

# Chapter 13: Standardization and the Parametric G-Formula

## Contents

<b>1</b>	<b>13.1 Standardization as an Alternative to IP Weighting (pp. 175-177)</b>	<b>2</b>
1.1	Standardization Review . . . . .	2
1.2	Example: Discrete Confounders . . . . .	2
<b>2</b>	<b>13.2 Estimating the Mean Outcome via Modeling (pp. 177-179)</b>	<b>2</b>
2.1	Parametric Outcome Model . . . . .	3
2.2	The Parametric G-Formula . . . . .	3
2.3	Example: NHEFS Data . . . . .	3
<b>3</b>	<b>13.3 Standardizing the Mean Outcome to the Confounder Distribution (pp. 179-181)</b>	<b>3</b>
3.1	Alternative Reference Distributions . . . . .	4
3.2	ATT vs ATE . . . . .	4
<b>4</b>	<b>13.4 IP Weighting or Standardization? (pp. 181-183)</b>	<b>4</b>
4.1	Comparison . . . . .	4
4.2	Which to Choose? . . . . .	4
<b>5</b>	<b>13.5 How Seriously Do We Take Our Models? (pp. 183-185)</b>	<b>5</b>
5.1	Model Misspecification . . . . .	5
5.2	Strategies for Model Selection . . . . .	5
5.3	Example: Polynomial Models in NHEFS . . . . .	5
<b>6</b>	<b>13.6 G-Formula for Continuous Treatments (pp. 185-186)</b>	<b>6</b>
6.1	Continuous Treatment . . . . .	6
6.2	Dose-Response Curve . . . . .	6
<b>7</b>	<b>13.7 Standardization or IP Weighting for Dichotomous Outcomes (pp. 186-188)</b>	<b>6</b>
7.1	Standardization for Binary Outcomes . . . . .	6
7.2	IP Weighting for Binary Outcomes . . . . .	7
<b>8</b>	<b>Summary</b>	<b>7</b>

This chapter describes **standardization** and the **parametric g-formula**, methods for computing standardized means and risks by outcome modeling. While IP weighting models the treatment assignment mechanism, standardization models the outcome mechanism. Both approaches can estimate the same causal effects under conditional exchangeability.

This chapter is based on Hernán and Robins (2020, chap. 13, pp. 175-188).

**Key insight:** Standardization and IP weighting are “doubly robust” in the sense that we need to correctly specify either the treatment model or the outcome model (but not necessarily both) for some estimators. However, standard implementations require correct specification of one or the other.

# 1 13.1 Standardization as an Alternative to IP Weighting (pp. 175-177)

---

We've seen two approaches to confounding adjustment:

1. **IP weighting** (Chapter 12): Model  $\Pr[A | L]$  and weight observations
2. **Standardization** (this chapter): Model  $E[Y | A, L]$  and compute weighted averages

Both can estimate  $E[Y^a]$  under conditional exchangeability.

## 1.1 Standardization Review

From Chapter 2, **standardization** computes:

$$E[Y^a] = \sum_{\ell} E[Y | A = a, L = \ell] \Pr[L = \ell]$$

This is a weighted average of stratum-specific means, with weights equal to the population distribution of  $L$ .

**Definition 1.1** (Standardization). **Standardization** estimates the mean outcome under treatment  $a$  by:

1. Computing  $E[Y | A = a, L = \ell]$  for all levels  $\ell$
2. Averaging over the population distribution of  $L$ :

$$\hat{E}[Y^a] = \sum_{\ell} \hat{E}[Y | A = a, L = \ell] \times \hat{\Pr}[L = \ell]$$

where  $\hat{\Pr}[L = \ell]$  is the observed proportion with  $L = \ell$ .

## 1.2 Example: Discrete Confounders

**Setting:** Binary  $A$ , binary  $Y$ , discrete  $L$  with  $k$  levels

**Step 1:** Compute proportion  $Y = 1$  within each stratum ( $A = a, L = \ell$ )

**Step 2:** Standardize to population distribution:

$$\hat{E}[Y^{a=1}] = \sum_{\ell=1}^k \hat{\Pr}[Y = 1 | A = 1, L = \ell] \times \hat{\Pr}[L = \ell]$$

$$\hat{E}[Y^{a=0}] = \sum_{\ell=1}^k \hat{\Pr}[Y = 1 | A = 0, L = \ell] \times \hat{\Pr}[L = \ell]$$

**Causal effect:**  $\hat{E}[Y^{a=1}] - \hat{E}[Y^{a=0}]$

**Comparison to Chapter 2:** In Part I (Chapter 2), we used **nonparametric standardization** where we directly computed sample proportions within each stratum. This chapter introduces **parametric standardization** using regression models, which is necessary when  $L$  is high-dimensional or continuous.

# 2 13.2 Estimating the Mean Outcome via Modeling (pp. 177-179)

---

When confounders are continuous or high-dimensional, we cannot compute stratum-specific means directly. Instead, we use **parametric models**.

## 2.1 Parametric Outcome Model

**Model:** Specify a model for  $E[Y | A, L]$ , such as:

$$E[Y | A, L] = \beta_0 + \beta_1 A + \beta_2^T L + \beta_3^T (A \times L)$$

This includes: - Main effects of  $A$  and  $L$  - Interactions between  $A$  and  $L$  to allow effect modification

**Estimation:** Fit the model using standard regression (e.g., linear regression for continuous  $Y$ , logistic regression for binary  $Y$ ).

## 2.2 The Parametric G-Formula

**Definition 2.1** (Parametric G-Formula). Given a model  $\hat{E}[Y | A, L]$ , the **parametric g-formula** estimates:

$$\hat{E}[Y^a] = \frac{1}{n} \sum_{i=1}^n \hat{E}[Y | A = a, L = L_i]$$

**Algorithm:**

1. Fit outcome model  $\hat{E}[Y | A, L]$  using all data
2. For each individual  $i$ , predict  $\hat{Y}_i^a = \hat{E}[Y | A = a, L = L_i]$
3. Average the predictions:  $\hat{E}[Y^a] = n^{-1} \sum_i \hat{Y}_i^a$
4. Repeat for each treatment level  $a$

## 2.3 Example: NHEFS Data

**Outcome model:** Linear regression for weight change

$$E[Y | A, L] = \beta_0 + \beta_1 A + \sum_j \beta_j L_j + \sum_j \gamma_j (A \times L_j)$$

**Procedure:**

1. Fit model using observed  $(A, L, Y)$
2. Predict  $\hat{Y}_i^{a=1}$  for all  $i$  by setting  $A = 1$ , keeping  $L_i$  as observed
3. Predict  $\hat{Y}_i^{a=0}$  for all  $i$  by setting  $A = 0$ , keeping  $L_i$  as observed
4. Average:  $\hat{E}[Y^{a=1}] = \bar{Y}^{\bar{a}=1}$ ,  $\hat{E}[Y^{a=0}] = \bar{Y}^{\bar{a}=0}$
5. Estimate causal effect:  $\hat{E}[Y^{a=1}] - \hat{E}[Y^{a=0}]$

**Why this works:** Under conditional exchangeability  $Y^a \perp\!\!\!\perp A | L$ :

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

The g-formula estimates this by averaging the conditional mean  $E[Y | A = a, L]$  over the empirical distribution of  $L$ .

## 3 13.3 Standardizing the Mean Outcome to the Confounder Distribution (pp. 179-181)

The g-formula standardizes to the **observed distribution** of confounders. We can also standardize to other distributions.

### 3.1 Alternative Reference Distributions

Options for standardization:

1. **Population distribution:**  $\sum_{\ell} E[Y | A = a, L = \ell] \Pr[L = \ell]$  (standard g-formula)
2. **Treated distribution:**  $\sum_{\ell} E[Y | A = a, L = \ell] \Pr[L = \ell | A = 1]$
3. **Untreated distribution:**  $\sum_{\ell} E[Y | A = a, L = \ell] \Pr[L = \ell | A = 0]$
4. **External distribution:**  $\sum_{\ell} E[Y | A = a, L = \ell] \Pr_{\text{ext}}[L = \ell]$

### 3.2 ATT vs ATE

Average treatment effect (ATE):

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

Standardized to the population (or sample) distribution of  $L$ .

Average treatment effect in the treated (ATT):

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

Standardized to the distribution of  $L$  among the treated.

G-formula for ATT:

$$\hat{E}[Y^a | A = 1] = \frac{1}{n_1} \sum_{i: A_i=1} \hat{E}[Y | A = a, L = L_i]$$

where  $n_1 = \sum_i I(A_i = 1)$ .

When to use ATT vs ATE:

- **ATE:** Answers “what if we intervened on the whole population?”
- **ATT:** Answers “what if we intervened on those who were actually treated?”

ATT is useful when: - Treatment is not feasible for some individuals - Policy question focuses on those currently receiving treatment - Positivity violations make ATE estimation unstable

## 4 13.4 IP Weighting or Standardization? (pp. 181-183)

Both IP weighting and standardization can estimate causal effects. How do they compare?

### 4.1 Comparison

Aspect	IP Weighting	Standardization
<b>Models</b>	$\Pr[A   L]$ (treatment)	$E[Y   A, L]$ (outcome)
<b>Target</b>	Marginal effect	Marginal effect (via averaging)
<b>Natural for</b>	Marginal structural models	Conditional models
<b>Handles</b>	Time-varying treatment easily	Time-varying treatment (complex)
<b>Efficiency</b>	Less efficient (if outcome model correct)	More efficient (if outcome model correct)
<b>Robustness</b>	Robust to outcome model misspec.	Robust to treatment model misspec.

### 4.2 Which to Choose?

Use **IP weighting when:** - Treatment mechanism is simple to model - Outcome is complex or multiply measured - Time-varying treatments - Interested in marginal effects explicitly

**Use standardization when:** - Outcome mechanism is simple to model - Treatment assignment is complex - Efficiency is important - Natural to think about outcome modeling

**Use both:** - Doubly robust estimation combines both approaches - Agreement between methods is reassuring - Disagreement suggests model misspecification

**Practical consideration:** Many researchers fit both methods as a sensitivity analysis. If results differ substantially, it suggests model misspecification in one or both approaches. This motivates the development of doubly robust methods that require only one model to be correct.

## 5 13.5 How Seriously Do We Take Our Models? (pp. 183-185)

---

Parametric models are **approximations**. How much does misspecification matter?

### 5.1 Model Misspecification

**Reality:** No model is exactly correct

- Linear models may be wrong for nonlinear relationships
- We may omit important interactions
- Functional form assumptions may be incorrect

**Consequences:**

1. **Bias:** Misspecified models give biased effect estimates
2. **Efficiency loss:** Correct models are most efficient
3. **Extrapolation problems:** Predictions far from data may be poor

### 5.2 Strategies for Model Selection

**Include product terms** (interactions): - Between treatment and confounders:  $A \times L$  - Allows effect modification - Helps model fit in treated and untreated separately

**Add polynomial terms:** - Quadratic:  $L + L^2$  - Cubic:  $L + L^2 + L^3$  - Flexible fit for continuous  $L$

**Use flexible methods:** - Splines - Generalized additive models - Machine learning methods (with care)

**Model checking:** - Residual plots - Goodness-of-fit tests - Cross-validation - Subject-matter knowledge

**Trade-offs:**

- **Complex models:** More flexible, less bias from misspecification, but more variance, potential overfitting
- **Simple models:** Less flexible, potential bias from misspecification, but less variance, easier to interpret

The **bias-variance trade-off** from Chapter 11 applies here. With large samples, lean toward flexibility. With small samples, lean toward parsimony.

### 5.3 Example: Polynomial Models in NHEFS

**Simple model:**

$$E[Y | A, L] = \beta_0 + \beta_1 A + \beta_2 \text{Age} + \beta_3 \text{Sex} + \dots$$

**Flexible model:**

$$E[Y | A, L] = \beta_0 + \beta_1 A + \beta_2 \text{Age} + \beta_3 \text{Age}^2 + \beta_4 A \times \text{Age} + \dots$$

Including  $A \times L$  interactions is especially important, as it allows the confounder-outcome relationship to differ between treated and untreated.

## 6 13.6 G-Formula for Continuous Treatments (pp. 185-186)

---

The g-formula extends naturally to **continuous treatments**.

### 6.1 Continuous Treatment

**Setting:** Treatment  $A$  is continuous (e.g., dose, duration, intensity)

**G-formula:**

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

Same as before, but now  $a$  can be any value in the continuous range.

**Estimation:**

1. Fit outcome model  $\hat{E}[Y | A, L]$  (e.g., linear regression)
2. For chosen dose  $a$ , predict  $\hat{Y}_i^a = \hat{E}[Y | A = a, L = L_i]$  for all  $i$
3. Average:  $\hat{E}[Y^a] = n^{-1} \sum_i \hat{Y}_i^a$
4. Repeat for different doses to trace out **dose-response curve**

### 6.2 Dose-Response Curve

**Definition 6.1** (Dose-Response Curve). The **dose-response curve** is the function  $a \mapsto E[Y^a]$  showing how the mean potential outcome varies with treatment level  $a$ . For continuous treatment, this is a smooth curve rather than discrete points.

**Example:** Effect of smoking intensity (cigarettes/day) on lung function

- Estimate  $\hat{E}[Y^a]$  for  $a = 0, 5, 10, 15, 20, \dots$  cigarettes/day
- Plot  $a$  vs  $\hat{E}[Y^a]$  to visualize dose-response

**Modeling considerations** for continuous  $A$ :

1. Include  $A$  and powers of  $A$  (e.g.,  $A, A^2, A^3$ ) for flexibility
2. Include interactions  $A \times L$  to allow effect modification
3. Use splines or generalized additive models for very flexible fits
4. Positivity: Need overlap in  $A$  distribution across  $L$  levels

## 7 13.7 Standardization or IP Weighting for Dichotomous Outcomes (pp. 186-188)

---

For binary outcomes, we can estimate **causal risk ratios** and **risk differences** using either approach.

### 7.1 Standardization for Binary Outcomes

**Outcome model:** Logistic regression (or log-binomial model)

$$\text{logit Pr}[Y = 1 | A, L] = \beta_0 + \beta_1 A + \beta_2^\top L + \beta_3^\top (A \times L)$$

**G-formula:**

$$\Pr[Y^a = 1] = \frac{1}{n} \sum_{i=1}^n \text{expit}(\hat{\beta}_0 + \hat{\beta}_1 a + \hat{\beta}_2^\top L_i + \hat{\beta}_3^\top (a \times L_i))$$

where  $\text{expit}(x) = \frac{e^x}{1+e^x}$ .

**Causal measures:**

- **Risk difference:**  $\Pr[\hat{Y}^{a=1} = 1] - \Pr[\hat{Y}^{a=0} = 1]$
- **Risk ratio:**  $\frac{\Pr[\hat{Y}^{a=1}=1]}{\Pr[\hat{Y}^{a=0}=1]}$
- **Odds ratio:**  $\frac{\Pr[\hat{Y}^{a=1}=1]/\Pr[\hat{Y}^{a=1}=0]}{\Pr[\hat{Y}^{a=0}=1]/\Pr[\hat{Y}^{a=0}=0]}$

## 7.2 IP Weighting for Binary Outcomes

**Marginal structural model:**

For risk difference:

$$\Pr[Y^a = 1] = \beta_0 + \beta_1 a$$

For risk ratio (log-binomial):

$$\log \Pr[Y^a = 1] = \beta_0 + \beta_1 a$$

For odds ratio (logistic):

$$\text{logit } \Pr[Y^a = 1] = \beta_0 + \beta_1 a$$

**Estimation:** Fit weighted model using IP weights

**Caution:** Logistic MSM models odds ratios, not risk ratios. For risk ratios, use log-binomial or Poisson models.

**Important distinction:**

- Conditional odds ratio from  $\text{logit } \Pr[Y = 1 | A, L] = \beta_0 + \beta_1 A + \dots$  is **NOT** generally a causal odds ratio (it's conditional on  $L$ )
- Marginal odds ratio from IP weighted logistic MSM **IS** a causal odds ratio (marginal over  $L$ )
- Standardization can compute any causal measure (risk difference, risk ratio, odds ratio)
- IP weighting model choice determines which causal measure is estimated

## 8 Summary

---

**Key concepts introduced:**

1. **Parametric standardization:** Use outcome regression models to compute standardized means
2. **Parametric g-formula:** Average predicted outcomes over covariate distribution
3. **Alternative standardization:** Can standardize to different reference populations (ATE, ATT, etc.)
4. **IP weighting vs standardization:** Two sides of the same coin, with different modeling and efficiency properties
5. **Model misspecification:** Always a concern; use flexible models and model checking
6. **Continuous treatments:** G-formula estimates dose-response curves
7. **Binary outcomes:** Can estimate risk differences, risk ratios, and odds ratios

**Relationship to IP weighting:**

- IP weighting models treatment, standardization models outcome
- Both estimate the same causal parameters under conditional exchangeability
- Neither is uniformly better; choice depends on context
- Doubly robust methods combine both approaches

**Practical advice:**

- Include product terms  $A \times L$  in outcome models
- Use flexible models (polynomials, splines) when sample size permits
- Check model fit with residual analysis and goodness-of-fit tests
- Consider both approaches as sensitivity analysis
- For binary outcomes, be clear about which causal measure you're estimating

**Looking ahead:**

- Chapter 14 introduces G-estimation for structural nested models, another approach that models neither the outcome nor the treatment directly
- Part III will show how the g-formula extends to time-varying treatments, where it becomes the “generalized” g-formula
- Doubly robust methods will combine IP weighting and outcome modeling for improved robustness

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