Chapter 5: Interaction

Contents

Chapter 4 introduced effect modification: the phenomenon where the effect of treatment A varies across levels of another variable V. In this chapter, we extend the concept to **interaction** between two treatments.

Interaction addresses a fundamentally different question: Does the joint effect of two treatments differ from the sum of their individual effects? Understanding interaction is crucial for:

- Evaluating combination therapies
- Understanding synergistic or antagonistic effects
- Designing multi-component interventions

This chapter is based on Hernán and Robins (2020, chap. 5, pp. 47-58).

1 5.1 Interaction Requires a Joint Intervention (pp. 47-49)

To define interaction, we need to consider the **joint effect** of two treatments, which requires imagining interventions on both treatments simultaneously.

1.1 Notation for Two Treatments

Suppose we have two binary treatments:

- A: First treatment (e.g., aspirin), with levels 0 or 1
- E: Second treatment (e.g., blood pressure medication), with levels 0 or 1

This creates four possible treatment combinations: 1. $(A = 0, E = 0)$: Neither treatment 2. $(A = 1, E = 0)$: Aspirin only 3. $(A = 0, E = 1)$: Blood pressure medication only
4. $(A = 1, E = 1)$: Both treatments

For each individual, we have four counterfactual outcomes:

- $Y^{a=0,e=0}$: Outcome if neither treatment received
- $Y^{a=1,e=0}$: Outcome if only A received
- $Y^{a=0,e=1}$: Outcome if only E received
- $Y^{a=1,e=1}$: Outcome if both treatments received

Superscript notation: $Y^{a,e}$ denotes the potential outcome under the joint intervention that sets $A = a$ and $E = e$ simultaneously.

This notation generalizes naturally to more than two treatments or to continuous treatments, but we focus on the binary case for clarity.

1.2 Definition of Interaction

Definition 1.1 (Additive Interaction)**.** There is **additive interaction** (on the risk difference scale) between treatments A and E if:

$$E[Y^{a=1,e=1}] - E[Y^{a=0,e=0}] \neq (E[Y^{a=1,e=0}] - E[Y^{a=0,e=0}]) + (E[Y^{a=0,e=1}] - E[Y^{a=0,e=0}])$$

In words: the joint effect of both treatments differs from the sum of the individual effects.

Equivalently, we can define additive interaction using the **interaction contrast**:

$$\text{IC} = E[Y^{a=1,e=1}] - E[Y^{a=1,e=0}] - E[Y^{a=0,e=1}] + E[Y^{a=0,e=0}]$$

If $\text{IC} \neq 0$, there is additive interaction.

Interpretation of IC:

- If $\text{IC} > 0$: **Synergism** or **superadditivity** (combined effect exceeds sum of individual effects)
- If $\text{IC} < 0$: **Antagonism** or **subadditivity** (combined effect is less than sum of individual effects)
- If $\text{IC} = 0$: **No interaction** or **additivity** (combined effect equals sum of individual effects)

Important distinction:

- **Effect modification** (Chapter 4): Effect of A varies across levels of a baseline variable V
- **Interaction** (Chapter 5): Joint effect of two treatments A and E differs from sum of individual effects

Effect modification involves one treatment and one baseline covariate; interaction involves two (or more) treatments.

1.3 Example: Interaction Between Aspirin and Exercise

Example 1.1 (Interaction Example)**.** Suppose we study the effect of aspirin and exercise on heart attack risk (with lower values being better):

Treatment Combination	Risk
No aspirin, no exercise: $E[Y^{a=0,e=0}]$	0.30
Aspirin only: $E[Y^{a=1,e=0}]$	0.25

Treatment Combination	Risk
Exercise only: $E[Y^{a=0,e=1}]$	0.20
Both aspirin and exercise: $E[Y^{a=1,e=1}]$	0.10

Individual effects:

- Effect of aspirin (without exercise): $0.25 - 0.30 = -0.05$
- Effect of exercise (without aspirin): $0.20 - 0.30 = -0.10$
- Sum of individual effects: $-0.05 + (-0.10) = -0.15$

Joint effect:

- $0.10 - 0.30 = -0.20$

Interaction contrast:

- IC $= -0.20 - (-0.15) = -0.05$

There is **negative** additive interaction (antagonism is not the right word here; rather, the combined benefit exceeds the sum). Actually, let me recalculate:

IC $= 0.10 - 0.25 - 0.20 + 0.30 = -0.05$

Since IC < 0, there is negative interaction, but in this case both treatments are beneficial, and their combined effect (-0.20) is actually more beneficial than the sum of individual effects (-0.15). This is **synergism**: the combination is more effective than expected from adding individual effects.

Confusion about signs: The sign of the interaction contrast depends on the coding of the outcome:

- If higher Y is worse (e.g., risk of disease), negative IC means synergistic benefit
- If higher Y is better (e.g., recovery), positive IC means synergistic benefit

To avoid confusion, always interpret the interaction contrast in the context of the specific outcome and treatment effects.

Alternative terminology:

- **Positive interaction / Synergism / Superadditivity**: |Joint effect| > |Sum of individual effects|
- **Negative interaction / Antagonism / Subadditivity**: |Joint effect| < |Sum of individual effects|

2 5.2 Identifying Interaction (pp. 49-51)

To identify interaction from observed data, we need the three identifiability conditions introduced in Chapter 3:

2.1 Identifiability Conditions for Interaction

1. **Exchangeability** for the joint treatment (A, E):

$$Y^{a,e} \perp\!\!\!\perp (A, E) \quad \text{for all } (a, e)$$

2. **Positivity** for the joint treatment:

$$Pr[A = a, E = e] > 0 \quad \text{for all } (a, e)$$

3. **Consistency**:

$$Y = Y^{A,E}$$

   No interference, well-defined treatments

Key point: We treat (A, E) as a single joint treatment with four levels: $(0, 0), (1, 0), (0, 1), (1, 1)$.

Randomized experiments:

- If both A and E are randomized independently, exchangeability and positivity typically hold
- This is a **factorial design**: randomize to all four combinations
- Allows unbiased estimation of interaction

Observational studies:

- Must adjust for confounders of (A, E) relationship with Y
- This may include confounders of A, confounders of E, and confounders of their joint distribution
- Positivity requires that all four treatment combinations occur at all levels of confounders

2.2 Estimating Interaction from Data

Under the identifiability conditions, we can estimate the interaction contrast:

In a randomized experiment:

$$\widehat{\text{IC}} = \bar{Y}_{A=1,E=1} - \bar{Y}_{A=1,E=0} - \bar{Y}_{A=0,E=1} + \bar{Y}_{A=0,E=0}$$

where $\bar{Y}_{a,e}$ is the average observed outcome in the group assigned to treatment combination (a, e).

In an observational study with confounders L: Use standardization, IP weighting, or regression to adjust for L before computing the interaction contrast.

Sample size considerations: Estimating interaction requires sufficient sample size in all four treatment groups. The standard error of the interaction contrast is typically larger than standard errors of main effects, so interaction estimates have less precision.

Multiple testing: When examining multiple potential interactions, be mindful of multiple comparisons and the increased risk of false positives.

3 5.3 Counterfactual Response Types and Interaction (pp. 51-53)

Another way to understand interaction is through **counterfactual response types**—classifications of individuals based on their pattern of potential outcomes.

3.1 Four Response Types (Binary Outcome)

For a binary outcome, individuals can be classified into one of four types based on their four potential outcomes:

1. **Doomed**: $Y^{0,0} = Y^{1,0} = Y^{0,1} = Y^{1,1} = 1$ (outcome occurs regardless of treatment)
2. **Preventive via A only**: $Y^{0,0} = Y^{0,1} = 1$, but $Y^{1,0} = Y^{1,1} = 0$
3. **Preventive via E only**: $Y^{0,0} = Y^{1,0} = 1$, but $Y^{0,1} = Y^{1,1} = 0$
4. **Preventive via both**: $Y^{0,0} = 1$, but $Y^{1,0} = Y^{0,1} = Y^{1,1} = 0$
5. **Immune**: $Y^{0,0} = Y^{1,0} = Y^{0,1} = Y^{1,1} = 0$ (outcome never occurs)

And several other combinations…

Actually, with four binary potential outcomes, there are $2^4 = 16$ possible response types. Many of these correspond to various patterns of interaction:

- **Synergistic** types: Individuals for whom the outcome is prevented only by the combination of both treatments
- **Redundant** types: Individuals who benefit from either treatment alone (treatments are substitutes)
- **Antagonistic** types: Individuals who benefit from one treatment alone but not from the combination

These response types are **unobservable**—we can never know which type an individual is because we only observe one potential outcome. However, the *distribution* of response types in the population determines whether interaction exists.

3.2 Relationship to Interaction

Interaction on the additive scale corresponds to certain response types being more or less common than others.

Example 3.1 (Response Types Example)**.** Suppose we have:

- 30% of population: Prevented only by A
- 20% of population: Prevented only by E
- 10% of population: Prevented only by both A and E together
- 40% of population: Immune (never get disease)

Then:

- $E[Y^{0,0}] = 0.30 + 0.20 + 0.10 = 0.60$
- $E[Y^{1,0}] = 0.20 + 0.10 = 0.30$ (the 30% prevented by A no longer get disease)
- $E[Y^{0,1}] = 0.30 + 0.10 = 0.40$ (the 20% prevented by E no longer get disease)
- $E[Y^{1,1}] = 0.10$ (only the synergistic group still gets disease)

Interaction contrast:

$$\text{IC} = 0.10 - 0.30 - 0.40 + 0.60 = 0$$

Despite the synergistic response type existing, there is no interaction on the additive scale because the effects balance out.

This example illustrates that:

- Existence of synergistic response types does **not** guarantee additive interaction will be detected
- Interaction depends on the *prevalence* of different response types
- Different scales (additive vs. multiplicative) may reveal interaction even when others do not

The response type framework provides a mechanistic interpretation of interaction but requires strong assumptions (monotonicity, no redundancy) to map cleanly onto observed interaction measures.

4 5.4 Sufficient Causes (pp. 53-55)

An alternative framework for thinking about interaction is the **sufficient cause model**, developed by Kenneth Rothman.

Definition 4.1 (Sufficient Cause)**.** A **sufficient cause** is a set of conditions that inevitably produces the outcome. Once all components of a sufficient cause are present, the outcome occurs.

4.1 Components of Sufficient Causes

Each sufficient cause consists of **component causes**—individual factors that, together, are sufficient to produce the outcome.

Example 4.1 (Sufficient Cause Example)**.** Consider lung cancer. Possible sufficient causes:

Sufficient Cause I: Smoking + Genetic variant A + Occupational asbestos exposure

Sufficient Cause II: Radon exposure + Genetic variant B + Poor diet

Sufficient Cause III: Smoking + Radon exposure + Genetic variant C

An individual will develop lung cancer if they complete any one (or more) of these sufficient causes by acquiring all of its component causes.

Key features of the sufficient cause model:

1. **Deterministic**: If all components of a sufficient cause are present, the outcome occurs with certainty (in this model)
2. **Multiple sufficient causes**: There may be many different ways to produce the same outcome

3. **Complementary components**: Within a sufficient cause, all components are necessary for that particular pathway to cause the outcome
4. **Unknown components**: We typically don't know all components of sufficient causes; many may be unmeasured or unknown

Critique: The deterministic nature is unrealistic for most health outcomes, which involve stochastic processes. However, the model is useful conceptually for understanding interaction.

5 5.5 Sufficient Cause Interaction (pp. 55-57)

Within the sufficient cause framework, we can define a specific type of interaction:

Definition 5.1 (Sufficient Cause Interaction)**.** There is **sufficient cause interaction** between A and E if they are both component causes of the same sufficient cause.

In other words, A and E interact if there exists some sufficient cause containing both A and E as components.

5.1 Sufficient Cause Interaction and Synergism

Sufficient cause interaction corresponds to **synergism**:

- Neither A alone nor E alone completes a sufficient cause
- But A and E together (plus other unknown components) form a sufficient cause
- Therefore, the joint effect exceeds the sum of individual effects

Example 5.1 (Sufficient Cause Interaction Example)**.** Consider the effect of smoking (A) and asbestos exposure (E) on lung cancer (Y):

Sufficient Cause I: Smoking + Asbestos + Unknown factors U **Sufficient Cause II**: Smoking + Unknown factors U
Sufficient Cause III: Asbestos + Unknown factors U

If an individual has Unknown factors U but not U or U :

- $Y^{a=0,e=0} = 0$ (no sufficient cause completed)
- $Y^{a=1,e=0} = 0$ (smoking alone doesn't complete a sufficient cause)
- $Y^{a=0,e=1} = 0$ (asbestos alone doesn't complete a sufficient cause)
- $Y^{a=1,e=1} = 1$ (Sufficient Cause I completed)

This is **pure synergism**: neither treatment alone has an effect, but the combination does.

Relationship between sufficient cause interaction and causal interaction:

- **Sufficient cause interaction is a mechanistic concept**: It refers to biological, chemical, or social mechanisms by which factors combine to produce an outcome
- **Causal interaction (in potential outcomes framework) is a statistical concept**: It refers to departures from additivity in population-level effect measures

These concepts are related but not identical:

- Sufficient cause interaction (synergism) implies positive additive interaction in the population *if* the sufficient cause is common enough
- However, positive additive interaction can occur even without sufficient cause interaction (e.g., if treatments prevent the outcome via independent pathways but with different magnitudes)

Modern perspective: The sufficient cause model has limitations (determinism, untestable assumptions) but provides useful intuition about mechanistic synergy. The potential outcomes framework is more general and forms the basis for modern causal inference methods.

6 5.6 Counterfactuals or Sufficient-Component Causes? (pp. 57-58)

Should we use the counterfactual (potential outcomes) framework or the sufficient cause framework to study interaction?

6.1 Advantages of the Counterfactual Framework

1. **More general**: Applies to any outcome (binary, continuous, time-to-event) and any treatments
2. **Testable implications**: Under identifiability conditions, interaction can be estimated from data
3. **Rigorous mathematical foundation**: Clear notation and assumptions
4. **Connects to statistical methods**: Directly links to regression, weighting, standardization

6.2 Advantages of the Sufficient Cause Framework

1. **Mechanistic intuition**: Provides a way to think about biological/physical mechanisms
2. **Useful for binary outcomes**: Particularly clear for presence/absence of disease
3. **Historical importance**: Widely used in epidemiology
4. **Qualitative insights**: Helps conceptualize synergism and prevention

Recommendation from Hernán & Robins: Use the counterfactual framework as the primary approach, but keep sufficient cause intuition in mind for biological interpretation.

Why prioritize counterfactuals?

- Sufficient cause model makes strong untestable assumptions (determinism, completeness of sufficient causes)
- Counterfactual framework is more flexible and connects better to modern statistical methods
- Most interaction research is in the potential outcomes tradition

When to use sufficient cause thinking?

- When seeking mechanistic explanations
- When biological synergism is the scientific question
- When teaching interaction to those unfamiliar with potential outcomes
- As a conceptual tool alongside formal potential outcomes analysis

Integration: Many modern papers use potential outcomes for formal definitions and estimation, then discuss findings in terms of sufficient cause mechanisms.

7 Summary

This chapter introduced **interaction** between two treatments.

Key concepts:

1. **Definition**: Interaction exists when the joint effect of two treatments differs from the sum of their individual effects:

$$\text{IC} = E[Y^{a=1,e=1}] - E[Y^{a=1,e=0}] - E[Y^{a=0,e=1}] + E[Y^{a=0,e=0}] \neq 0$$

2. **Identification**: Requires exchangeability, positivity, and consistency for the joint treatment (A, E)
3. **Response types**: Interaction can be understood through the distribution of counterfactual response types in the population

4. **Sufficient causes**: The sufficient cause model provides a mechanistic framework where interaction corresponds to synergism
5. **Frameworks**: The counterfactual (potential outcomes) framework is recommended for formal analysis; sufficient causes provide mechanistic intuition
6. **Scale dependence**: Whether interaction exists depends on the scale (additive, multiplicative, etc.)

Distinction from Chapter 4:

- **Effect modification** (Chapter 4): Effect of treatment A varies by baseline variable V
- **Interaction** (Chapter 5): Joint effect of treatments A and E differs from sum of individual effects

Looking ahead:

- **Chapter 6**: Causal diagrams provide graphical tools for understanding when and how to adjust for confounding
- **Chapter 7**: Detailed treatment of confounding
- **Chapter 13**: The g-formula provides a flexible method for estimating effects when both confounding and interaction are present
- **Chapter 14**: G-estimation of structural nested models for treatment effect heterogeneity

Practical takeaway:

- When studying combination therapies or multiple exposures, consider whether interaction exists
- Use factorial designs (randomize both treatments independently) when possible
- Report joint effects and interaction contrasts, not just main effects
- Interpret interaction in light of both statistical and mechanistic frameworks

8 References

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