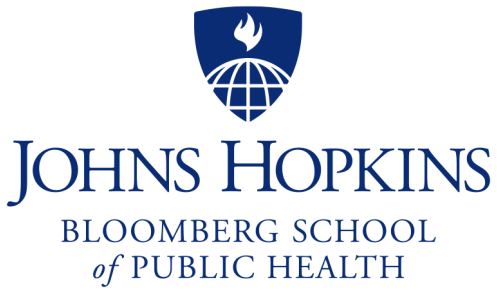
# Introduction to Causal Inference from an Observational Study for a Single Time Point Intervention

Michael Rosenblum

Department of Biostatistics

February 27, 2017



# Outline

1. Definition of Causal Effects and 2 Examples
2. Assumptions Needed to Identify Causal Effects from Observed Data Distribution
3. Estimation Methods and Assumptions Needed for Consistency of Estimators
4. Computing Standard Errors (using bootstrap)
5. Potential Challenges You May Encounter

# Definitions and Goal

- X=Baseline Variables, Z=Binary treatment or Exposure, Y=Outcome.
- Observed Data Structure: $(X_i, Z_i, Y_i)$ for each study participant $i = 1, \ldots, n$.
- Goal is to estimate the effect of the treatment/exposure on the outcome.
- We focus on **population average treatment effect (abbreviated ATE)**, a contrast between what the mean outcome would be **if everyone in population were assigned to treatment versus everyone assigned to control**.
- Main challenges we address: how to account for measured confounding
- We assume no unmeasured confounding (i.e., X contains all confounders)

# Example 1

- Population: HIV infected individuals; Data from cohort study.
- X=age, sex, ART-naive; Z=Indicator of Adherence > 50% to Antiretroviral Therapy during month; Y=Indicator of Virologic Failure.
- Data structure: observe $(X_i, Z_i, Y_i)$ for each study participant $i$

# Example 2

- Population: individuals with intracerebral hemorrhage (ICH); data from randomized trial
- X=ICH volume, ICH location, age, NIH Stroke Scale; Z=Indicator of Received Surgical Intervention; Y=Modified Rankin Scale < 4 at 180 days.
- Data structure: observe $(X_i, Z_i, Y_i)$ for each study participant $i$

# Main Challenge

- X=Baseline Variables, Z=Binary treatment or Exposure, Y=Outcome.
- Observed Data Structure: $(X_i, Z_i, Y_i)$ for each study participant $i = 1, \ldots, n$.
- Problem in observational study: those with Z=1 may not be comparable to those with Z=0 in baseline characteristics related to $Y$.
- Difference in sample proportions with $Y = 1$ comparing Z=1 and Z=0 groups can have confounding/selection bias for estimating ATE.

# Need to Introduce Potential Outcomes

- Potential outcomes $Y_0$, $Y_1$, i.e., outcome **under assignment to** $Z = 0, 1$, respectively.
- Goal is to estimate causal effect, e.g., difference of proportions $P[Y_1 = 1] - P[Y_0 = 1]$
  $P[Y_0 = 1]$ is population proportion under hypothetical intervention where everyone **assigned** $Z = 0$.
  $P[Y_1 = 1]$ is population proportion under hypothetical intervention where everyone **assigned** $Z = 1$.
- The fundamental challenge of causal inference: only one of $Y_0$, $Y_1$ is observed for each person, i.e., the one corresponding to their $Z$ value.
- Therefore, half the potential outcomes are missing. Goal is inferences about $Y_0$ and $Y_1$ in a hypothetical population where none of these missing.
- If there are confounders, then $P[Y = 1|Z = 1] \neq P[Y_1 = 1]$.

# Defining Causal Effect using Potential Outcomes

- Potential outcomes $Y_0$, $Y_1$, i.e., outcome **under assignment to** $Z = 0, 1$, respectively.
- Goal is to estimate causal effect, e.g.,
  risk difference $P[Y_1 = 1] - P[Y_0 = 1]$,
  risk ratio $P[Y_1 = 1]/P[Y_0 = 1]$,
  log odds ratio $\text{logit}[P(Y_1 = 1)] - \text{logit}[P(Y_0 = 1)]$ where
  $\text{logit}(x) = \log[x/(1-x)]$.
- Note: $\text{expit} = \text{logit}^{-1}$.

# Note on Interpretation of Conditioning

- $P[Y = 1|Z = 1]$ is read as “Probability of Y=1 GIVEN (conditioned on) Z=1".
- “GIVEN" can be interpreted as “among those in the population with" or "among strata with"
- $P[Y = 1|Z = 1]$ is read as “Probability of Y equals 1 among strata with Z=1"
- $P[Y = 1|Z = 1, X = x]$ is read as “Probability of Y=1 among strata with Z=1, X=x"
- The above are population quantities, which we could in principle learn by measuring $(X, Z, Y)$ on everyone in the population; in practice we just get a sample from the population and try to infer from this about the population.

# Assumptions

- X=Baseline Variables, Z=Binary treatment, Y=Outcome.
- Goal is to estimate causal effect, e.g., $P[Y_1 = 1] - P[Y_0 = 1]$.
- Key assumptions that allow identifiability of causal effect based on observed data distribution:
  - Consistency: $Y = Y_Z = (1 - Z)Y_0 + ZY_1$ (connects observed and potential outcomes)
  - Strong ignorability: $Y_0, Y_1$ independent of $Z$ given $X$. Also called: no unmeasured confounders assumption (i.e., $X$ has all confounders). Roughly speaking, confounder of effect of $Z$ on $Y$ is a variable that impacts both.
  - Experimental Treatment Assignment (ETA): P(Z|X)>0, i.e., no stratum of $X$ where exposure/non-exposure impossible.
  - Assume each triple $(X_i, Z_i, Y_i)$ is independent, identically distributed draw from unknown joint distribution $P_{X,Z,Y}$.
  - Let $\mathcal{X}$ denote all possible values of $X$.
- For clarity of presentation we estimate one of $P[Y_1 = 1]$, $P[Y_0 = 1]$ at a time. Can then plug into the desired contrast.

# Identifiability of Causal Effects from Observational Data

Goal: Estimate $P(Y_1 = 1)$, which under the assumptions of consistency and ignorability, equals

$$\sum_{x \in \mathcal{X}} P(Y = 1 \mid Z = 1, X = x) P(X = x).$$

This follows from:

$$\begin{aligned}
& P(Y_1 = 1) \\
= & \sum_{x \in \mathcal{X}} P(Y_1 = 1 \mid X = x) P(X = x) \\
= & \sum_{x \in \mathcal{X}} P(Y_1 = 1 \mid X = x, Z = 1) P(X = x) \text{ (by ignorability)} \\
= & \sum_{x \in \mathcal{X}} P(Y = 1 \mid X = x, Z = 1) P(X = x) \text{ (by consistency).}
\end{aligned}$$

We expressed $P(Y_1 = 1)$ in terms of observed data distribution.

# Identifiability of Causal Effects from Observational Data

Goal: Estimate $P(Y_1 = 1)$, which under the assumptions of consistency and ignorability, equals

$$\sum_{x \in \mathcal{X}} P(Y = 1 \mid Z = 1, X = x) P(X = x).$$

Can do similarly to estimate $P(Y_0 = 1)$ by changing to $Z = 0$ in above.
Note: in general

$$\sum_{x \in \mathcal{X}} P(Y = 1 \mid Z = 1, X = x) P(X = x) \qquad \neq \qquad P(Y = 1 | Z = 1).$$

This is because in general $P(X = x|Z = 1) \neq P(X = x)$ due to selection bias.

# Estimation

Goal: Estimate $P(Y_1 = 1)$, which was shown to equal

$$\sum_{x \in \mathcal{X}} P(Y = 1 \mid Z = 1, X = x) P(X = x).$$

1. Standardization (a.k.a. g-computation) Fit outcome regression working model $m_1(X, \alpha)$ for $P(Y = 1 \mid Z = 1, X)$. Estimator is $\frac{1}{n} \sum_{i=1}^{n} m_1(X_i, \hat{\alpha})$.
2. Inverse Probability Weighting (Horvitz-Thompson): Fit working model $g_1(X, \gamma)$ for $P(Z = 1 \mid X)$. Estimator is $\frac{1}{n} \sum_{i=1}^{n} Z_i Y_i / g_1(X_i, \hat{\gamma})$.
3. Double Robust Estimator: Involves fitting both models. Many options. E.g., if both models are logistic regression, first fit $g_1$, then fit $m_1$ using weights $1/g_1(X_i, \hat{\gamma})$ and denote fitted coefficients by $\bar{\alpha}$. Estimator is $\frac{1}{n} \sum_{i=1}^{n} m_1(X_i, \bar{\alpha})$. (Due to Marshall Joffe.)

# Estimation

Goal: Estimate $P(Y_1 = 1)$, which was shown to equal

$$\sum_{x \in \mathcal{X}} P(Y = 1 \mid X = x, Z = 1) P(X = x).$$

Requirements for consistency of estimators:

1. Standardization (a.k.a. g-computation) Outcome regression working model $m_1(X, \alpha)$ for $P(Y = 1 \mid Z = 1, X)$ must be correctly specified.
2. Inverse Probability Weighting: Propensity score working model $g_1(X, \gamma)$ for $P(Z = 1 \mid X)$ must be correctly specified.
3. Double Robust Estimator: Involves fitting both models. At least one working model must be correctly specified.

Note: our ultimate goal is to estimate causal effect, not coefficient vectors $\alpha, \gamma$. Causal effect is generally not equal to any of these coefficients.

# Estimation

Goal: Estimate $P(Y_1 = 1)$, which was shown to equal

$$\sum_{x \in \mathcal{X}} P(Y = 1 \mid X = x, Z = 1) P(X = x).$$

1. Standardization: Fit a logistic regression model $m_1(X, \alpha)$ for $P(Y = 1 \mid Z = 1, X)$ (using only those with $Z = 1$). Estimator is $\frac{1}{n} \sum_{i=1}^{n} m_1(X_i, \hat{\alpha})$.
   That is, the empirical average, **over all subjects (even those with $Z = 0$) of their predicted outcomes if they'd gotten $Z = 1$, based only on their baseline variables $X_i$, using the outcome regression model fit.** For example, if you fit model

$$P(Y = 1 | Z = 1, X) = \text{expit}(\alpha_0 + \alpha_1 X + \alpha_2 X^2),$$

   this estimator is: $\frac{1}{n} \sum_{i=1}^{n} \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 X_i + \hat{\alpha}_2 X_i^2)$.

# Estimation

Goal: Estimate $P(Y_1 = 1)$, which was shown to equal

$$\sum_{x \in \mathcal{X}} P(Y = 1 \mid X = x, Z = 1) P(X = x).$$

1. Inverse Weighted Estimator (also called IPW, IPTW): Fit a logistic regression model $g_1(X, \gamma)$ for $P(Z = 1|X)$ using all participants; this is called propensity score model. Estimator is $\frac{1}{n} \sum_{i=1}^{n} Z_i Y_i / g_1(X_i, \hat{\gamma})$.
   For example, if you fit model

$$P(Z = 1|X) = \text{expit}(\gamma_0 + \gamma_1 X + \gamma_2 X^2),$$

   this estimator is: $\frac{1}{n} \sum_{i=1}^{n} Z_i Y_i / \text{expit}(\hat{\gamma}_0 + \hat{\gamma}_1 X_i + \hat{\gamma}_2 X_i^2)$.

# Estimation

Goal: Estimate $P(Y_1 = 1)$, which was shown to equal

$$\sum_{x \in \mathcal{X}} P(Y = 1 \mid X = x, Z = 1) P(X = x).$$

1. Double Robust Estimator of Joffe: First fit propensity score logistic regression model $g_1(X, \gamma)$. Next, fit outcome regression logistic regression model $m_1(X, \alpha)$ using weights $1/g_1(X_i, \hat{\gamma})$ and denote fitted coefficients by $\bar{\alpha}$. Estimator is $\frac{1}{n} \sum_{i=1}^{n} m_1(X_i, \bar{\alpha})$.

# Estimation

Goal: Estimate $P(Y_1 = 1)$, which was shown to equal

$$\sum_{x \in \mathcal{X}} P(Y = 1 \mid X = x, Z = 1) P(X = x).$$

Another Double Robust Estimator: Fit $g_1$, then fit logistic regression model $m_1$ with additional term $Z/g_1(X, \hat{\gamma})$. Estimator is $\frac{1}{n} \sum_{i=1}^{n} m_1(X_i, \hat{\alpha})$.

# Computing Standard Errors for Various Estimators

Consider any of the above estimators we've discussed.
In general, can use nonparametric bootstrap to estimate the standard error, when data has sample size *n*:

- Repeatedly (say, 10,000 times) resample *n* units with replacement from your data set to created a replicated data set of size *n*.
- Compute estimator on replicated data set.
- Compute the standard deviation of the 10,000 estimates–this is the estimate of the standard error.

Note: for each replicated data set, when computing the estimator, you should refit the models. This captures the variability due to the model parameters being estimated rather than known a priori.
Recommendation: use BCa method for confidence interval.

# Potential Challenges

1. Very small estimated values of $P(Z = z|X)$; called “practical Experimental Treatment Assignment violation". Leads to very large weights. May need to truncate weights; or can modify the quantity being estimated.
2. Too many variables to adjust for and not enough participants $n$. Watch out for model overfit.
3. Assumption Violations (which can be hard or sometimes impossible to detect)

References:


- Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for non-ignorable drop-out using semiparametric non-response models (with discussion). Journal of the American Statistical Association 1999; 94:1096–1146.
- van der Laan, M. J. and D. Rubin (2006, October). Targeted Maximum Likelihood Learning. The International Journal of Biostatistics 2 (1).
- Robins JM, Sued M, Lei-Gomez Q, Rotnizky A. Comment: performance of double-robust estimators when inverse probability weights are highly variable. Statistical Science 2007; 22(4):544–559.
- Robins JM, Sued M, Lei-Gomez Q, Rotnitzky A. Double-robustness with improved efficiency in missing and causal inference models. Technical Report, Harvard School of Public Health, 2007

Acknowledgments: Slides are from a course co-taught with Constantine Frangakis.
Thank you to Jacob Fiksel who provided this beamer template.