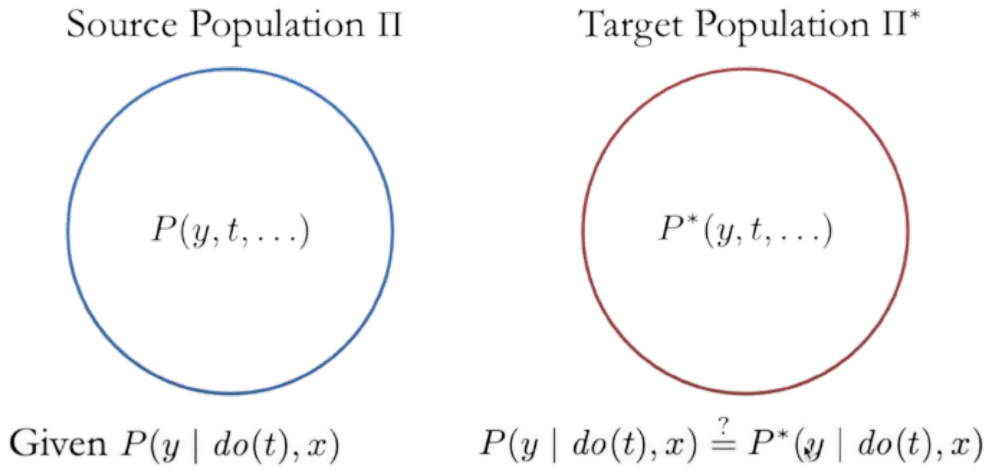
External Validity: From Do-Calculus to Transportability Across Populations

by Judea Pearl and Elias Bareinboim

Presented by Pooya Khosravi
CS 295 - Winter 2022-23, Causal Inference

Transportability of Causal Effects Across Populations

- We know the causal effect of one population
- What do we need to transfer/transport and get the causal effect in a different population?



Examples of Transportability in Medicine

- Use of clinical trials to test the effectiveness of a drug in a specific population and then generalizing the results to other populations
- Use of electronic health records to study the effectiveness of a treatment across different healthcare systems and populations
- Transportability is important in medical research to ensure that treatments are effective and safe for all patients, regardless of their background or health status

Examples of Transportability in Social Sciences

- Use of randomized controlled trials to test the effectiveness of an intervention in a specific population and then generalizing the results to other populations
- Use of surveys to study the attitudes and behaviors of a population and then generalizing the results to other populations
- Transportability is important in social science research to ensure that policies and interventions are effective and relevant for all populations, regardless of their demographic characteristics or cultural background

Question for the talk

- Can you think of another example that benefits from causal transportability?
    - Why would external validity be important for the given problem?
    - What are some factors that limit the generalizability of causal effects for the given problem?

Outline

- External validity
  - Explanation of external validity and its importance in scientific research
- Challenges of achieving external validity
  - Discussion of the challenges researchers face when trying to achieve external validity
- The Do-Calculus method
  - Explanation of the Do-Calculus method and how it can be used to address confounding variables
- Transportability across populations
  - Definition of transportability and its importance in achieving external validity

INTRODUCTION:
THREATS VS. ASSUMPTIONS

INTRODUCTION: THREATS VS. ASSUMPTIONS

- Why generalization?
- Arbitrary or drastically different environments, or Sufficiently similar environments
- Prior methods:
  - Meta analysis, or Hierarchical models
- Results of diverse studies are pooled together by standard statistical procedures
- Rarely make explicit distinction between experimental and observational regimes

INTRODUCTION: THREATS VS. ASSUMPTIONS

This paper provide theoretical guidance on:

- Limits on what can be achieved in practice
- Problems that are likely to be encountered when populations differ significantly
- What population differences can be circumvented
- What differences constitute theoretical impediments

INTRODUCTION: THREATS VS. ASSUMPTIONS

Studying "threats" over "licensing assumptions". Why?

- Safer to cite, little risk related to endorsing something
- Assumptions are self-destructive in their honesty.
    - The more explicit the assumption, the more criticism it invites
- Threats can be communicated in plain English
- Assumptions require a formal language with precise characterization

INTRODUCTION: THREATS VS. ASSUMPTIONS

Formal language would consist of:

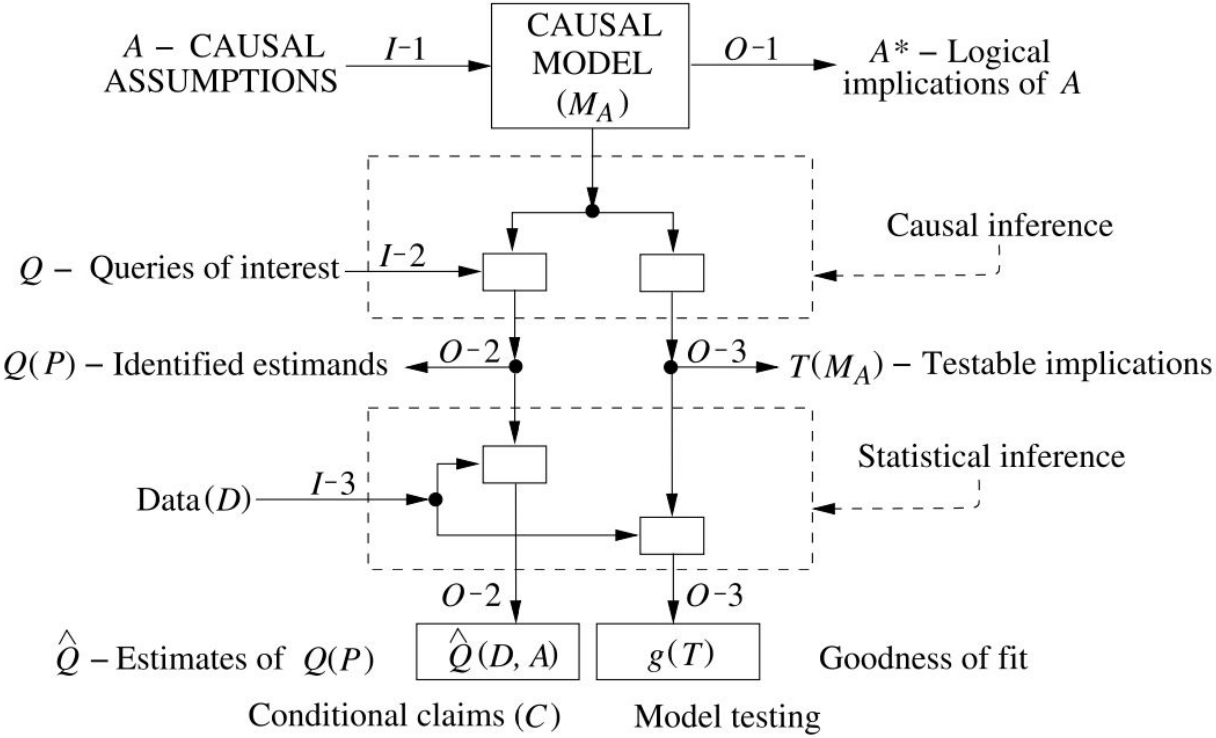
- Causal diagrams
- Models of interventions
- Counterfactuals

Using Do-Calculus to:

- Test the feasibility of transport
- Estimating causal effects in the target population

PRELIMINARIES: THE LOGICAL FOUNDATIONS OF CAUSAL INFERENCE

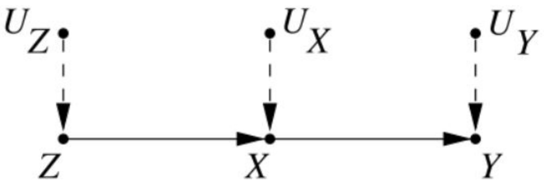
THE LOGICAL FOUNDATIONS OF CAUSAL INFERENCE



Assumptions in Nonparametric Models (SEM)

- A set U of background or exogenous variables, representing factors outside the model.
- A set V = {V1, . . . , Vn} of endogenous variables, assumed to be observable.
- A set F of functions {f1,...,fn} such that each fi determines the value of Vi ∈ V.
- A joint probability distribution P(u) over U.

Assumptions in Nonparametric Models (SEM)

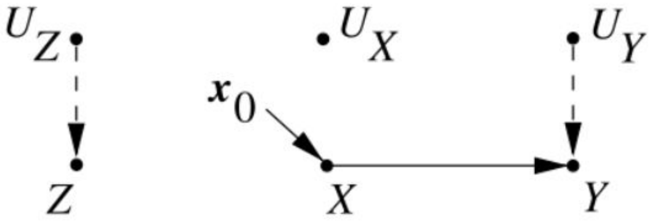


(a)

$$z = f_Z(u_Z),$$
$$x = f_X(z, u_X),$$
$$y = f_Y(x, u_Y),$$

Interventions, Counterfactuals and Causal Effects

- Interventions through a mathematical operator called do(x)
- Example: do(x0)



(b)

$$P_M\big(y \mid \text{do}(x)\big) = P_{M_x}(y).$$

$$z = f_Z(u_Z),$$

$$x = x_0,$$

$$y = f_Y(x, u_Y),$$

Identification, d-Separation and Causal Calculus

A causal query Q(M) is identifiable, given a set of assumptions A, if for any two (fully specified) models, M1 and M2, that satisfy A, we have:

$$P(M_1) = P(M_2) \Rightarrow Q(M_1) = Q(M_2).$$

The Rules of do-Calculus

RULE 1 (Insertion/deletion of observations).

$$P(y \mid \mathrm{do}(x), z, w) = P(y \mid \mathrm{do}(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}. \tag{2.5}$$

RULE 2 (Action/observation exchange).

$$P(y \mid \mathrm{do}(x), \mathrm{do}(z), w) = P(y \mid \mathrm{do}(x), z, w) \quad \text{if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}}. \tag{2.6}$$

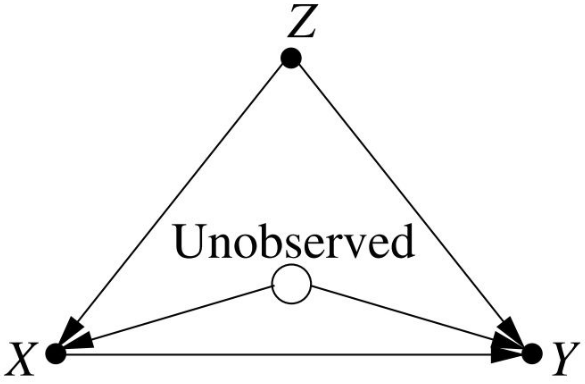
RULE 3 (Insertion/deletion of actions).

$$P(y \mid \mathrm{do}(x), \mathrm{do}(z), w) = P(y \mid \mathrm{do}(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X Z(W)}}}, \tag{2.7}$$

INFERENCE ACROSS POPULATIONS: MOTIVATING EXAMPLES

Example 1

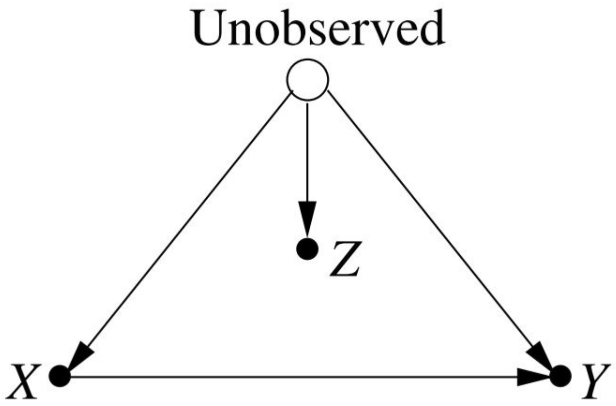
- Figure represents cause-effect relationships in the pretreatment population in Los Angeles
  - Z represents “age.”
- We conduct a randomized trial in Los Angeles
  - to estimate the causal effect of exposure X on outcome Y for every age group Z = z
- If we assume that age-specific effects are invariant across cities
  - if the LA study provides us with (estimates of) age specific causal effects
  - the overall causal effect in NYC should be
  - $$(3.1) \quad P^*(y|\operatorname{do}(x)) = \sum_z P(y|\operatorname{do}(x), z)P^*(z).$$



(a)

Example 2

- Let the variable Z stand for subjects language proficiency
- Assume that Z does not affect exposure (X) or outcome (Y), yet it correlates with both, being a proxy for age
- Age (the hollow circle) is not measured in either study
- The inequality $P(z) \neq P^*(z)$ may reflect different reasons

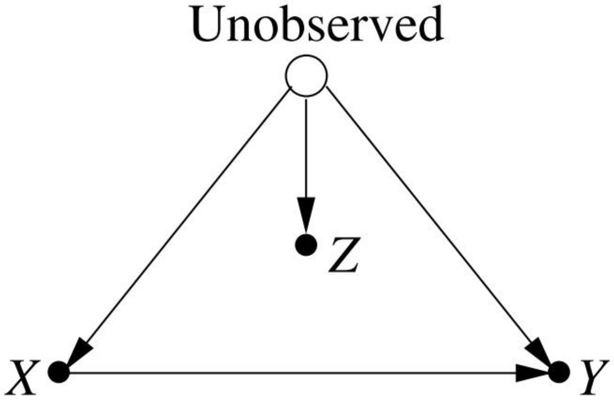


(b)

Example 2

- If the two cities have identical age distributions and NYC residents acquire linguistic skills at a younger age
- Then Z has no effect on X and Y
  - The inequality can be ignored and transport formula would be:

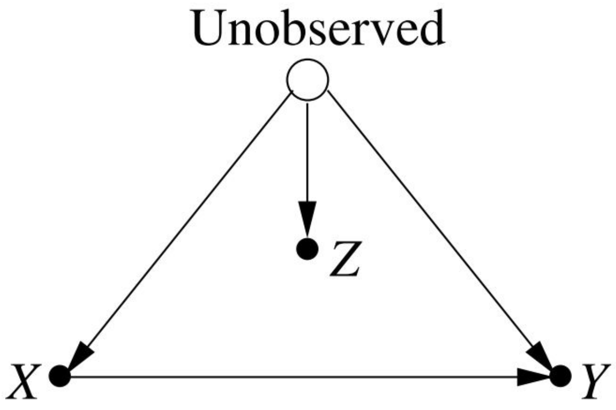
$$P^*(y|\,\mathrm{do}(x)) = P(y|\,\mathrm{do}(x)). \tag{3.2}$$



(b)

Example 2

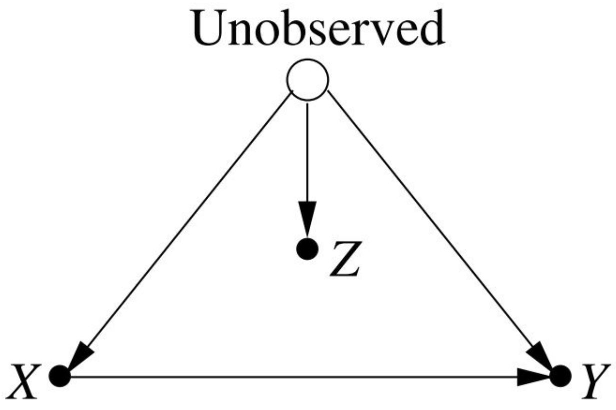
- If the conditional probabilities $P(z|\text{age})$ and $P^*(z|\text{age})$ are the same in both cities
  - $P(z) \neq P^*(z)$ would represent genuine age differences
  - 3.2 would no longer be valid
- The choice of the proper transport formula depends on the causal context in which population differences are embedded.



(b)

Example 2

- If the conditional probabilities $P(z|\text{age})$ and $P^*(z|\text{age})$ are the same in both cities
  - $P(z) \neq P^*(z)$ would represent genuine age differences
  - 3.2 would no longer be valid
- The choice of the proper transport formula depends on the causal context in which population differences are embedded.

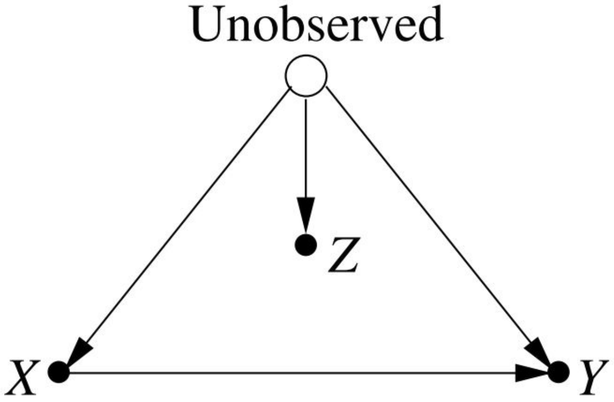


(b)

Example 2

The Z-specific effect of X on Y in NYC is given by

$$
\begin{aligned}
&P^*(y|\,\mathrm{do}(x), z) \\
&\quad = \sum_{\text{age}} P^*(y|\,\mathrm{do}(x), z, \text{age}) P^*(\text{age}|\,\mathrm{do}(x), z) \\
&\quad = \sum_{\text{age}} P^*(y|\,\mathrm{do}(x), \text{age}) P^*(\text{age}|z) \\
&\quad = \sum_{\text{age}} P(y|\,\mathrm{do}(x), \text{age}) P^*(\text{age}|z).
\end{aligned}
$$



(b)

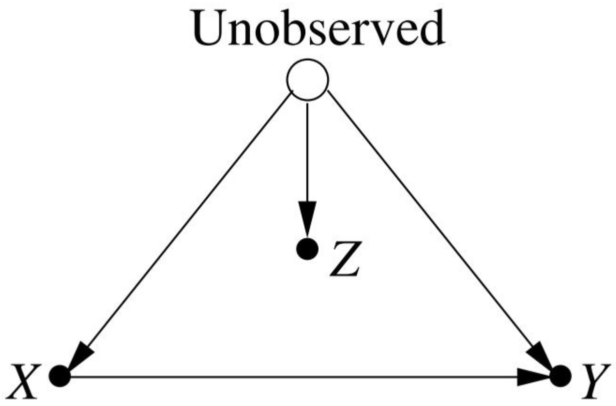
Example 2

Thus, if the two populations differ in the relation between age and skill, that is,

$$P(\text{age}|z) \neq P^*(\text{age}|z)$$

the skill-specific causal effect would differ as well.

Intuition: A NYC person at skill level Z = z is likely to be in a totally different age group from his skill-equals in Los Angeles



(b)

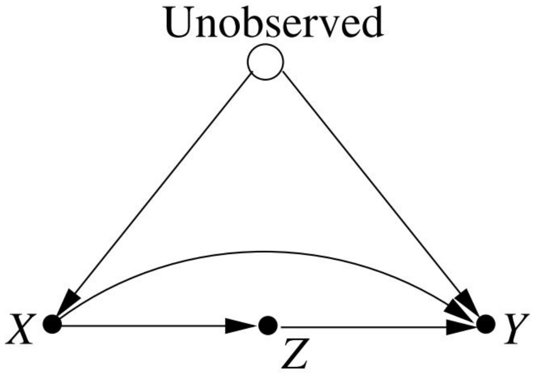
Difference between Example 1 and 2

- Age is normally taken to be an exogenous variable
- While skills may be indicative of earlier factors capable of modifying the causal effect.
- Therefore, conditional on skill, the effect may be **different** in the two populations.

Example 3

- Z is a X-dependent variable, say a disease biomarker, standing on the causal pathways between X and Y as shown in Figure.
- Overall causal effect (in both LA and NYC) is no longer a simple average of the z-specific causal effects
  - 3.1 is wrong, correct weighing rule is:

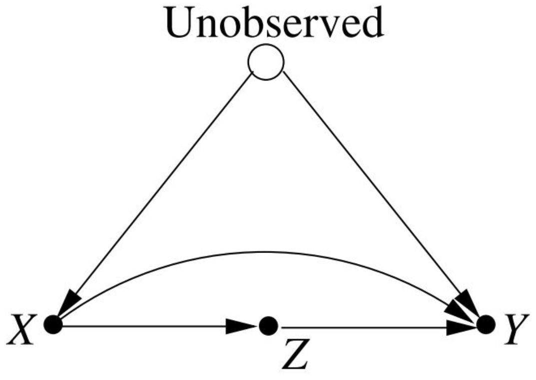
$$P^*(y|\,\mathrm{do}(x)) = \sum_z P^*(y|\,\mathrm{do}(x), z) P^*(z|\,\mathrm{do}(x)), \tag{3.3}$$



(c)

Example 3 - More Realistic Setting

- We wish to use Z as a “surrogate endpoint” to predict the efficacy of treatment X on outcome Y, where Y is too difficult and/or expensive to measure routinely
  - Instead of conducting experiment/observational studies at two locations, we consider two such studies taking place at the **same** location, but at **different** times
  - In the first study, we measure P(y, z| do(x)) and discover that Z is a good surrogate
  - Once Z is proclaimed a “surrogate endpoint,” it invites efforts to find direct means of controlling Z



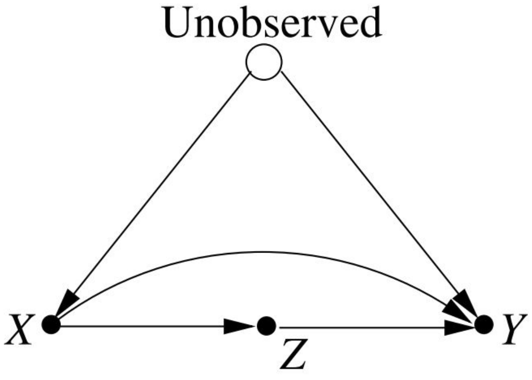
(c)

Example 3 - More Realistic Setting

- Y = heart disease
- Z = cholesterol levels
- X is cholestrol-reducing substances

- Assuming the difference observed is only in people's susceptibility to X
  - The correct transport formula:

$$(3.4) \quad P^*(y|\,\mathrm{do}(x)) = \sum_z P(y|\,\mathrm{do}(x), z) P^*(z|x),$$



(c)

FORMALIZING TRANSPORTABILITY

Selection Diagrams and Selection Variables

- Pattern from examples show transportability is a causal and not a statistical notion.
  - For example 3: Important to ascertain that the change in $P(z|x)$ was due to the change in the way Z is affected by X, but not due to a change in confounding conditions between the two.
- Licensing transportability requires knowledge of the mechanisms, or processes, through which population differences come about;
- Different localization of these mechanisms yield different transport formulae.

Selection Diagrams and Selection Variables

- A representation in which the causal mechanisms are explicitly encoded
- Differences in populations are represented as local modifications of those mechanisms

Solution:

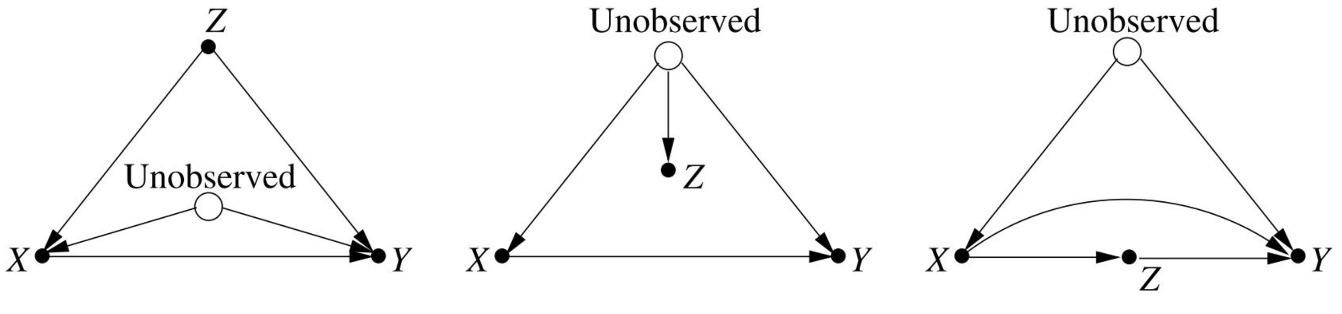
- Causal diagrams augmented with a set, S, of “selection variables,” where each member of S corresponds to a mechanism by which the two populations differ

Selection Diagram

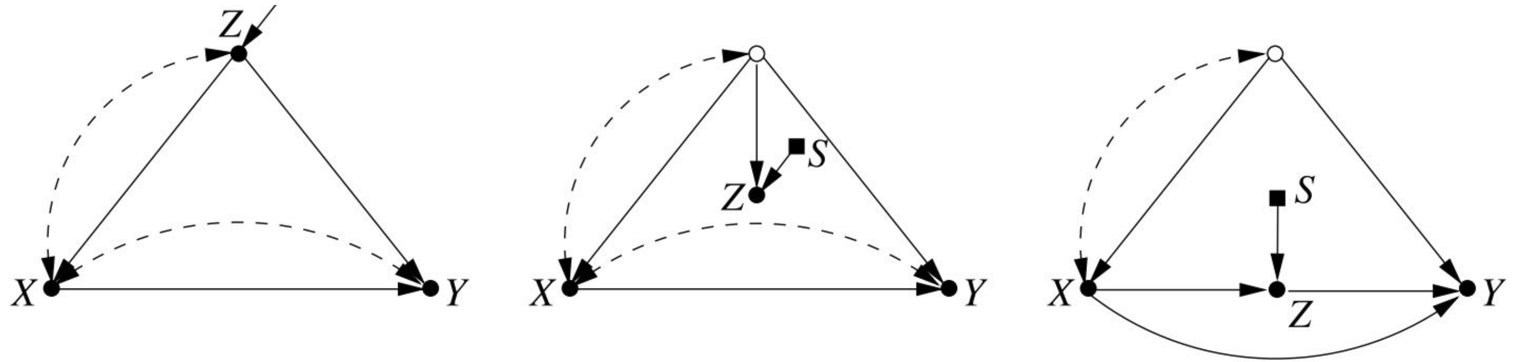
DEFINITION 4 (Selection diagram). Let $\langle M, M^* \rangle$ be a pair of structural causal models (Definition 1) relative to domains $\langle \Pi, \Pi^* \rangle$, sharing a causal diagram G. $\langle M, M^* \rangle$ is said to induce a selection diagram D if D is constructed as follows:

1. Every edge in G is also an edge in D.
2. D contains an extra edge $S_i \to V_i$ whenever there might exist a discrepancy $f_i \neq f_i^*$ or $P(U_i) \neq P^*(U_i)$ between M and M^*.

Selection Diagrams for Examples 1-3



(a) (b) (c)



(a) (b) (c)

Transportability: Definitions and Examples

Definition 5 (Transportability). Let D be a selection diagram relative to domains $\langle \Pi, \Pi^* \rangle$. Let $\langle P, I \rangle$ be the pair of observational and interventional distributions of Π, and P^* be the observational distribution of Π^*. The causal relation $R(\Pi^*) = P^*(y | \mathrm{do}(x), z)$ is said to be transportable from Π to Π^* in D if $R(\Pi^*)$ is uniquely computable from P, P^*, I in any model that induces D.

Transportability Theorem

THEOREM 1. *Let D be the selection diagram characterizing two populations, Π and Π^*, and S a set of selection variables in D. The relation $R = P^*(y|\operatorname{do}(x), z)$ is transportable from Π to Π^* if the expression $P(y|\operatorname{do}(x), z, s)$ is reducible, using the rules of do-calculus, to an expression in which S appears only as a conditioning variable in do-free terms.*

Theorem 1 Proof

PROOF. Every relation satisfying the condition of Theorem 1 can be written as an algebraic combination of two kinds of terms, those that involve S and those that do not. The former can be written as P^*-terms and are estimable, therefore, from observations on Π^*, as required by Definition 5. All other terms, especially those involving do-operators, do not contain S; they are experimentally identifiable therefore in Π. $\square$

Trivial Transportability

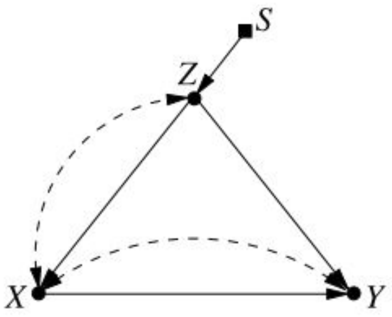
DEFINITION 6 (Trivial transportability). A causal relation R is said to be *trivially transportable* from Π to Π^*, if $R(\Pi^*)$ is identifiable from (G^*, P^*).

- We can estimate $R(\Pi^*)$ directly from observational studies on Π^*, unaided by causal information from Π.

Direct Transportability

DEFINITION 7 (Direct transportability). A causal relation R is said to be *directly transportable* from Π to Π^*, if $R(\Pi^*) = R(\Pi)$.

- Meaning: X blocks all paths from S to Y once we remove all arrows pointing to X and condition on Z.



(a)

TRANSPORTABILITY OF CAUSAL EFFECTS—A GRAPHICAL CRITERION

Theorem 2

THEOREM 2. *Let D be the selection diagram characterizing two populations, Π and Π^*, and S the set of selection variables in D. The strata-specific causal effect $P^*(y|\,\mathrm{do}(x), z)$ is transportable from Π to Π^* if Z d-separates Y from S in the X-manipulated version of D, that is, Z satisfies $(Y \perp\!\!\!\perp S|Z, X)_{D_{\overline{X}}}$.*

Theorem 2 Proof

PROOF.

$$P^*(y|\,\mathrm{do}(x), z) = P(y|\,\mathrm{do}(x), z, s^*).$$

From Rule 1 of do-calculus we have: $P(y|\,\mathrm{do}(x), z, s^*) = P(y|\,\mathrm{do}(x), z)$ whenever Z satisfies $(Y \perp\!\!\!\perp S|Z, X)$ in $D_{\overline{X}}$. This proves Theorem 2. $\square$

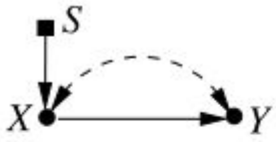
S-admissibility

DEFINITION 8 (S-admissibility). A set T of variables satisfying $(Y \perp\!\!\!\perp S | T, X)$ in $D_{\overline{X}}$ will be called S-admissible (with respect to the causal effect of X on Y).

Corollaries of Theorem 2

COROLLARY 1. *The average causal effect* $P^*(y|\mathrm{do}(x))$ *is transportable from* Π *to* Π^* *if there exists a set* Z *of observed pretreatment covariates that is* S*-admissible. Moreover, the transport formula is given by the weighting of equation (3.1).*

COROLLARY 2. *Any* S *variable that is pointing directly into* X *as in Figure 6(a), or that is d-separated from* Y *in* $D_{\overline{X}}$ *can be ignored.*



(a)

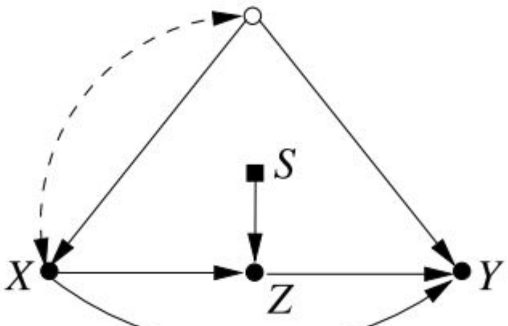
- Randomization used in the experimental study washes away propensity to receive treatment

Theorem 3

Generalizing Theorem 2 to cases involving treatment-dependent Z variables:

THEOREM 3. *The average causal effect* $P^*(y|\text{do}(x))$ *is transportable from* Π *to* Π^* *if either one of the following conditions holds:*

1. $P^*(y|\text{do}(x))$ *is trivially transportable.*
2. *There exists a set of covariates,* Z *(possibly affected by* X*) such that* Z *is* S*-admissible and for which* $P^*(z|\text{do}(x))$ *is transportable.*
3. *There exists a set of covariates,* W *that satisfy* $(X \perp\!\!\!\perp Y|W)_{D_{\overline{X(W)}}}$ *and for which* $P^*(w|\text{do}(x))$ *is transportable.*



(c)

Applying Theorems

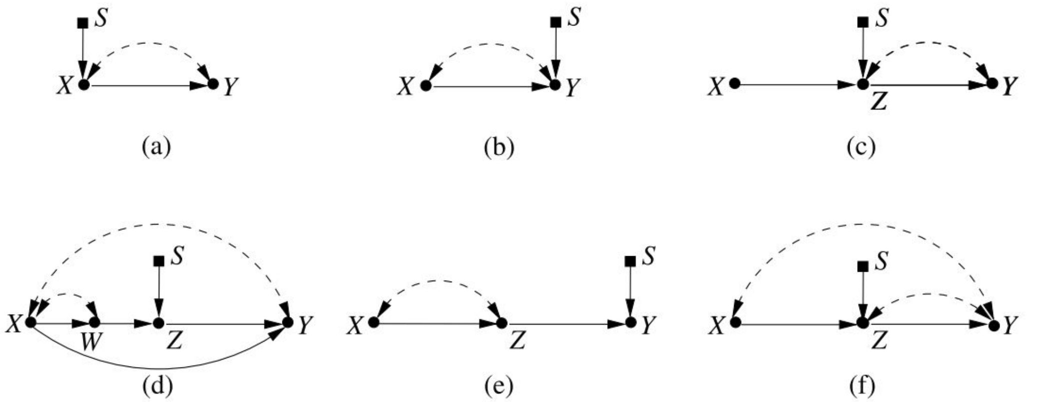
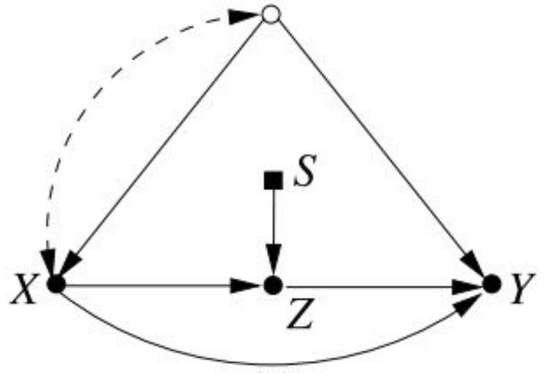


FIG. 6. *Selection diagrams illustrating transportability. The causal effect* $P(y|\text{do}(x))$ *is (trivially) transportable in* (c) *but not in* (b) *and* (f). *It is transportable in* (a), (d) *and* (e) *(see Corollary 2).*

Applying Theorem 3

- Our goal is to estimate P*(y| do(x))—the effect of X on Y in the new population created by changes in how Z responds to X.
- The structure of the problem permits us to satisfy condition 2 of Theorem 3, since:
    - Z is S-admissible
        - $(S \perp\!\!\!\perp Y | X, Z)_{G_{\overline{X}}}$
    - P*(z|do(x)) is trivially transportable.
        - X and Z are unconfounded
    - We get:

$$(5.8) \quad P^*(y|\operatorname{do}(x)) = \sum_z P(y|\operatorname{do}(x), z) P^*(z|x),$$



(c)

CONCLUSIONS

Conclusions

- External validity is important in scientific research to ensure that causal effects are generalizable and applicable to different populations
- The Do-Calculus method is a powerful tool for addressing confounding variables and achieving external validity in causal inference
- Transportability is the ability to generalize causal effects from one population to another, and it is important for achieving external validity in many fields, including medicine and social sciences
- The examples presented illustrate the importance of external validity and the different methods that can be used to achieve it

Repeat: Question for the talk

- Can you think of another example that benefits from causal transportability?
  - Why would external validity be important for the given problem?
  - What are some factors that limit the generalizability of causal effects for the given problem?

Thank you!