

# Chapter 2: Randomized Experiments

## Contents

<b>1</b>	<b>2.1 Randomization (pp. 13-16)</b>	<b>1</b>
1.1	Exchangeability . . . . .	2
1.2	Why Randomization Achieves Exchangeability . . . . .	2
1.3	Association Equals Causation Under Exchangeability . . . . .	3
<b>2</b>	<b>2.2 Conditional Randomization (pp. 16-18)</b>	<b>3</b>
2.1	Stratified Randomization . . . . .	3
2.2	Conditional Exchangeability . . . . .	3
2.3	Why Use Conditional Randomization? . . . . .	4
<b>3</b>	<b>2.3 Standardization (pp. 18-21)</b>	<b>4</b>
3.1	The Standardization Formula . . . . .	4
3.2	Worked Example: Standardization in Zeus’s Family . . . . .	5
3.3	Interpretation . . . . .	5
<b>4</b>	<b>2.4 Inverse Probability Weighting (pp. 21-24)</b>	<b>6</b>
4.1	The IPW Approach . . . . .	6
4.2	IPW Estimation . . . . .	6
4.3	Worked Example: IPW in Zeus’s Family . . . . .	7
4.4	Comparison: Standardization vs. IPW . . . . .	7
4.5	Positivity . . . . .	8
<b>5</b>	<b>Summary</b>	<b>8</b>
<b>6</b>	<b>Fine Point</b>	<b>9</b>
6.1	2.1 The intention-to-treat effect . . . . .	9
6.2	2.2 The null hypothesis . . . . .	9
<b>7</b>	<b>Technical Point</b>	<b>9</b>
7.1	2.1 Identification and exchangeability . . . . .	9
7.2	2.2 Derivation of the IPW estimator . . . . .	9
<b>8</b>	<b>References</b>	<b>10</b>

In Chapter 1, we concluded that association is not causation, and we asked under which conditions we could use data to estimate causal effects. This chapter provides one answer: conduct a randomized experiment. We use randomized experiments to illustrate fundamental concepts of causal inference, including exchangeability and the distinction between conditional and unconditional effects.

This chapter is based on Hernán and Robins (2020, chap. 2, pp. 13-24).

## 1 2.1 Randomization (pp. 13-16)

---

Suppose we conduct a randomized experiment using Zeus’s family as our study population. To determine who receives a heart transplant, Zeus flips a (fair) coin for each of the 20 individuals. If

heads, the individual receives a heart transplant ( $A = 1$ ); if tails, they do not ( $A = 0$ ).

Table 2.1 shows the observed data from this randomized experiment.

From Table 2.1, we can compute associational measures:

- Risk in the treated:  $Pr[Y = 1|A = 1] = 7/13 \approx 0.54$
- Risk in the untreated:  $Pr[Y = 1|A = 0] = 3/7 \approx 0.43$

Unlike Table 1.2 in Chapter 1 (where treatment was not randomized), Table 2.1 shows data from a randomized experiment. The treatment assignment is now determined by chance rather than by pre-existing characteristics.

## 1.1 Exchangeability

The key feature of a randomized experiment is that treatment assignment is independent of an individual's baseline characteristics, including their potential outcomes. We say the treated and untreated are **exchangeable** when:

$$Y^a \perp\!\!\!\perp A \quad \text{for all } a$$

This notation means the counterfactual outcome  $Y^a$  is independent of the treatment actually received  $A$ . Exchangeability implies:

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

for all values of  $a$ .

**Definition 1.1** (Exchangeability). Treatment groups are **exchangeable** if the distribution of counterfactual outcomes is the same in both groups:

$$Y^a \perp\!\!\!\perp A \quad \text{for all } a$$

Under exchangeability, individuals in the treated group would have experienced the same distribution of outcomes if they had remained untreated (and vice versa).

Exchangeability is also called **no confounding**. When exchangeability holds, association equals causation. That is:

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

The left side is the associational risk difference (observed), and the right side is the causal risk difference (counterfactual). Under exchangeability, we can interpret the observed association as a causal effect.

## 1.2 Why Randomization Achieves Exchangeability

In our randomized experiment, each individual had a 0.5 probability of receiving treatment, determined by a fair coin flip. This random assignment ensures that treatment is independent of all variables—whether measured or unmeasured, known or unknown.

Specifically, randomization ensures:

$$Pr[A = 1|Y^{a=1} = 1, Y^{a=0} = 1] = Pr[A = 1|Y^{a=1} = 0, Y^{a=0} = 0] = Pr[A = 1] = 0.5$$

The probability of treatment is the same regardless of an individual's potential outcomes. Therefore, the treated and untreated are exchangeable by design.

Randomization is often described as “eliminating confounding” or “balancing baseline covariates.” These are informal ways of saying randomization achieves exchangeability.

In practice, we cannot verify exchangeability directly because we don't observe all potential outcomes. However, we can check whether measured baseline covariates are balanced between treatment groups. Imbalance suggests (but doesn't prove) problems with randomization.

### 1.3 Association Equals Causation Under Exchangeability

When exchangeability holds, we can identify the average causal effect from observed data:

$$E[Y^{a=1}] - E[Y^{a=0}] = E[Y|A = 1] - E[Y|A = 0]$$

The right side is the observed associational difference, which we can compute from Table 2.1. Under the exchangeability guaranteed by randomization, this associational difference equals the causal effect.

In our example:

- Observed associational risk difference:  $0.54 - 0.43 = 0.11$
- This estimates the true causal risk difference (from Table 1.1):  $0.5 - 0.5 = 0$

The discrepancy is due to random variability from the coin flips, not systematic bias.

In a randomized experiment, the observed association is an **unbiased estimate** of the causal effect. With only 20 individuals, random variability can be substantial. As sample size increases, the estimate converges to the true causal effect (law of large numbers).

Chapter 10 discusses statistical inference (confidence intervals, p-values) to quantify this random variability.

## 2 2.2 Conditional Randomization (pp. 16-18)

---

Not all randomized experiments use simple randomization with equal treatment probabilities. Often, randomization is **conditional** on measured covariates.

### 2.1 Stratified Randomization

Consider a modified version of our experiment. Suppose Zeus decides to ensure exactly half the women and half the men receive transplants. He creates two strata: women and men. Within each stratum, he randomly assigns half to treatment.

Table 2.2 shows a possible outcome from this stratified randomization.

In Table 2.2,  $L$  denotes sex (0 for female, 1 for male). Exactly 5 of the 10 women and 5 of the 10 men received transplants.

### 2.2 Conditional Exchangeability

In a conditionally randomized experiment, exchangeability holds **within each stratum** but not necessarily overall. We have:

$$Y^a \perp\!\!\!\perp A|L \quad \text{for all } a$$

This is **conditional exchangeability** (or **conditional randomization**).

**Definition 2.1** (Conditional Exchangeability). Treatment groups are **conditionally exchangeable** given covariates  $L$  if:

$$Y^a \perp\!\!\!\perp A|L \quad \text{for all } a$$

Within each stratum of  $L$ , the treated and untreated are exchangeable.

Conditional exchangeability means:

$$Pr[Y^a = 1|A = 1, L = l] = Pr[Y^a = 1|A = 0, L = l] = Pr[Y^a = 1|L = l]$$

for all values of  $a$  and  $l$ .

Conditional randomization guarantees conditional exchangeability but **not marginal exchangeability** (unconditional exchangeability). That is,  $Y^a \perp\!\!\!\perp A|L$  does not imply  $Y^a \perp\!\!\!\perp A$ .

In Table 2.2, if women have higher baseline risk than men, and we ensure equal numbers of women in each treatment group, then the treated and untreated groups are balanced with respect to sex. However, the marginal (overall) association may not equal the causal effect if sex is related to the outcome.

### 2.3 Why Use Conditional Randomization?

Conditional randomization serves several purposes:

1. **Ensures balance:** Guarantees equal distribution of important prognostic factors across treatment groups
2. **Improves precision:** Reduces variability in estimates by controlling for predictors of the outcome
3. **Enables subgroup analysis:** Allows examination of effect modification (heterogeneous effects)

Common forms include:

- **Stratified randomization:** Randomize within strata defined by covariates
- **Block randomization:** Randomize in blocks to ensure balanced sample sizes
- **Matched-pair randomization:** Match individuals on covariates, then randomize within pairs

## 3 2.3 Standardization (pp. 18-21)

---

When exchangeability is conditional rather than marginal, we cannot simply compare outcomes between treatment groups. We must account for the stratification variable  $L$ .

### 3.1 The Standardization Formula

Under conditional exchangeability  $Y^a \perp\!\!\!\perp A|L$ , the average causal effect can be computed using **standardization**:

$$E[Y^a] = \sum_l E[Y|A = a, L = l] \times Pr[L = l]$$

This formula weights the stratum-specific mean outcomes by the population distribution of  $L$ .

**Definition 3.1** (Standardization Formula). Under conditional exchangeability  $Y^a \perp\!\!\!\perp A|L$ , the mean counterfactual outcome is:

$$E[Y^a] = \sum_l E[Y|A = a, L = l] \times Pr[L = l]$$

where the sum is over all values  $l$  of the covariate  $L$ .

This formula is sometimes called the **g-formula** (generalized formula). It expresses the counterfactual mean as a weighted average of conditional means.

The standardization formula requires: 1. **Conditional exchangeability:**  $Y^a \perp\!\!\!\perp A|L$  2. **Positivity:**  $Pr[A = a|L = l] > 0$  for all values  $l$  with  $Pr[L = l] > 0$  3. **Consistency:**  $Y = Y^A$  (the observed outcome equals the counterfactual under treatment received)

### 3.2 Worked Example: Standardization in Zeus’s Family

Using Table 2.2, let’s compute the standardized causal risk difference.

**Step 1:** Compute stratum-specific risks

*Among women ( $L = 0$ ):*

- $Pr[Y = 1|A = 1, L = 0] = 3/5 = 0.6$  (3 of 5 treated women died)
- $Pr[Y = 1|A = 0, L = 0] = 2/5 = 0.4$  (2 of 5 untreated women died)

*Among men ( $L = 1$ ):*

- $Pr[Y = 1|A = 1, L = 1] = 4/5 = 0.8$  (4 of 5 treated men died)
- $Pr[Y = 1|A = 0, L = 1] = 3/5 = 0.6$  (3 of 5 untreated men died)

**Step 2:** Compute population distribution of  $L$

- $Pr[L = 0] = 10/20 = 0.5$  (10 women)
- $Pr[L = 1] = 10/20 = 0.5$  (10 men)

**Step 3:** Apply standardization formula

$$E[Y^{a=1}] = 0.6 \times 0.5 + 0.8 \times 0.5 = 0.7 \tag{1}$$

$$E[Y^{a=0}] = 0.4 \times 0.5 + 0.6 \times 0.5 = 0.5 \tag{2}$$

**Step 4:** Compute causal effect

$$E[Y^{a=1}] - E[Y^{a=0}] = 0.7 - 0.5 = 0.2$$

The standardized causal risk difference is 0.2. In this particular example, this equals the crude (unstandardized) associational difference:  $(7/10) - (5/10) = 0.2$ .

In general, the crude association  $Pr[Y = 1|A = 1] - Pr[Y = 1|A = 0]$  does **not** equal the standardized causal effect when conditional randomization is used.

The standardization formula “adjusts for” or “controls for”  $L$  by reweighting each stratum to match the population distribution. This is analogous to direct standardization in epidemiology.

The standardization formula can be used with continuous outcomes by replacing probabilities with expected values, and can extend to multiple covariates and continuous covariates (by replacing the sum with an integral).

### 3.3 Interpretation

The standardized mean  $E[Y^{a=1}]$  answers: “What would the mean outcome be if everyone in the population received treatment  $a = 1$ ?” It’s a weighted average of the stratum-specific mean outcomes, where weights are the proportion of the population in each stratum.

Similarly,  $E[Y^{a=0}]$  answers: “What would the mean outcome be if everyone in the population received treatment  $a = 0$ ?”

The difference  $E[Y^{a=1}] - E[Y^{a=0}]$  is the **average causal effect** in the population.

## 4 2.4 Inverse Probability Weighting (pp. 21-24)

---

Standardization is not the only method to adjust for conditional randomization. An alternative approach is **inverse probability weighting** (IPW).

### 4.1 The IPW Approach

The idea of IPW is to create a **pseudo-population** in which treatment is marginally randomized (unconditional). In this pseudo-population, exchangeability holds marginally, so we can estimate the causal effect by a simple comparison of means.

Each individual is weighted by the inverse of their probability of receiving the treatment they actually received:

$$W^A = \frac{1}{Pr[A|L]}$$

**Definition 4.1** (Inverse Probability Weights). For an individual who received treatment  $A$  and has covariates  $L$ , the **inverse probability weight** is:

$$W^A = \frac{1}{Pr[A|L]}$$

where  $Pr[A|L]$  is the probability (propensity) of receiving the treatment actually received, given covariates  $L$ .

Individuals whose treatment assignment is unlikely given their covariates receive higher weights. Individuals whose treatment assignment is likely receive lower weights. This reweighting creates balance in the pseudo-population.

The weights are also called **stabilized weights** when modified to improve efficiency. The basic IPW weights above are sometimes called **unstabilized weights**.

### 4.2 IPW Estimation

Under conditional exchangeability  $Y^a \perp\!\!\!\perp A|L$ , the mean counterfactual outcome can be estimated as:

$$E[Y^a] = E\left[\frac{I(A = a) \times Y}{Pr[A = a|L]}\right]$$

where  $I(A = a)$  is an indicator function equal to 1 if  $A = a$  and 0 otherwise.

The IPW estimator of the causal risk difference is:

$$E[Y^{a=1}] - E[Y^{a=0}] = \frac{1}{n} \sum_{i=1}^n \frac{I(A_i = 1) \times Y_i}{Pr[A_i = 1|L_i]} - \frac{1}{n} \sum_{i=1}^n \frac{I(A_i = 0) \times Y_i}{Pr[A_i = 0|L_i]}$$

This is equivalent to computing a weighted mean in each treatment group:

- Treated:  $\frac{\sum_{i:A_i=1} W_i Y_i}{\sum_{i:A_i=1} W_i}$  where  $W_i = 1/Pr[A_i = 1|L_i]$
- Untreated:  $\frac{\sum_{i:A_i=0} W_i Y_i}{\sum_{i:A_i=0} W_i}$  where  $W_i = 1/Pr[A_i = 0|L_i]$

Then subtract the two weighted means.

### 4.3 Worked Example: IPW in Zeus's Family

Using Table 2.2, let's compute the IPW estimate of the causal effect.

**Step 1:** Compute treatment probabilities

In the stratified randomization:

- $Pr[A = 1|L = 0] = 5/10 = 0.5$  (among women)
- $Pr[A = 1|L = 1] = 5/10 = 0.5$  (among men)
- $Pr[A = 0|L = 0] = 5/10 = 0.5$  (among women)
- $Pr[A = 0|L = 1] = 5/10 = 0.5$  (among men)

**Step 2:** Compute weights

Since all probabilities equal 0.5, all weights equal  $1/0.5 = 2$ .

**Step 3:** Compute weighted means

*Treated group (A = 1):*

Based on Table 2.2, there are 10 treated individuals, of whom 7 died (Y=1). The weighted mean is:

$$\frac{\sum_{i:A_i=1} W_i Y_i}{\sum_{i:A_i=1} W_i} = \frac{7 \times 2}{10 \times 2} = \frac{14}{20} = 0.7$$

*Untreated group (A = 0):*

Based on Table 2.2, there are 10 untreated individuals, of whom 5 died (Y=1). The weighted mean is:

$$\frac{\sum_{i:A_i=0} W_i Y_i}{\sum_{i:A_i=0} W_i} = \frac{5 \times 2}{10 \times 2} = \frac{10}{20} = 0.5$$

**Step 4:** Compute IPW estimate

- Weighted mean in treated: 0.7
- Weighted mean in untreated: 0.5
- IPW estimate:  $0.7 - 0.5 = 0.2$

In this example, IPW and standardization give the same answer (0.2) because the design is balanced. With equal treatment probabilities within strata and equal numbers in each stratum, the two methods are equivalent.

In general, standardization and IPW can give different estimates when: 1. Treatment probabilities vary across strata 2. Stratum sizes are unequal 3. There is model misspecification (in observational studies)

Both methods are consistent (converge to the true causal effect) under the same assumptions: conditional exchangeability, positivity, and consistency.

### 4.4 Comparison: Standardization vs. IPW

Feature	Standardization	IPW
<b>Idea</b>	Weight stratum-specific outcomes by population distribution	Create pseudo-population with marginal randomization
<b>Formula</b>	$E[Y^a] = \sum_l E[Y A = a, L = l]Pr[L = l]$	$E[Y^a] = E[Y \cdot I(A = a)/Pr[A = a L]]$

Feature	Standardization	IPW
<b>Weights</b>	Stratum probabilities $Pr[L = l]$	Inverse treatment probabilities $1/Pr[A L]$
<b>Model</b>	Outcome model $E[Y A, L]$	Treatment model $Pr[A L]$
<b>Extensions</b>	G-computation, parametric g-formula	Marginal structural models

In randomized experiments, treatment probabilities  $Pr[A|L]$  are known by design. In observational studies, we must estimate them from data, which introduces additional assumptions and potential for bias.

Standardization requires correctly specifying the outcome model  $E[Y|A, L]$ . IPW requires correctly specifying the treatment model  $Pr[A|L]$ . **Doubly robust** methods combine both and remain consistent if *either* model is correct.

## 4.5 Positivity

Both standardization and IPW require the **positivity assumption**: For all values of  $L$  that occur in the population, there must be a non-zero probability of receiving each treatment level.

Formally: If  $Pr[L = l] > 0$ , then  $0 < Pr[A = a|L = l] < 1$  for all  $a$ .

**Definition 4.2** (Positivity). The **positivity assumption** (also called **experimental treatment assignment** or **overlap**) requires:

$$0 < Pr[A = a|L = l] < 1$$

for all values  $l$  such that  $Pr[L = l] > 0$  and for all treatment levels  $a$ .

Positivity means there are both treated and untreated individuals at every level of  $L$  that occurs in the population. Without positivity, we cannot estimate  $E[Y|A = a, L = l]$  for some strata because we have no data.

Randomized experiments guarantee positivity if the randomization probabilities are strictly between 0 and 1. Observational studies often violate positivity, leading to non-identifiability and extreme IPW weights.

## 5 Summary

This chapter introduced fundamental concepts using randomized experiments:

1. **Randomization** ensures exchangeability  $Y^a \perp\!\!\!\perp A$ , allowing causal effects to be identified from associational contrasts
2. **Conditional randomization** achieves conditional exchangeability  $Y^a \perp\!\!\!\perp A|L$  within strata, requiring adjustment methods
3. **Standardization** computes  $E[Y^a] = \sum_l E[Y|A = a, L = l]Pr[L = l]$  by weighting stratum-specific outcomes
4. **Inverse probability weighting** creates a pseudo-population where treatment is marginally randomized by weighting individuals as  $W^A = 1/Pr[A|L]$
5. Both standardization and IPW require **conditional exchangeability**, **positivity**, and **consistency**

In randomized experiments, these assumptions are met by design. In observational studies (Chapters 3-7), these assumptions are less plausible and require careful justification.

## 6 Fine Point

---

### 6.1 2.1 The intention-to-treat effect

In randomized experiments, individuals may not receive or adhere to their assigned treatment. For example, someone randomized to receive a transplant may refuse it. The **intention-to-treat (ITT) effect** is the causal effect of treatment *assignment* (not treatment received).

Let  $Z$  denote random assignment and  $A$  denote treatment actually received. The ITT effect is  $E[Y^{z=1}] - E[Y^{z=0}]$ , while the **per-protocol effect** is  $E[Y^{a=1}] - E[Y^{a=0}]$ .

ITT effects are always identified in randomized experiments (because  $Y^z \perp\!\!\!\perp Z$  by randomization). Per-protocol effects require additional assumptions about non-adherence.

### 6.2 2.2 The null hypothesis

Chapter 1 distinguished the **average causal null hypothesis** ( $E[Y^{a=1}] = E[Y^{a=0}]$ ) from the **sharp causal null hypothesis** ( $Y^{a=1} = Y^{a=0}$  for all individuals).

In randomized experiments, the sharp null allows for **exact inference** using permutation tests (Fisher's randomization test). These tests are valid in finite samples without distributional assumptions.

Under the average null (but not the sharp null), standard large-sample inference (normal approximation, t-tests) is typically used.

## 7 Technical Point

---

### 7.1 2.1 Identification and exchangeability

A causal quantity is **identified** if it can be expressed as a function of the observed data distribution. Exchangeability is a sufficient condition for identifying average causal effects.

Under  $Y^a \perp\!\!\!\perp A$ , we have:

$$E[Y^a] = E[Y|A = a]$$

This is because:

$$E[Y^a] = E[Y^a|A = a] \quad (\text{by exchangeability}) = E[Y|A = a] \quad (\text{by consistency})$$

Similarly, under  $Y^a \perp\!\!\!\perp A|L$ :

$$E[Y^a|L = l] = E[Y|A = a, L = l]$$

The standardization formula follows by taking expectations over  $L$ .

### 7.2 2.2 Derivation of the IPW estimator

The IPW estimator can be derived from first principles. Under  $Y^a \perp\!\!\!\perp A|L$  and positivity:

$$E[Y^a] = E[E[Y^a|L]] \quad (\text{law of total expectation}) \quad (3)$$

$$= E[E[Y^a|A = a, L]] \quad (\text{by conditional exchangeability}) \quad (4)$$

$$= E[E[Y|A = a, L]] \quad (\text{by consistency}) \quad (5)$$

$$= E \left[ \frac{I(A = a)}{\text{Pr}[A = a|L]} E[Y|A = a, L] \right] \quad (\text{inverse weighting}) \quad (6)$$

$$= E \left[ \frac{I(A = a) \cdot Y}{\text{Pr}[A = a|L]} \right] \quad (\text{by law of total expectation}) \quad (7)$$

The key step is the third equality, which uses the fact that  $E[I(A = a)/\text{Pr}[A = a|L]] = 1$ .

## 8 References

---

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