Chapter 4: Effect Modification

Contents

In previous chapters, we focused on estimating average causal effects in a population. But what if the effect of treatment varies across individuals? This chapter introduces **effect modification**: the phenomenon where the causal effect of treatment differs across levels of another variable.

Understanding effect modification is crucial for:

- Identifying subgroups that benefit more (or less) from treatment
- Targeting interventions to those who need them most
- Understanding biological or social mechanisms

This chapter is based on Hernán and Robins (2020, chap. 4, pp. 37-46).

1 4.1 Heterogeneity of Treatment Effects (pp. 37-39)

Consider again Zeus's family from Chapter 1. Suppose we estimate that heart transplant has no average causal effect on mortality (the average treatment effect is zero). Does this mean the treatment has no effect on anyone?

Not necessarily. The average treatment effect can be zero even when there are large individual effects, as long as beneficial effects for some individuals are balanced by harmful effects for others.

Definition 1.1 (Effect Modification)**.** **Effect modification** (also called **heterogeneity of treatment effects**) is present when the average causal effect of treatment differs across levels of another variable V.
For a binary modifier V, effect modification exists when:

$$\mathrm{E}[Y^{a=1} - Y^{a=0} | V = 1] \neq \mathrm{E}[Y^{a=1} - Y^{a=0} | V = 0]$$

1.1 Example: Effect Modification by Sex

Suppose in Zeus's family:

- Among males: $\mathrm{E}[Y^{a=1} - Y^{a=0} | \text{Sex} = \text{male}] = -0.3$ (benefit from transplant)
- Among females: $\mathrm{E}[Y^{a=1} - Y^{a=0} | \text{Sex} = \text{female}] = +0.3$ (harm from transplant)

The average effect across the whole family is zero, but there is strong effect modification by sex. The treatment has opposite effects in males and females.

Important terminology:

- V is called an **effect modifier** or **modifier**
- The causal effect **varies** or is **heterogeneous** across levels of V
- Effect modification is a **property of the population**, not of individuals

Notation: We write $\mathrm{E}[Y^{a=1} - Y^{a=0} | V = v]$ to denote the average causal effect in the subpopulation with $V = v$. This is a **conditional average treatment effect**.

1.2 Effect Modification Depends on the Scale

Whether effect modification exists can depend on the **scale** (measure) used:

- **Risk difference scale**: $\mathrm{E}[Y^{a=1} | V] - \mathrm{E}[Y^{a=0} | V]$
- **Risk ratio scale**: $\mathrm{E}[Y^{a=1} | V] / \mathrm{E}[Y^{a=0} | V]$

Example 1.1 (Scale-Dependent Effect Modification)**.** Suppose:

- When $V = 0$: $\mathrm{E}[Y^{a=1}] = 0.2$, $\mathrm{E}[Y^{a=0}] = 0.1$
- When $V = 1$: $\mathrm{E}[Y^{a=1}] = 0.4$, $\mathrm{E}[Y^{a=0}] = 0.2$

Risk difference:

- When $V = 0$: $0.2 - 0.1 = 0.1$
- When $V = 1$: $0.4 - 0.2 = 0.2$

There is effect modification on the risk difference scale (effect is 0.1 in one group, 0.2 in the other).
Risk ratio:

- When $V = 0$: $0.2/0.1 = 2.0$
- When $V = 1$: $0.4/0.2 = 2.0$

There is NO effect modification on the risk ratio scale (effect is 2.0 in both groups).

This example shows that effect modification is **measure-dependent**. The same data can show:

- Effect modification on one scale (risk difference)
- No effect modification on another scale (risk ratio)

Which scale should we use?

- From a **biological/mechanistic** perspective, one scale may be more natural
- From a **public health** perspective, risk differences are often most relevant
- There is no single "correct" answer—the choice depends on the scientific question

This highlights that effect modification is not simply a yes/no property, but depends on how we measure effects.

2 4.2 Stratification to Identify Effect Modification (pp. 39-41)

The standard approach to identify effect modification is **stratification**: estimate the treatment effect separately within levels (strata) of the modifier variable V.

2.1 Stratified Analysis

To detect effect modification by variable V:

1. Divide the population into strata defined by V (e.g., males and females)
2. Estimate the average causal effect within each stratum
3. Compare effects across strata

If effects differ across strata, V is an effect modifier.

Example 2.1 (Stratified Analysis Example)**.** Using data from a randomized trial of heart transplant in Zeus's family:
Stratum 1 (Males):
- Risk in treated: $Pr[Y = 1|A = 1, V = \text{male}] = 0.40$
- Risk in untreated: $Pr[Y = 1|A = 0, V = \text{male}] = 0.70$
- Risk difference: $0.40 - 0.70 = -0.30$ (benefit)

Stratum 2 (Females):
- Risk in treated: $Pr[Y = 1|A = 1, V = \text{female}] = 0.65$
- Risk in untreated: $Pr[Y = 1|A = 0, V = \text{female}] = 0.35$
- Risk difference: $0.65 - 0.35 = +0.30$ (harm)

Since $-0.30 \neq +0.30$, there is effect modification by sex.

Key points about stratification:

1. **Randomized experiments**: In a randomized experiment, stratification by any baseline variable V yields valid estimates of stratum-specific causal effects (because randomization ensures exchangeability within each stratum).

2. **Observational studies**: In observational studies, stratification by V only yields valid estimates if:

   - We also adjust for all confounders
   - Confounders may differ across strata of V

3. **Multiple modifiers**: We can stratify by multiple variables simultaneously (e.g., sex and age), but:

   - The number of strata grows quickly (e.g., 2 sexes $\times$ 4 age groups = 8 strata)
   - Sample sizes within strata may become small
   - This is sometimes called **fine stratification**

4. **Statistical testing**: Formal statistical tests can assess whether observed differences across strata are consistent with chance (e.g., interaction tests in regression). However, lack of statistical significance does not mean absence of effect modification—it may simply reflect lack of power.

3 4.3 Why Care About Effect Modification (pp. 41-42)

Why is identifying effect modification important?

3.1 1. Improving Precision of Effect Estimates

If treatment effects vary substantially across subgroups, reporting only an average effect can be misleading. Stratum-specific estimates provide more precise information about who benefits from treatment.

3.2 2. Targeting Interventions

When effect modification exists, we can:

- Target treatment to subgroups that benefit most
- Avoid treating subgroups that experience harm
- Allocate limited resources more efficiently

Example 3.1 (Precision Medicine Example)**.** If genetic testing reveals that a drug benefits patients with genotype AA but harms patients with genotype BB, we should:

- Prescribe the drug only to AA patients
- Use alternative treatments for BB patients

This is the foundation of **precision medicine** or **personalized medicine**.

3.3 3. Understanding Mechanisms

Effect modification can provide clues about biological or social mechanisms:

- If an effect is modified by sex, hormones may play a role
- If an effect is modified by age, developmental processes may be involved
- If an effect is modified by socioeconomic status, access or adherence may matter

Caution: Effect modification is **descriptive**, not **mechanistic**. Just because V modifies the effect of A on Y does not necessarily mean V is part of the causal mechanism. There may be other variables correlated with V that are the true modifiers.

Example: Suppose the effect of a drug is modified by zip code. This doesn't mean zip code *causes* differential effects. Rather, zip code may be a proxy for socioeconomic status, healthcare access, environmental exposures, etc.

To understand mechanisms, we need:

- Subject-matter knowledge
- Careful measurement of mechanistic variables
- Additional assumptions (sometimes formalized using mediation analysis, covered in later chapters)

4 4.4 Stratification as a Form of Adjustment (pp. 42-43)

Stratification serves two related but distinct purposes:

1. **Identifying effect modification**: Are effects different across strata of V?
2. **Controlling for confounding**: Is V a confounder that biases the marginal (unstratified) effect?

These are different scientific questions with different implications.

Definition 4.1 (Confounder vs. Effect Modifier)**.**

- A **confounder** is a variable that, if not adjusted for, biases the estimate of the average causal effect
- An **effect modifier** is a variable across which the causal effect differs

A variable can be:

- A confounder only

- An effect modifier only

- Both a confounder and effect modifier
- Neither

4.1 Example: Confounder vs. Modifier

Confounder only: Age affects both treatment and outcome, but the treatment effect is the same at all ages.

- We must adjust for age to get an unbiased marginal effect
- But we don't need to report age-specific effects (they're all the same)

Effect modifier only: In a randomized trial, sex does not confound (randomization handles that), but the treatment effect differs by sex.

- No need to adjust for confounding
- But we should report sex-specific effects

Both: In an observational study of surgery, age affects who gets surgery (confounding) and how well surgery works (effect modification).

- We must adjust for age
- We should report age-specific effects

Key distinction:

- **Confounding is a bias** to be removed through adjustment
- **Effect modification is a finding** to be reported and understood

When a variable is both a confounder and an effect modifier: 1. We must adjust for it to eliminate bias 2. We should present stratum-specific effects (not just an overall adjusted effect)

Common mistake: Treating an effect modifier as if it were only a confounder, adjusting for it and reporting only an overall effect, thereby obscuring important heterogeneity.

5 4.5 Matching as Another Form of Adjustment (pp. 43-45)

In addition to stratification, **matching** is another method for adjustment that can also reveal effect modification.

5.1 Matching Methods

Matching creates treatment and control groups that are similar with respect to measured covariates L. Common approaches:

1. **Individual matching**: For each treated individual, find one (or more) untreated individuals with similar values of L
2. **Caliper matching**: Match treated and untreated individuals whose L values are within some distance threshold
3. **Propensity score matching**: Match on the probability of treatment given L (covered in Chapter 15)

Example 5.1 (Matching Example)**.** Suppose we want to study the effect of smoking on lung cancer. We: 1. Identify 1000 smokers 2. For each smoker, find a non-smoker matched on age, sex, occupation, and family history 3. Compare lung cancer rates in smokers vs. their matched non-smokers
If matching successfully balances confounders, the comparison estimates the causal effect.

5.2 Matching and Effect Modification

Like stratification, matching can identify effect modification:

- Match separately within subgroups defined by V
- Estimate effects within each subgroup
- Compare effects across subgroups

Matching vs. Stratification:

Similarities:

- Both create comparable treatment and control groups
- Both can identify effect modification
- Both can eliminate confounding

Differences:

- **Stratification** uses all data, dividing into non-overlapping strata
- **Matching** may discard data (unmatched individuals), creating matched pairs/sets
- **Matching** can be more efficient when controls far outnumber treated individuals
- **Stratification** is conceptually simpler and more transparent

Modern approach: Matching has largely been superseded by more flexible methods like:

- Regression adjustment
- Inverse probability weighting (Chapter 12)
- Propensity score methods (Chapter 15)

However, matching is still useful pedagogically and in certain applications (e.g., case-control studies in epidemiology).

6 4.6 Effect Modification and Adjustment Methods (pp. 45-46)

When both confounding and effect modification are present, we need methods that can: 1. Adjust for confounders 2. Allow treatment effects to vary across modifiers

6.1 Stratification with Confounding

If V is both a confounder and effect modifier, simple stratification may not suffice if there are additional confounders L.

Solution: Adjust for L within each stratum of V. This can be done via:

- **Stratification on both V and L**: Creates many narrow strata
- **Regression within strata**: Fit separate regression models for each level of V
- **Inverse probability weighting within strata**: Weight by propensity score within each V stratum

> **Example 6.1** (Stratification with Additional Confounders)**.** To estimate the effect of exercise on heart disease, modified by age, while adjusting for sex and smoking:
> **Approach 1**: Stratify by age, sex, and smoking
> - Too many strata (e.g., 3 age groups $\times$ 2 sexes $\times$ 2 smoking status = 12 strata)
> - Sample sizes become small
>
> **Approach 2**: Stratify by age, then adjust for sex and smoking within each age stratum
> - Fit regression model within young, middle-aged, and elderly separately
> - Each model controls for sex and smoking
> - Compare effects across age groups

6.2 Regression Models for Effect Modification

Regression models provide a flexible framework for handling effect modification:

Model with effect modification:

$$\mathrm{E}[Y|A, V, L] = \beta_0 + \beta_1 A + \beta_2 V + \beta_3 A \times V + \beta_4^T L$$

The **interaction term** $\beta_3 A \times V$ captures effect modification:

- If $\beta_3 = 0$: No effect modification by V
- If $\beta_3 \neq 0$: Effect of A differs by V

Important distinctions:

Statistical interaction (in regression models):

- Refers to product terms like $A \times V$ in a model
- Is model-dependent and scale-dependent
- Does not have an inherent causal interpretation

Effect modification (in causal inference):

- Refers to heterogeneity of causal effects across subgroups
- Is defined in terms of potential outcomes
- Has a clear causal interpretation

These concepts are related but not identical:

- Effect modification on the risk difference scale corresponds to interaction in a linear model
- Effect modification on the risk ratio scale corresponds to interaction in a log-linear model
- Absence of statistical interaction does NOT guarantee absence of effect modification (it depends on the model and scale)

Best practice: Report stratum-specific effects directly rather than relying solely on interaction terms. This makes the findings more transparent and interpretable.

7 Summary

This chapter introduced **effect modification**: the phenomenon where causal effects differ across levels of another variable.

Key concepts:

1. **Definition**: Effect modification exists when $\mathrm{E}\big[Y^{a=1} - Y^{a=0}|V\big]$ differs across levels of V

2. **Scale dependence**: Whether effect modification exists depends on the scale (risk difference, risk ratio, etc.)

3. **Identification**: Stratification and matching can identify effect modification by estimating effects separately in subgroups

4. **Importance**: Effect modification is crucial for:

   - Targeting interventions
   - Improving precision
   - Understanding mechanisms

5. **Confounding vs. modification**:

   - Confounders bias marginal effects (we adjust for them)
   - Modifiers create heterogeneous effects (we report them)
   - Variables can be both

6. **Regression models**: Interaction terms in regression models can capture effect modification, but must be interpreted carefully with respect to scale

Looking ahead:

- **Chapter 5** introduces **interaction** between two treatments (different from effect modification by baseline variables)
- **Chapter 6** introduces **causal diagrams**, which help identify confounders and understand when stratification is appropriate
- **Chapter 13** covers the **parametric g-formula**, a flexible method for estimating effects while accounting for both confounding and effect modification
- **Chapter 15** discusses **propensity scores**, which can be combined with stratification to handle confounding while preserving effect modification

Practical takeaway: When analyzing observational data: 1. Think carefully about which variables might modify effects (use subject-matter knowledge) 2. Stratify or use regression with interactions to assess effect modification 3. Report stratum-specific effects when they differ meaningfully 4. Don't hide important heterogeneity behind an average effect

8 References

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