Lecture 11: Introduction to TMLE; Interpretation; Wrap up

Outline

- Recap of Estimators
- Intro to TMLE
- Back to the Roadmap
- Interpreting results
- Overview of course, what we have (and have not) covered in this course

References

- TLB. Chapters 4-6

The Roadmap

1. Specify **Causal Model** representing real background knowledge
2. Specify **Causal Question**
3. Specify **Observed Data** and link to causal model
4. **Identify** : Knowledge + data sufficient?
5. Commit to an **estimand** as close to question as possible, and a **statistical model** representing real knowledge.
6. **Estimate**
7. **Interpret** Results

Estimate the Chosen Parameter of the Observed Data Distribution

- For illustration we are focusing primarily on a single statistical estimation problem
  - $O=(W,A,Y)\sim P_{0;}$ Statistical model is non or semiparametric
  - $\Psi(P_0)=E_{W,0}(E_0(Y|A=1,W)-E_0(Y|A=1,W))$
    - Equal under backdoor criteria to the ATE
- Additional discussion of estimation of parameter defined using a marginal structural model
- We are focusing on three classes of estimator

Overview of Estimators

- Each class of estimator in turn requires for its implementation an estimator of a distinct factor of the observed data distribution
- Distribution of the observed data:
- $P_0(O)=P_0(W,A,Y) = P_0(W)\ P_0(A|W)\ P_0(Y|A,W)$

Different Estimators require estimators of distinct factors of the observed data distribution

- $P_0(O)=P_0(W,A,Y)=P_0(W)P_0(A|W)\ P_0(Y|A,W)$
- Simple substitution estimators
  - Also referred to as "G-computation" estimators
  - Actually, rather than full $P_0(Y|A,W)$, only require estimators of $E_0(Y|A,W)$ (and $P_0(W)$)
- Consistency depends on consistent estimation of $E_0(Y|A,W)$
  - Super Learning can help here, but…

IPTW Estimators

- $P_0(O)=P_0(W,A,Y)=P_0(W)P_0(A|W)\ P_0(Y|A,W)$

Inverse probability weighted estimators

- Consistency of IPTW estimators depends on consistent estimation of $g_0(A|W)$
  - Super learning can help here, but….

Coming next: Double Robust Estimators

- $P_0(o)=P_0(w,a,y)=P_0(y|a,w)\ P_0(w)P_0(a|w)$

Double Robust estimators
-A-IPTW
-TMLE

- Implementation requires estimators of <u>both</u>
  $E_0(Y|A,W)$ and $g_0(A|W)$
  - Super learning
- Double Robust Estimators are consistent if <u>either</u> $E_0(Y|A,W)$ <u>or</u> $g_0(A|W)$ are estimated consistently

Double Robust Estimators

- $P_0(o)=P_0(w,a,y)=P_0(y|a,w)\ P_0(w)P_0(a|w)$

Double Robust estimators

- If <u>both</u> $E_0(Y|A,W)$ and $g_0(A|W)$ are estimated consistently, these estimators are efficient
  - Achieve lowest asymptotic variance possible for any estimator
- These asymptotic properties (describe what happens when n goes to infinity) also typically translate into lower bias and variance in finite samples…

Augmented IPTW

- Like IPTW, defined as a solution to an estimating equation
- The estimating equation corresponding to the efficient influence curve
- Efficient and double robust
- Will not cover in this class

Targeted Maximum Likelihood Estimation

- TMLE is a general methodology
- As with other estimators, we will illustrate for "G comp estimand"

$$\Psi(P_0)=E_{W,0}(E_0(Y|A=1,W)-E_0(Y|A=1,W))$$

General Overview: TMLE

- TMLE is a substitution estimator

1. Estimate the portion of P_0 that the target parameter is a function of (recall, we refer to this as Q_0)
   - $\Psi(P_0)=\Psi(Q_0)$
2. Update this initial fit of Q_0 in a step targeted at achieving the optimal bias trade off for $\Psi(Q_0)$
3. Plug in updated (targeted) fit of Q_0 into the parameter mapping Ψ to generate estimate

Overview of TMLE for ATE estimand

1. Estimate $E_0(Y|A,W)$
   - Use machine learning to respect statistical model
   - Gives "best" estimate of $E_0(Y|A,W)$
2. Modify this initial estimate of $E_0(Y|A,W)$
   - Target it to give less biased estimate of $\Psi(P_0)=E_W(E_0(Y|A=1,W)-E_0(Y|A=0,W))$
   - This targeting requires estimation of $g_0(A|W)$
3. Implement substitution estimator with new targeted estimate of $E_0(Y|A,W)$

Step by Step Overview: TMLE

1. Estimate $E_0(Y|A,W) \equiv \bar{Q}_0(A,W)$
   - Eg using super learner
- Notation for this initial estimate of $E_0(Y|A,W)$:

$$\bar{Q}_n^0(A,W)$$

"n" because it is an estimate of the true parameter value

"0" refers to initial (non-targeted) estimate

2. Generate predicted values for Y for each subject, given that subject's A,W
   - For Subject i: $\bar{Q}_n^0(A_i,W_i)$

Step by Step Overview: TMLE

3. Estimate treatment mechanism
$$g_0(a|W) \equiv P_0(A = a|W) \text{ for } a \in \mathcal{A}$$
4. Use this estimate to create a new "clever covariate" $H_n^*(A,W)$ for each subject
   - For subject i

$$H_n^*(A_i, W_i) \equiv \left(\frac{I(A_i = 1)}{g_n(A_i = 1|W_i)} - \frac{I(A_i = 0)}{g_n(A_i = 0|W_i)} \right)$$

   - We will use this clever covariate to update our initial estimate

Step by Step Overview: TMLE

5. Update the initial estimate of $E_0(Y|A,W)$

- Run a logistic regression of Y on $H_n^*(A,W)$ using $logit(\bar{Q}_n^0(A,W))$ as offset

  (suppressing additional intercept term)

- Let ε_n denote the resulting MLE estimate of the coefficient on $H_n^*(A,W)$
- Update the initial estimate

$$logit(\bar{Q}_n^1(A,W)) = logit(\bar{Q}_n^0(A,W)) + \epsilon_n H_n^*(A,W)$$

  - Could iterate, but here we have convergence in one step

Step by Step Overview: TMLE

6. Calculate predicted values for each subject under each treatment level of interest using the updated estimate $\bar{Q}_n^1(A, W)$

- For each subject, set a=1 and a=0 and generate predicted outcome with updated estimate

$$logit(\bar{Q}_n^1(1, W_i)) = logit(\bar{Q}_n^0(1, W_i)) + \epsilon_n H_n^*(1, W_i)$$

$$logit(\bar{Q}_n^1(0, W_i)) = logit(\bar{Q}_n^0(0, W_i)) + \epsilon_n H_n^*(0, W_i)$$

Step by Step Overview: TMLE

7. Estimate $\Psi(P_0)$ as the empirical mean of the predicted values of Y for a=1 and a=0, based on the updated fit

$$\hat{\Psi}_{TMLE}(P_n) = \frac{1}{n}\sum_{i=1}^{n}\left[\bar{Q}_n^1(1, W_i) - \bar{Q}_n^1(0, W_i)\right]$$

Influence curve-based Inference

- Variance of an asymptotically linear estimator is well-approximated by the variance of its Influence curve/n
  - Conditions for asymptotic linearity of TMLE: Ch. 27 TLB

1. Estimate the Influence curve by plugging in estimates of g_0 and Q_0
   - Resulting estimate of the influence curve is a function of the observed data only
2. Take sample variance and divide by sample size
   - If estimator of g_0 is consistent, this will give an asymptotically conservative estimator of the variance of the TMLE
     - See p. 96 TLB

TMLE: Some take home messages

- TMLE is Double Robust: Consistent if either g_0 or Q_0 are estimated consistently
  - This can translate into real bias gains in finite samples
- TMLE is efficient if g_0 and Q_0 are both estimated consistently
  - This can translate into real variance gains in finite samples

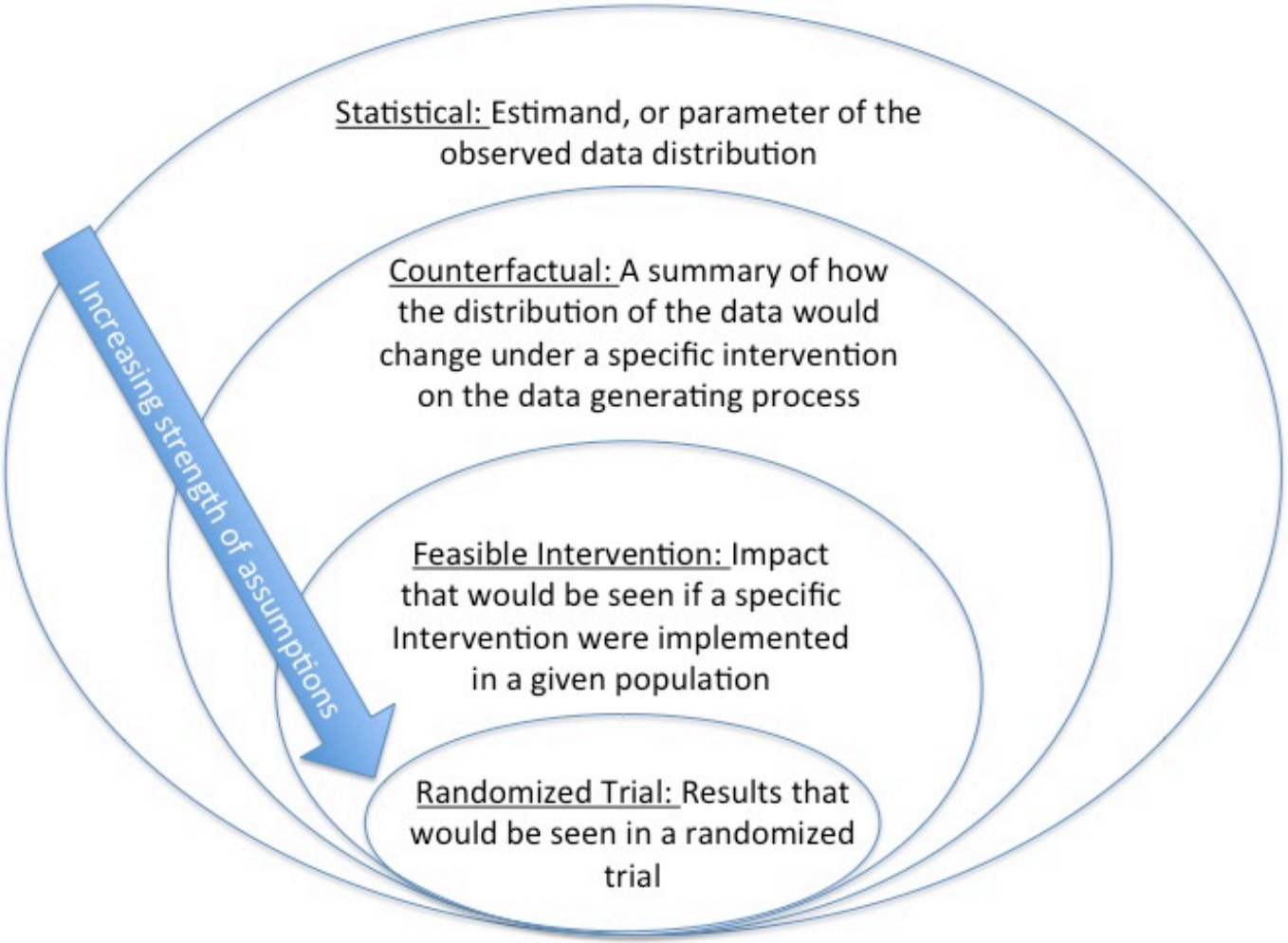
TMLE: Some take home messages

- Use data-adaptive estimation (Super Learning) for g and Q
  - Asymptotic linearity relies on bias disappearing at a fast enough rate
  - Influence curve based inference relies on g_0 being estimated consistently
    - Conservative variance estimate
  - Consistent estimation of both g_0 and Q_0 gives us efficiency

Back to the Roadmap

1. Specify **Causal Model** representing real background knowledge
2. Specify **Causal Question**
3. Specify **Observed Data** and link to causal model
4. **Identify** : Knowledge + data sufficient?
5. Commit to an **estimand** as close to question as possible, and a **statistical model** representing real knowledge.
6. **Estimate**
7. **Interpret** Results

A Hierarchy of Interpretations

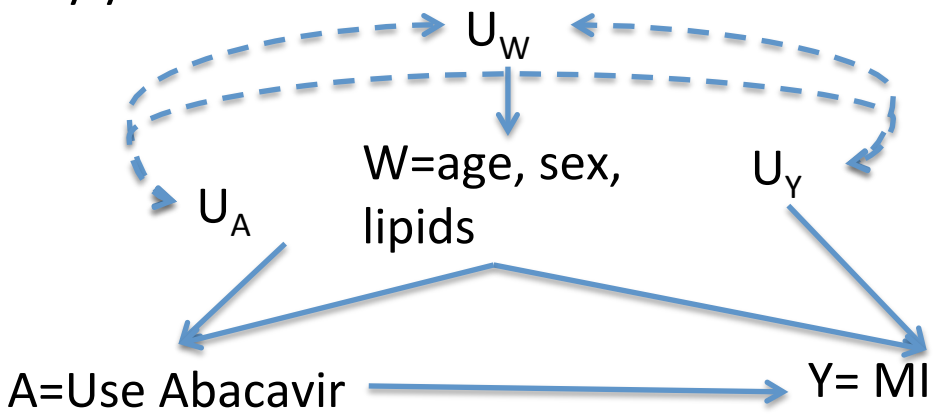


Example: Abacavir and Cardiovascular Disease

- Analysis of observational data from several cohorts suggested abacavir use associated with increased risk of myocardial infarction among treated HIV-infected population
  - Other analyses found no evidence of such an association….
- Example of a causal question: Does use of abacavir (ABC) increase risk of myocardial infarction (MI)?

Example: Abacavir and Cardiovascular Disease, point treatment version

- X=O=(W,A,Y)
  - W= baseline covariates (age, sex, lipid profile) measured at start of ART
  - A= Indicator First ART Regimen contains Abacavir
  - Y= Myocardial Infarction by year 5



- Target Causal Parameter: $E_{U,X}(Y_1-Y_0)$

$$\Psi(P_0) = \sum_w E_0(Y \mid A=1, W=w) - E_0(Y \mid A=0, W=w) P_0(W=w)$$

$$\hat{\Psi}(P_n) = 0.02 \ (95\% \text{ CI: } 0.01, 0.03)$$

Statistical Interpretation

- An estimate of our statistical target parameter
  - Ex: Difference in probability of developing MI by year 5 among subjects with identical age, sex, and lipids who started ART with vs. without Abacavir, standardized to the age, sex and lipid distribution of the whole population
- Quality of the estimate depends on
  - Whether statistical model contains the truth
  - Sample size/ data support for the estimand
  - Estimator

Counterfactual Interpretation

- Change in (some aspect of) the outcome distribution under hypothetical modification to conditions under which data were generated
  - Ex. Difference in counterfactual probability of MI by year 5 under under hypothetical intervention in which whole population started an ART regimen with abacavir versus if no one did
- Moving from statistical to counterfactual interpretation requires that untestable identifiability assumptions hold
  - Ex. Under the assumption that age, sex, and lipids satisfy the backdoor criteria (ie are sufficient to adjust for confounding)

Real World Interpretation

- What would we see if an intervention were implemented in the real world
- Moving from counterfactual to real world interpretation requires
  - Same intervention
    - Relaxing this: "Treatment variation irrelevance" (vanderWeele)
  - Same population
    - Relaxing this: Transportability (vanderWeele, Pearl)
- Ex: Same use of other drugs in the regimen, how the assignment occurs (ie. via a policy vs. patient/ provider preference) doesn't change the effect…

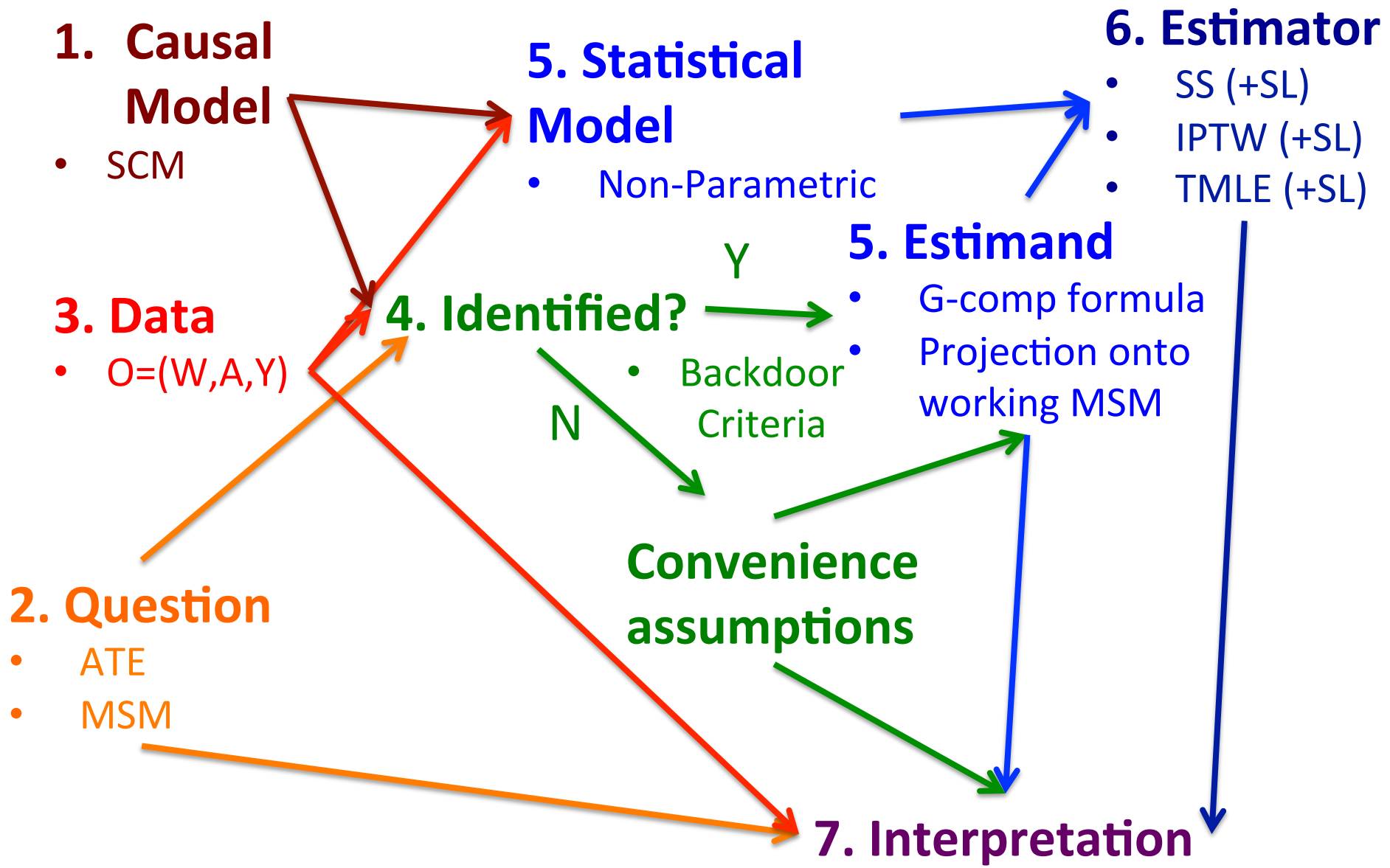
RCT Interpretation

- What would we see if an intervention were evaluated in a randomized trial
  - Ex: Subjects starting ART were randomly assigned to regimen with versus without abacavir
- Moving from real world to RCT interpretation requires
  - Effective randomization
  - Perfect compliance
  - Perfect follow up

Interpreting Results: Take Home Points

- Always have a statistical interpretation
  - If your statistical model contains the truth, you have enough support in your data, and you choose a good estimator
- How far to go beyond this is up to you/reader/policy maker
- Should be based on a frank evaluation of the plausibility of the assumptions required

What have we accomplished?



This is just the beginning…

1. Specify a Casual Model

- We have focused on SCM of Pearl
- Other formal Casual frameworks
  - "Neyman-Rubin" Potential Outcome
  - Dawid: Decision Theoretic
  - Robins: FRCISTG
  - Robins & Richardson: Minimal Causal Model
  - Etc…
  - Differ in extent and type of non testable assumptions, assumptions about the nature of causality, etc…

2. Specify Causal Question

- We have focused on using counterfactuals to define
  1. "Point treatment effects": Static intervention on a single variable
  2. ATE and parameters defined using a (working) MSM
- LOTS more options
  - Interventions on multiple nodes
  - Dynamic (ie personalized or adaptive) interventions
  - Mediation
  - Etc…
  - Review: Petersen & van der Laan *Epidemiology* 2014

3. Specify observed data and its link to the casual model

- We have focused on independent random sample of $O=(W,A,Y)\sim P_0$
- Lots of more complex data structures and links
  - Longitudinal data, Missingness
  - Case control sampling
  - "Adaptive randomization"
  - Etc…

4. Identify

- We have focused on the Back-door criteria/ Randomization assumption
- Many more identifiability results
  - Ex. Front door criteria, Instrumental variables
  - Ex. Sequential back door criteria for multiple intervention nodes
  - Etc.
- Causal frameworks provide a tool for developing these-> new statistical estimand that under specific assumptions give us a wished for causal quantity

5. Commit to a Statistical Model and Estimand (Target parameter of the observed data distribution)

- We have focused on a non-parametric statistical model for P_0
- If your have real model knowledge, by all means use it
  - Straightforward to incorporate in SCM
  - Ex: You know something how the exposure was assigned
- Statistical model should contain the truth

6. Estimate

- We have focused on three estimators
  - Simple (or non-targeted) substitution estimator
  - Inverse probability of treatment weighted estimator
  - TMLE
  - Inference based on NP- bootstrap or IC
- Each of these requires doing a good job estimating some part of the observed data distribution well
  - $E_0(Y|A,W)$, $g_0(A|W)$, or both
  - We focused on data adaptive methods (and in particular Super Learning) to help ensure this
- Other estimators for same quantity exist
  - Ex. Propensity score matching, using the estimated propensity score as a dimension reduction…

7. Interpret.

- Our perspective: A target causal parameter need not correspond to feasible randomized experiment, or hypothetical intervention in order to be of interest
- There is lots of debate on this topic! Decide for yourself….
  - See Petersen & van der Laan *Epidemiology* 2014 for a brief review and some key references to get started

Formal Causal Frameworks provide a very general toolbox to…

1. Represent background knowledge and uncertainty more accurately
2. Frame sharper questions
3. Evaluate/improve plausibility of assumptions
4. Optimize analysis to give best possible answer to motivating question
5. More accurately evaluate uncertainty/make better inferences

Use your tools well!