Design: The Elusive Principle of Statistics

Qingyuan Zhao

Statistical Laboratory, University of Cambridge

March 5, 2024 @ University of Oxford

The concept of design

- In statistics, design is usually associated with randomized experiments.
  - R. A. Fisher's *Design of Experiments* (1935).
  - Parallel literature of survey sampling.
- There is growing awareness that design is a much broader concept.
- My tentative definition:

  Design = **HOW** data are **collected** and **preprocessed**.
- There is very little formal discussion about design in AI and machine learning, although it seems crucial in many trendy topics such as transfer learning, domain adaptation, distribution shift, generalizability, transportability, fairness, accountability, explanability...

This talk

Examples from my recent work, otherwise completely unrelated, echo a common theme:

Design trumps analysis.[1]

1. Meta-research of biodiversity conservation.
2. Early outbreak analysis of COVID-19 in China.
3. Racial discrimination in policing in New York City.
4. Mendelian randomization—a popular method in genetic epidemiology.
5. Graphical approaches to confounder selection in observational studies.

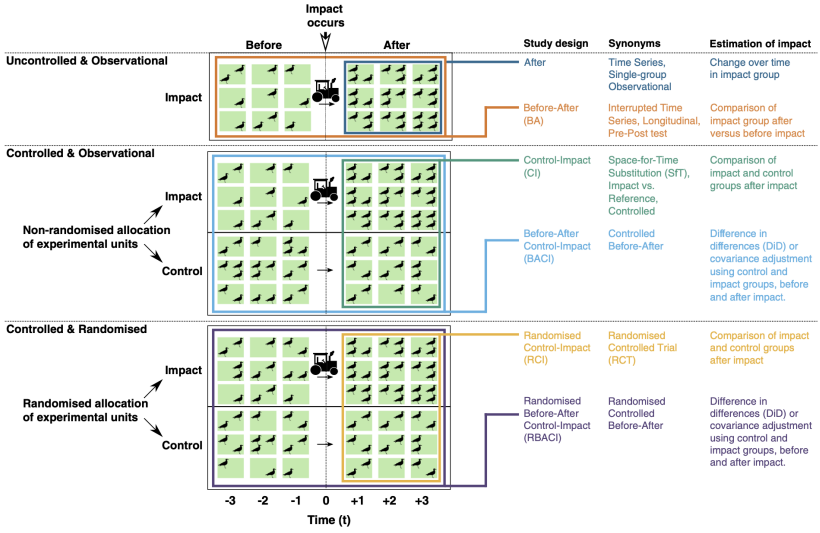
[1] This motto is taken from D. B. Rubin's 2008 article "For objective causal inference, design trumps analysis", although a much broader concept of "design" will be used here.

Example 1: Meta-research of biodiversity conservation

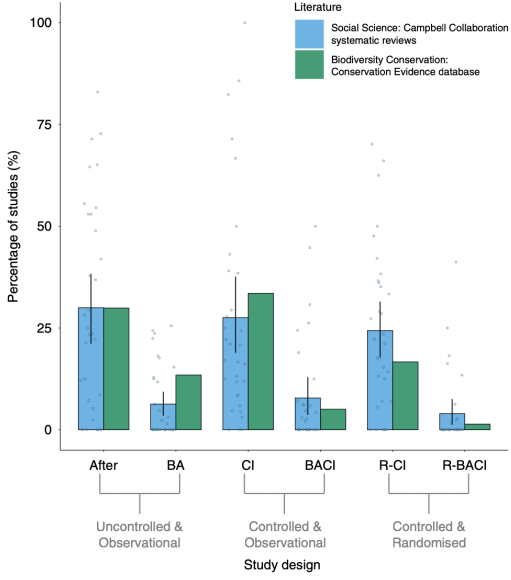
Based on:

- Alec P. Christie, ..., Qingyuan Zhao, and William J. Sutherland (2020). "Quantifying and Addressing the Prevalence and Bias of Study Designs in the Environmental and Social Sciences". In: *Nature Communications* 11.1, p. 6377. DOI: 10.1038/s41467-020-20142-y.
- This was a collaboration with ecologists at Cambridge who made painstaking effort to build a database of conservation studies.
- Check out their website: `https://www.conservationevidence.com/`.

Common study designs in ecology



Prevalence of study designs



Statistical model

We fitted a hierarchical Bayesian model for **within-study comparisons**.

- Data: $\hat{\beta}_{ij}$ effect size estimator in study i with design $j \in \{\text{BA}, \text{CI}, \text{BACI DiD}, \text{BACI CA}\}$,[2] with standard error $\hat{\beta}_{ij}$ obtained from a generalized linear (mixed) model.
- Bayesian model (prior is omitted):

$$\begin{aligned}
&\hat{\beta}_{ij} = \beta_i + \gamma_{ij} + \epsilon_{ij}, \\
&\beta_i \sim \text{N}(0, \sigma^2_\beta),\ \gamma_{ij} \sim \text{N}(0, \sigma^2_j),\ \epsilon_i \sim \text{N}(0, \Lambda), \\
&\Lambda = \lambda \cdot \text{diag}(\hat{\sigma}_i)\, \Omega\, \text{diag}(\hat{\sigma}_i).
\end{aligned}$$

- The most interesting parameters are σ^2_j, which describes the average magnitude of bias for design j.

[2] DiD = Difference in Differences; CA = Covariate Adjustment.

Results

Term	Posterior mean	95% Credible Interval
Randomised (R-)		
σ_β	0.746	[0.679, 0.813]
λ	1.119	[0.980, 1.276]
σ[BACI DiD]	0.029	[0.005, 0.097]
σ[BACI CA]	0.005	[0.002, 0.008]
σ[CI]	0.005	[0.002, 0.008]
σ[BA]	0.773	[0.699, 0.846]
Non-randomised		
σ_β	0.700	[0.628, 0.776]
λ	1.822	[1.595, 2.098]
σ[BACI DiD]	0.017	[0.004, 0.049]
σ[BACI CA]	0.049	[0.005, 0.128]
σ[CI]	0.091	[0.008, 0.137]
σ[BA]	0.645	[0.573, 0.720]

- Takeaway: having a control group makes a huge difference; randomization reduces bias.
- Limitations: strong assumptions on the bias and noise correlation.

Example 2: Early outbreak analysis of COVID-19 in China

Based on:

- Qingyuan Zhao, Yang Chen, and Dylan S Small (Feb. 2020). "Analysis of the Epidemic Growth of the Early 2019-nCoV Outbreak Using Internationally Confirmed Cases". In: DOI: 10.1101/2020.02.06.20020941.
- Qingyuan Zhao, Nianqiao Ju, Sergio Bacallado, and Rajen D. Shah (2021). "BETS: The Dangers of Selection Bias in Early Analyses of the Coronavirus Disease (COVID-19) Pandemic". In: *Annals of Applied Statistics* 15.1, pp. 363–390. DOI: 10.1214/20-aoas1401.
- Qingyuan Zhao (2022). "Small Data, Big Time—A Retrospect of the First Weeks of COVID-19 (with Discussion)". In: *Journal of the Royal Statistical Society (Series A, Statistics in Society)* 185.4, pp. 1793–1814. DOI: 10.1111/rssa.12874.

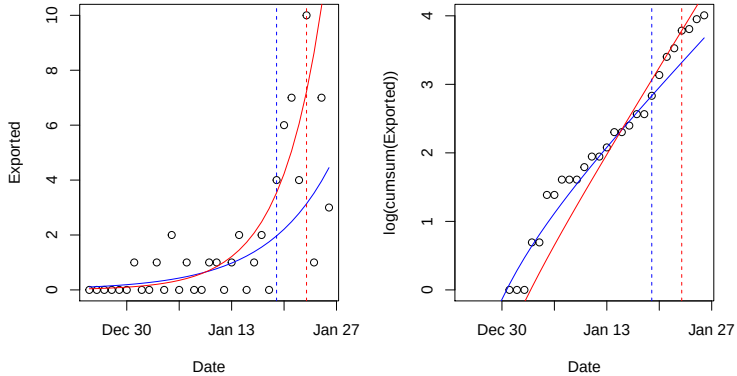
What went wrong?

- medRxiv paper: "Our estimated epidemiological parameters are higher than an earlier report using confirmed cases in Wuhan. This indicates the **2019-nCoV could have been spreading faster than previous estimates**."
- AoAS paper: "We gave a detailed illustration of why some early and highly influential analyses of the COVID-19 pandemic were **severely biased**."
- JRSSA paper: "Further reanalyses of some published COVID-19 studies show that the **epidemic growth was dramatically underestimated** by compartmental models, and the lack of fit **could have been clearly identified by simple data visualization**."

Naive method

- Joseph T. Wu, Kathy Leung, and Gabriel M. Leung (Feb. 2020). "Nowcasting and Forecasting the Potential Domestic and International Spread of the 2019-nCoV Outbreak Originating in Wuhan, China: A Modelling Study". In: *The Lancet* 395.10225, pp. 689–697. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(20)30260-9.
- They used a **SEIR (Susceptible-Exposed-Infectious-Recovered) model** for the epidemic in Wuhan and a **Poisson process** to model case exportation.
- They fitted the model using **17 (!) international cases** who showed symptoms before January 20, 2020.
- To replicate their analysis, I fitted some simple Poisson log-linear models.

Initial doubling time



- **Blue** (symptom onsets before January 20): 5.9 days (95% CI 3.4–15.7).
- **Red** (symptom onsets before January 24): 3.9 days (2.9–5.5).
- Original study: 6.4 days (5.8–7.1).

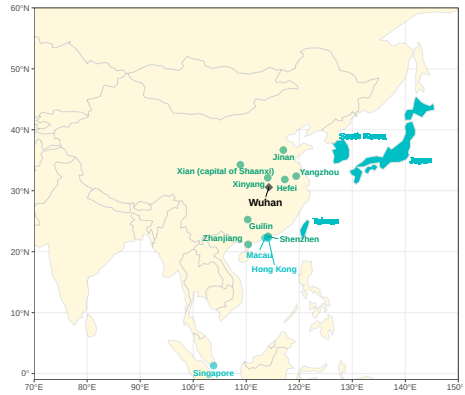
What else was wrong?

The model employed by Wu, Leung, and Leung (2020)

- does **NOT** take into account Wuhan's travel ban on January 23.
- ignores the rich information available for the individual cases.

Data collection

- Nianqiao Ju and I spent a lot of time collecting a total of 1,460 individual case reports.
- Confirmed before 29th Feb for mainland China and 15th Feb for international locations.
- We found 378 cases exported from Wuhan.



Overview of the dataset

Column name	Description	Example	Summary statistics
`Case`	Unique identifier for each case	HongKong-05	1460 in total
`Residence`	Nationality or residence of the case	Wuhan	21.5% reside in Wuhan
`Gender`	Gender	Male/Female	52.1%/47.7% (0.2% NA)
`Age`	Age	63	Mean=45.6, IQR=[34, 57]
`Known Contact`	Known epidemiological contact?	Yes/No	84.7%/15.3%
`Cluster`	Relationship with other cases	Husband of HongKong-04	32.1% known
`Outside`	Transmitted outside Wuhan?	Yes/Likely/No	58.5%/7.7%/33.8%
`Begin Wuhan`	**Begin of stay in Wuhan (B)**	30-Nov[4]	
`End Wuhan`	**End of stay in Wuhan (E)**	22-Jan	
`Exposure`	Period of exposure	1-Dec to 22-Jan	58.9% known period/date 8.2% known date
`Arrived`	Final arrival date at the location where confirmed a COVID-19 case	22-Jan	40.6% did not travel
`Symptom`	**Date of symptom onset (S)**	23-Jan	9.0% NA
`Initial`	Date of first medical visit	23-Jan	6.5% NA
`Confirmed`	Date confirmed	24-Jan	

Design

- The most time-consuming part was to **transcribe paragraphs of text to a data frame**.

  The other two cases are a married couple of residents of in Wuhan, a 62-year-old female [HongKong-04] and a 63-year-old male [HongKong-05], with good prior health conditions. Based on information provided by the patients, They took a high-speed train departing from Wuhan at 2:20pm, January 22, and arrived at the West Kowloon station around 8pm. The female patient had a fever since yesterday with no respiratory symptoms. The male patient started to cough yesterday and had a fever today. They went to the emergency department at the Prince of Wales Hospital yesterday and were admitted to the hospital for treatment in isolation. Currently their health conditions are stable. Respiratory samples of the two patients were tested positive for the novel coronavirus.

- Actually even more time-consuming and mentally defeated was to **define the columns**.

Generative BETS model (T = time of transmission)

$$f(b,e,t,s) = \underbrace{f_B(b)\cdot f_E(e \mid b)}_{\text{travel}}\cdot \underbrace{f_T(t \mid b,e)}_{\text{disease transmission}} \cdot \underbrace{f_S(s \mid b,e,t)}_{\text{disease progression}}.$$

To allow extrapolation, the BETS model makes two basic assumptions

Assumption 1: Disease transmission independent of travel

$$f_T(t \mid b,e) = \begin{cases} g(t), & \text{if } b < t < e, \\ 1 - \int_b^e g(x)\,dx, & \text{if } t = \infty. \end{cases}$$

Here $g(\cdot)$ models the **epidemic growth** in Wuhan before the lockdown.

Assumption 2: Disease progression independent of travel

$$f_S(s \mid b,e,t) = \begin{cases} \nu \cdot h(s-t), & \text{if } t < s < \infty, \\ 1-\nu, & \text{if } s = \infty. \end{cases}$$

Here $h(\cdot)$ is the density of the **incubation period** $S - T$ (for symptomatic cases).

Results

Location	Sample size	Doubling time (in days)	Incubation period	
			Median	95% quantile
China - Hefei	34	2.1 (1.2–3.7)	4.3 (2.9–6.0)	12.0 (9.1–17.3)
China - Shaanxi	53	1.7 (1.0–2.8)	4.5 (3.1–6.2)	14.6 (11.5–19.8)
China - Shenzhen	129	2.2 (1.7–3.0)	3.5 (2.8–4.3)	11.2 (9.5–13.6)
China - Xinyang	74	2.3 (1.5–3.5)	6.8 (5.4–8.2)	16.4 (13.8–20.1)
China - Other	42	2.0 (1.1–3.4)	5.1 (3.6–6.7)	12.3 (9.8–16.4)
International	46	2.1 (1.4–3.4)	3.8 (2.5–5.3)	10.9 (8.4–15.1)
All locations	378	2.1 (1.8–2.5)	4.5 (4.0–5.0)	13.4 (12.2–14.8)

(Point estimates obtained by MLE. Confidence intervals obtained by inverting LRT.)

Takeaway

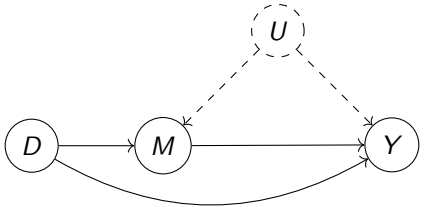
- The paper identifies five sources of biases: **under-ascertainment**, **non-random sample selection**, **travel quarantine**, **epidemic growth**, **right-truncation**.
- A really interesting experiment on estimating the incubation period: **upward bias** due to epidemic growth and **downward bias** due to right-truncation.

Example 3: Racial discrimination in policing in New York City

Based on

- Qingyuan Zhao, Luke J Keele, Dylan S Small, and Marshall M Joffe (2021). "A Note on Posttreatment Selection in Studying Racial Discrimination in Policing". In: *American Political Science Review* 116.1, pp. 337–350. DOI: 10.1017/s0003055421000654.
- This started from a friend telling me about a heated exchange on Twitter between two groups of political scientists, one taking the potential outcomes approach (Gaebler, Cai, Basse, Shroff, Goel, and Hill 2022) and one taking the graphical approach (Knox, Lowe, and Mummolo 2020).

Setup in Knox, Lowe, and Mummolo (2020)



- D: binary, 1 means minority.
- M: binary, 1 means police detainment.
- Y: binary, 1 means use of force.

Key challenges

1. Only observe data with $M = 1$ in police admin data.
2. There can be unmeasured M-Y confounders.

$\Longrightarrow$ Collider bias (when conditioning on $M = 1$) in previous studies.

Our reanalysis of the NYPD stop-and-frisk dataset

- New identification formula:

$$\text{Causal risk ratio} = \frac{\mathbb{E}[Y(1)]}{\mathbb{E}[Y(0)]} = \underbrace{\frac{\mathbb{E}[Y \mid D=1, M=1]}{\mathbb{E}[Y \mid D=0, M=1]}}_{\text{naive estimator}} \cdot \underbrace{\left\{\frac{\mathbb{P}(D=1 \mid M=1)}{\mathbb{P}(D=0 \mid M=1)}\right\} \Big/ \left\{\frac{\mathbb{P}(D=1)}{\mathbb{P}(D=0)}\right\}}_{\text{selection bias factor}}.$$

- We estimated $\mathbb{P}(D=1)$ using two external surveys (CPS and PPCS).

External dataset	Estimated risk ratio	95% Confidence interval
	Naive estimator	
None	1.29	1.28–1.30
	Adjusted for selection bias	
CPS	13.6	12.8–14.3
PPCS	32.3	31.3–33.3
PPCS (Large Metro)	16.7	15.4–18.4

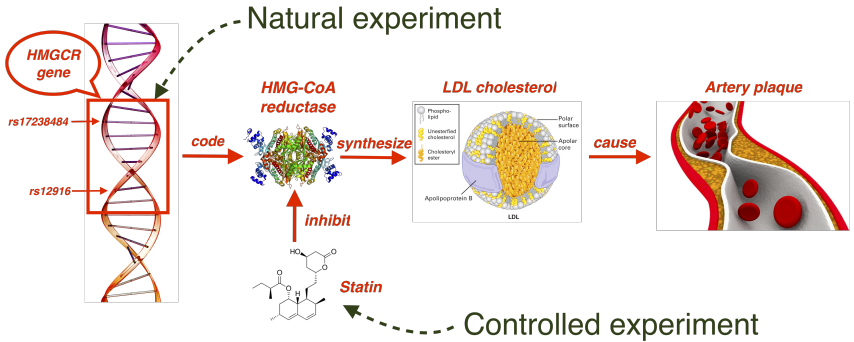
- The selection bias could be $>$ **10-fold!!**

Example 4: Mendelian randomization

Based on

- Qingyuan Zhao, Jingshu Wang, Gibran Hemani, Jack Bowden, and Dylan S. Small (2020). "Statistical Inference in Two-Sample Summary-Data Mendelian Randomization Using Robust Adjusted Profile Score". In: *Annals of Statistics* 48.3, pp. 1742–1769. DOI: 10.1214/19-aos1866.
- Matthew J Tudball, George Davey Smith, and Qingyuan Zhao (2022). "Almost Exact Mendelian Randomization". In: arXiv: 2208.14035 [stat.ME].
- This is an **extremely popular** design in genetic epidemiology that leverages the natural "experiment" in genetic inheritance.

Illustration: Causal effect of the LDL-cholesterol



- Basic idea: People who inherited certain alleles of *rs17238484* and *rs12916* have **naturally lower LDL cholesterol that mimic the effect of statin**.

Population-based Mendelian randomization (MR)

- MR can be understood as using genetic variants as **instrumental variables**.
- Issue: Pleiotropy (same gene affects multiple phenotypes).

Model for MR using summary data of genome-wide association studies (GWAS)

- $\hat{\gamma}_j \overset{ind.}{\sim} \mathrm{N}(\gamma_j, \sigma_{1j}^2)$: genetic effect on treatment A.
- $\hat{\Gamma}_j \overset{ind.}{\sim} \mathrm{N}(\Gamma_j, \sigma_{2j}^2)$: genetic effect on outcome Y.
- We assume

$$\Gamma_j = \beta\gamma_j + \alpha_j,\ j = 1, \ldots, p.$$

- β is causal effect of A on Y.
- $\alpha_j \sim \mathrm{N}(0, \tau^2)$ is direct pleiotropic effect of jth genetic variant on Y.

Example of population-based MR

- GWAS summary data from UK BioBank. Sample size around 500,000. 160 SNPs.
- "Treatment" is Body Mass Index (BMI); Outcome is systolic blood pressure (SBP).

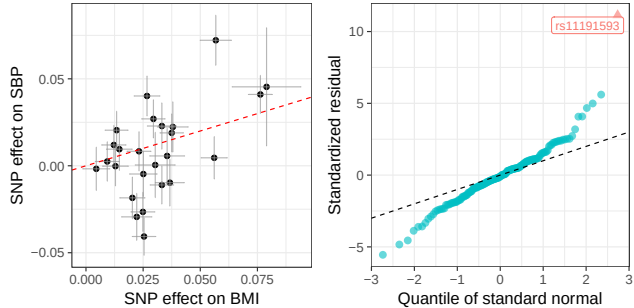


Figure: Left: $\hat{\Gamma}_j$ vs $\hat{\gamma}_j$; Right: Q-Q plot for $\hat{\alpha}_j = \hat{\Gamma}_j - \hat{\beta}\hat{\gamma}_j$ after standardisation.

- Estimated $\hat{\beta} = 0.402$ (standard error $= 0.106$). BMI and SBP were standardised.

Within-family MR

- Genes are not randomized at the population level.
- Rather, in Mendel's model they are randomized **conditional on** parent's haplotypes.
- Thus, MR is best justified in within-family design (Davey Smith and Ebrahim 2003).

Almost exact MR (Tudball, Davey Smith, and Zhao 2022)

1. Use graphical models to identify the exact "natural experiment".
2. Precisely describe various sources of biases in population-based MR: **population structure**, **assortative mating**, **dynastic effects**, and **horizontal pleiotropy**.
3. Use randomization as the sole basis of inference.

Example 5: Graphical approaches to confounder selection

Based on

- F. Richard Guo, Anton Rask Lundborg, and Qingyuan Zhao (2022). *Confounder Selection: Objectives and Approaches*. arXiv: 2208.13871 [stat.ME].
- F. Richard Guo and Qingyuan Zhao (Oct. 2023). *Confounder Selection via Iterative Graph Expansion*. DOI: 10.48550/arXiv.2309.06053. arXiv: 2309.06053 [math, stat].
- This started from a question by Anton's medical client in our Statistics Clinic.

Confounder selection

- Arguably the most important task in observational studies.
- Many criteria and methods, often loosely stated, sometimes ill-advised.

"Guideline" used by Anton's client

Austin (2011) claimed that there are four choices:

1. all measured baseline covariates;
2. all baseline covariates that are associated with treatment assignment;
3. all covariates that affect the outcome (i.e., the potential confounders),
4. all covariates that affect both treatment assignment and the outcome (i.e., the true confounders).

Citing simulation studies, he concluded that "there were merits to including **only the potential confounders** or **the true confounders** in the propensity score model."

Two common heuristics

The conjunction heuristic (a.k.a. the common cause principle)

Contrlling for all covariates "related" to both the treatment and the outcome.

- Very common in practice (Glymour, Weuve, and Chen 2008) and methodological development (Koch, Vock, Wolfson, and Vock 2020; Shortreed and Ertefaie 2017).
- Well known that this may select too few.

The pre-treatment heuristic

Controlling for all covariates that precede the treatment temporally.

- Defended in Rubin (2009): "I cannot think of a credible real-life situation where I would intentionally allow substantially different observed distributions of a true covariate in the treatment and control groups."
- Counter-examples from graphical models: e.g. M-bias (Greenland, Pearl, and Robins 1999).

Graphical approaches

Theorem (Back-door criterion (Pearl 1993, 2009))

Given a treatment X and an outcome Y, a set of covariates S controls for confounding if

1. *S contains no descendant of X;*
2. *S blocks all back-door paths from X to Y.*

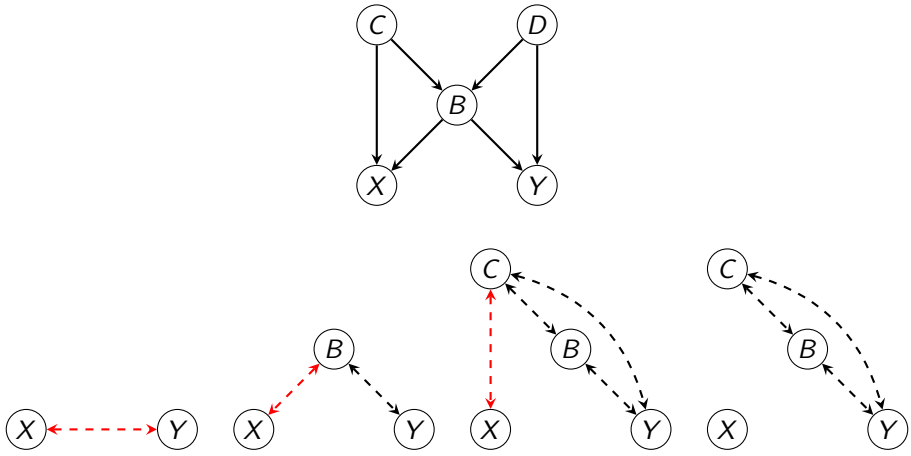
- This is **complete** in the sense that if S is a sufficient adjustment set, then $S \setminus \mathrm{de}(X)$ satisfies the backdoor criterion (Shpitser, VanderWeele, and Robins 2010).
- **Limitation**: requires "full" structural knowledge.

Theorem (Disjunctive criterion (VanderWeele and Shpitser 2011))

Suppose the causal graph is faithful. If at least one subset of S controls for confounding, then $S \cap [(X) \cup (Y)]$ controls for confounding.

- **Limitation**: verifying the assumption can be as difficult as the task of confounder selection itself.

Our approach in a nutshell



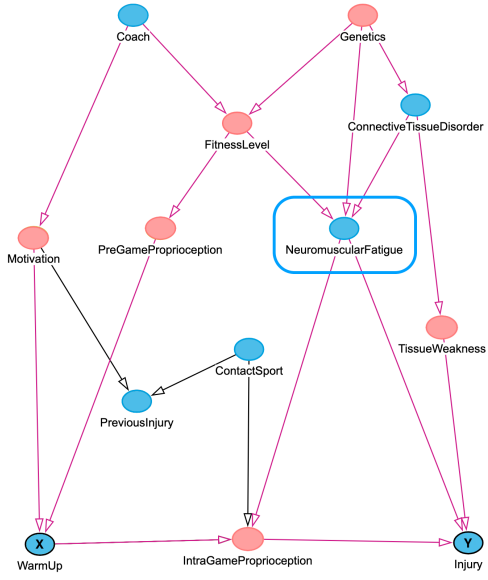
A more complex example

Suppose we are interested in whether `WarmUp` has a causal effect on `Injury`. We managed to collect a dataset with the following variables:

- `WarmUp`;
- `Injury`;
- `Coach`;
- `PreviousInjury`;
- `ContactSport`;
- `NeuromuscularFatigue`;
- `ConnectiveTissueDisorder`.

Demo link: `https://ricguo.shinyapps.io/InteractiveConfSel/`

"True" full graph behind the demo (Shrier and Platt 2008)



Summary

- All five examples echo a common theme: **design trumps analysis**.
  1. Meta-research of biodiversity conservation.
  2. Early outbreak analysis of COVID-19 in China.
  3. Racial discrimination in policing in New York City.
  4. Mendelian randomization—a popular method in genetic epidemiology.
  5. Graphical approaches to confounder selection in observational studies.
- As a principle of statistics, design is well understood in some contexts:
  - **Randomized experiments** and **survey sampling**;
  - **Causal identification** theory (using potential outcomes/graphs).

Discussion

- But design is a much broader concept. Related discussion include
  - **Exploratory data analysis** (Tukey 1977).
  - **Design sensitivity** and **evidence factors** (Rosenbaum 2010, 2021).
  - **Triangulation** in social science (Campbell and Fiske 1959; Mathison 1988) and epidemiology (Lawlor, Tilling, and Davey Smith 2016).
  - **Measurement theory** in psychometrics.
- Design is still and perhaps will always be elusive, but that is the beauty of statistics.

References I

- Austin, Peter C. (2011). "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies". In: *Multivariate Behavioral Research* 46.3, pp. 399–424.
- Campbell, Donald T. and Donald W. Fiske (1959). "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix". In: *Psychological Bulletin* 56.2, pp. 81–105. ISSN: 1939-1455. DOI: 10.1037/h0046016.
- Christie, Alec P., ..., Qingyuan Zhao, and William J. Sutherland (2020). "Quantifying and Addressing the Prevalence and Bias of Study Designs in the Environmental and Social Sciences". In: *Nature Communications* 11.1, p. 6377. DOI: 10.1038/s41467-020-20142-y.
- Davey Smith, George and Shah Ebrahim (Feb. 2003). "'Mendelian Randomization': Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease?*". In: *International Journal of Epidemiology* 32.1, pp. 1–22. ISSN: 0300-5771. DOI: 10.1093/ije/dyg070.

References II

- Gaebler, Johann, William Cai, Guillaume Basse, Ravi Shroff, Sharad Goel, and Jennifer Hill (Dec. 2022). "A Causal Framework for Observational Studies of Discrimination". In: *Statistics and Public Policy* 9.1, pp. 26–48. ISSN: null. DOI: 10.1080/2330443X.2021.2024778.
- Glymour, M. Maria, Jennifer Weuve, and Jarvis T. Chen (2008). "Methodological Challenges in Causal Research on Racial and Ethnic Patterns of Cognitive Trajectories: Measurement, Selection, and Bias". In: *Neuropsychology Review* 18.3, pp. 194–213. DOI: 10.1007/s11065-008-9066-x.
- Greenland, Sander, Judea Pearl, and James M. Robins (1999). "Causal Diagrams for Epidemiologic Research". In: *Epidemiology (Cambridge, Mass.)* 10.1, pp. 37–48.
- Guo, F. Richard, Anton Rask Lundborg, and Qingyuan Zhao (2022). *Confounder Selection: Objectives and Approaches*. arXiv: 2208.13871 [stat.ME].
- Guo, F. Richard and Qingyuan Zhao (Oct. 2023). *Confounder Selection via Iterative Graph Expansion*. DOI: 10.48550/arXiv.2309.06053. arXiv: 2309.06053 [math, stat].
- Knox, Dean, Will Lowe, and Jonathan Mummolo (2020). "Administrative Records Mask Racially Biased Policing". In: *American Political Science Review* 114.3, pp. 619–637. DOI: 10.1017/s0003055420000039.

References III

- Koch, Brandon, David M Vock, Julian Wolfson, and Laura Boehm Vock (2020). "Variable Selection and Estimation in Causal Inference Using Bayesian Spike and Slab Priors". In: *Statistical Methods in Medical Research* 29.9, pp. 2445–2469. DOI: 10.1177/0962280219898497.
- Lawlor, Debbie A, Kate Tilling, and George Davey Smith (Dec. 2016). "Triangulation in Aetiological Epidemiology". In: *International Journal of Epidemiology* 45.6, pp. 1866–1886. ISSN: 0300-5771. DOI: 10.1093/ije/dyw314.
- Mathison, Sandra (1988). "Why Triangulate?" In: *Educational Researcher* 17.2, pp. 13–17. ISSN: 0013-189X. DOI: 10.2307/1174583. JSTOR: 1174583. (Visited on 03/06/2024).
- Pearl, Judea (1993). "Comment: Graphical Models, Causality and Intervention". In: *Statistical Science* 8.3, pp. 266–269.
- — (2009). *Causality: Models, Reasoning, and Inference*. 2nd ed. New York: Cambridge University Press.
- Rosenbaum, Paul R. (2010). *Design of Observational Studies*. Springer Series in Statistics. New York: Springer. DOI: 10.1007/978-1-4419-1213-8.
- — (Mar. 2021). *Replication and Evidence Factors in Observational Studies*. Boca Raton: Chapman and Hall/CRC. ISBN: 978-1-00-303964-8. DOI: 10.1201/9781003039648.

References IV

- Rubin, Donald B. (2009). "Should Observational Studies Be Designed to Allow Lack of Balance in Covariate Distributions across Treatment Groups?" In: *Statistics in Medicine* 28.9, pp. 1420–1423.
- Shortreed, Susan M. and Ashkan Ertefaie (2017). "Outcome-Adaptive Lasso: Variable Selection for Causal Inference". In: *Biometrics. Journal of the International Biometric Society* 73.4, pp. 1111–1122. DOI: 10.1111/biom.12679.
- Shpitser, Ilya, Tyler VanderWeele, and James M. Robins (2010). "On the Validity of Covariate Adjustment for Estimating Causal Effects". In: *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*. UAI'10. Arlington, Virginia, USA: AUAI Press, pp. 527–536. ISBN: 978-0-9749039-6-5.
- Shrier, Ian and Robert W. Platt (Oct. 2008). "Reducing Bias through Directed Acyclic Graphs". In: *BMC Medical Research Methodology* 8.1, p. 70. ISSN: 1471-2288. DOI: 10.1186/1471-2288-8-70.
- Tudball, Matthew J, George Davey Smith, and Qingyuan Zhao (2022). "Almost Exact Mendelian Randomization". In: arXiv: 2208.14035 [stat.ME].
- Tukey, John Wilder (1977). *Exploratory Data Analysis*. Addison-Wesley Publishing Company. ISBN: 978-0-201-07616-5.

References V

- VanderWeele, Tyler and Ilya Shpitser (2011). "A New Criterion for Confounder Selection". In: *Biometrics. Journal of the International Biometric Society* 67.4, pp. 1406–1413.
- Wu, Joseph T., Kathy Leung, and Gabriel M. Leung (Feb. 2020). "Nowcasting and Forecasting the Potential Domestic and International Spread of the 2019-nCoV Outbreak Originating in Wuhan, China: A Modelling Study". In: *The Lancet* 395.10225, pp. 689–697. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(20)30260-9.
- Zhao, Qingyuan (2022). "Small Data, Big Time—A Retrospect of the First Weeks of COVID-19 (with Discussion)". In: *Journal of the Royal Statistical Society (Series A, Statistics in Society)* 185.4, pp. 1793–1814. DOI: 10.1111/rssa.12874.
- Zhao, Qingyuan, Yang Chen, and Dylan S Small (Feb. 2020). "Analysis of the Epidemic Growth of the Early 2019-nCoV Outbreak Using Internationally Confirmed Cases". In: DOI: 10.1101/2020.02.06.20020941.
- Zhao, Qingyuan, Nianqiao Ju, Sergio Bacallado, and Rajen D. Shah (2021). "BETS: The Dangers of Selection Bias in Early Analyses of the Coronavirus Disease (COVID-19) Pandemic". In: *Annals of Applied Statistics* 15.1, pp. 363–390. DOI: 10.1214/20-aoas1401.

References VI


Zhao, Qingyuan, Luke J Keele, Dylan S Small, and Marshall M Joffe (2021). "A Note on Posttreatment Selection in Studying Racial Discrimination in Policing". In: *American Political Science Review* 116.1, pp. 337–350. DOI: 10.1017/s0003055421000654.


Zhao, Qingyuan, Jingshu Wang, Gibran Hemani, Jack Bowden, and Dylan S. Small (2020). "Statistical Inference in Two-Sample Summary-Data Mendelian Randomization Using Robust Adjusted Profile Score". In: *Annals of Statistics* 48.3, pp. 1742–1769. DOI: 10.1214/19-aos1866.