Introduction to Causal Inference
Based on Ch. 1&2, What If by H. Robins

Gauranga Kumar Baishya, CMI

May 14, 2025

Outline

What is Causal Effect?

Causal inference is an innate human ability, essential for survival.We distinguish between **association** and **causation** in everyday reasoning. Chapter 1 introduces formal mathematical notation to formalize causal inference. Understanding counterfactual is key to defining causal effects rigorously.

Mathematical Notation for Individual Causal Effect

- Define a treatment variable $A \in \{0, 1\}$ (1: treated, 0: untreated).
- Define an outcome variable $Y \in \{0, 1\}$ (1: death, 0: survival).
- Define potential outcomes:
  - $Y^{a=1}$: Y under treatment a=1.
  - $Y^{a=0}$: Y under treatment a=0.
  - The outcome variable that would have been observed under the treatment value a = 1 & a = 0 respectively.
- Individual causal effect exists if:

$$Y^{a=1} \neq Y^{a=0}. \tag{1}$$

- Only one of these potential outcomes is observed, making causal inference challenging.

Concept of Individual Causal Effects

Causal effect means comparing outcomes under different interventions. Compare the outcome under treatment vs. no treatment. If $Y^{a=1} \neq Y^{a=0}$, the treatment has a causal effect on the individual.

Example: Zeus receives a heart transplant and dies in 5 days, but without the transplant, would he have lived?

Different Causal Effect Measures

Individual causal effects are often unobservable due to missing counterfactual. Instead, we define the **average causal effect** (ACE):

$$E[Y^{a=1}] - E[Y^{a=0}] \neq 0. \tag{2}$$

This measures the difference in expected outcomes if everyone received treatment vs. if no one did.

Risk difference: $\text{RD} = P(Y^{a=1} = 1) - P(Y^{a=0} = 1)$. **Risk ratio**: $\text{RR} = \frac{P(Y^{a=1}=1)}{P(Y^{a=0}=1)}$. **Odds ratio**: $\text{OR} = \frac{\frac{P(Y^{a=1}=1)}{P(Y^{a=1}=0)}}{\frac{P(Y^{a=0}=1)}{P(Y^{a=0}=0)}}$.

Key Differences: Association vs Causation

- **Association** is correlation between treatment and outcome.
- **Causation** implies a counterfactual comparison.

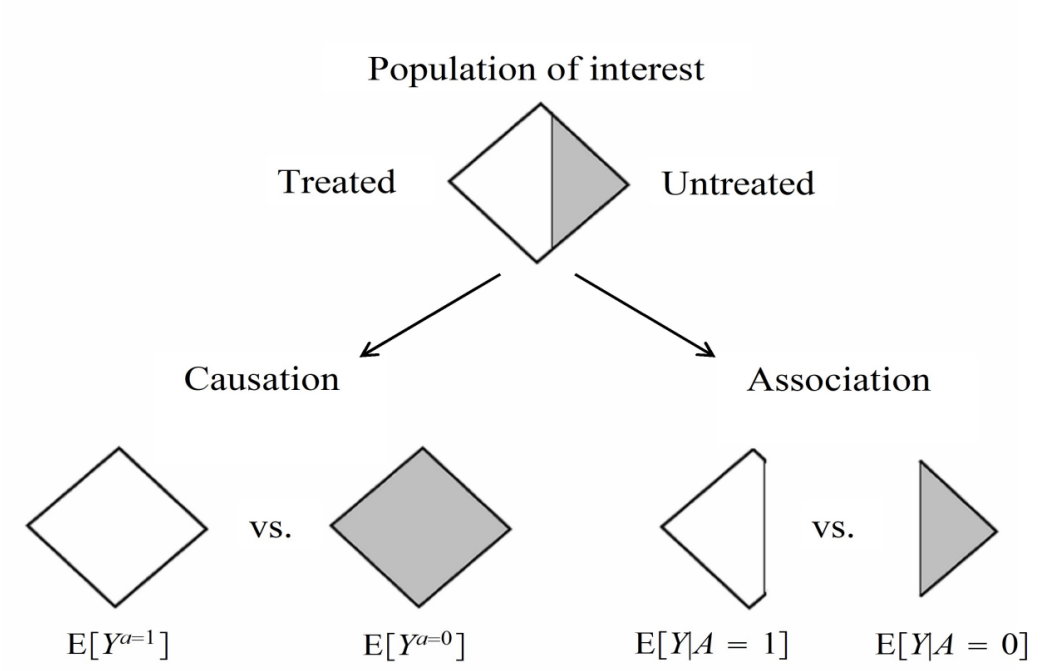
The risk $P[Y = 1|A = a]$ is a conditional probability: the risk of Y in the subset of the population that meets the condition *'having actually received treatment value a'* (i.e., $A = a$). In contrast the risk $P[Y^a = 1]$ is an unconditional probability, also known as marginal-probability, the risk of Y^a in the entire population. **Association** is defined by differences in risk in two disjoint subsets of the population determined by the individual's actual treatment value ($A = 1$ or $A = 0$), whereas **causation** is defined by differences in risk in the same population under two different treatment values ($a = 1$ or $a = 0$).

Example: Aspirin study

Suppose the **causal risk ratio** of 5-year mortality for aspirin vs. no aspirin is **0.5**. The corresponding **associational risk ratio** is **1.5** because individuals at high risk of cardiovascular death are more likely to be prescribed aspirin. A physician, observing the higher risk in the treated group, incorrectly concludes that aspirin increases mortality. As a result, she decides to **withhold aspirin** from her patients. This decision is flawed because it ignores the difference between **causation** and **association**. The physician's action leads to harm, as aspirin actually reduces mortality risk.

Summary of Chapter 1

Causal effects require counterfactual comparisons. Individual causal effects are difficult to observe directly. Average causal effects provide population-level estimates. Association is not equivalent to causation. Careful study design and statistical adjustments are needed. Confounders can create spurious associations. Only randomized experiments or strong statistical methods can establish causation.



The definition of causation implies a contrast between the whole white diamond (all individuals treated) and the whole grey diamond (all individuals untreated), whereas association implies a contrast between the white (the treated) and the grey (the untreated) areas of the original diamond. That is, inferences about causation are concerned with *what if* questions in counterfactual worlds, such as "what would be the risk if everybody had been treated?"

Exchangeability

In a real-world study, we do not observe both potential outcomes $Y^{a=1}$ under treatment and $Y^{a=0}$ under no treatment for an individual. Instead, we only know the observed outcome Y under the treatment value A that the individual actually received. Randomized experiments inherently contain missing counterfactual data, but randomization ensures that missing values occur by chance, allowing for consistent estimation of effect measures.

Suppose we flip a biased coin to assign individuals to groups: heads (to the grey group, untreated) or tails (to the white group, treated). Treatment assignment is randomized. The observed risks in the treated and untreated groups, $Pr[Y = 1|A = 1] = 0.3$ and $Pr[Y = 1|A = 0] = 0.6$, lead to an associational risk ratio of 0.5 and a risk difference of -0.3. If the treatment assignment were reversed, the computed risks and association measures would remain the same.

Exchangeability

Exchangeability implies that the risk of death in the white group would be the same as in the grey group had the white group received the grey group's treatment, meaning:

$$Pr[Y^a = 1|A = 1] = Pr[Y^a = 1|A = 0] = Pr[Y^a = 1].$$

Randomization ensures that treatment does not predict the counterfactual outcome, making treatment and counterfactual outcomes independent: $Y^a \perp A$ for all a.

Ideal randomized experiments allow us to compute counterfactual risks under treatment and no treatment for the population, yielding a causal risk ratio of 0.5 and a causal risk difference of -0.3. **Thus, in such experiments, association equals causation**.

Randomization makes all counterfactual outcomes jointly independent of treatment, leading to full exchangeability.

Exchangeability: Difference between $Y^a \perp A$ & $Y \perp A$

Another explanation for exchangeability $Y^a \perp A$ in a randomized experiment is that the counterfactual outcome Y^a, like genetic makeup, is a fixed characteristic of an individual that exists before treatment is assigned.

Since Y^a represents what would have been the outcome under treatment a, it does not depend on the treatment that is actually received.

It is crucial to distinguish between $Y^a \perp A$ and $Y \perp A$:

- $Y \perp A$ means independence between the observed outcome and the observed treatment.
- $Y^a \perp A$ means that the treated and the untreated would have experienced the same risk of death if they had received the same treatment level (either $a = 0$ or $a = 1$).
- For example, if the treatment has a causal effect, then $Y^{a=1} \neq Y^{a=0}$. Since $Y = Y^A$, then Y^a with a evaluated at the observed treatment A is the observed Y^A, which depends on A, and thus will not be independent of A.

Conditional Randomization

- **Study Overview:**
  - Treatment (A): 1 if transplanted, 0 otherwise.
  - Prognostic Factor (L): 1 if critical, 0 otherwise.
  - Outcome (Y): 1 if died, 0 otherwise.
- **Types of Randomization:**
  - **Marginal Randomization:** refers to randomly assigning treatments to individuals within a population without considering any other variables, essentially looking at the overall effect across the entire group.
    A single probability (65%) is used for all individuals.
  - **Conditional Randomization:** randomly assigning treatments within subgroups of the population, where randomization is conditioned on the values of specific variables, allowing us to analyze treatment effects within those subgroups. Treatment probability depends on L (75% if $L = 1$, 50% if $L = 0$).
- **Key Implications:**
  - Marginal design ensures exchangeability: $Y^a \perp\!\!\!\perp A$.
  - Conditional design lead to imbalance: $Y^a \perp\!\!\!\perp A|L$, not $Y^a \perp\!\!\!\perp A$.
  - Requires stratification to estimate causal effects correctly.

Example: Heart Transplant Study

- **Observed Data:**
  - 69% of treated patients were in critical condition, versus 43% untreated.
  - Indicates that treated individuals had a higher baseline risk of death.
- **Causal Effect Estimation:**
  - Marginal risk ratio: $\frac{7/13}{3/7} = 1.26$ (incorrect under conditional design).
  - Stratification approach:
    - Compute risk ratios separately for $L = 1$ and $L = 0$.
    - Weighted average provides the true causal effect.
- **Conclusion:**
  - Conditional randomization requires stratification for valid causal inference.
  - Helps adjust for baseline differences in risk factors.

Crossover Experiments: Concept and Example

- **Concept:**
  - Individuals are observed under different treatment conditions over multiple periods.
  - Helps estimate individual causal effects.
- **Example: Zeus's Blood Pressure Study**
  - Treatment (A): Use of lightning bolt.
  - Outcome (Y): Blood pressure elevation.
  - Day 1: Zeus calls a lightning strike ($A = 1$), blood pressure rises ($Y = 1$).
  - Day 2: Zeus refrains ($A = 0$), blood pressure remains normal ($Y = 0$).
  - Counterfactual outcomes: $Y^{a=1} = 1$, $Y^{a=0} = 0$.

Crossover Experiments: Mathematical Formulation

- **Mathematical Model:**
  - Individuals are observed over two periods $t = 0, 1$.
  - An individual i receives treatment A_{it} at each period t.
  - Let the (deterministic) counterfactual outcome at $t = j$ under treatment sequence for individual i if treated with a_1 at time 1 and a_0 at time 0 : $Y_{ij}^{a_0,a_1}\ \forall j = 0, 1$
  - The individual causal effect: $Y_{it}^{a_t=1} - Y_{it}^{a_t=0}$
- **Key Assumptions:**
  - (i) **No carryover effect:** $Y_{it=1}^{a_0,a_t=1} = Y_{it=1}^{a_1}$
  - (ii) **Time-invariant causal effect:** $Y_{it}^{a_t=1} - Y_{it}^{a_t=0} = \alpha_i, \quad t = 0, 1$
  - (iii) **Stable untreated outcome:** $Y_{it}^{a_t=0} = \beta_i, \quad t = 0, 1$

  Here, if individual is treated at time $1(A_{i1} = 1)$ but not time $0(A_{i0} = 0)$, then: $Y_{i1} - Y_{i0}$ is individual causal effect. And if $A_{i1} = 0$ and $A_{i0} = 1$ then: $Y_{i1} - Y_{i0}$ is individual causal effect.
- **Limitation:**
  - Crossover experiments cannot be used to study the effect of heart transplant, an irreversible action, on death, an irreversible outcome.

Risk Periods

Risk is defined as the proportion of individuals who develop the outcome of interest during a specified period.
Example: The 5-day mortality risk in the treated group is:

$$P(Y = 1 \mid A = 1)$$

which represents the proportion of treated individuals who die within the first 5 days.

Importance of Risk Period Specification:

The risk period must be clearly stated to avoid misinterpretation.
Example: Study- antibiotic therapy for elderly plague patients:

- Investigator 1: Computes 1-year risk ratio and finds a causal risk ratio of 0.05 (95% mortality reduction).
- Investigator 2: Computes 100-year risk ratio and finds a causal risk ratio of 1 (no effect, since all patients eventually die).

Both results are correct but differ due to the risk period choice.

Conclusion: When we say a treatment affects mortality, we mean it delays death, not prevents it indefinitely.

Standardization: Concept

Definition: Standardization is a method used to estimate population-level causal effects by averaging stratum-specific risks.

Heart Transplant Study: Treatment is assigned based on the patient's condition:

- $P(A = 1 \mid L = 0) = 0.5$: Noncritical patients ($L = 0$) receive a transplant ($A = 1$) with 50% probability.
- $P(A = 1 \mid L = 1) = 0.75$: Critical patients ($L = 1$) receive a transplant ($A = 1$) with 75% probability.

Observed death risks for each group:

- $P(Y = 1 \mid L = 0, A = 1) = \frac{1}{4}$: Among noncritical patients who received a transplant, 25% died.
- $P(Y = 1 \mid L = 0, A = 0) = \frac{1}{4}$: Among noncritical patients who did not receive a transplant, 25% died.
- $P(Y = 1 \mid L = 1, A = 1) = \frac{2}{3}$: Among critical patients who received a transplant, 67% died.
- $P(Y = 1 \mid L = 1, A = 0) = \frac{2}{3}$: Among critical patients who did not receive a transplant, 67% died.

Key Assumption: Conditional exchangeability

Standardization: Computing the Causal Risk Ratio

- **Goal:** Estimate the **causal risk ratio**:

$$\frac{P(Y^{a=1}=1)}{P(Y^{a=0}=1)}$$

where:

  - $P(Y^{a=1}=1)$: The risk if **everyone** in the population had been treated.
  - $P(Y^{a=0}=1)$: The risk if **no one** in the population had been treated.

- **Compute the Marginal Counterfactual Risk**

$$P(Y^a=1)=\sum_l P(Y^a=1 \mid L=l)P(L=l)$$

  - This formula takes a weighted average of risks across subgroups.
  - $P(Y^a=1 \mid L=l)$ is the risk in group $L=l$ if they all received treatment a.
  - $P(L=l)$ is the proportion of the population in group $L=l$.

Standardization: Computing the Causal Risk Ratio

- **Estimate Using Observed Data**

$$P(Y^a = 1) = \sum_l P(Y = 1 \mid L = l, A = a)P(L = l)$$

  - Under conditional exchangeability, we replace counterfactual risks with observed risks.
  - This gives an estimate of what the population-level risk would be under full treatment or no treatment.

- **Compute the Causal Risk Ratio**

$$\frac{P(Y^{a=1} = 1)}{P(Y^{a=0} = 1)} = \frac{\sum_l P(Y = 1 \mid L = l, A = 1)P(L = l)}{\sum_l P(Y = 1 \mid L = l, A = 0)P(L = l)}$$

  - This ratio compares the estimated risk under universal treatment vs. no treatment.
  - In our heart transplant study, both numerator and denominator equal 0.5, so the causal risk ratio is 1.
  - Interpretation: The transplant had **no causal effect** on mortality in this population.

Inverse Probability Weighting (IPW)

- **Goal:** Estimate the causal risk ratio by creating a pseudo-population.
- **Key Idea:** Assign weights to individuals based on their probability of treatment.
- **Exchangeability Condition:** $Y^a \perp\!\!\!\perp A|L$ (i.e., given L, treatment assignment is as good as random).
- **Pseudo-Population:**
  - We construct a hypothetical population where treatment A is independent of L.
  - This allows us to estimate counterfactual outcomes.
- **Weight Calculation:**

$$w = \frac{1}{P(A|L)}$$

where individuals with low probability of treatment receive higher weights.

Computing Counterfactual Risks Using IPW

Compute Risk if Everyone Were Untreated

- Probability of being untreated:

$$P(A = 0|L = 0) = 0.5, \quad P(A = 0|L = 1) = 0.25$$

- Deaths in the untreated group:

$$P(Y^{a=0} = 1) = \frac{1}{4} \times 2 + \frac{2}{3} \times 4 = 0.5$$

Compute Risk if Everyone Were Treated

- Probability of being treated:

$$P(A = 1|L = 0) = 0.5, \quad P(A = 1|L = 1) = 0.75$$

- Deaths in the treated group:

$$P(Y^{a=1} = 1) = \frac{1}{4} \times 2 + \frac{2}{3} \times 4 = 0.5$$

Computing the Causal Risk Ratio

Compute the Causal Risk Ratio

- Formula:

$$\text{Causal Risk Ratio} = \frac{P(Y^{a=1} = 1)}{P(Y^{a=0} = 1)}$$

- Substituting values:

$$\frac{0.5}{0.5} = 1$$

- **Interpretation:** The treatment has no causal effect on the outcome.

Why Use IP Weighting?

- **IP weighting vs. Standardization**
  - Both methods estimate the same causal effect.
  - Standardization uses $P(L)$ and $P(Y|A, L)$.
  - IP weighting uses $P(A|L)$.
- **Application to Observational Studies**
  - Randomized trials may be unethical or impractical.
  - IP weighting helps adjust for confounding in observational studies.