Machine Learning for Healthcare

6.871, HST.956

Lecture 10: Causal Inference Part 1

David Sontag

CSAIL

INSTITUTE FOR MEDICAL ENGINEERING & SCIENCE

HEALTH SCIENCES & TECHNOLOGY

Recall from lecture 3…

Patient/Provider Goals of Clinical Data Science

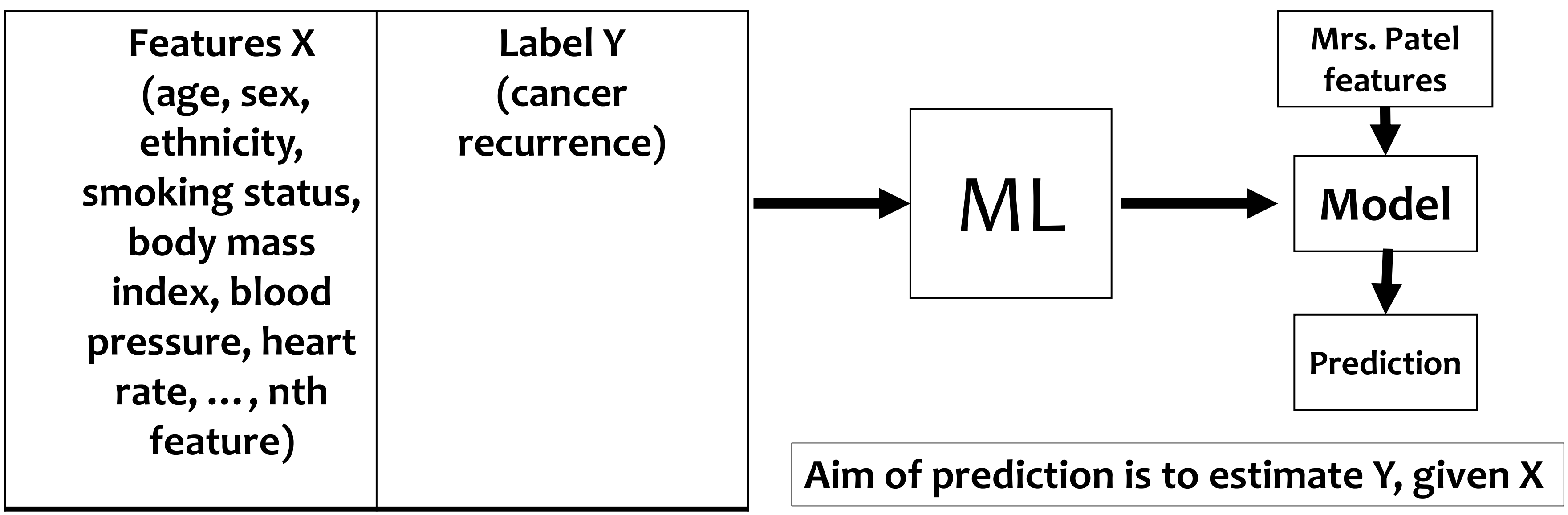
- Mrs. Patel is a 65 year old who was recently diagnosed with kidney cancer. She returns to your office to discuss treatment and has some questions.
  - After treatment, **what is the risk** of my cancer coming back before the Ultimate World Cruise (December 2023)?
  - **Will the risk** of my cancer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

How would you answer these questions using clinical data science?

Recall from lecture 3…

Will my cancer come back?

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

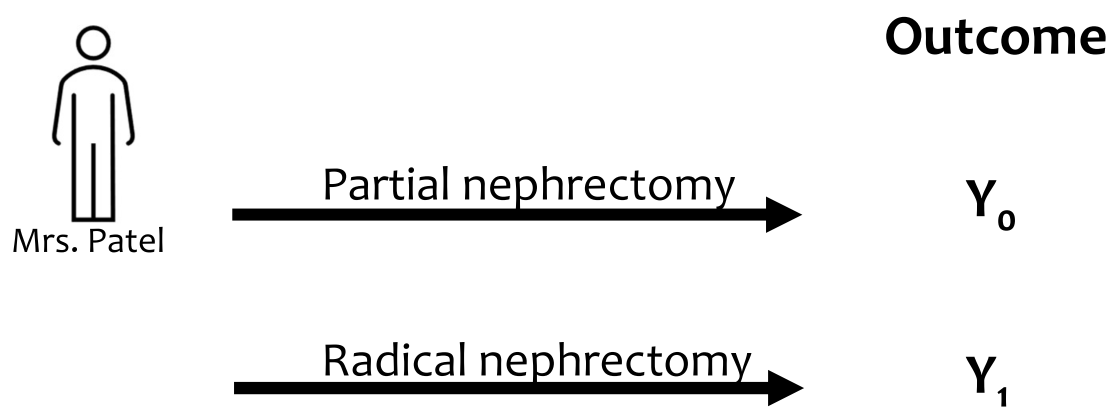


Recall from lecture 3…

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



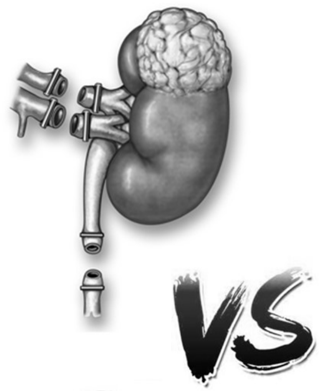
- Reality: We cannot know the ground truth

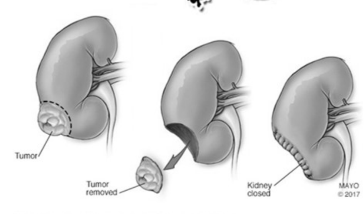
Recall from lecture 3…

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

Recall from lecture 3…

Clinical Research Study Designs

Descriptive
- Case report
- Case series
- Survey

Analytic

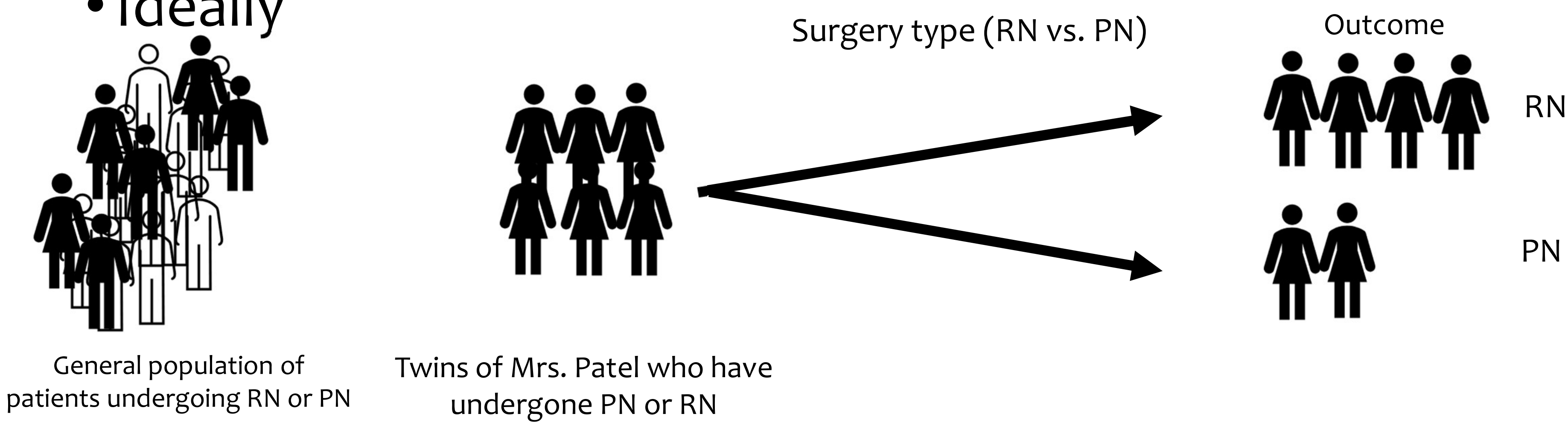
Observational
- Cohort studies
- Cross sectional
- Case-control

Experimental
- Randomized controlled trials

Recall from lecture 3…

Change the risk of my cancer coming back?

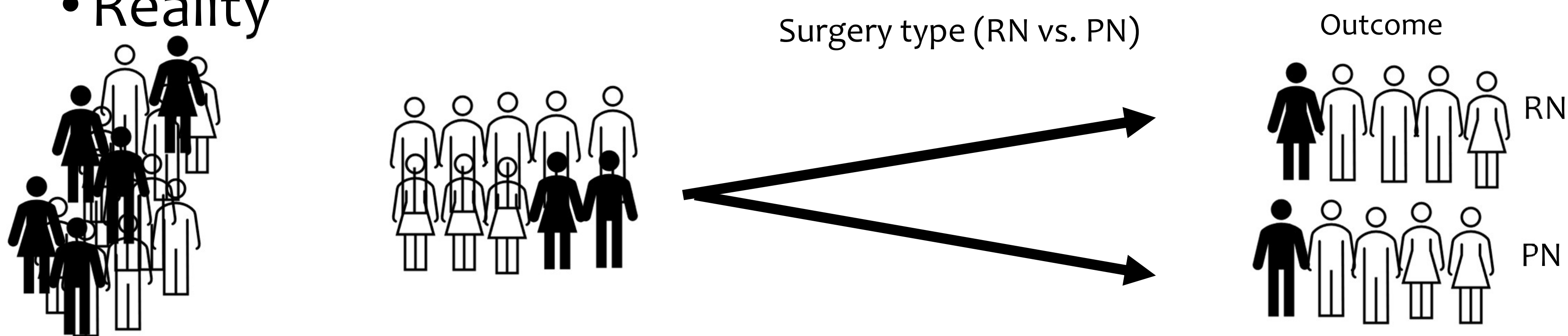
- You hypothesize that type of surgery (partial vs. radical) will change her risk of cancer recurrence. How do you evaluate this hypothesis?
  - Ideally



Recall from lecture 3…

Change the risk of my cancer coming back?

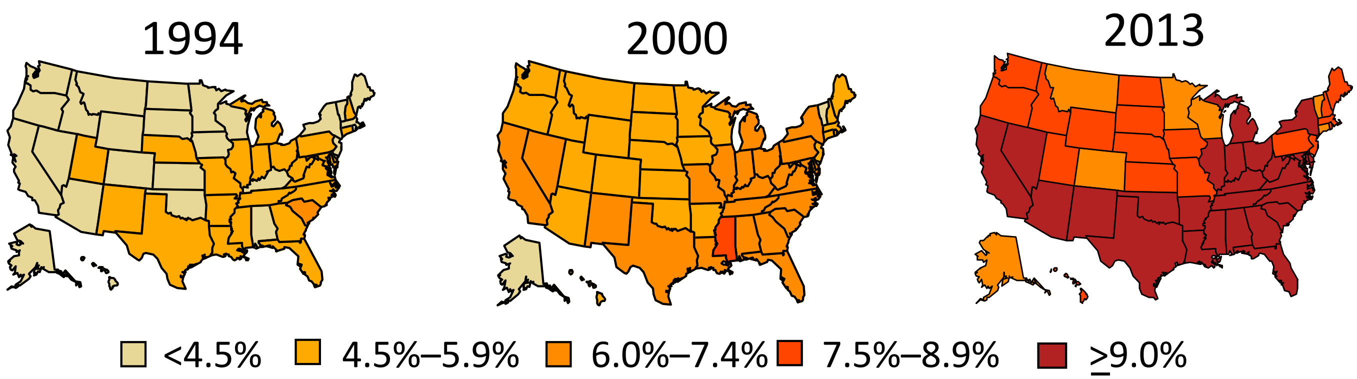
- You hypothesize that type of surgery (partial vs. radical) will change her risk of cancer recurrence. How do you evaluate this hypothesis?
  - Reality



Selected population of patients undergoing RN or PN

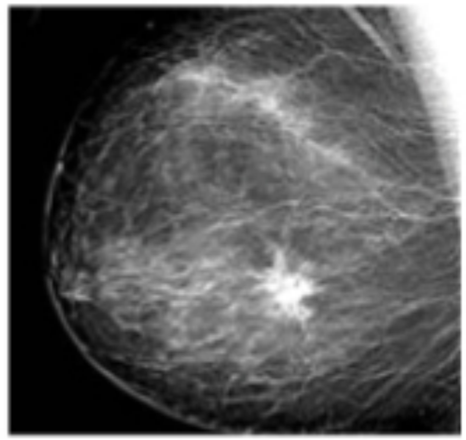
Patients **"similar"** to Mrs. Patel who have undergone PN or RN

Does gastric bypass surgery prevent onset of diabetes?

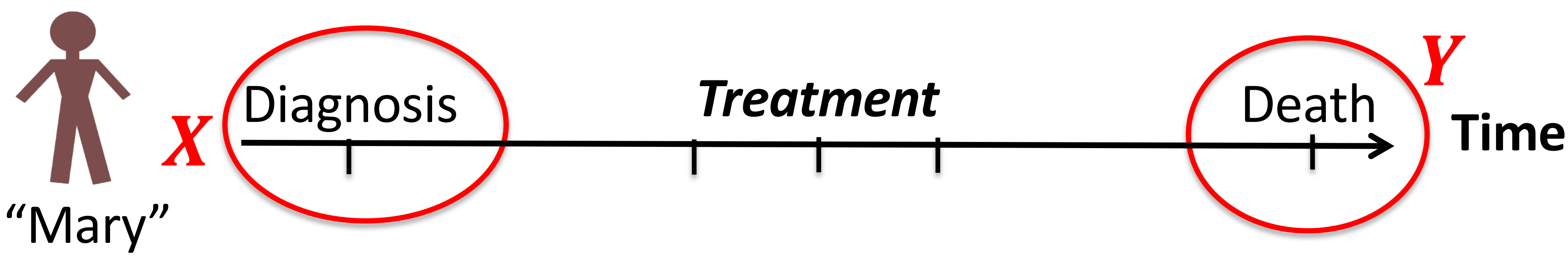


- In Lecture 4, we discussed case study of machine learning for early detection of Type 2 diabetes
- Health system doesn't want to know how to predict diabetes – they want to know how to *prevent it*
- Gastric bypass surgery is the highest negative weight (9th most predictive feature)
  - Does this mean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

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

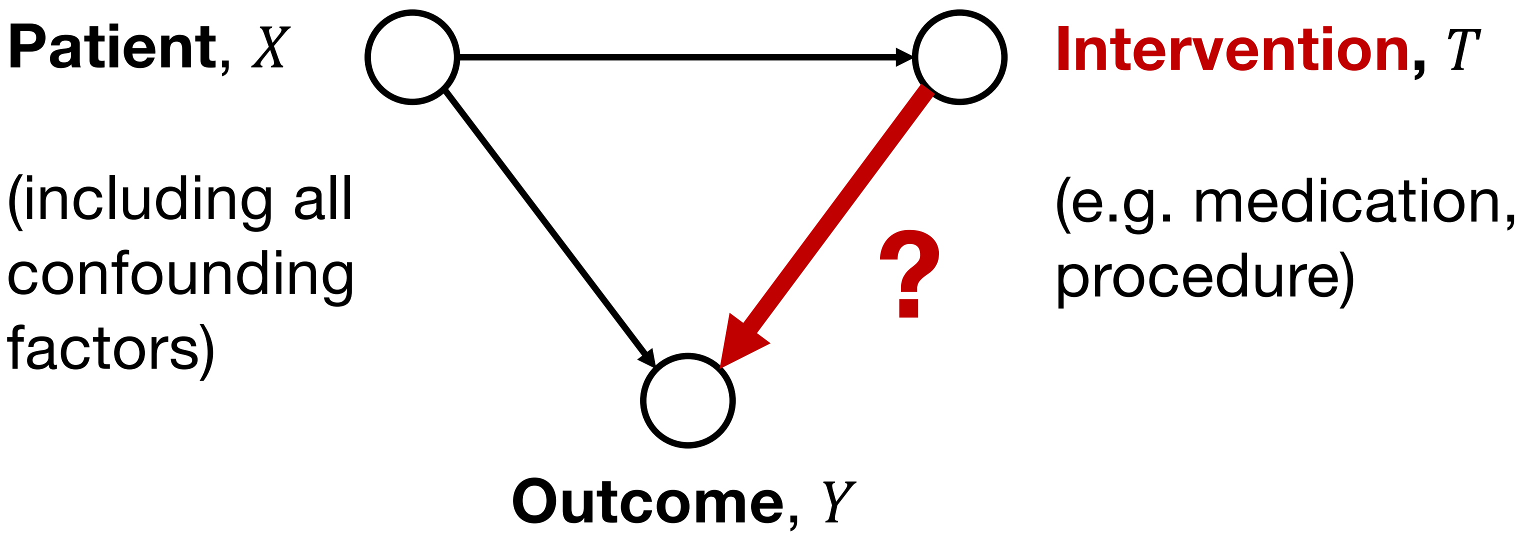


- Such predictive models widely used to stage patients. Should we initiate treatment? How aggressive?
- What could go wrong if we trained to predict survival, and then used to guide patient care?



A long survival time may be because of treatment!

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



High dimensional

Observational data

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:

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

* other commonly used notations for Y_0 are Y(0) and Y | do(T)=0

Potential Outcomes Framework (Rubin-Neyman Causal Model)

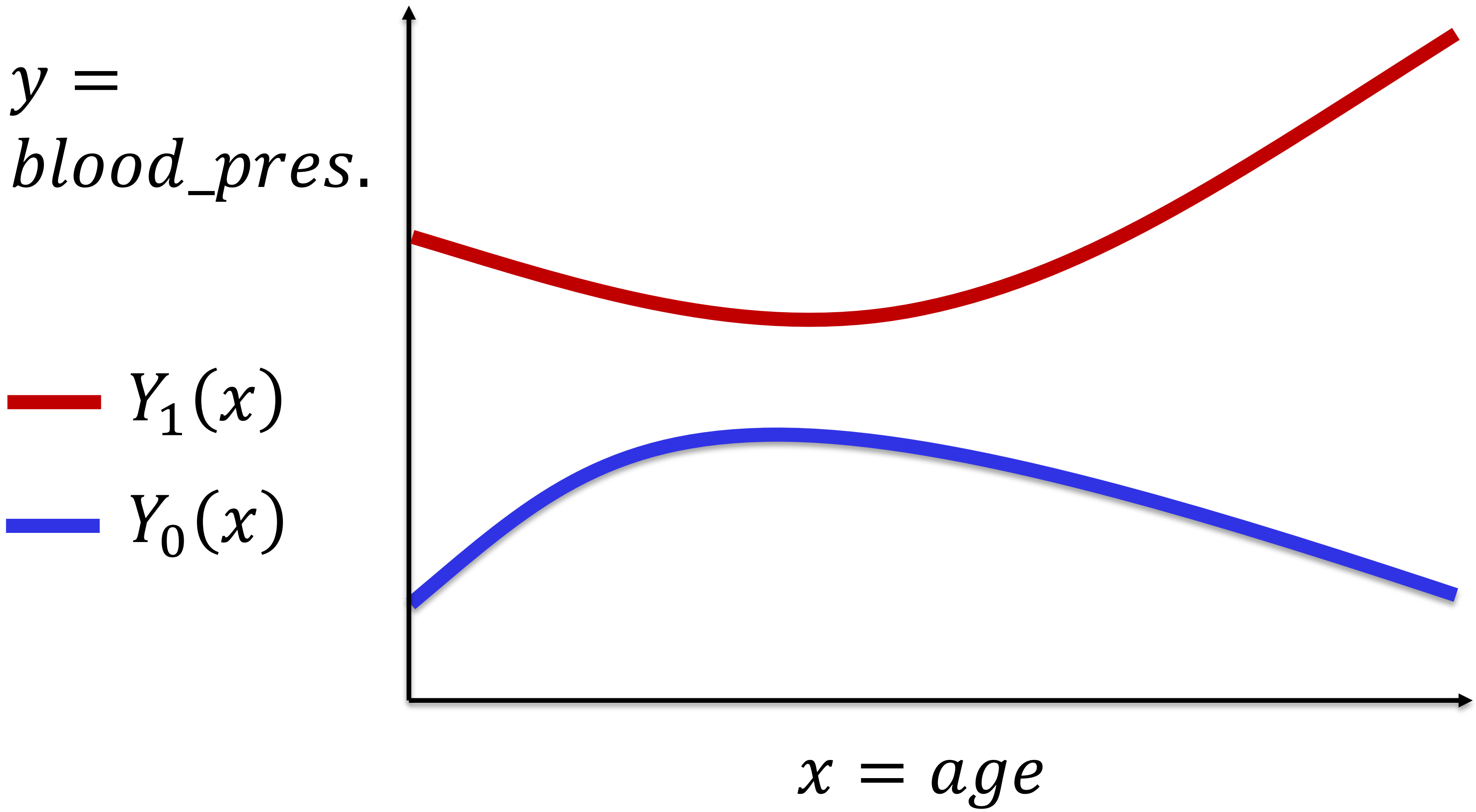
- Each unit (individual) x_i has two potential outcomes:
  - $Y_0(x_i)$ is the potential outcome had the unit not been treated: "***control outcome***"
  - $Y_1(x_i)$ is the potential outcome had the unit been treated: "***treated outcome***"

- Observed factual outcome:
  $$y_i = t_i Y_1(x_i) + (1 - t_i) Y_0(x_i)$$
- Unobserved counterfactual outcome:
  $$y_i^{CF} = (1 - t_i) Y_1(x_i) + t_i Y_0(x_i)$$

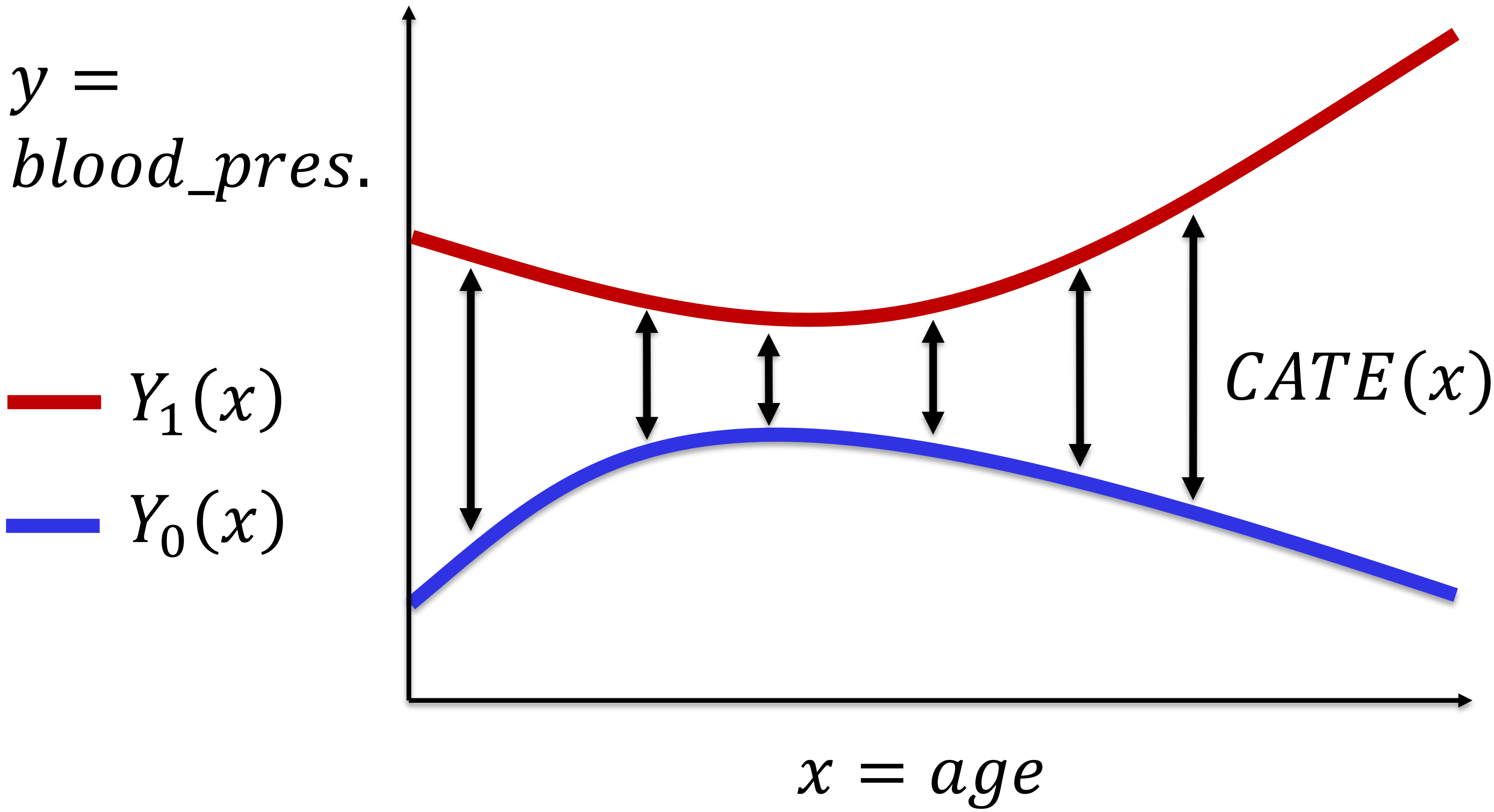
“The fundamental problem of causal inference”

We only ever observe one of the two outcomes

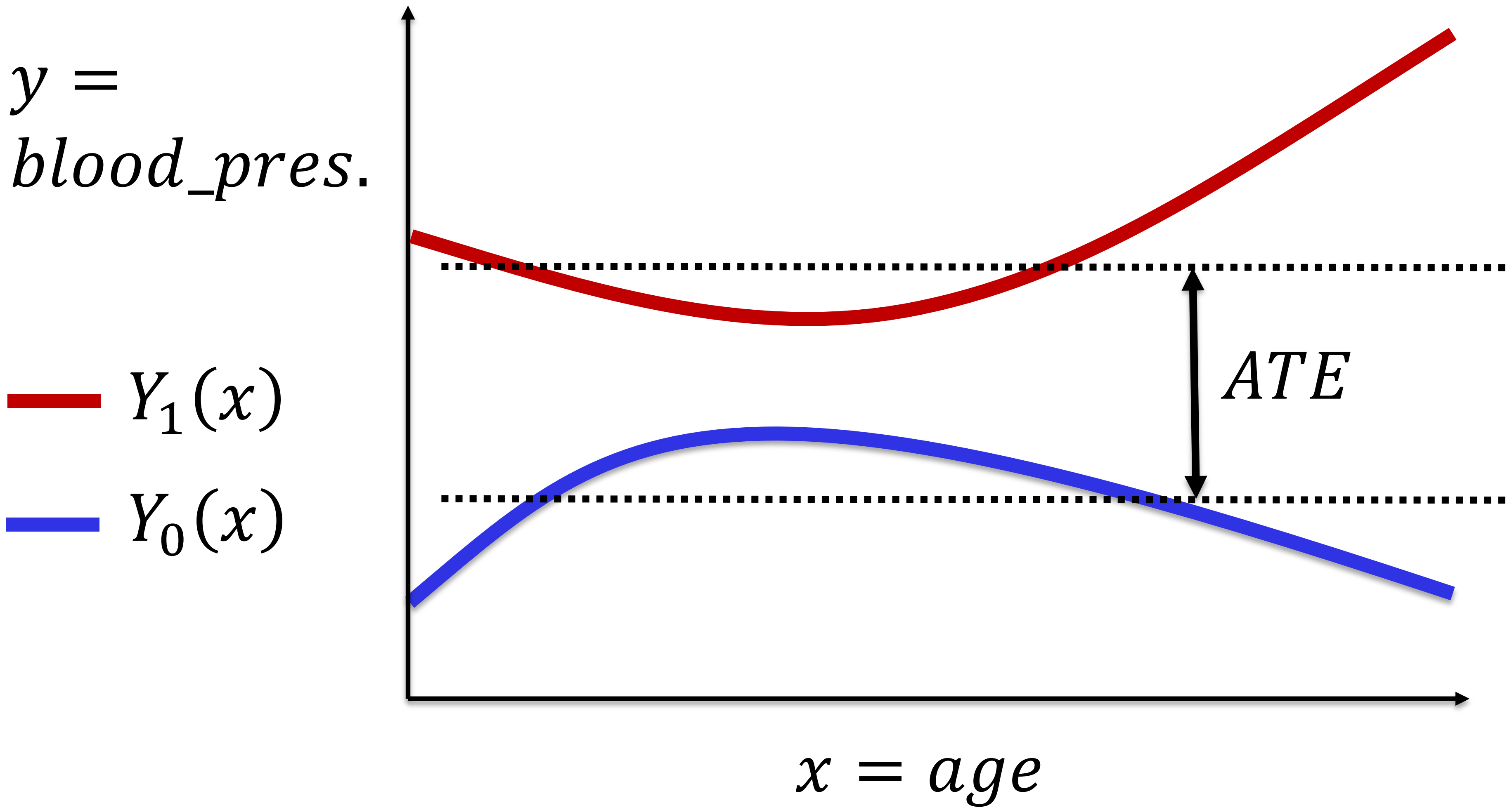
Example – Blood pressure and age



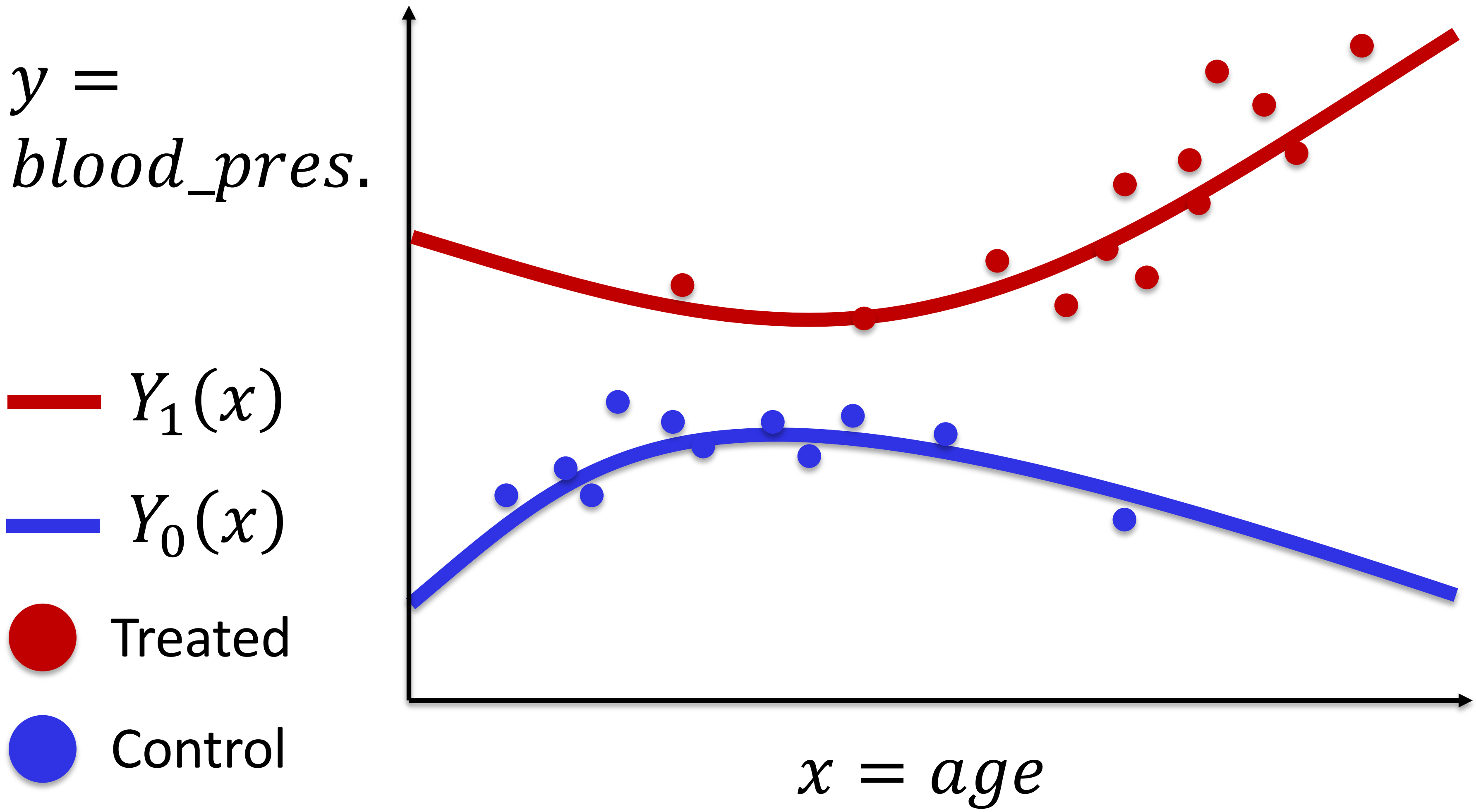
Blood pressure and age



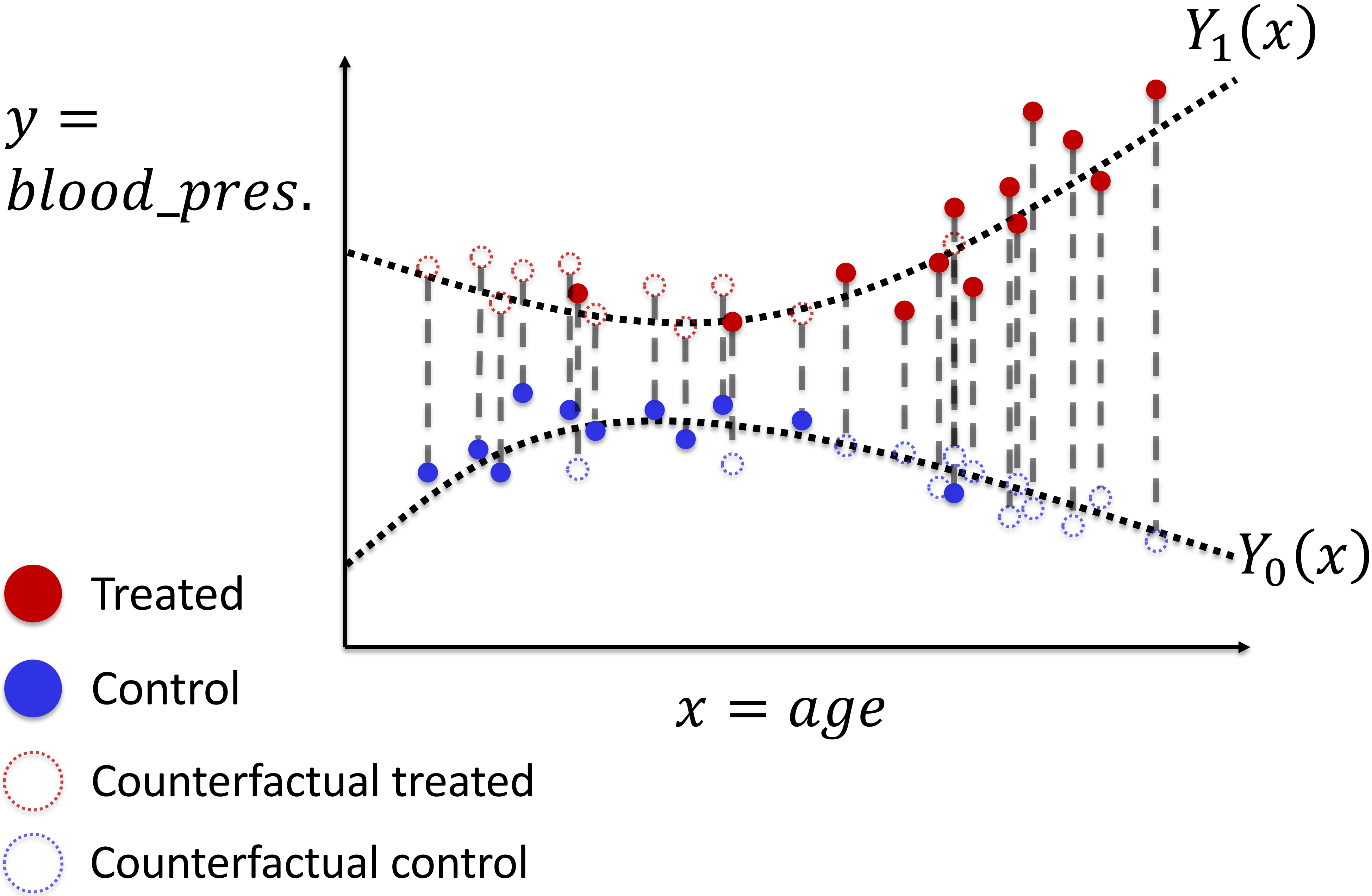
Blood pressure and age



Blood pressure and age



Blood pressure and age



(age, gender, exercise,treatment)			Observed sugar levels
(45, F, 0, **A**)			6
(45, F, 1, **B**)			6.5
(55, M, 0, **A**)			7
(55, M, 1, **B**)			8
(65, F, 0, **B**)			8
(65,F, 1, **A**)			7.5
(75,M, 0, **B**)			9
(75,M, 1, **A**)			8

(Example from Uri Shalit)

(age, gender, exercise)			Observed sugar levels
(45, F, 0)			6
(45, F, 1)			6.5
(55, M, 0)			7
(55, M, 1)			8
(65, F, 0)			8
(65,F, 1)			7.5
(75,M, 0)			9
(75,M, 1)			8

(Example from Uri Shalit)

(age, gender, exercise)	Y_0: Sugar levels *had they received* medication A	Y_1: Sugar levels *had they received* medication B	Observed sugar levels
(45, F, 0)	**6**	5.5	6
(45, F, 1)	7	**6.5**	6.5
(55, M, 0)	**7**	6	7
(55, M, 1)	9	**8**	8
(65, F, 0)	8.5	**8**	8
(65,F, 1)	**7.5**	7	7.5
(75,M, 0)	10	**9**	9
(75,M, 1)	**8**	7	8

(Example from Uri Shalit)

(age,gender, exercise)	Sugar levels *had they received* medication A	Sugar levels *had they received* medication B	Observed sugar levels
(45, F, 0)	**6**	5.5	6
(45, F, 1)	7	**6.5**	6.5
(55, M, 0)	**7**	6	7
(55, M, 1)	9	**8**	8
(65, F, 0)	8.5	**8**	8
(65,F, 1)	**7.5**	7	7.5
(75,M, 0)	10	**9**	9
(75,M, 1)	**8**	7	8

mean(sugar|medication B) –
mean(sugar|medicaton A) =
?

mean(sugar|*had they received* B) –
mean(sugar|*had they received* A) =
?

(Example from Uri Shalit)

Most important 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

Most important assumption – no unmeasured confounders

Y_0, Y_1: potential outcomes for control and treated

x: unit covariates (features)

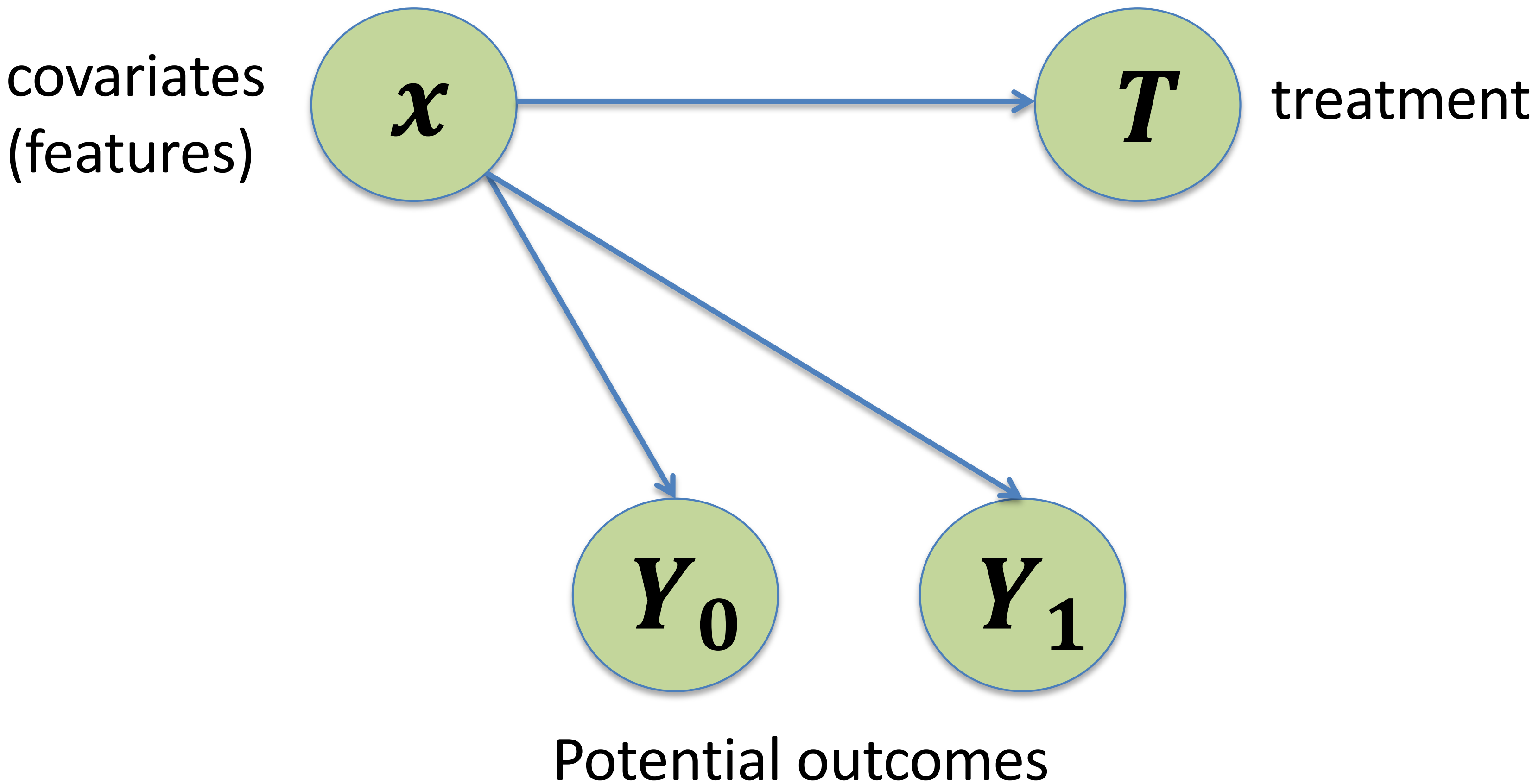
T: treatment assignment

We assume:

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

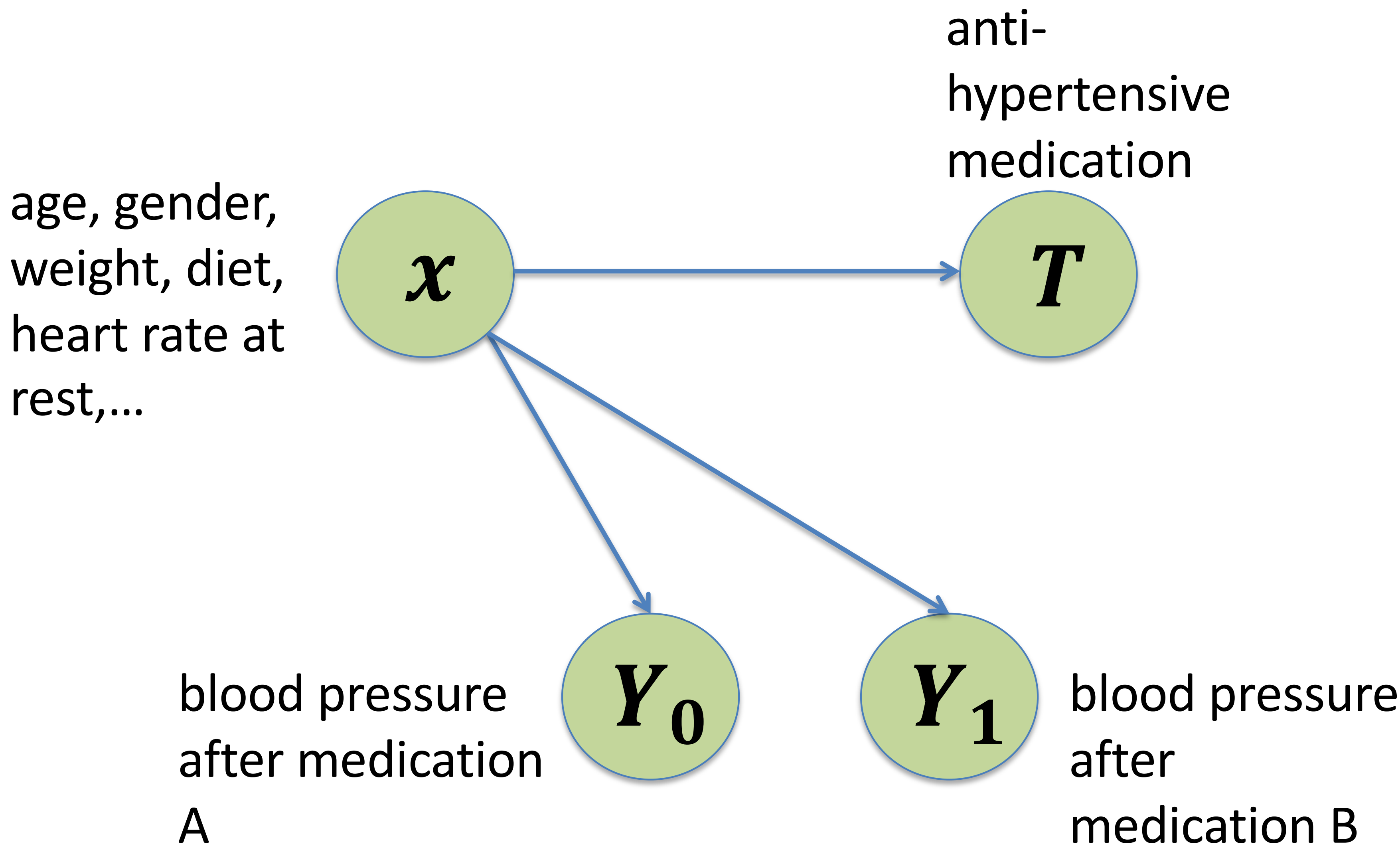
Ignorability

Ignorability



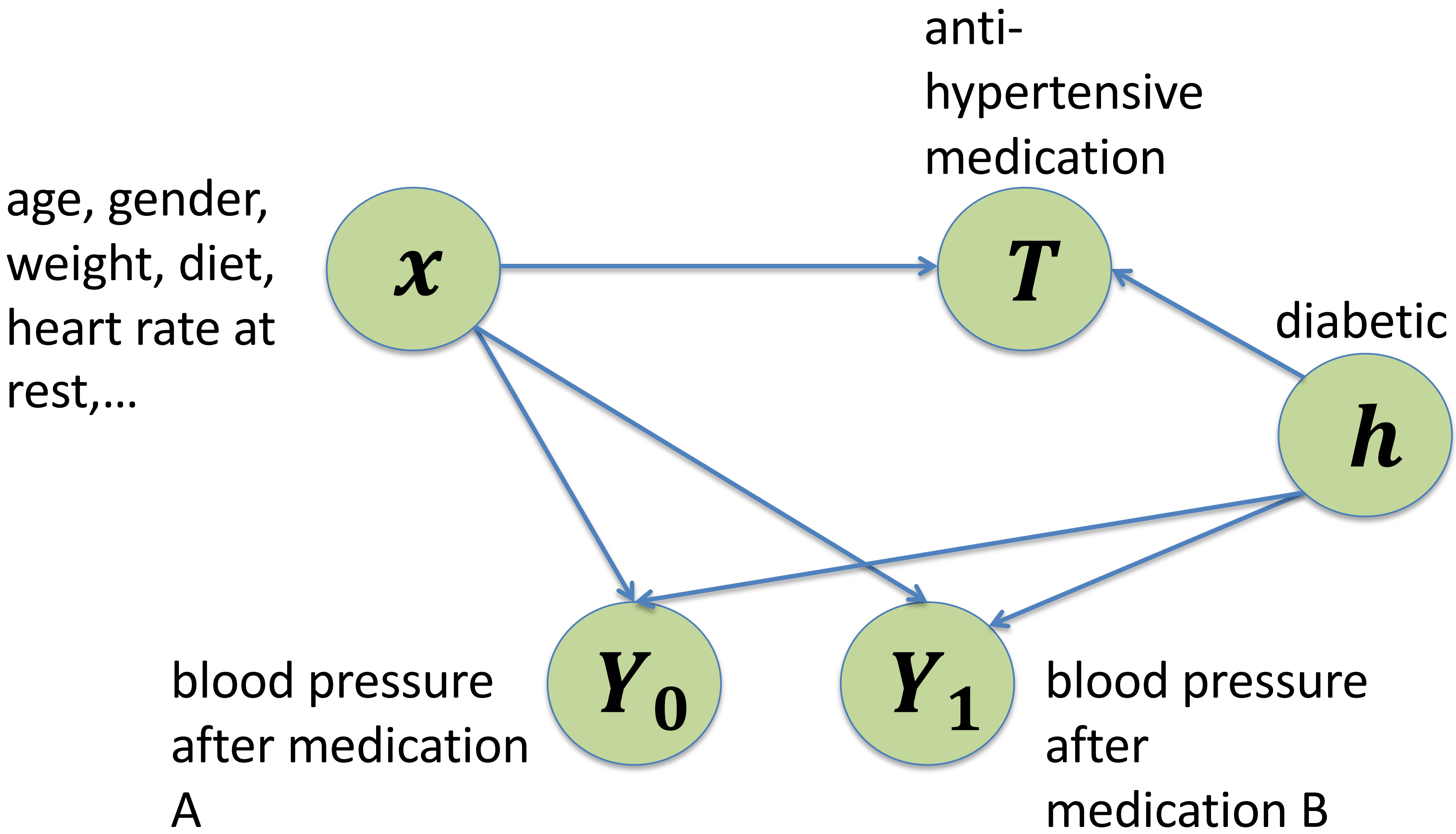
$$(Y_0, Y_1) \perp\!\!\!\perp T \mid x$$

Ignorability

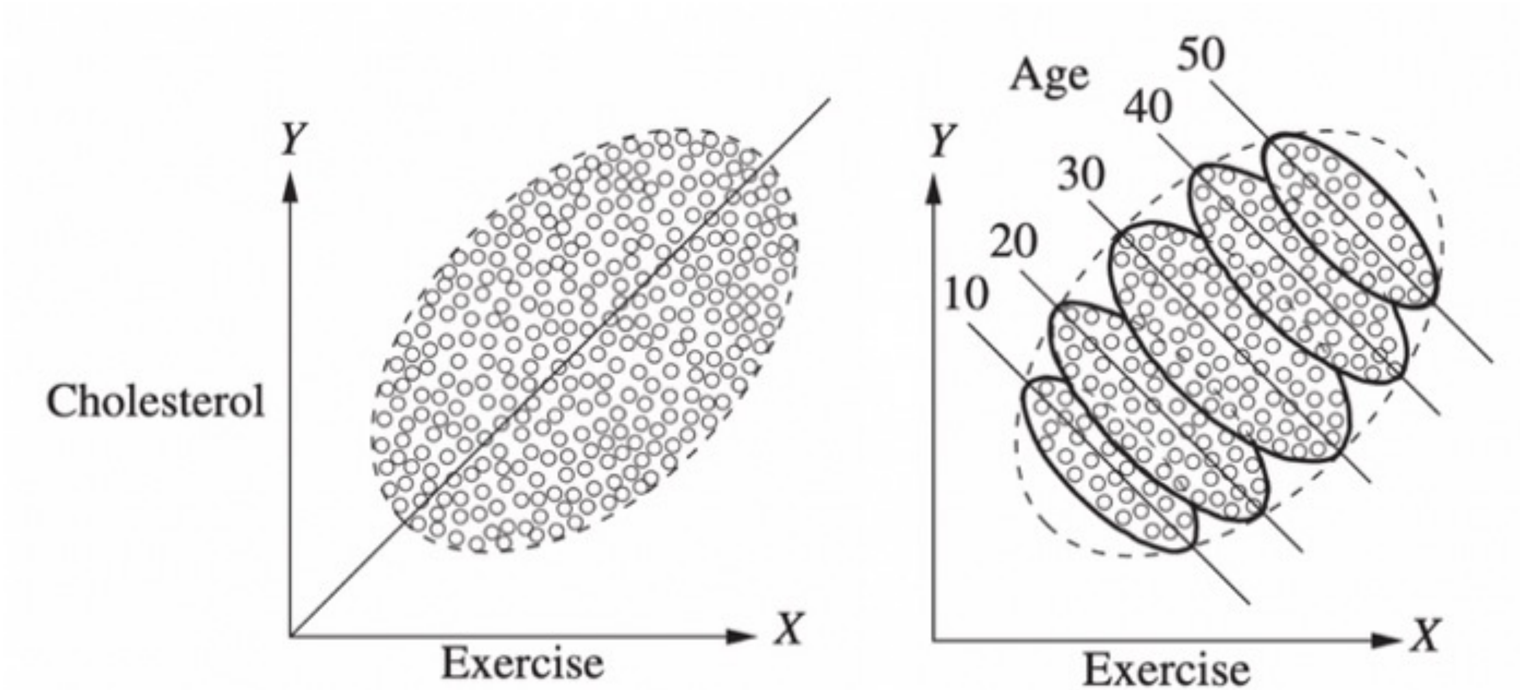


$$(Y_0, Y_1) \perp\!\!\!\perp T \mid x$$

No Ignorability



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



Weekly exercise effect on cholesterol?

(Pearl et al. 2016)

Outline for lecture

- How to recognize a causal inference problem
- Potential outcomes framework
  - Average treatment effect (ATE)
  - Conditional average treatment effect (CATE)
- **Covariate adjustment: A method for estimating ATE and CATE**
- Theory – when/why does this work?

Many methods!

Covariate adjustment

Propensity score re-weighting

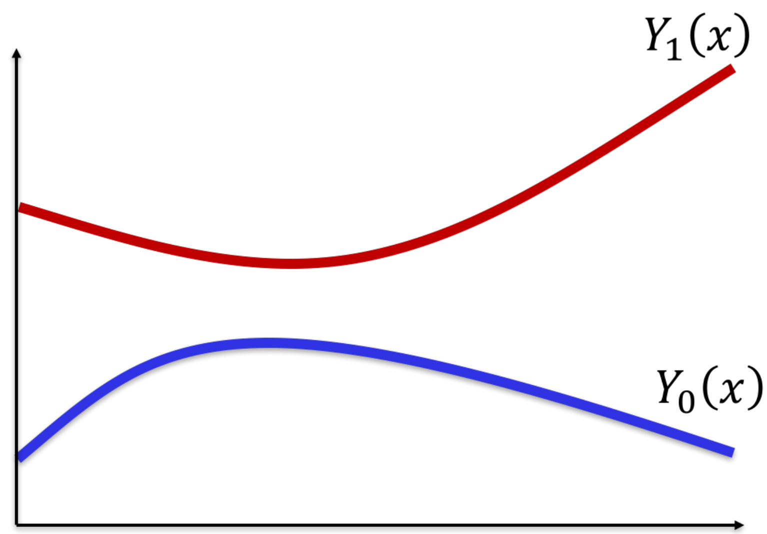
Doubly robust estimators

Matching

…

Covariate adjustment

- Explicitly model the relationship between treatment, confounders, and outcome
- Also called "Response Surface Modeling"
- Used for both CATE and ATE
- A regression problem



Covariate adjustment

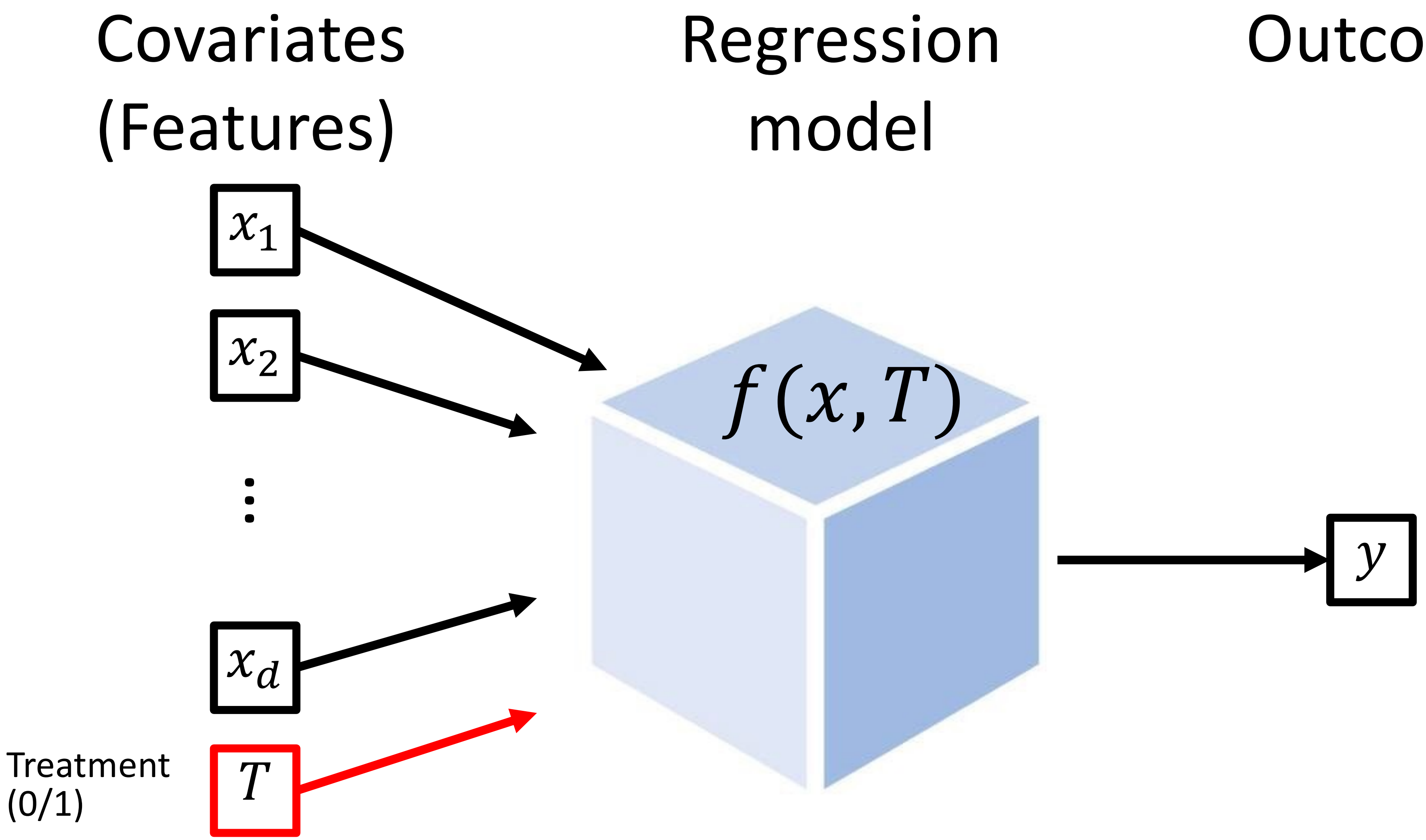
- Explicitly model the relationship between treatment, confounders, and outcome
- Under ignorability, the expected causal effect of T on Y:

$$\mathbb{E}_{x\sim p(x)}\Big[\, \mathbb{E}[Y|T=1,x] - \mathbb{E}[Y|T=0,x] \Big]$$

- Fit a model $f(x,t) \approx \mathbb{E}[Y|T=t,x]$

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

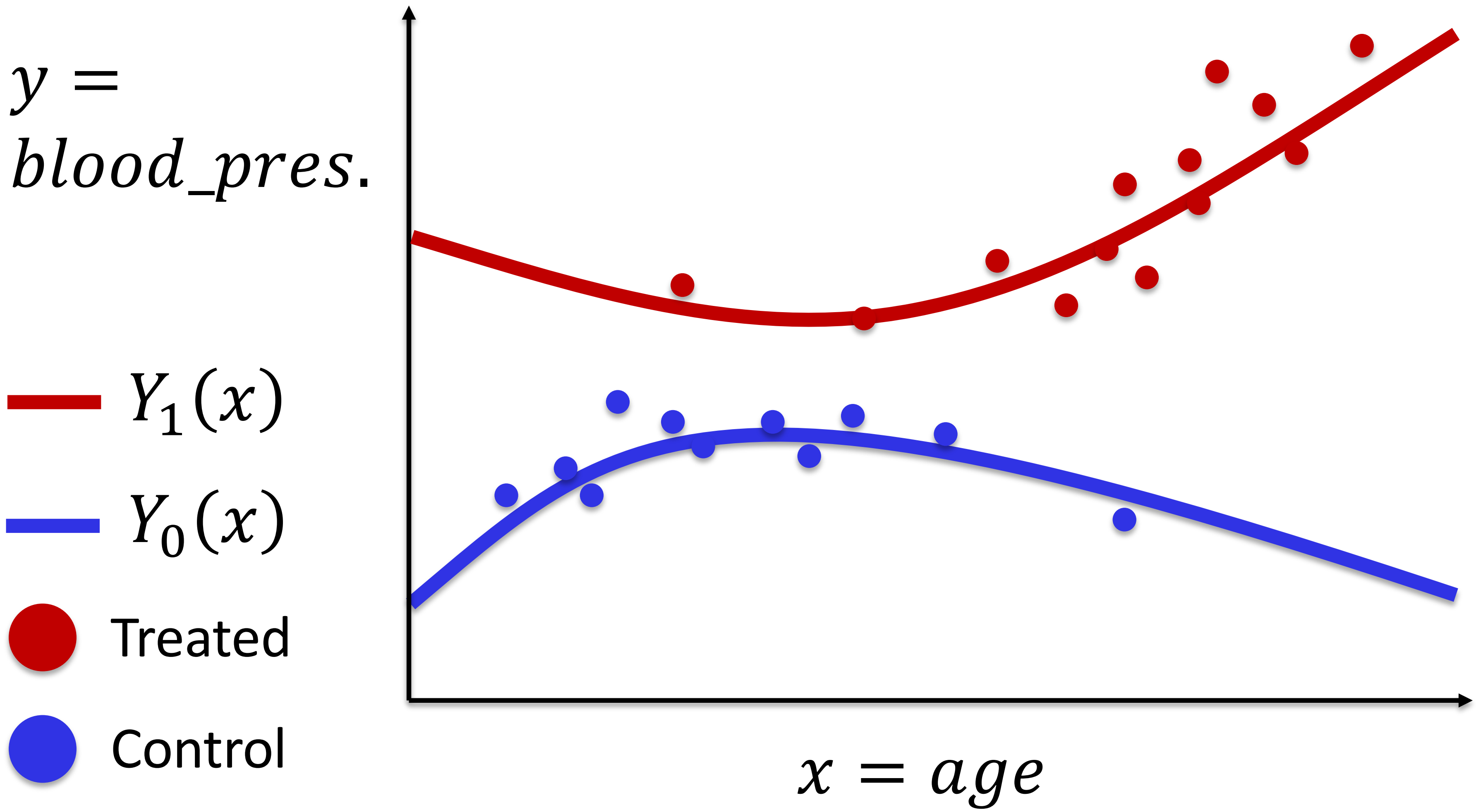
$$\widehat{CATE}(x^i) = f(x^i,1) - f(x^i,0)$$



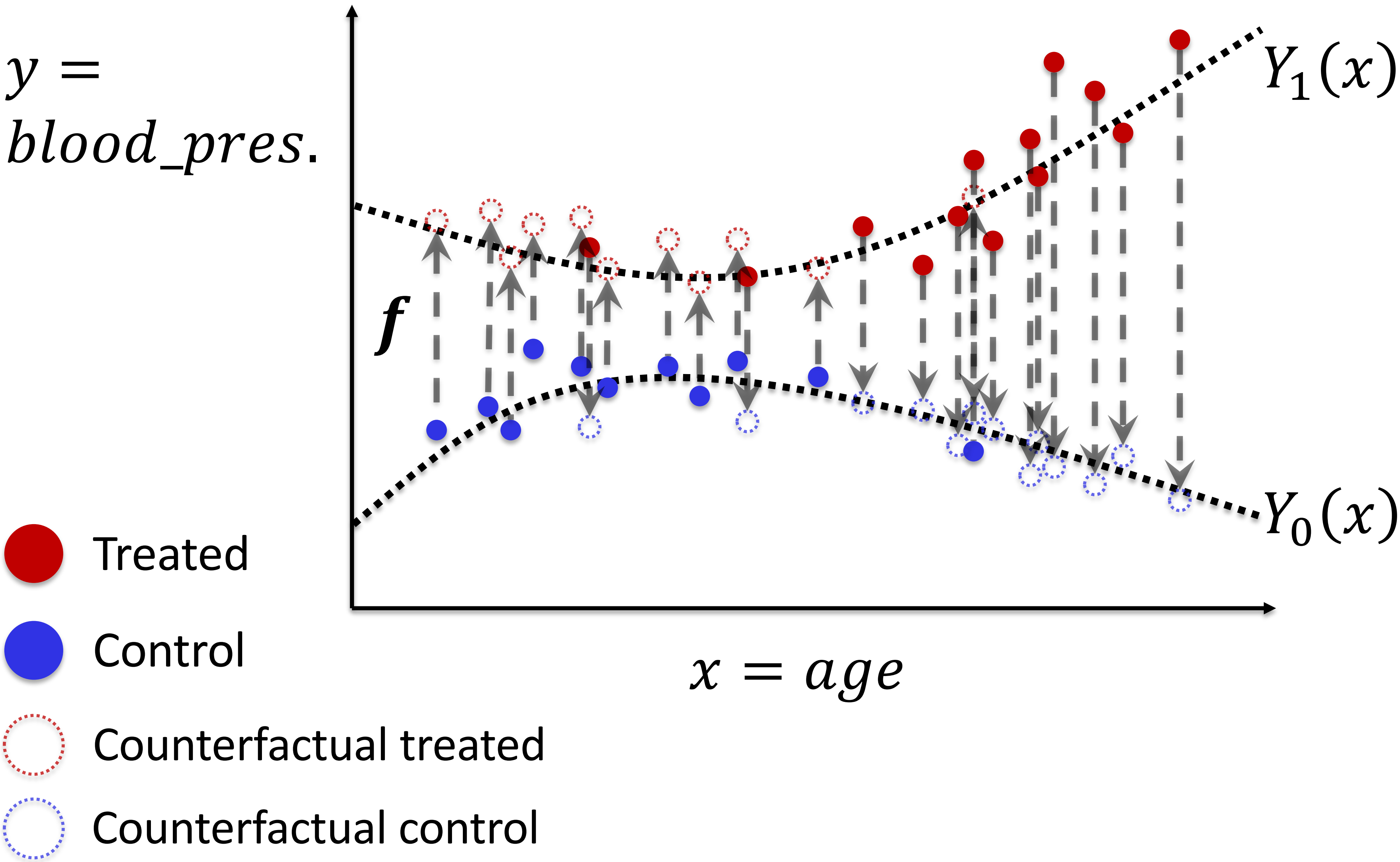
Fit a model $f(x,t) \approx \mathbb{E}[Y|X,T]$

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

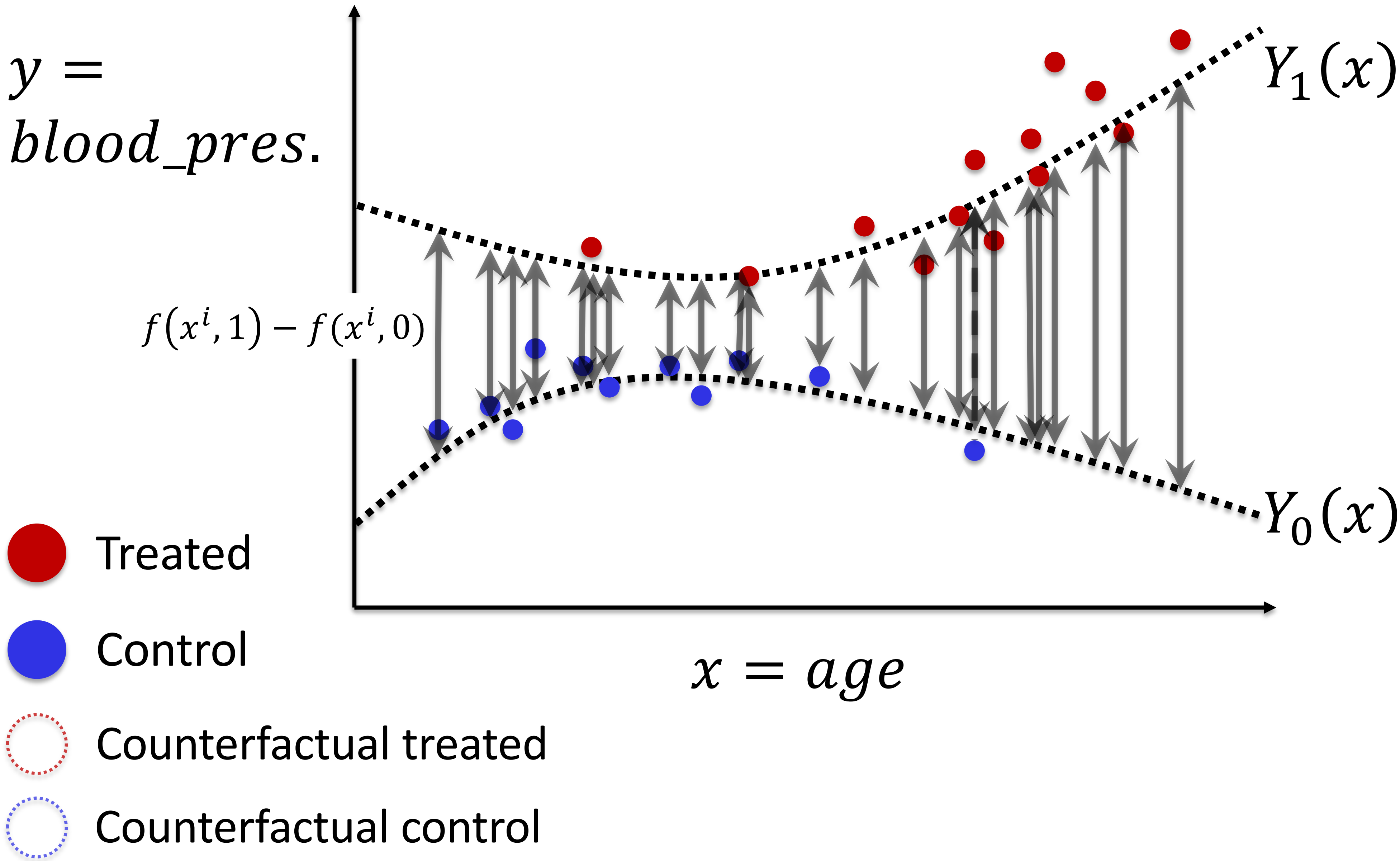
Recall this example…



Covariate adjustment (intuition): imputing the counterfactual



Covariate adjustment (reality): estimating difference in means



Outline for lecture

- How to recognize a causal inference problem
- Potential outcomes framework
  - Average treatment effect (ATE)
  - Conditional average treatment effect (CATE)
- Covariate adjustment: A method for estimating ATE and CATE
- **Theory – when/why does this work?**

Average Treatment Effect – the adjustment formula

- Assuming ignorability, we will derive the *adjustment formula* (Hernán & Robins 2010, Pearl 2009)
- Also called *(parametric) G-formula*

$$ATE = \mathbb{E}\left[Y_1 - Y_0\right] =$$

$$\mathbb{E}_{x\sim p(x)}\left[\ \mathbb{E}\left[Y_1 | x, T=1\right] - \mathbb{E}\left[Y_0 | x, T=0\right]\ \right]$$

Average Treatment Effect

The expected causal effect of T on Y:

$$ATE := \mathbb{E}\left[Y_1 - Y_0\right]$$

Average Treatment Effect

The expected causal effect of T on Y:

$$ATE := \mathbb{E}\left[Y_1 - Y_0\right]$$

law of total expectation

$$\mathbb{E}\left[Y_1\right] =$$

$$\mathbb{E}_{x\sim p(x)}\left[\mathbb{E}_{Y_1\sim p(Y_1|x)}\left[Y_1|x\right]\right] =$$

Average Treatment Effect

The expected causal effect of T on Y:

$$ATE := \mathbb{E}\left[Y_1 - Y_0\right]$$

$$\mathbb{E}\left[Y_1\right] =$$

$$\mathbb{E}_{x \sim p(x)}\left[\mathbb{E}_{Y_1 \sim p(Y_1|x)}\left[Y_1|x\right]\right] =$$

$$\mathbb{E}_{x \sim p(x)}\left[\mathbb{E}_{Y_1 \sim p(Y_1|x, T=1)}\left[Y_1|x, T=1\right]\right] =$$

ignorability

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

Average Treatment Effect

The expected causal effect of T on Y:

$$ATE := \mathbb{E}\left[Y_1 - Y_0\right]$$

$$\mathbb{E}\left[Y_1\right] =$$

$$\mathbb{E}_{x \sim p(x)}\left[\mathbb{E}_{Y_1 \sim p(Y_1|x)}\left[Y_1|x\right]\right] =$$

$$\mathbb{E}_{x \sim p(x)}\left[\mathbb{E}_{Y_1 \sim p(Y_1|x, T=1)}\left[Y_1|x, T=1\right]\right] =$$

$$\mathbb{E}_{x \sim p(x)}\left[\mathbb{E}\left[Y_1|x, T=1\right]\right]$$

shorter notation

Average Treatment Effect

The expected causal effect of T on Y:

$$ATE := \mathbb{E}\left[Y_1 - Y_0\right]$$

$$\mathbb{E}\left[Y_0\right] =$$

$$\mathbb{E}_{x \sim p(x)}\left[\mathbb{E}_{Y_0 \sim p(Y_0|x)}\left[Y_0|x\right]\right] =$$

$$\mathbb{E}_{x \sim p(x)}\left[\mathbb{E}_{Y_0 \sim p(Y_0|x, T=0)}\left[Y_0|x, T=0\right]\right] =$$

$$\mathbb{E}_{x \sim p(x)}\left[\mathbb{E}\left[Y_0|x, T = 0\right]\right]$$

The adjustment formula

Under the assumption of ignorability, we have that:

$$ATE = \mathbb{E}\left[Y_1 - Y_0\right] =$$

$$\mathbb{E}_{x\sim p(x)}\left[\ \mathbb{E}\left[Y_1|x, T=1\right] - \mathbb{E}\left[Y_0|x, T=0\right]\ \right]$$

$$\mathbb{E}\left[Y_1|x, T=1\right]$$

$$\mathbb{E}\left[Y_0|x, T=0\right]$$

} Quantities we can estimate from data

The adjustment formula

Under the assumption of ignorability, we have that:

$$ATE = \mathbb{E}\left[Y_1 - Y_0\right] =$$

$$\mathbb{E}_{x \sim p(x)}\left[\ \mathbb{E}\left[Y_1 | x, T = 1\right] - \mathbb{E}\left[Y_0 | x, T = 0\right]\ \right]$$

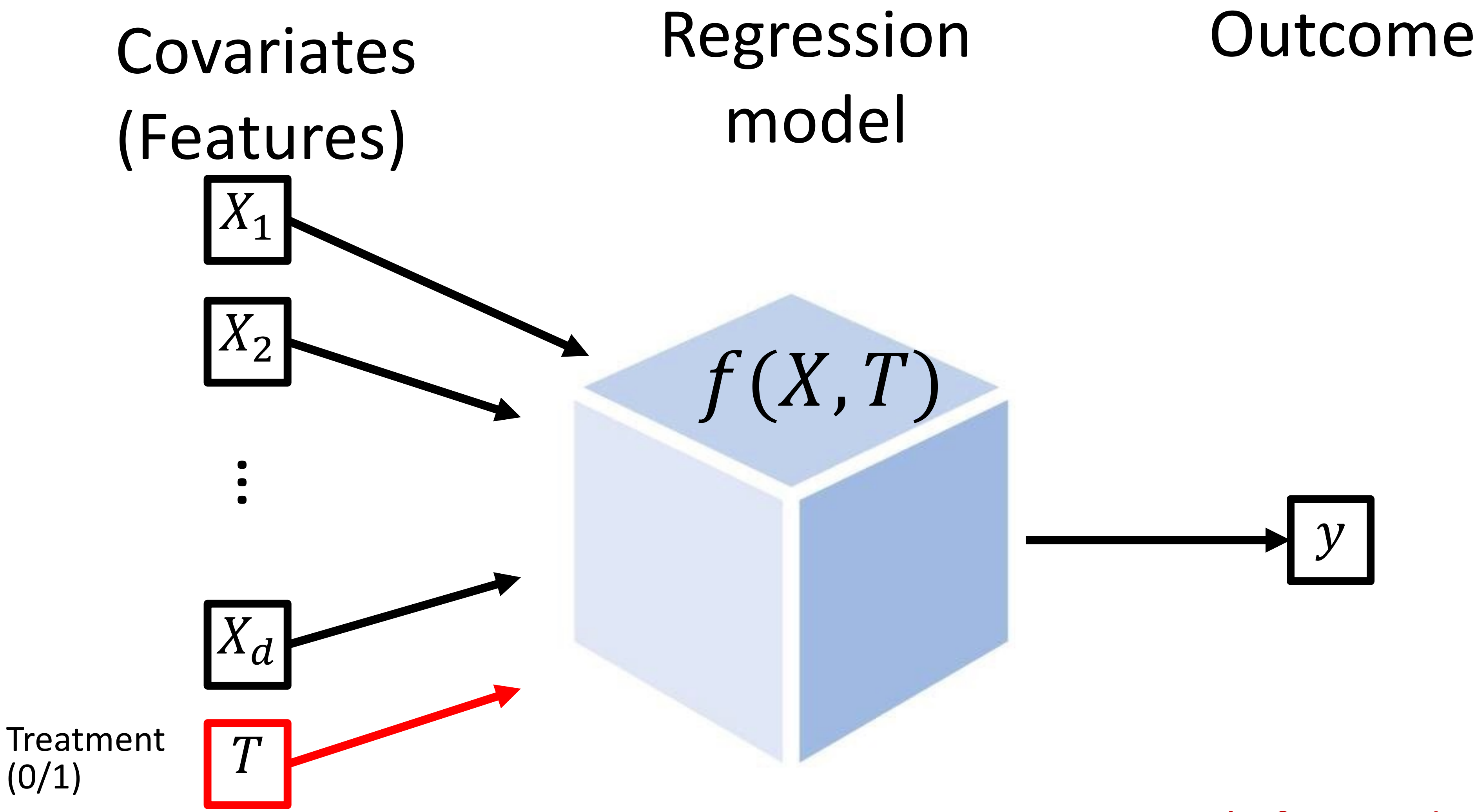
$$\mathbb{E}\left[Y_0 | x, T = 1\right]$$

$$\mathbb{E}\left[Y_1 | x, T = 0\right]$$

$$\mathbb{E}\left[Y_0 | x\right]$$

$$\mathbb{E}\left[Y_1 | x\right]$$

Quantities we *cannot* directly estimate from data



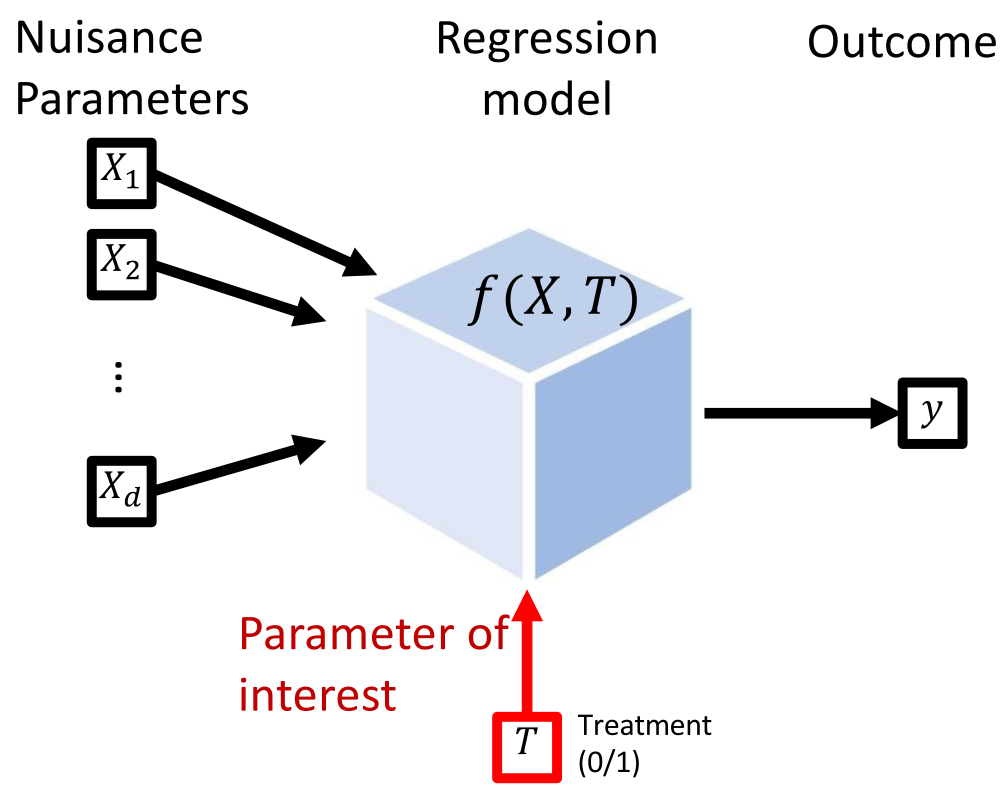
Fit a model $f(x,t) \approx \mathbb{E}[Y|X,T]$

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

Straightforward application of machine learning, right?

Covariate adjustment relies on being able to extrapolate correctly

- Correctly estimating the (C)ATE depends on being able to tell the difference between T=1 and T=0
- Either need to make strong parametric assumptions about the form of $\mathbb{E}[Y|X,T]$, *or*
- Make no assumptions about form of $\mathbb{E}[Y|X,T]$ (use black-box ML method for *f*); instead, make assumptions about *p(t | x)*



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

Summary

- One approach to use machine learning for causal inference
  - Predict outcome given features and treatment, then use resulting model to impute counterfactuals (*covariate adjustment*)
- Consistency of estimates depend on:
  - Causal graph being correct (i.e., no **unobserved confounding**)
  - Identifiability of causal effect; more on this in Thursday's lecture

References

- Recent work from ML community:
  https://sites.google.com/view/nips2018causallearning/ and
  http://tripods.cis.cornell.edu/neurips19_causalml/
- Recent book on causal inference by Miguel Hernan and Jamie Robins:
  https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
  Recent book on causal inference by Jonas Peters, Dominik Janzing and Bernhard Schölkopf:
  https://mitpress.mit.edu/books/elements-causal-inference
  (download PDF for free on left: “Open Access Title”)
- Examples of recent papers in this research field:
  https://arxiv.org/abs/1906.02120
  https://arxiv.org/abs/1705.08821
  https://arxiv.org/abs/1510.04342
  https://arxiv.org/abs/1810.02894