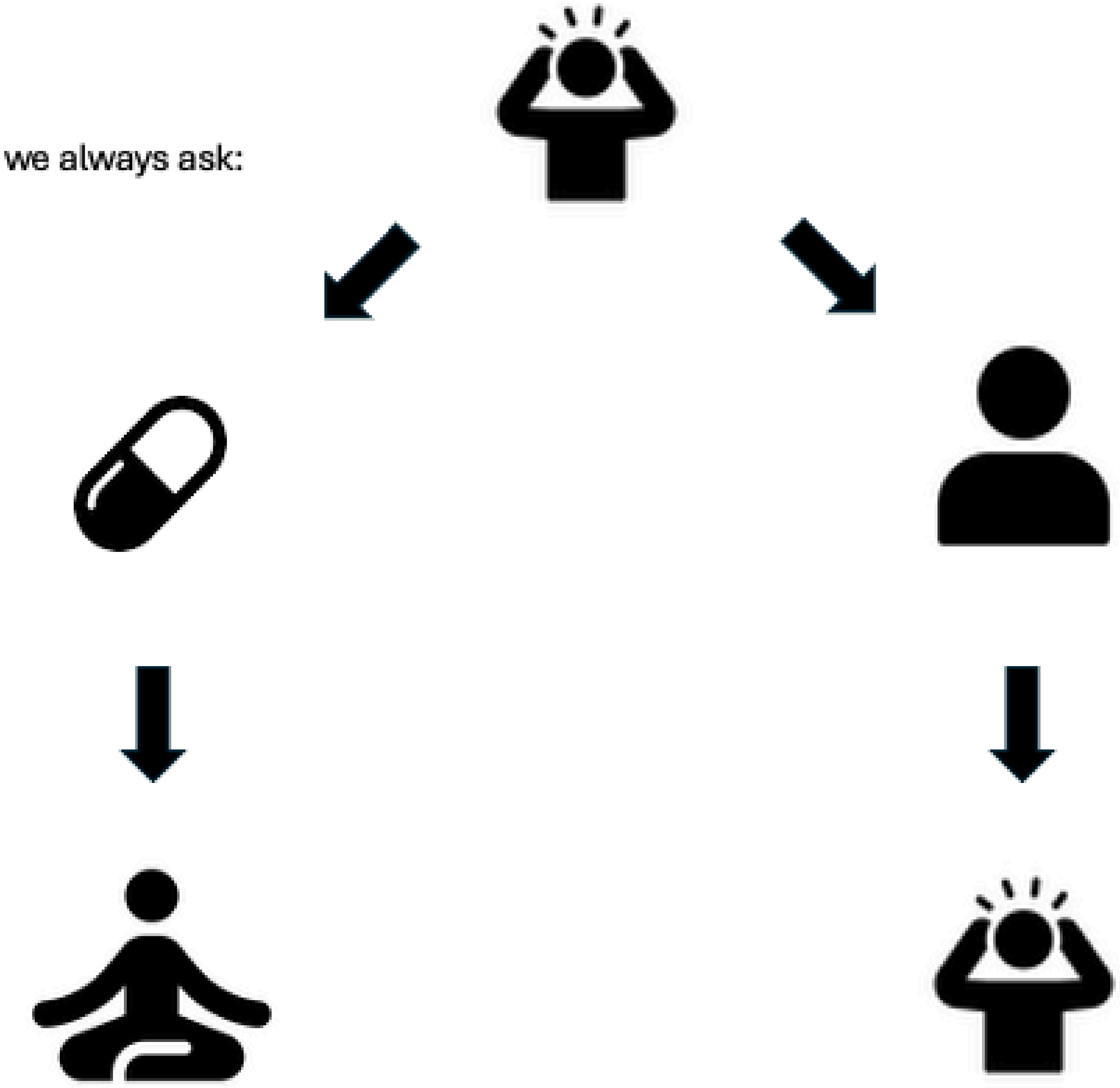


Image: Flaticon.com

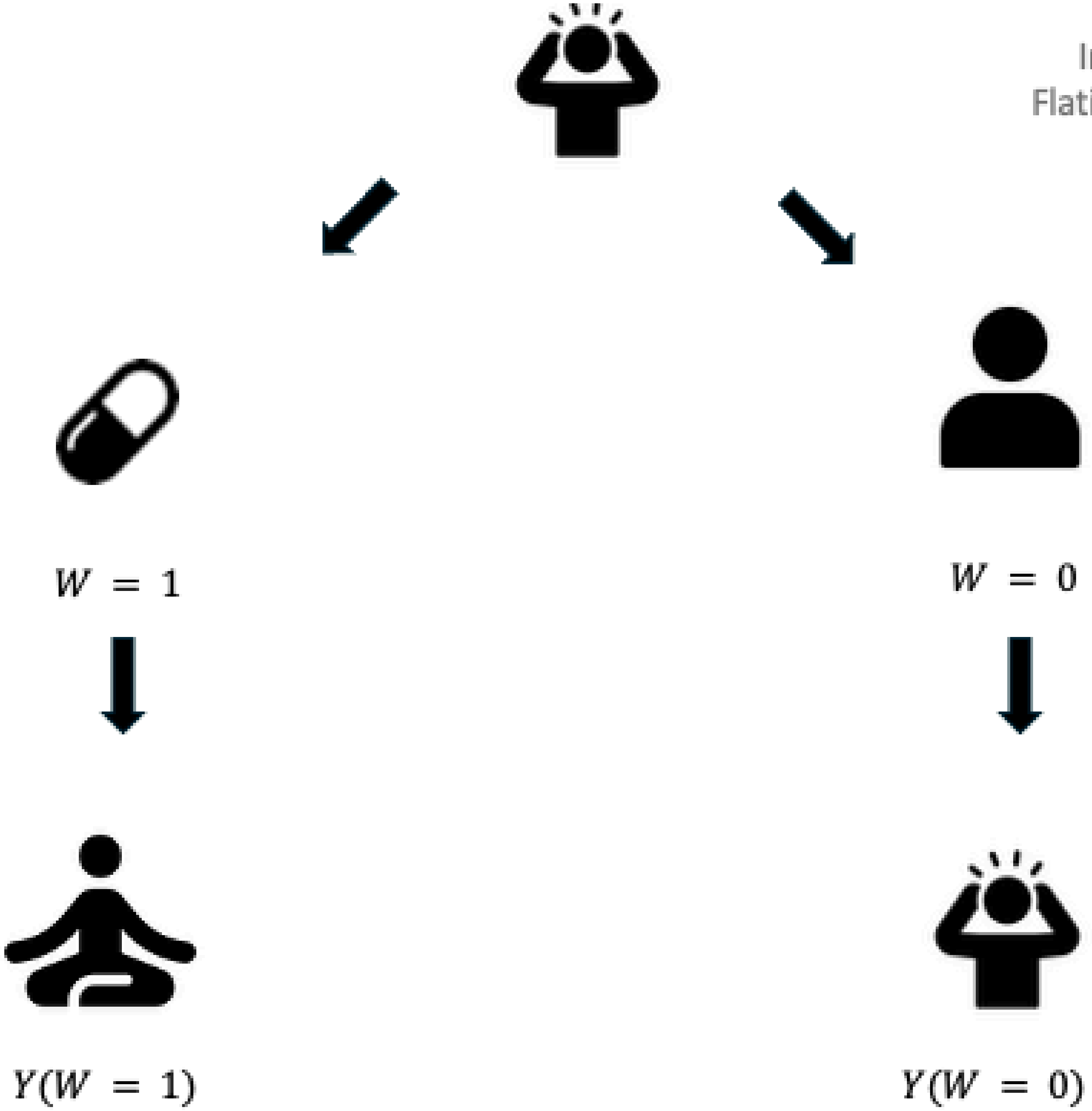
Causal effect

Image: Flaticon.com

Image:
Flaticon.com

Treatment/Intervention
W

Potential outcomes
$Y(W)$



Causal effect

$Y(W = 1)$ $Y(W = 0)$

Image: Flaticon.com

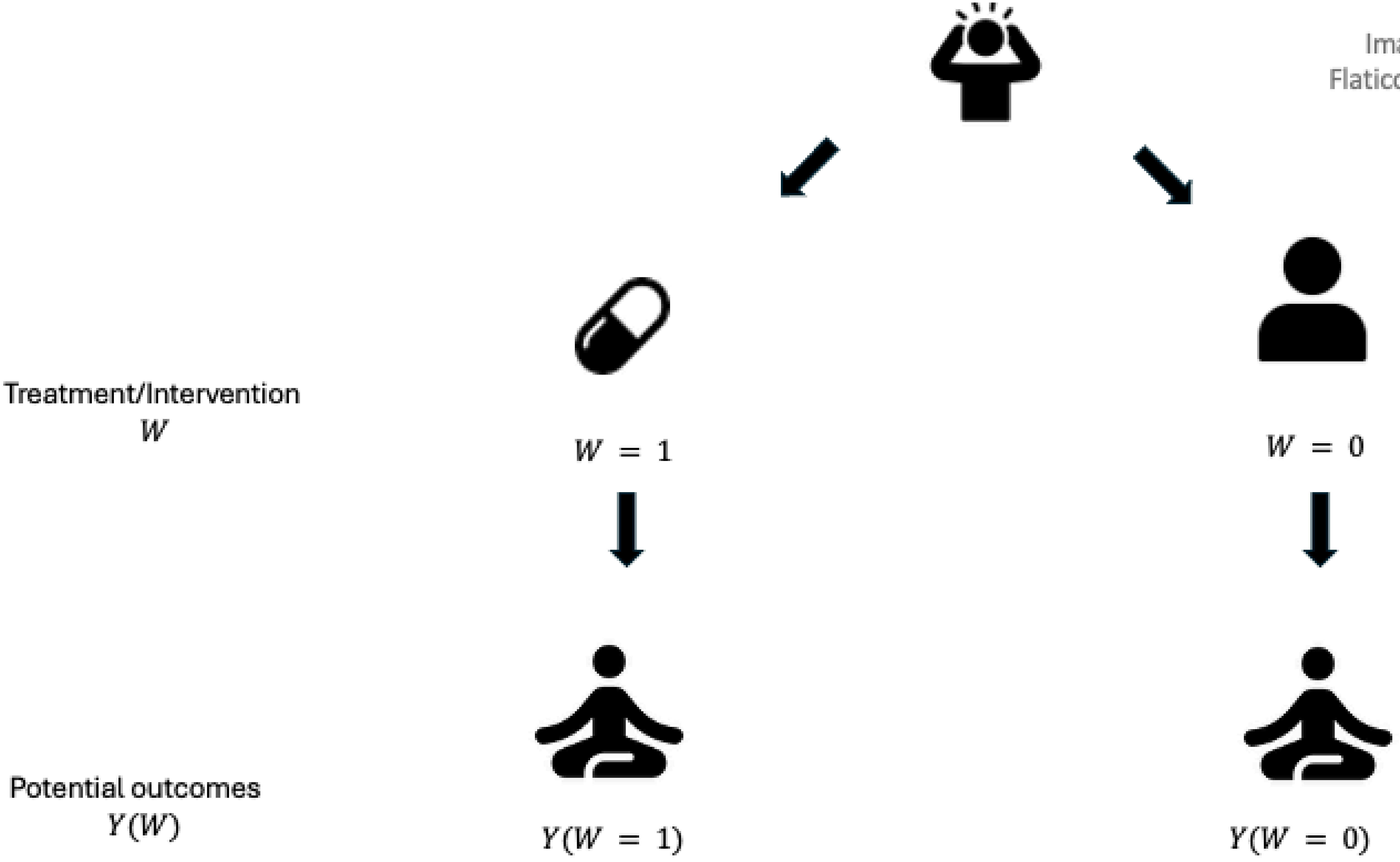


Image: Flaticon.com

Causal effect

$Y(W = 1)$ $Y(W = 0)$

Image: Flaticon.com

Fundamental problem

Fundamental problem

The fundamental problem of causal inference is that we only only observe (at most) one potential outcome.

Causal inference is a **missing data problem.**

Moving to many units

- How do we solve the fundamental problem of causal inference?
  - We don't know the missing value!
- Instead of thinking about individual-level treatment effects, focus on the **average** treatment effect (ATE).

$$\begin{aligned}\tau &= \bar{Y}(1) - \bar{Y}(0) \\ &= \frac{1}{N}\sum_{i=1}^{N}[Y_i(1) - Y_i(0)]\end{aligned}$$

Moving to many units

Average causal effect

$\bar{Y}(W = 1)$

$-$

$\bar{Y}(W = 0)$

Image: Flaticon.com

Categories of causal studies

Randomised experiments

- Treatment assignment is controlled by the researcher.
- Clinical trials, education experiments, psychology, industrial experiments, agriculture, A/B testing in tech.
- Treatment groups tend to have similar covariates. (We sometimes call this balance.)
- Causal inference is straightforward.
- We still want to take care in design and analysis to get best possible answers.

Observational studies

- Treatment assignment is not controlled by the researcher.
- Epidemiology, criminology, political science, economics.
- Treatment groups usually do not have similar covariates X.
- Causal inference is difficult (sometimes impossible).
- Careful design & analysis is required.

Factorial experiments in environmental studies

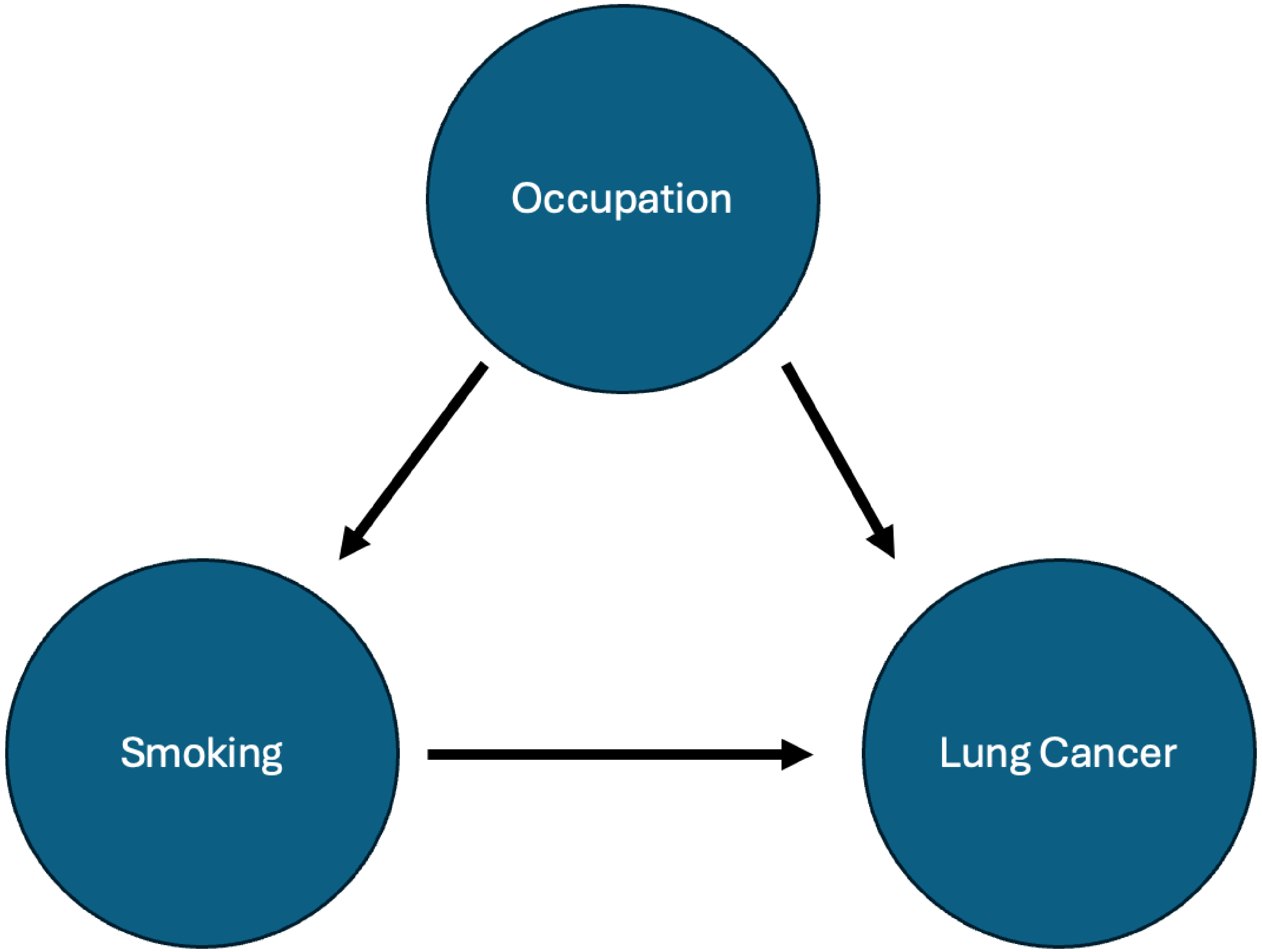
- What is the effect of different weed management strategies on ecosystem function? (Iddris et al. 2023)
  - 2^2 field experiment: high v. low fertilization, mechanical v. herbicide weeding
- What is the support level in the Global North for stricter global supply chain sustainability standards? (Kolcava, Smith, and Bernauer 2023)
  - Survey experiment to test 36 combinations of 3 different factors: scope, transparency, and enforcement.

Complex experiments in environmental studies

- Fractional factorial: I want to run a factorial experiment with many factors, but I don't have enough units for every possible combination.
  - Toxicity evaluation of 10 different microplastics to aquatic organisms (Enyoh et al. 2022).
- Latin square: I want to reduce my variance/increase power by blocking on two potential sources of unwanted variation.
  - Evaluate the effectiveness of a bycatch reduction devices by blocking on net and tow (Burridge and Robins 2000).
- Split-plot design: My units fall into batches and I want to avoid ``bad'' randomisations.
  - Evaluate the effect of tillage on phosphorus leaching. For each parcel of land, randomly assign half of parcel to tillage and half to no tillage (Butler and Coale 2005).

Observational studies

Confounding: In observational data, background characteristics can influence both (1) which treatment a unit receives, and (2) its potential outcomes.



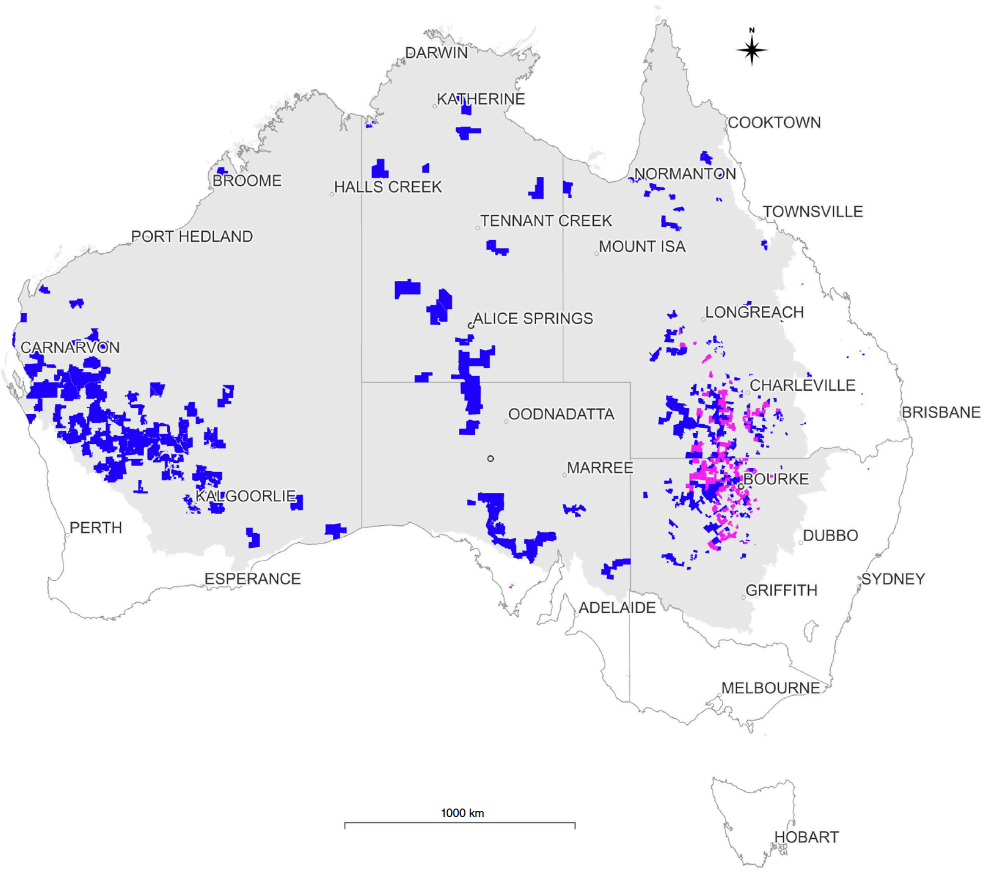
Matching

The goal of designing an observational study is to approximate a randomised experiment as closely as possible, and mitigate the effects of confounding. One approach is **matching**.

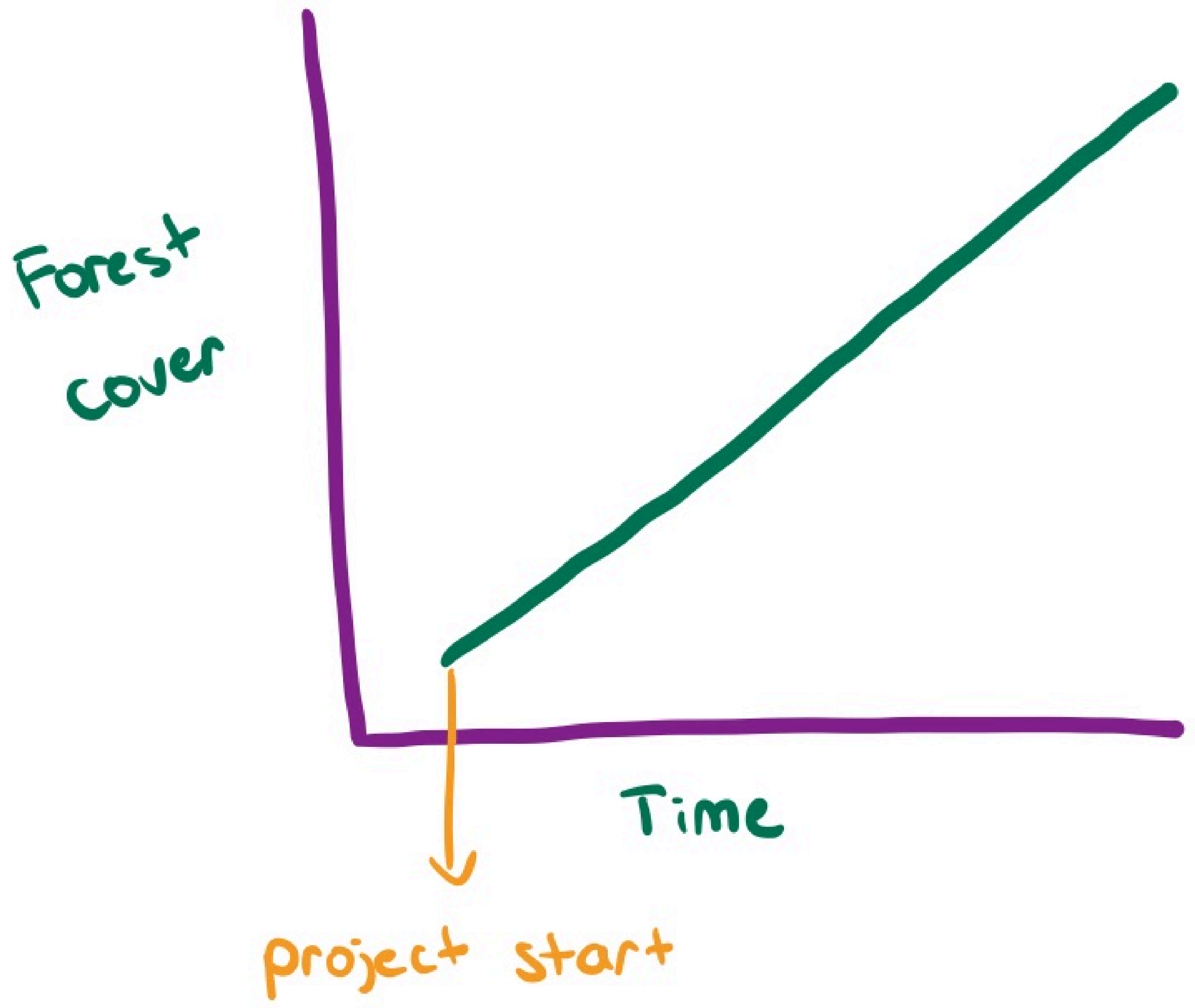
Case study: evaluating carbon offsets

Question: What is the impact of human-induced regeneration (HIR) carbon offset projects on forest cover? (Macintosh et al. 2024)

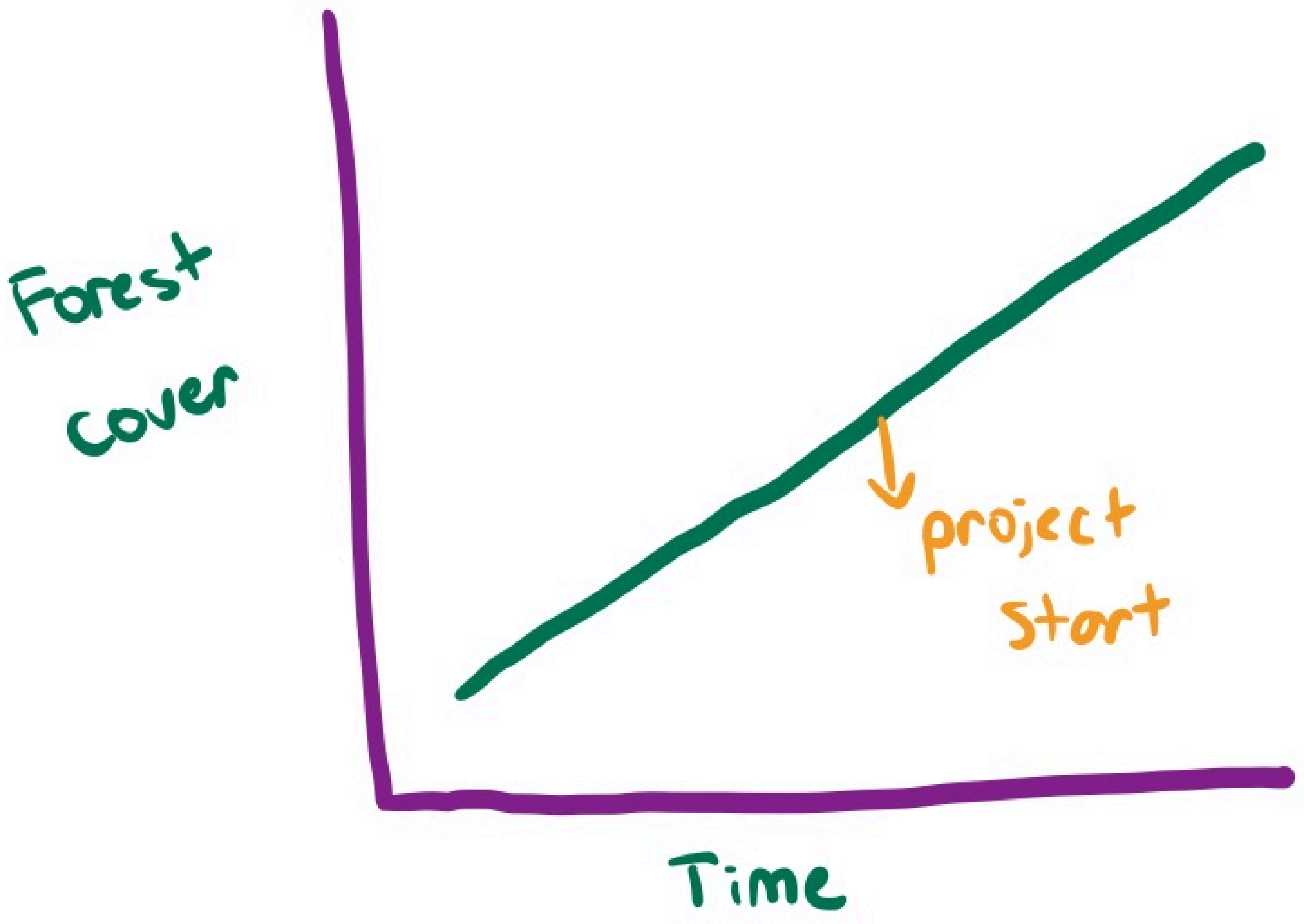
- HIR projects: regeneration of even-aged native forests through changes in land management on land that previously contained forest cover.
- 5th largest nature-based solution offset in the world by carbon credit issuances.



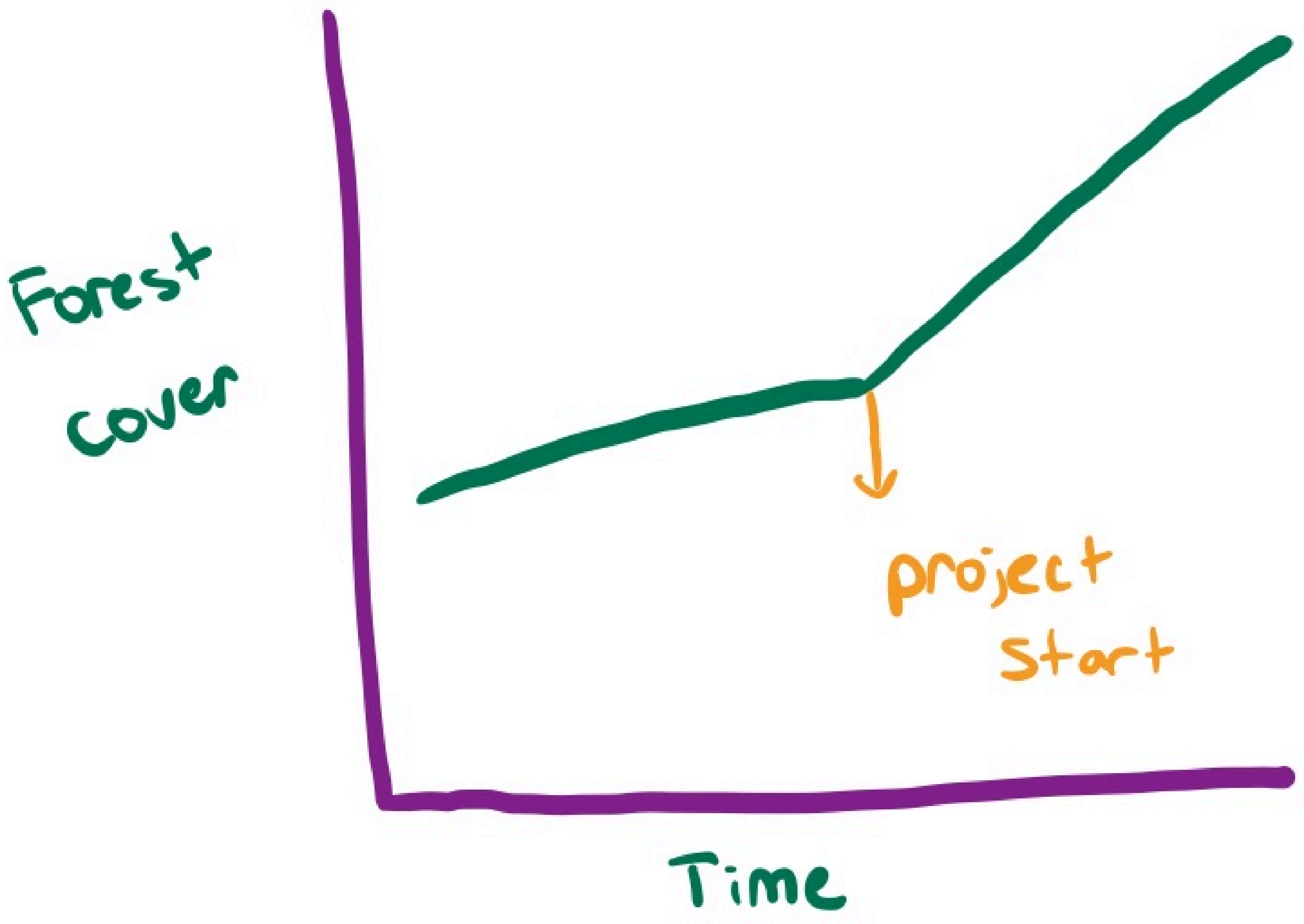
Scenario 1



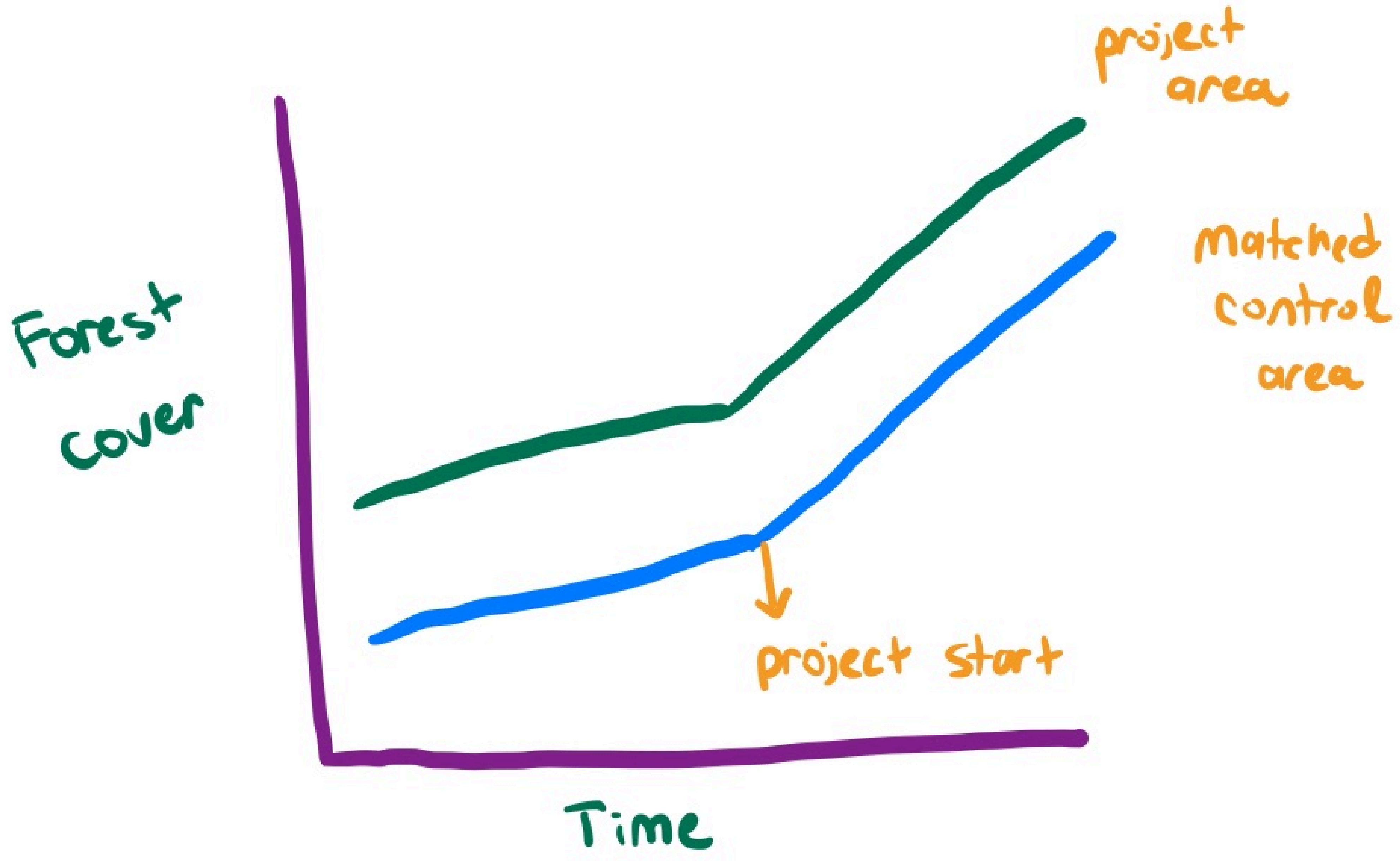
Scenario 2



Scenario 3



Scenario 4



project area
matched control area
Forest cover
project start
Time

Matching

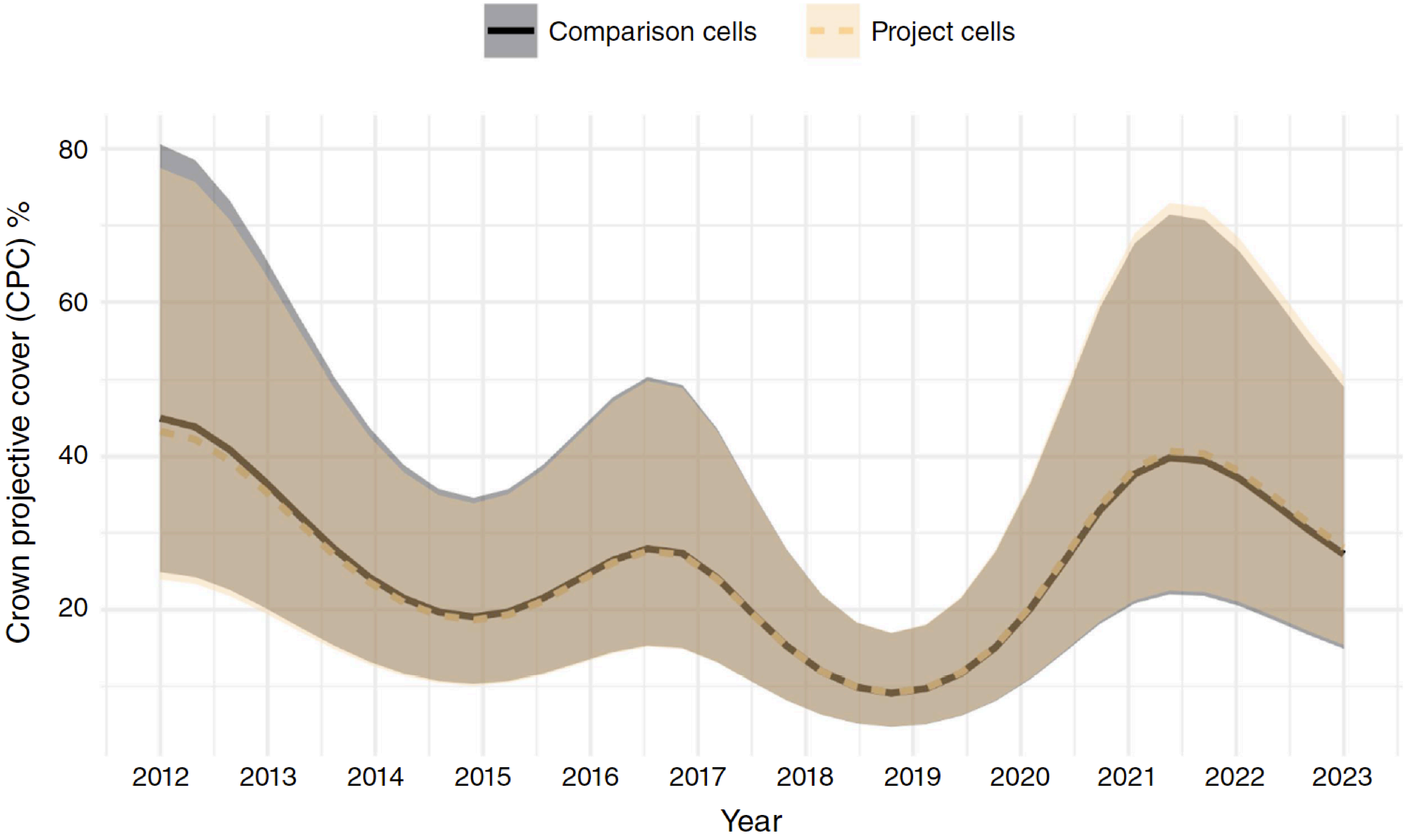
- **Matching goal**: for each pixel in a HIR project area, match it to a control unit not part of any HIR projects.
- **Procedure**: match based on soil, rainfall, biome, distance to project pixel, etc.
- **Result**: A good set of matches isolates the causal effect of the HIR land management.

Matching

Statistical interpretation:

- If we produce high quality matches, we can consider ourselves to approximate a randomised experiment.
- We can consider it a coin flip which unit happened to be in the HIR project, and which one happened to be outside the HIR project.
- We have controlled for all observed confounding variables.

Results for HIR projects



Some select challenges in causal inference

- Incorporating time series and spatial information into causal methods
- Matching based on high-dimensional data
- Mediation
- Adaptive experiments
- Noncompliance to treatment
- Continuous treatments
- Time-varying treatments
- Complex outcome response surfaces
- Interference between units

Thank you!

References

Burridge, CY, and JB Robins. 2000. “Benefits of Statistical Blocking Techniques in the Design of Gear Evaluation Trials: Introducing the Latin Square Design.” *FISHERIES RESEARCH* 47 (1): 69–79. https://doi.org/10.1016/S0165-7836(99)00125-3.

Butler, JS, and FJ Coale. 2005. “Phosphorus Leaching in Manure-Amended Atlantic Coastal Plain Soils.” *JOURNAL OF ENVIRONMENTAL QUALITY* 34 (1): 370–81.

Enyoh, Christian Ebere, Qingyue Wang, Prosper E. Ovuoraye, and Tochukwu Oluwatosin Maduka. 2022. “Toxicity Evaluation of Microplastics to Aquatic Organisms Through Molecular Simulations and Fractional Factorial Designs.” *CHEMOSPHERE* 308 (2).

Iddris, Najeeb Al-Amin, Greta Formaglio, Carola Paul, Volker von Groß, Guantao Chen, Andres Angulo-Rubiano, Dirk Berkelmann, et al. 2023. “Mechanical Weeding Enhances Ecosystem Multifunctionality and Profit in Industrial Oil Palm.” *Nature Sustainability* 6 (6): 683–95.

Kolcava, Dennis, E Keith Smith, and Thomas Bernauer. 2023. “Cross-National Public Acceptance of Sustainable Global Supply Chain Policy Instruments.” *Nature Sustainability* 6 (1): 69–80.

Macintosh, Andrew, Megan C Evans, Don Butler, Pablo Larraondo, Chamith Edirisinghe, Kristen B Hunter, Maldwyn J Evans, Dean Ansell, Marie Waschka, and David Lindenmayer. 2024. “Non-Compliance and Under-Performance in Australian Human-Induced Regeneration Projects.” *The Rangeland Journal* 46 (5).