Precision Medicine: Lecture 02
Causal Inference

Lina Montoya, 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

Acknowledgments

Drawn from lectures by:

- Maya Petersen, MD PhD
- Laura Balzer, PhD
- Jen Ahern, PhD

Resources:

- Introduction to Causal Inference course `https://www.ucbbiostat.com/`
- Petersen, Maya L., and Mark J. van der Laan. "Causal models and learning from data: integrating causal modeling and statistical estimation." Epidemiology (Cambridge, Mass.) 25.3 (2014): 418.
- Balzer, Laura B., and Maya L. Petersen. "Invited Commentary: Machine Learning in Causal Inference – How Do I Love Thee? Let Me Count the Ways." American Journal of Epidemiology (2021).

Outline

Introduction

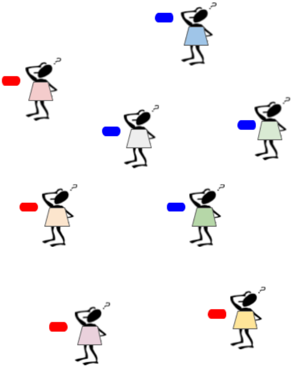
Defining causal questions and inference

The Causal Roadmap applied to the average treatment effect

The Causal Roadmap applied to Precision Medicine causal questions

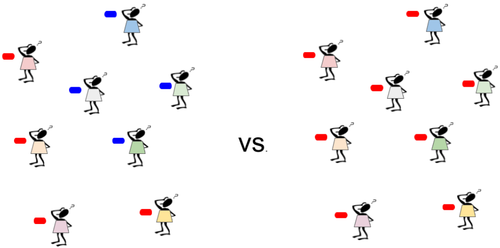
Introduction

- In precision medicine, we often want to know: **who should get which treatment?**
  - And at what time?
  - And at what dose?
  - And in what sequence?
  - Etc.



Introduction

- In precision medicine, we often want to know: who should get which treatment?
  - And at what time?
  - And at what dose?
  - And in what sequence?
  - Etc.
- Further: **Is giving treatment in an individualized way better than giving it in a non-individualized way (e.g., giving everyone the same treatment)?**



Introduction

- In precision medicine, we often want to know: who should get which treatment?
  - And at what time?
  - And at what dose?
  - And in what sequence?
  - Etc.
- Further: Is giving treatment in an individualized way better than giving it in a non-individualized way (i.e., giving everyone the same treatment)?
- **These are fundamentally causal questions**
- **The field of causal inference provides formal frameworks for answering causal questions**

Outline

Introduction

Defining causal questions and inference

The Causal Roadmap applied to the average treatment effect

The Causal Roadmap applied to Precision Medicine causal questions

What makes a question causal?

- A causal question asks about the world under manipulated conditions
  - For example, manipulating the way a treatment was administered
  - Question: What would outcomes look like if all participants were exposed VS. if the same participants, over the same time-frame, and under the same conditions were not exposed?

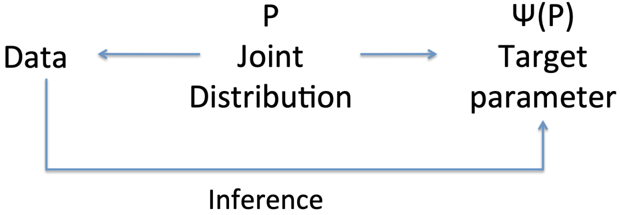
We are not time travelers → answering causal questions can be *hard*

Contrast this with other kinds of questions

- Descriptive questions
  - E.g.: What was the average age of participants in our study?
  - Think: Table 1 in applied journal papers
- Statistical questions
  - E.g.: Which risk factors are associated with breast cancer?
  - E.g.: How likely were women who took vitamins to have breast cancer?
  - Think: risk scores/prediction problems

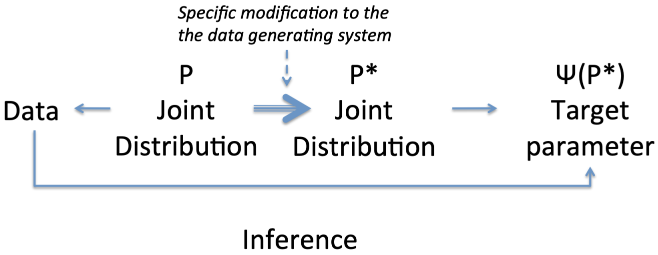
Statistical vs. Causal Inference

- Statistical inference - inference on a statistical parameter
  - Statistical parameter - summary of an observed data distribution based on a sample drawn from that distribution
    - May require statistical assumptions, e.g., on the distribution of the observed data



Statistical vs. Causal Inference

- Causal inference - inference on a causal parameter
  - Causal parameter - summary of a distribution we do not (fully) observe based on a sample drawn from that distribution
    - Causal parameters provide a summary of how a data distribution would change under manipulated conditions, so they require assumptions beyond statistical ones, namely, about the processes that generated the data



Statistical vs. Causal Inference

Example: We sample individuals from some underlying population and on each subject observe: A=vitamin use, Y=breast cancer

- Statistical parameter
  - How likely were women who took vitamins to have breast cancer vs. those who did not take vitamins?
    - $P(Y = 1|A = 1) - P(Y = 1|A = 0)$

Statistical vs. Causal Inference

Example: We sample individuals from some underlying population and on each subject observe: A=vitamin use, Y=breast cancer

- Statistical parameter
  - How likely were women who took vitamins to have breast cancer vs. those who did not take vitamins?
    - $P(Y = 1|A = 1) - P(Y = 1|A = 0)$
- Causal parameter
  - What is the probability of breast cancer had all women had taken vitamins vs. if all women had not taken vitamins?
    - ???

Statistical vs. Causal Inference

- Statistical questions (based on what we observe) are different than causal questions (based on what we wish to have observed)
- Answering causal questions is hard – yet, in public health (and beyond), we ask causal questions all the time...

OK, so I think I'm asking a causal question...?

RCTs?
Observational study?
Quasi-experiment?
SMARTs?

DAGs?
SWIGs?
Potential outcomes?

Logistic regression?
Linear assumptions?
Machine learning?

Mediators?
Modifiers?
Moderators?
Instrumental variables?
Colliders?

Double robust?
Double machine learning?

M-bias?
Collider bias?
Confounding bias?

No unmeasured confounding?
E-values?
Sensitivity analysis?

Estimators?
Estimands?
Estimates?

IPTW? TMLE?
AIPW? IPCW?
G-computation?

SUTVA?
Positivity?
Exchangeability?
Cross-world independence?
Interference?

Missing data?
Censoring?
Non-complianc
e?

Parametric?
Non-parametric?
Semi-parametric?

Causal Roadmap to the Rescue

A general Roadmap[1] for tackling causal questions...

1. Specify data and causal model representing *real* background knowledge
2. Specify causal question
3. Specify causal parameter that answers causal question
4. Specify observed data and link to causal model
5. Identification
6. Statistical estimation and inference
7. Interpret results

[1] Petersen & van der Laan, 2014

Outline

Introduction

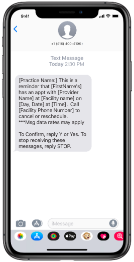
Defining causal questions and inference

The Causal Roadmap applied to the average treatment effect

The Causal Roadmap applied to Precision Medicine causal questions

Running example - Average Treatment Effect

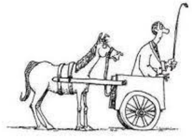
- Antiretroviral Therapy (ART) potent and effective, but 20-40% of HIV-infected patients in sub-Saharan Africa lost to follow-up within two years after enrolled in care[2]
- Want to know the impact of **SMS text message reminders** on patient **retention in HIV care**



[2] Geng, et. al., 2010

Why use The Roadmap?

- Usual approach might be...
  - Get EHR data/questionnaires measured at first clinic visit, SMS app information, clinic retention records starting at first visit
  - Since the outcome is binary, decide to use logistic regression for analysis
  - Estimate conditional odds ratios by exponentiating the regression coefficient on the prevention package
- What's the problem here?
  - You're allowing the tool (e.g., logistic regression) to define the question, rather than starting with the question you care about and picking a tool that allows you to answer that question
- Solution: causal roadmap!



The Roadmap

1. **Specify data and causal model representing real background knowledge**
2. Specify causal question
3. Specify causal parameter that answers causal question
4. Specify observed data and link to causal model
5. Identification
6. Statistical estimation and inference
7. Interpret results

Step 1: Data and causal model

What are the data, and how do those variables in the data relate to each other?

Step 1: Data and causal model

Data:

- Measured variables
  - X : the set of baseline covariates
    - age, sex, SES, viral load at baseline, time on ART, CD4 count...
  - A : the exposure/intervention
    - intervention to send SMS messages, where $a \in \mathcal{A} = \{$send SMS $= 1$, standard of care $= 0\}$
  - Y : the outcome
    - lapse in care one year after baseline, where $y \in \mathcal{Y} = \{$no lapse in care $= 1$, lapse in care$= 0\}$
- Unmeasured variables
  - U : the set of unmeasured factors
    - genetic factors, cultural factors...

Step 1: Data and causal model

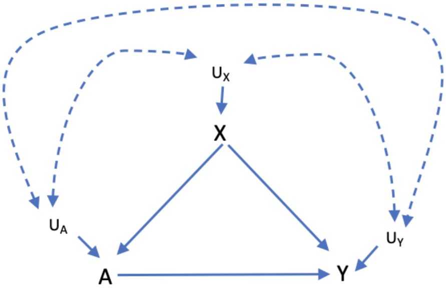
- Causal modeling formalizes what we actually know about how the variables relate to each other - however limited
  - What measured variables may affect each other?
  - What is the role of background factors?
  - What is the functional form of these relationships?
- Can encode this knowledge with directed acyclic graphs (DAGs) and structural causal models

Step 1: Data and causal model

- Structural causal model:

$$X = f_X(U_X)$$
$$A = f_A(U_A, X)$$
$$Y = f_Y(U_Y, A, X)$$

- No assumptions on the background factors (U_X, U_A, U_Y)
- No assumptions on the functional form of the equations (f_X, f_A, f_Y)



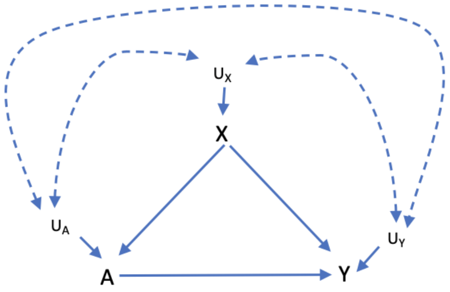
Double-headed arrows = shared causes

Step 1: Data and causal model

No unmeasured confounding setting? How would SCM and graph change?

- Structural causal model:

$$X = f_X(U_X)$$
$$A = f_A(U_A, X)$$
$$Y = f_Y(U_Y, A, X)$$



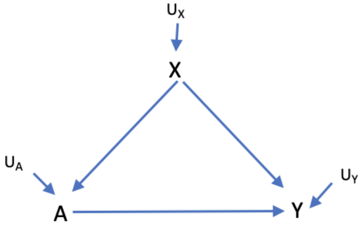
Step 1: Data and causal model

No unmeasured confounding setting? How would SCM and graph change?

- Structural causal model:

$$X = f_X(U_X)$$
$$A = f_A(U_A, X)$$
$$Y = f_Y(U_Y, A, X)$$

- Background factors are all independent (no unmeasured common causes)
- Still no functional form assumptions

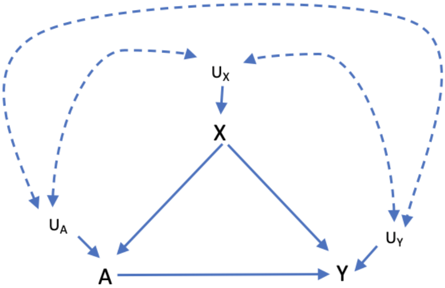


Step 1: Data and causal model

RCT setting? How would SCM and graph change?

- Structural causal model:

$$X = f_X(U_X)$$
$$A = f_A(U_A, X)$$
$$Y = f_Y(U_Y, A, X)$$



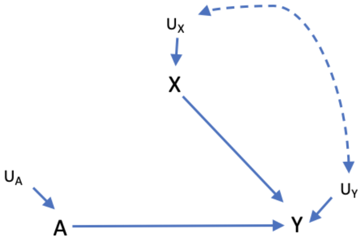
Step 1: Data and causal model

RCT setting? How would SCM and graph change?

- Structural causal model:

$$X = f_X(U_X)$$
$$A = U_A$$
$$Y = f_Y(U_Y, A, X)$$

- A determined by flip of a coin $U_A \sim Bernoulli(0.5)$
- A not a function of X (exclusion restriction)

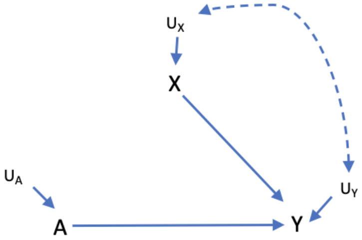


Step 1: Data and causal model

In this example, this will be our SCM

- Structural causal model:

$$X = f_X(U_X)$$
$$A = U_A$$
$$Y = f_Y(U_Y, A, X)$$



  - RCT part of **ADAPT-R** = **Ada**ptive Strategies for **P**reventing and **T**reating Lapses of **R**etention in HIV Care

The Roadmap

1. Specify data and causal model representing *real* background knowledge
2. **Specify causal question**
3. Specify causal parameter that answers causal question
4. Specify observed data and link to causal model
5. Identification
6. Statistical estimation and inference
7. Interpret results

Step 2: Causal question

- What is the **effect** of SMS text message reminders on patient retention in HIV care?
- Consider one hypothetical experiment
  - What would be the difference in one-year patient retention had everyone received SMS reminders versus if no one had received SMS reminders?
  - Why is this an example of a causal question?

Step 2: Causal question

- What is the **effect** of SMS text message reminders on patient retention in HIV care?
- Consider one hypothetical experiment
  - What would be the difference in patient retention had everyone received SMS reminders versus if no one had received SMS reminders?
  - Why is this an example of a causal question? Because we're asking about the probability of retention under different conditions than observed (we only observe retention under one SMS condition)
- Many other hypothetical experiments possible
- Note: giving SMS to everyone (or no one) is an example of a *static* (vs. dynamic) treatment rule
  - Assigning SMS, regardless of patient covariates

The Roadmap

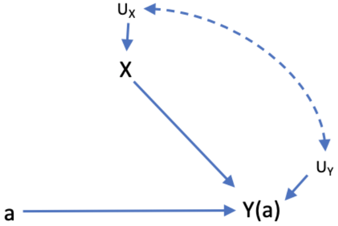
1. Specify data and causal model representing *real* background knowledge
2. Specify causal question
3. **Specify causal parameter that answers causal question**
4. Specify observed data and link to causal model
5. Identification
6. Statistical estimation and inference
7. Interpret results

Step 3: Causal parameter

- Counterfactual (potential) outcomes are defined by modifications to the data generating process described by the model

$$X = f_X(U_X)$$
$$A = a$$
$$Y_a = f_Y(U_Y, a, X)$$

- Y_a: the counterfactual outcome, if possibly contrary to fact, a person received exposure level A=a
  - Y_1: counterfactual one-year retention outcome under SMS (A=1)
  - Y_0: counterfactual one-year retention outcome under standard of care (A=0)



"Surgically modified" model

Step 3: Causal parameter

Note: other notation for counterfactual outcomes...

- $Y(a)$, Y^a, $Y^{(a)}$
- In this lecture, we'll use Y_a

Step 3: Causal parameter

- Use counterfactuals to define the target causal parameter $=$ summary of distribution we do not (fully) observe
- Examples:
  - Treatment-specific mean
    - E.g.: The expected counterfactual one-year retention outcome had all patients received SMS

$$\mathbb{E}[Y_1]$$

Step 3: Causal parameter

- Use counterfactuals to define the target causal parameter = summary of distribution we do not (fully) observe
- Examples:
  - Treatment-specific mean
    - E.g.: The expected counterfactual one-year retention outcome had all patients received SMS

$$\mathbb{E}[Y_1]$$

  - Average treatment effect
    - E.g.: The difference in the expected counterfactual one-year retention outcome had all patients received SMS vs. standard of care:

$$\mathbb{E}[Y_1] - \mathbb{E}[Y_0]$$

Step 3: Causal parameter

- Use counterfactuals to define the target causal parameter $=$ summary of distribution we do not (fully) observe
- Examples:
  - Treatment-specific mean
    - E.g.: The expected counterfactual one-year retention outcome had all patients received SMS

$$\mathbb{E}[Y_1]$$

  - Average treatment effect
    - E.g.: The difference in the expected counterfactual one-year retention outcome had all patients received SMS vs. standard of care:

$$\mathbb{E}[Y_1] - \mathbb{E}[Y_0]$$

  - Causal relative risk: $\mathbb{E}[Y_1]/\mathbb{E}[Y_0]$

Step 3: Causal parameter

- Use counterfactuals to define the target causal parameter = summary of distribution we do not (fully) observe
- Examples:
  - Treatment-specific mean
    - E.g.: The expected counterfactual one-year retention outcome had all patients received SMS

$$\mathbb{E}[Y_1]$$

  - Average treatment effect
    - E.g.: The difference in the expected counterfactual one-year retention outcome had all patients received SMS vs. standard of care:

$$\mathbb{E}[Y_1] - \mathbb{E}[Y_0]$$

  - Causal relative risk: $\mathbb{E}[Y_1]/\mathbb{E}[Y_0]$
  - Causal odds ratio: $\frac{\mathbb{E}[Y_1]/(1-\mathbb{E}[Y_1])}{\mathbb{E}[Y_0]/(1-\mathbb{E}[Y_0])}$

Step 3: Causal parameter

- Use counterfactuals to define the target causal parameter = summary of distribution we do not (fully) observe
- Examples:
  - Treatment-specific mean
    - E.g.: The expected counterfactual one-year retention outcome had all patients received SMS

$$\mathbb{E}[Y_1]$$

  - Average treatment effect
    - E.g.: The difference in the expected counterfactual one-year retention outcome had all patients received SMS vs. standard of care:

$$\mathbb{E}[Y_1] - \mathbb{E}[Y_0]$$

  - Causal relative risk: $\mathbb{E}[Y_1]/\mathbb{E}[Y_0]$
  - Causal odds ratio: $\frac{\mathbb{E}[Y_1]/(1-\mathbb{E}[Y_1])}{\mathbb{E}[Y_0]/(1-\mathbb{E}[Y_0])}$
  - etc...

Step 3: Causal parameter

- Use counterfactuals to define the target causal parameter = summary of distribution we do not (fully) observe
- Examples:
  - Treatment-specific mean
    - E.g.: The expected counterfactual retention outcome had all patients received SMS
$$\mathbb{E}[Y_1]$$
  - Average treatment effect (i.e., causal risk difference if binary outcome)
    - E.g.: The difference in the expected counterfactual one-year retention outcome had all patients received SMS vs. standard of care:
$$\mathbb{E}[Y_1] - \mathbb{E}[Y_0]$$
  - Causal relative risk: $\mathbb{E}[Y_1]/\mathbb{E}[Y_0]$
  - Causal odds ratio: $\frac{\mathbb{E}[Y_1]/(1-\mathbb{E}[Y_1])}{\mathbb{E}[Y_0]/(1-\mathbb{E}[Y_0])}$
  - etc...

Which of the above parameters are more causal?

Step 3: Causal parameter

- Use counterfactuals to define the target causal parameter = summary of distribution we do not (fully) observe
- Examples:
  - Treatment-specific mean
    - E.g.: The expected counterfactual retention outcome had all patients received SMS
      $$\mathbb{E}[Y_1]$$
  - Average treatment effect (i.e., causal risk difference if binary outcome)
    - E.g.: The difference in the expected counterfactual one-year retention outcome had all patients received SMS vs. standard of care:
      $$\mathbb{E}[Y_1] - \mathbb{E}[Y_0]$$
  - Causal relative risk: $\mathbb{E}[Y_1]/\mathbb{E}[Y_0]$
  - Causal odds ratio: $\frac{\mathbb{E}[Y_1]/(1-\mathbb{E}[Y_1])}{\mathbb{E}[Y_0]/(1-\mathbb{E}[Y_0])}$
  - etc...

Nothing "more" or "less causal" about any of these parameters

Step 3: Causal parameter

Focus first on **average treatment effect (ATE)**

- E.g.: The difference in the expected counterfactual one-year retention outcome had all patients received SMS vs. standard of care:

$$\mathbb{E}[Y_1] - \mathbb{E}[Y_0]$$

The Roadmap

1. Specify data and causal model representing *real* background knowledge
2. Specify causal question
3. Specify causal parameter that answers causal question
4. **Specify observed data and link to causal model**
5. Identification
6. Statistical estimation and inference
7. Interpret results

Step 4a: Specify observed data

- Since we are not time-travelers, we need to specify the data that can *actually* be observed
- Remember: for one patient, we actually observe:

$$O = (X, A, Y) \sim P$$

  - X : baseline covariates (e.g., age, sex, SES, viral load at baseline, time on ART, CD4 count...)
  - A : exposure (e.g., indicator of receiving SMS text messages)
  - Y : outcome (e.g., indicator of HIV care retention during one year)
  - With probability distribution P

Step 4b: Link observed data to causal model

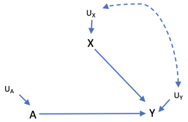
- How does the data we observe O relate to the manipulated, causal model?

Step 4b: Link observed data to causal model

- How does the data we observe O relate to the manipulated, causal model?
- We assume the observed data were generated by sampling n times from a data generating system specified by (described by) our causal model
  - Here we will assume (but don't have to) that our dataset are $1,153$ i.i.d. copies $O_1, O_2, ..., O_n$ drawn from P

Step 4c: Specify a statistical model

- Formally, a statistical model $\mathcal{M}$ is the set of possible distributions of the observed data
  - $P \in \mathcal{M}$
- Our causal model implies the statistical model
- Here, our causal model only places restrictions on our statistical model through the treatment mechanism
  - No other functional form assumptions
  - E.g.: our causal model says exposure $A = U_A \sim Bernoulli(0.5)$, i.e., $P(A = 1|X) = 0.5$
  - We specified this in our causal model (in step 1)



Step 4c: Continuum of Statistical Models (informally)

- Non-parametric model: no restrictions on the set of possible distributions P
- Semi-parametric model: some restrictions on the set of possible distributions P
- Parametric model: assumes that we know P up to a finite number of unknown parameters

Step 4c: Continuum of Statistical Models (informally)

- Non-parametric model: no restrictions on the set of possible distributions P
- Semi-parametric model: some restrictions on the set of possible distributions P
- Parametric model: assumes that we know P up to a finite number of unknown parameters

Note: The point is not that parametric (or semiparametric) models are bad...the point is that our statistical model should accurately reflect our knowledge (it should contain the true distribution of O)

The Roadmap

1. Specify data and causal model representing *real* background knowledge
2. Specify causal question
3. Specify causal parameter that answers causal question
4. Specify observed data and link to causal model
5. **Identification**
6. Statistical estimation and inference
7. Interpret results

Step 5: Identification

- What we want: $\mathbb{E}[Y_1 - Y_0]$ (causal parameter - function of counterfactual distribution)
- What we have: sample from observed data distribution

Step 5: Identification

- What we want: $\mathbb{E}[Y_1 - Y_0]$ (causal parameter - function of counterfactual distribution)
- What we have: sample from observed data distribution

Can we write our causal parameter as something we can estimate with the observed data?

Step 5: Identification

- What we want: $\mathbb{E}[Y_1 - Y_0]$ (causal parameter - function of counterfactual distribution)
- What we have: sample from observed data distribution

Can we write our causal parameter as something we can estimate with the observed data?

What are the **assumptions** required to do this?

Step 5: Identification

We already made the following assumptions:

- **Temporality**:
  - E.g.: exposure precedes outcome
  - Indicated by arrows on the causal graph from A to Y
  - Equivalently Y as a function of A in the structural equations
- **Consistency**: $Y_a = Y$ when $A = A$
  - Counterfactuals defined through modifications to the causal model, which also provides a description of the study under existing conditions (i.e., observed exposure)
- **Stable Unit Treatment Value Assumption (SUTVA)**: no interference between units
  - Indicated by the causal model where the outcome Y is only a function of each patient's exposure A

Step 5: Identification

We need the following assumptions:

- **No unmeasured confounding** (AKA randomization/strong ignorability assumption):

  $$Y_a \perp A | X \text{ for } a \in \mathcal{A}$$

  - X is the adjustment set (i.e., confounders) - determined by backdoor criterion
    - Glossing over here- but this takes careful consideration!
- **Positivity** (AKA experimental treatment assignment): sufficient support in the exposure within possible values of our adjustment set X

  $$\min_{a \in \mathcal{A}} P(A = a | X = x) > 0 \text{ for all } w \text{ for which } P(X = x) > 0$$

Step 5: Identification

We need the following assumptions:

- **No unmeasured confounding** (AKA randomization assumption):
$$Y_a \perp A|X \text{ for } a \in \mathcal{A}$$
  - X is the adjustment set (i.e., confounders)
- **Positivity** (AKA experimental treatment assignment): sufficient support in the exposure within possible values of our adjustment set X

$$\min_{a \in \mathcal{A}} P(A = a|X = x) > 0 \text{ for all } w \text{ for which } P(X = x) > 0$$

Do these assumptions hold in our running example?

Step 5: Identification

We need the following assumptions:

- **No unmeasured confounding** (AKA randomization assumption):
$$Y_a \perp A|X \text{ for } a \in \mathcal{A}$$
  - X is the adjustment set (i.e., confounders)
- **Positivity** (AKA experimental treatment assignment or ETA): sufficient support in the exposure within possible values of our adjustment set X
$$\min_{a \in \mathcal{A}} P(A = a|X = x) > 0 \text{ for all } w \text{ for which } P(X = x) > 0$$

Yes! Both hold by RCT design (SMS randomized, $P(A|X)$ is known)

Step 5: Identification

With the aforementioned assumptions:

$$
\begin{aligned}
\mathbb{E}[Y_1] &= \mathbb{E}[\mathbb{E}[Y_1|X]] \text{ by the tower rule} \\
&= \mathbb{E}[\mathbb{E}[Y_1|A=1,X]] \text{ under randomization} \\
&= \mathbb{E}[\mathbb{E}[Y|A=1,X]] \text{ by consistency and positivity}
\end{aligned}
$$

Step 5: Identification

With these assumptions:

$$\begin{aligned}\mathbb{E}[Y_1] &= \mathbb{E}[\mathbb{E}[Y_1|X]] \text{ by the tower rule}\\ &= \mathbb{E}[\mathbb{E}[Y_1|A=1,X]] \text{ under randomization}\\ &= \mathbb{E}[\mathbb{E}[Y|A=1,X]] \text{ by consistency and positivity}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[Y_0] &= \mathbb{E}[\mathbb{E}[Y_0|X]] \text{ by the tower rule}\\ &= \mathbb{E}[\mathbb{E}[Y_1|A=0,X]] \text{ under randomization}\\ &= \mathbb{E}[\mathbb{E}[Y|A=0,X]] \text{ by consistency and positivity}\end{aligned}$$

Step 5: Identification

So under the aforementioned assumptions:

$$\begin{aligned}\mathbb{E}[Y_1 - Y_0] =& \mathbb{E}[\mathbb{E}[Y|A=1,X] - \mathbb{E}[Y|A=0,X]] \\ =& \sum_x [\mathbb{E}[Y|A=1,X=x] \\ & - \mathbb{E}[Y|A=0,X=x]]P(X=x)\end{aligned}$$

- The RHS is our statistical parameter
- Also called the G-computation formula[3]
- "Difference in the expected outcome given the exposure and the confounders and the expected outcome given no exposure and the confounders, averaged (standardized) with respect to the confounder distribution"
- (summation generalizes to integral for continuous X)

[3] Robins, 1986

The Roadmap

1. Specify data and causal model representing *real* background knowledge
2. Specify causal question
3. Specify causal parameter that answers causal question
4. Specify observed data and link to causal model
5. Identification
6. **Statistical estimation and inference**
7. Interpret results

Step 6: Statistical estimation and inference

- Recall our statistical parameter:

$$\Psi(P) = \mathbb{E}[\mathbb{E}[Y|A=1,X] - \mathbb{E}[Y|A=0,X]]$$

- Many estimators available
  - Simple substitution estimator based on G-computation formula
  - Inverse probability of treatment weighting (IPTW)[4]
  - Augmented IPTW (AIPW)[5]
  - Targeted maximum likelihood estimation (TMLE)[6]
  - ...

[5] Hernan & Robins, 2006; Rosenbaum & Rubin, 1983

[6] Robins, Rotnitzky, & Zhao, 1994; Scharfstein, Rotnitzky, & Robins, 1999; Robins, 1999

[6] Rosenblum & van der Laan, 2010; van der Laan & Rose, 2011

Step 6: Statistical estimation and inference

- Pause to recall the "usual" approach
  1. Run logistic regression of the outcome (retention) Y on the exposure (SMS messages) A and baseline confounders X

     $$\text{Logit}(\mathbb{E}[Y|A,X]) = \beta_0 + \beta_1 A + \beta_2 X_1 + ... + \beta_{19} X_{18}$$

  2. Exponentiate the coefficient on the exposure and interpret the association in terms of a conditional odds ratio that "holds other factors constant"
- But: our target parameter $\Psi(P)$ does not equal e^{β_1}!
  - Relies on outcome regression model being correctly specified
  - Letting the estimation approach drive the question...

Step 6: Estimation with simple sub. estimator based on G-comp.

- Focuses on estimation of the "outcome regression"
  - Expected outcome, given the exposure and confounders
  - $Q(A, X) = \mathbb{E}[Y|A, X]$

$$\frac{1}{n}\sum_{i=1}^{n}\left[\hat{Q}(A_i = 1, X_i) - \hat{Q}(A_i = 0, X_i)\right]$$

Step 6: Estimation with IPTW

- Can think of confounding as a problem of biased sampling
  - Certain exposure-covariate subgroups are over-represented relative to what we would see in a randomized trial
  - Other exposure-covariate subgroups are under-represented
- Apply weights to up-weight under-represented observations and down-weight over-represented observations
- Average weighted outcomes and compare

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\mathbb{I}[A_i=1]}{\hat{P}(A_i=1|X_i)}Y_i\right]-\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\mathbb{I}[A_i=0]}{\hat{P}(A_i=0|X_i)}Y_i\right]$$

Step 6: Challenge of estimation

- But we often do not know enough to correctly specify these regressions
  - G-comp. relies on a regression based on a correctly specified parametric model of the conditional expected outcome $Q(A, X)$
  - Inverse weighting also relies on a regression based on a correctly specified parametric model of the propensity score $P(A = 1|X)$
- But parametric restrictions $\rightarrow$ misspecification $\rightarrow$ bias and misleading inference
  - Wrong answers and wrong conclusions
  - Increasing sample size (i.e., "big data") will not solve this problem $\rightarrow$ more precision around the wrong answer

Step 6: Incorporating machine learning in estimation?

- Estimate complex relationships in data flexibly
- Avoid introducing unsubstantiated assumptions during estimation of outcome regression or propensity score

Step 6: Prediction vs. causal inference

Can we use ML instead of parametric regressions in IPTW or G-computation?

- No reliable way to obtain statistical inference
  - No theory to support that ML in place of parametric regressions in “single robust” estimators have a normal (or any limit) distribution
- ML has different goals than causal effect estimation

Step 6: Incorporating ML in estimation

- AIPW

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{I}[A_i=1]}{\hat{P}(A_i|X_i)}(Y_i-\hat{Q}(A_i,X_i))+\hat{Q}(1,X_i)$$
$$-\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{I}[A_i=0]}{\hat{P}(A_i|X_i)}(Y_i-\hat{Q}(A_i,X_i))+\hat{Q}(0,X_i)$$

- TMLE

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{I}[A_i=1]}{\hat{P}(A_i|X_i)}(Y_i-\hat{Q}^*(A_i,X_i))+\hat{Q}(1,X_i)$$
$$-\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{I}[A_i=0]}{\hat{P}(A_i|X_i)}(Y_i-\hat{Q}^*(A_i,X_i))+\hat{Q}^*(0,X_i)$$

  - $\hat{Q}^*$ targeted toward parameter we care about (not Q)

Step 6: Incorporating ML in estimation

AIPW and TMLE

- Double-robust: can incorporate machine learning to avoid unsubstantiated assumptions, while maintaining valid statistical inference
  - Consistent estimate if either the outcome regression or propensity score is consistently estimate
  - Asymptotically linear under regularity conditions (by the Central Limit Theorem, normally distributed)
  - Efficient: lowest possible variance if both outcome regression and propensity score are consistently estimated at reasonable rates

Step 6: Statistical estimation and inference

- Recall our statistical parameter:

$$\Psi(P) = \mathbb{E}[\mathbb{E}[Y|A=1,X] - \mathbb{E}[Y|A=0,X]]$$

- Many estimators available
  - Simple substitution estimator based on G-computation formula
  - Inverse probability of treatment weighting (IPTW)
  - Augmented IPTW (AIPW)
  - Targeted maximum likelihood estimation (TMLE)
  - ...

Which of the above estimators are more causal?

Step 6: Statistical estimation and inference

- Recall our statistical parameter:

$$\Psi(P) = \mathbb{E}[\mathbb{E}[Y|A=1,X] - \mathbb{E}[Y|A=0,X]]$$

- Many estimators available
  - Simple substitution estimator based on G-computation formula
  - Inverse probability of treatment weighting (IPTW)
  - Augmented IPTW (AIPW)
  - Targeted maximum likelihood estimation (TMLE)
  - ...

Nothing "more" or "less causal" about these estimators – by themselves, these are not "causal methods"

The Roadmap

1. Specify data and causal model representing *real* background knowledge
2. Specify causal question
3. Specify causal parameter that answers causal question
4. Specify observed data and link to causal model
5. Identification
6. Statistical estimation and inference
7. **Interpret results**

Step 7: Interpretation

- To what degree have identifiability assumptions have been met? This informs the strength of interpretation

Step 7: Interpretation

- If there are concerns about identifiability (e.g., temporal ordering unclear, key confounder not measured) the parameter can be interpreted as an **association** (statistical interpretation)
  - Interpretation: The difference in the outcome associated with everyone exposed, compared to everyone unexposed, accounting for the measured confounders
  - E.g.: The difference in the probability retention associated with SMS messages was X, after controlling for age, sex, SES, viral load, CD4 count,...

Step 7: Interpretation

- If we believe identifiability assumptions are met, the parameter can be interpreted as the **average treatment effect**
  - Interpretation: The difference in the expected outcome if everyone were exposed compared with if everyone were unexposed
  - E.g.: There would be a X difference in the probability of retention if everyone had received SMS messages vs. standard of care

Step 7: Interpretation

- If we believe identifiability assumptions are met, the parameter can be interpreted as the **average treatment effect**
  - Interpretation: The difference in the expected outcome if everyone were exposed compared with if everyone were unexposed
  - E.g.: There would be a X difference in the probability of retention if everyone had received SMS messages vs. standard of care
- Note that we are in the RCT setting, but RCT interpretation requires:
  - Effective randomization
  - Perfect compliance
  - Perfect follow up

Step 7: Interpretation

- If we believe identifiability assumptions are met, the parameter can be interpreted as the **average treatment effect**
  - Interpretation: The difference in the expected outcome if everyone were exposed compared with if everyone were unexposed
  - E.g.: There would be a X difference in the probability of retention if everyone had received SMS messages vs. standard of care
- Note that we are in the RCT setting, but RCT interpretation requires:
  - Effective randomization
  - Perfect compliance
  - Perfect follow up

Where can we incorporate this knowledge?

Step 7: Interpretation

- If we believe identifiability assumptions are met, the parameter can be interpreted as the **average treatment effect**
  - Interpretation: The difference in the expected outcome if everyone were exposed compared with if everyone were unexposed
  - E.g.: There would be a X difference in the probability of retention if everyone had received SMS messages vs. standard of care
- Note that we are in the RCT setting, but RCT interpretation requires:
  - Effective randomization
  - Perfect compliance
  - Perfect follow up

  Step 1 (data and causal model) and step 2 (causal question), which will inform later steps.

Outline

Introduction

Defining causal questions and inference

The Causal Roadmap applied to the average treatment effect

The Causal Roadmap applied to Precision Medicine causal questions

Tie back to Precision Medicine

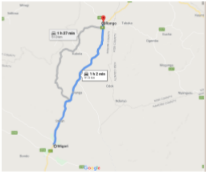
So what does this have to do with Precision Medicine?

Remember...

- In precision medicine, we often want to know: who should get which treatment?
  - And at what time?
  - And at what dose?
  - And in what sequence?
  - Etc.
- Further: Is giving treatment in an individualized way better than giving it in a non-individualized way (i.e., giving everyone the same treatment)?
- **These are fundamentally <u>causal</u> questions**
- **The field of <u>causal inference</u> provides formal frameworks for answering causal questions**

Running example

- Antiretroviral Therapy (ART) potent and effective, but 20-40% of HIV-infected patients in sub-Saharan Africa lost to follow-up within two years after enrolled in care
- Many <u>*diverse*</u> barriers to HIV care retention
- Want to know the impact of SMS text message reminders on patient retention in HIV care
- **Further: Who benefits the most from SMS text messages? Is sending SMS reminders to those who respond well to them better than SMS for all?**




Step 1: Data and causal model

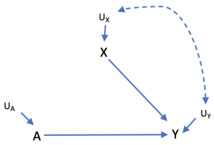
Data:

- U : the set of unmeasured factors
  - genetic factors, cultural factors...
- X : the set of baseline covariates
  - age, sex, SES, viral load at baseline, time on ART, CD4 count...
- A : the exposure
  - intervention to stay in care, where $a \in \mathcal{A} = \{\text{SMS} = 1, \text{standard of care} = 0\}$
- Y : the outcome
  - lapse in care one year after follow-up, where $y \in \mathcal{Y} = \{\text{no lapse in care} = 1, \text{lapse in care} = 0\}$

Step 1: Causal model

- Structural causal model:

$$
\begin{aligned}
X &= f_X(U_X)\\
A &= U_A \sim \mathit{Bernoulli}(0.5)\\
Y &= f_Y(U_Y, A, X)
\end{aligned}
$$



Step 2 & 3: Causal Questions & Parameters

- "Simple" Dynamic Treatment Rule & its Value
- Dynamic Treatment Rule & its Value
- Optimal Dynamic Treatment Rule & its Value

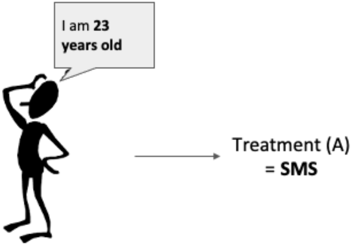
Step 2: Causal Questions - "Simple" Dynamic Treatment Rule

"Simple" Dynamic Treatment Rule

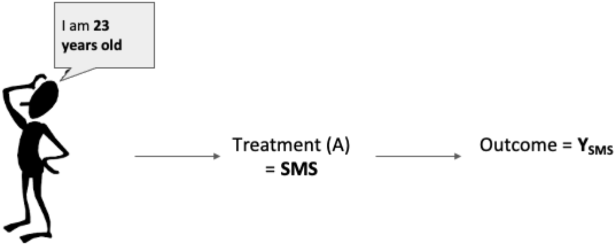
- If all participants had been given SMS **based on their age**, what proportion of participants would not have had a lapse in care?

Step 2: Causal Questions - “Simple” Dynamic Treatment Rule

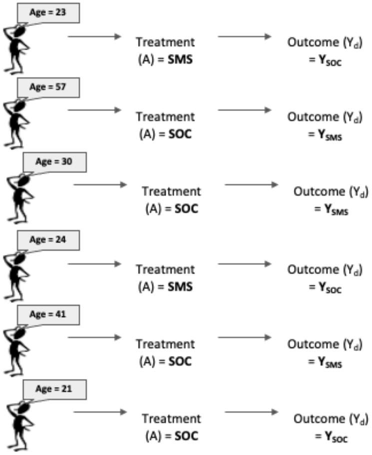
Step 2: Causal Questions - "Simple" Dynamic Treatment Rule



Step 2: Causal Questions - “Simple” Dynamic Treatment Rule



Step 2: Causal Questions - "Simple" Dynamic Treatment Rule



Step 3: Causal Parameters - “Simple” Dynamic Treatment Rule

- Define the simple dynamic treatment rule:
  - $d(age)$ = a dynamic treatment rule for assigning A based on each participants's age.

$$A = \begin{cases} \text{SMS}, & \text{if } age \leq 24 \\ \text{standard of care}, & \text{if } age > 24 \end{cases}$$

Why is this a *dynamic* (vs. static) treatment rule?

Step 3: Causal Parameters - "Simple" Dynamic Treatment Rule

- Define the simple dynamic treatment rule:
  - $d(age)$ = a dynamic treatment rule for assigning A based on each participants's age.

$$A = \begin{cases} \text{SMS}, & \text{if } age \leq 24 \\ \text{standard of care}, & \text{if } age > 24 \end{cases}$$

Why is this a *dynamic* (vs. static) treatment rule? Because we're assigning treatment based on some function of X, the covariates (e.g., age)

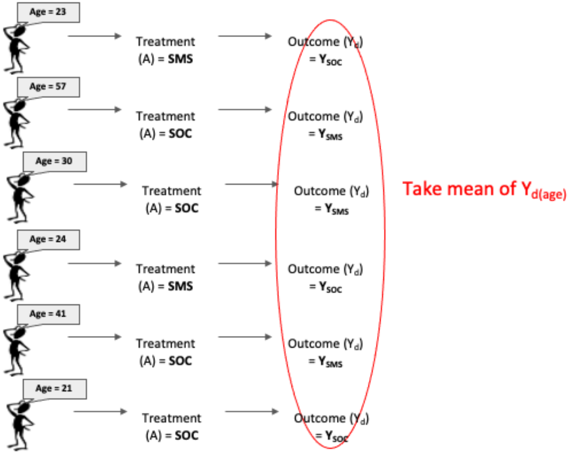
Step 3: Causal Parameters - "Simple" Dynamic Treatment Rule

- Counterfactuals

$$X = f_X(U_X)$$
$$A = d(age)$$
$$Y_{d(age)} = f_Y(U_Y, X, d(age))$$

- $Y_{d(age)}$ is the counterfactual retention status for a participant if his/her/their intervention A were assigned based on his/her/their *age*

Step 3: Causal Parameters - "Simple" Dynamic Treatment Rule



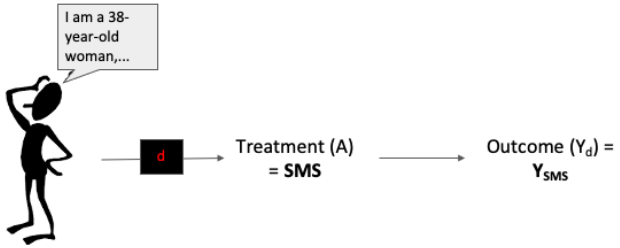
Causal parameter: $\mathbb{E}[Y_{d(age)}]$

Step 2: Causal Questions - Dynamic Treatment Rule

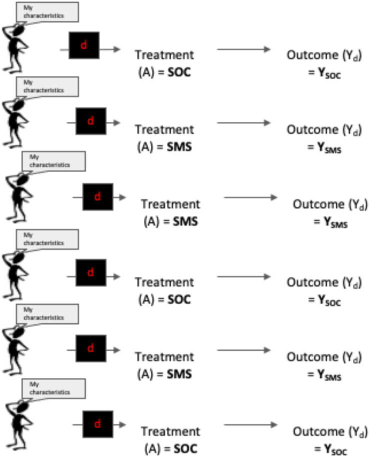
Dynamic Treatment Rule

- If all participants had been assigned SMS **based on their characteristics** (i.e., covariates), what proportion of participants would not have had a lapse in care within one year?

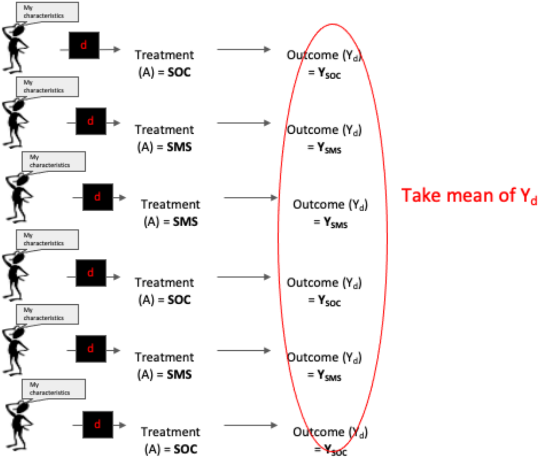
Step 2: Causal Questions - Dynamic Treatment Rule



Step 2: Causal Questions - Dynamic Treatment Rule



Step 2: Causal Questions - Dynamic Treatment Rule



Step 3: Causal Parameters - Dynamic Treatment Rule

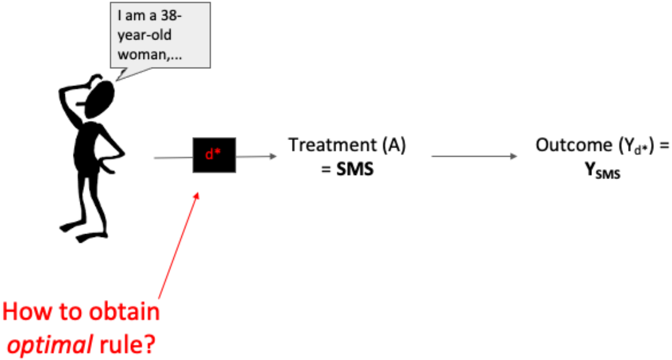
- Define $d(X)$ as a dynamic treatment rule for assigning A based on the covariates X
  - X ={age, sex, SES, viral load, CD4 count, etc...}
- The counterfactual outcomes are now:

$$
\begin{aligned}
X &= f_X(U_X) \\
A &= d(X) \\
Y_d &= f_Y(U_Y, X, d(X))
\end{aligned}
$$

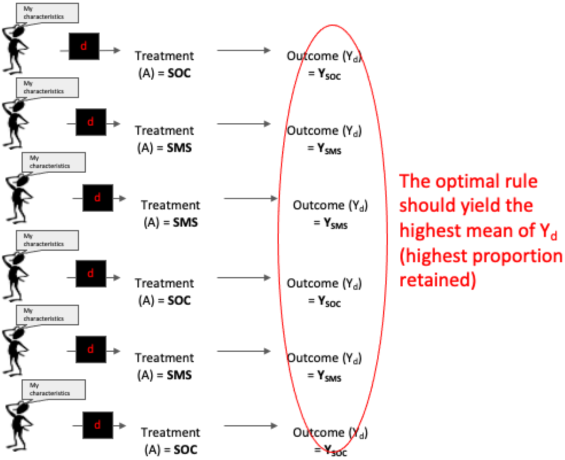
- Y_d is the counterfactual one year retention status for an individual if the intervention A were assigned using $d(X)$
- Causal parameter: probability of no lapse under dynamic treatment rule

$$\mathbb{E}[Y_d]$$

Step 2: Causal Questions - Optimal Dynamic Treatment Rule



Step 2: Causal Questions - Optimal Dynamic Treatment Rule



Step 2: Causal Questions - Optimal Dynamic Treatment Rule

Optimal Dynamic Treatment Rule

- What is the dynamic rule for assignment to SMS messages that yields the highest probability of no lapse **based on participants' characteristics** (i.e., covariates)?

Step 3: Causal Parameters - Optimal Dynamic Treatment Rule

- Define the optimal dynamic treatment rule[7]:

$$d^*(X) \in \underset{d \in \mathcal{D}}{\operatorname{argmax}} \, \mathbb{E}[Y_d]$$

  where $\mathcal{D}$ is the set of all dynamic treatment rules
- Can also define as function of conditional average treatment effect (CATE)

$$d^*(X) \equiv \mathbb{I}\Big[\mathbb{E}[Y_1 - Y_0 | X] > 0\Big]$$

[7] Murphy, 2003; Robins, 2004

Step 3: Causal Parameters - Optimal Dynamic Treatment Rule

- Counterfactuals

$$X = f_X(U_X)$$
$$A = d^*(X)$$
$$Y_{d^*} = f_Y(U_Y, X, d^*(X))$$

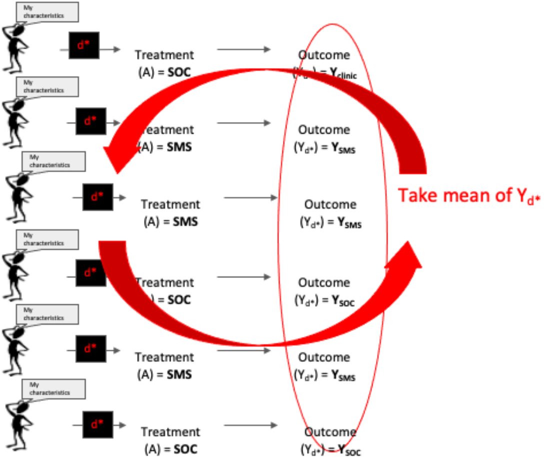
- Y_{d^*} is the counterfactual one year retention status for an individual if the intervention A were assigned based on the optimal treatment rule $d^*(X)$

Step 2: Causal Questions - Value of Optimal Dynamic Treatment Rule

Value of Optimal Dynamic Treatment Rule

- Had **all participants followed their optimal rule for SMS messages**, what proportion of participants would have been retained within a year?

Step 3: Causal Parameters - Value of Optimal Dynamic Treatment Rule



Causal parameter: $\mathbb{E}[Y_{d*}]$

Step 2 & 3: Causal Questions and Parameters - Value of Optimal Dynamic Treatment Rule

Evaluating a personalized vs. non-personalized intervention

- Had all participants followed their optimal rule for SMS messages, what proportion of participants would have been retained within a year?
- How does this compare to giving everyone SMS?

Causal parameter: $\mathbb{E}[Y_{d^*}] - \mathbb{E}[Y_1]$

(can do the same contrast with the other values of the rule)

Step 4: Specify Observed Data

- Assume that observed data $O = (X, A, Y) \sim P$ were generated by sampling $1,153$ $i.i.d.$ times from a data generating system described by the SCM

Step 5: Identification

- Identification criteria
  - Randomization (no unmeasured confounding)

$$Y_d \perp A|X \qquad \forall\ d \in \mathcal{D}$$

  - Positivity

$$\min_{a \in \mathcal{A}} P(A = a|X = x) > 0 \text{ for all } w \text{ for which } P(X = x) > 0$$

Both hold by RCT design (SMS randomized, $P(A|X)$ is known)

Step 5: Statistical Estimands

1. True ODTR
   - $d^*(X) \in \text{argmax}_{d \in \mathcal{D}} \mathbb{E}[Q(A = d(X), X)]$
   - Identify CATE as $B(X) = Q(1, X) - Q(0, X)$, the **"blip function"**, then

$$d^*(X) \equiv \mathbb{I}[B(X) > 0]$$

Step 5: Statistical Estimands

W1	W2	A	Y	E[Y|A=1,W]	E[Y|A=0,W]	Blip	d*(W)
0	0.867	1	1	0.566	0.5	0.066	1
1	1.426	1	0	0.171	0.668	-0.497	0
0	0.113	0	1	0.739	0.5	0.355	1
...	...	...	...	...	...	...	...

Step 5: Statistical Estimands

This minus this = this

W1	W2	A	Y	E[Y|A=1,W]	E[Y|A=0,W]	Blip	d*(W)
0	0.867	1	1	0.566	0.5	0.066	1
1	1.426	1	0	0.171	0.668	-0.497	0
0	0.113	0	1	0.739	0.5	0.355	1
...	...	...	...	...	...	...	...

Step 5: Statistical Estimands

Is blip > 0?

W1	W2	A	Y	E[Y|A=1,W]	E[Y|A=0,W]	Blip	d*(W)
0	0.867	1	1	0.566	0.5	0.066	1
1	1.426	1	0	0.171	0.668	-0.497	0
0	0.113	0	1	0.739	0.5	0.355	1
...	...	...	...	...	...	...	...

Step 5: Statistical Estimands

1. True ODTR
   - $d^*(X) \in \text{argmax}_{d \in \mathcal{D}} \mathbb{E}[Q(A = d(X), X)]$
   - Identify CATE as $B(X) = Q(1, X) - Q(0, X)$, the **"blip function"**, then

$$d^*(X) \equiv \mathbb{I}[B(X) > 0]$$

2. True value of true ODTR: $\mathbb{E}[Q(A = d^*(X), X)]$

Step 6 & 7: Estimation and Interpretation

Up next in the class!