“Causal Inference and the Role of Machine Learning”

David Benkeser

Assistant Professor

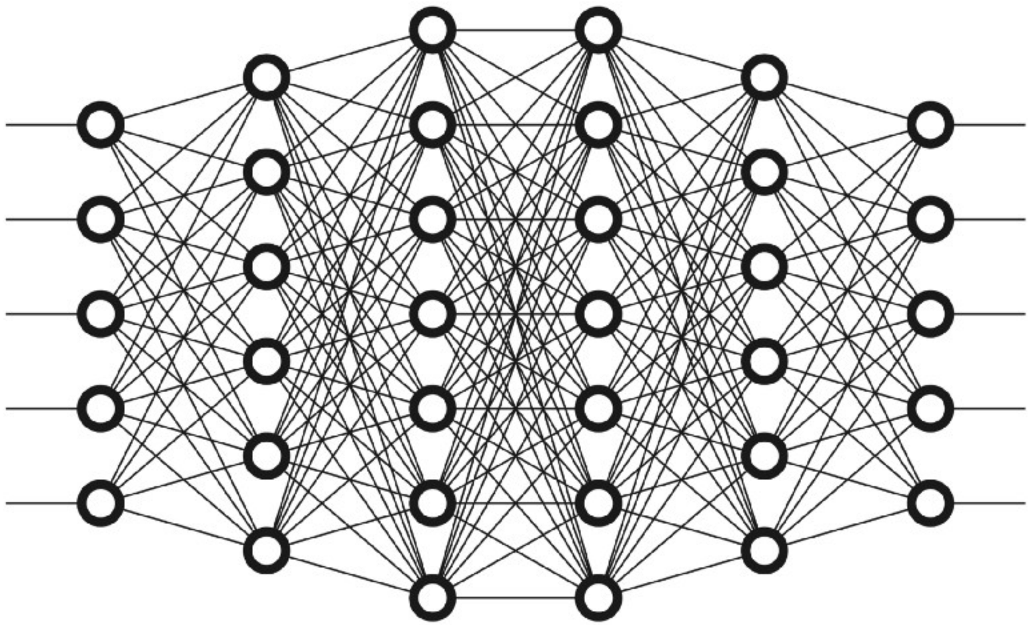
Department of Biostatistics and Bioinformatics

Rollins School of Public Health

Emory University

Disclosure

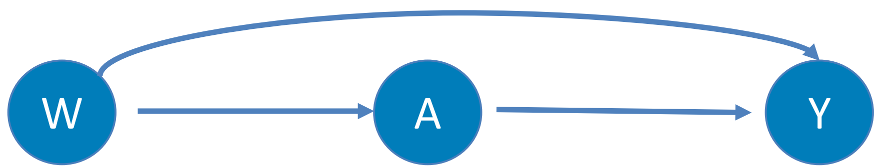
Dr. Benkeser discloses that this image was not used in his talk.



EMORY | ROLLINS SCHOOL OF PUBLIC HEALTH

Basic causal inference

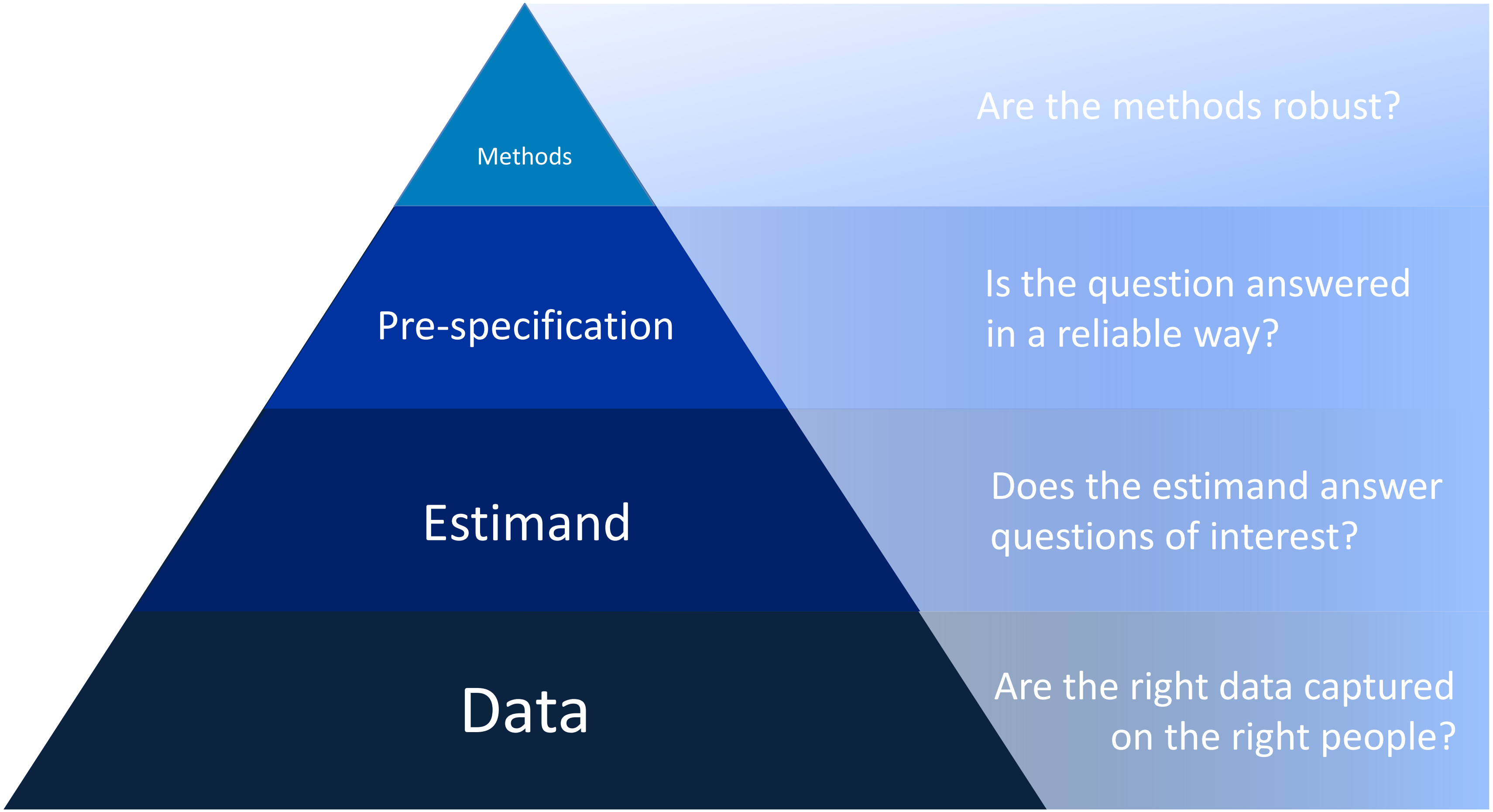
Observed data:
W = putative confounders
A = binary treatment
Y = binary outcome



Counterfactual outcomes:
Y(1) = mean we would see under treatment A = 1
Y(0) = mean we would see under treatment A = 0

Population-level causal effects:
Average treatment effect: E[Y(1) – Y(0)]

Pyramid of Good Science (patent pending)



EMORY | ROLLINS SCHOOL OF PUBLIC HEALTH

Randomized vs. observational studies

Why are randomized trials the gold standard of evidence?

- Data are collected specifically to answer question
- Simple estimands* for drawing causal inference
- Regulatory oversight of analysis
- Simple methods for drawing causal inference

How can observational/secondary data analysis have similar rigor?

- Scrutinize data source and assumptions
- Scrutinize estimands
- Pre-specify analyses
- Use robust methodology

*assuming no missing data, perfect compliance, etc…

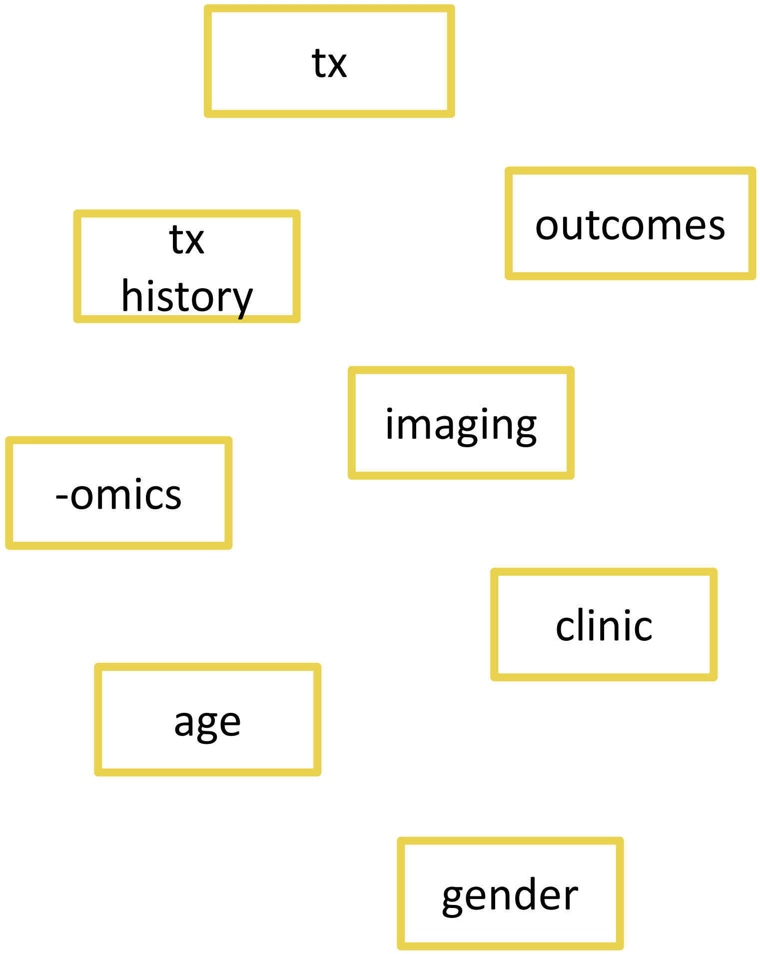
Roadmap for causal inference

1. Specify a causal model representing scientific knowledge.
2. Specify observed data and link to causal model.
3. Specify causal question and causal parameter.
4. Assess identifiability of quantity of interest.
5. Specify a statistical parameter and statistical model.
6. Estimate statistical parameter with robust, pre-specified approach.
7. Interpret results.

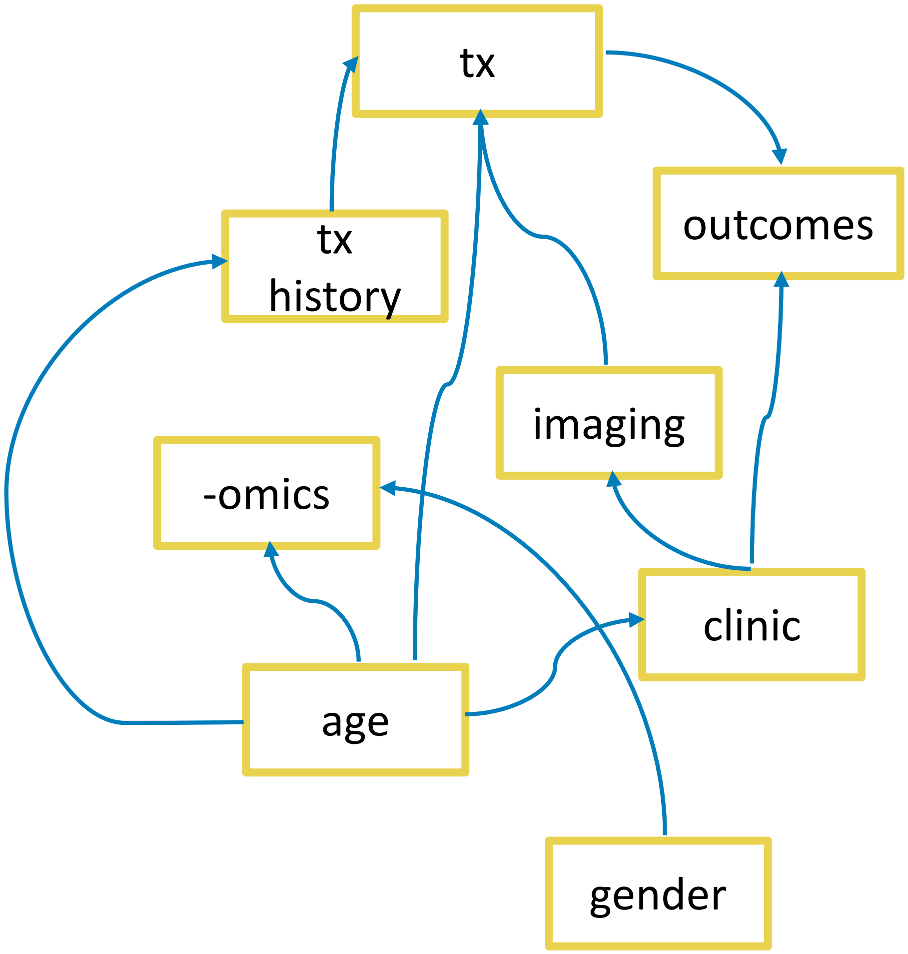
Roadmap for causal inference

1. **Specify a causal model representing scientific knowledge.**
2. **Specify observed data and link to causal model.**
3. Specify causal question and causal parameter.
4. Assess identifiability of quantity of interest.
5. Specify a statistical parameter and statistical model.
6. Estimate statistical parameter with robust, pre-specified approach.
7. Interpret results.

Causal models

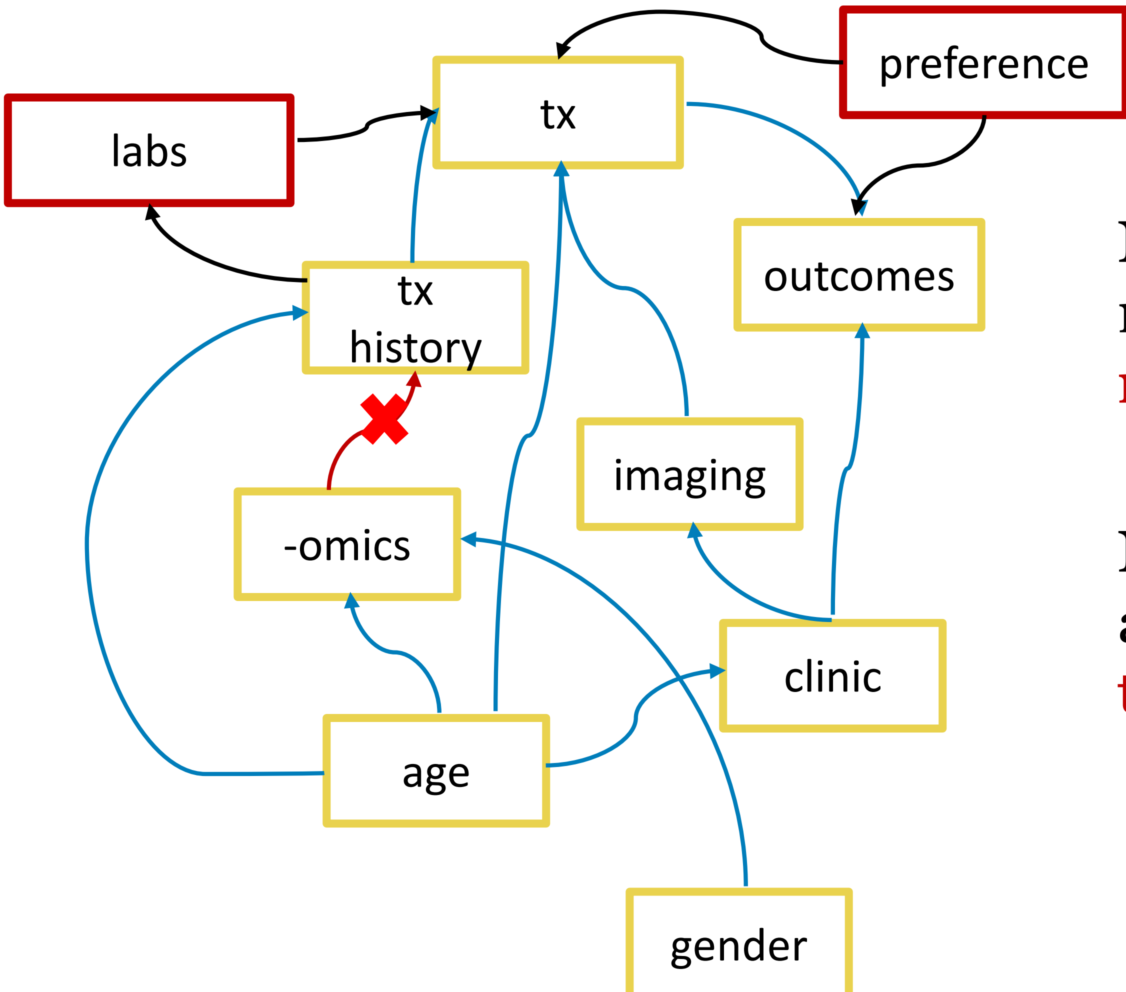


Causal models



Causal models encode what we know about our experiment.

Causal models



Not just what is measured, but **what is not measured**.

Not just arrows that are there, but **arrows that are not there**.

Causal models

Causal models should make us uncomfortably aware of how little we know and/or how little we measured.

Roadmap for causal inference

1. Specify a causal model representing scientific knowledge.
2. Specify observed data and link to causal model.
3. **Specify causal question and causal parameter.**
4. Assess identifiability of quantity of interest.
5. Specify a statistical parameter and statistical model.
6. Estimate statistical parameter with robust, pre-specified approach.
7. Interpret results.

Causal interventions

How would be modify the experiment in an ideal world?

Examples:

- Everyone (no one) gets this seasons flu vaccine
  - E.g., average treatment effect
- Everyone not contraindicated gets flu vaccine
  - E.g., treatment rule
- Everyone gets slightly higher odds of receiving vaccine
  - E.g., incremental propensity score intervention
- Everyone gets a slightly higher dose of drug than they otherwise would receive
  - E.g., stochastic interventions

Roadmap for causal inference

1. Specify a causal model representing scientific knowledge.
2. Specify observed data and link to causal model.
3. Specify causal question and causal parameter.
4. **Assess identifiability of quantity of interest.**
5. Specify a statistical parameter and statistical model.
6. Estimate statistical parameter with robust, pre-specified approach.
7. Interpret results.

Identifiability

What do we need to assume to estimate causal parameters using observed data?

Average treatment effect: $E[Y(1) - Y(0)]$

CF data assumptions: Conditional exchangeability

$$(Y(1), Y(0)) \perp A \mid W$$

Observed data assumptions: Positivity

$$0 < P(A = 1 \mid W) < 1$$

Roadmap for causal inference

1. Specify a causal model representing scientific knowledge.
2. Specify observed data and link to causal model.
3. Specify causal question and causal parameter.
4. Assess identifiability of quantity of interest.
5. **Specify a statistical parameter and statistical model.**
6. Estimate statistical parameter with robust, pre-specified approach.
7. Interpret results.

Statistical parameter

<u>IF</u> we can use the observed data to draw causal inference, how? What do we estimate?

$E[Y(1) - Y(0)] =$

$$\sum_{w} (\underbrace{E[Y \mid A = 1, W=w] - E[Y \mid A = 0, W = w]}_{\text{subgroup-specific effect}})\ \underbrace{P(W= w)}_{\text{subgroup size}}$$

EMORY | ROLLINS SCHOOL OF PUBLIC HEALTH

Roadmap for causal inference

1. Specify a causal model representing scientific knowledge.
2. Specify observed data and link to causal model.
3. Specify causal question and causal parameter.
4. Assess identifiability of quantity of interest.
5. Specify a statistical parameter and statistical model.
6. **Estimate statistical parameter with robust, pre-specified approach.**
7. Interpret results.

Estimation

Identification expresses our counterfactual quantities of interest as quantities defined using the observed data.

- This required causal assumptions; cannot be relaxed for free.

This is certainly progress, we can answer our scientific questions of interest using the observed data!

Estimation

Practitioners make statistical assumptions of varying degrees to tackle the resulting estimation/inference problem.

- Most of these are verifiable and **thus unnecessary** (except for convenience).

$$E[Y(1) - Y(0)] =$$

$$\sum_{w} \big(\underbrace{E[Y \mid A = 1, W=w]} - \underbrace{E[Y \mid A = 0, W = w]}\big)\, P(W= w)$$

Linear/logistic regression?

Estimation

We can instead use **modern statistical learning** to reduce the risk of misleading conclusions due to inappropriate statistical assumptions.

$$E[Y(1) - Y(0)] \;=$$

$$\sum_{w} \Big(\underbrace{E[Y \mid A = 1, W=w]}_{} - \underbrace{E[Y \mid A = 0, W = w]}_{}\Big)\, P(W= w)$$

Machine learning?

Machine learning

Other identifications lead to alternative estimation strategies.

- E.g., inverse probability of treatment weighting, propensity score matching

Many estimators are built on an estimate of

- "Outcome regression" = E[Y | A, W]
  - G-computation, standardization
- "Propensity score" = P(A | W)
  - IPTW, matching
- Or both (doubly robust)

Machine learning

How can we decide ***a priori*** how to estimate these quantities?

Many choices available for fitting regressions:
- parametric regression (logistic, linear, splines)
- additive models, partially linear additive models
- machine learning, deep learning 😍

The best approach depends on the truth! What to do?

Regression stacking (super learning) particularly appealing.
- Pre-specified, objective competition between regression estimators
- All types of estimators can be included
- Oracle inequalities endow some optimality to procedure

Machine learning

Regardless, of your learning approach, there is a challenge associated with using machine learning to draw statistical inference:

How should we select tuning parameters?

Machine learning

Cross-validation selects tuning parameters that are best for estimating e.g., E[Y | A, W].

- But we need a good estimate of E(E[Y | A = a, W])

This generally results in estimators with **too much bias**.

- Undersmooth? Difficult in practice.

Doubly-robust methods generally allow for tuning parameter selection using cross-validation.

- Can be used in fully pre-specified analyses!

Current research topics

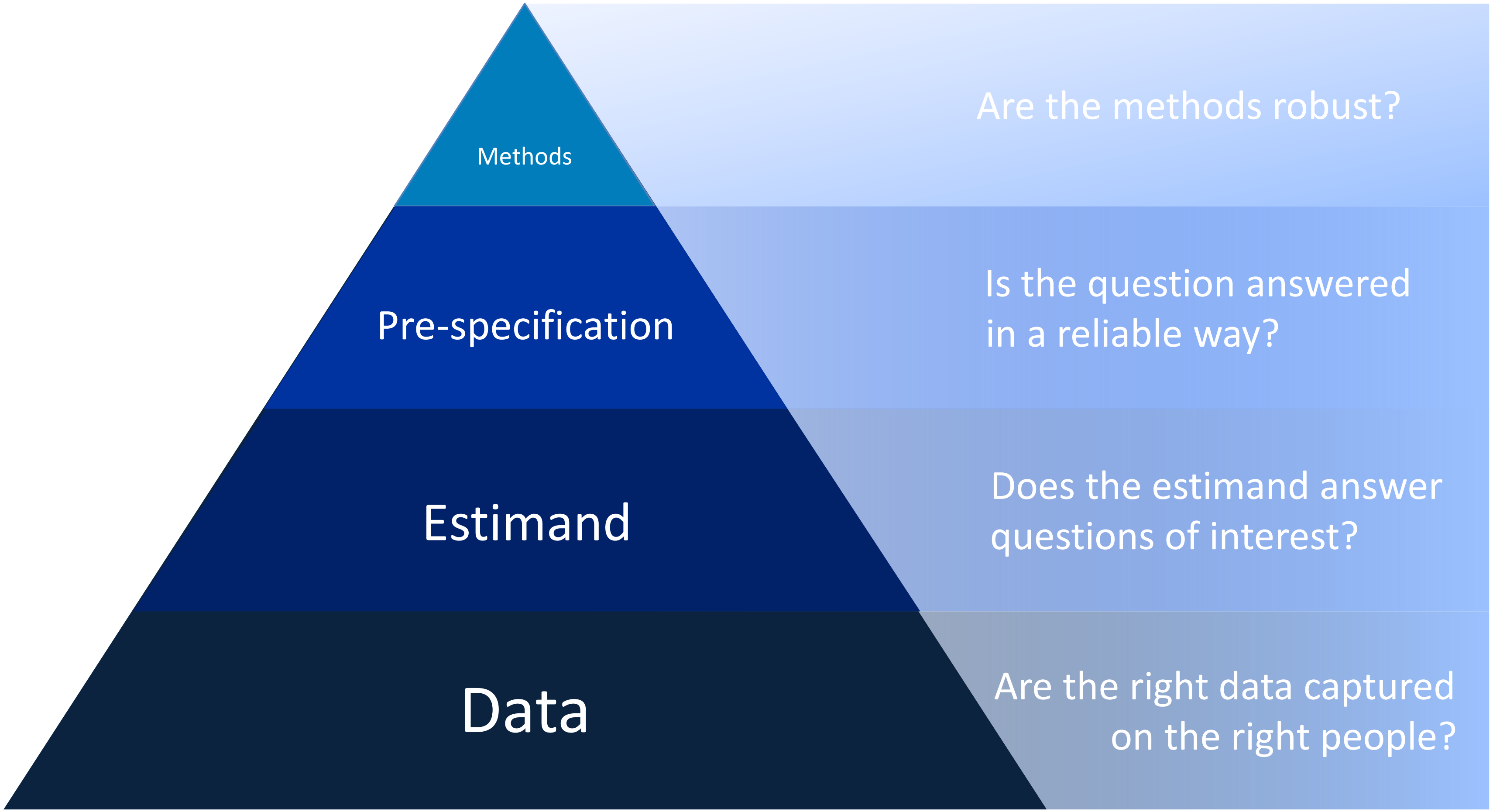
Doubly robust inference
- van der Laan (2014) Int. J. Biost.; Benkeser et al (2017) Biometrika

"Double machine learning"
- Zheng et al (2011) Targeted Learning; Chernozhukov (2017) Econom. J.

Making bootstrap "work" for ML estimators
- Wager and Athey (2015); Cai et al (2019+) arxiv: 1905.10299

Counterfactual fairness in ML
- Kusner et al (2017) NIPS Proc.

Pyramid of Good Science (patent pending)



EMORY | ROLLINS SCHOOL OF PUBLIC HEALTH

Pyramid of Good Science (patent pending)