LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

LMU MUNICH SCHOOL OF MANAGEMENT

Munich Center for Machine Learning

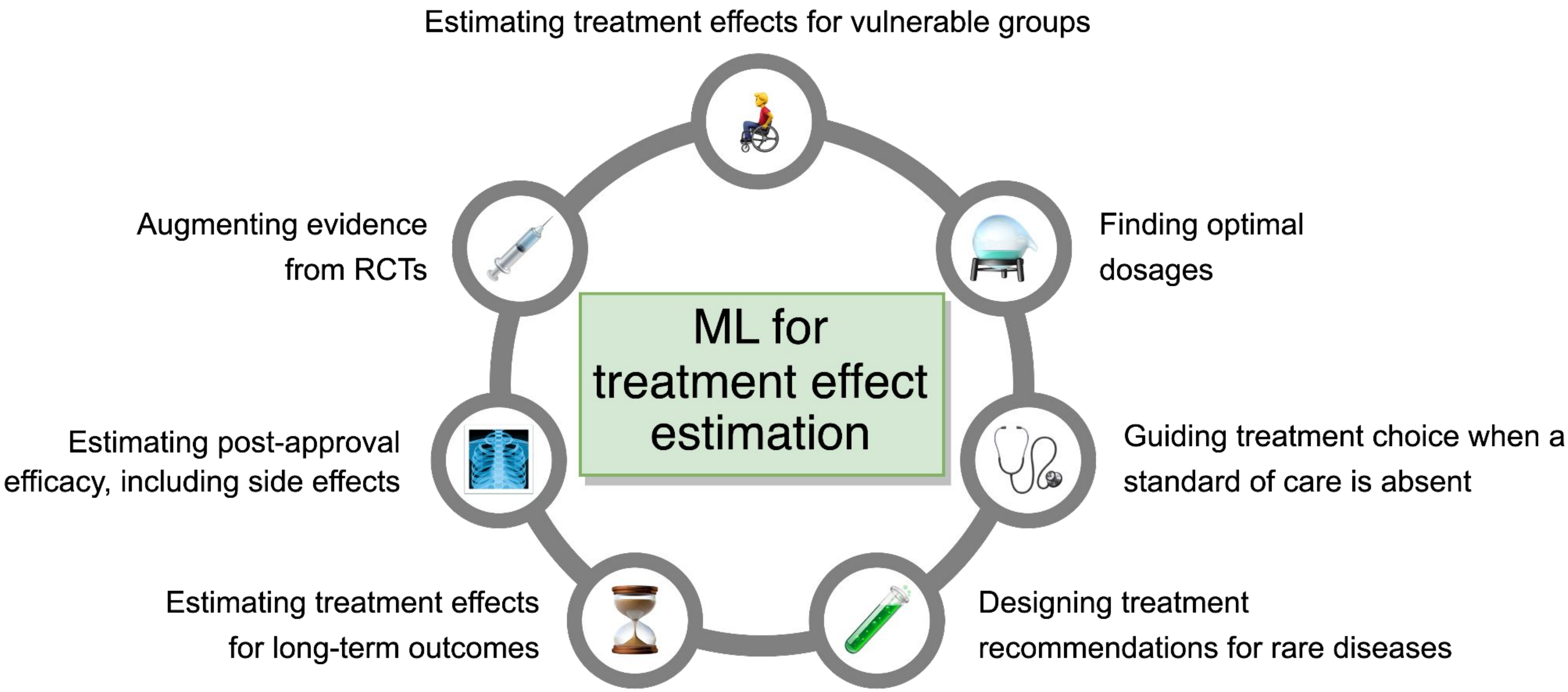
Causal ML for predicting treatment outcomes

Valentyn Melnychuk

Institute of AI in Management
LMU Munich
https://www.ai.bwl.lmu.de

VISION

Promises of Causal ML

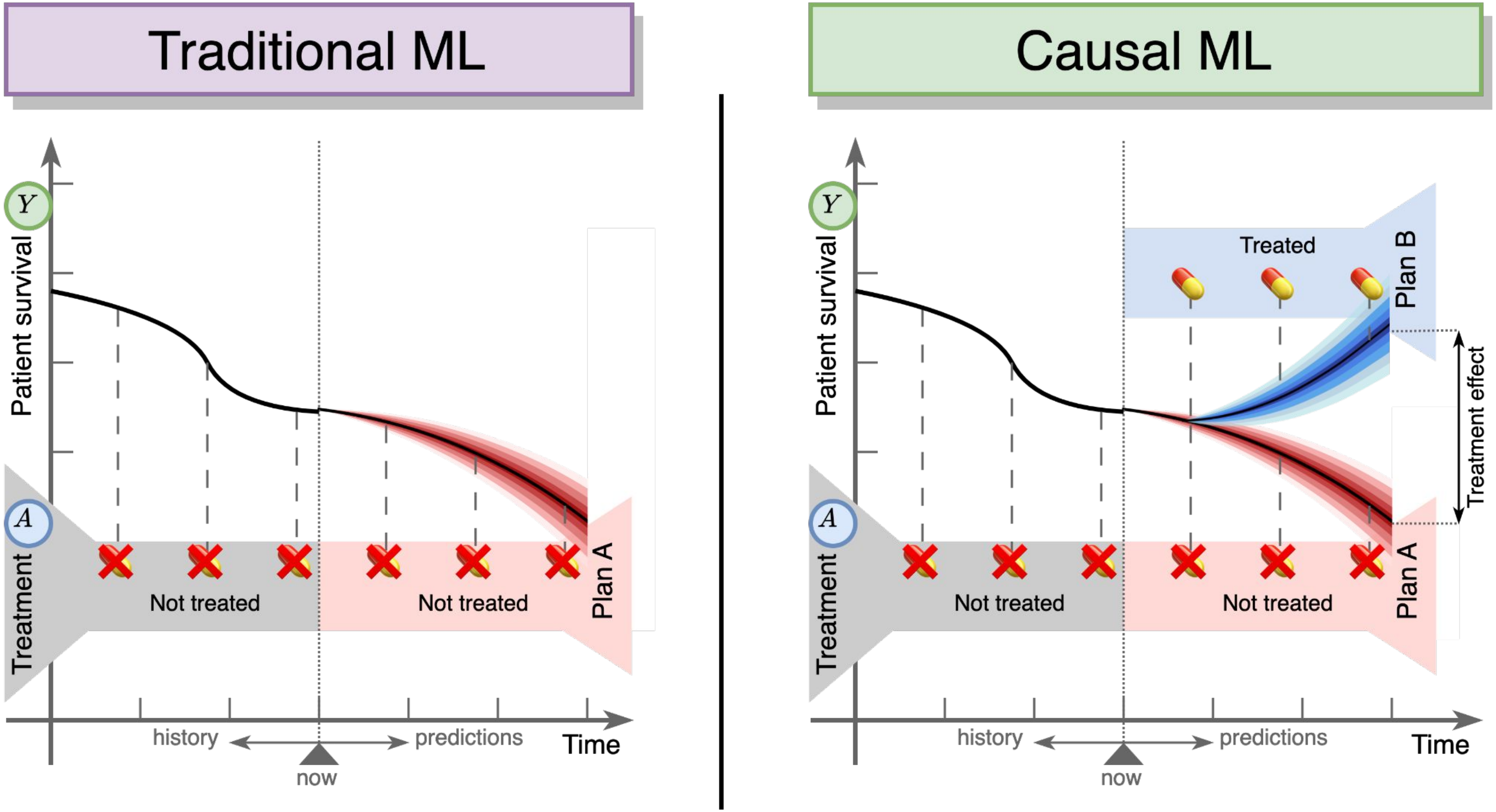


Why do we need Causal ML in medicine?

Reference:
Feuerriegel, S., Frauen, D., Melnychuk, V., Schweisthal, J., Hess, K., Curth, A., Bauer, S., Kilbertus, N., Kohane, I.S. and van der Schaar, M., 2024. **Causal machine learning for predicting treatment outcomes**. Nature Medicine, 30(4), pp.958-968.

VISION

Moving from diagnostics to therapeutics: Estimating treatment effects with ML



TERMINOLOGY

Real-world data (RWD) vs. real-world evidence (RWE) to support medicine

The US Food and Drug Administration (FDA) defines [1,2,3]:

Real-world data (RWD)

- Data relating to patient health status and the delivery of healthcare
- **Examples:** electronic health records (EHRs), claims and billing activities, disease registries, …
- Naming: observational data (≠ experimental data)

Real-world evidence (RWE)

- Analysis of RWD regarding usage and effectiveness
- Vision: greater personalization of care
- Disclaimer: should not replace but augment RCTs

1) Real-World Evidence — Where Are We Now? https://www.nejm.org/doi/full/10.1056/NEJMp2200089
2) Real-World Evidence — What Is It and What Can It Tell Us? https://www.nejm.org/doi/full/10.1056/nejmsb1609216
3) Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness https://jamanetwork.com/journals/jama/fullarticle/2697359

TERMINOLOGY

Real-world data (RWD) vs. real-world evidence (RWE) to support medicine

The US Food and Drug Administration (FDA) defines [1,2,3]:

Real-world data (RWD)

- Data relating to patient health status and the delivery of healthcare
- **Examples:** electronic health records (EHRs), claims and billing activities, disease registries, …
- Naming: observational data (≠ experimental data)

- **Aim:** estimate treatment effectiveness
- **Challenges:** representativeness (selection bias), no proper randomization, …
- **Custom methodologies:** target trial emulation, **causal machine learning**, …

Real-world evidence (RWE)

- Analysis of RWD regarding usage and effectiveness
- Vision: greater personalization of care
- Disclaimer: should not replace but augment RCTs

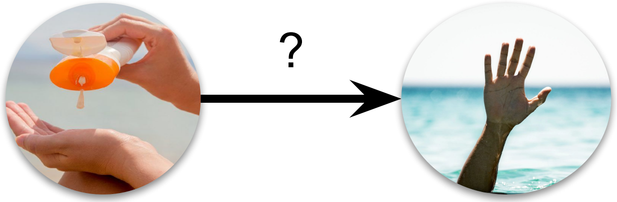
1) Real-World Evidence — Where Are We Now? https://www.nejm.org/doi/full/10.1056/NEJMp2200089
2) Real-World Evidence — What Is It and What Can It Tell Us? https://www.nejm.org/doi/full/10.1056/nejmsb1609216
3) Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness https://jamanetwork.com/journals/jama/fullarticle/2697359

Real-world data (RWD) vs. real-world evidence (RWE) to support medicine

Why is getting a **meaningful** RWE challenging?

Real-world (observational) data (RWD)

- Observational data of
  - sunscreen usage (binary treatment)
  - number of drowning-related deaths (outcome)



- **Aim:** effect of sunscreen on the chance of drowning

Real-world evidence (RWE)

- Evidence: The higher the usage of sunscreen -> the more likely is the chance of drowning
- This is counterintuitive: Is there something we didn't account for?

Real-world data (RWD) vs. real-world evidence (RWE) to support medicine

Why is getting a **meaningful** RWE challenging? -> **Hidden confounding**

Real-world data (RWD)

- Observational data of
  - sunscreen usage (binary treatment)
  - number of drowning-related deaths (outcome)
  - **intensity of sunlight (covariates)**

- **Aim:** effect of sunscreen on the chance of drowning for **different intensities of sunlight**

Real-world evidence (RWE)

- Evidence: no association between sunscreen usage and chance of drowning in each group of sunlight
- Comparing with the previous slide: Intensity of sunlight is a **confounder**

Application scenarios of RWD

RWD helps to guide decision-making (beyond RCTs):

1 … in the absence of a standard of care

- Specific subtypes of diseases with no standard of care yet (e.g., oncology)
- New or experimental drugs (e.g., orphan drugs, is Biontech vs. Moderna vaccine more effective for subcohort X?)

2 … in complex, high-dimensional decision problems

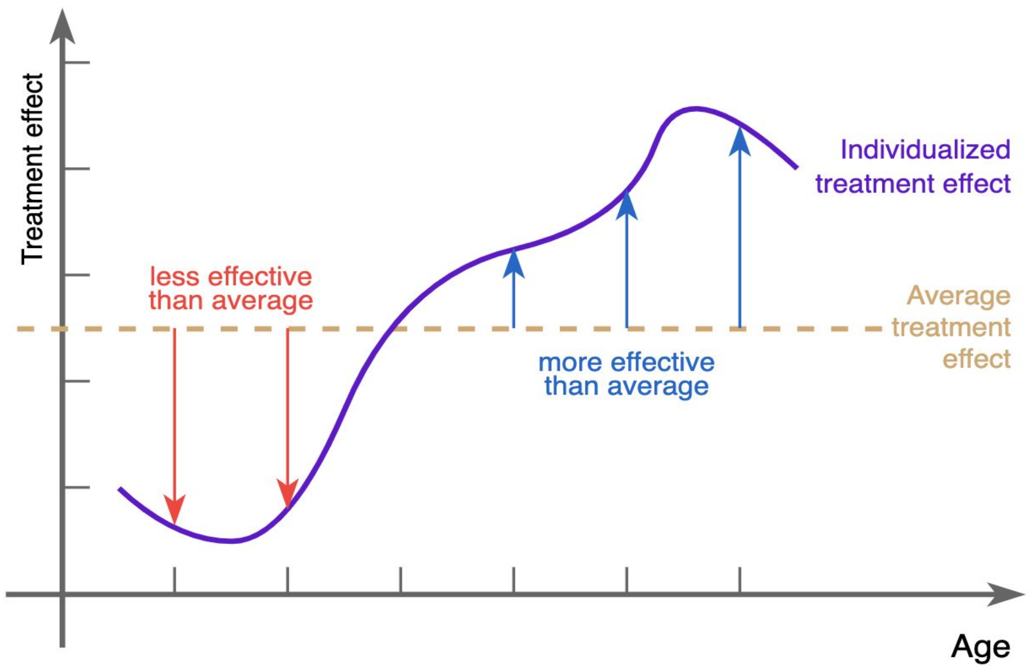
- Complex dosaging problems

3 … when RCTs are unethical

- Vulnerable populations (e.g., pregnant women) [1]

4 … when a greater personalization is desired

- Highly granular subpopulations that cannot be really placed in RCTs (e.g., women, above 60, with comorbidity etc.)
- Personalization based on genome data (e.g., precision medicine)

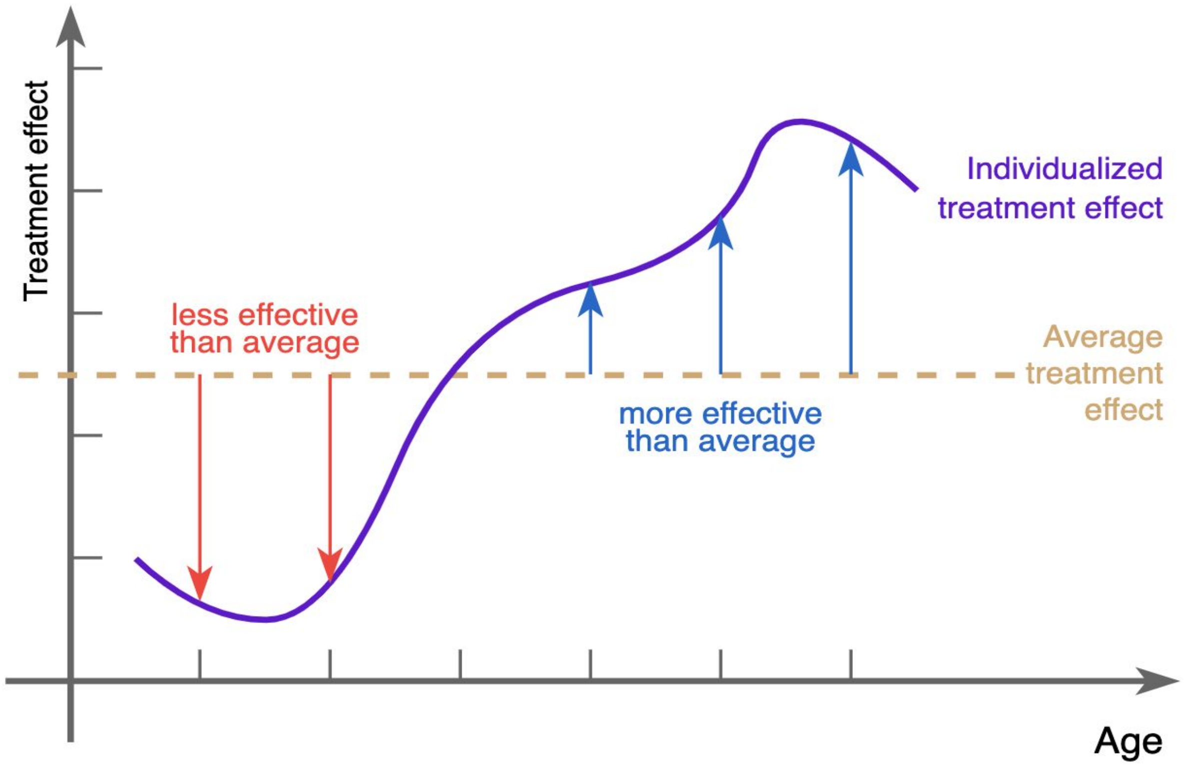


1) The Effectiveness of Right Heart Catheterization in the Initial Care of Critically Ill Patients https://jamanetwork.com/journals/jama/article-abstract/407990

AIM

Understanding heterogeneity in the treatment effect

- Focus is often on **average** treatment effect (ATE)
- ATE is aggregated across the population
- ATE **cannot** tell whether a treatment works for some or not
  → e.g., medication works only for women but not for men, but RCT was done with all patients
- NB: both RCTs and target trial emulation focus on ATEs



To personalize treatment recommendations, we need to understand the **individualized** treatment effect (ITE)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN LMU MUNICH SCHOOL OF MANAGEMENT

Munich Center for Machine Learning

Short introduction to causal machine learning

Reference:
Feuerriegel, S., Frauen, D., Melnychuk, V., Schweisthal, J., Hess, K., Curth, A., Bauer, S., Kilbertus, N., Kohane, I.S. and van der Schaar, M., 2024. Causal machine learning for predicting treatment outcomes. Nature Medicine, 30(4), pp.958-968.

PRIMER

Ambiguity of the definition

“Causal ML” could be both:

Causal inference for machine learning

Causal inference concepts

ML / DL problems

- Explainability
- Fairness
- Algorithmic recourse
- Robustness / domain adaptation
- …

Machine learning for causal inference

Causal inference problems

- Predicting treatment outcomes
- Counterfactual inference
- Causal discovery
- …

ML / DL tools

PRIMER

Ambiguity of the definition

"Causal ML" could be both:

Causal inference for machine learning

Causal inference concepts

ML / DL problems

- Explainability
- Fairness
- Algorithmic recourse
- Robustness / domain adaptation
- …

Machine learning for causal inference

Causal inference problems

- Predicting treatment outcomes
- Counterfactual inference
- Causal discovery
- …

ML / DL tools

PRIMER

Ladder of causation

Pearl's layers of causation

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Causal Hierarchy Theorem: statistical inference for a layer requires the information from the same or higher layer. For the inference from lower layer data, we need to make **additional assumptions**.

[1] Elias Bareinboim et al. "On Pearl's hierarchy and the foundations of causal inference". In: Probabilistic and Causal Inference: The Works of Judea Pearl. Association for Computing Machinery, 2022, pp. 507–556.

Ladder of causation

Pearl's layers of causation

Traditional ML

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Causal Hierarchy Theorem: statistical inference for a layer requires the information from the same or higher layer. For the inference from lower layer data, we need to make **additional assumptions**.

[1] Elias Bareinboim et al. "On Pearl's hierarchy and the foundations of causal inference". In: Probabilistic and Causal Inference: The Works of Judea Pearl. Association for Computing Machinery, 2022, pp. 507–556.

PRIMER

Ladder of causation

Pearl's layers of causation

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a s about the electi
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Causal ML

Causal Hierarchy Theorem: statistical inference for a layer requires the information from the same or higher layer. For the inference from lower layer data, we need to make **additional assumptions**.

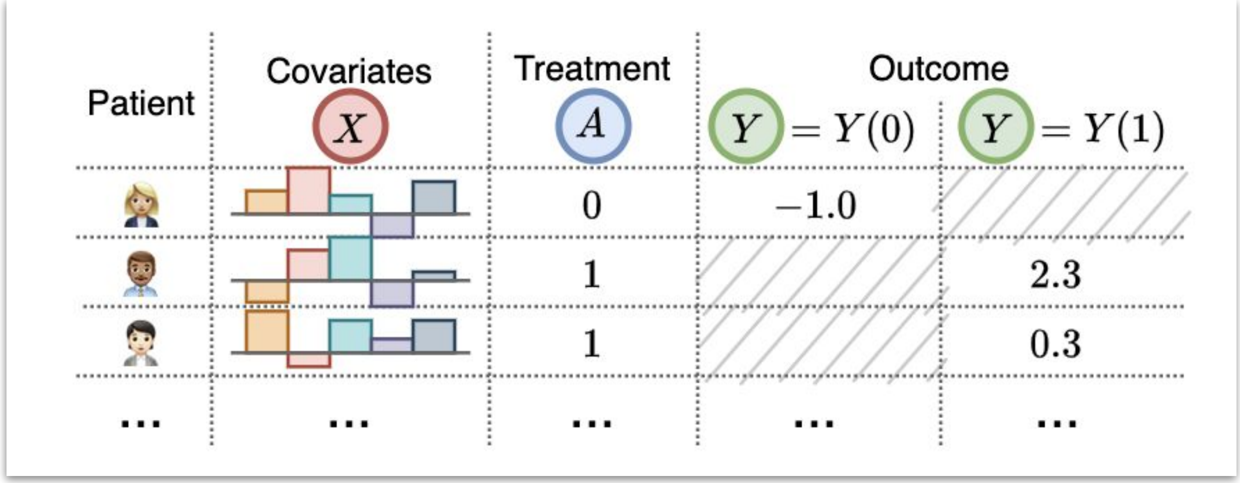
[1] Elias Bareinboim et al. "On Pearl's hierarchy and the foundations of causal inference". In: Probabilistic and Causal Inference: The Works of Judea Pearl. Association for Computing Machinery, 2022, pp. 507–556.

Predicting treatment outcomes (treatment effects or potential outcome)

Problem formulation

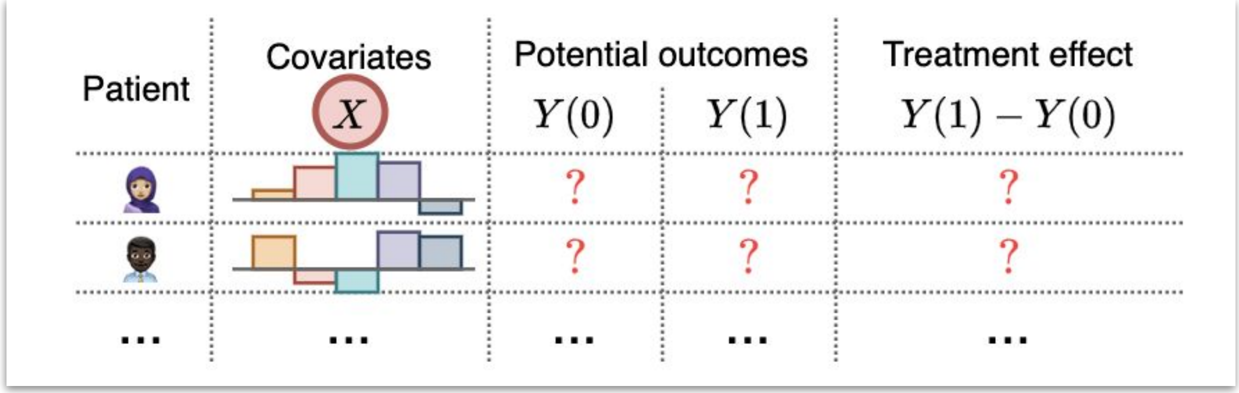
- Given i.i.d. observational dataset $\mathcal{D} = \{x_i, a_i, y_i\}_{i=1}^{n} \sim \mathbb{P}(X, A, Y)$

  X covariates
  A (binary) treatments
  Y continuous (factual) outcomes

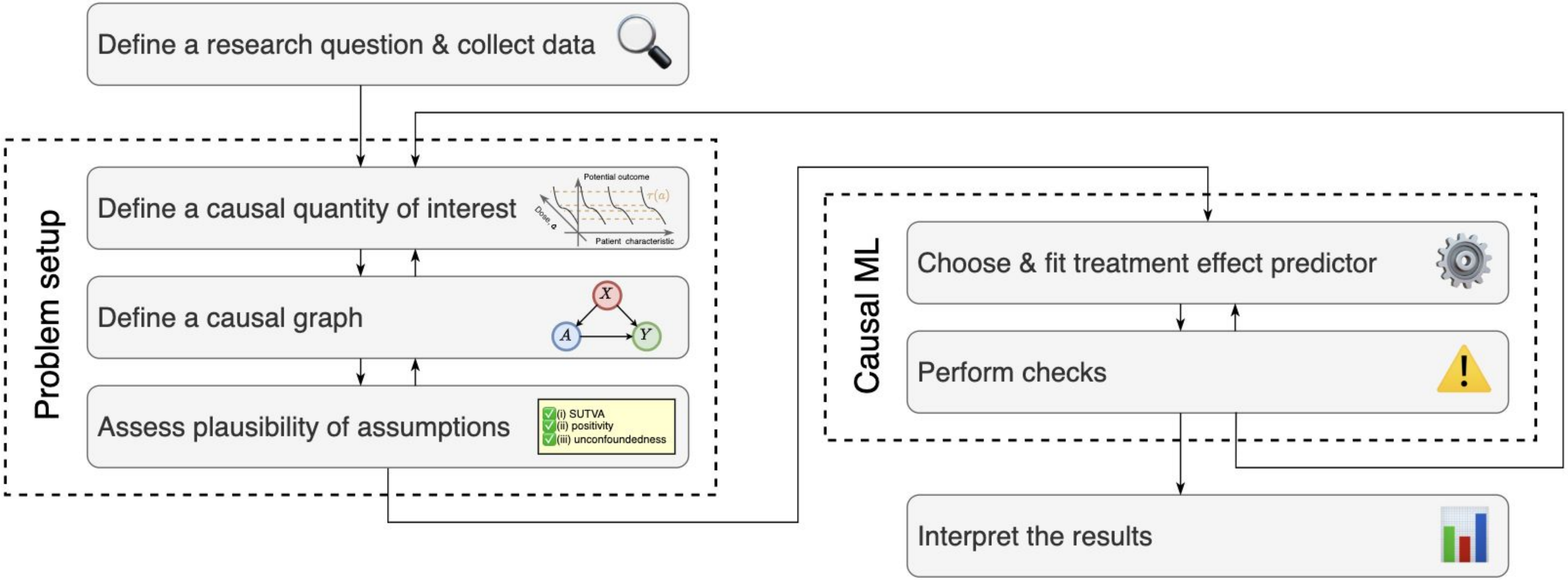


- We want to identify & estimate treatment outcomes:
  - **treatment effects** $Y[1] - Y[0]$
  - **potential outcomes** (separately) $Y[0]$ $Y[1]$

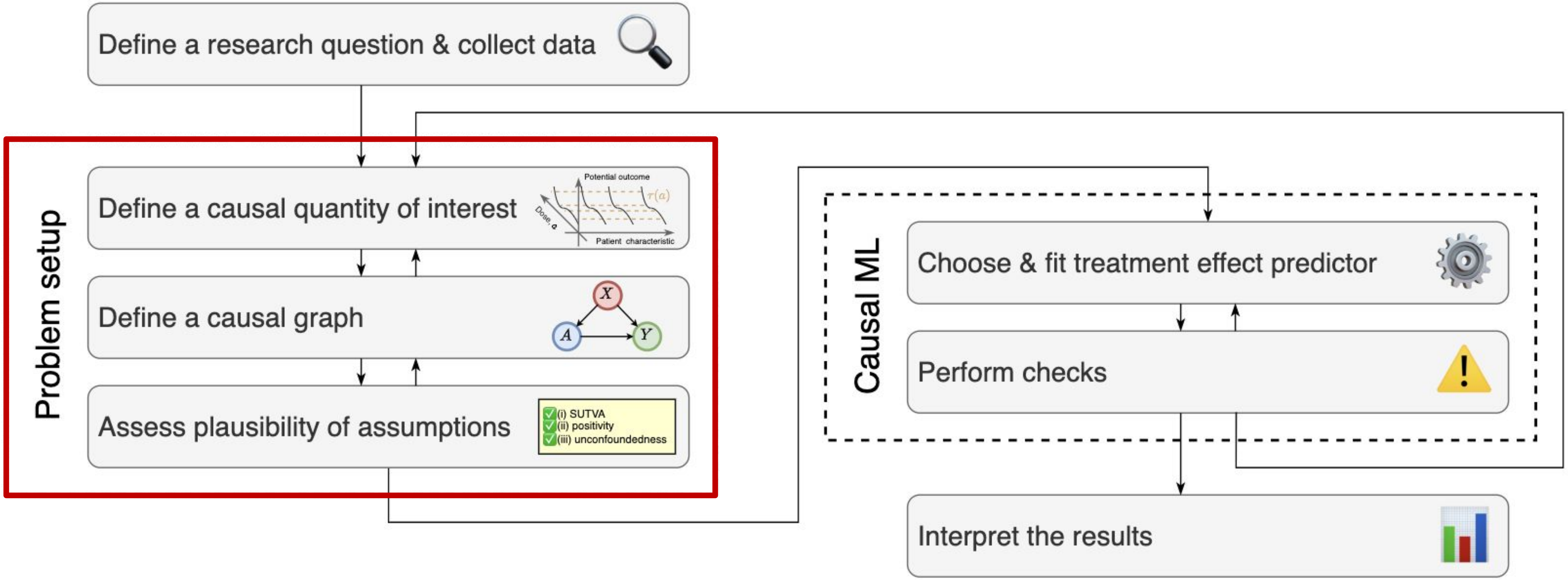
- **Fundamental problem**: never observing both potential outcomes!



Causal ML Workflow

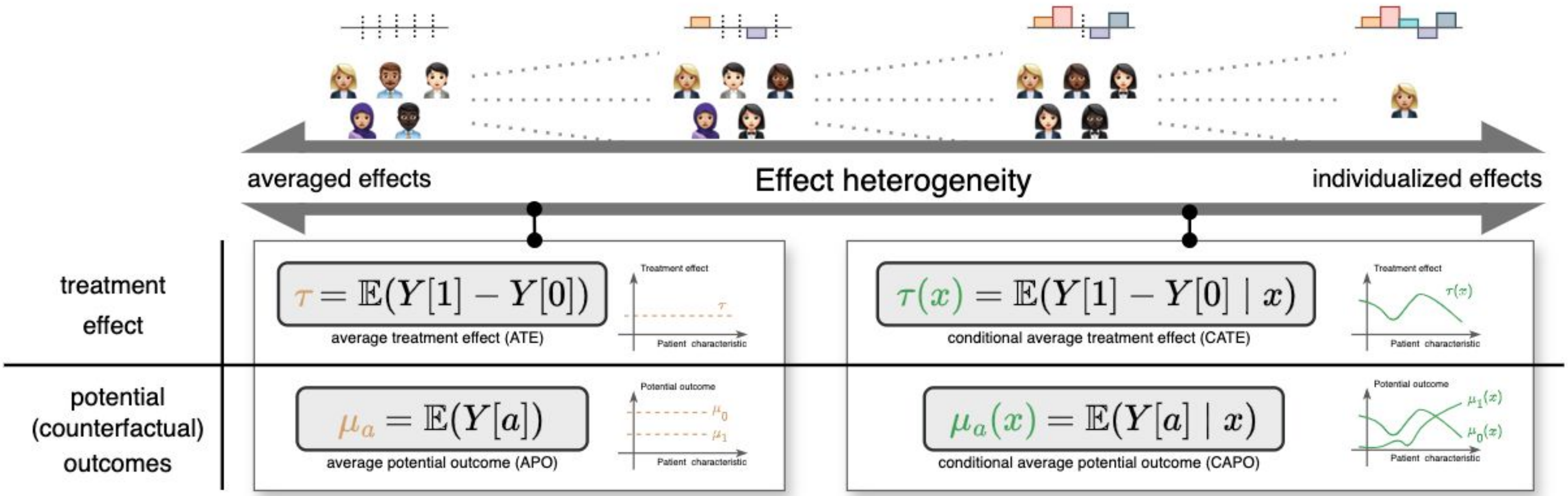


Causal ML Workflow



PROBLEM SETUP

Causal quantities of interest



PROBLEM SETUP

Assumption frameworks

$$\mathcal{D} = \{x_i, a_i, y_i\}_{i=1}^{n} \sim \mathbb{P}(X, A, Y)$$

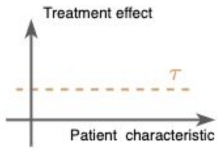
Potential outcomes framework (Neyman-Rubin)

Structural causal model (SCM) (Pearl-Bareinboim)

Causal graph

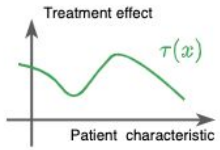
$$\tau = \mathbb{E}(Y[1] - Y[0])$$

average treatment effect (ATE)



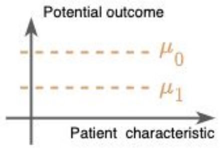
$$\tau(x) = \mathbb{E}(Y[1] - Y[0] \mid x)$$

conditional average treatment effect (CATE)



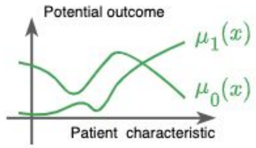
$$\mu_a = \mathbb{E}(Y[a])$$

average potential outcome (APO)



$$\mu_a(x) = \mathbb{E}(Y[a] \mid x)$$

conditional average potential outcome (CAPO)



PROBLEM SETUP

Assumption frameworks: SCMs and causal graphs

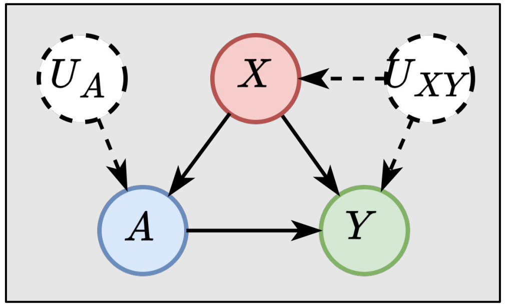
$$\mathcal{D} = \{x_i, a_i, y_i\}_{i=1}^{n} \sim \mathbb{P}(X, A, Y)$$

Potential outcomes framework (Neyman-Rubin)

Structural causal model (SCM) (Pearl-Bareinboim)

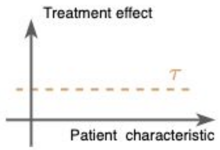
Causal graph

Assumptions stem from structural knowledge



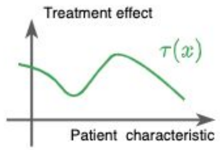
$$\tau = \mathbb{E}(Y[1] - Y[0])$$

average treatment effect (ATE)



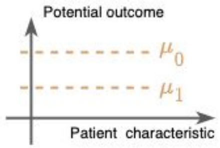
$$\tau(x) = \mathbb{E}(Y[1] - Y[0] \mid x)$$

conditional average treatment effect (CATE)



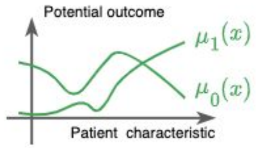
$$\mu_a = \mathbb{E}(Y[a])$$

average potential outcome (APO)



$$\mu_a(x) = \mathbb{E}(Y[a] \mid x)$$

conditional average potential outcome (CAPO)



Assumption frameworks: Potential outcomes framework

$$\mathcal{D} = \{x_i, a_i, y_i\}_{i=1}^{n} \sim \mathbb{P}(X, A, Y)$$

More general

(i) Consistency
(ii) Positivity (Overlap)
(iii) Exchangeability (Ignorability)

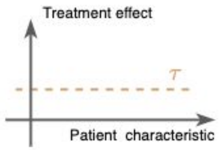
Potential outcomes framework (Neyman-Rubin)

Structural causal model (SCM) (Pearl-Bareinboim)

Causal graph

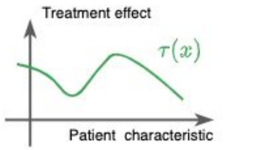
$$\tau = \mathbb{E}(Y[1] - Y[0])$$

average treatment effect (ATE)



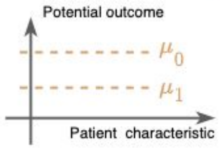
$$\tau(x) = \mathbb{E}(Y[1] - Y[0] \mid x)$$

conditional average treatment effect (CATE)



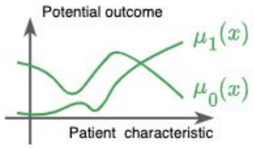
$$\mu_a = \mathbb{E}(Y[a])$$

average potential outcome (APO)



$$\mu_a(x) = \mathbb{E}(Y[a] \mid x)$$

conditional average potential outcome (CAPO)



Assumption frameworks

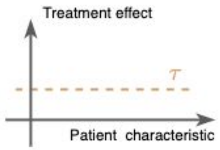
$$\mathcal{D} = \{x_i, a_i, y_i\}_{i=1}^{n} \sim \mathbb{P}(X, A, Y)$$

Potential outcomes framework (Neyman-Rubin) = Structural causal model (SCM) (Pearl-Bareinboim)

Causal graph

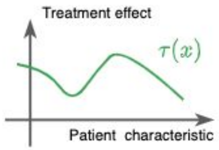
$$\tau = \mathbb{E}(Y[1] - Y[0])$$

average treatment effect (ATE)



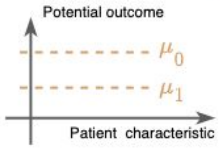
$$\tau(x) = \mathbb{E}(Y[1] - Y[0] \mid x)$$

conditional average treatment effect (CATE)



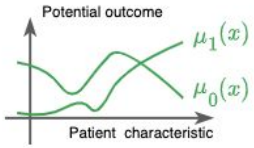
$$\mu_a = \mathbb{E}(Y[a])$$

average potential outcome (APO)



$$\mu_a(x) = \mathbb{E}(Y[a] \mid x)$$

conditional average potential outcome (CAPO)

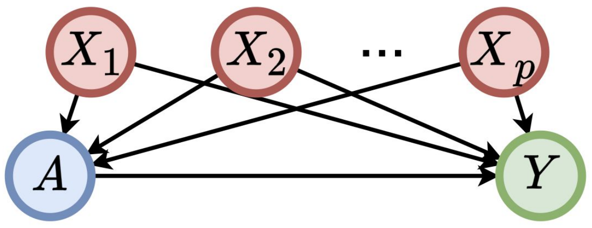


Example of a case study

Aim: estimate heterogeneous treatment effect of development aid on SDG outcomes

- Treatment *A*: development aid earmarked to end the HIV/AIDS epidemic
- Outcome *Y*: relative reduction in HIV infection rate
- Covariates *X*: control for differences in country characteristics

Causal graph



Causal quantity of interest

$$\boxed{\mu_a(x) = \mathbb{E}(Y[a] \mid x)}$$

conditional average potential outcome (CAPO)

Assumptions

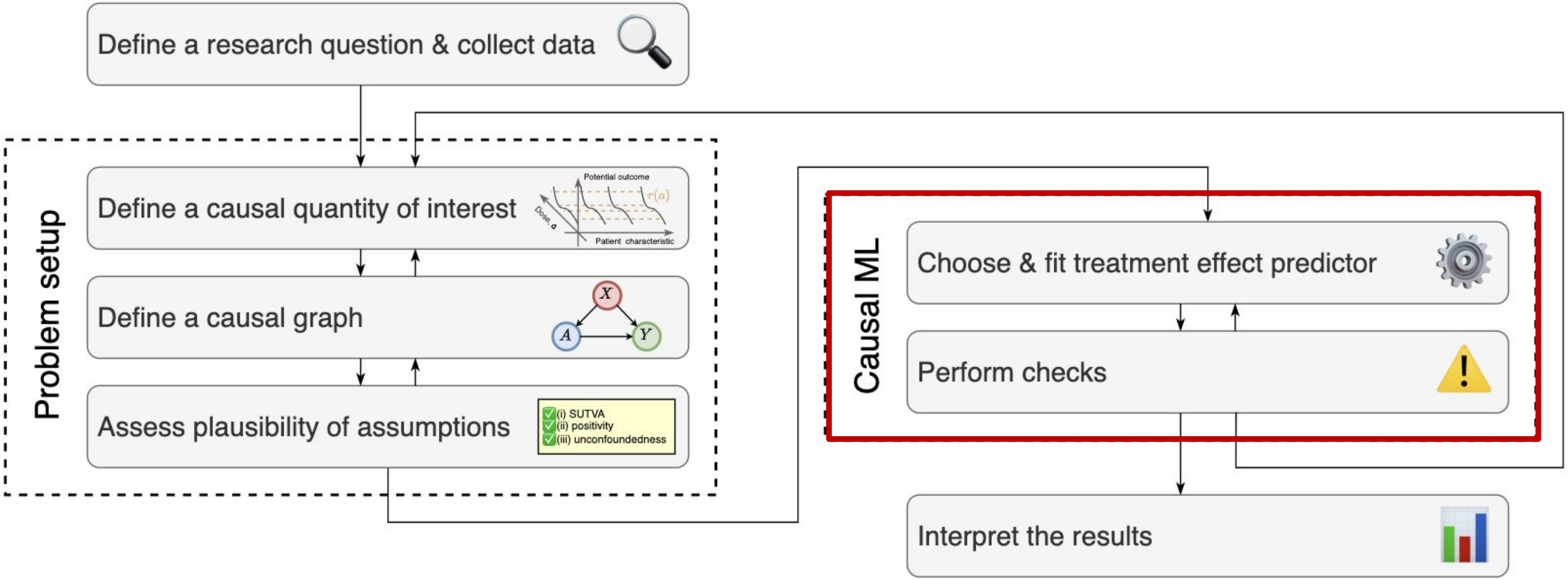
Potential outcomes framework (Neyman-Rubin)

Consistency: $Y = Y(a)$ if $A = a$

Positivity: $0 < p(A = a \mid X = x) < 1, \forall a \in \mathcal{A}$

Ignorability: $Y(a) \perp\!\!\!\perp A \mid X = x \;, \forall a \in \mathcal{A}$

Causal ML Workflow



Identification vs. estimation / learning

Identification (infinite data)

$$\mathbb{P}(X, A, Y)$$

observational distribution

$$\mu_a = \mathbb{E}(Y[a])$$

average potential outcome (APO)

target quantity

Potential outcomes framework (Neyman-Rubin)

$$\mu_a = \mathbb{E}(\mathbb{E}[Y \mid a, X])$$

back-door adjustment

$$\mu_a = \mathbb{E}\left[\frac{1(A=a)}{\pi_a(X)} Y\right]$$

inverse propensity of treatment weighting (IPTW)

identification formulas

Estimation (finite data)

$$\mathcal{D} = \{x_i, a_i, y_i\}_{i=1}^n \sim \mathbb{P}(X, A, Y)$$

sample from observational distribution

$$\mu_a = \mathbb{E}(\mathbb{E}[Y \mid a, X])$$

$$\mu_a = \mathbb{E}\left[\frac{1(A=a)}{\pi_a(X)} Y\right]$$

identification formulas

Semi-parametric efficiency theory / Neyman-orthogonal learning

$$\hat{\mu}_{a,\text{A-IPTW}} = \frac{1}{n} \sum_{i=1}^n \frac{a_i = a}{\hat{\pi}_a(x_i)} \Big(y_i - \hat{\mu}_a(x_i) \Big) + \hat{\mu}_a(x_i)$$

$$\hat{\eta} = \{\hat{\mu}_a(x) = \hat{\mathbb{E}}[Y \mid A = a, X = x];$$
$$\hat{\pi}_a(x) = \hat{\mathbb{P}}[A = a \mid X = x]\}$$

augmented inverse propensity of treatment weighting (A -IPTW)

efficient estimator

CAUSAL ML

Identification vs. estimation / learning

Identification (infinite data)

$$\mathbb{P}(X, A, Y)$$

observational distribution

$$\mu_a = \mathbb{E}(Y[a])$$

average potential outcome (APO)

target quantity

Potential outcomes framework (Neyman-Rubin)

$$\mu_a = \mathbb{E}(\mathbb{E}[Y \mid a, X])$$

back-door adjustment

$$\mu_a = \mathbb{E}\left[\frac{1(A=a)}{\pi_a(X)} Y\right]$$

inverse propensity of treatment weighting (IPTW)

identification formulas

Estimation (finite data)

$$\mathcal{D} = \{x_i, a_i, y_i\}_{i=1}^n \sim \mathbb{P}(X, A, Y)$$

sample from observational distribution

$$\mu_a = \mathbb{E}(\mathbb{E}[Y \mid a, X])$$

$$\mu_a = \mathbb{E}\left[\frac{1(A=a)}{\pi_a(X)} Y\right]$$

identification formulas

Semi-parametric efficiency theory / Neyman-orthogonal learning

$$\hat{\mu}_{a,\text{A-IPTW}} = \frac{1}{n}\sum_{i=1}^n \frac{a_i = a}{\hat{\pi}_a(x_i)}\Big(y_i - \hat{\mu}_a(x_i)\Big) + \hat{\mu}_a(x_i)$$

$$\hat{\eta} = \{\hat{\mu}_a(x) = \hat{\mathbb{E}}[Y \mid A = a, X = x];$$
$$\hat{\pi}_a(x) = \hat{\mathbb{P}}[A = a \mid X = x]\}$$

augmented inverse propensity of treatment weighting (A -IPTW)

efficient estimator

CAUSAL ML

Challenges and open questions fitting an ML model

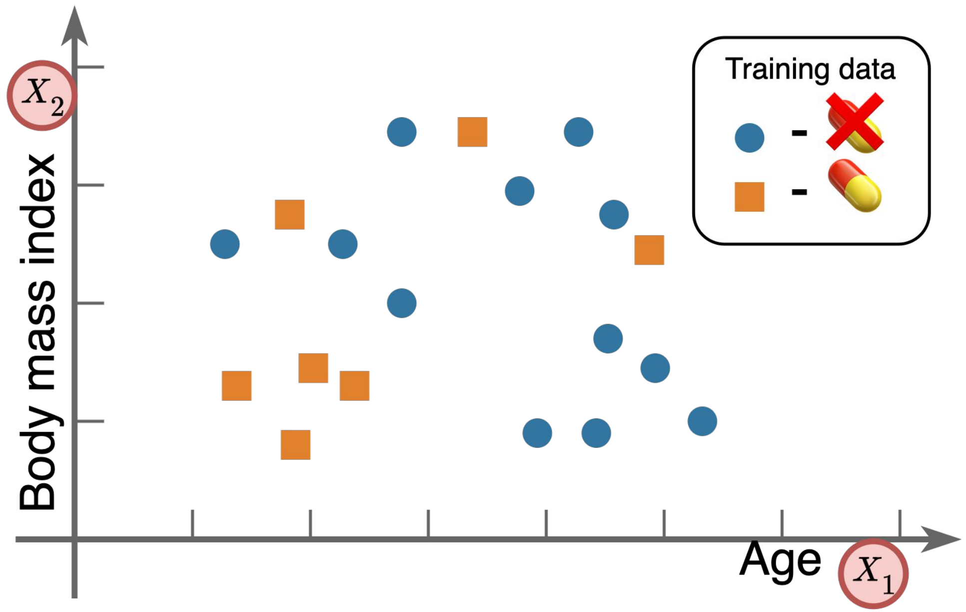
Challenges

$$\mu_a(x) = \mathbb{E}(Y[a] \mid x)$$

conditional average potential outcome (CAPO)

$$\tau(x) = \mathbb{E}(Y[1] - Y[0] \mid x)$$

conditional average treatment effect (CATE)



CAUSAL ML

Challenges and open questions fitting an ML model

Challenges

$$\mu_a(x) = \mathbb{E}(Y[a] \mid x)$$

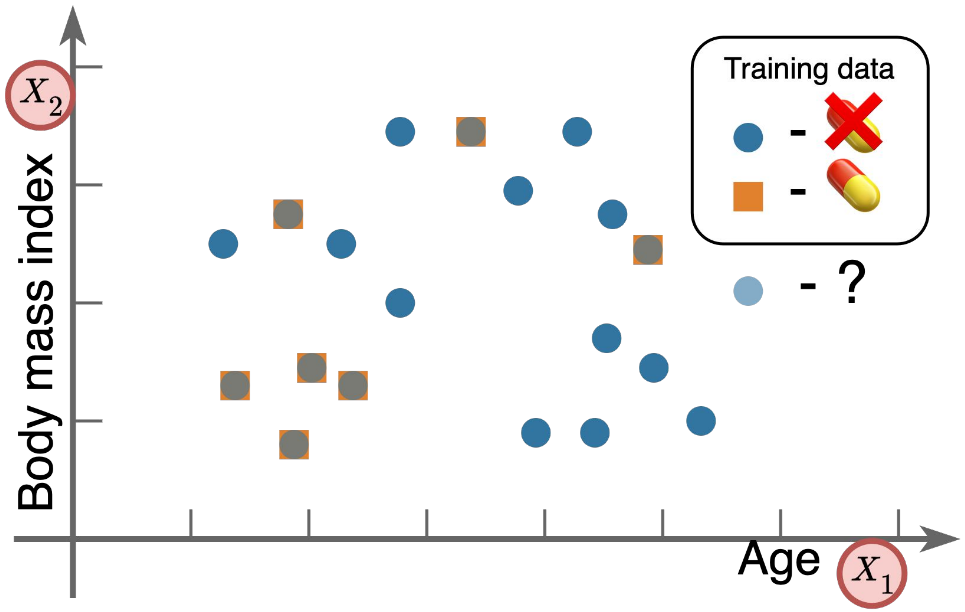
conditional average potential outcome (CAPO)

- **Selection bias**: parts of the population rarely gets treated

$$\tau(x) = \mathbb{E}(Y[1] - Y[0] \mid x)$$

conditional average treatment effect (CATE)

- **Selection bias**: parts of the population rarely gets treated



Challenges and open questions fitting an ML model

Challenges

$$\mu_a(x) = \mathbb{E}(Y[a] \mid x)$$

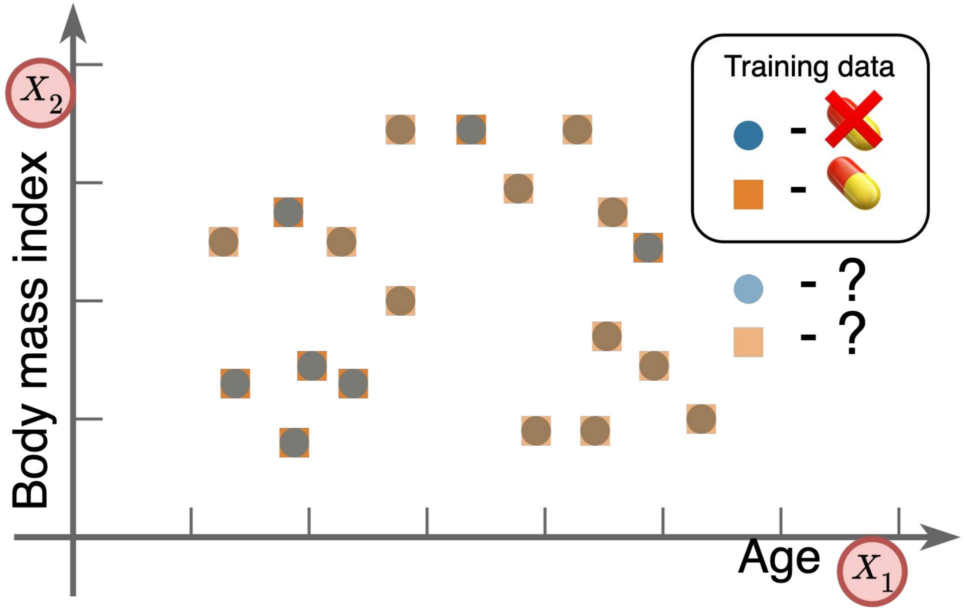
conditional average potential outcome (CAPO)

- **Selection bias**: parts of the population rarely gets treated

$$\tau(x) = \mathbb{E}(Y[1] - Y[0] \mid x)$$

conditional average treatment effect (CATE)

- **Selection bias**: parts of the population rarely gets treated
- **Fundamental problem**: never observing a difference of potential outcomes



Challenges and open questions fitting an ML model

Challenges

$$\mu_a(x) = \mathbb{E}(Y[a] \mid x)$$

conditional average potential outcome (CAPO)

- **Selection bias**: parts of the population rarely gets treated

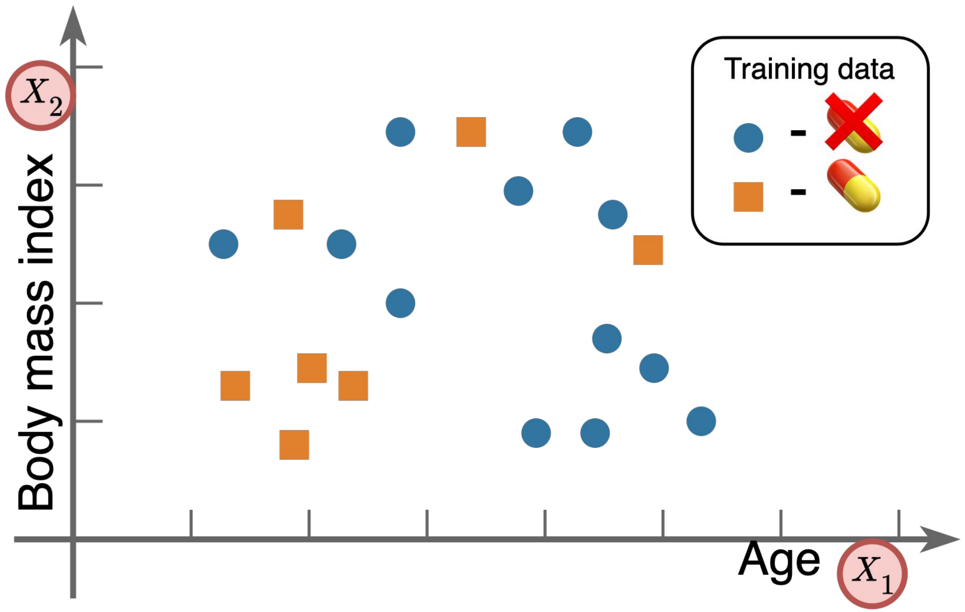
$$\tau(x) = \mathbb{E}(Y[1] - Y[0] \mid x)$$

conditional average treatment effect (CATE)

- **Selection bias**: parts of the population rarely gets treated
- **Fundamental problem**: never observing a difference of potential outcomes

Open problems

- How to effectively address selection bias?
- How to incorporate inductive biases, e.g., regularize CAPO / CATE models?



Methods

Meta-learners

- Meta-learners (Kunzel 2019) are model-agnostic methods for CATE estimation
- Can be used for treatment effect estimation in combination with an arbitrary ML model of choice (e.g., a decision tree, a neural network)

Model-based learners

- Model-specific methods make adjustments to existing ML models to address statistical challenges arising in treatment effect estimation
- Prominent **examples** are the causal tree (Athey 2016) and the causal forest (Wager 2018, Athey 2019)
- Others adapt representation learning to leverage neural networks (Shalit 2017, Shi 2019)

1. Künzel, Sören R., et al. "Metalearners for estimating heterogeneous treatment effects using machine learning." Proceedings of the national academy of sciences 116.10 (2019): 4156-4165.
2. Athey, Susan, and Guido Imbens. "Recursive partitioning for heterogeneous causal effects." Proceedings of the National Academy of Sciences 113.27 (2016): 7353-7360.
3. Athey, Susan, and Stefan Wager. "Estimating treatment effects with causal forests: An application." Observational studies 5.2 (2019): 37-51.
4. Shalit, Uri, Fredrik D. Johansson, and David Sontag. "Estimating individual treatment effect: generalization bounds and algorithms." International conference on machine learning. PMLR, 2017.
5. Shi, Claudia, David Blei, and Victor Veitch. "Adapting neural networks for the estimation of treatment effects." Advances in neural information processing systems 32 (2019).

Methods

Meta-learners

One-stage learners

- "Plug-in learners": fit a **s**ingle regression model with a treatment as an input or **t**wo regression models for each treated and control sub-groups
- Examples: S-learner and T-learner

Two-stage learners

- Two-stages of learning: derive and estimate pseudo-outcomes as surrogates, which has the same expected value as the CATE
- Examples: DR-learner and R-learner

Model-based learners

- Model-specific methods make adjustments to existing ML models to address statistical challenges arising in treatment effect estimation
- Prominent **examples** are the causal tree (Athey 2016) and the causal forest (Wager 2018, Athey 2019)
- Others adapt representation learning to leverage neural networks (Shalit 2017, Shi 2019)

1. Künzel, Sören R., et al. "Metalearners for estimating heterogeneous treatment effects using machine learning." Proceedings of the national academy of sciences 116.10 (2019): 4156-4165.
2. Athey, Susan, and Guido Imbens. "Recursive partitioning for heterogeneous causal effects." Proceedings of the National Academy of Sciences 113.27 (2016): 7353-7360.
3. Athey, Susan, and Stefan Wager. "Estimating treatment effects with causal forests: An application." Observational studies 5.2 (2019): 37-51.
4. Shalit, Uri, Fredrik D. Johansson, and David Sontag. "Estimating individual treatment effect: generalization bounds and algorithms." International conference on machine learning. PMLR, 2017.
5. Shi, Claudia, David Blei, and Victor Veitch. "Adapting neural networks for the estimation of treatment effects." Advances in neural information processing systems 32 (2019).

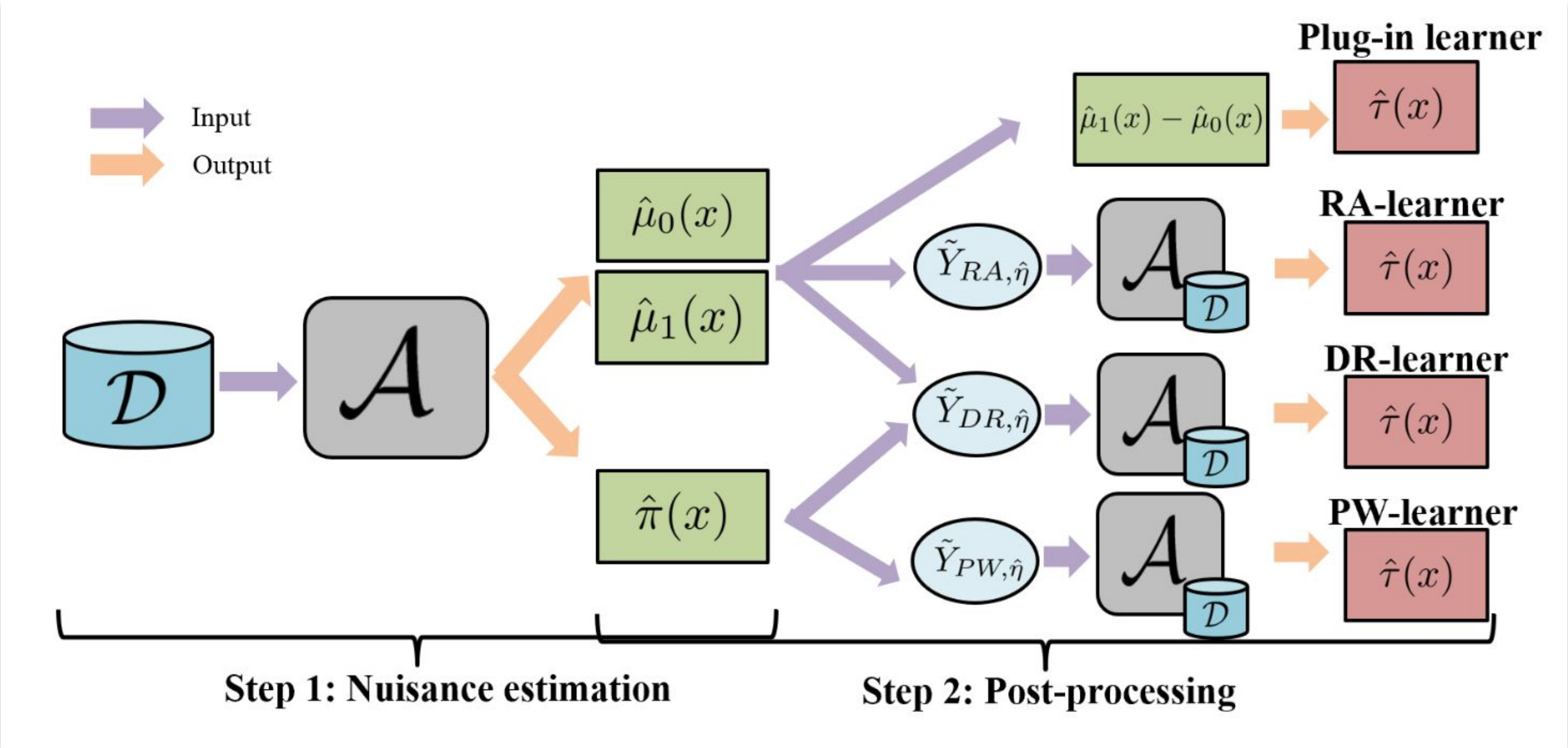
One-stage and two-stage meta-learners

Example: meta-learners for CATE

$$\tau(x) = \mathbb{E}(Y[1] - Y[0] \mid x)$$

conditional average treatment effect (CATE)

Method: Using any ML model to fit relevant parts of the observed distribution, namely, **nuisance functions**. Then, we can use the nuisance functions estimators for the final CATE model.



Curth, Alicia, and Mihaela Van der Schaar. "Nonparametric estimation of heterogeneous treatment effects: From theory to learning algorithms." International Conference on Artificial Intelligence and Statistics. PMLR, 2021.

Comparison of meta-learners

	Plug-in learners		Two-step learners			
	S-learner	T-learner	RA-learner	IPW-learner	DR-learner	R-learner
a. Large sample size? —Yes→	=	=	=	=	⇑	⇑
b. Need to regularize CATE predictor? —Yes→	×	×	⇑	⇑	⇑	⇑
c. Strong overlap violations (e.g., patients with certain covariates are never treated)? —Yes→	=	=	=	⇓	⇓	=
d. Propensity score is known (e.g., RCT) or easy to estimate? —Yes→	=	=	=	⇑	⇑	⇑
e. Interpretability through a single model for predicting CATE and potential outcomes? —Yes→	⇑	⇑	×	×	×	×

⇑ Beneficial for relative performance

= Has no influence on relative performance

⇓ Detrimental for relative performance

× Not available

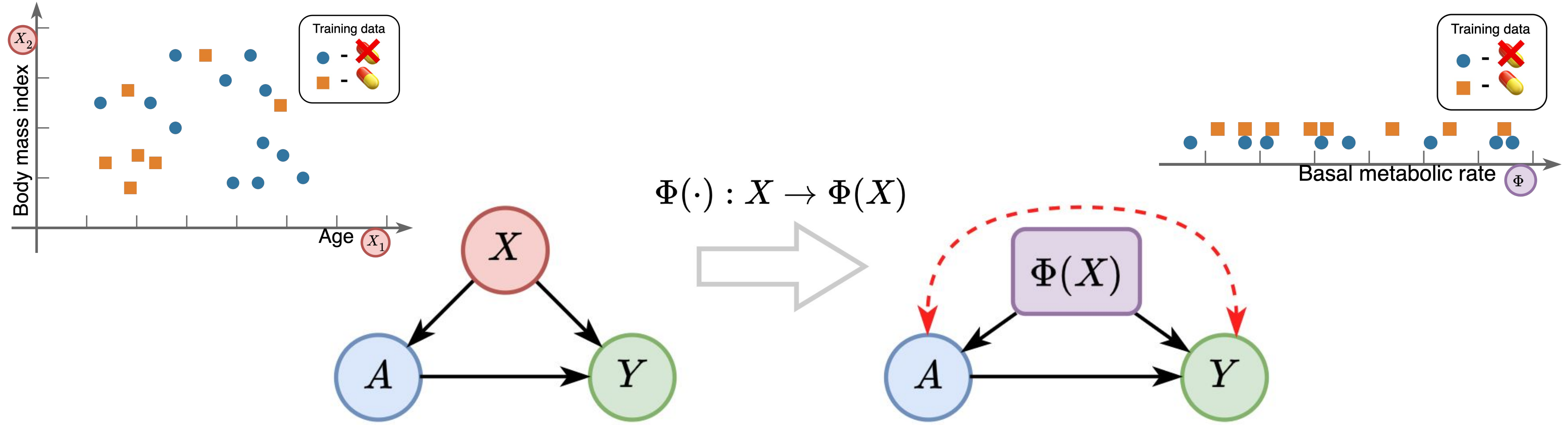
Model-based learners: Representation learning

Example: TarNET / CFRNet for CATE

$$\boxed{\tau(x) = \mathbb{E}(Y[1] - Y[0] \mid x)}$$

conditional average treatment effect (CATE)

Method: Learning a low-dimensional (balanced) representation Φ() of high-dimensional covariates. Then, we can fit a CATE model based on the representations.



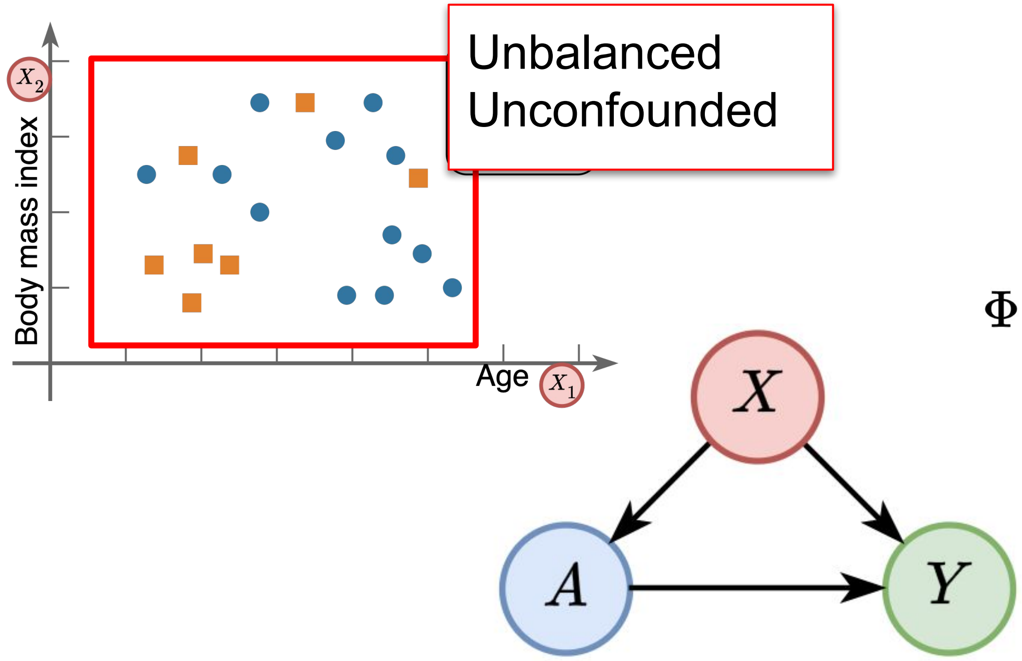
Model-based learners: Representation learning

Example: TarNET / CFRNet for CATE

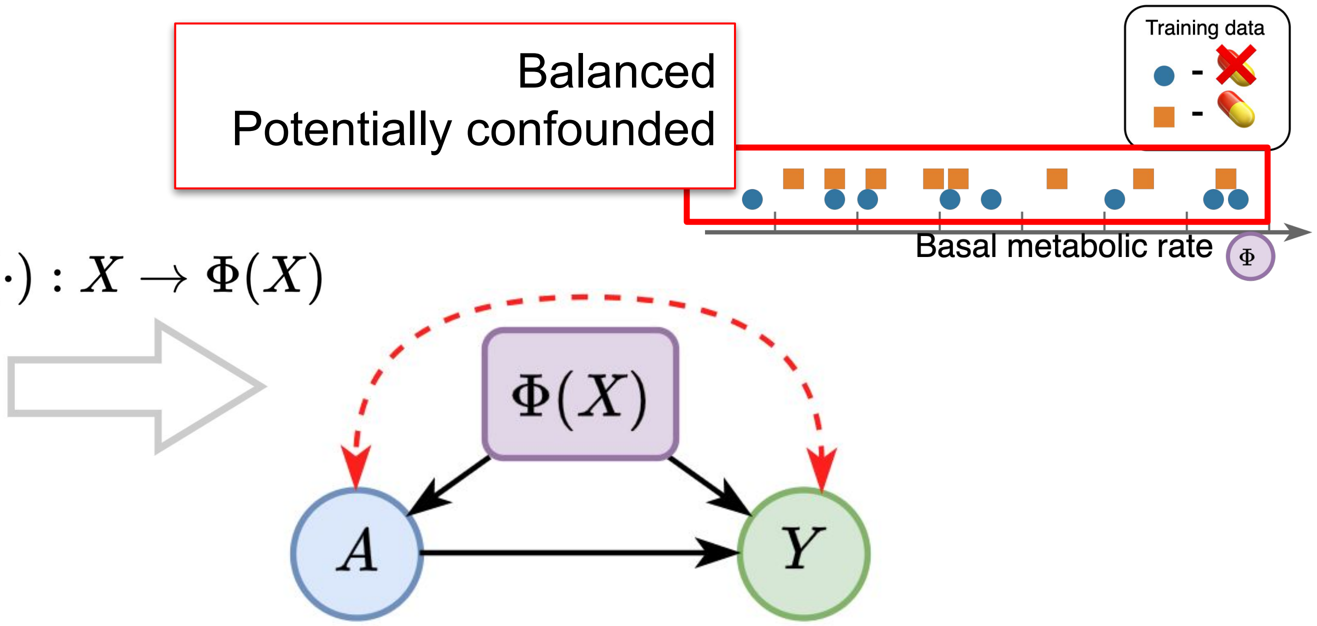
$$\boxed{\tau(x) = \mathbb{E}(Y[1] - Y[0] \mid x)}$$

conditional average treatment effect (CATE)

Method: Learning a low-dimensional (balanced) representation Φ() of high-dimensional covariates. Then, we can fit a CATE model based on the representations.

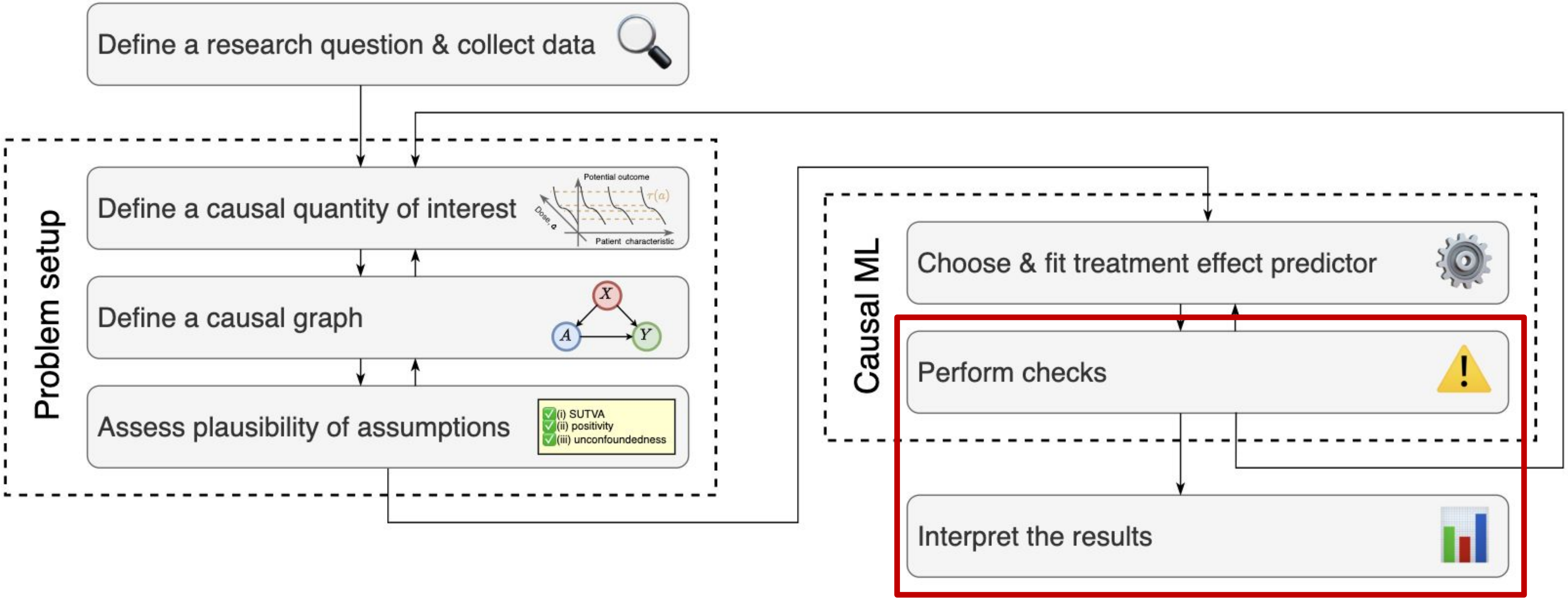


$\Phi(\cdot) : X \to \Phi(X)$



Where we are (and what is still needed): Current state of causal ML research

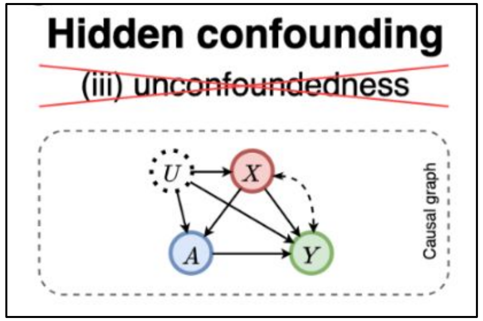
Causal ML Workflow



Extensions & Open research problems

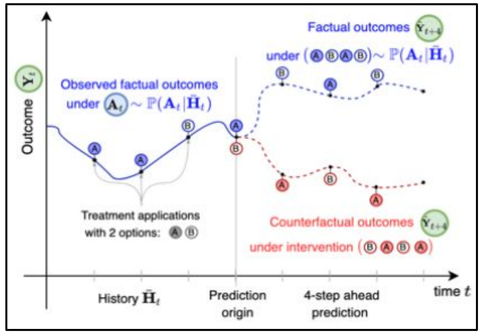
1 Model validity

- Selection and validation of CATE models
  - Unlike traditional ML, we do not have a ground truth validation subset
- Robustness checks wrt. violation of assumptions
  - Sensitivity models
  - Spillover effects



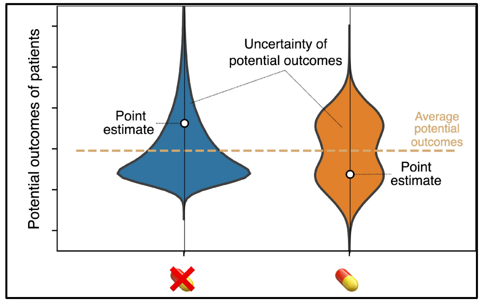
2 Flexibility

- Extensions to more complicated settings
  - continuous / high-dimensional treatments
  - time-varying potential outcomes and treatment effects
- Data fusion from multiple environments
- Constrained ML: interpretability, privacy enforcement

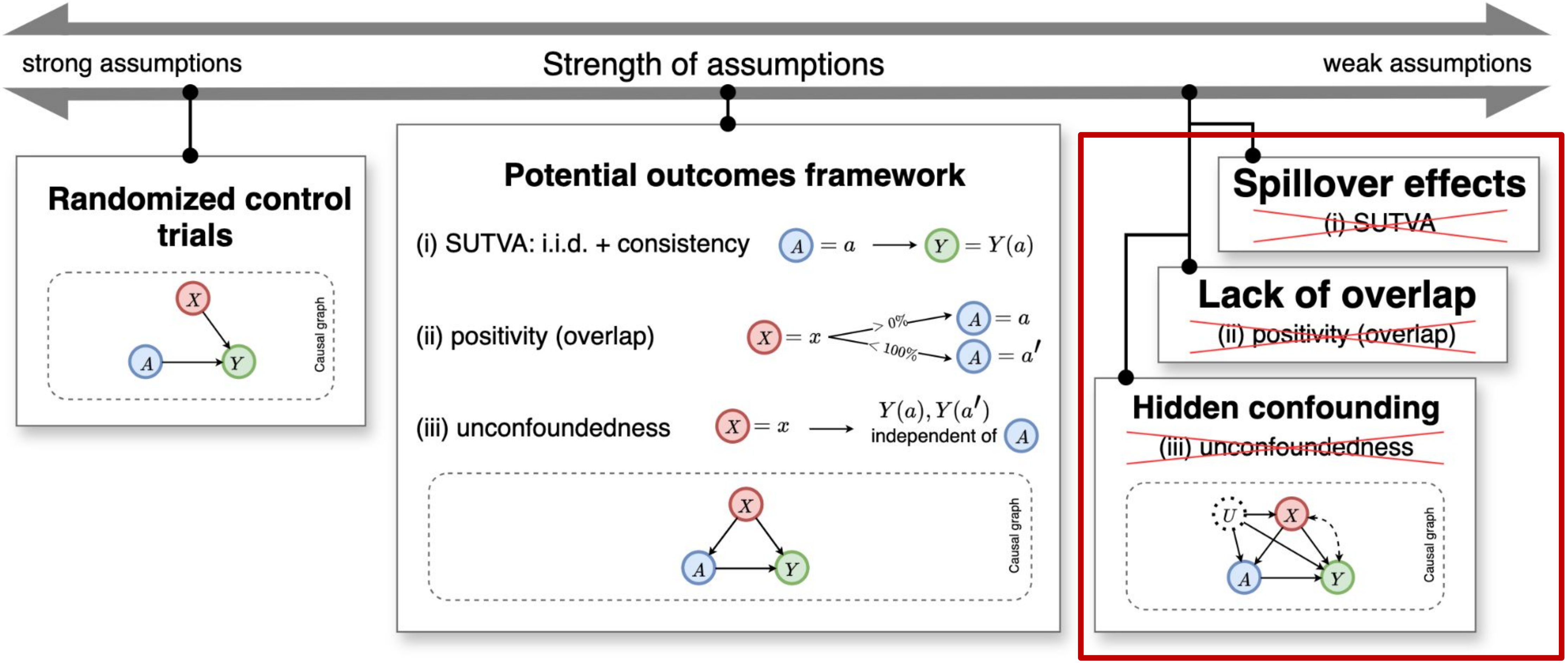


3 Uncertainty quantification

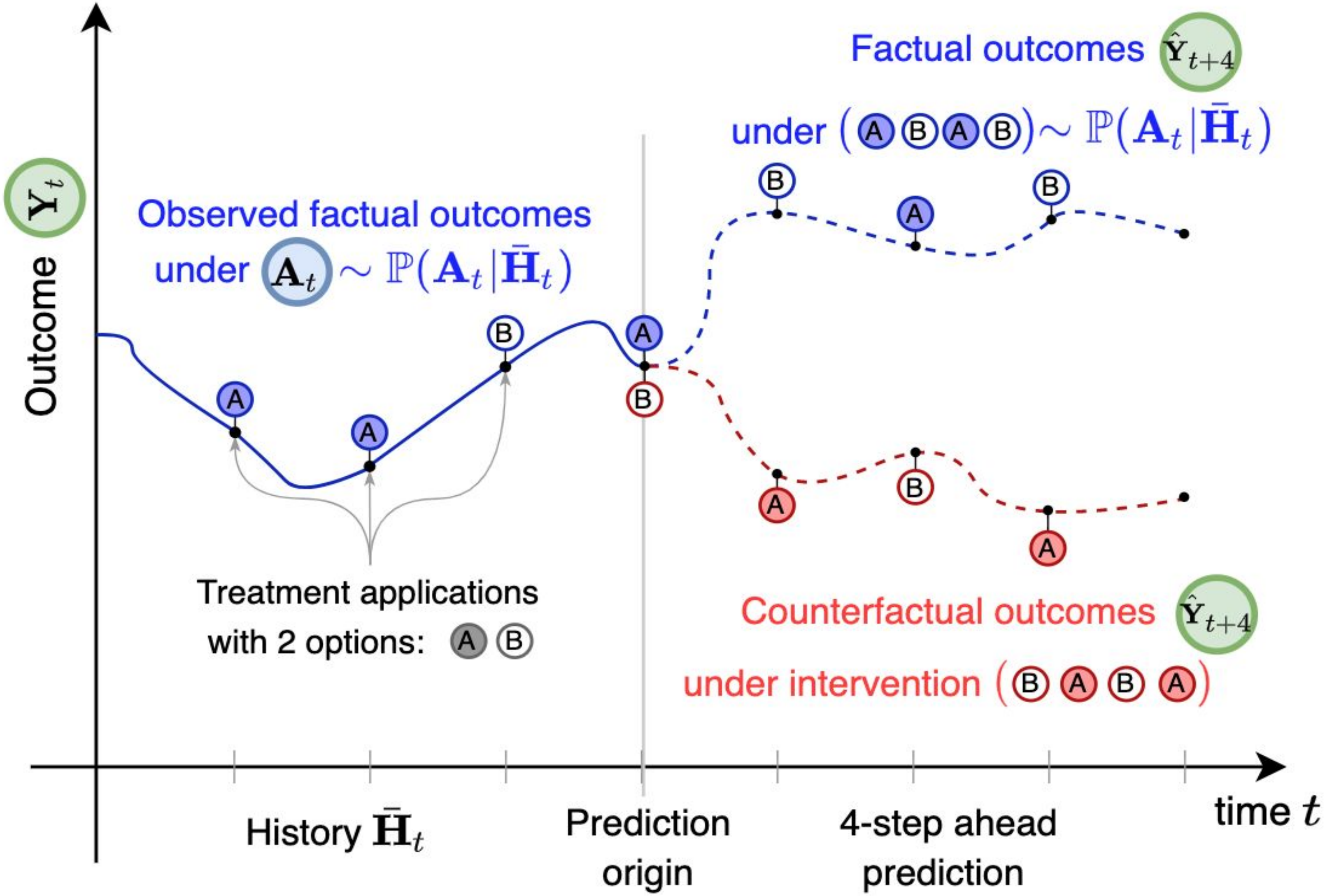
- Uncertainty quantification
  - uncertainty of estimation (aka confidence intervals)
  - predictive uncertainty (aka predictive intervals)



Model validity: Robustness checks wrt. violation of assumptions

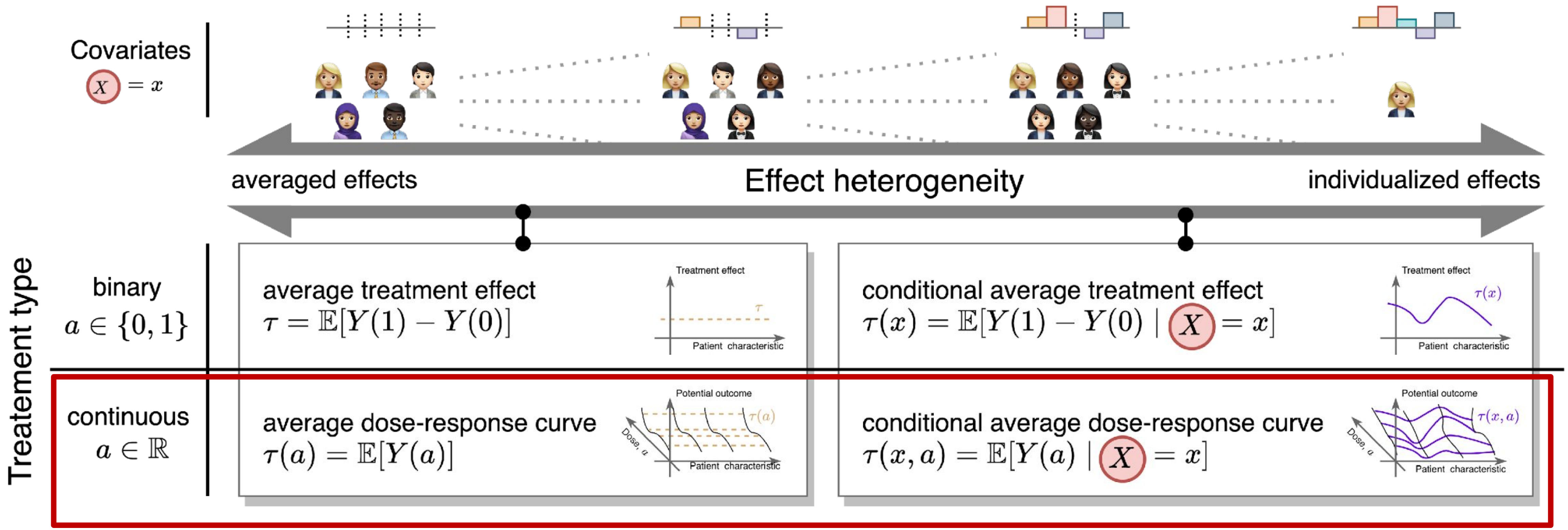


Flexibility: Causal ML for predicting outcomes over time

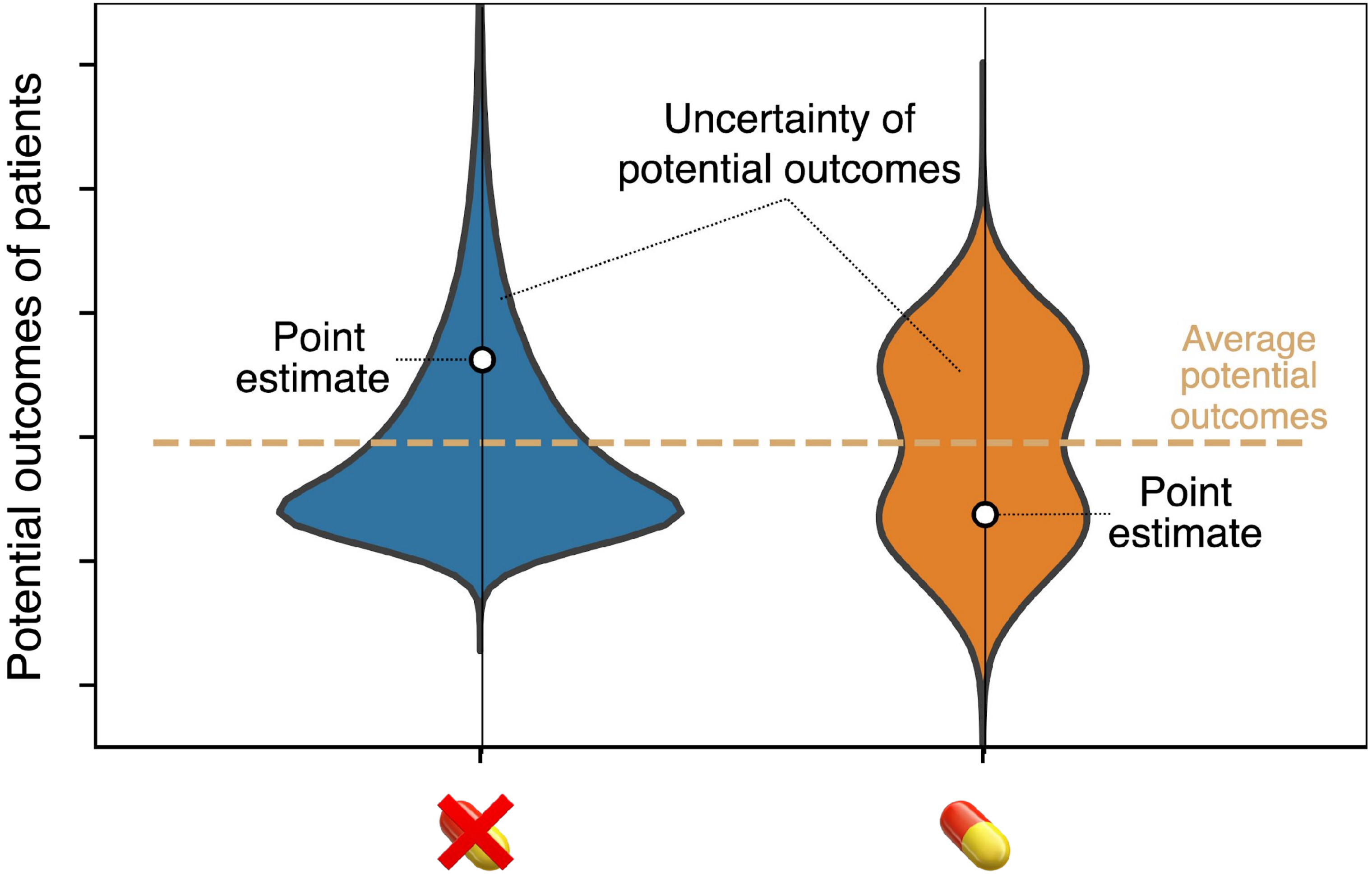


Melnychuk, Valentyn, Dennis Frauen, and Stefan Feuerriegel. "Causal transformer for estimating counterfactual outcomes." International Conference on Machine Learning. PMLR, 2022.

Flexibility: Continuous / high-dimensional treatments

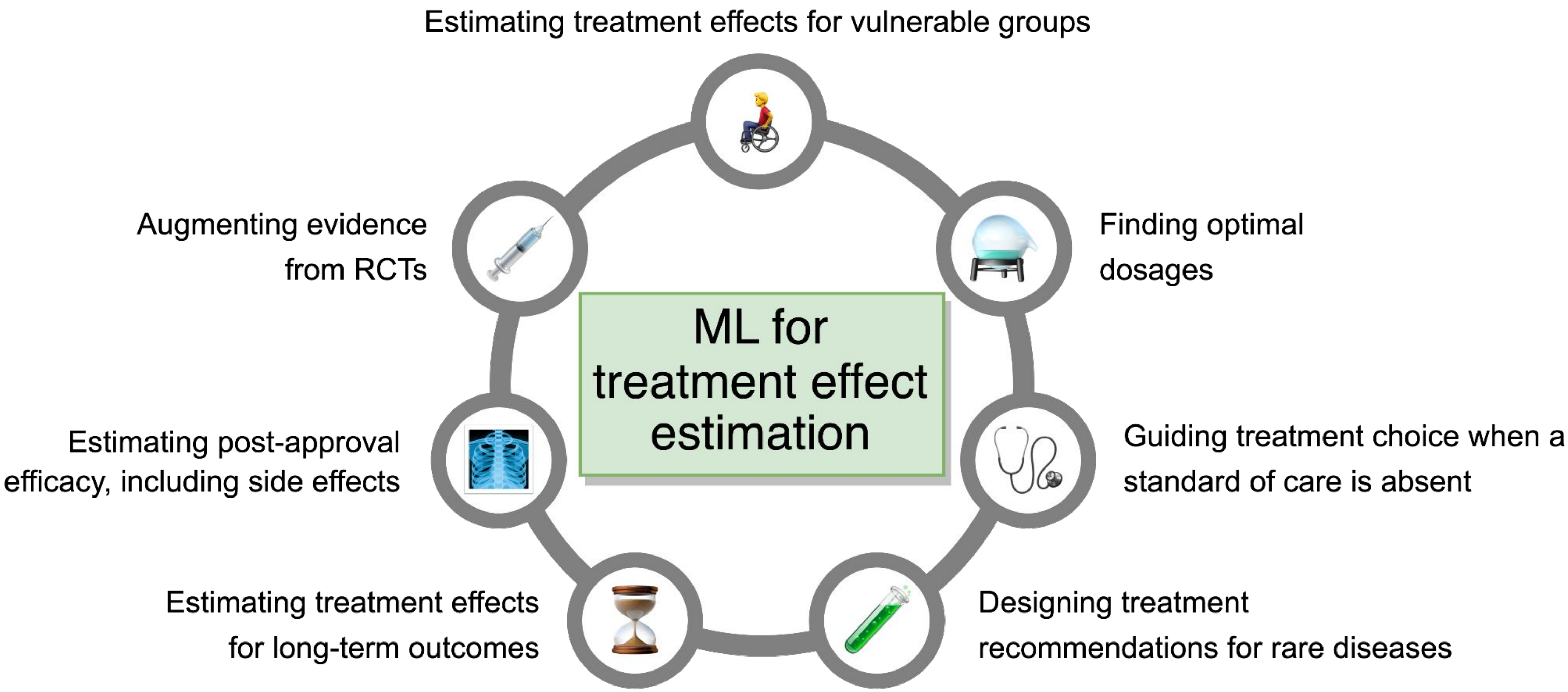


Uncertainty quantification



VISION

Promises of Causal ML



LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

LMU MUNICH SCHOOL OF MANAGEMENT

INSTITUTE OF AI IN MANAGEMENT

Valentyn Melnychuk

Institute of AI in Management
LMU Munich
https://valentyn1997.github.io/
https://www.ai.bwl.lmu.de