Causal Inference

PPOL 8602: Research Design for Public Policy

Kingsford Onyina

February 9, 2026

Causal Inference vs Internal Validity

Causal Inference

- Goal: Does X cause Y?
- Estimating causal effects (counterfactual logic)
- Captures all aspects of validity: including internal, measurement and external validity.

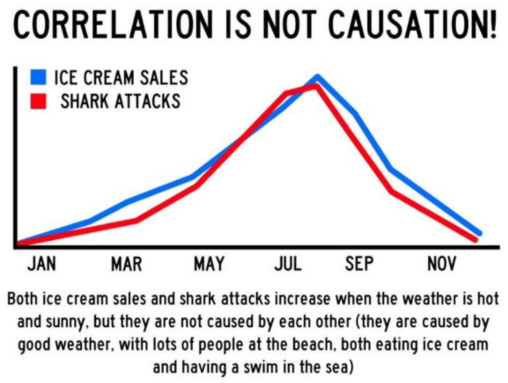
Internal Validity

- Property of the study design
- Ability to rule out alternative explanations
- Involves making specific steps to give us confidence that X causes Y.

Internal validity disciplines causal inference.

Now Focusing on Causal Inference

- Goes beyond correlation and description
- Core task in policy evaluation and program design
- Requires:
  - Counterfactual thinking
  - Identification strategy (design)
  - Explicit assumptions



Fundamental Problem of Causal Inference

- A unit cannot be observed both treated and untreated
- Only one potential outcome is realized
- The counterfactual outcome is missing

This is the core identification problem.

Rubin Causal Model (1974)

- Each unit has two potential outcomes:
  - $Y_i(1)$: outcome if treated
  - $Y_i(0)$: outcome if untreated
- Individual causal effect:

$$Y_i(1) - Y_i(0)$$

- Never observed directly

Causal inference targets average effects under assumptions and design.

Average Treatment Effects

Since individual effects are unobservable, we focus on **average (nomothetic) effects**:

- **Average Treatment Effect (ATE):**

$$\text{ATE} = \mathbb{E}\big[Y_i(1) - Y_i(0)\big]$$

Nomothetic effect for a randomly selected individual in the population

- **Average Treatment Effect on the Treated (ATT):**

$$\text{ATT} = \mathbb{E}\big[Y_i(1) - Y_i(0) \mid D_i = 1\big]$$

Nomothetic effect for units that actually received treatment

- **Average Treatment Effect on the Untreated (ATU):**

$$\text{ATU} = \mathbb{E}\big[Y_i(1) - Y_i(0) \mid D_i = 0\big]$$

Nomothetic effect for units that did not receive treatment

Solving the Fundamental Problem (Examples)

Big idea: approximate the missing counterfactual

- **Social experiments (RCTs)**: random assignment balances confounders
- **Natural experiments**: as-if random variation in treatment
- **Instrumental variables**: exogenous source of variation in X
- **Fixed effects**: control for time-invariant unobservables...etc.

Assumptions: Unit Homogeneity

- We cannot rerun history: each unit has only one realized path
- Units should be comparable **in expectation**

KKV-style idea

Two units are homogeneous when the expected values of the dependent variable from each unit are the same when the explanatory variable takes on a particular value.

Assumptions: Conditional Independence

- Assignment to X is independent of potential outcomes **given covariates**
- Intuition: after conditioning, treatment is "as good as random"
- Supports common strategies:
  - Regression adjustment
  - Matching / weighting
  - Difference-in-differences (under parallel trends)

Assumptions: SUTVA

Stable Unit Treatment Value Assumption

- **No interference**: one unit's treatment does not affect another's outcome
- **No multiple versions**: treatment is well-defined (same "dose/type")

Ensures treatment effects are well-defined and interpretable.

Conditional independence handles confounding; SUTVA handles interference and ambiguity.

Problems of Causal Inference: Endogeneity

Endogeneity occurs when X is correlated with the error term.

- **Omitted variable bias**: unobservables affect both X and Y
- **Reverse causality (simultaneity)**: Y affects X
- **Measurement error**: error in X biases estimates

Other Common Problems

- Small samples $\Rightarrow$ low statistical power
- Selection bias (who gets treated is not random)
- Attrition (especially in panel data)
- Limited external validity (generalizability)

Criteria for Causation (Internal Validity)

- **Temporal ordering**: cause precedes effect
- **Covariation**: X and Y move together
- **Non-spuriousness**: no alternative explanation
- **Mechanism**
- **Context**: It implies that whether X causes Y depends on the setting — the social, institutional, temporal, or population context.
- A causal effect may exist in one context but not another.

Qualifying Exam Essentials

- **Variables**
  - Dependent variable (Y): outcome of interest
  - Independent variable (X): explanatory variable
  - Control Variables (Z)
- **Hypotheses**
  - Null hypothesis H_0 vs Alternative hypothesis H_1
  - One-sided vs Two-sided tests
- **Design & Interpretation**
  - Unit of analysis
  - Ecological fallacy vs Reductionist fallacy

Data Structures and Causal Leverage

- **Cross-sectional**: time order often unclear
- **Time series**: dynamics over time (single unit)
- **Pooled cross-sections**: repeated snapshots, different units
- **Panel/longitudinal**: strongest for causal designs (FE, DiD)

Key Takeaway / Way forward

- Causality is unavoidable in policy research (KKV)
- Rubin formalizes the counterfactual logic
- We estimate **average** effects (ATE/ATT/ATU), not individual effects
- Credible inference = strong design + transparent assumptions + honest uncertainty
- The way forward? There's a world beyond OLS

Why Causal Inference? (Conclusion: KKV)

King, Keohane & Verba (1995, p. 76)

"Avoiding causal language when causality is the real subject of investigation either renders the research irrelevant or permits it to remain undisciplined by the rules of scientific inference. Our uncertainty about causal inferences will never be eliminated. But this uncertainty should not suggest that we avoid attempts at causal inference. Rather we should draw causal inferences where they seem appropriate but also provide the reader with the best and most honest estimate of the uncertainty of that inference. It is appropriate to be bold in drawing causal inferences as long as we are cautious in detailing the uncertainty of the inference."

What KKV Is Saying

- Policy questions are inherently **causal**
- Avoiding causal language makes research either **irrelevant** or **undisciplined**
- Uncertainty is inevitable $\Rightarrow$ be transparent, not evasive
- Strong practice:
    - Make causal claims explicit
    - State assumptions clearly
    - Report uncertainty honestly

Thank You

Thank you for your time and engagement.

Your feedback is appreciated.

Share Feedback