Precision Medicine: Lecture 02
Causal Inference

Michael R. Kosorok,
Nikki L. B. Freeman and Owen E. Leete

Department of Biostatistics
Gillings School of Global Public Health
University of North Carolina at Chapel Hill

Fall, 2019

Outline

Introduction

- Precision medicine leverages patient heterogeneity through data-driven methods so that the "right treatment is given to the right patient at the right time."
- What exactly does "right treatment is given to the right patient at the right time" mean?
- This is a fundamentally causal question.
- By data-driven methods we mean that we want to answer causal questions empirically in a principled, statistical way
- The field of **causal inference** provides a framework for thinking about causal inference and estimating causal effects.

We will use the potential outcomes framework

- Causal inference is an active area of research across many different disciplines including biostatistics, statistics, economics, epidemiology, and philosophy
- There are many "flavors" of causal inference...Pearl, Robbins, Dawid, Rubin,...
- We are going to use the potential outcomes framework (Neyman (1923), Rubin (1974, 2005), Robins (1986, 1987)).

Outline

Causal treatment effect

- Example: Outcome Y, two possible treatments (exposures, interventions) $\mathcal{A} = \{0, 1\}$
- We compare the outcome that occurs when action $A = 1$ is taken compared to the outcome that occurs when action $A = 0$ is taken
- If the outcomes differ, we say that A has a **causal effect** on Y.
- Let $Y^*(a)$ denote the outcome that would occur if action a, $a \in \{0, 1\}$, is taken.
- Then we say that there is an (individual) **causal treatment effect** when

$$Y^*(1) - Y^*(0) \neq 0.$$

Potential outcomes

- Only one $a \in \mathcal{A}$ can be observed for an individual
- Thus only one of $Y^*(0)$ and $Y^*(1)$ can be observed. Because of this $Y^*(0)$ and $Y^*(1)$ are called **potential outcomes** or **counterfactual outcomes**
  - Potential outcomes from the before-treatment perspective
  - After treatment, the treatment that the individual does not receive is the counterfactual treatment
- In general, individual causal treatment effects are not identifiable
- We can, however, try to identify **average causal effects** in the target population.

$$\tau^* = E\{Y^*(1) - Y^*(0)\}$$

General set-up

- We observe iid (Y_i, X_i, A_i), $i = 1, \ldots, n$
  - Y is the outcome of interest
  - X are observed baseline covariates
  - A is the observed treatment (action, exposure, intervention) with $\mathcal{A}$ denoting the finite set of possible treatments
- Our goal is to estimate τ^*
  - We need to estimate $E\{Y^*(a)\}$
- Note that we will focus primarily on the causal average difference to elucidate the ideas of causal inference. However, one could similarly consider the causal risk ratio or causal odds ratio.
- We will also focus on the two-treatment setting for ease of exposition.
- Under what conditions can we estimate τ^*?

Causal consistency and SUTVA

- **Causal consistency**:
  $Y_i = Y_i^*(0)(1 - I(A_i = 1)) + Y_i^*(1)I(A_i = 1)$,
  $i = 1, \ldots, n$.
  - The outcome *observed* for individual i who received treatment A_i is the same as his or her *potential* outcome for that treatment.
- **Stable unit treatment value assumption (SUTVA)**
  - No interference
  - The potential outcomes for an individual are not affected by the treatments received by others nor by others' potential outcomes

Outline

Identification of causal effects in the randomized treatment setting

- Under causal consistency and SUTVA, we can identify causal effects using data from a randomized trial
- Randomization ensures that treatment assignment is independent of all other factors, including the outcome an individual would have under any of the treatment options:

$$\{Y^*(0), Y^*(1)\} \perp\!\!\!\perp A.$$

This is called **exchangeability**.

Identification of causal effects in the randomized treatment setting

Claim: For data from a randomized trial, under our assumptions

$$\tau^* = E[Y|A=1] - E[Y|A=0].$$

Observe that:

$$\begin{aligned} E[Y|A=1] &= E[Y^*(1)I(A=1) + Y^*(0)I(A=0)|A] \\ &= E[Y^*(1)|A] \\ &= E[Y^*(1)]. \end{aligned}$$

Similarly, $E[Y|A=0] = E[Y^*(0)]$

Estimation of causal effects in the randomized treatment setting

- A natural estimator for the average causal treatment effect is

$$\hat{\tau} = \frac{\sum_{i=1}^{n} Y_i A_i}{\sum_{i=1}^{n} A_i} - \frac{\sum_{i=1}^{n} Y_i (1 - A_i)}{\sum_{i=1}^{n} (1 - A_i)}.$$

- This estimator is consistent by the law of large numbers and continuous mapping theorem.

Outline

Identification in the observational study setting

- Without randomized treatment assignment, it is no longer reasonable to assume that exchangeability holds.
- Individuals who receive a particular treatment may be fundamentally different from those that receive the other treatment.
- Without exchangeability, our previously derived estimator is probably not consistent for the causal treatment effect.

No unmeasured confounders

- Suppose that we could find a subset of covariates, X^*, that are associated with both prognosis and treatment assignment.
- If we look at individuals with the same level of all of those covariates, treatment assignment is effectively at random:

$$\{Y^*(0), Y^*(1)\} \perp\!\!\!\perp A | X^*.$$

This is called **conditional exchangeability**.

No unmeasured confounders

To identify causal effects in the observational study setting, we will need two additional assumptions:

- The **no unmeasured confounders** assumptions (sometimes called strong ignorability) assumes that all variables X^* used to make treatment decisions are measured in the data $X^* \subseteq X$ and

$$\{Y^*(0), Y^*(1)\} \perp\!\!\!\perp A|X.$$

  - This assumption cannot be verified emprically. We must rely on subject-matter expertise.

- The **positivity** assumption is that

$$P(A = a|X = x) > 0 \text{ for all x such that } P(X = x) > 0,$$

  where we have treated X as discrete to avoid measure-theoretic arguments.

Identification in the observational study setting

Claim: For data from an observational study and under our assumptions (consistency, SUTVA, no unmeasured confounders, and positivity)

$$\tau^* = E[E[Y|X, A = 1]] - E[E[Y|X, A = 0]]$$

Observe that

$$\begin{aligned} E\{Y^*(1)\} &= E[E[Y^*(1)|X]] \\ &= E[E[Y^*(1)|X, A = 1]] \\ &= E[E[Y|X, A = 1]] \end{aligned}$$

Similarly, $E\{Y^*(0)\} = E[E[Y|X, A = 0]]$.

Causal diagrams represent causal relationships

- Before discussing estimation for causal effects in the observational study setting, we will highlight causal graphs
- Causal graphs provide a way to summarize our knowledge and assumptions about causal problems
- This is important in most realistic settings when the relationship between variables is complex
- They are also useful when working with non-statisticians
- We will not delve into the rich details of causal graphs. Our presentation will focus on the fundamentals.

DAGs

- An arrow from one variable to another indicates a direct causal effect (i.e. not mediated through another variable) for at least one individual
- A lack of an arrow between two variables indicates that we know or are assuming that there is no direct causal effect between the two variables for any individuals
- Causal diagrams are **directed acyclical graphs (DAGs)**
  - **Directed** because the edges imply a direction
  - **Acyclical** because there can be no cycles (a variable cannot cause itself directly or through another variable(s))
- A DAG is a *causal* DAG if the common causes of any pair of variables in the graph are also in the graph

DAG Examples 1

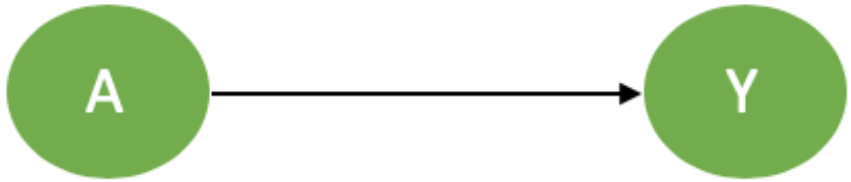

Figure 1: DAG for a simple RCT

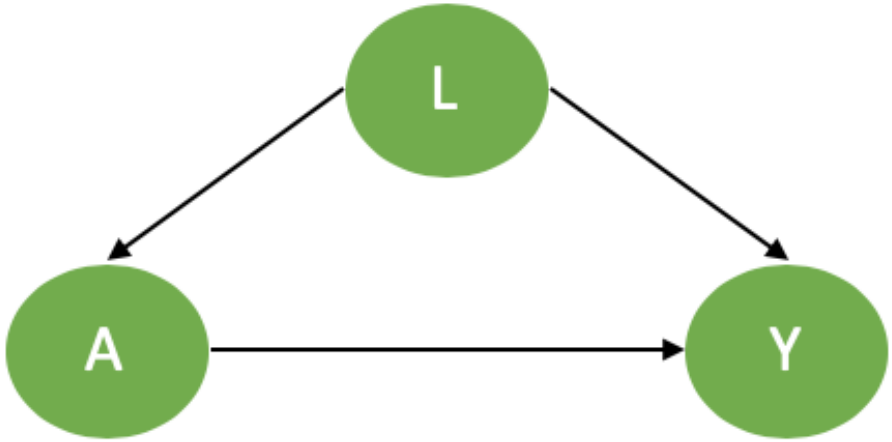

Figure 2: DAG for a simple observational study

Confounders, mediators, and colliders

Variables in a causal analysis are classified by how they affect the causal relationship of interest. DAGs can help us understand these terms.

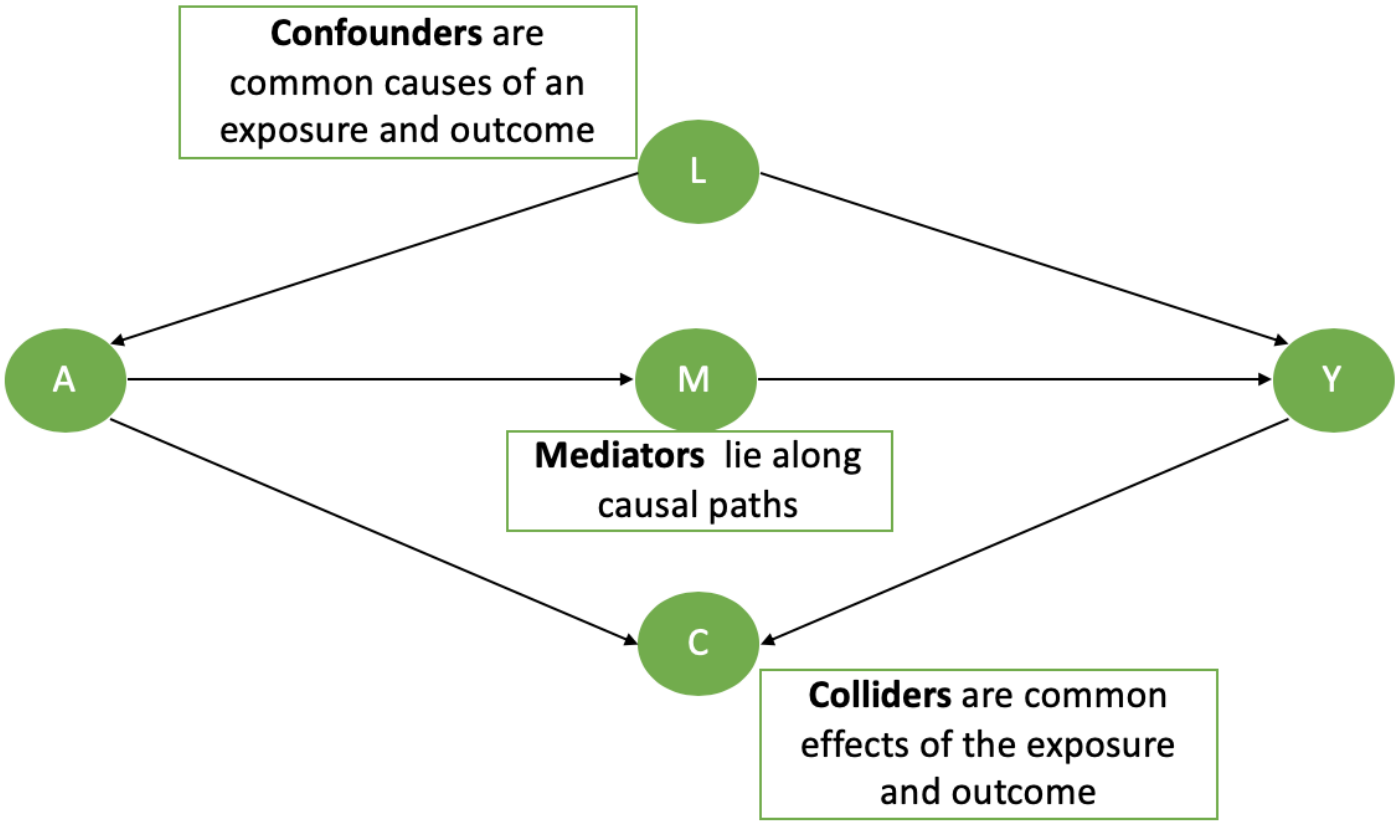

Figure 3: Confounders, mediators, and colliders

DAGs encode marginal and conditional independence between variables

- Marginal and conditional independence are critical for making arguments for exchangeability and conditional exchangeability, respectively, and thus for making causal claims
- Two variables are (marginally) associated if one causes the other, or if they share common causes. Otherwise they will be (marginally) independent.
- Conditioning on a common cause, denoted on a DAG by drawing a box around the variable being conditioned on, "breaks the flow" of association
- It seems like we should just condition on everything to do causal inference...
- Conditioning on a collider or a descendant of a collider induces an association between two variables

DAG Examples 2

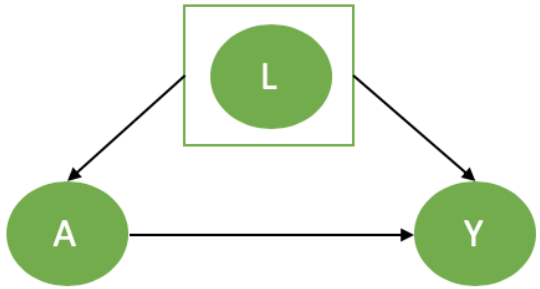

Figure 4: Simple example of conditioning

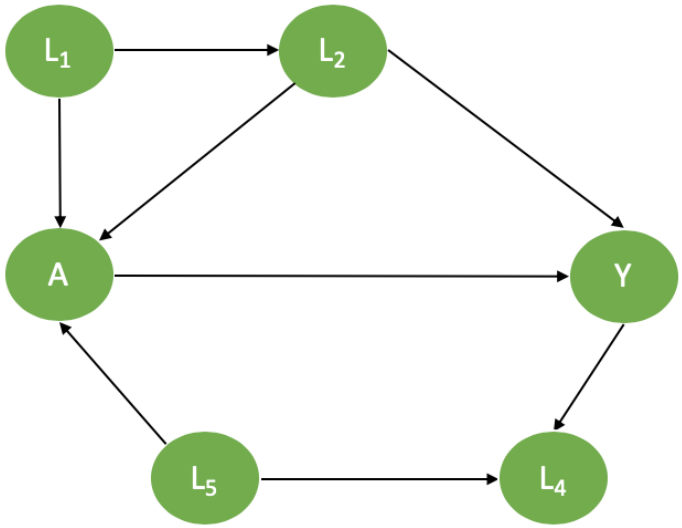

Figure 5: On which variable(s) should we condition?

Why use DAGs?

- DAGs force us to be explicit in our assumptions when doing causal inference
- DAGs help us visualize whether variables are marginally or conditionally independent. This is fundamental for establishing (conditional) exchangeability
- Much more could be said about DAGs, how to characterize them, and how to utilize them
- We note that we have described causal effects in terms of potential outcomes, yet potential outcomes do not show up in the DAGs we have looked at so far
- Richardson and Robins (2013) developed Single World Intervention Graphs (SWIGs) which marge the "DAG world" with the "potential outcomes" world.

Estimation via outcome regression

- Recall that under our assumptions, causal effects in the observational setting are identified and

$$\tau^* = E[E[Y|X, A = 1]] - E[E[Y|X, A = 0]]$$

- This depends on the regression of the observed outcome on covariates and observed treatment received
- If $E[Y|X = x, A = a] = Q(x, a)$ is the true regression relationship then

$$\tau^* = E[Q(X = x, A = 1)] - E[Q(X = x, A = 0)]$$

- The function $Q(x, a)$ is usually unknown in practice

Estimation via outcome regression

- Even though $Q(x, a)$ is unknown, we may posit a parametric model $Q(x, a; \beta)$. For example

$$Q(x, a; \beta) = \beta_0 + \beta_1^\intercal x + \beta_2 a + \beta_3^\intercal xa$$

if Y is continuous or

$$\text{logit}(Q(x, a; \beta)) = \beta_0 + \beta_1^\intercal x + \beta_2 a + \beta_3^\intercal xa$$

if Y is binary.

- Estimation of the posited parametric model can be done via standard methods (least squares regression, weighted least squares, maximum likelihood, etc.)

Estimation via outcome regression

- Under the assumption that the posited model is correctly specified, and
- For an estimator $\hat{\beta}$ for β, the outcome regression estimator is given by

$$\hat{\tau}^*_{\text{OR}} = \frac{1}{n}\sum_{i=1}^{n}\{Q(X_i, 1; \hat{\beta}) - Q(X_i, 0; \hat{\beta})\}$$

Propensity score methods

- An alternative to outcome regression for estimating causal effects is using propensity score methods
- For the two treatment case, the propensity score is defined as

$$\pi(X) = P(A = 1|X),$$

  the probability of treatment conditional on baseline covariates. This can be extended to the multi-treatment case.
- The propensity score is key because under the assumption of no unmeasured confounders

$$\{Y^*(1), Y^*(0)\} \perp\!\!\!\perp A|\pi(X)$$

Propensity score methods

- By conditioning on the propensity score, we can identify the causal effect using the observed data

$$\tau^* = E[E\{Y|\pi(X), A = 1\}] - E[E\{Y|\pi(X), A = 1\}]$$

- Methods using this idea include
  - Propensity score stratification
  - Regression adjustment on the propensity score
  - Propensity score matching
- A very appealing approach is the IPW estimator

Inverse propensity score weighting

- Because we only see $Y^*(0)$ or $Y^*(1)$, the problem of causal inference can be cast as a missing data problem and suggests the use of inverse probability weighted estimators proposed by Horvitz and Thompson (1952)
- Based on the semiparametric theory for missing data, consider the estimator for $E\{Y^*(1)\}$

$$E\{Y^*(1)\} = \frac{1}{n}\sum_{i=1}^{n}\frac{A_i Y_i}{\pi(X_i)}$$

and the estimator for $E\{Y^*(0)\}$

$$E\{Y^*(0)\} = \frac{1}{n}\sum_{i=1}^{n}\frac{(1-A_i)Y_i}{1-\pi(X_i)}$$

Justification for the IPW estimator

Under causal consistency, SUTVA, and NUC

$$
\begin{aligned}
E\left\{\frac{AY}{\pi(X)}\right\} &= E\left\{\frac{AY^*(1)}{\pi(X)}\right\} \\
&= E\left[E\left\{\frac{AY^*(1)}{\pi(X)}|Y^*(1),X\right\}\right] \\
&= E\left[\frac{E\{A|Y^*(1),X\}Y^*(1)}{\pi(X)}\right] \\
&= E\{Y^*(1)\}
\end{aligned}
$$

because

$$E\{A|Y^*(1),X\} = E\{A|X\} = P(A=1|X) = \pi(X).$$

Similarly, $\left\{\frac{(1-A)Y}{1-\pi(X)}\right\} = E\{Y^*(0)\}$.

IPW estimator

With unbiased estimators for $E\{Y^*(1)\}$ and $E\{Y^*(0)\}$, a natural estimator for the average causal effect is

$$\hat{\tau}^*_{\text{IPW}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i Y_i}{\pi(X_i)} - \frac{(1-A_i)Y_i}{1-\pi(X_i)}\right\}.$$

In an observational study setting, $\pi(X)$ is unknown. A solution is to posit and estimate a propensity model, $\pi(x;\gamma)$ (e.g. logistic). Assuming that the propensity model is well-specified,

$$\hat{\tau}^*_{\text{IPW}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i Y_i}{\pi(X_i;\hat{\gamma})} - \frac{(1-A_i)Y_i}{1-\pi(X_i;\hat{\gamma})}\right\}.$$

The approximate sampling distribution for this estimator can be derived using M-estimation theory.

AIPWE

Based on semiparametric efficiency theory, the augmented inverse propensity weighted estimator "augments" the IPW estimator to increase efficiency. Among estimators of the form

$$\hat{\tau}^*_{\text{AIPWE}} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{A_i Y_i}{\pi(X_i;\hat{\gamma})} - \frac{(1-A_i)Y_i}{1-\pi(X_i;\hat{\gamma})} - \{A_i - \pi(X_i;\hat{\gamma})\}h(X_i)\right],$$

the optimal efficient estimator has

$$h(X) = \frac{E(Y|X, A=1)}{\pi(X)} + \frac{E(Y|X, A=0)}{1-\pi(X)}.$$

Doubly robust estimator

Given fitted models $\pi(x;\hat{\gamma})$ and $Q(x,a;\hat{\beta})$, a logical estimator for the average causal effect is

$$\hat{\tau}^*_{\text{DR}} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{A_i Y_i}{\pi(X_i;\hat{\gamma})} - \frac{(1-A_i)Y_i}{1-\pi(X_i;\hat{\gamma})} - \frac{A_i - \pi(X_i;\hat{\gamma}}{\pi(X_i;\hat{\gamma})}Q(X_i,1;\hat{\beta}) - \frac{A_i - \pi(X_i;\hat{\gamma})}{1-\pi(X_i;\hat{\gamma})}Q(X_i,0;\hat{\beta})\right].$$

- $\hat{\tau}^*_{\text{DR}}$ is consistent for τ^* if one of $\pi(x;\gamma)$ or $Q(x,a;\beta)$ is correctly specified.
- For this reason, it is often called the "**doubly robust**" estimator.
- In an RCT, $\pi(X)$ is known and the doubly robust estimator is consistent regardless of whether the outcome model is correctly specified.

Instrumental variables

- The methods we have discussed so far have required that we have no unmeasured confounders
- Instrument variables (IVs) offer an alternative way to estimate causal effects without having measured all adjustment factors
- A variable Z is an **instrument** if it satisfies the following conditions
  1. Z is associated with A,
  2. Z does not affect Y except through its potential effect on A, and
  3. Z and Y do not share causes.
- To identify causal effects using an IV requires a fourth assumption, either *homogenity* or *monotonicity* (sometimes called the *no defiers* assumption).

Outline

Precision medicine and causal inference

- Precision medicine uses the tools of causal inference to find dynamic treatment regimes
- One way to think about precision medicine is to cast it as a multi-stage causal inference problem
- However, this occludes some of the more fundamental differences between the two fields
- In particular, a usual target of causal inference is

$$\tau^* = E[\tau(X)]$$

where $\tau(X) = E[Y|X, A = 1] - E[Y|X, A = 0]$ is the **conditional average treatment effect**. For DTRs, we care mostly about

$$\text{sign}(\tau(x)).$$