Causality

Mar 21, 2023
6.7930[6.871]/HST.956: Machine Learning for Healthcare

Goals for Today: Causality

1. Introduction: real-world evidence, challenges of inference and causality
- Observational data are shaped by informed interventions, feedback loops

2. Methods of Causal Inference
- Potential Outcomes Framework
- Typical Underlying Assumptions: No unmeasured confounders, Ignorability, Common support
- Statistical methods of causal inference: Covariate adjustment, Propensity score re-weighting
- Natural Experiments

3. From Association to Causality: Principles of causality in real-world evidence
- Strength: Causal effects often have stronger associations.
- Consistency: Causal effects can be repeatedly observed by different persons, in different places, circumstances, and times?
- Specificity: Causal effects are often condition-specific.
- Temporality: Causal effects respect temporal ordering.
- Biological gradient: Causal effects are influenced by underlying biology which has continuously-varying characteristics.
- Plausibility: Causal effects should be plausible.
- Coherence: Causal effects should not seriously conflict with the generally known mechanisms.
- Experiment: Causal effects can predict experimental evidence.
- Analogy: Causal effects may be transferable between systems.

4. Causality in practice
- Bias and Confounding: Handling sensitive variables, explicitly sampling across groups
- All models are wrong, but some are useful -> Do we always want a biologically causal model?
- Useful models require careful consideration of setting and use cases to include generalizability, interventions, and risks.

Patient/Provider Goals of Clinical Data Science

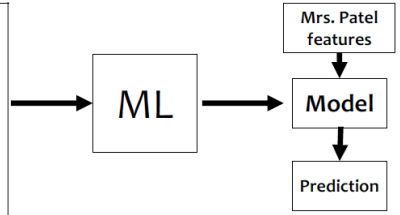
- Mrs. Patel is a 65 year old who was recently diagnosed with kidney cancer. She returns to your office to discuss treatment and has some questions.
  - After treatment, **what is the risk** of my cancer coming back before the Ultimate World Cruise (December 2023)?
  - **Will the risk** of my cancer coming back **change** if I get a partial nephrectomy instead of a radical nephrectomy?

How would you answer these questions using clinical data science?

Will my cancer come back?

- How would you the estimate of Mrs. Patel's risk of cancer recurrence?

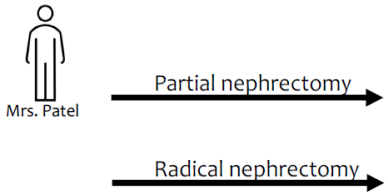
Features X (age, sex, ethnicity, smoking status, body mass index, blood pressure, heart rate, …, nth feature)	Label Y (cancer recurrence)



Aim of prediction is to estimate Y, given X

Change the risk of my cancer coming back?

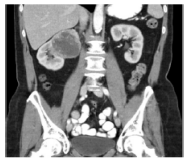
- You hypothesize that type of surgery (partial vs. radical) will change her risk of cancer recurrence.
- Ground truth



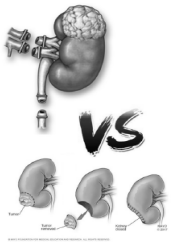
- Reality: We cannot know the ground truth

RCT: Radical vs. Partial Nephrectomy

- EORTC 30904



Population: 541 patients with tumors <5cm suspicious for kidney cancer



Randomized to RN vs. PN

Results

Local recurrence

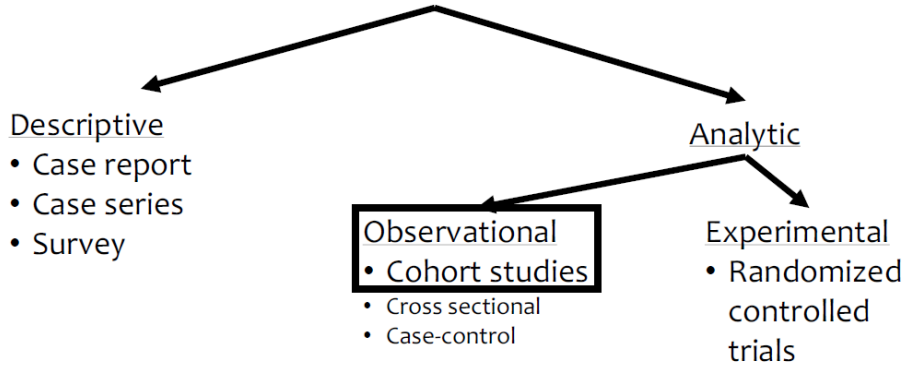
RN 1/273 = 0.37%

PN 6/278 = 2.16%

Van Poppel, Hendrik, et al. "A prospective, randomised EORTC intergroup phase 3 study comparing the oncologic outcome of elective nephron-sparing surgery and radical nephrectomy for low-stage renal cell carcinoma." *European urology* 59.4 (2011): 543-552.
https://www.fairbanksurology.com/robotic-radical-nephrectomy
https://www.mayoclinic.org/tests-procedures/nephrectomy/multimedia/img-20332175

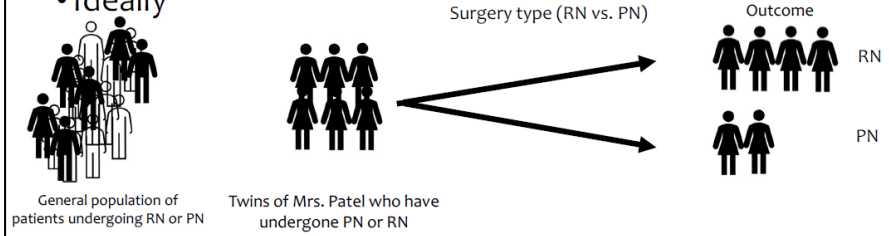
Conclusion from randomized control trial:
On average, radical nephrectomy has a *lower* rate of local recurrence than partial

Clinical Research Study Designs



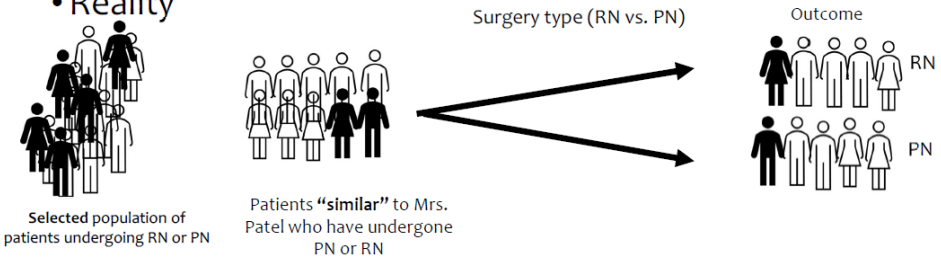
Change the risk of my cancer coming back?

- You hypothesize that type of surgery (partial vs. radical) will change her risk of cancer recurrence. How do you evaluate this hypothesis?
  - Ideally

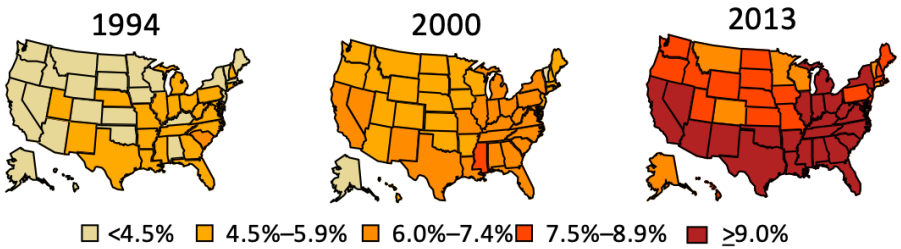


Change the risk of my cancer coming back?

- You hypothesize that type of surgery (partial vs. radical) will change her risk of cancer recurrence. How do you evaluate this hypothesis?
  - Reality

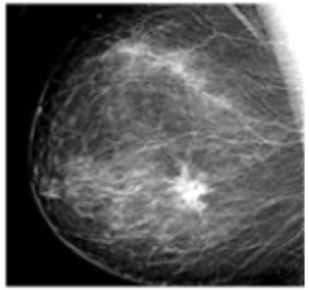


Does gastric bypass surgery prevent onset of diabetes?

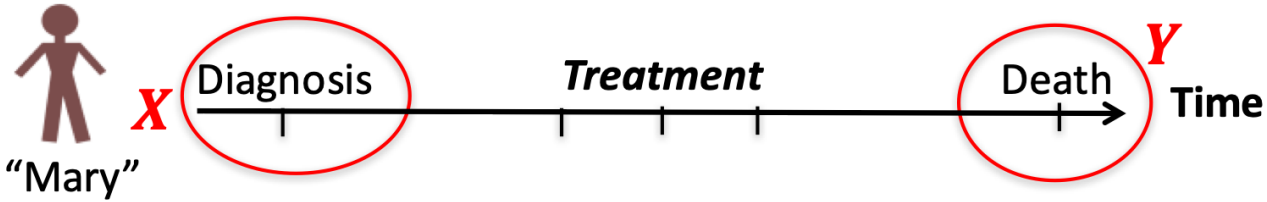


- In Lecture 4 & PS2 we used machine learning for early detection of Type 2 diabetes
- Health system doesn't want to know how to predict diabetes – they want to know how to *prevent it*
- Gastric bypass surgery is the highest negative weight (9th most predictive feature)
  - Does this mean it would be a good intervention?

What is the likelihood this patient, with breast cancer, will survive 5 years?

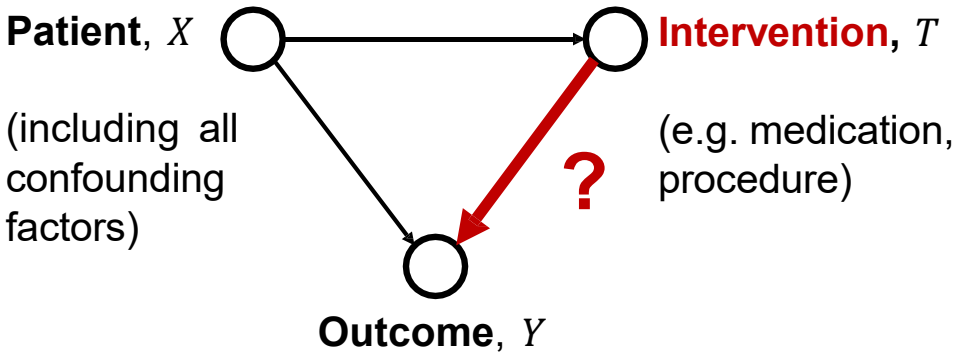


- Such predictive models widely used to stage patients. Should we initiate treatment? How aggressive?
- What could go wrong if we trained to predict survival, and then used to guide patient care?

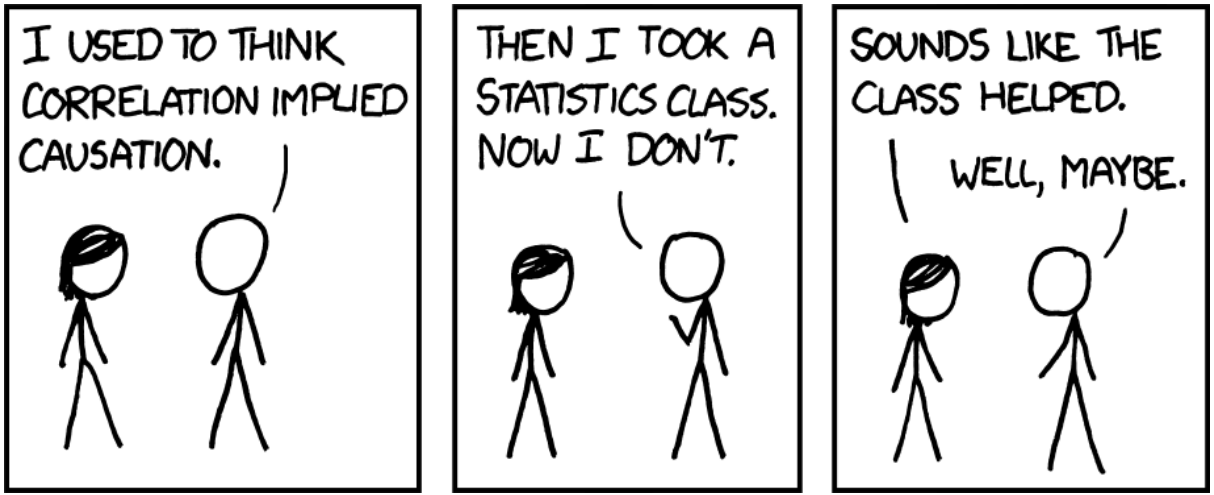


A long survival time may be because of treatment!

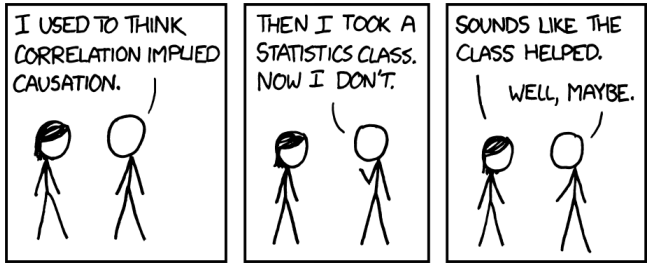
To properly answer, need to formulate as *causal* questions:



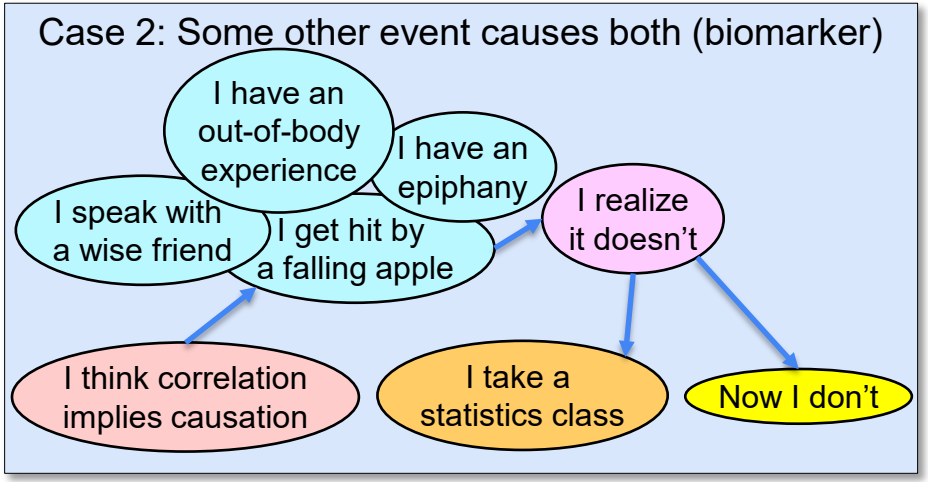
High dimensional *Observational data*


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

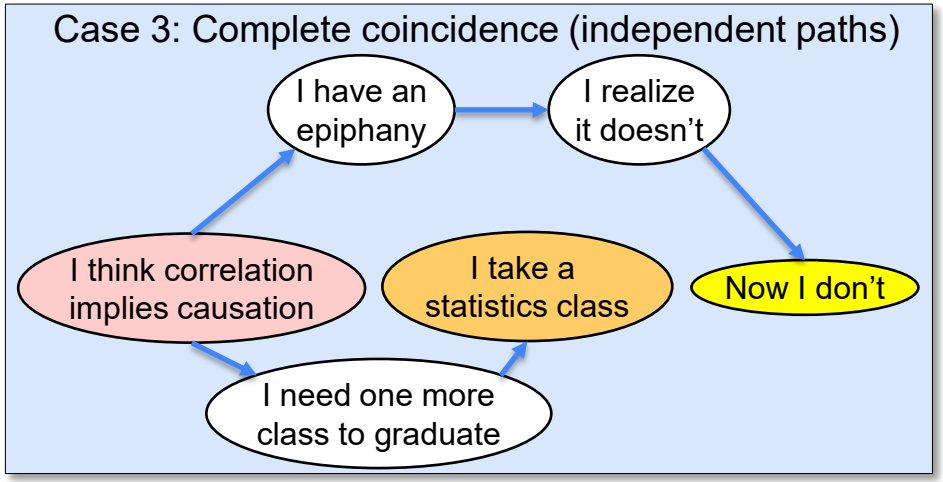
XKCD



XKCD



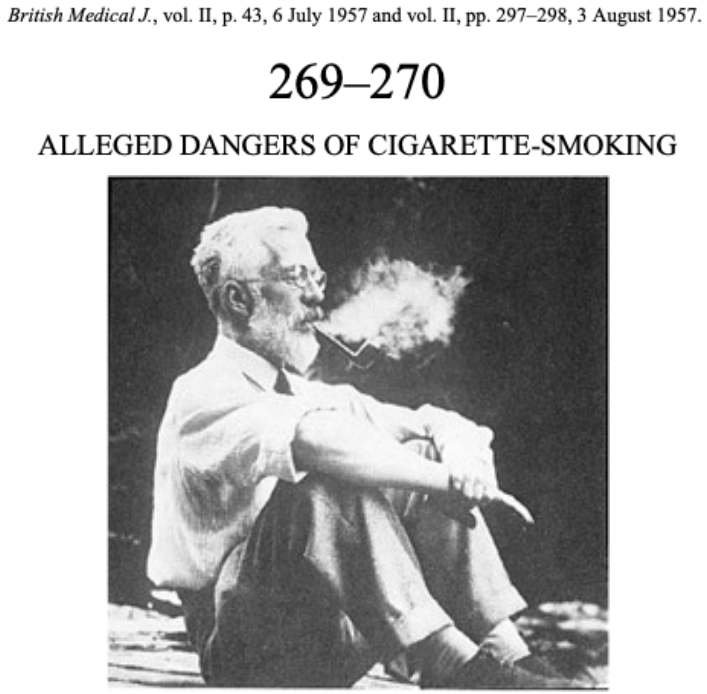
Intuition: not everyone takes a statistics class
Perhaps something pushed me to take one.
Perhaps that same something led to the outcome



Intuition: Sometimes even the correlation is fortuitous
(solution: increase sample size ➔ correlation goes away)

What should we do with associational evidence?

- RA Fisher - famous statistician, rejected smoking ➔ cancer causality
- Claim: Only associational studies have been run so far.
- Monozygotic twins have more similar smoking patterns than dizygotic twins, so maybe a genetic propensity to smoke instead of a causal link?
- How many cancers were caused by this wrong interpretation?



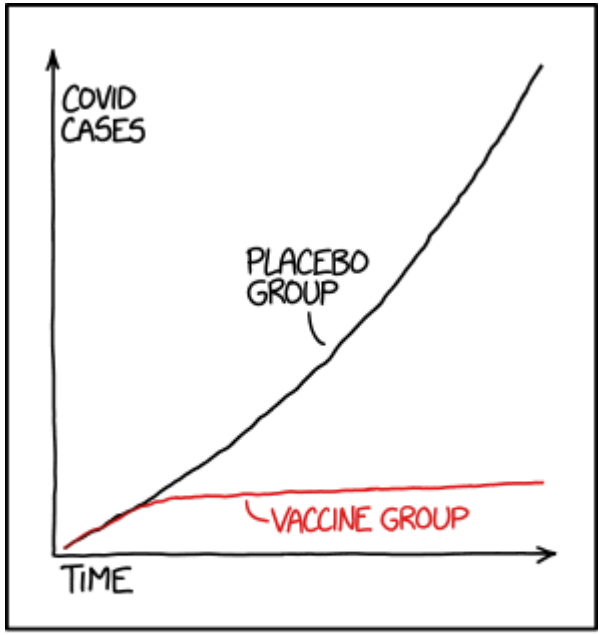
British Medical J., vol. II, p. 43, 6 July 1957 and vol. II, pp. 297–298, 3 August 1957.

269–270

ALLEGED DANGERS OF CIGARETTE-SMOKING

"Alleged benefits of covid vaccination"

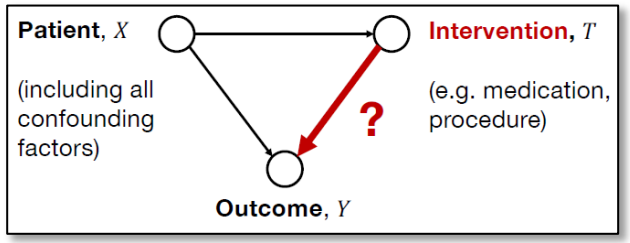
Statistics



STATISTICS TIP: ALWAYS TRY TO GET DATA THAT'S GOOD ENOUGH THAT YOU DON'T NEED TO DO STATISTICS ON IT

We reject the null hypothesis based on the 'hot damn, check out this chart' test

xkcd



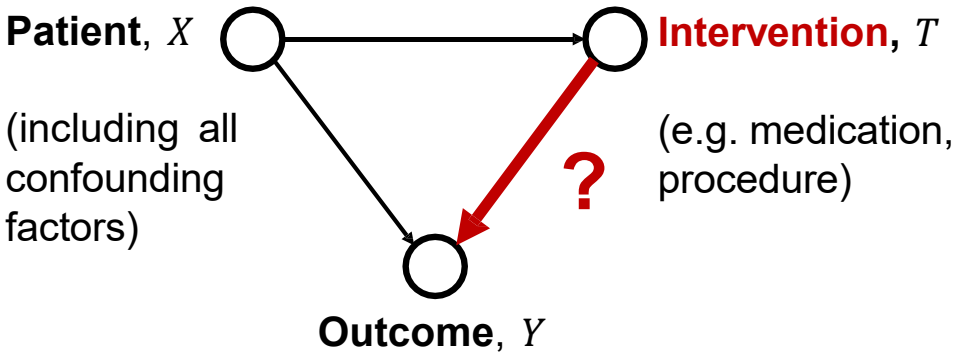
Alleged benefits of mask-wearing to protect against covid spread:

- Yes, there is plausibility
- Yes, there is correlation
- Yes, there are interventional studies

But many confounders:

- Counties who choose to mask also choose other measures
- Individuals who choose to mask also take other precautions
- Can we untangle these effects?

To properly answer, need to formulate as *causal* questions:



High dimensional

Observational data

ATE = Average Treatment Effect
CATE = Conditional Average Treatment Effect

Each unit (individual) x_i has two potential outcomes*:

- $Y_0(x_i)$ is the potential outcome had the unit not been treated: "***control outcome***"
- $Y_1(x_i)$ is the potential outcome had the unit been treated: "***treated outcome***"

Conditional average treatment effect for unit i:

$$CATE(x_i) = \mathbb{E}_{Y_1 \sim p(Y_1|x_i)}[Y_1|x_i] - \mathbb{E}_{Y_0 \sim p(Y_0|x_i)}[Y_0|x_i]$$

Average Treatment Effect:

$$ATE := \mathbb{E}[Y_1 - Y_0] = \mathbb{E}_{x \sim p(x)}[CATE(x)]$$

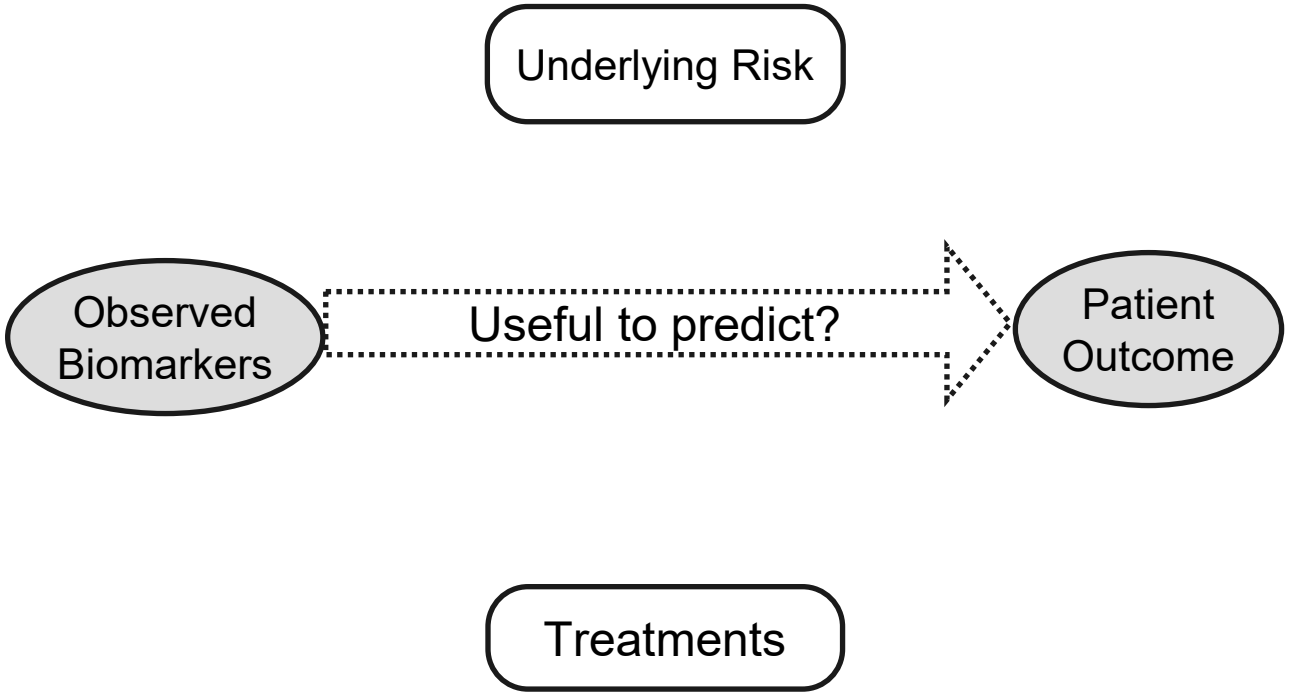
Observed factual outcome:

$$y_i = t_i Y_1(x_i) + (1 - t_i) Y_0(x_i)$$

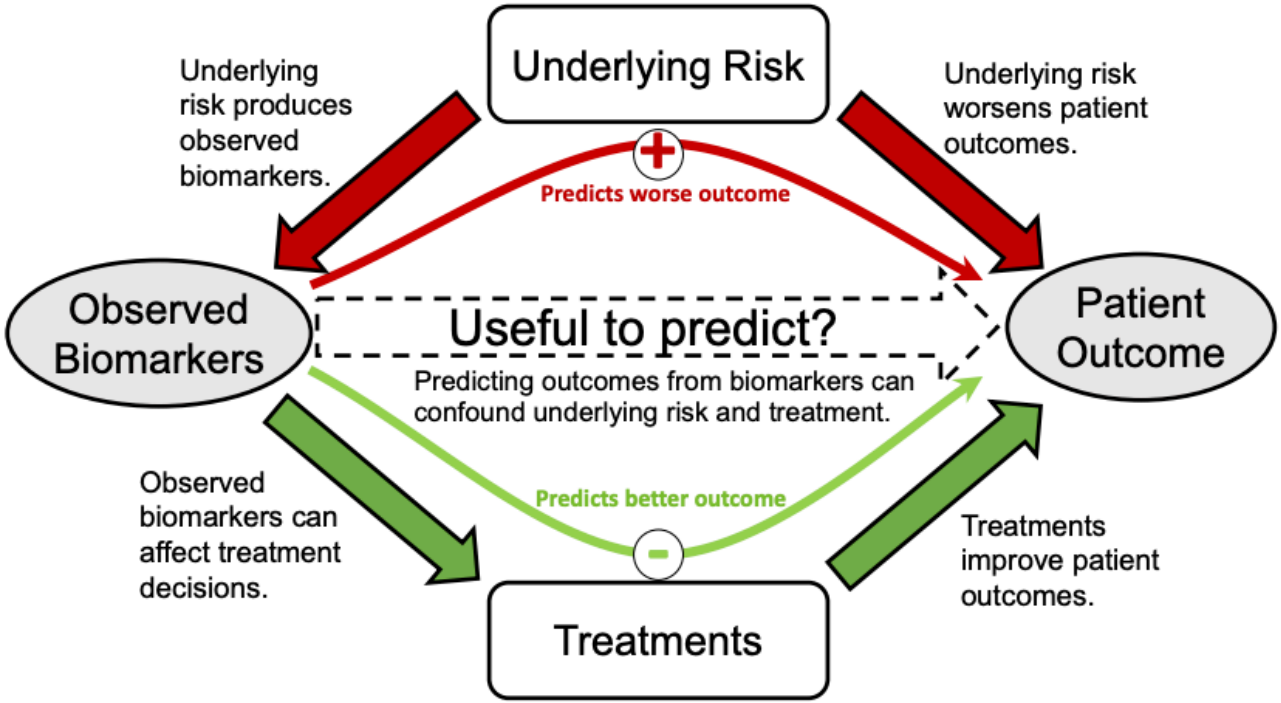
Unobserved counterfactual outcome:

$$y_i^{CF} = (1 - t_i) Y_1(x_i) + t_i Y_0(x_i)$$

Real-world evidence comes from complex human behaviors



Real-world evidence comes from complex human behaviors



Goals for Today: Causality

1. Introduction: real-world evidence, challenges of inference and causality

- Observational data are shaped by informed interventions, feedback loops

2. Methods of Causal Inference

- Potential Outcomes Framework
- Typical Underlying Assumptions: No unmeasured confounders, Ignorability, Common support
- Statistical methods of causal inference: Covariate adjustment, Propensity score re-weighting
- Natural Experiments

3. From Association to Causality: Principles of causality in real-world evidence

- Strength: Causal effects often have stronger associations.
- Consistency: Causal effects can be repeatedly observed by different persons, in different places, circumstances, and times?
- Specificity: Causal effects are often condition-specific.
- Temporality: Causal effects respect temporal ordering.
- Biological gradient: Causal effects are influenced by underlying biology which has continuously-varying characteristics.
- Plausibility: Causal effects should be plausible.
- Coherence: Causal effects should not seriously conflict with the generally known mechanisms.
- Experiment: Causal effects can predict experimental evidence.
- Analogy: Causal effects may be transferable between systems.

4. Causality in practice

- Bias and Confounding: Handling sensitive variables, explicitly sampling across groups
- All models are wrong, but some are useful -> Do we always want a biologically causal model?
- Useful models require careful consideration of setting and use cases to include generalizability, interventions, and risks.

Potential Outcomes Framework (Rubin-Neyman Causal Model)

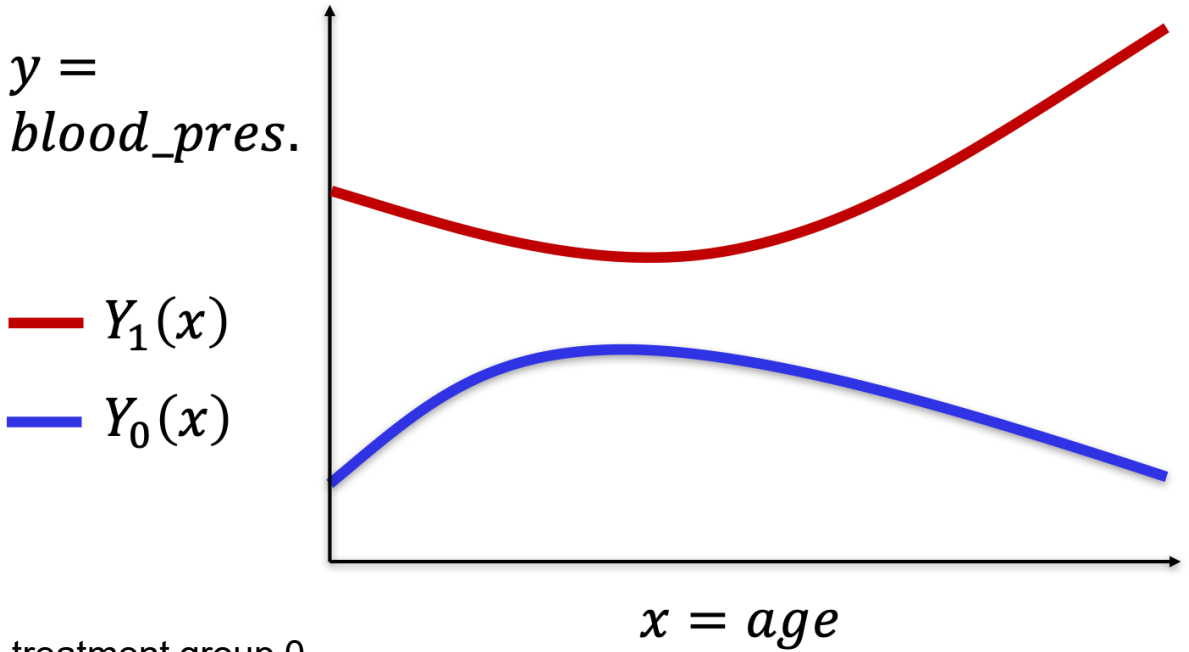
- Each unit (individual) x_i has two potential outcomes:
  - $Y_0(x_i)$ is the potential outcome had the unit not been treated: "***control outcome***"
  - $Y_1(x_i)$ is the potential outcome had the unit been treated: "***treated outcome***"

- Conditional average treatment effect for unit i:
  $$CATE(x_i) = \mathbb{E}_{Y_1 \sim p(Y_1|x_i)}[Y_1|x_i] - \mathbb{E}_{Y_0 \sim p(Y_0|x_i)}[Y_0|x_i]$$
- Average Treatment Effect:
  $$ATE := \mathbb{E}[Y_1 - Y_0] = \mathbb{E}_{x \sim p(x)}[CATE(x)]$$

“The fundamental problem of causal inference”

We only ever observe one of the two outcomes

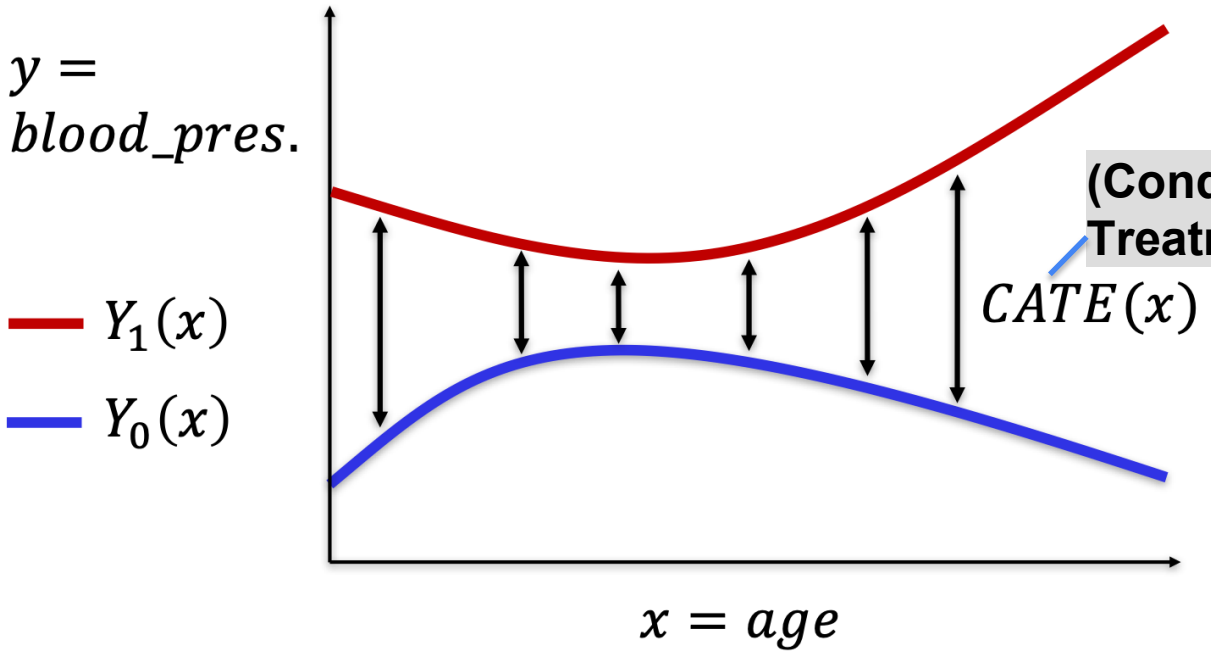
Example – Blood pressure and age



Y_0 = Hypothetical treatment group 0
Y_1 = Hypothetical treatment group 1

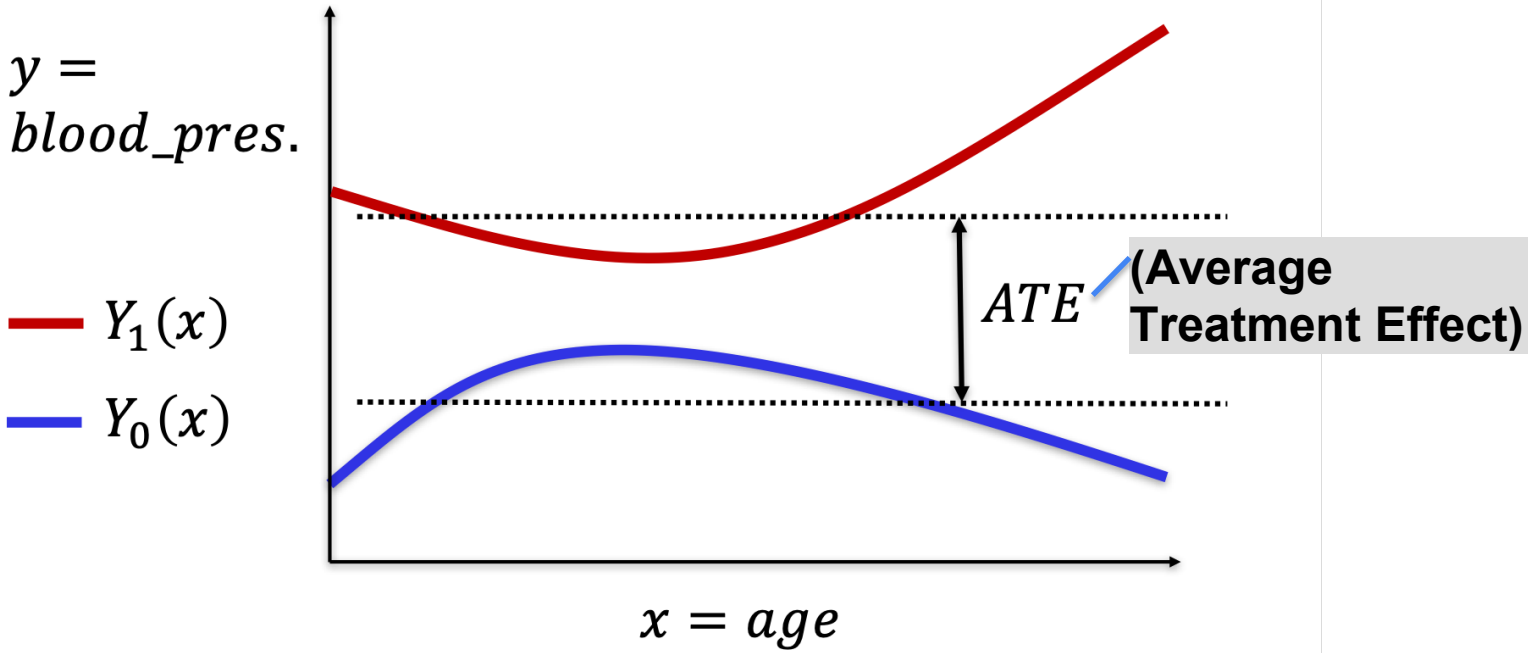
Only information about these individuals
(usually highly multidimentional)

Blood pressure and age

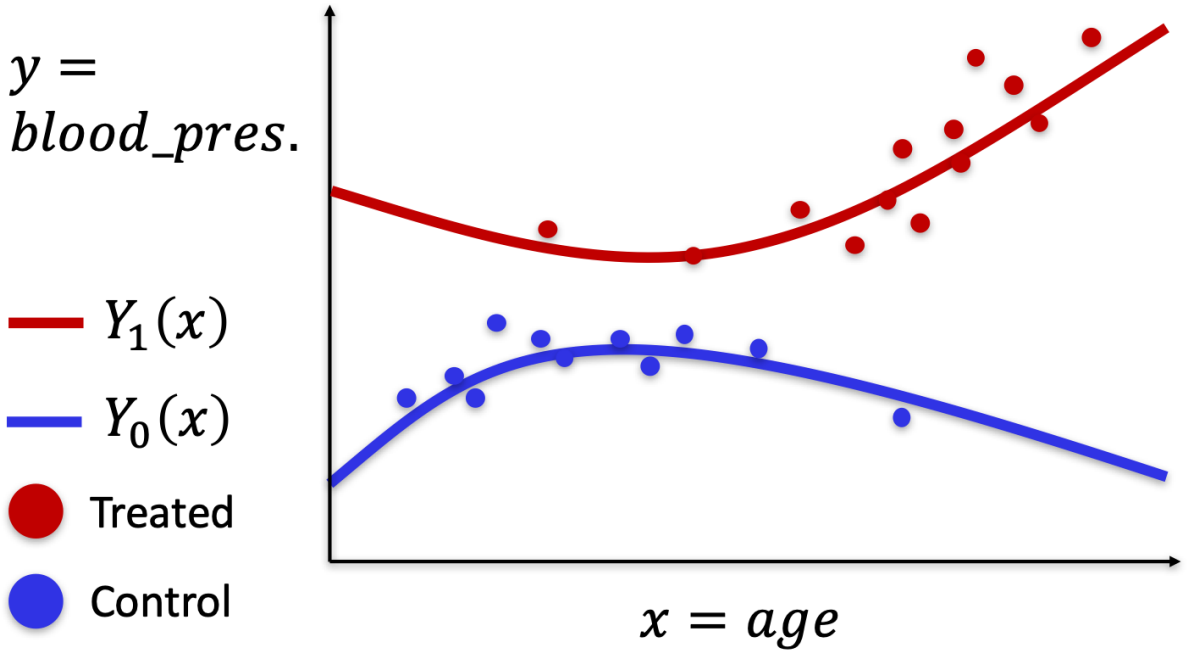


(Conditional Average Treatment Effect)

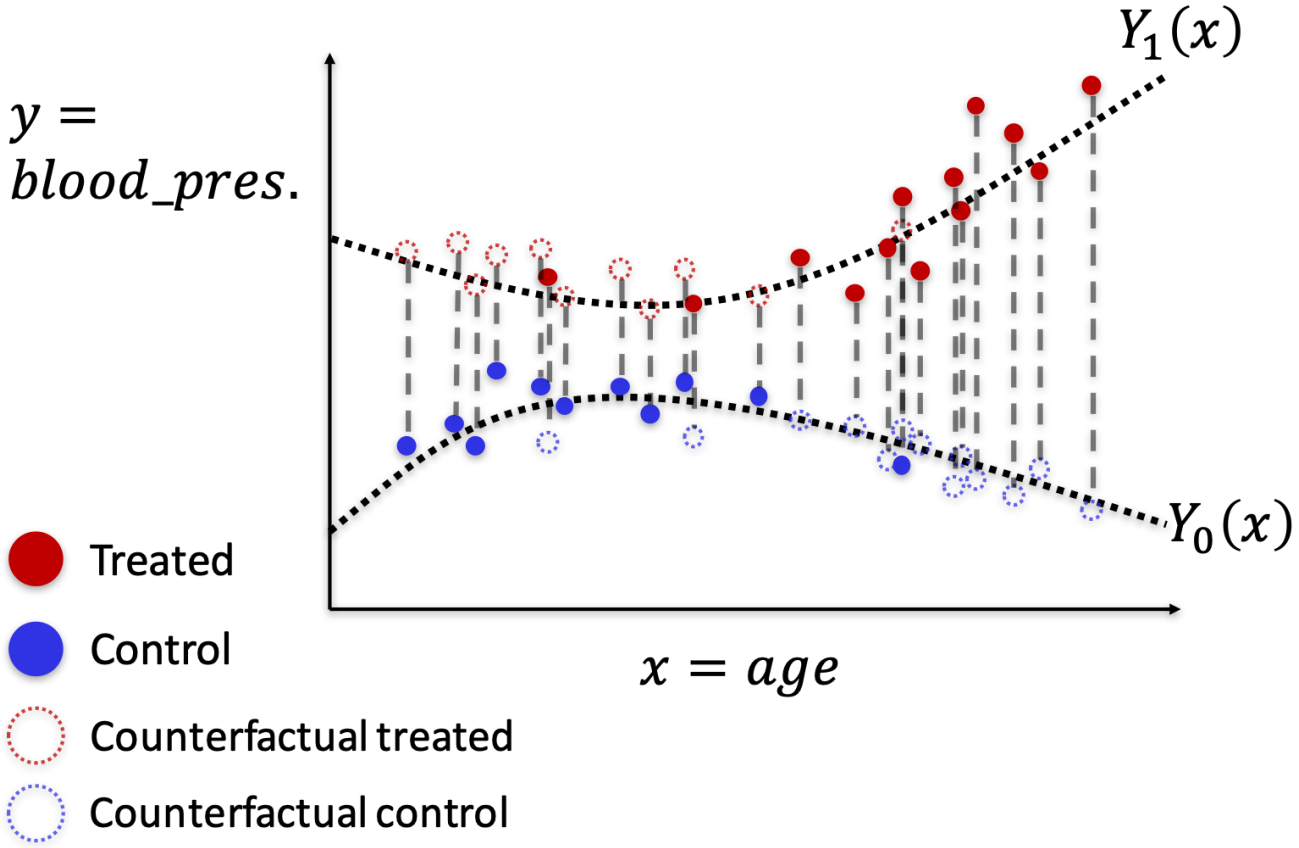
Blood pressure and age



Blood pressure and age



Blood pressure and age



Typical assumption – no unmeasured confounders

Y_0, Y_1: potential outcomes for control and treated

x: unit covariates (features)

T: treatment assignment

We assume:

$$(Y_0, Y_1) \perp\!\!\!\perp T \mid x$$

The potential outcomes are independent of treatment assignment, conditioned on covariates x

Note: We only care about confounders **that lead to treatment decisions**

Slide Adapted from: David Sontag (MIT)

Typical assumption – no unmeasured confounders

Y_0, Y_1: potential outcomes for control and treated

x: unit covariates (features)

T: treatment assignment

We assume:

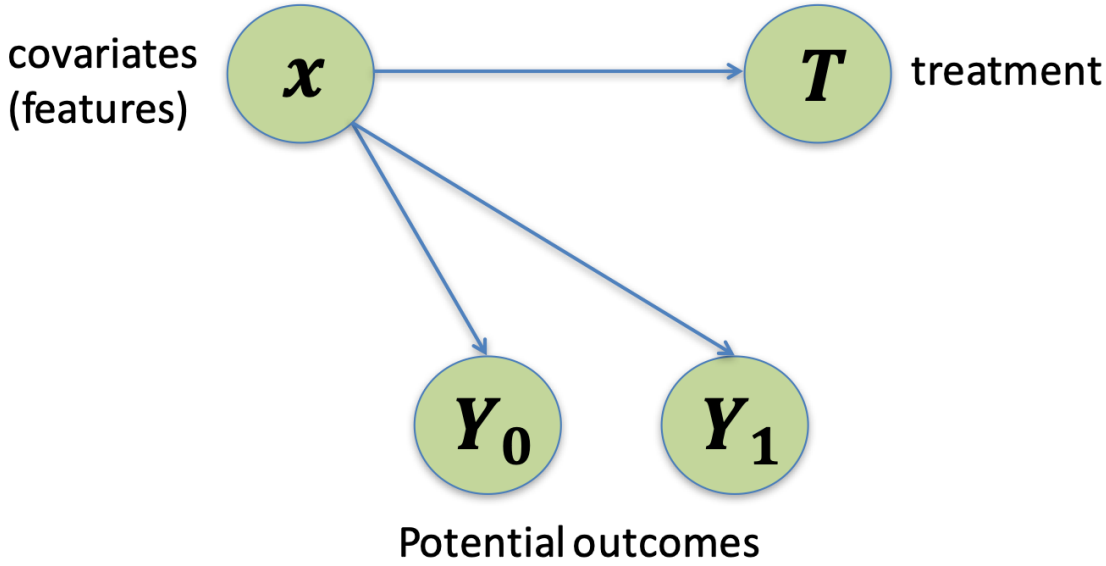
$$(Y_0, Y_1) \perp\!\!\!\perp T \mid x$$

Ignorability

Note: We care about confounders **that lead to treatment decisions*

Slide Adapted from: David Sontag (MIT)

Ignorability

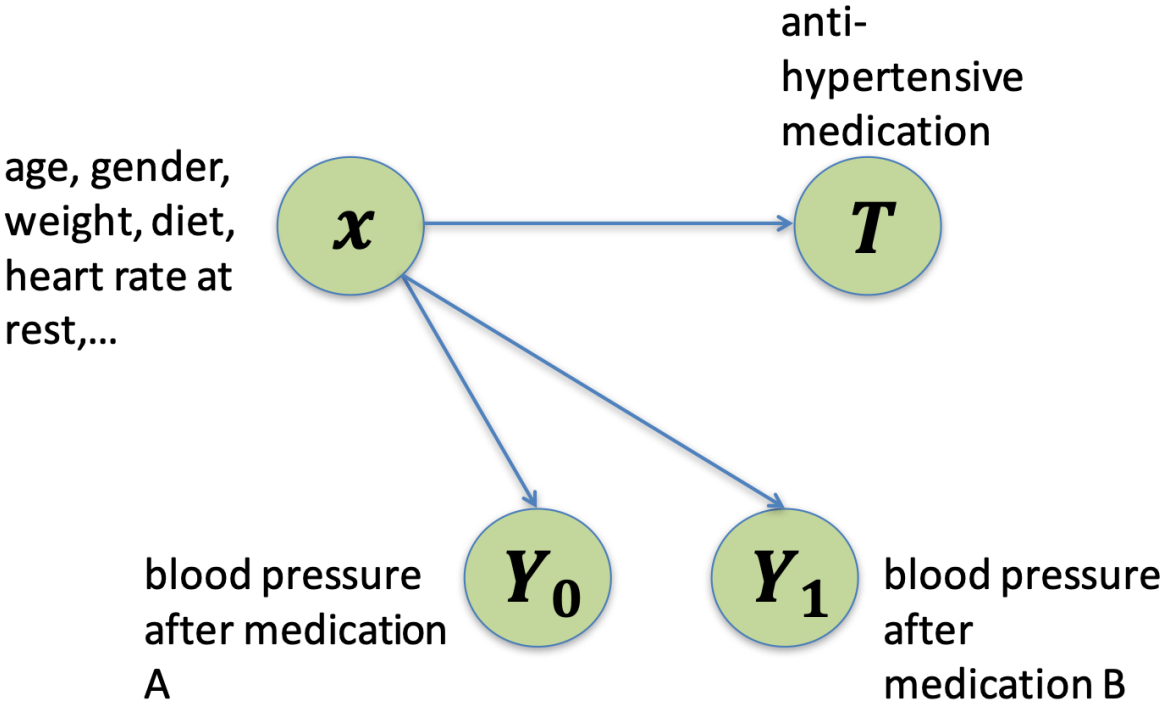


$$(Y_0, Y_1) \perp\!\!\!\perp T \mid x$$

Note: We care about confounders **that lead to treatment decisions*

Slide Adapted from: David Sontag (MIT)

Ignorability

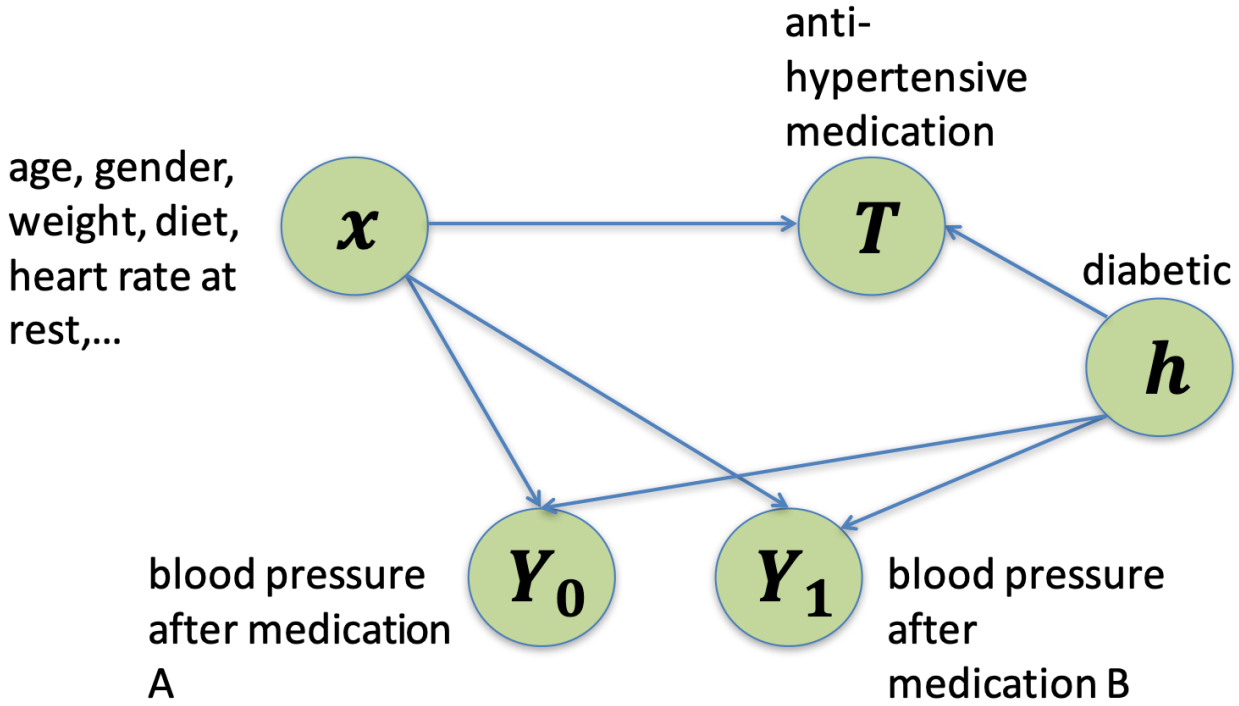


$$(Y_0, Y_1) \perp\!\!\!\perp T \mid x$$

Note: We care about confounders **that lead to treatment decisions*

Slide Adapted from: David Sontag (MIT)

No Ignorability



$$(Y_0, Y_1) \not\perp T \mid x$$

Note: We care about confounders **that lead to treatment decisions**

Slide Adapted from: David Sontag (MIT)

Typical assumption – common support

Y_0, Y_1: potential outcomes for control and treated

x: unit covariates (features)

T: treatment assignment

We assume:

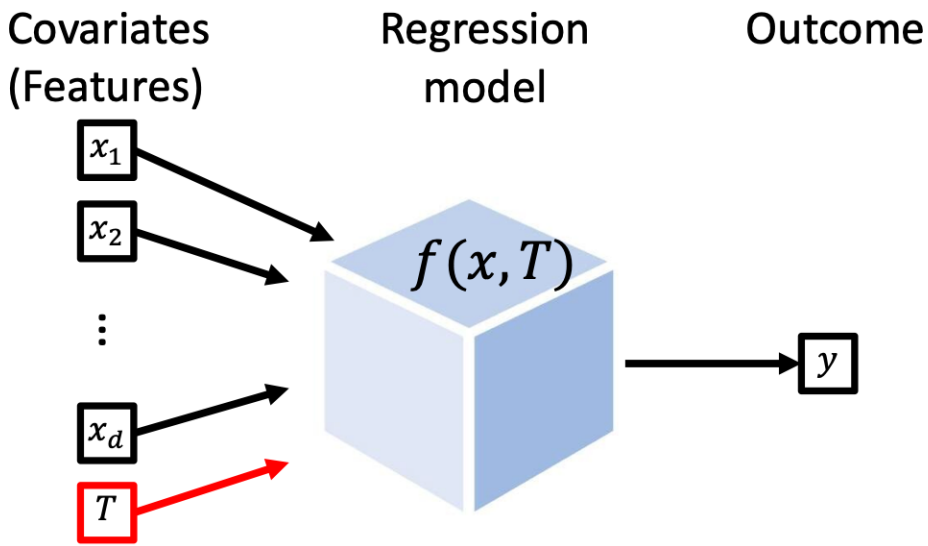
$$p(T = t|X = x) > 0 \; \forall t, x$$

Note: We care about confounders **that lead to treatment decisions**

Slide Adapted from: David Sontag (MIT)

Covariate adjustment

Explicitly model the relationship between treatment, confounders, and outcome:



Slide Adapted from: David Sontag (MIT)

Covariate adjustment (parametric g-formula)

- Explicitly model the relationship between treatment, confounders, and outcome
- Under ignorability, the expected causal effect of T on Y:
  $\mathbb{E}_{x \sim p(x)}\left[\mathbb{E}[Y_1|T=1,x] - \mathbb{E}[Y_0|T=0,x]\right]$
- Fit a model $f(x,t) \approx \mathbb{E}[Y_t|T=t,x]$

$$\widehat{ATE} = \frac{1}{n}\sum_{i=1}^{n} f(x_i,1) - f(x_i,0)$$

Inferred treated *Inferred untreated*

ATE = average treatment effect

Slide Adapted from: David Sontag (MIT)

Covariate adjustment (parametric g-formula)

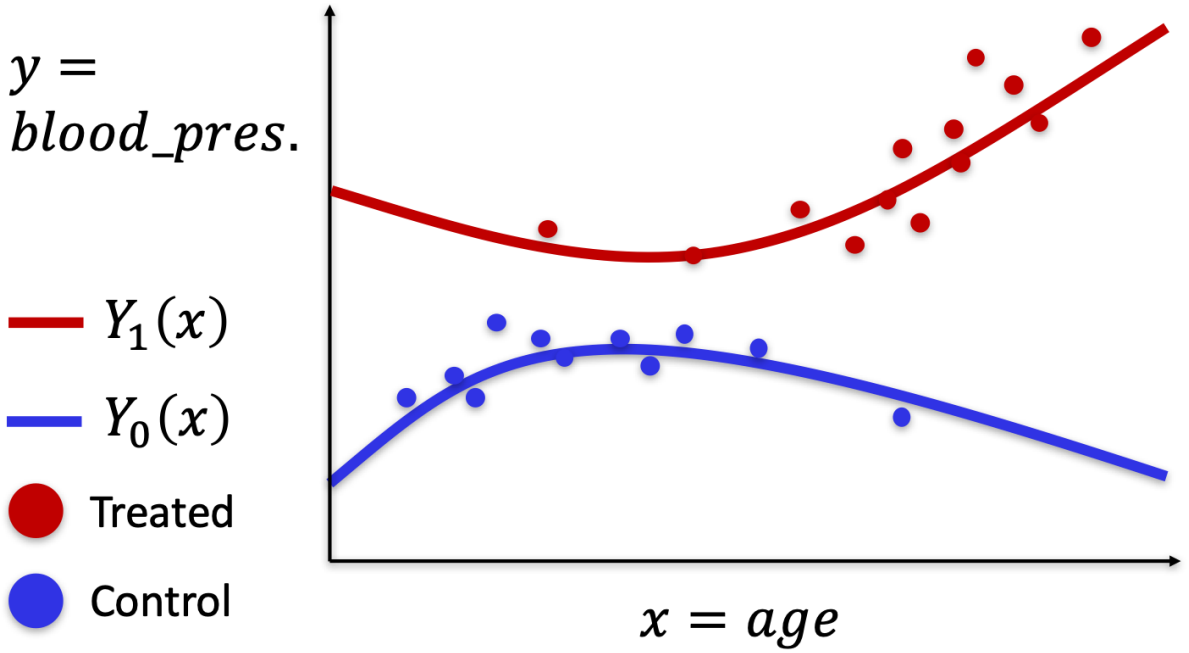
- Explicitly model the relationship between treatment, confounders, and outcome
- Under ignorability, the expected causal effect of T on Y:
  $\mathbb{E}_{x \sim p(x)}\left[\mathbb{E}[Y_1 | T = 1, x] - \mathbb{E}[Y_0 | T = 0, x]\right]$
- Fit a model $f(x, t) \approx \mathbb{E}[Y_t | T = t, x]$

$$\widehat{CATE}(x_i) = f(x_i, 1) - f(x_i, 0)$$

CATE = conditional average treatment effect

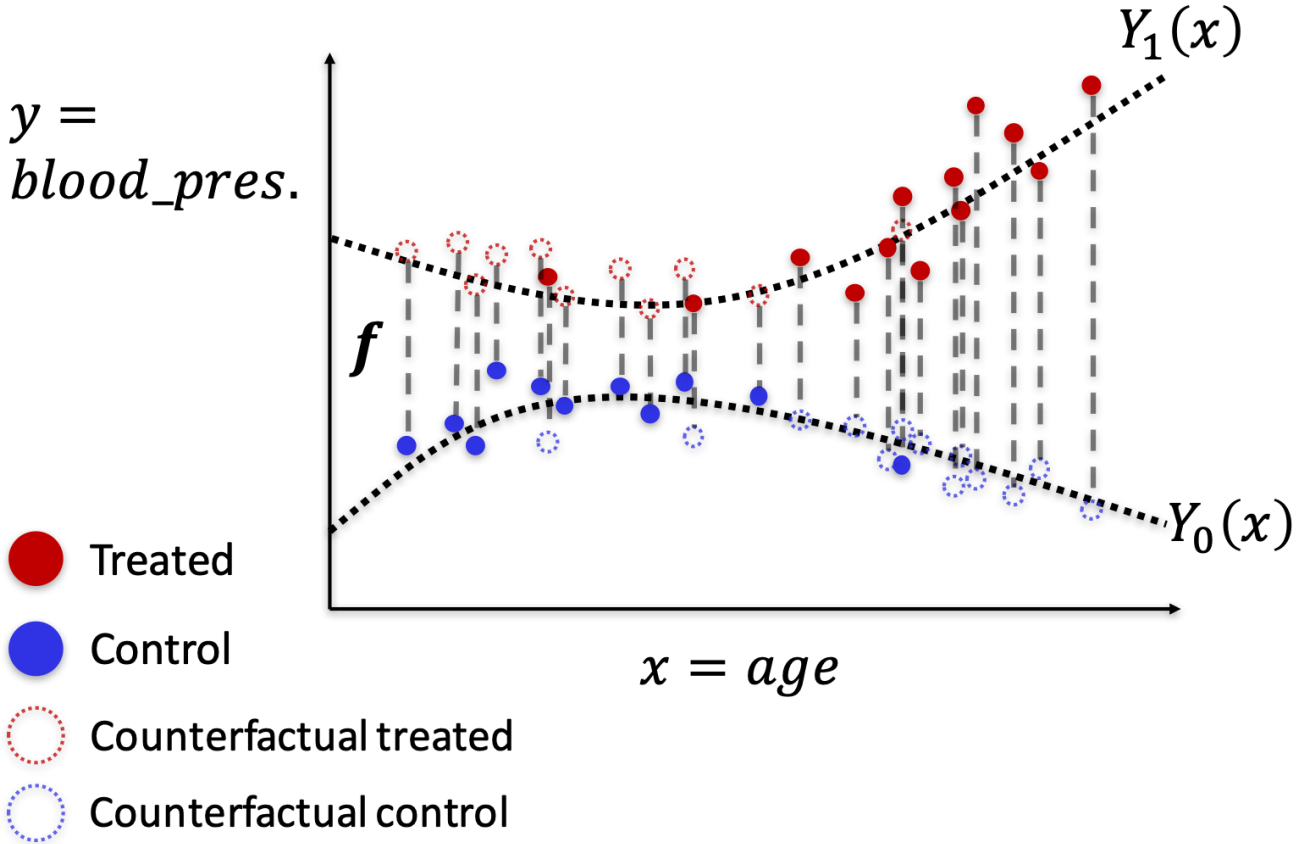
Slide Adapted from: David Sontag (MIT)

Covariate adjustment



Slide Adapted from: David Sontag (MIT)

Covariate adjustment



Slide Adapted from: David Sontag (MIT)

Propensity scores

- Tool for estimating ATE **(Average Treatment Effect)**
- Basic idea: turn observational study into a pseudo-randomized trial by re-weighting samples, similar to importance sampling

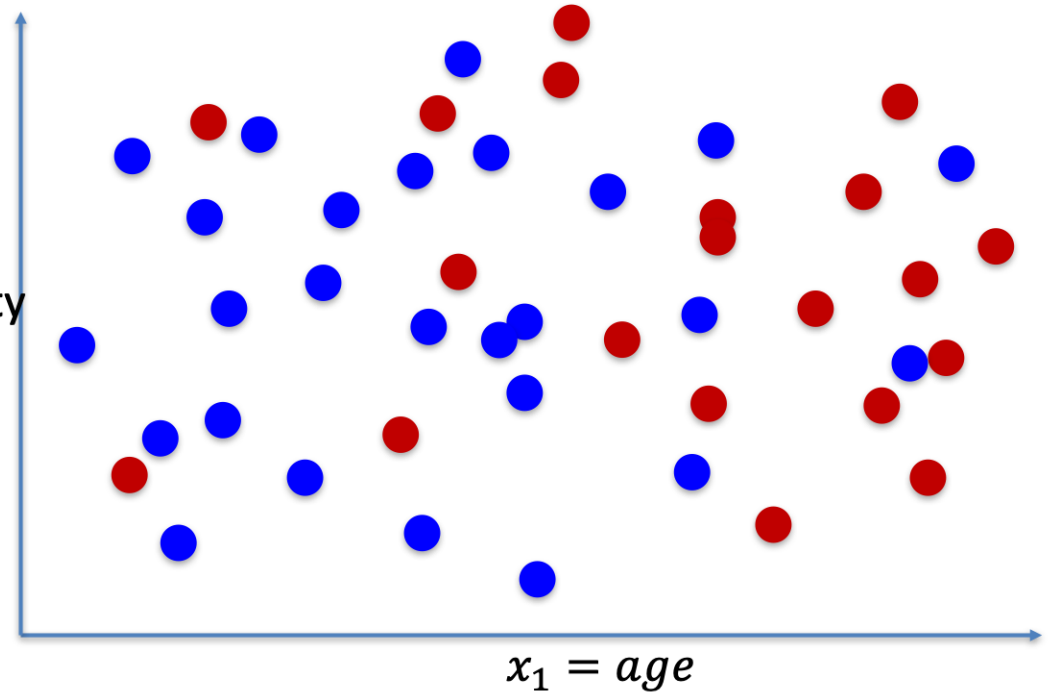
Propensity score: the probability of receiving one of the treatments being compared, given the measured covariates
PropensityScore = Prob(Treatment | Covariates)

Slide Adapted from: David Sontag (MIT)

Inverse propensity score re-weighting

$$p(x|t=0) \neq p(x|t=1)$$

control *treated*



CCI: 10-year survival in patients with multiple comorbidities

Treated

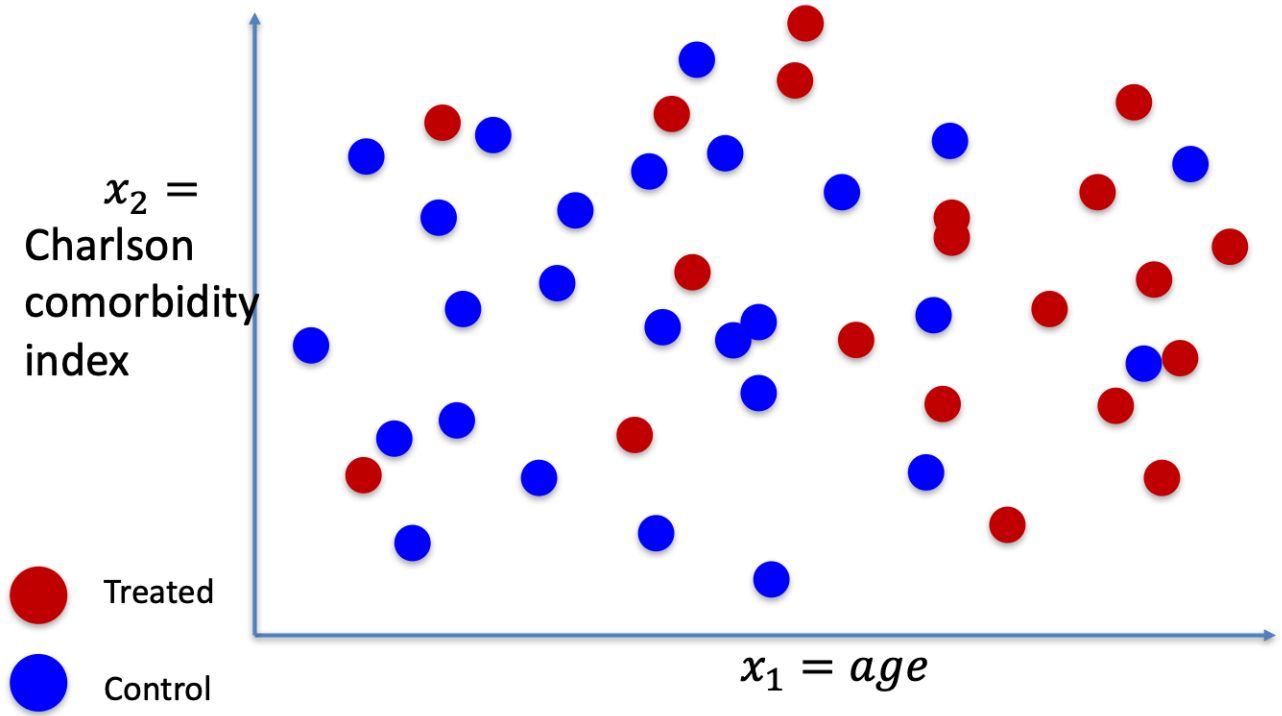
Control

Slide Adapted from: David Sontag (MIT)

Inverse propensity score re-weighting

$$p(x|t=0) \cdot w_0(x) \approx p(x|t=1) \cdot w_1(x)$$

reweighted control *reweighted treated*



Slide Adapted from: David Sontag (MIT)

Propensity score

- Propensity score: $p(T = 1|x)$, using machine learning tools
- Samples re-weighted by the inverse propensity score of the treatment they received

Propensity score: the probability of receiving one of the treatments being compared, given the measured covariates
PropensityScore = Prob(Treatment | Covariates)

Slide Adapted from: David Sontag (MIT)

Propensity scores – algorithm

Inverse probability of treatment weighted estimator

(Average Treatment Effect)

How to calculate ATE with propensity score

PropensityScore = Prob(Treatment | Covariates)

for sample $(x_1, t_1, y_1), \ldots, (x_n, t_n, y_n)$

1. Use any ML method to estimate $\hat{p}(T = t|x)$

2. $$\hat{ATE} = \frac{1}{n} \sum_{i \text{ s.t. } t_i=1} \frac{y_i}{\hat{p}(t_i = 1|x_i)} - \frac{1}{n} \sum_{i \text{ s.t. } t_i=0} \frac{y_i}{\hat{p}(t_i = 0|x_i)}$$

treated

untreated

Slide Adapted from: David Sontag (MIT)

Propensity scores – algorithm

Inverse probability of treatment weighted estimator

(Average Treatment Effect)

How to calculate ATE with propensity score

PropensityScore = Prob(Treatment | Covariates)

for sample $(x_1, t_1, y_1), \ldots, (x_n, t_n, y_n)$

1. Randomized trial $p(T = t|x) = 0.5$

2. $$\hat{ATE} = \frac{1}{n} \sum_{i \text{ s.t. } t_i=1} \frac{y_i}{\hat{p}(t_i = 1|x_i)} - \frac{1}{n} \sum_{i \text{ s.t. } t_i=0} \frac{y_i}{\hat{p}(t_i = 0|x_i)}$$

treated

untreated

Slide Adapted from: David Sontag (MIT)

Propensity scores – algorithm

Inverse probability of treatment weighted estimator

(Average Treatment Effect)

How to calculate ATE with propensity score

PropensityScore = Prob(Treatment | Covariates)

for sample $(x_1, t_1, y_1), \ldots, (x_n, t_n, y_n)$

1. Randomized trial $p(T = t|x) = 0.5$

2. $\hat{ATE} = \frac{1}{n} \sum_{i \text{ s.t. } t_i=1} \frac{y_i}{0.5} - \frac{1}{n} \sum_{i \text{ s.t. } t_i=0} \frac{y_i}{0.5} =$

Slide Adapted from: David Sontag (MIT)

Propensity scores – algorithm

Inverse probability of treatment weighted estimator

How to calculate ATE with propensity score

for sample $(x_1, t_1, y_1), \ldots, (x_n, t_n, y_n)$

1. Randomized trial $p = 0.5$

2. $$\hat{ATE} = \frac{1}{n} \sum_{i \text{ s.t. } t_i=1} \frac{y_i}{0.5} - \frac{1}{n} \sum_{i \text{ s.t. } t_i=0} \frac{y_i}{0.5} =$$
$$\frac{2}{n} \sum_{i \text{ s.t. } t_i=1} y_i - \frac{2}{n} \sum_{i \text{ s.t. } t_i=0} y_i$$

Slide Adapted from: David Sontag (MIT)

Propensity scores – algorithm

Inverse probability of treatment weighted estimator

How to calculate ATE with propensity score

for sample $(x_1, t_1, y_1), \ldots, (x_n, t_n, y_n)$

1. Randomized trial $p = 0.5$

Sum over $\sim \frac{n}{2}$ terms

2. $$\hat{ATE} = \frac{1}{n} \sum_{i \text{ s.t. } t_i=1} \frac{y_i}{0.5} - \frac{1}{n} \sum_{i \text{ s.t. } t_i=0} \frac{y_i}{0.5} =$$

$$\frac{2}{n} \sum_{i \text{ s.t. } t_i=1} y_i - \frac{2}{n} \sum_{i \text{ s.t. } t_i=0} y_i$$

Slide Adapted from: David Sontag (MIT)

(Inverse Prob. Weighting)

Problems with IPW

- Need to estimate propensity score (problem in all propensity score methods)
- If there's not much overlap, propensity scores become non-informative and easily mis-calibrated
- Weighting by inverse can create large variance and large errors for small propensity scores
  - Exacerbated when more than two treatments

Slide Adapted from: David Sontag (MIT)

Many more ideas and methods – Natural experiments

- Does stress during pregnancy affect later child development?
- Confounding: genetic, mother personality, economic factors…
- Natural experiment: the Cuban missile crisis of October 1962. Many people were afraid a nuclear war is about to break out.
- Compare children who were in utero during the crisis with children from immediately before and after

Slide Adapted from: David Sontag (MIT)

Many more ideas and methods – Instrumental variables

- Informally: a variable which affects treatment assignment but not the outcome
- Example: are private schools better than public schools?
- Confounding: different student population, different teacher population
- Can’t force people which school to go to

Slide Adapted from: David Sontag (MIT)

Many more ideas and methods – Instrumental variables

- Informally: a variable which affects treatment assignment but not the outcome
- Example: are private schools better than public schools?
- Can't force people which school to go to
- Can *randomly* give out vouchers to some children, giving them an opportunity to attend private schools
- The voucher assignment is the instrumental variable

Slide Adapted from: David Sontag (MIT)

Goals for Today: Causality

1. Introduction: real-world evidence, challenges of inference and causality

- Observational data are shaped by informed interventions, feedback loops

2. Methods of Causal Inference

- Potential Outcomes Framework
- Typical Underlying Assumptions: No unmeasured confounders, Ignorability, Common support
- Statistical methods of causal inference: Covariate adjustment, Propensity score re-weighting
- Natural Experiments

3. From Association to Causality: Principles of causality in real-world evidence

- Strength: Causal effects often have stronger associations.
- Consistency: Causal effects can be repeatedly observed by different persons, in different places, circumstances, and times?
- Specificity: Causal effects are often condition-specific.
- Temporality: Causal effects respect temporal ordering.
- Biological gradient: Causal effects are influenced by underlying biology which has continuously-varying characteristics.
- Plausibility: Causal effects should be plausible.
- Coherence: Causal effects should not seriously conflict with the generally known mechanisms.
- Experiment: Causal effects can predict experimental evidence.
- Analogy: Causal effects may be transferable between systems.

4. Causality in practice

- Bias and Confounding: Handling sensitive variables, explicitly sampling across groups
- All models are wrong, but some are useful -> Do we always want a biologically causal model?
- Useful models require careful consideration of setting and use cases to include generalizability, interventions, and risks.

Principles of Causality in Observational Data

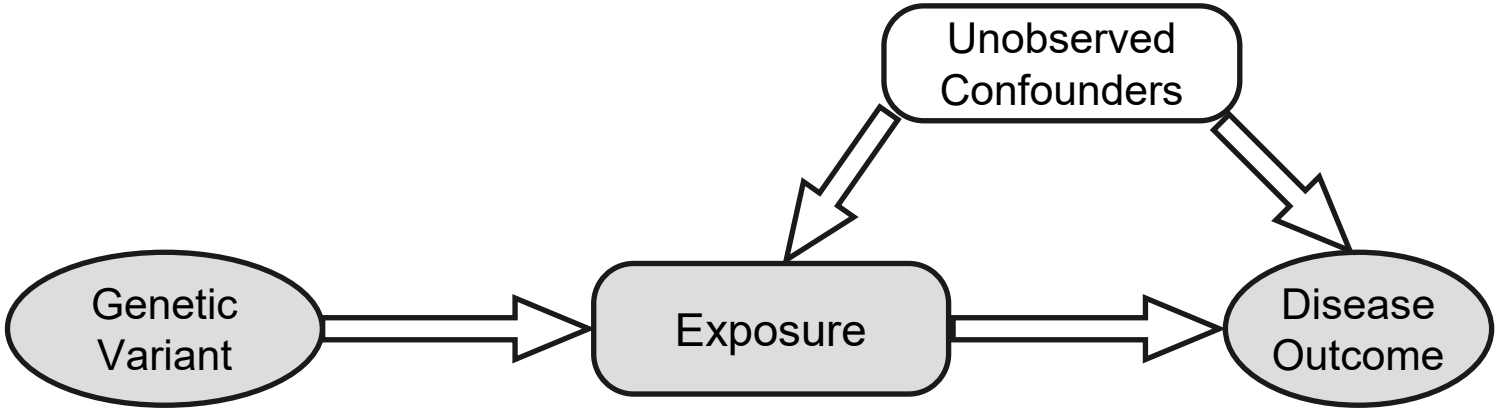
1. **Temporality**: Causal effects respect temporal ordering. (Mendelian Randomization)
2. **Biological gradient**: Causal effects are influenced by underlying biology which has continuously-varying characteristics. (Threshold-based effects)
3. **Plausibility**: Causal effects should be plausible. (Comorbidities)
4. **Coherence**: Causal effects should not seriously conflict with the generally known mechanisms. (Systematically aligning)
5. **Strength**: Causal effects often have stronger associations.
6. **Consistency**: Causal effects can be repeatedly observed by different persons, in different places, circumstances, and times.
7. **Specificity**: Causal effects are often condition-specific.
8. **Experiment**: Causal effects can predict experimental evidence.
9. **Analogy**: Causal effects may be transferable between systems.

Hill AB. Proceedings of the Royal Society of Medicine. 1965

Temporality: Mendelian Randomization and Natural Temporality of Genetics

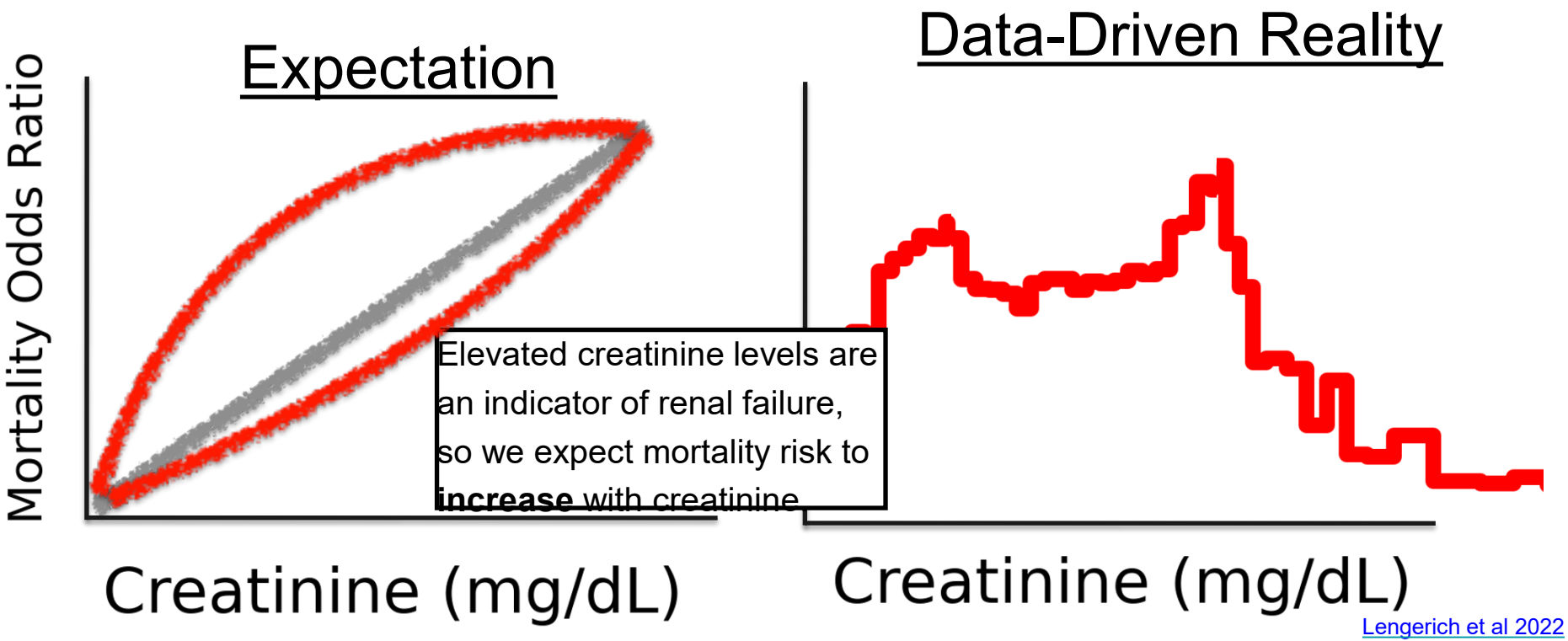
- Use inherited genetic variants as pseudo-randomized assignment to infer causal relationship between exposure and disease outcome.
- The fundamental idea:
  - Genetic instrumental variables help extract the variation in intermediate phenotype that is **independent** of confounding variables.
  - Use this variation to estimate causal effect.
- Common method of MR: 2-Step Least Squares

Temporality: Mendelian Randomization and Natural Temporality of Genetics



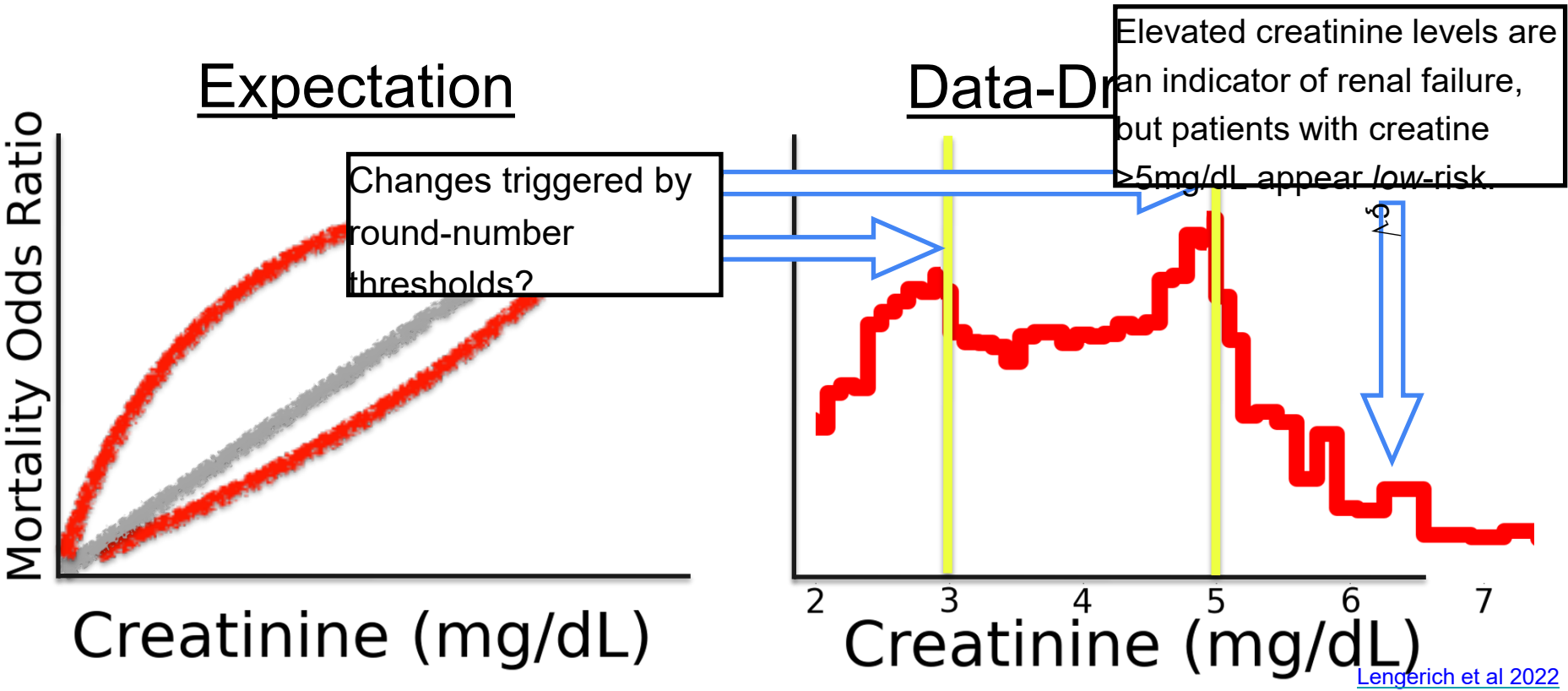
The fundamental idea: If we can't randomize assignment of exposure, we can use the *portion* of the exposure which is predictable from the naturally randomized genetic variant.

Biological Gradient: Causal Effects Should Be Smooth and Often Monotone



Lengerich et al 2022

Biological Gradient: Causal Effects Should Be Smooth and Often Monotone

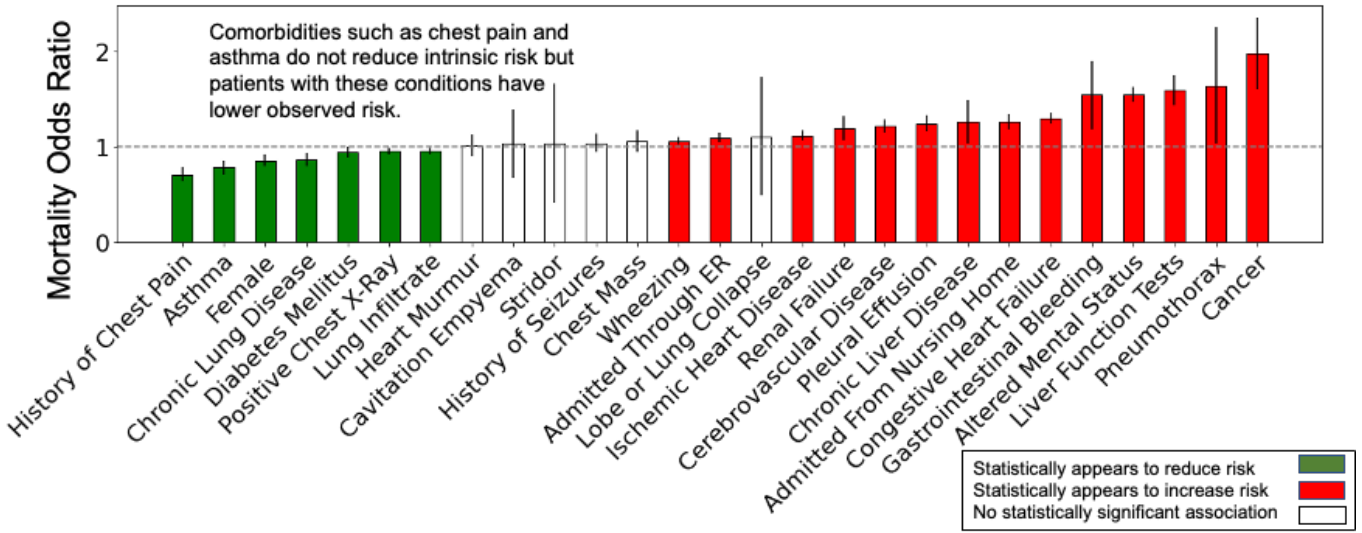


Lengerich et al 2022

Plausibility: Causal Biological Effects Should Make Be Plausible

- Real-world associations of chronic comorbidities in EHRs are often the opposite sign that they would be plausible.

For patients hospitalized with pneumonia:



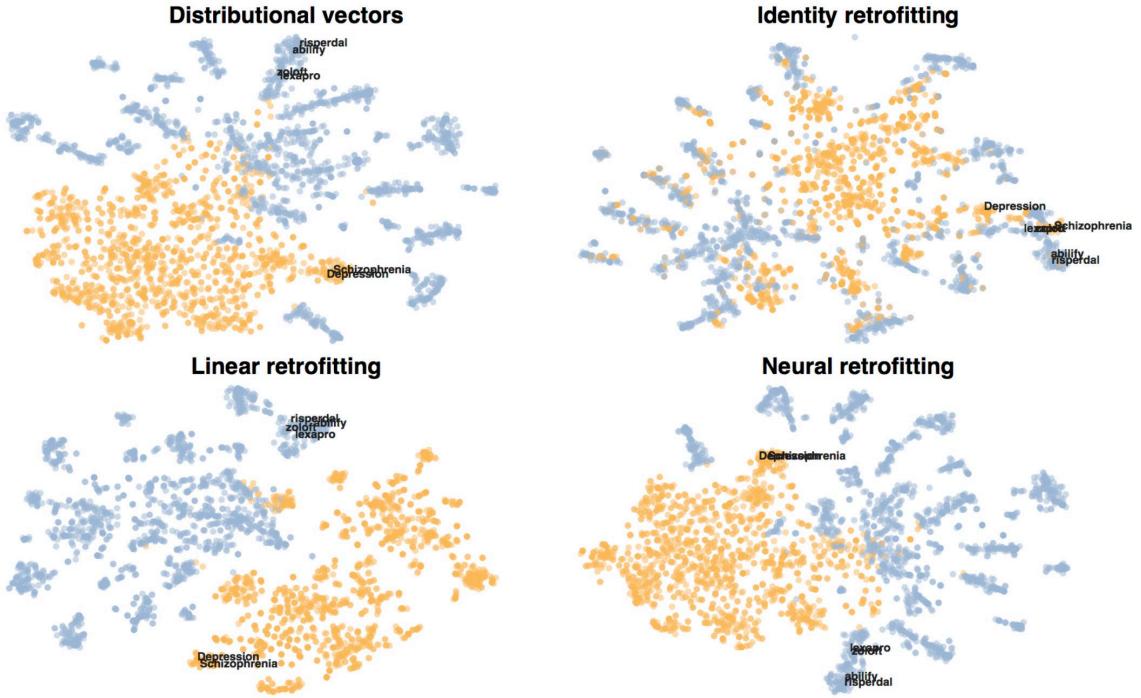
Lengerich et al 2022

Coherence: Systematically Aligning Real-World Evidence with Domain Knowledge

- Human-in-the-loop training
- Simulations
- Knowledge graphs
- Multi-task models
- Data augmentation policies
- Teacher-student model training to encode logic rules
- Re-usable representations / embeddings

Coherence: Retrofitting Distributional Embeddings to Knowledge Graphs

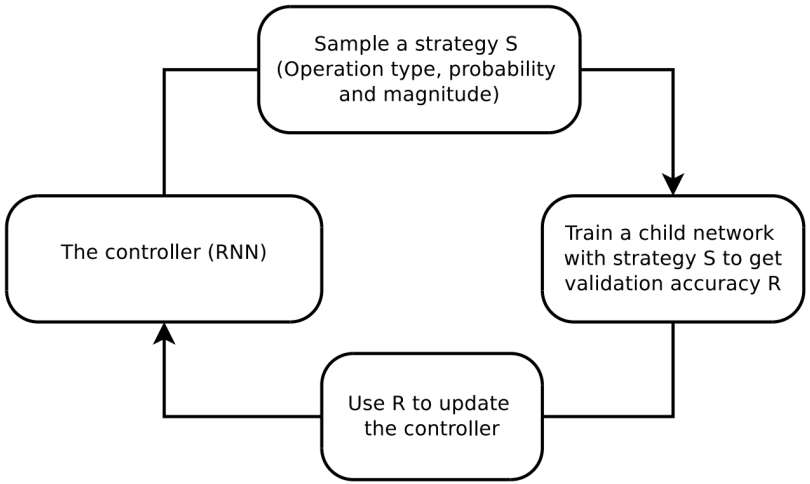
- Forces distributional (co-occurrence, etc) embeddings to respect known relations encoded as knowledge graph edges



Lengerich et al 2018

Coherence: Data Augmentation

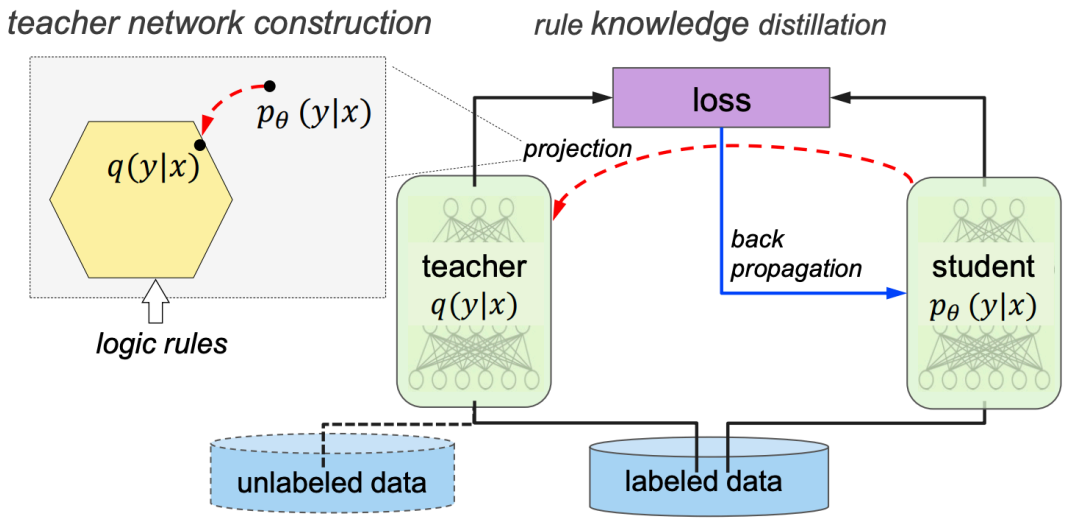
- Augmentation policies: orthogonal to labels
- Good augmentation policies for images: resize, rotate, crop, color
- Can we learn to augment from data?
- Current bottleneck: requires a set of atomic augmentation techniques to mix.



Cubic et al 2018

Coherence: Logic Rules with Teacher-Student Models

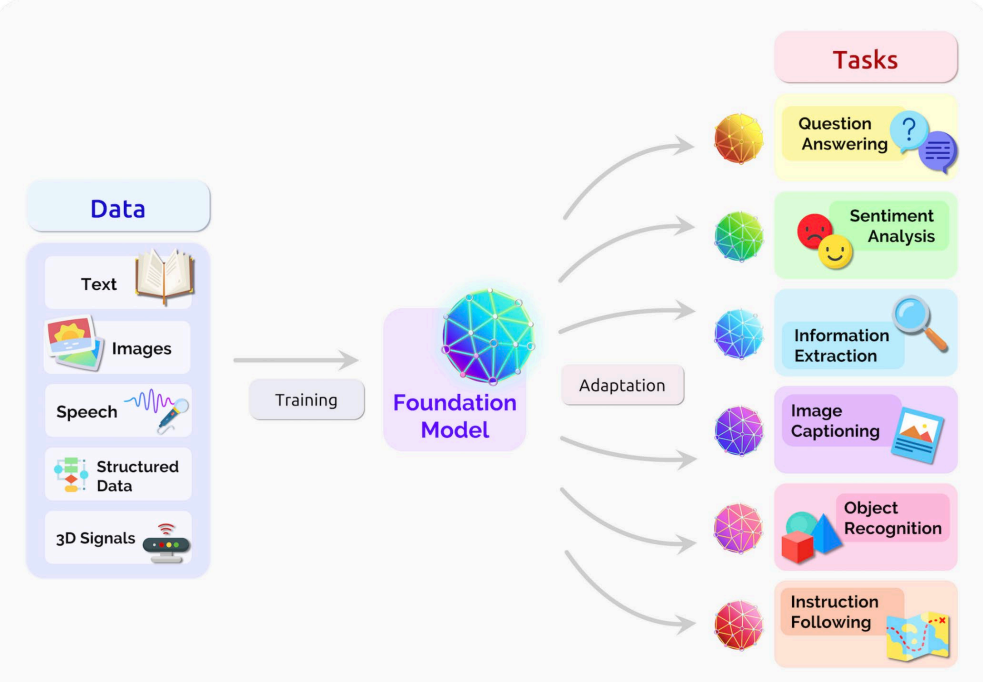
- How can we make a NN that respects logic rules?
- One solution: build a teacher model that is forced to respect the logic rules, and then learn to trade off against data.



Hu et al 2020

Coherence: Foundation Models

- Pretext task: masked token prediction
- If we observe many examples and routinely fill in the missing *syntax*, then some underlying *semantics* are implicitly learned.
- Do new inferences respect the semantics learned in training corpus?



Bommasani et al. 2022

Goals for Today: Causality

1. Introduction: real-world evidence, challenges of inference and causality

- Observational data are shaped by informed interventions, feedback loops

2. Methods of Causal Inference

- Potential Outcomes Framework
- Typical Underlying Assumptions: No unmeasured confounders, Ignorability, Common support
- Statistical methods of causal inference: Covariate adjustment, Propensity score re-weighting
- Natural Experiments

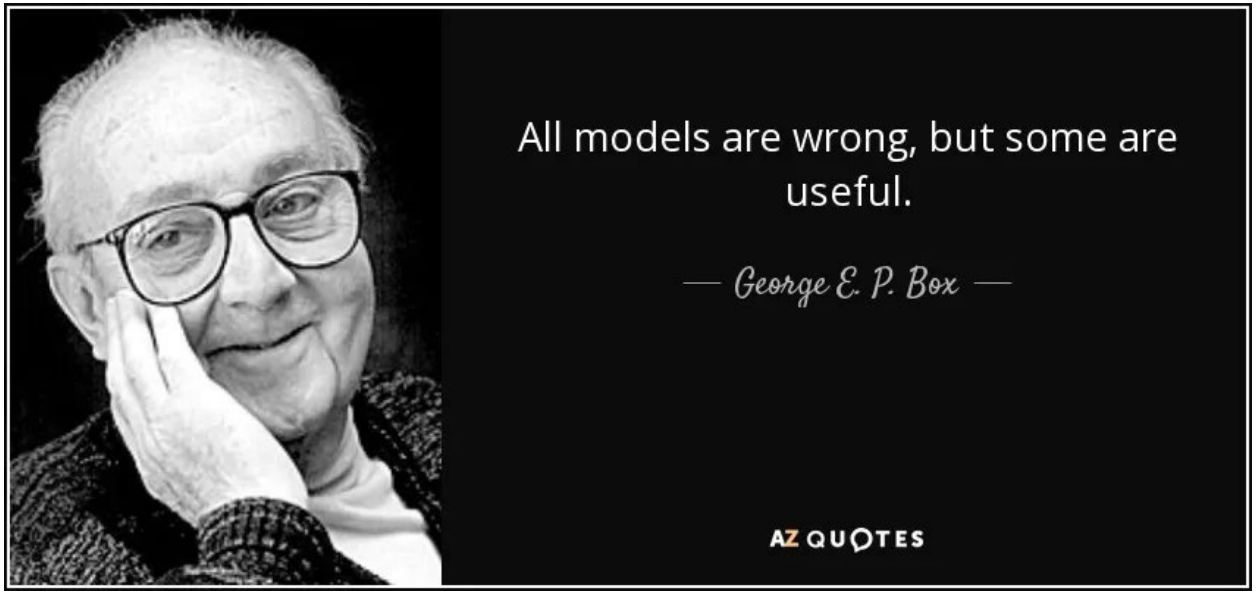
3. From Association to Causality: Principles of causality in real-world evidence

- Strength: Causal effects often have stronger associations.
- Consistency: Causal effects can be repeatedly observed by different persons, in different places, circumstances, and times?
- Specificity: Causal effects are often condition-specific.
- Temporality: Causal effects respect temporal ordering.
- Biological gradient: Causal effects are influenced by underlying biology which has continuously-varying characteristics.
- Plausibility: Causal effects should be plausible.
- Coherence: Causal effects should not seriously conflict with the generally known mechanisms.
- Experiment: Causal effects can predict experimental evidence.
- Analogy: Causal effects may be transferable between systems.

4. Causality in practice

- Bias and Confounding: Handling sensitive variables, explicitly sampling across groups
- All models are wrong, but some are useful -> Do we always want a biologically causal model?
- Useful models require careful consideration of setting and use cases to include generalizability, interventions, and risks.

Causality in practice



Causality in practice: Bias and Confounding

- Suppose we want to learn causal effect invariant to a sensitive feature.
  - Should we hold that feature out of training?
- At test time, we can:
  - Throw away the effect of the feature
  - Sweep over all values of the sensitive feature

Causality in practice: Do we always want a biologically causal model?

- Accuracy isn't enough for medical impact:
  - Medicine is a continuous process: Should we include prior lab tests as a predictor in our model?
  - Should we include prior treatments as a predictor in our model? They are missing not-at-random.
- Techniques depend on our goal:
  - learning biological causality?
  - maximizing predictive accuracy?
  - guiding interventions?

Resources

- [Bommasani et al 2022] On the Opportunities and Risks of Foundation Models
- [Cubuk et al 2019] AutoAugment: Learning Augmentation Policies from Data
- [Faruqui et al 2015] Retrofitting Word Vectors to Semantic Lexicons
- [Henry et al 2022] Factors driving provider adoption of the TREWS machine learning-based early warning system and its effects on sepsis treatment timing
- [Hill AB 1965] The Environment and Disease: Association or Causation?
- [Hsu] Mendelian Randomization
- [Hu et al 2020] Harnessing Deep Neural Networks with Logic Rules.
- [Lengerich et al 2018] Retrofitting Distributional Embeddings to Knowledge Graphs with Functional Relations
- [Lengerich et al 2022] Death by Round Numbers: Glass-Box Machine Learning Uncovers Biases in Medical Practice
- [Wong et al 2021] External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients