Machine Learning for Healthcare

6.871, HST.956

Lecture 14: Causal Inference Part 1

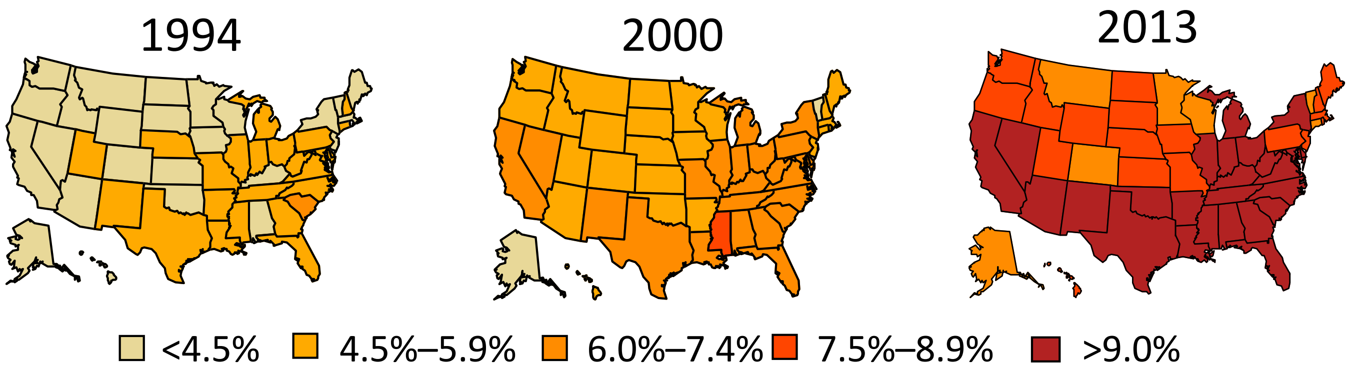
David Sontag

CSAIL

INSTITUTE FOR MEDICAL ENGINEERING & SCIENCE

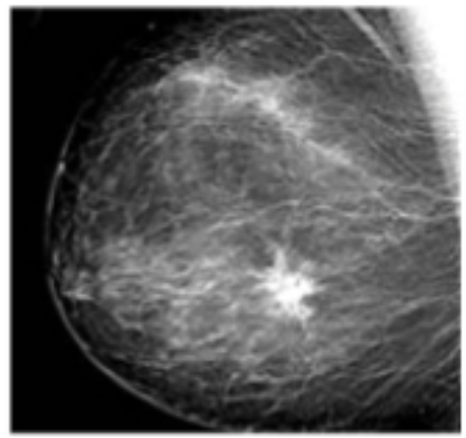
HEALTH SCIENCES & TECHNOLOGY

Does gastric bypass surgery prevent onset of diabetes?

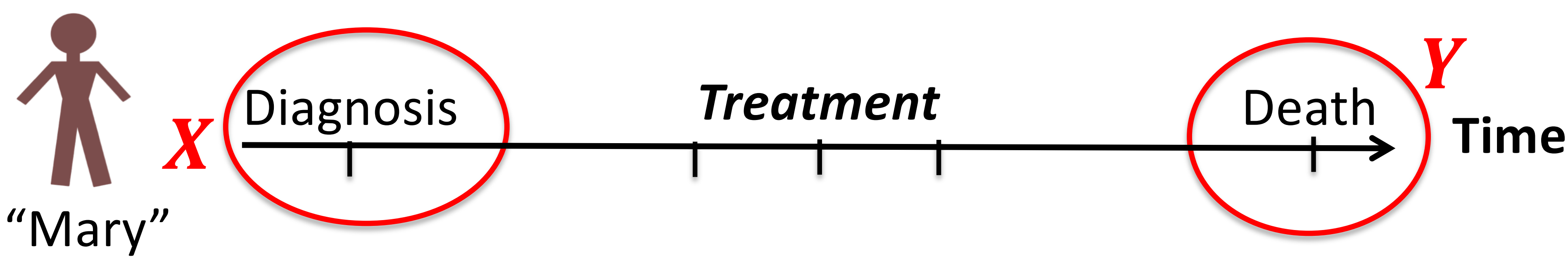


- In Lecture 4 & PS2 we used machine learning for early detection of Type 2 diabetes
- Health system doesn't want to know how to predict diabetes – they want to know how to *prevent it*
- Gastric bypass surgery is the highest negative weight (9th most predictive feature)
  - Does this mean it would be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

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

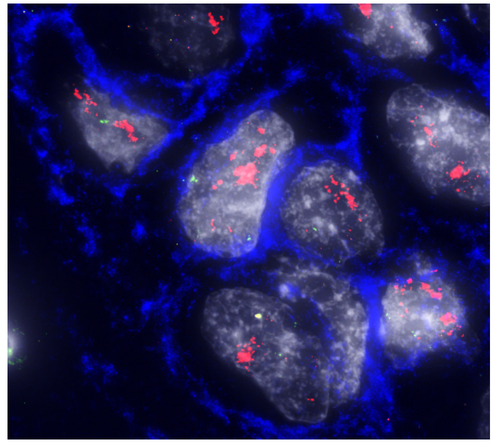


- Such predictive models widely used to stage patients. Should we initiate treatment? How aggressive?
- What could go wrong if we trained to predict survival, and then used to guide patient care?



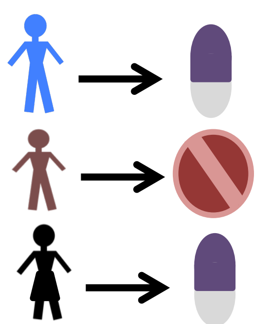
A long survival time may be because of treatment!

What treatment should we give this patient?



Expansion pathology
(image from Andy Beck)

- People respond differently to treatment
- Goal: use data from other patients and their journeys to guide future treatment decisions
- What could go wrong if we trained to predict (past) treatment decisions?



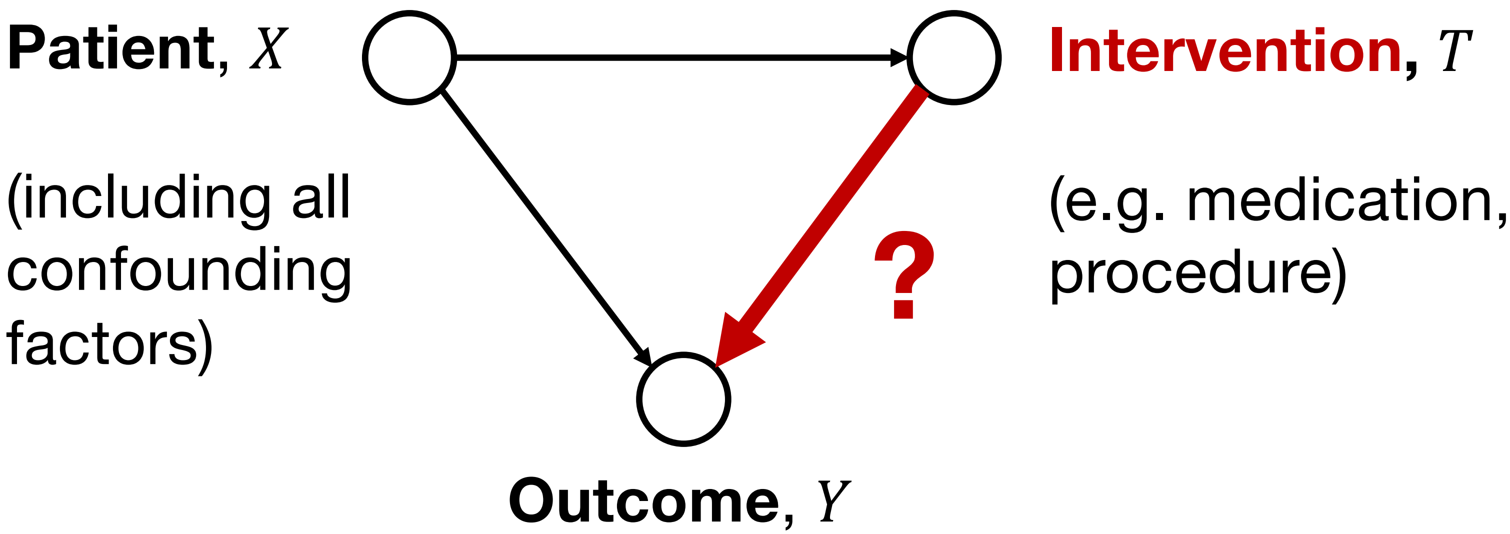
"David" → Treatment A
"John" → Treatment B
"Juana" → Treatment A

Best this can do is match current medical practice!

Does smoking cause lung cancer?

- Doing a randomized control trial is unethical
- Could we simply answer this question by comparing Pr(lung cancer | smoker) vs Pr(lung cancer | nonsmoker)?
- No! Answering such questions from observational data is difficult because of *confounding*

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

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:

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

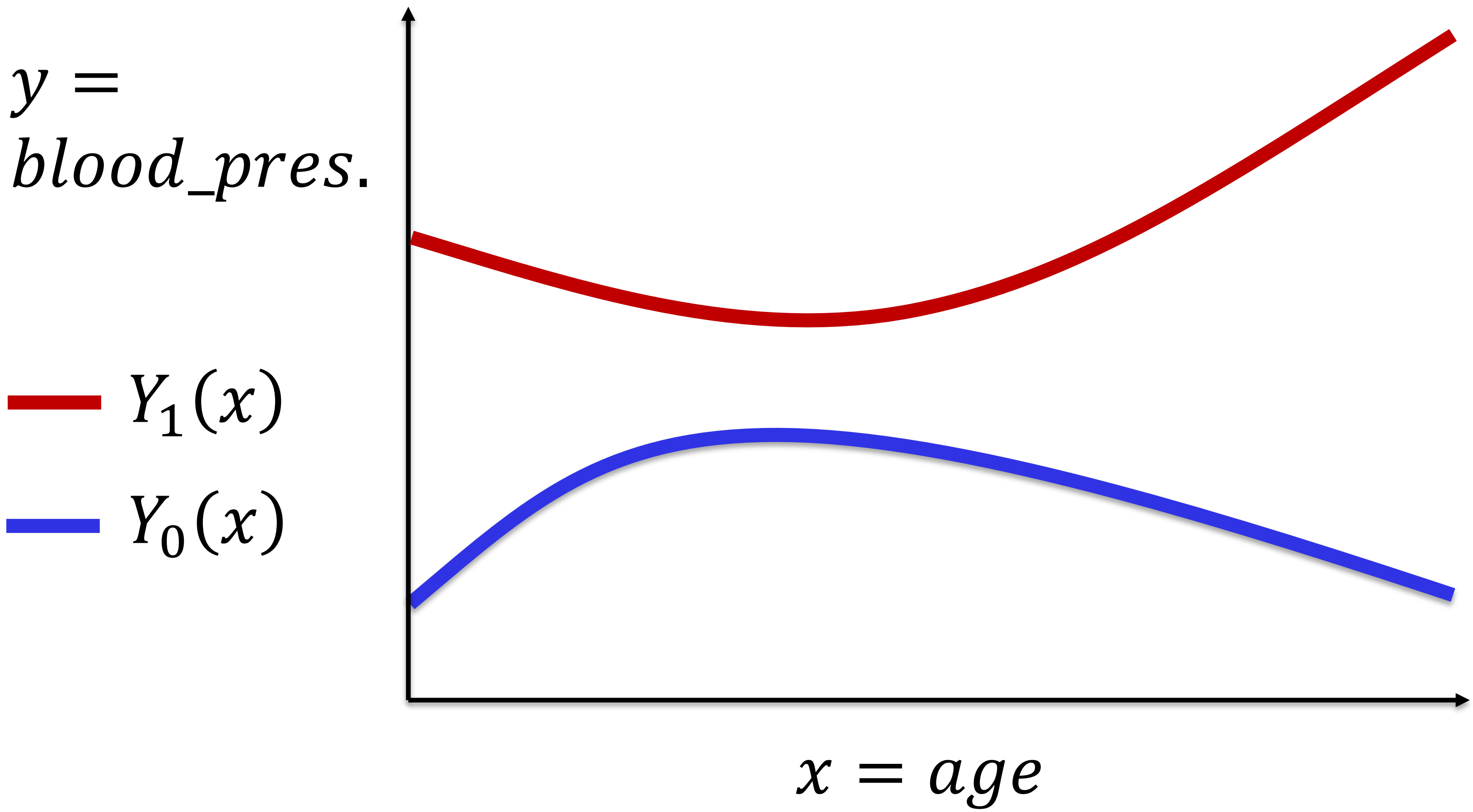
- Unobserved counterfactual outcome:

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

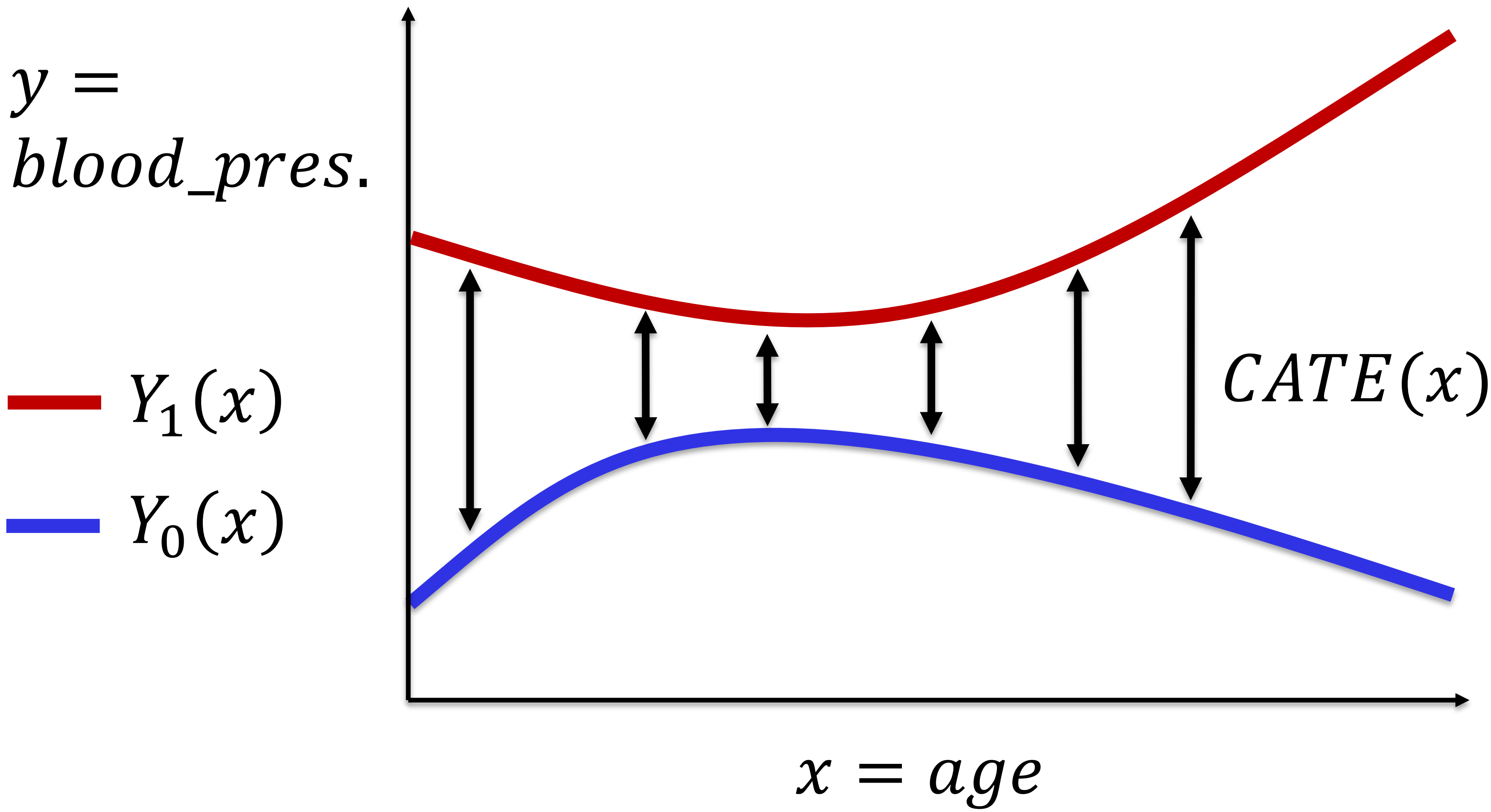
"The fundamental problem of causal inference"

We only ever observe one of the two outcomes

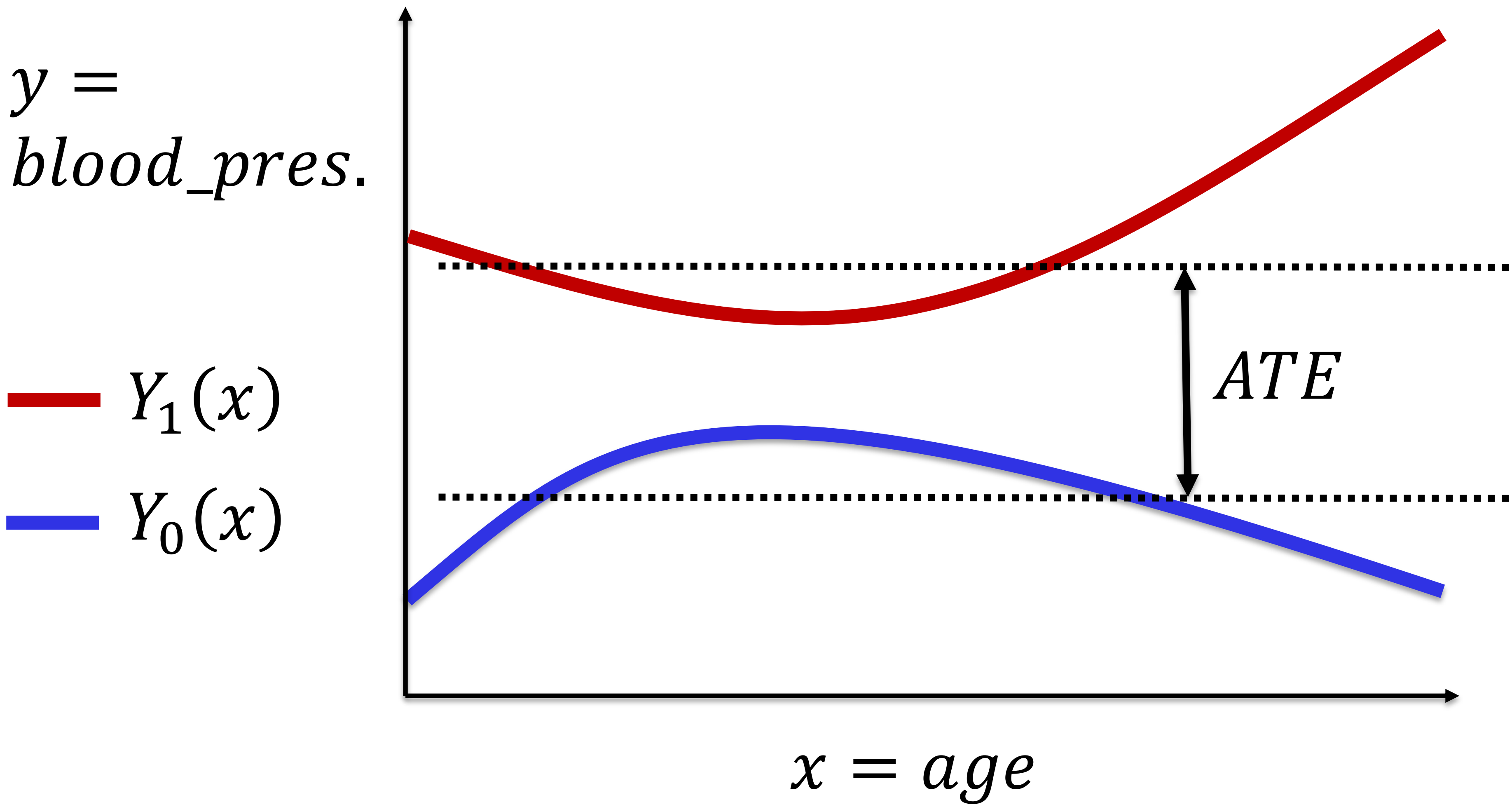
Example – Blood pressure and age



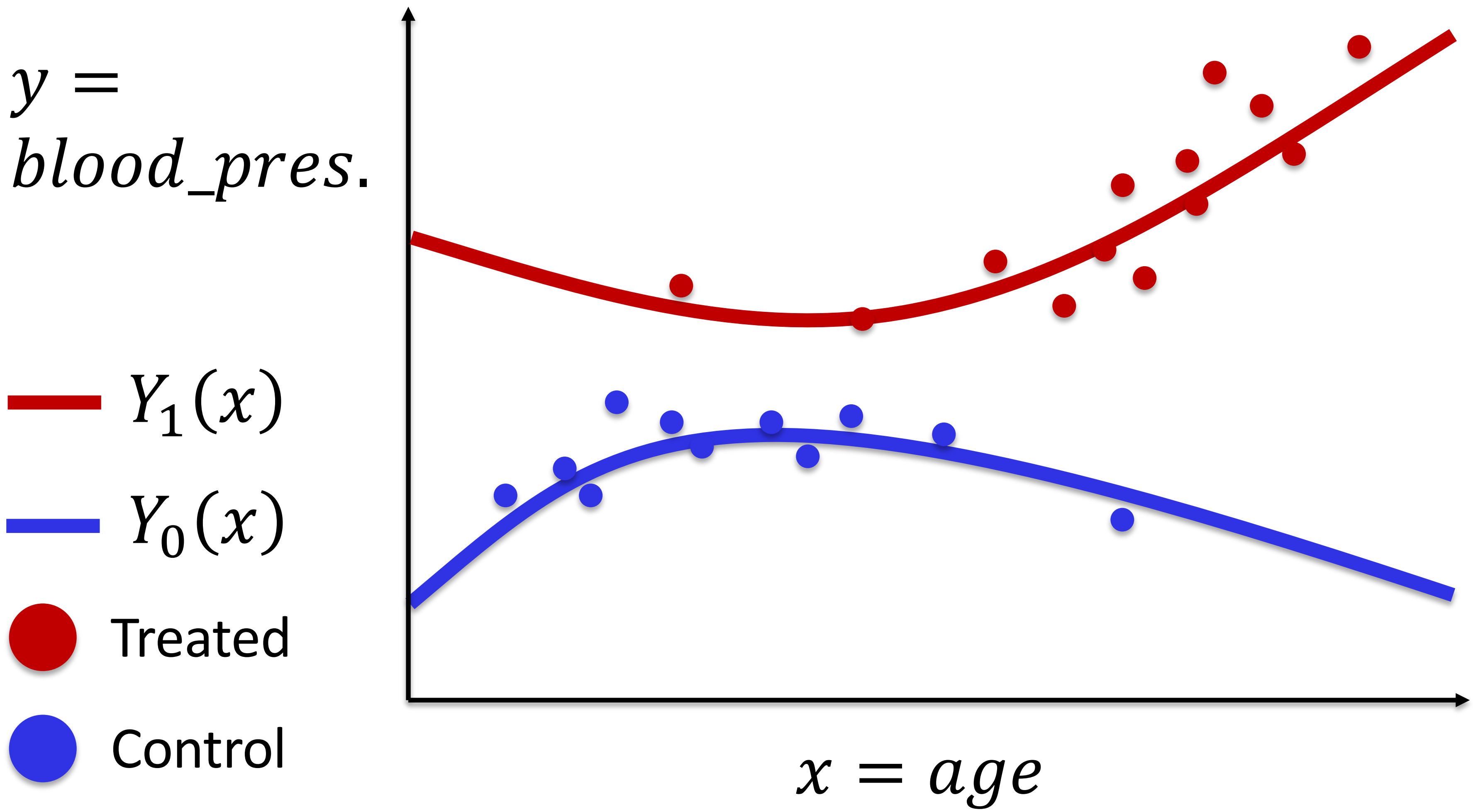
Blood pressure and age



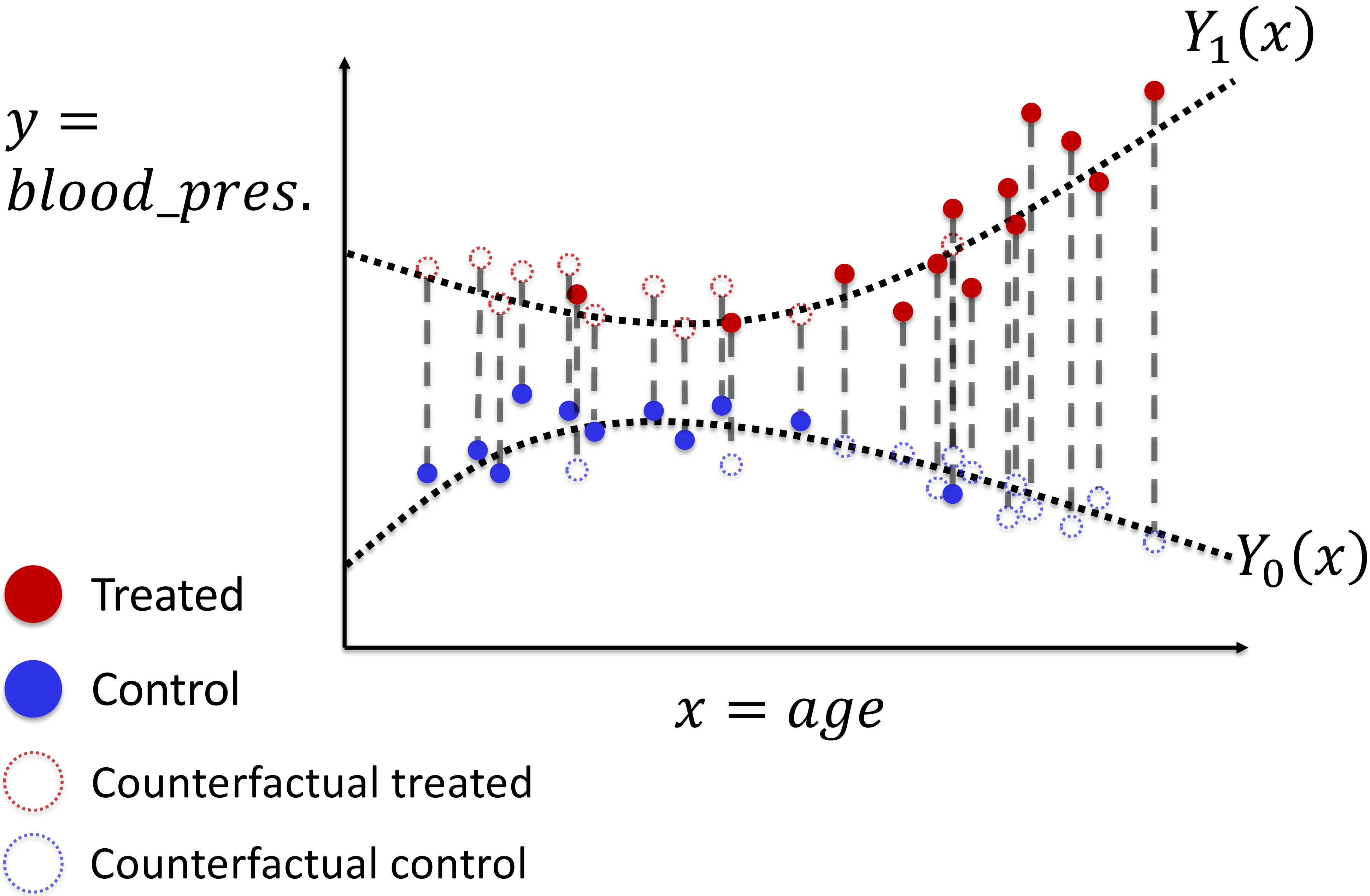
Blood pressure and age



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)

(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) =
7.875 - 7.125 = 0.75

mean(sugar|*had they received* B) –
mean(sugar|*had they received* A) =
7.125 - 7.875 = -0.75

(Example from Uri Shalit)

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

Typical assumption – no unmeasured confounders

Y_0, Y_1: potential outcomes for control and treated

x: unit covariates (features)

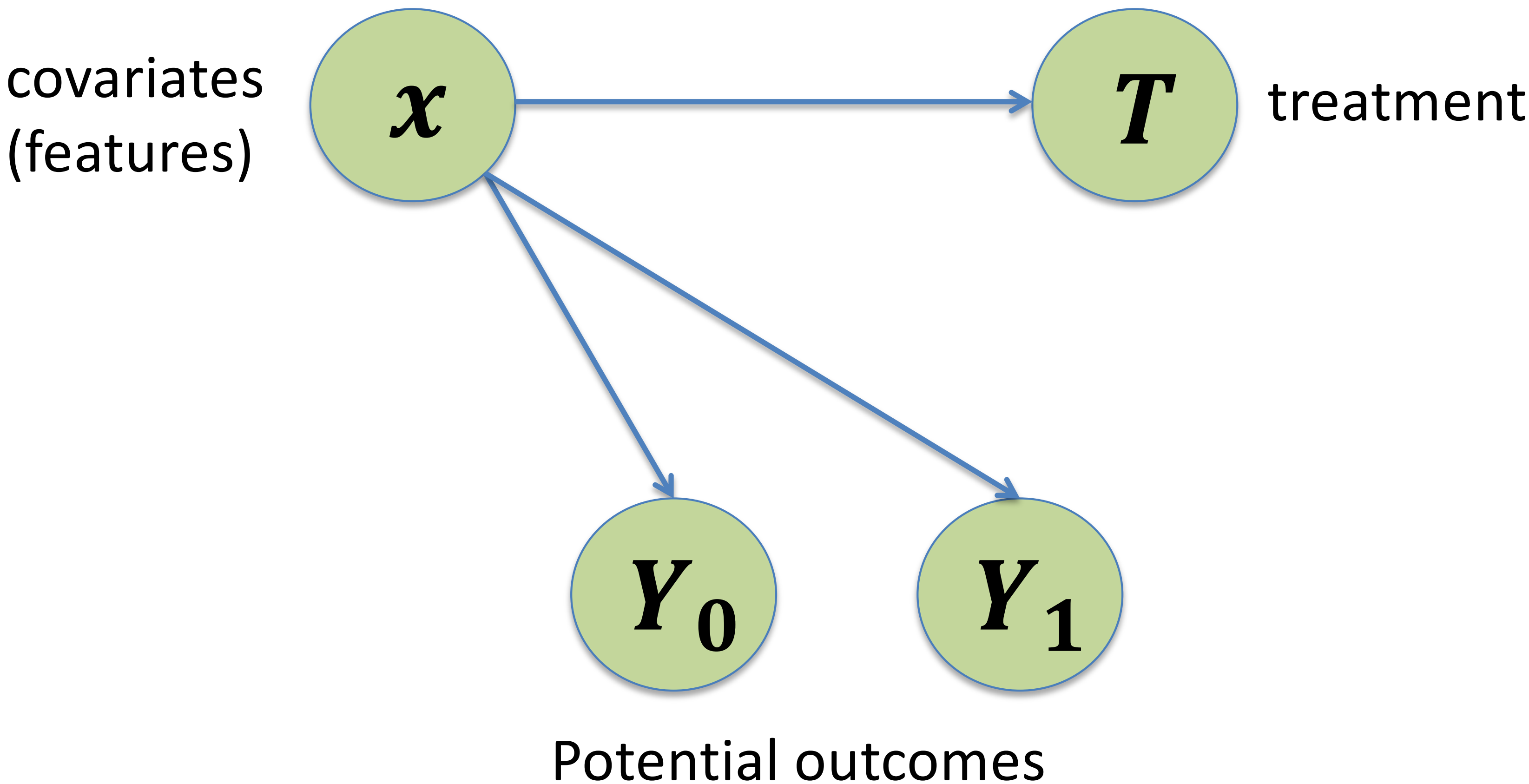
T: treatment assignment

We assume:

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

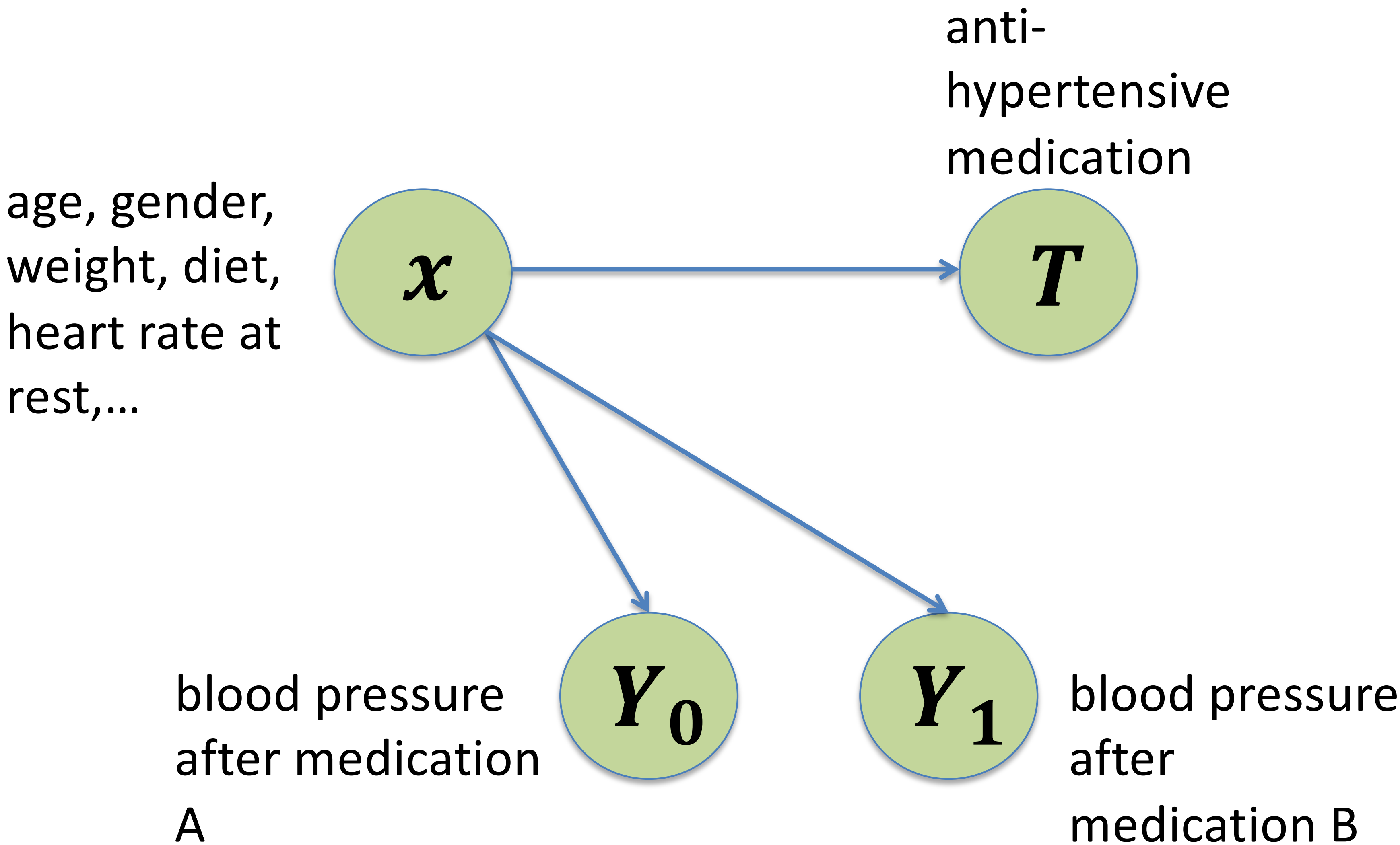
Ignorability

Ignorability



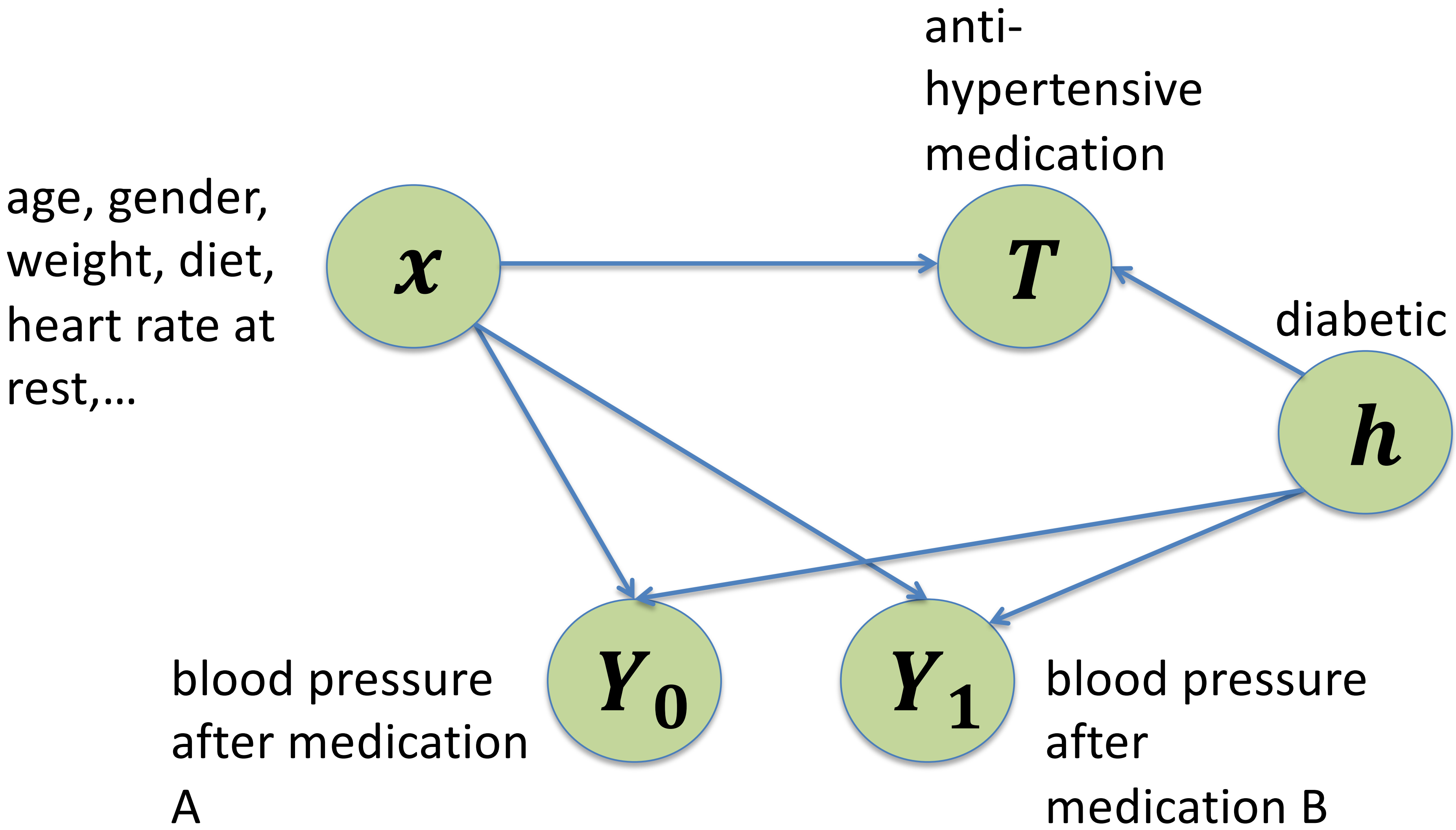
$$(Y_0, Y_1) \perp\!\!\!\perp T \mid x$$

Ignorability



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

No Ignorability



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

Typical assumption – common support

Y_0, Y_1: potential outcomes for control and treated

x: unit covariates (features)

T: treatment assignment

We assume:

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

Framing the question

1. Where could we go to for data to answer these questions?
2. What should **X**, T, and Y be to satisfy ignorability?
3. What is the specific causal inference question that we are interested in?
4. Are you worried about common support?

Outline for lecture

- How to recognize a causal inference problem
- Potential outcomes framework
  - Average treatment effect (ATE)
  - Conditional average treatment effect (CATE)
- **Algorithms for estimating ATE and CATE**

Average Treatment Effect

The expected causal effect of T on Y:

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

Average Treatment Effect – the adjustment formula

- Assuming ignorability, we will derive the *adjustment formula* (Hernán & Robins 2010, Pearl 2009)
- The adjustment formula is extremely useful in causal inference
- Also called *G-formula*

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:

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

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

ignorability

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

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

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

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:

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

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

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

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

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

The adjustment formula

Under the assumption of ignorability, we have that:

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

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

$$\left.\begin{array}{l} \mathbb{E}\left[Y_1 | x, T=1\right] \\ \mathbb{E}\left[Y_0 | x, T=0\right] \end{array}\right\} \text{Quantities we can estimate from data}$$

The adjustment formula

Under the assumption of ignorability, we have that:

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

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

$$\left.\begin{array}{l} \mathbb{E}\left[Y_0 | x, T = 1\right] \\ \mathbb{E}\left[Y_1 | x, T = 0\right] \\ \mathbb{E}\left[Y_0 | x\right] \\ \mathbb{E}\left[Y_1 | x\right] \end{array}\right\}$$

Quantities we *cannot* directly estimate from data

The adjustment formula

Under the assumption of ignorability, we have that:

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

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

$$\left.\begin{array}{l} \mathbb{E}[Y_1|x, T=1] \\ \mathbb{E}[Y_0|x, T=0] \end{array}\right\}$$

Quantities we can estimate from data

Empirically we have samples from $p(x|T=1)$ or $p(x|T=0)$. *Extrapolate* to $p(x)$

Many methods!

Covariate adjustment

Propensity score re-weighting

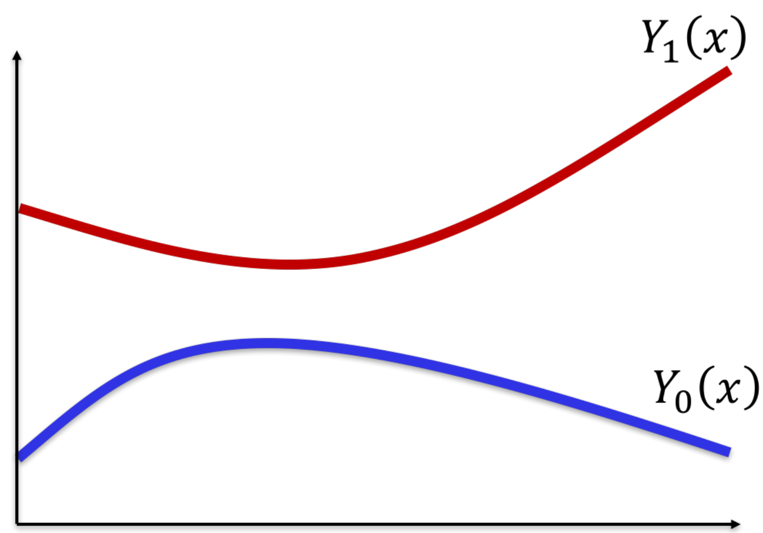
Doubly robust estimators

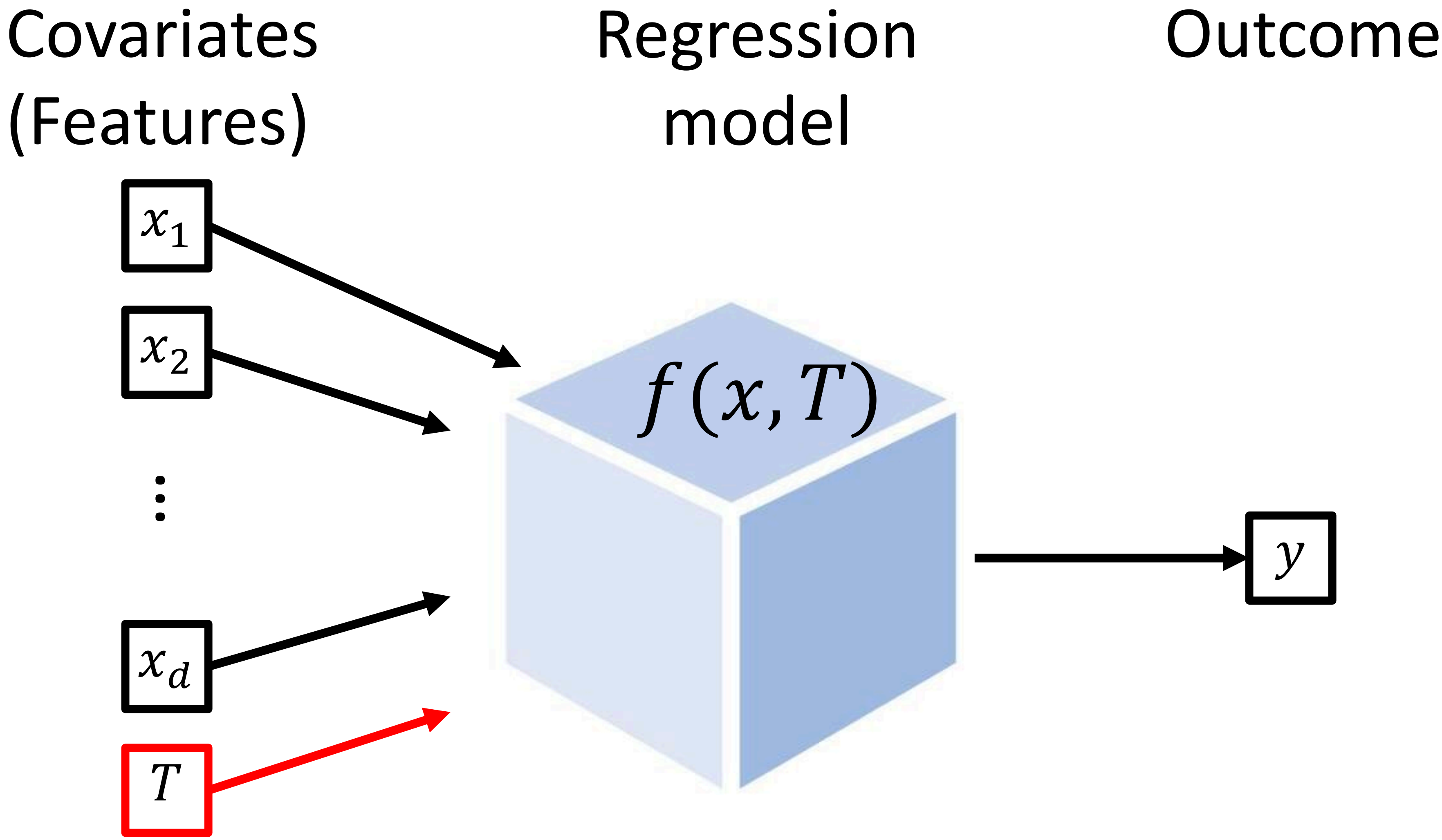
Matching

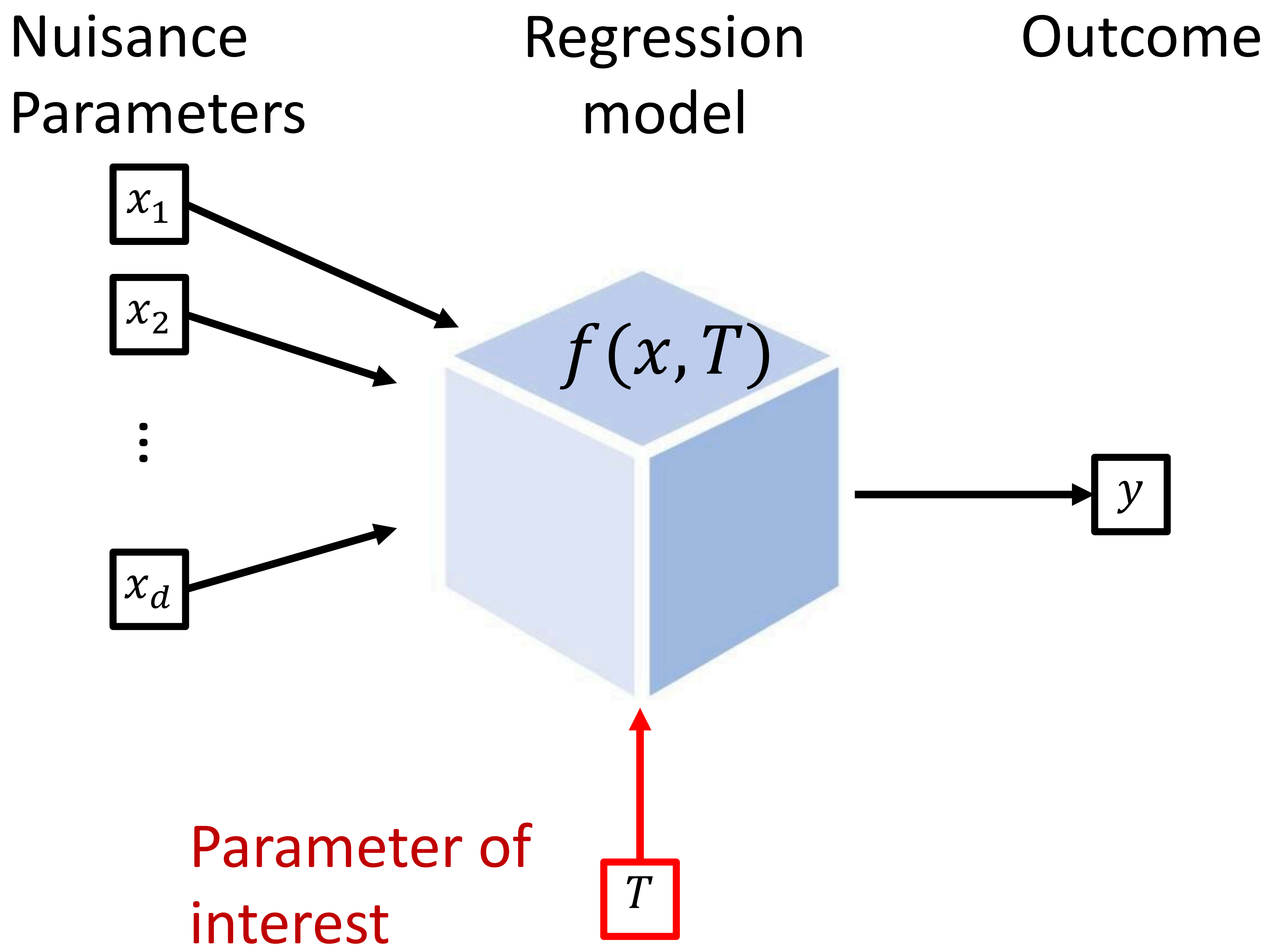
…

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




Covariates
(Features)
Regression
model
Outcome
x_1
x_2
$\vdots$
x_d
T
$f(x,T)$
y


Nuisance Parameters
Regression model
Outcome
x_1
x_2
x_d
$f(x, T)$
y
Parameter of interest
T

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 $f(x,t) \approx \mathbb{E}[Y_t|T=t,x]$

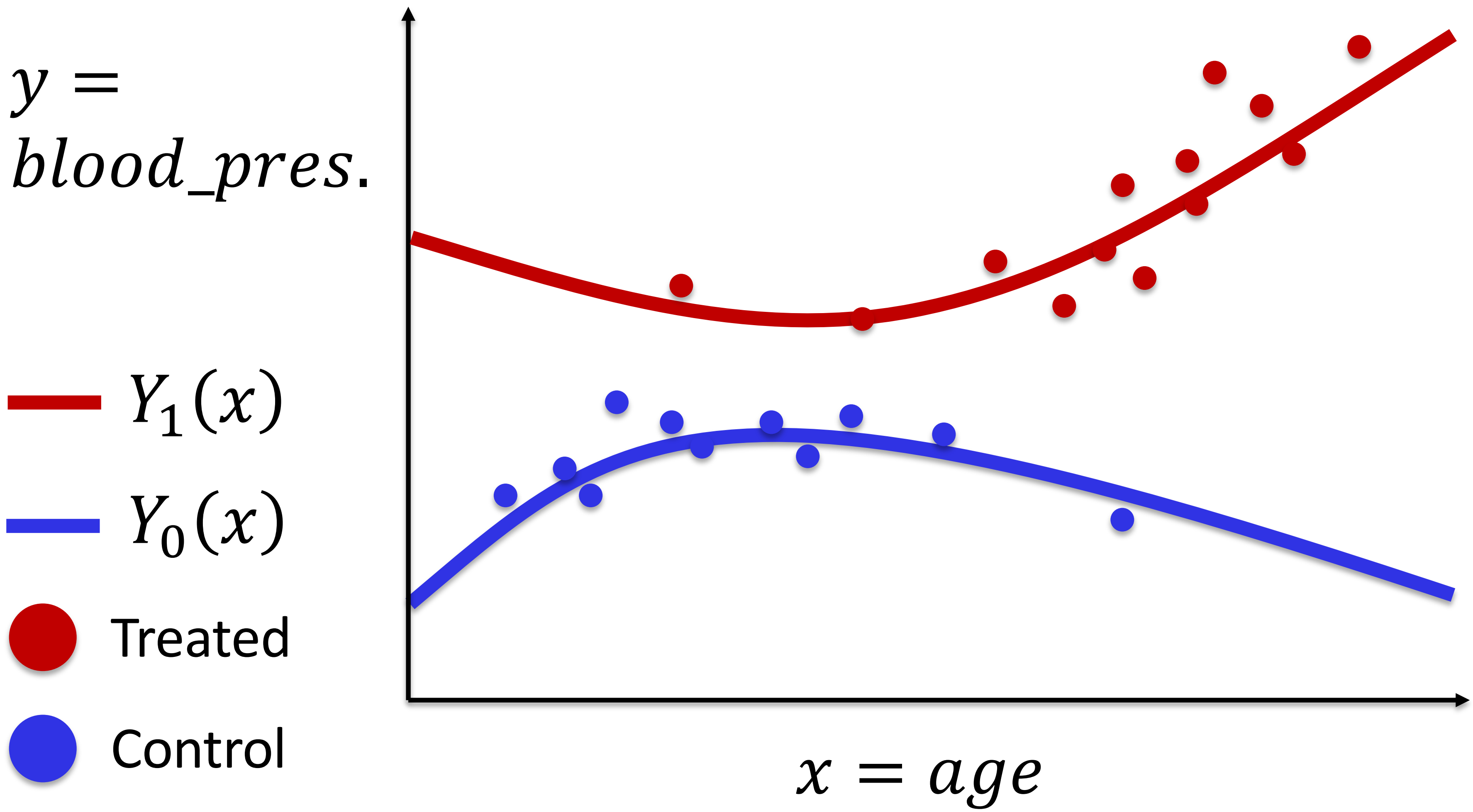
$$\widehat{ATE} = \frac{1}{n}\sum_{i=1}^{n} f(x_i,1) - f(x_i,0)$$

Covariate adjustment (parametric g-formula)

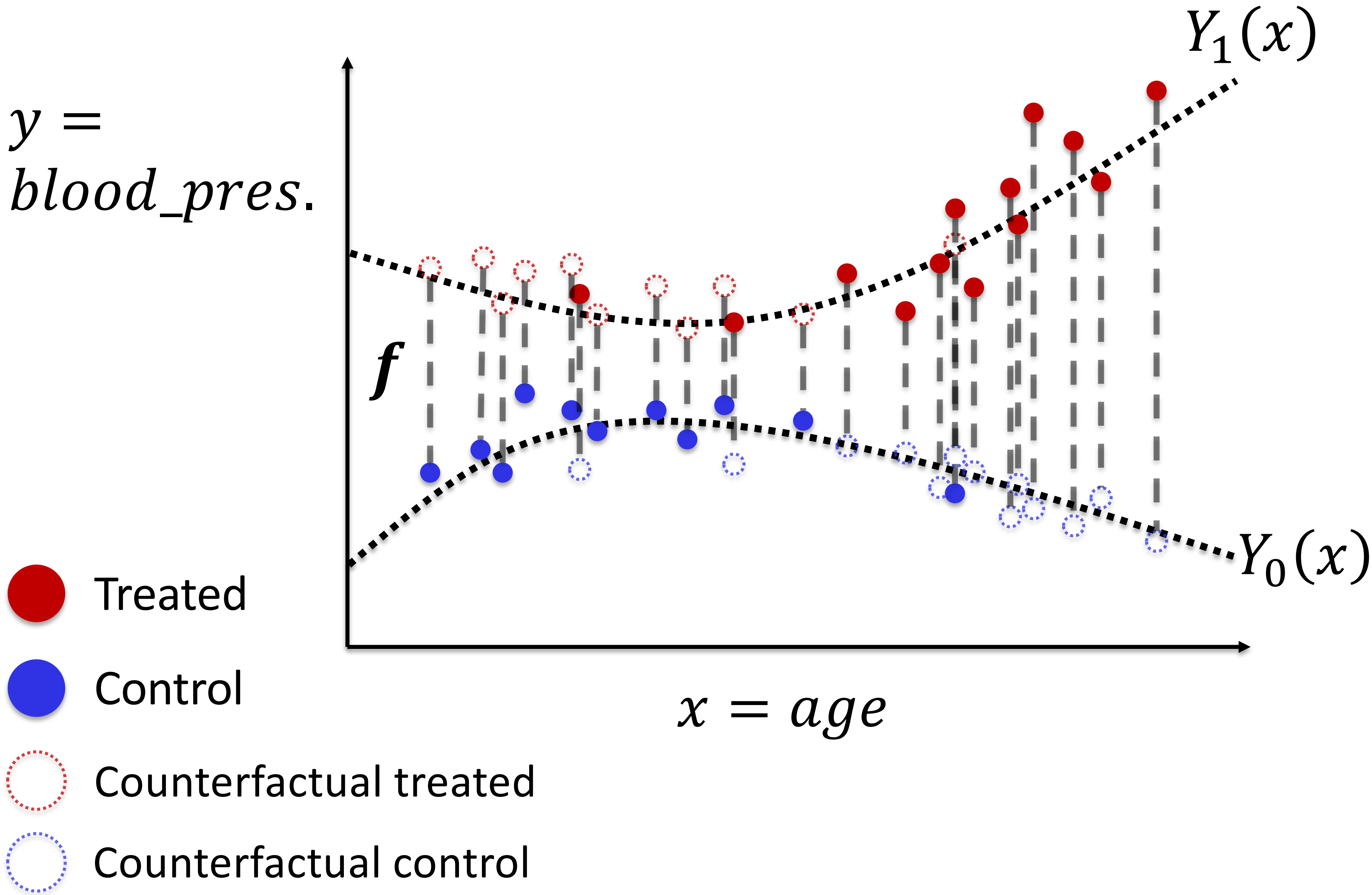
- Explicitly model the relationship between treatment, confounders, and outcome
- Under ignorability, the expected causal effect of T on Y:
  $\mathbb{E}_{x\sim p(x)}\big[\,\mathbb{E}[Y_1|T=1,x]-\mathbb{E}[Y_0|T=0,x]\big]$
- Fit a model $f(x,t)\approx\mathbb{E}[Y_t|T=t,x]$

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

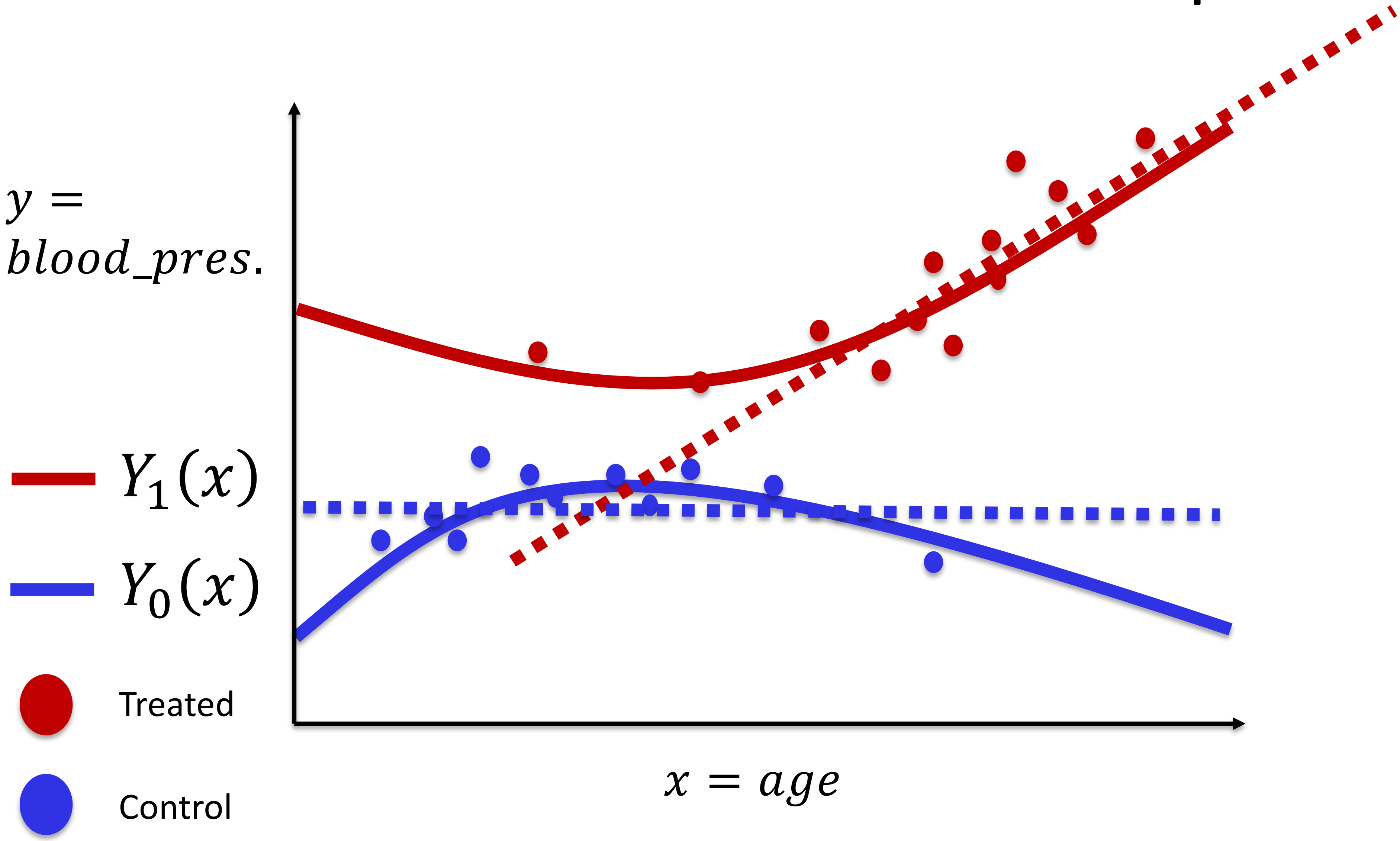
Covariate adjustment



Covariate adjustment



Example of how covariate adjustment fails when there is no overlap



Summary

- One approaches 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 (i.e., overlap)
  - Nonparametric regression is used (or correctly specified model); 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