

# Chapter 1: A Definition of Causal Effect

## Contents

<b>1</b>	<b>1.1 Individual Causal Effects (pp. 3-4)</b>	<b>1</b>
1.1	The Fundamental Problem . . . . .	2
<b>2</b>	<b>1.2 Average Causal Effects (pp. 4-7)</b>	<b>2</b>
2.1	Definition of Average Causal Effect . . . . .	2
<b>3</b>	<b>1.3 Measures of Causal Effect (pp. 7)</b>	<b>3</b>
3.1	Causal Risk Difference . . . . .	3
3.2	Causal Risk Ratio . . . . .	3
3.3	Causal Odds Ratio . . . . .	3
<b>4</b>	<b>1.4 Random Variability (pp. 7-9)</b>	<b>4</b>
4.1	Sample vs. Superpopulation . . . . .	4
4.2	Two Types of Error . . . . .	4
<b>5</b>	<b>1.5 Causation versus Association (pp. 9-12)</b>	<b>4</b>
5.1	Association Measures . . . . .	4
5.2	Definitions of Association . . . . .	5
5.3	Causation vs. Association . . . . .	5
5.4	The Challenge of Causal Inference . . . . .	5
<b>6</b>	<b>Summary</b>	<b>5</b>
<b>7</b>	<b>References</b>	<b>6</b>

As a human being, you are already familiar with causal inference’s fundamental concepts. Through sheer existence, you know what a causal effect is, understand the difference between association and causation, and you have used this knowledge consistently throughout your life. Had you not, you’d be dead. Without basic causal concepts, you would not have survived long enough to read this chapter, let alone learn to read. As a toddler, you would have jumped right into the swimming pool after seeing those who did were later able to reach the jam jar. As a teenager, you would have skied down the most dangerous slopes after seeing those who did won the next ski race. As a parent, you would have refused to give antibiotics to your sick child after observing that those children who took their medicines were not at the park the next day.

Since you already understand the definition of causal effect and the difference between association and causation, do not expect to gain deep conceptual insights from this chapter. Rather, the purpose of this chapter is to introduce mathematical notation that formalizes the causal intuition that you already possess. Make sure that you can match your causal intuition with the mathematical notation introduced here. This notation is necessary to precisely define causal concepts, and will be used throughout the book.

This content is based on Hernán and Robins (2020, chap. 1, pp. 3-12).

## 1 1.1 Individual Causal Effects (pp. 3-4)

We use a hypothetical example to introduce causal effects. Consider Zeus's extended family as our population of interest. Suppose all 20 family members have a life-threatening disease and we want to study the causal effect of heart transplant (treatment  $A$ ) on death (outcome  $Y$ ) within 5 years.

We represent treatment and outcome as binary variables:

- Let  $A = 1$  if the individual receives a heart transplant,  $A = 0$  if not
- Let  $Y = 1$  if the individual dies within 5 years,  $Y = 0$  if survives

To define a causal effect, we need to compare two **counterfactual outcomes** (also called **potential outcomes**) for each individual:

- $Y^{a=1}$ : The outcome if the individual receives treatment ( $a = 1$ )
- $Y^{a=0}$ : The outcome if the individual does not receive treatment ( $a = 0$ )

An **individual causal effect** exists when  $Y^{a=1} \neq Y^{a=0}$  for an individual. For example, if Zeus would die if transplanted ( $Y^{a=1} = 1$ ) but survive if not transplanted ( $Y^{a=0} = 0$ ), then heart transplant has a causal effect on Zeus's outcome.

These counterfactual outcomes represent what *would* happen under each treatment condition. The notation  $Y^a$  denotes the counterfactual outcome under treatment level  $a$ . This framework was formally developed in the statistical literature and is fundamental to causal inference.

## 1.1 The Fundamental Problem

The fundamental problem of causal inference is that we can only observe one of the two counterfactual outcomes for each individual. If Zeus receives a heart transplant, we observe  $Y^{a=1}$  but not  $Y^{a=0}$ . The unobserved counterfactual outcome remains unknown.

Formally, for each individual, the observed outcome  $Y$  equals the counterfactual outcome  $Y^a$  corresponding to the treatment actually received:  $Y = Y^A$ . This equality is called **consistency**.

Because of this missing data problem, individual causal effects cannot generally be identified (i.e., computed from observed data). We can never simultaneously observe both  $Y^{a=1}$  and  $Y^{a=0}$  for the same individual. This is why causal inference is fundamentally a missing data problem.

However, we can define and sometimes estimate **average causal effects** in populations, which is the focus of the next section.

## 2 1.2 Average Causal Effects (pp. 4-7)

---

Since individual causal effects cannot be identified, we focus on **average causal effects** in a population. Table 1.1 shows the counterfactual outcomes for all 20 members of Zeus's family.

From Table 1.1, we can compute:

- Risk if all treated:  $Pr[Y^{a=1} = 1] = 10/20 = 0.5$
- Risk if all untreated:  $Pr[Y^{a=0} = 1] = 10/20 = 0.5$

### 2.1 Definition of Average Causal Effect

An **average causal effect** of treatment  $A$  on outcome  $Y$  is present if:

$$Pr[Y^{a=1} = 1] \neq Pr[Y^{a=0} = 1]$$

or equivalently (using expected values):

$$E[Y^{a=1}] \neq E[Y^{a=0}]$$

In our population, treatment does *not* have an average causal effect because both risks equal 0.5. The **null hypothesis of no average causal effect** holds. However, this does not mean there are no individual effects.

Even though the average causal effect is null, 12 individuals in Table 1.1 have individual causal effects. Six were harmed by treatment ( $Y^{a=1} - Y^{a=0} = 1$ ), including Zeus, and six were helped ( $Y^{a=1} - Y^{a=0} = -1$ ). The harmful and beneficial effects cancel out in the average.

The average causal effect  $E[Y^{a=1}] - E[Y^{a=0}]$  always equals the average of individual causal effects  $E[Y^{a=1} - Y^{a=0}]$ , because a difference of averages equals the average of the differences.

When there is no causal effect for *any* individual (i.e.,  $Y^{a=1} = Y^{a=0}$  for all individuals), we say the **sharp causal null hypothesis** is true. The sharp null implies the null hypothesis of no average effect.

### 3 1.3 Measures of Causal Effect (pp. 7)

---

When a causal effect exists, we can quantify its magnitude using different **effect measures**. The three most common for binary outcomes are:

#### 3.1 Causal Risk Difference

$$Pr[Y^{a=1} = 1] - Pr[Y^{a=0} = 1]$$

This additive measure equals zero under the null hypothesis. It measures the absolute difference in risk.

#### 3.2 Causal Risk Ratio

$$\frac{Pr[Y^{a=1} = 1]}{Pr[Y^{a=0} = 1]}$$

This multiplicative measure equals one under the null hypothesis. It measures how many times treatment increases (or decreases) the risk.

#### 3.3 Causal Odds Ratio

$$\frac{Pr[Y^{a=1} = 1]/Pr[Y^{a=1} = 0]}{Pr[Y^{a=0} = 1]/Pr[Y^{a=0} = 0]}$$

This also equals one under the null hypothesis.

Different effect measures serve different purposes. For example, suppose 3 in a million would develop the outcome if treated, and 1 in a million if untreated:

- Causal risk ratio = 3 (treatment triples the risk)
- Causal risk difference = 0.000002 (2 additional cases per million treated)

The risk ratio is useful for understanding relative risk, while the risk difference is useful for understanding the absolute number of cases attributable to treatment.

The causal risk difference is the average of individual causal effects  $Y^{a=1} - Y^{a=0}$  on the difference scale. However, the causal risk ratio is *not* the average of individual causal effects  $Y^{a=1}/Y^{a=0}$  on the ratio scale—it is a measure of causal effect in the population but not an average of individual effects.

## 4 1.4 Random Variability (pp. 7-9)

---

In practice, we do not observe the counterfactual outcomes in Table 1.1. We only observe data from a sample of individuals. This introduces **random variability**.

### 4.1 Sample vs. Superpopulation

We can view our study population in two ways:

1. **Fixed population:** The 20 individuals are the entire population of interest
2. **Random sample:** The 20 individuals are a random sample from a larger “superpopulation”

Under the superpopulation perspective, even if the true average causal effect is zero, our sample estimate might not be exactly zero due to **sampling variability**.

For example, suppose we randomly sample 20 individuals from a superpopulation where  $E[Y^{a=1}] = E[Y^{a=0}] = 0.5$  (true null effect). We might observe  $\hat{E}[Y^{a=1}] = 0.55$  and  $\hat{E}[Y^{a=0}] = 0.45$  in our sample just by chance.

**Statistical inference** helps us determine whether an observed association is due to a true causal effect or random variability. We use:

- **Confidence intervals:** Range of plausible values for the causal effect
- **Hypothesis tests:** Probability of observing our data if the null is true
- **P-values:** Strength of evidence against the null hypothesis

Until Chapter 10, we assume our population is extremely large to avoid statistical complications and focus on causal inference concepts.

### 4.2 Two Types of Error

It’s critical to distinguish:

1. **Random error:** Due to sampling variability; reduced by larger sample sizes
2. **Systematic error (bias):** Due to flaws in study design; NOT reduced by larger samples

With a very large sample, random error becomes negligible, and systematic error dominates. This is why proper study design and appropriate methods are essential for causal inference.

## 5 1.5 Causation versus Association (pp. 9-12)

---

A key distinction in causal inference is between **causation** and **association**.

### 5.1 Association Measures

Unlike causal effects (defined by counterfactuals), **associations** are defined using observed data. Table 1.2 shows the observed treatment and outcomes for Zeus’s family.

In reality, we observe only one outcome per person—the one corresponding to their actual treatment. In Table 1.2, 13 individuals received transplants ( $A = 1$ ) and 7 did not ( $A = 0$ ). Among the 13 treated, 7 died ( $Y = 1$ ). Among the 7 untreated, 3 died ( $Y = 1$ ).

The **associational risk** in the treated:  $Pr[Y = 1|A = 1] = 7/13$

The **associational risk** in the untreated:  $Pr[Y = 1|A = 0] = 3/7$

## 5.2 Definitions of Association

Treatment  $A$  and outcome  $Y$  are **independent** (not associated) when:

$$Pr[Y = 1|A = 1] = Pr[Y = 1|A = 0]$$

or equivalently:  $E[Y|A = 1] = E[Y|A = 0]$

This is denoted  $Y \perp\!\!\!\perp A$ .

When this equality does not hold,  $A$  and  $Y$  are **associated** or **dependent**. Association measures include:

- **Associational risk difference:**  $Pr[Y = 1|A = 1] - Pr[Y = 1|A = 0]$
- **Associational risk ratio:**  $Pr[Y = 1|A = 1]/Pr[Y = 1|A = 0]$
- **Associational odds ratio:**  $\frac{Pr[Y=1|A=1]/Pr[Y=0|A=1]}{Pr[Y=1|A=0]/Pr[Y=0|A=0]}$

## 5.3 Causation vs. Association

In our example:

- **No causal effect:**  $Pr[Y^{a=1} = 1] = Pr[Y^{a=0} = 1] = 0.5$
- **Association present:**  $Pr[Y = 1|A = 1] = 0.54 \neq 0.43 = Pr[Y = 1|A = 0]$

This demonstrates the fundamental principle: **Association is not causation**.

The distinction is:

- **Causation:** Compares the *same* population under different treatment values
  - “What if everyone had been treated?” vs. “What if everyone had been untreated?”
  - Uses counterfactual risks:  $Pr[Y^{a=1} = 1]$  vs.  $Pr[Y^{a=0} = 1]$
- **Association:** Compares *different subsets* of the population
  - “What is the risk in those who were treated?” vs. “What is the risk in those who were untreated?”
  - Uses conditional risks:  $Pr[Y = 1|A = 1]$  vs.  $Pr[Y = 1|A = 0]$

In our example, there’s an association because those who received transplants were sicker on average than those who didn’t. This discrepancy between causation and association is called **confounding** (discussed in Chapter 7).

## 5.4 The Challenge of Causal Inference

Causal inference requires data like Table 1.1 (all counterfactual outcomes), but we only have data like Table 1.2 (observed outcomes). The question is: Under which conditions can real-world data be used for causal inference?

Chapter 2 provides one answer: conduct a randomized experiment.

## 6 Summary

---

This chapter introduced fundamental concepts:

1. **Individual causal effects:** Defined as  $Y^{a=1} \neq Y^{a=0}$ , but cannot be identified due to the fundamental problem (missing counterfactuals)
2. **Average causal effects:** Defined as  $E[Y^{a=1}] \neq E[Y^{a=0}]$ , can sometimes be identified from data
3. **Effect measures:** Include risk difference (additive), risk ratio (multiplicative), and odds ratio

4. **Random variability:** Distinguishes sampling variability (reduced by larger  $n$ ) from systematic bias (not reduced by larger  $n$ )
5. **Causation vs. association:** Causation compares counterfactual risks in the same population; association compares observed risks in different subsets

The key insight: Association does not imply causation. The challenge is to use observed associations to make valid causal inferences.

## 7 References

---

Hernán, Miguel A, and James M Robins. 2020. *Causal Inference: What If*. Chapman & Hall/CRC. <https://miguelhernan.org/whatifbook>.