

Machine Learning for Healthcare

Mihaela van der Schaar

John Humphrey Plummer Professor of
Machine Learning, Artificial Intelligence and Medicine, **University of Cambridge**
Chancellor's Professor, **University of California Los Angeles**



vanderschaar-lab.com



mv472@cam.ac.uk

@MihaelaVDS



linkedin.com/in/mihaela-van-der-schaar/



Big Thanks to our research team!



Ahmed Alaa



Alex Chan



Alexis Bellot



Alicia Curth



Alihan Hüyük



Boris van Breugel



Changhee Lee



Dan Jarrett



Ioana Bica



James Jordon



Jeroen Berrevoets



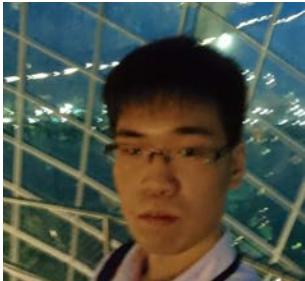
Jonathan Crabbé



Trent Kyono



Yao Zhang



Yuchao Qin



Zhaozhi Qian



Evgeny Saveliev



Nick Maxfield



van_der_Schaar
\\ LAB

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

**“If it were not for the great variability between individuals,
medicine might as well be a science, not an art”**

Sir William Osler (1892)

Our vision: Turn Medicine from Art to Science using AI & ML

- 1) Precision medicine → Bespoke medicine
- 2) Empowering healthcare professionals (clinicians, medical personnel, policy makers)
- 3) Systems, pathways and processes
- 4) Population health and public health policy
- 5) New discoveries – clinical, therapeutics

Precision Medicine to Bespoke Medicine

Patients are complex

- Different genetic background
- Different environmental exposures
- Different lifestyles
- Different histories and interventions etc.

This translates into

- Different risks (and different risks over time)
- Variation in symptoms
- Different health and disease trajectories
- Different responses to treatment etc.

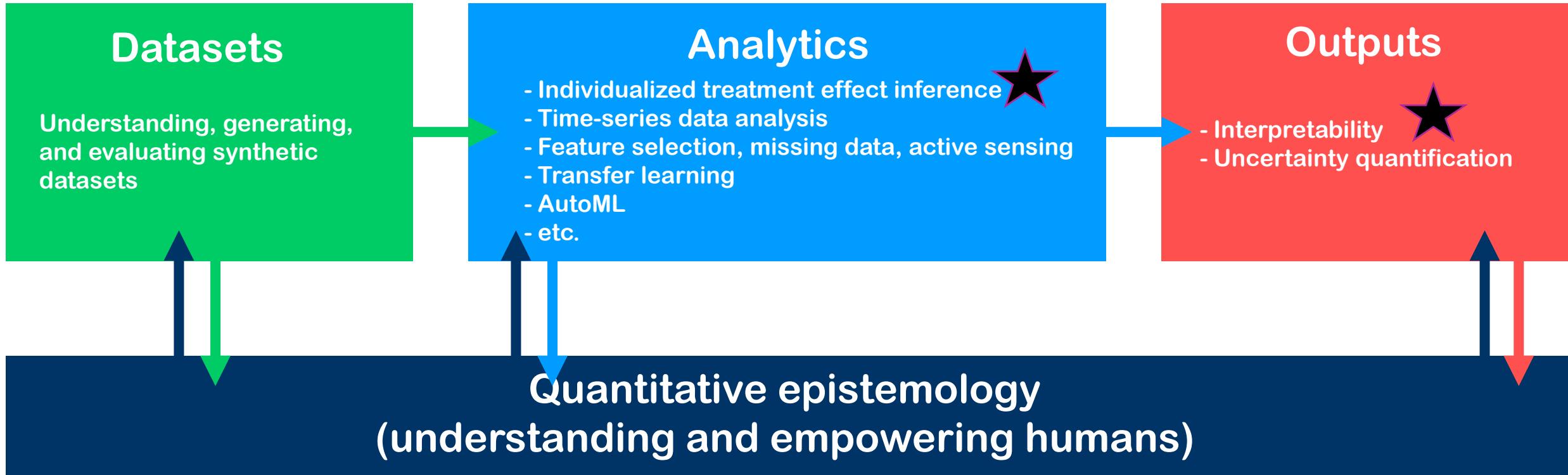
Personalized medicine fits the patient to a pattern

Bespoke medicine recognizes and adapts to changes in patterns with

- age
- lifestyle
- onset of new conditions (or not)
- progress in course of treatment

**How?
Machine learning!**

Our group's research agenda: New ML aimed at revolutionizing healthcare



The cancer pathway





ML-AIM Predictor for Risk Prognosis

Making more informed and dynamic estimates about cancer survival
by learning on diagnosis data and patient events over time

[TRY THE DEMO](#)

Public Health
England

<https://www.youtube.com/watch?v=TWI-WloWvfk>

Personalized therapeutics

Goal: estimate the effect of a **treatment/intervention** on an **individual patient**

Mary



Diagnosed with
Disease X

Which treatment/intervention is best for Mary?



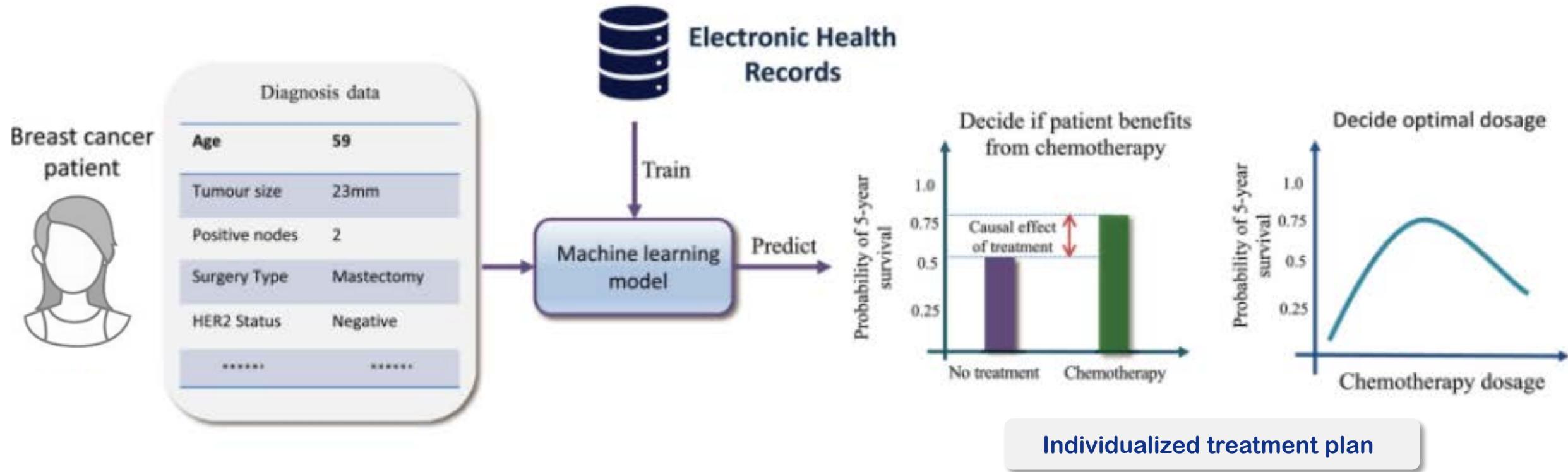
van_der_Schaar
\ LAB

vanderschaar-lab.com

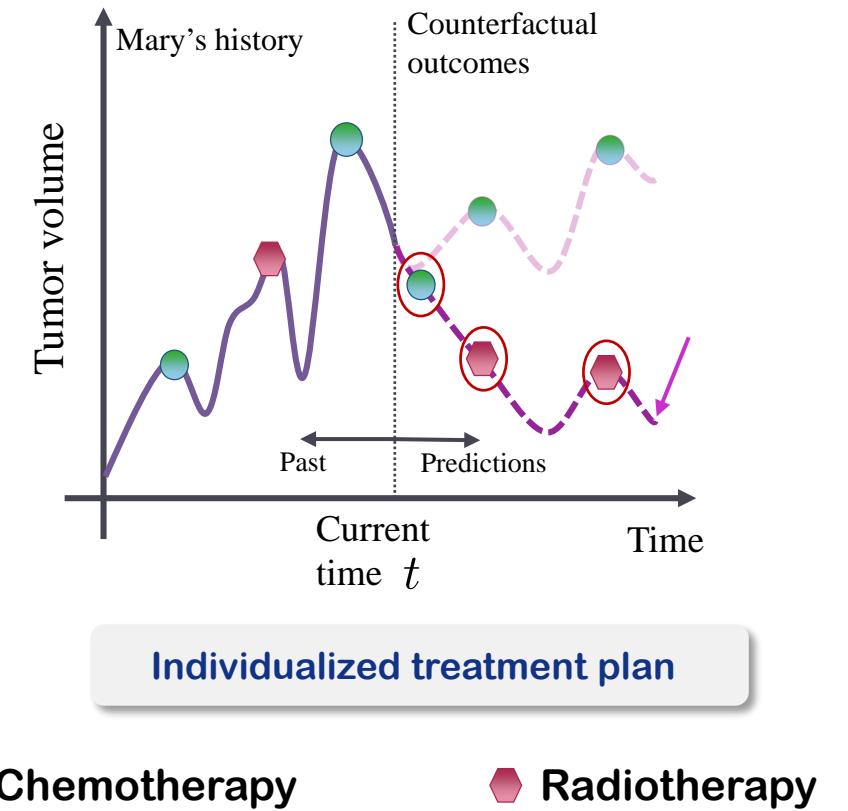
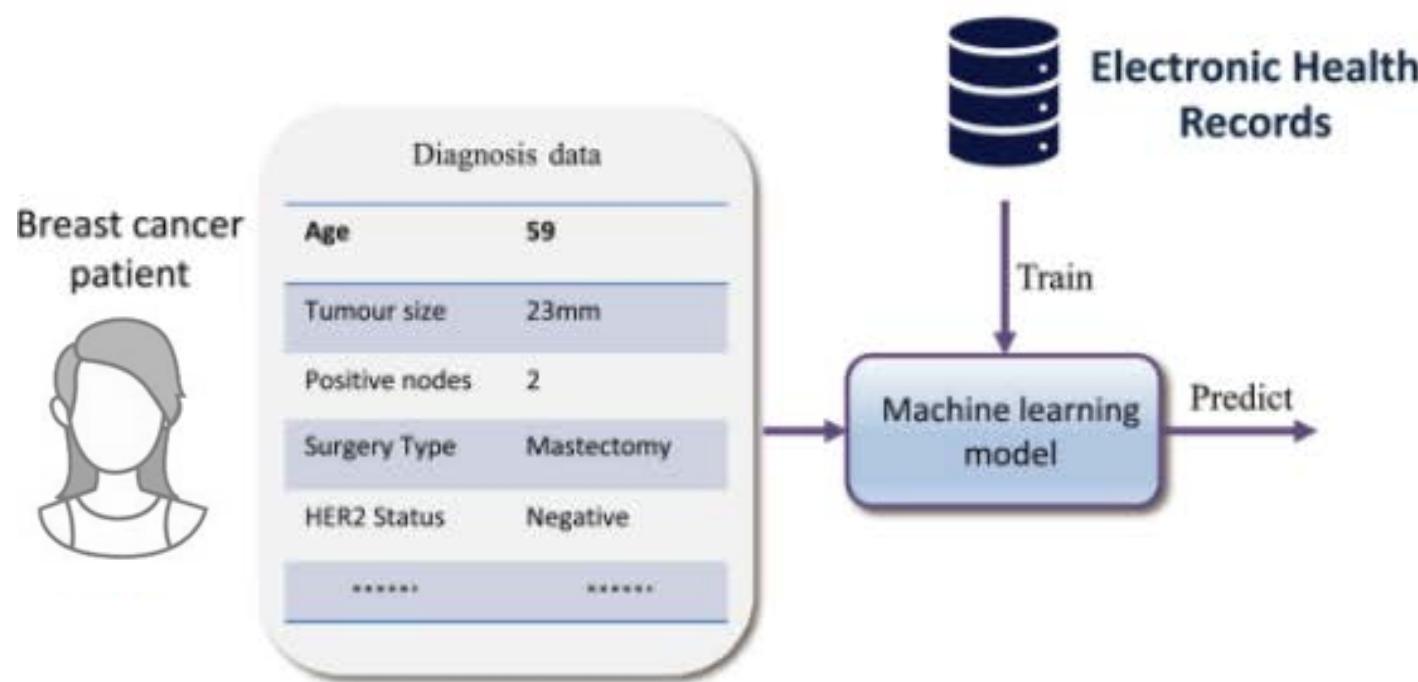


UNIVERSITY OF
CAMBRIDGE

Personalized therapeutics: individualized treatment effects (ITE)



Personalized therapeutics: individualized treatment effects (ITE)



van_der_Schaar
\ LAB

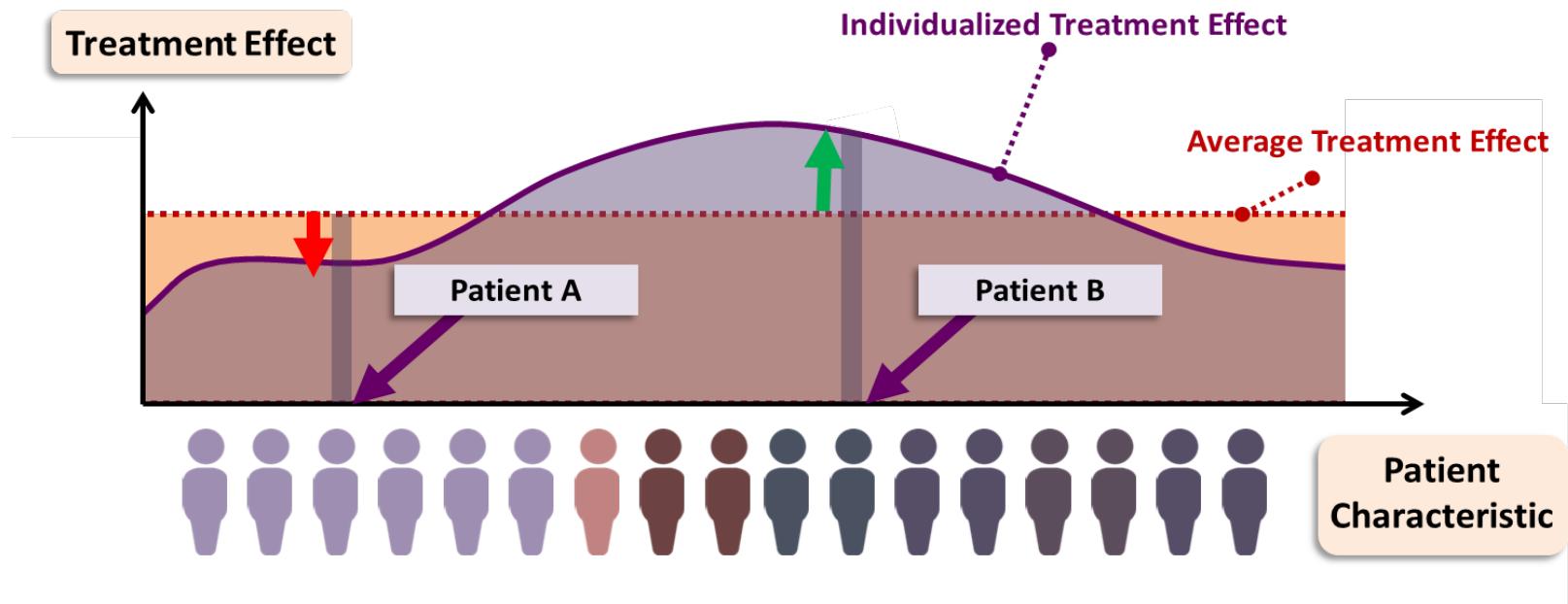
vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Randomized controlled trials - Limitations

Current treatment decisions are based on randomized controlled trials (RCTs)



Treatment effects are often heterogeneous [Willke et al., 2011]



Randomized controlled trials - Limitations

Average treatment effects

Population-level



1. Small sample size
2. Not representative set of patients
3. Focused on very specific question & unable to capture complexity of care
4. Time consuming
5. Costly (in money and resources)
6. Ethical issues

Individualized Treatment Effects from Observational & RCT data

Patient-centric



1. Large sample size
2. Representative set of patients
3. Captures many aspects of care
4. Fast
5. Inexpensive
6. Implementation can be scalable and adaptive



van_der_Schaar
\ LAB

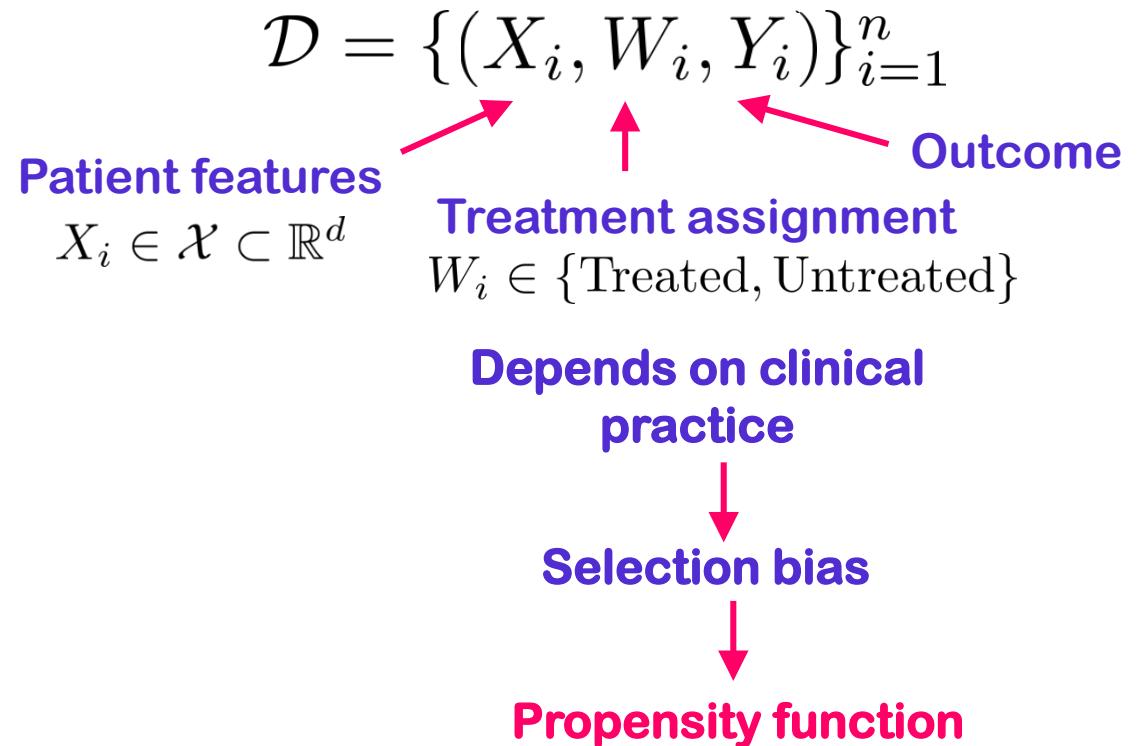
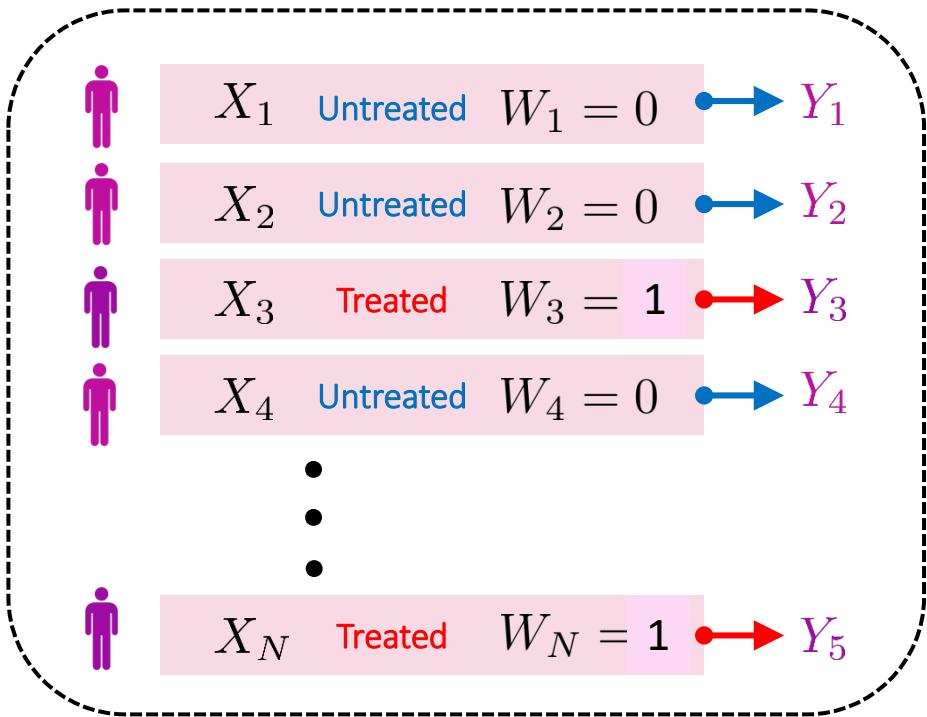
vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Observational data

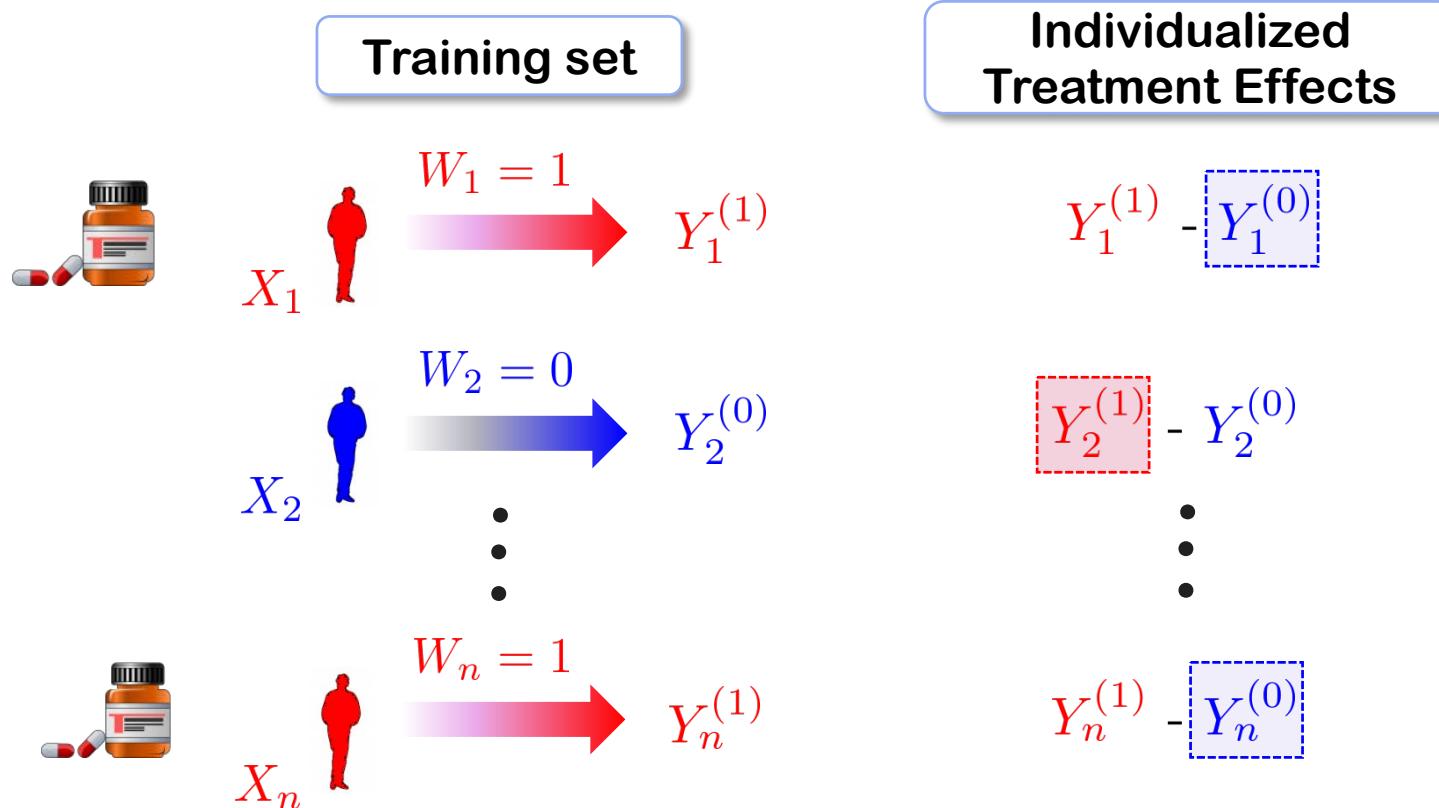
- Observational data: collected from actual clinical practice!



$$p(x) = \mathbb{P}(W_i = 1 | X_i = x)$$

Challenges

Counterfactuals – answering “What if?” questions
We never observe counterfactual outcomes!



A complicated ML problem

Not a simple, “supervised” ML problem - no explicit label!



van_der_Schaar
\ LAB

vanderschaar-lab.com

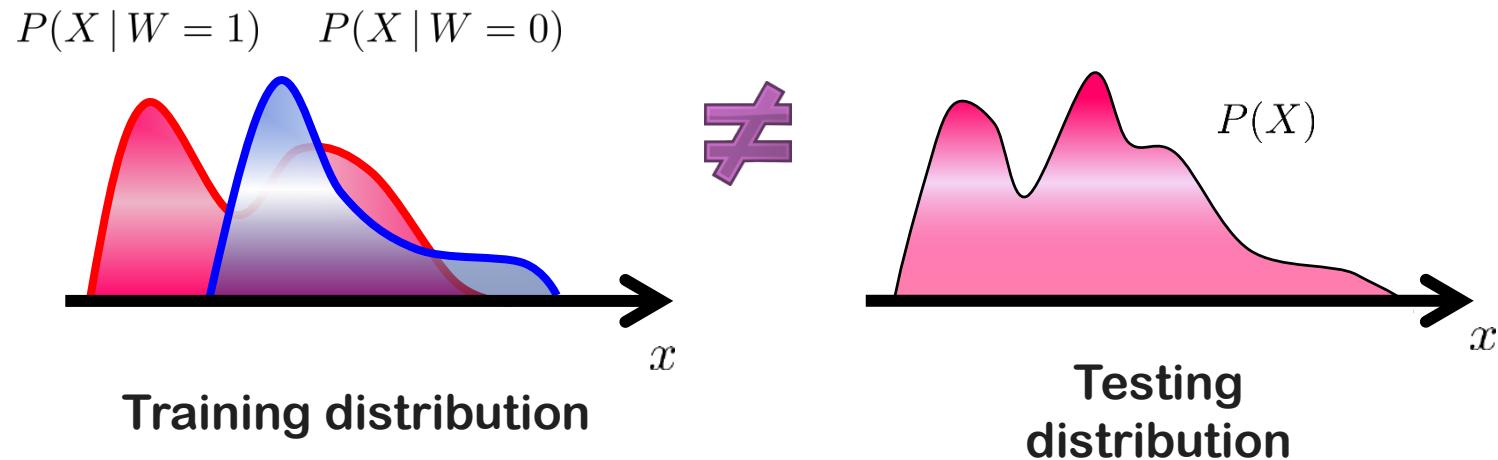


UNIVERSITY OF
CAMBRIDGE

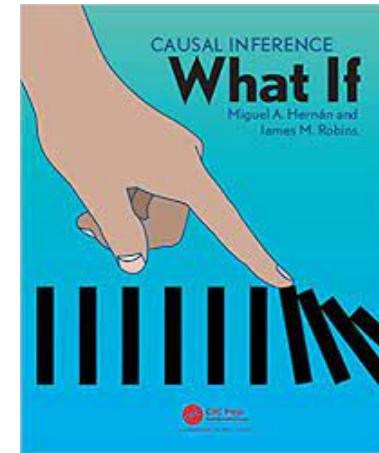
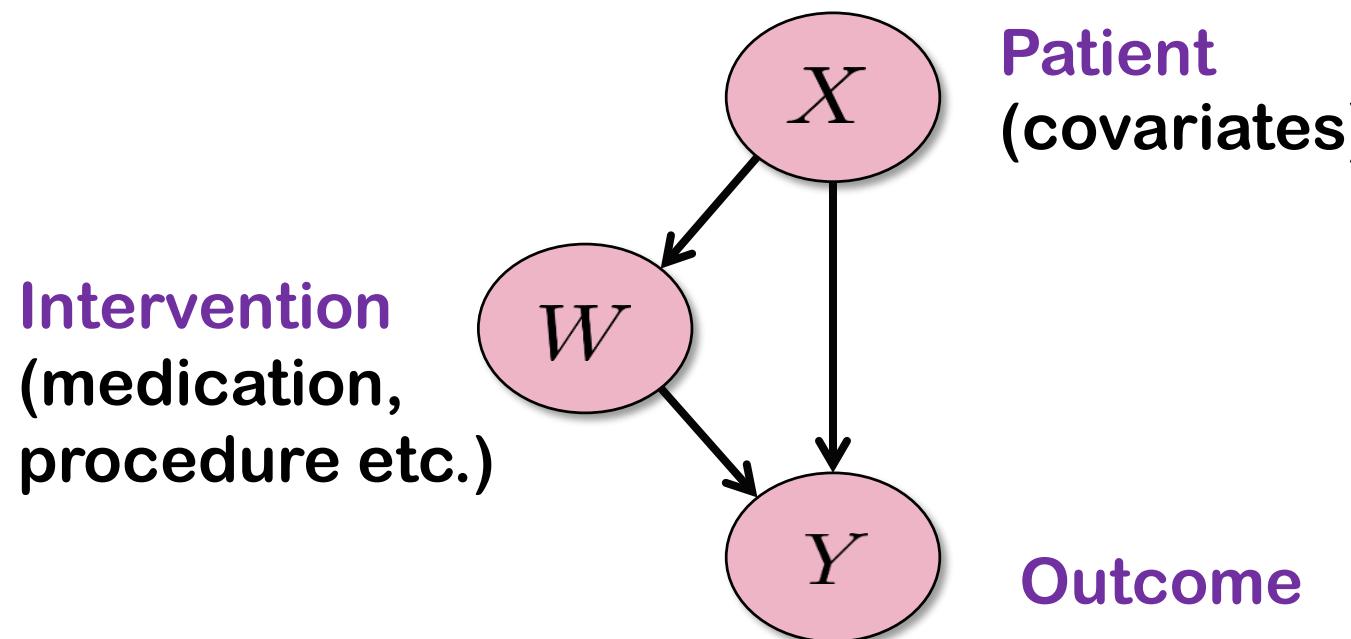
Challenges

1- Need to model interventions (X_i, W_i, Y_i)

2- Selection bias \rightarrow covariate shift:
training distribution \neq testing distribution



Learning ITEs: A causal inference problem



Extensive work in:
- Statistics
- Econometrics



van_der_Schaar
\ LAB

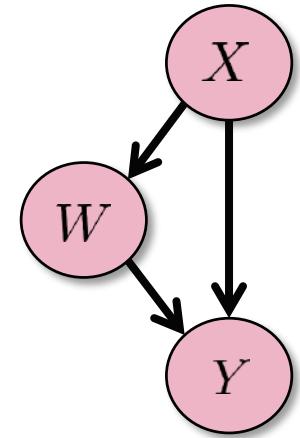
vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Potential outcomes framework [Neyman-Rubin]

Observational data (X_i, W_i, Y_i)



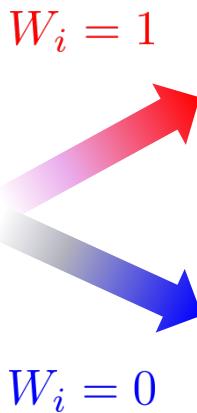
Each patient i has features

$$X_i \in \mathcal{X} \subset \mathbb{R}^d$$



Treatment assignment

$$W_i \in \{0, 1\}$$



Two potential outcomes

$$Y_i^{(1)}, Y_i^{(0)} \in \mathbb{R}$$

$Y_i^{(1)}$ **Treated outcome**

Patient was treated

$Y_i^{(0)}$ **Control outcome**

Patient was not treated



van_der_Schaar
\ LAB

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

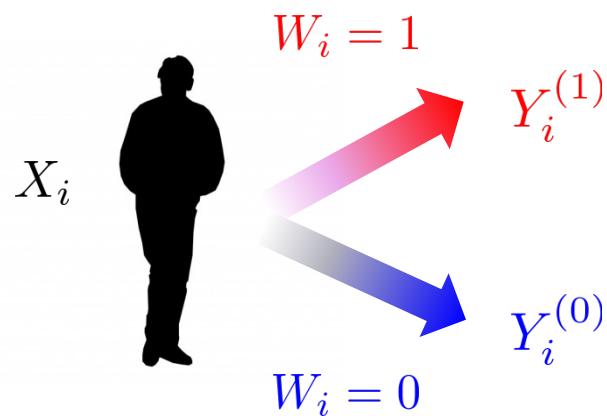
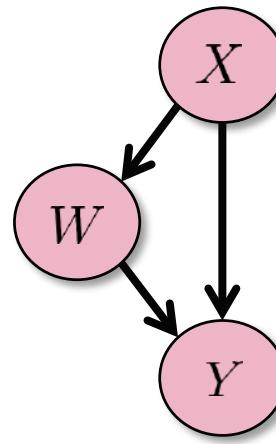
Potential outcomes framework [Neyman-Rubin]

Observational data (X_i, W_i, Y_i)

Each patient i has features $X_i \in \mathcal{X} \subset \mathbb{R}^d$

Two potential outcomes $Y_i^{(1)}, Y_i^{(0)} \in \mathbb{R}$

Treatment assignment $W_i \in \{0, 1\}$



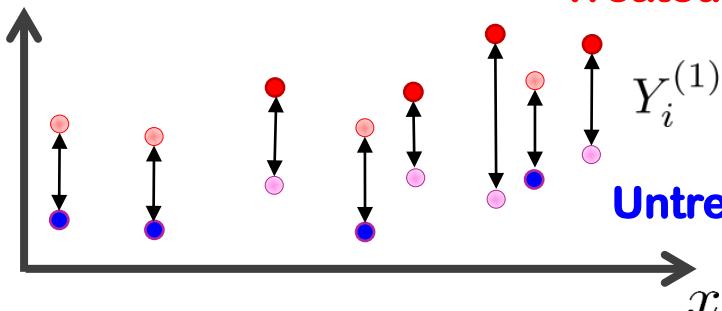
Observed
factual outcomes

$$Y_i = W_i Y_i^{(1)} + (1 - W_i) Y_i^{(0)}$$

Unobserved
counterfactual outcomes

$$Y_i^{\text{CF}} = (1 - W_i) Y_i^{(1)} + W_i Y_i^{(0)}$$

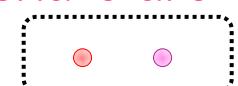
Outcomes



Factual outcomes



Counterfactual outcomes



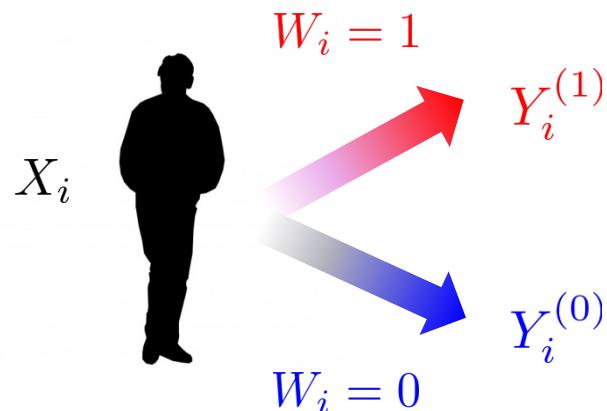
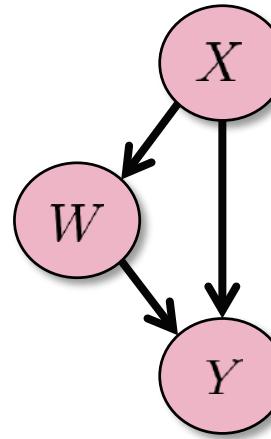
Potential outcomes framework [Neyman-Rubin]

Observational data (X_i, W_i, Y_i)

Each patient i has features $X_i \in \mathcal{X} \subset \mathbb{R}^d$

Two potential outcomes $Y_i^{(1)}, Y_i^{(0)} \in \mathbb{R}$

Treatment assignment $W_i \in \{0, 1\}$



Average treatment effect (ATE)

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

Conditional average treatment effect (CATE) / Individualized treatment effect (ITE)

$$T(x) = \mathbb{E} [Y_i^{(1)} - Y_i^{(0)} \mid X_i = x]$$



van_der_Schaar
\ LAB

vanderschaar-lab.com

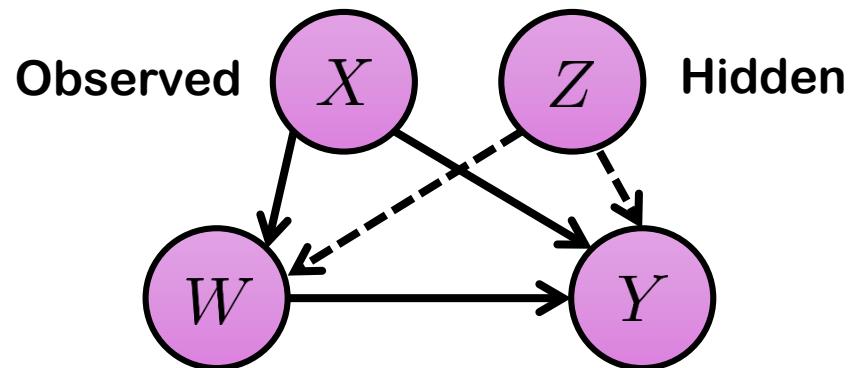


UNIVERSITY OF
CAMBRIDGE

Potential outcomes framework [Neyman-Rubin]

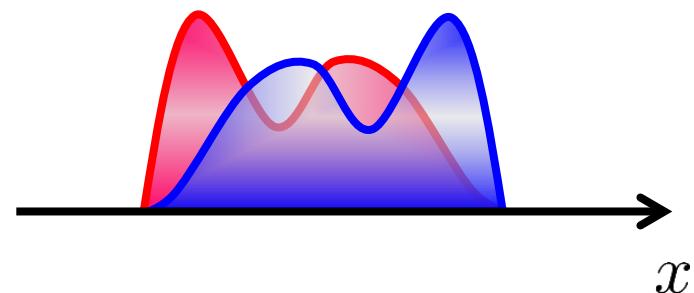
Assumptions

No unmeasured confounders (Ignorability)



Common support

$$\mathbb{P}(W = 1 | X = x) \quad \mathbb{P}(W = 0 | X = x)$$



$$(Y_i^{(0)}, Y_i^{(1)}) \perp\!\!\!\perp W_i | X_i$$

$$\mathbb{P}(W = w | X = x) > 0, \forall x, w$$



van_der_Schaar
\ LAB

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Covariate adjustment

Explicitly model the relation between treatments, covariates and outcomes

Response surfaces

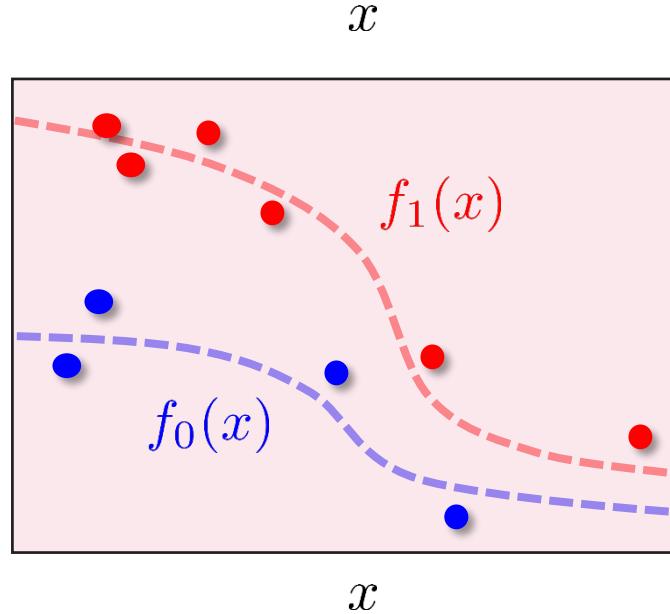
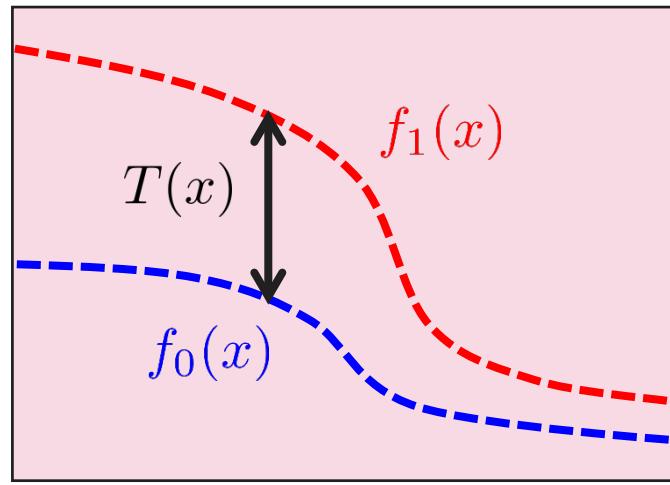
$$f_1(x) = \mathbb{E}[Y^{(1)} | X = x]$$

$$f_0(x) = \mathbb{E}[Y^{(0)} | X = x]$$

Causal effects

$$T(x) = f_1(x) - f_0(x)$$

Used for both ATE and CATE/ITE



Modeling ITEs: Key questions

How should we

- model the treatment assignment variable W ?
- deal with selection bias?



van_der_Schaar
\ LAB

vanderschaar-lab.com



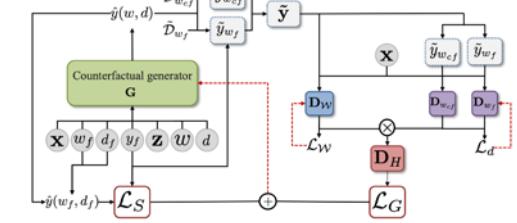
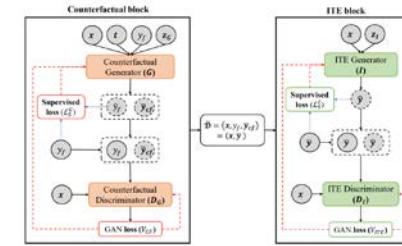
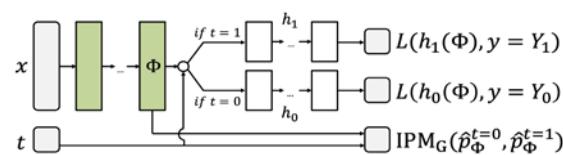
UNIVERSITY OF
CAMBRIDGE

How can machine learning help?

Very flexible, end-to-end model architectures with loss functions crafted for the problem at hand

Modeling the treatment assignment variable W

A large variety of architectures!



Dealing with selection bias

Crafted loss to train any given architecture



$$L(X, W, Y)$$



van_der_Schaar
\\ LAB

vanderschaar-lab.com



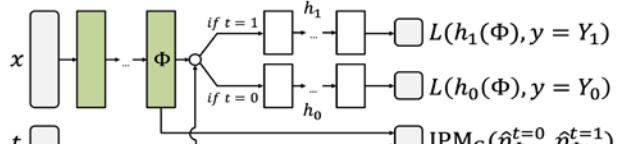
UNIVERSITY OF
CAMBRIDGE

Progress in ML...

Binary treatments

Counterfactual regression

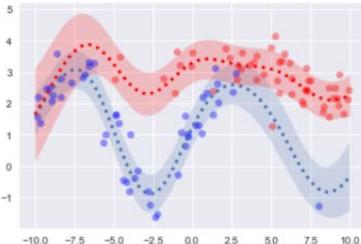
[Shalit, Johansson & Sontag, 2016]



Balanced representations

Multi-task Gaussian processes

[Alaa & van der Schaar, 2017]

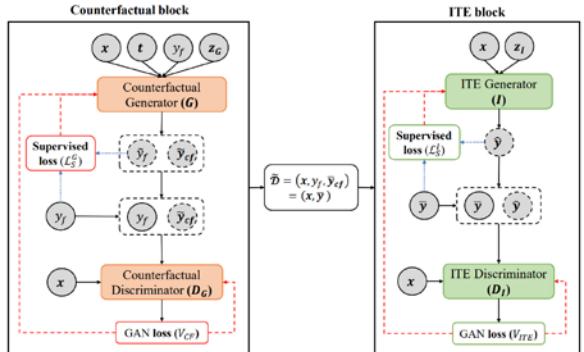


Multi-task learning

Multiple treatments

GANITE

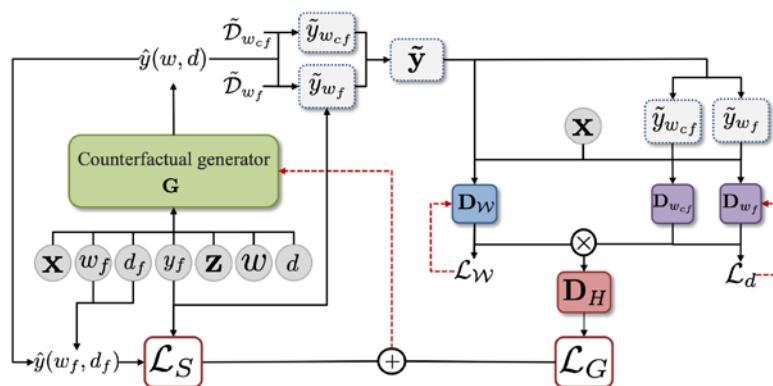
[Yoon, Jordon & van der Schaar, 2018]



Continuous treatments (dosage)

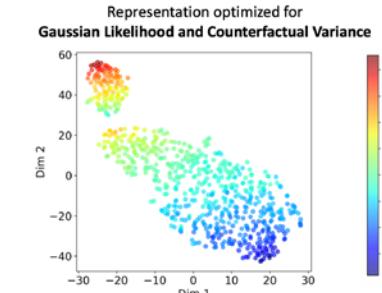
SCIGAN

[Bica, Jordon & van der Schaar, 2020]



Deep Kernel Learning

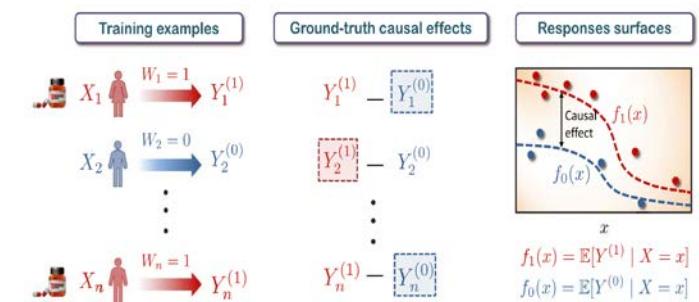
[Zhang, Bellot & van der Schaar, 2020]



Overlapping representations

Review + Next frontiers

Bica, Alaa, Lambert & van der Schaar,
Clinical Pharmacology & Therapeutics, 2020



Our lab's related publications

Causal Inference

28. T. Kyono, I. Bica, Z. Qian, M. van der Schaar "Selecting Treatment Effects Models for Domain Adaptation Using Causal Knowledge", *submitted*, 2021. [\[Link\]](#)
27. A. Curth, M. van der Schaar, "Nonparametric Estimation of Heterogeneous Treatment Effects: From Theory to Learning Algorithms," *accepted to International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021. [\[Link\]](#)
26. C. Xu, A. Alaa, I. Bica, B. D. Ershoff, M. Cannesson, M. van der Schaar, "Learning Matching Representations for Individualized Organ Transplantation Allocation," *accepted to International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021. [\[Link\]](#)
25. Z. Qian, A. Alaa, M. van der Schaar, "When and How to Lift the Lockdown? Global COVID-19 Scenario Analysis and Policy Assessment using Compartmental Gaussian Processes," *Neural Information Processing Systems (NeurIPS)*, 2020. [\[Link\]](#)
24. H.-S. Lee, Y. Zhang, W. Zame, C. Shen, J.-W. Lee, M. van der Schaar, "Robust Recursive Partitioning for Heterogeneous Treatment Effects with Uncertainty Quantification," *Neural Information Processing Systems (NeurIPS)*, 2020. [\[Link\]](#)
23. J. Berrevoets, J. Jordon, I. Bica, A. Gimson, M. van der Schaar, "OrganITE: Optimal transplant donor organ offering using an individual treatment effect," *Neural Information Processing Systems (NeurIPS)*, 2020. [\[Link\]](#)
22. T. Kyono, Y. Zhang, M. van der Schaar, "CASTLE: Regularization via Auxiliary Causal Graph Discovery," *Neural Information Processing Systems (NeurIPS)*, 2020. [\[Link\]](#)
21. I. Bica, J. Jordon, M. van der Schaar, "Estimating the Effects of Continuous-valued Interventions using Generative Adversarial Networks," *Neural Information Processing Systems (NeurIPS)*, 2020. [\[Link\]](#)
20. Y. Zhang, M. van der Schaar, "Gradient Regularized V-Learning for Dynamic Treatment Regimes," *Neural Information Processing Systems (NeurIPS)*, 2020. [\[Link\]](#)
19. I. Bica, A. Alaa, M. van der Schaar, "Time Series Deconfounder: Estimating Treatment Effects over Time in the Presence of Hidden Confounders," *International Conference on Machine Learning (ICML)*, 2020. [\[Link\]](#)
18. W. R. Zame, I. Bica, C. Shen, A. Curth, H.-S. Lee, S. Bailey, J. Weatherall, D. Wright, F. Bretz, M. van der Schaar, "Machine learning for clinical trials in the era of COVID-19," *Statistics in Biopharmaceutical Research - Special Issue on Covid-19*, 2020. [\[Link\]](#)
17. Z. Qian, A. M. Alaa, A. Bellot, J. Rashbass, M. van der Schaar, "Learning Dynamic and Personalized Comorbidity Networks from Event Data using Deep Diffusion Processes," *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020. [\[Link\]](#)
16. Y. Zhang, A. Bellot, M. van der Schaar, "Learning Overlapping Representations for the Estimation of Individualized Treatment Effects," *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020. [\[Link\]](#)
15. I. Bica, A. M. Alaa, J. Jordon, M. van der Schaar, "Estimating Counterfactual Treatment Outcomes over Time through Adversarially Balanced Representations," *International Conference on Learning Representations (ICLR)*, 2020. [\[Link\]](#) - Selected as spotlight presentation
14. I. Bica, A. Alaa, C. Lambert, M. van der Schaar, "From real-world patient data to individualized treatment effects using machine learning: Current and future methods to address underlying challenges," *Clinical Pharmacology & Therapeutics*, 2020. [\[Link\]](#)
13. A. Bellot, M. van der Schaar, "Conditional Independence Testing using Generative Adversarial Networks," *Neural Information Processing Systems (NeurIPS)*, 2019. [\[Link\]](#) [\[Supplementary Materials\]](#)

<https://www.vanderschaar-lab.com/publications/causal-inference>



van_der_Schaar
\\ LAB

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Modeling ITEs: Key questions

How should we

- model the treatment assignment variable W ?
- deal with selection bias?



van_der_Schaar
\ LAB

vanderschaar-lab.com

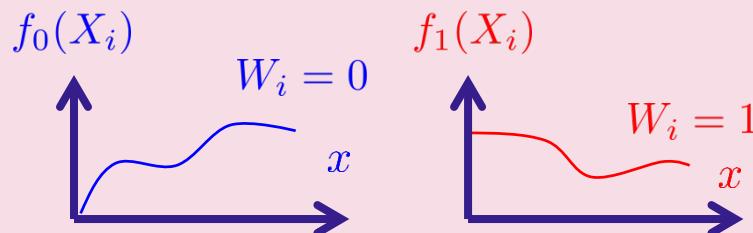


UNIVERSITY OF
CAMBRIDGE

ITE models: A classification [Alaa, van der Schaar, JSTSP 2017]

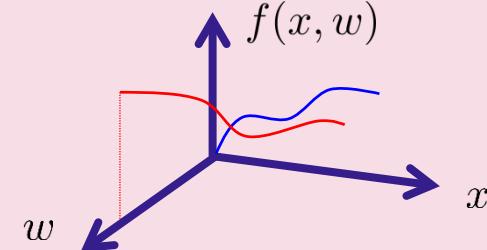
Separate models

W indexes separate models



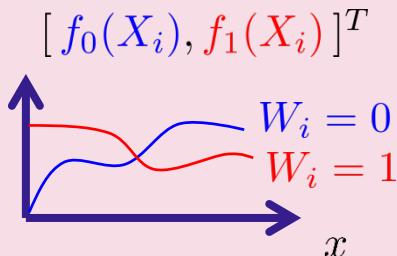
Augment feature space

W is an augmented feature dimension



Shared representations

W indexes shared representations



Many ITE estimation models!

Model	Separate models	Augment to feature space	Shared representations	Handles selection bias?
Causal Multitask GPs [Alaa & van der Schaar, 2017]			X	X
X-learner [Kunzel et. al, 2018]			X	
Balancing neural nets [Johansson et al., 2016]		X		X
Counterfactual regression [Shalit et al., 2017]			X	X
Causal forests [Wager & Athey, 2015]		X		X
BART [Hill, 2011]		X		
Virtual twin random forest [Lu et. al, 2017]	X			



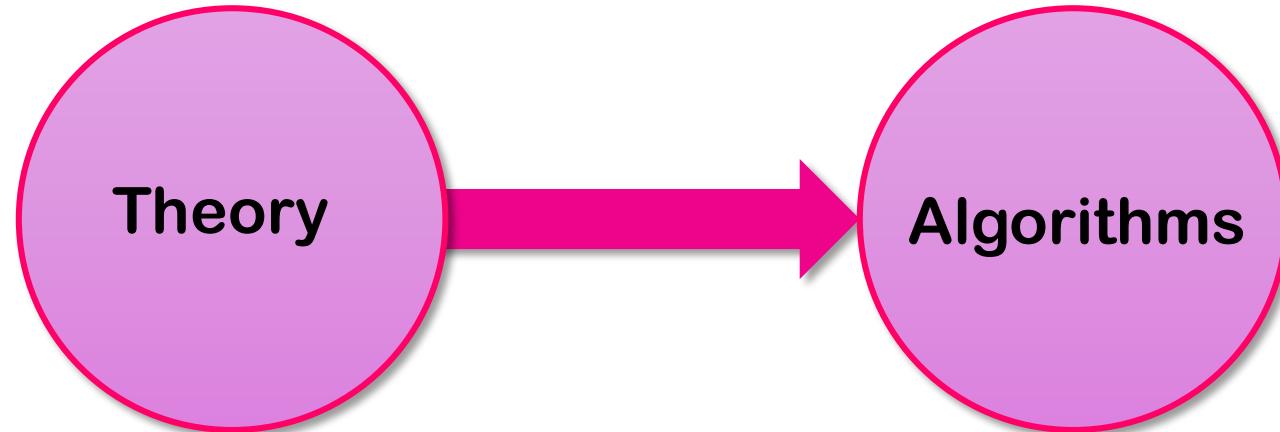
A first theory for individualized treatment effect inference

What is possible?

(Fundamental limits)

How can it be achieved?

(Practical implementation)



[Alaa, van der Schaar, JSTSP 2017][ICML 2018]



**van_der_Schaar
\ LAB**

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Bayesian nonparametric ITE estimation

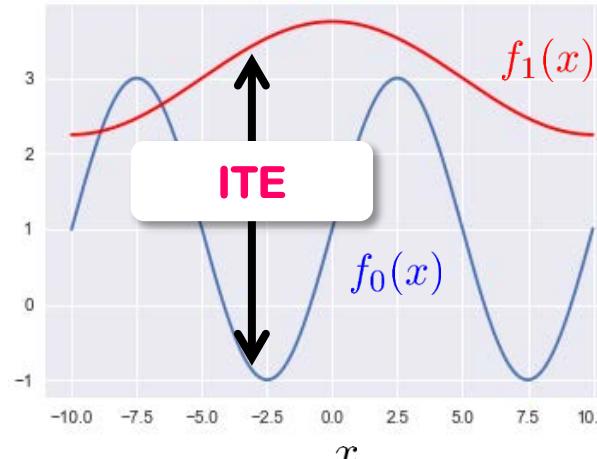
- True ITE model

- Response without treatment

$$Y_i^{(0)} = f_0(X_i) + \epsilon_{i,0}, \quad \epsilon_{i,0} \sim \mathcal{N}(0, \sigma_0^2)$$

- Response with treatment

$$Y_i^{(1)} = f_1(X_i) + \epsilon_{i,1}, \quad \epsilon_{i,1} \sim \mathcal{N}(0, \sigma_1^2)$$



$$T(x) = f_1(x) - f_0(x)$$

- ITE estimation

- Prior over response functions: $f_0, f_1 \sim \Pi$

- “Frequentist” loss of point estimator $\hat{T}(\cdot)$ induced by Bayesian posterior $d\Pi_n(T | \mathcal{D})$

- Precision of estimating heterogeneous effects $\text{PEHE}(\hat{T}) \triangleq \mathbb{E} \| \hat{T} - T \|_{L^2(\mathbb{P})}^2$



van_der_Schaar
\ LAB

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Bayesian nonparametric ITE estimation

Why?

Robust to misspecification & Theoretical analysis

What can be achieved? “Hardness” of a nonparametric estimation problem

Minimax risk = information-theoretic quantity, independent of the model (Stone, 1982)

Minimax estimation risk: $\inf_{\hat{T}} \sup_{f_0, f_1} \text{PEHE}(\hat{T})$

Best estimate

Most “difficult” response surfaces

Minimax rate - fastest rate by which any (Bayesian) estimator \hat{T} can approximate ITE function T



van_der_Schaar
\LAB

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

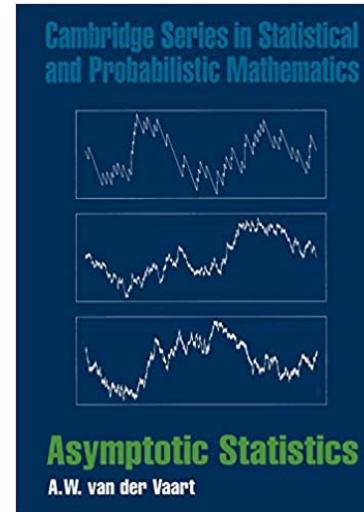
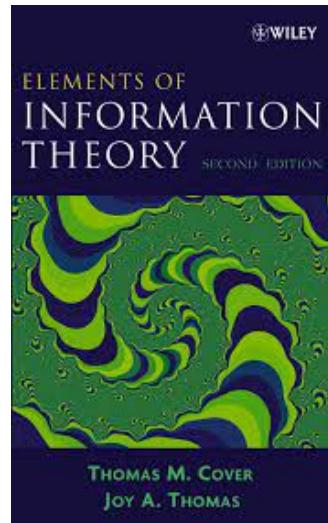
Minimax Rate for ITE Estimation [Alaa & vdS, JSTSP, 2017]

Theorem 1

The minimax rate for ITE estimation is given by:

$$\inf_{\hat{T}} \sup_{f_0, f_1} \text{PEHE}(\hat{T}) \asymp n^{-\left(1 + \frac{1}{2} \left(\frac{d_0}{\alpha_0} \vee \frac{d_1}{\alpha_1}\right)\right)^{-1}}$$

Proof. Information-theoretic techniques based on Fano's method



van_der_Schaar
\ LAB

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Minimax Rate for ITE Estimation

Depends on the “complexity” of $f_0(x)$ and $f_1(x)$...

Sparsity d

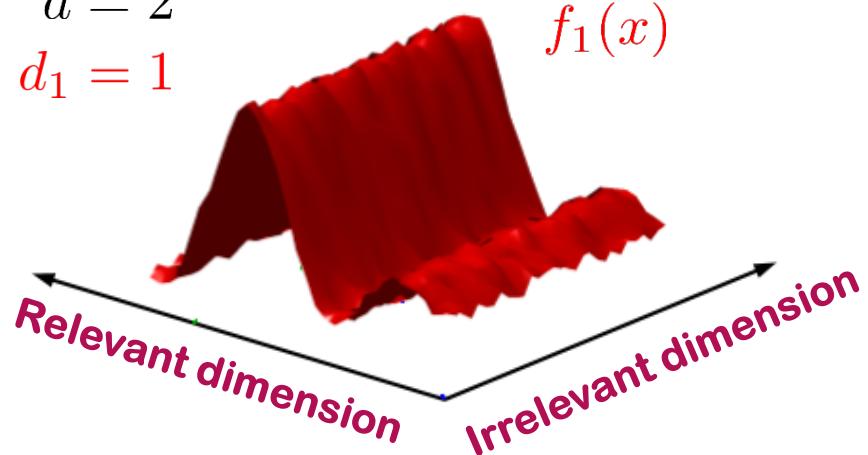
$f_0(x) \rightarrow d_0$ relevant dimensions

$f_1(x) \rightarrow d_1$ relevant dimensions

$$x \in [0, 1]^d, d_0, d_1 \leq d$$

$$d = 2$$

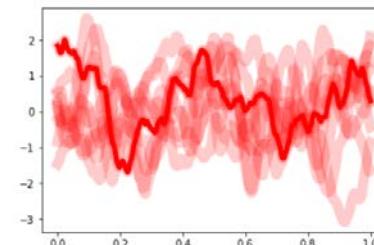
$$d_1 = 1$$



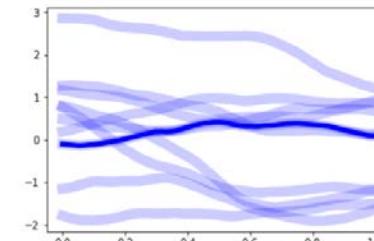
Smoothness α

$f_0(x) \rightarrow$ Hölder ball H^{α_0}

$f_1(x) \rightarrow$ Hölder ball H^{α_1}



$\alpha_1 \downarrow \downarrow$
Rough
functions



$\alpha_0 \uparrow \uparrow$
Smooth
functions



Minimax Rate for ITE Estimation

Theorem 1

The minimax rate for ITE estimation is given by:

$$\inf_{\hat{T}} \sup_{f_0, f_1} \text{PEHE}(\hat{T}) \asymp n^{-\left(1 + \frac{1}{2} \left(\frac{d_0}{\alpha_0} \vee \frac{d_1}{\alpha_1} \right)\right)^{-1}}$$

- To achieve optimal ITE learning rate models should
 - effectively incorporate W such that it encodes different relevant dimensions and smoothness levels of $f_0(x)$ and $f_1(x)$
 - effectively tune “hyperparameters” (i.e. smoothness of the prior)
- Minimax rate does not depend on selection bias



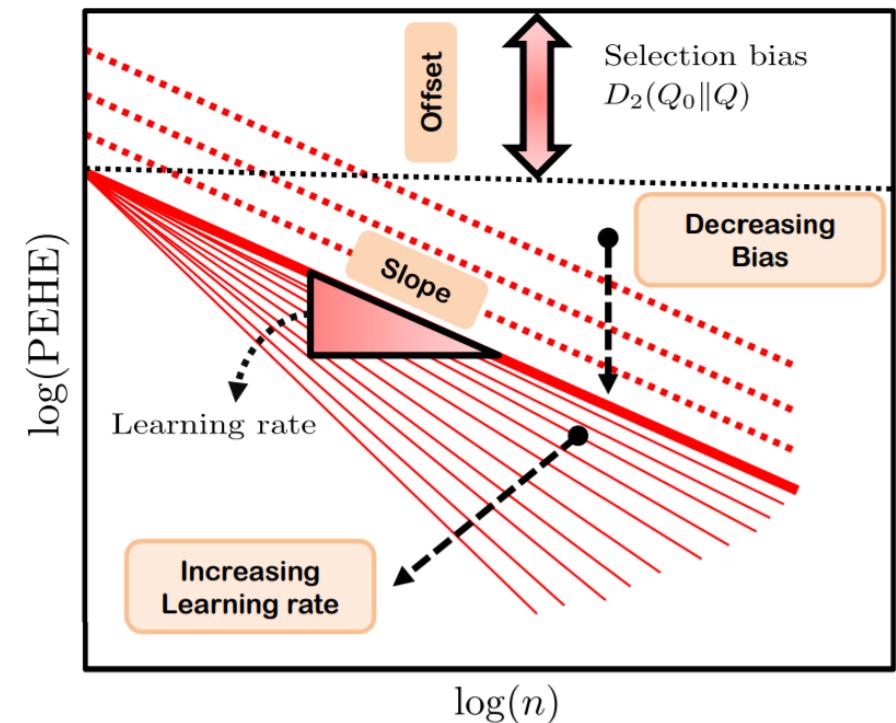
Should we care about selection bias? [Alaa & vdS, ICML 2018]

Assume that $\alpha_0 = \alpha_1$ and $d_0 = d_1$
Minimax-optimal estimator

$$\log(\text{PEHE}(\hat{T})) \approx D_2(\mathbb{P}(X | W = 0) \| \mathbb{P}(X)) + D_2(\mathbb{P}(X | W = 1) \| \mathbb{P}(X)) \\ + \log(C) - \frac{2\alpha_0}{2\alpha_0 + d_0} \log(n).$$

Slope

Rényi
Divergence
↓
Offset
↓



van_der_Schaar
\ LAB

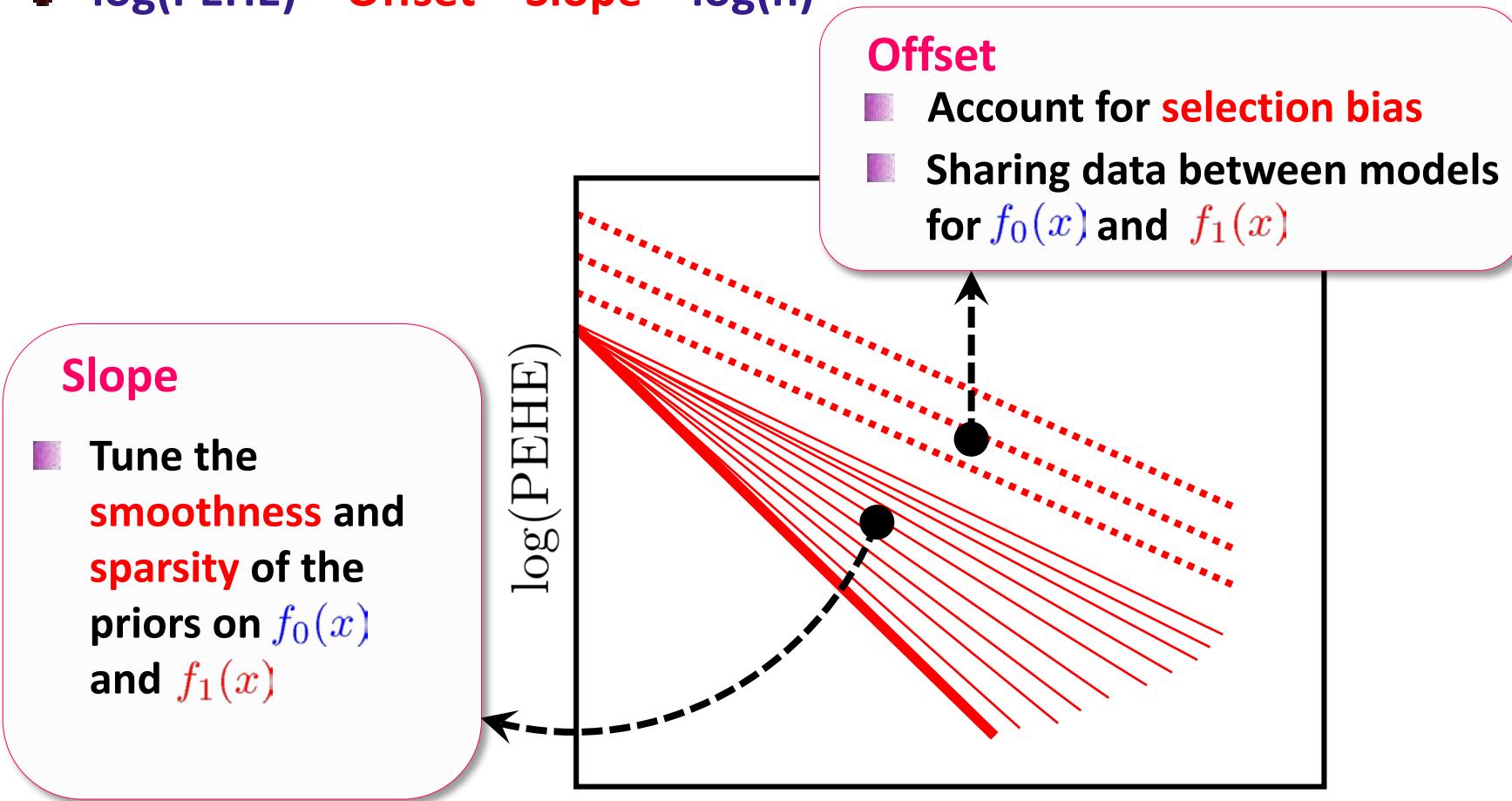
vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Theory guides model design

- $\log(\text{PEHE}) = \text{Offset} - \text{Slope} * \log(n)$



ITE using Non-stationary Gaussian Processes (NSGP)

[Alaa & vdS, NIPS 2017, ICML 2018]

Gaussian process prior (GP) on $f_0(x)$ and $f_1(x)$

$$g \sim \mathcal{GP}(0, \mathbf{K}_\beta((x, w), (x', w')))$$

Multi-tasking! Shared representations!

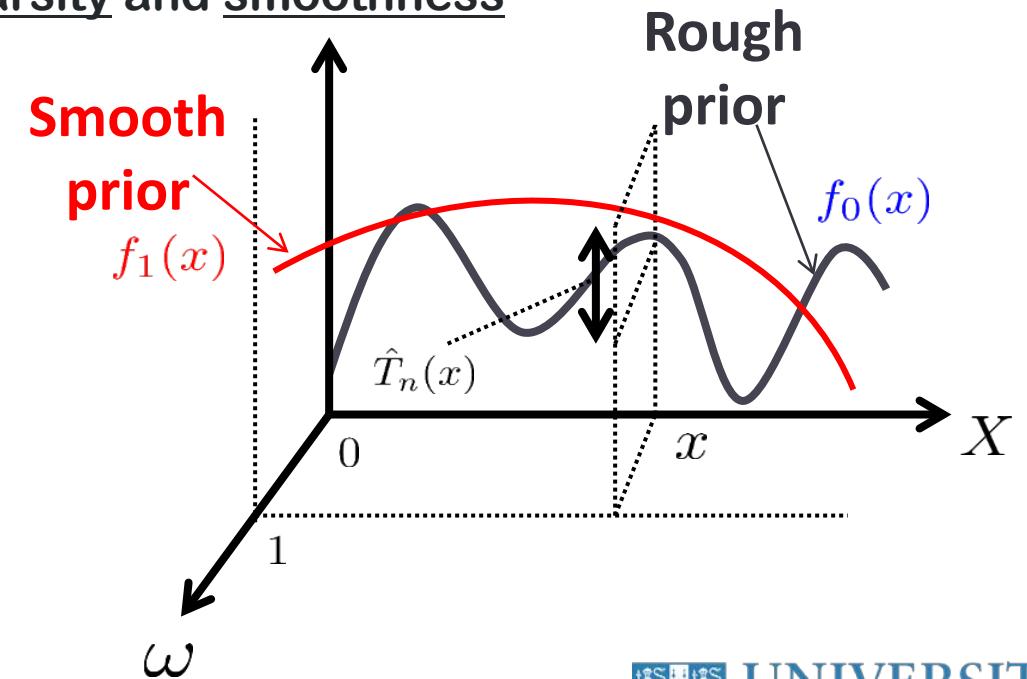
The treatment assignment variable w indexes 2 conditional priors

$f_1(x) = g(x, 1)$ and $f_0(x) = g(x, 0)$ with different sparsity and smoothness

Non-stationary Matérn kernel with ARD

$$\mathbf{K}_\beta((x, 1), (x', 1)) = k_{\beta_1}(x, x')$$

$$\mathbf{K}_\beta((x, 0), (x', 0)) = k_{\beta_0}(x, x')$$



van_der_Schaar
\\ LAB

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Achievable ITE estimation rate

- **Theorem 2**

ITE estimation rate achieved by NSGP:

$$\text{PEHE}(\hat{T}) \lesssim n^{-\frac{2(\alpha_0 \wedge \beta_0)}{2\beta_0 + d_0}} \vee n^{-\frac{2(\alpha_1 \wedge \beta_1)}{2\beta_1 + d_1}}$$

- **Matching condition**

Optimal ITE estimation rate is achieved if:

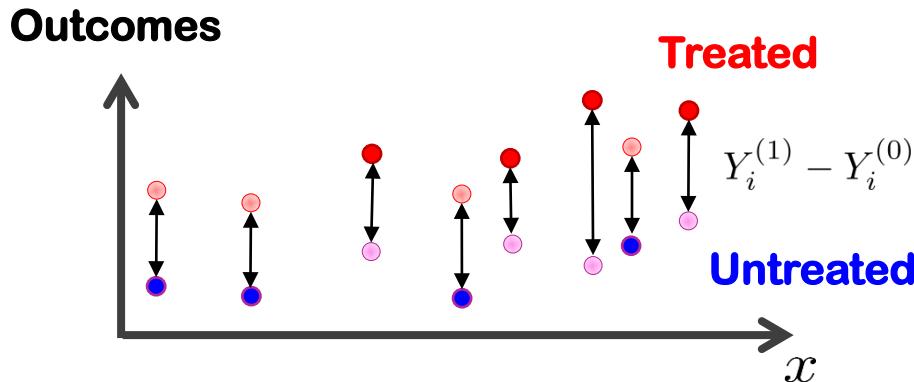
- Only relevant dimensions for $f_0(x)$ and $f_1(x)$ are selected
- Smoothness of prior matches the smoothness of $f_0(x)$ and $f_1(x)$

We need to learn relevant features and smoothness of surfaces!



Hyper-parameter optimization

- No access to empirical PEHE!



- Use efficient PEHE estimator (unbiased, minimum variance) as a cross-validation objective [J. Robbins, 2008]:

$$\hat{\text{PEHE}}^*(\hat{T}) = \sum_{i=1}^n \left(\frac{Y_i^{(W_i)} - (W_i - p(X_i)) \cdot \hat{T}(X_i)}{p(X_i) \cdot (1 - p(X_i))} \right)^2$$

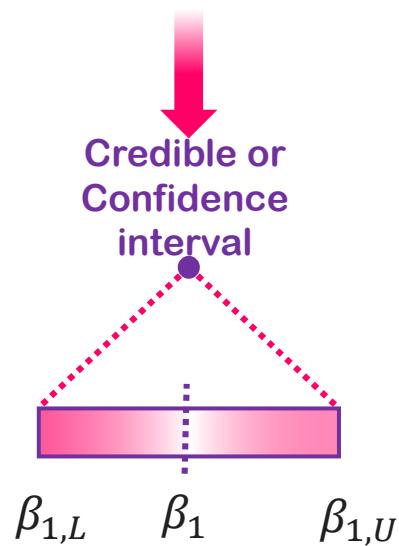
Estimated propensity score

- Matching condition satisfied when $n \rightarrow \infty$
- Accounts for selection bias



How trustworthy are the predictions of the model?

Trustworthiness



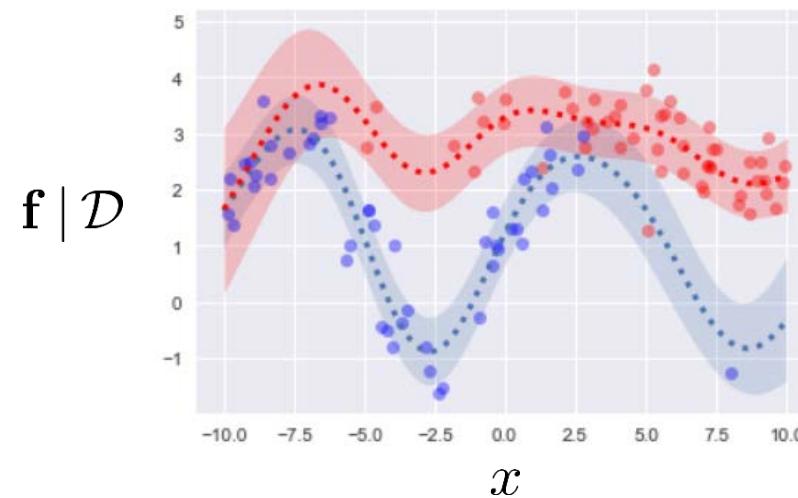
Confidence intervals

- AutoNCP [Zhang, Zame, vdS, 2020]

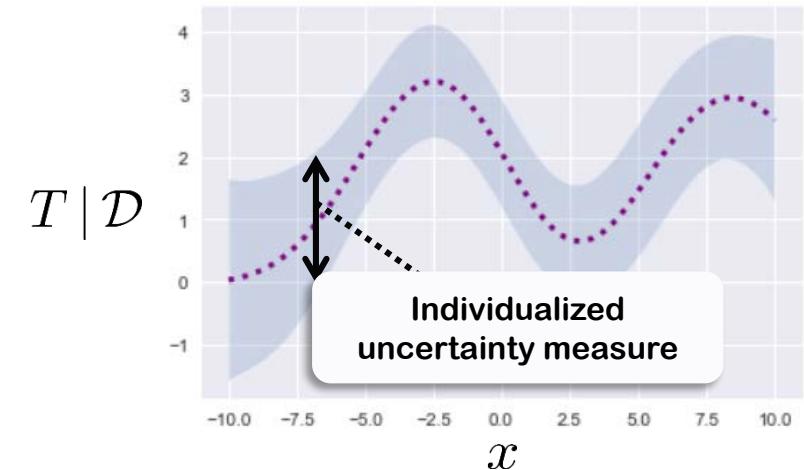
Credible intervals

- Bayesian: NSGP, CMGP [Alaa, vdS, 2017/2018]
- GANITE [Yoon, Jordon, vdS, 2018]

Posterior potential outcomes distribution



Posterior ITE distribution



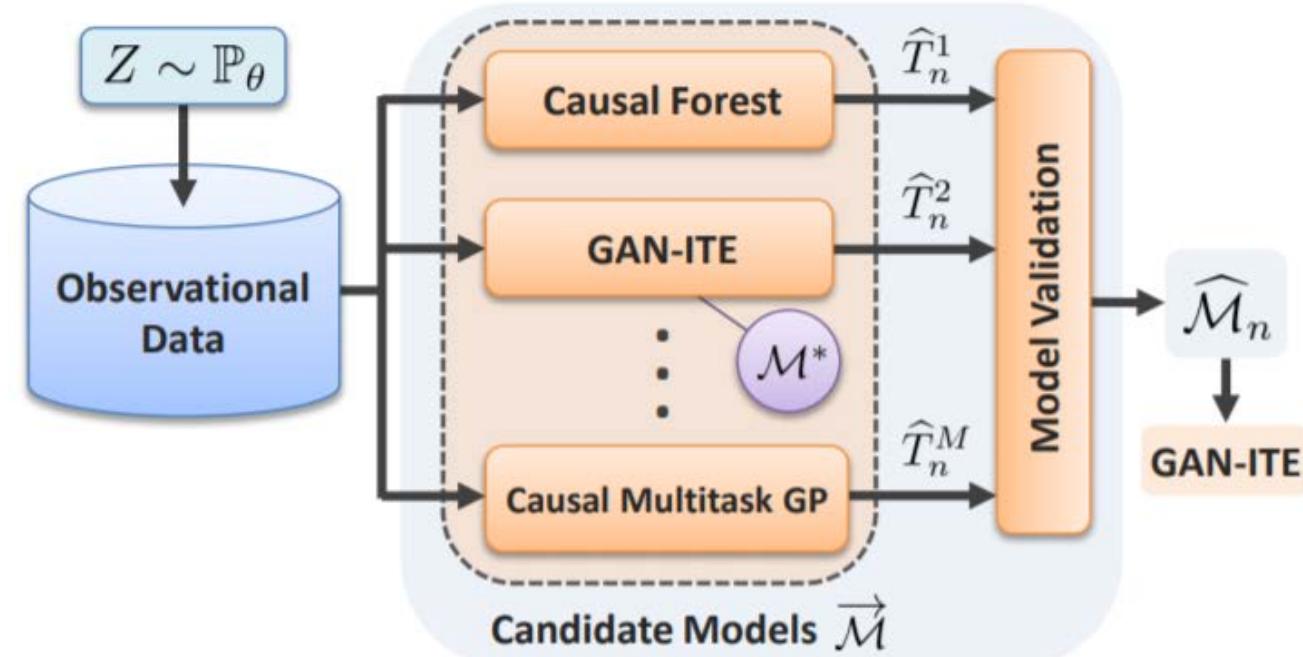
How to select the best model? Automating causal inference! [Alaa&vdS, ICML 2019]

Numerous competing models

Different models may be best for different observational studies

Which one to select?

BNN	ICML 2016
CMGP	NIPS 2017
TARNet	ICML 2017
CFR Wass.	ICML 2017
CFR MMD	ICML 2017
NSGP	ICML 2018
GANITE	ICLR 2018
SITE	NIPS 2018
BART	
Causal Forest	

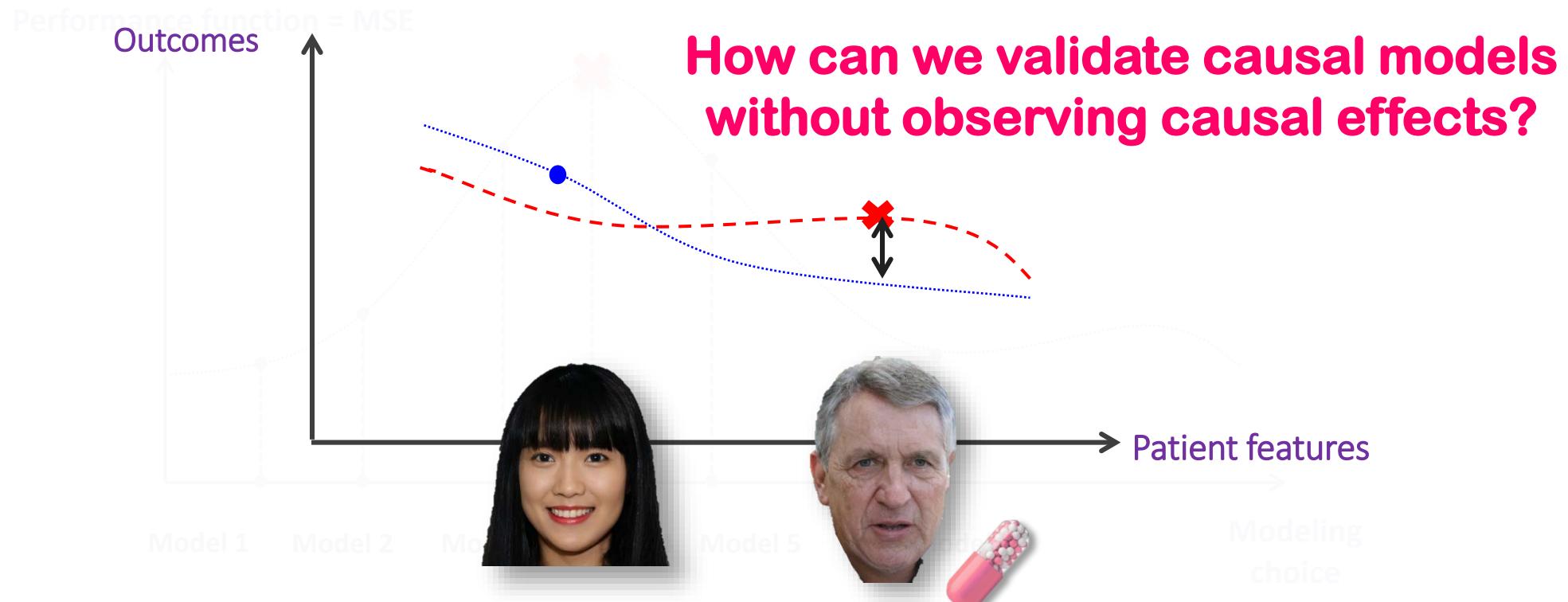


Key Challenge: How to do cross-validation?

Automating causal inference! [Alaa&vdS, ICML 2019]

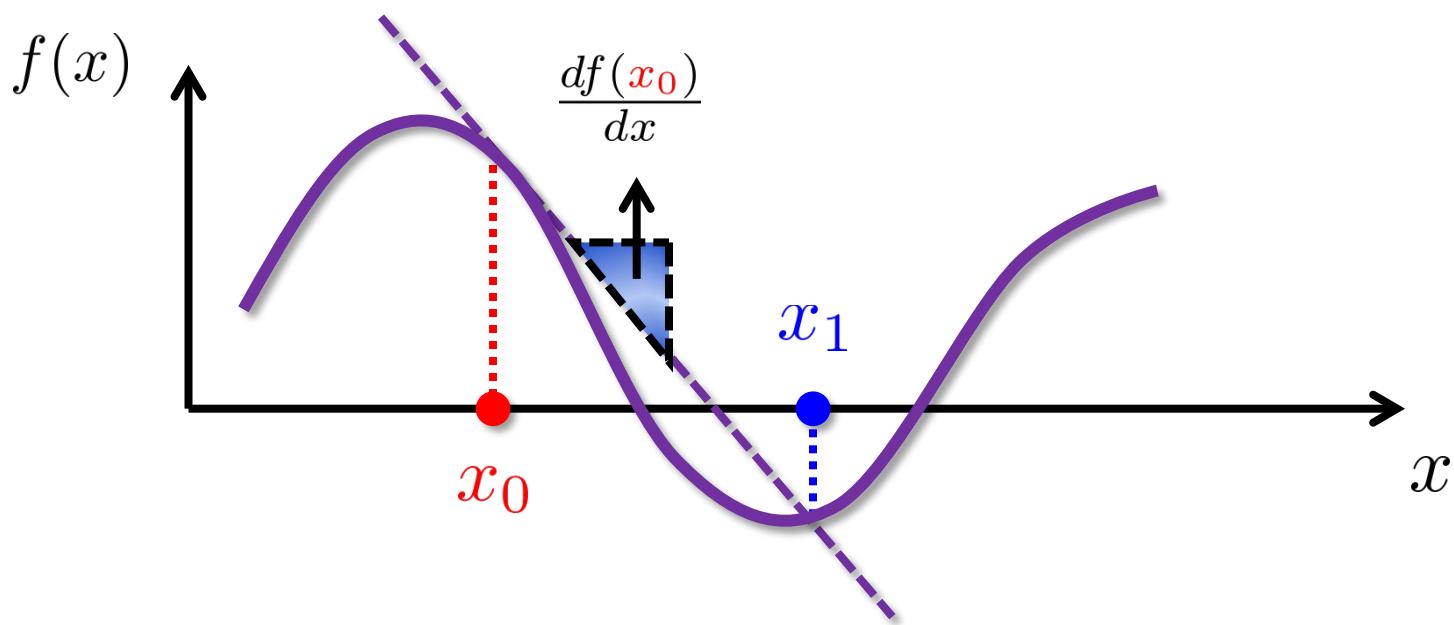
- PEHE: MSE of a causal model

$$\ell_\theta(\hat{T}) = \|T(X) - \hat{T}(X)\|_\theta^2$$



Taylor series approximation

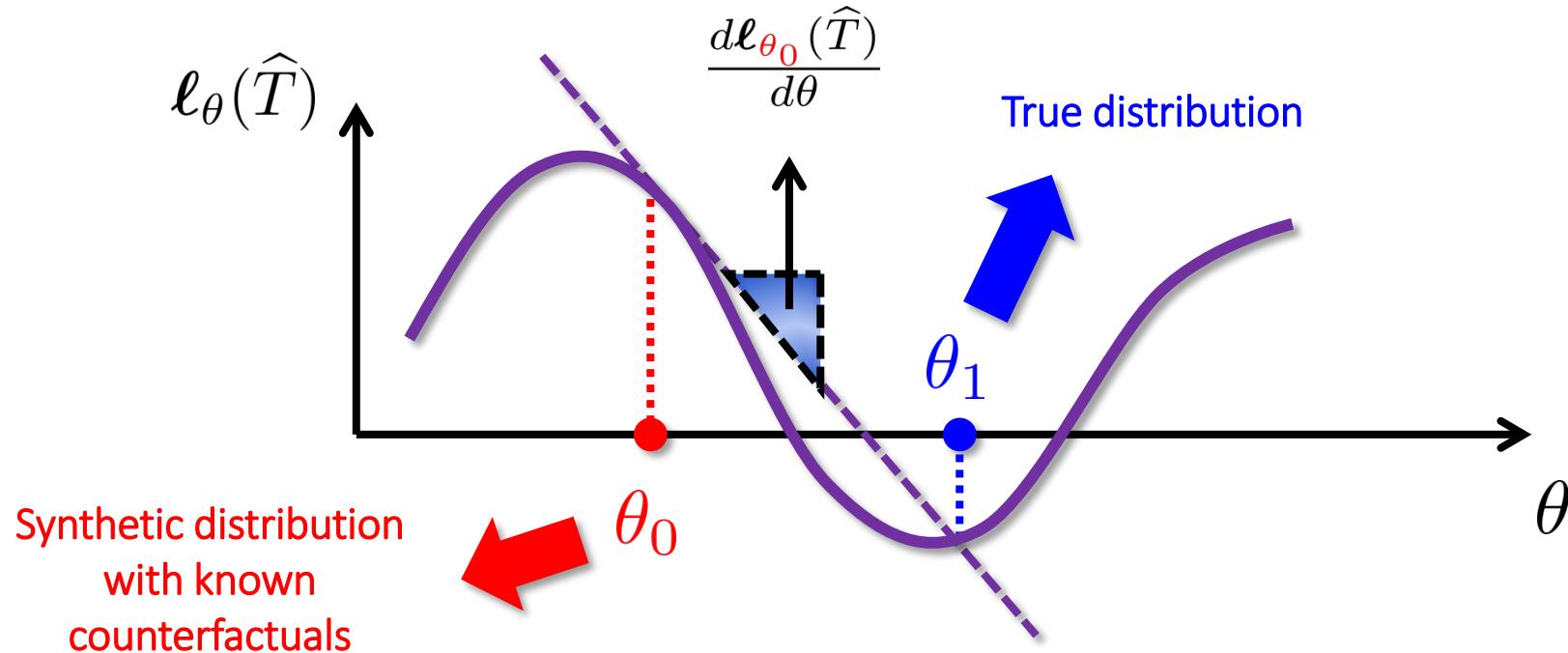
- Value of a function at a given input can be predicted using its value and (higher-order derivatives) at a proximal input



$$f(x_1) = f(x_0) + (x_1 - x_0) \frac{df(x_0)}{dx} + \frac{1}{2!} (x_1 - x_0)^2 \frac{d^2 f(x_0)}{dx^2} + \dots$$

Analogy with Taylor series approximation

- Performance of a causal inference model is a **functional** of data-generating distribution \mathbb{P}_θ



Functional calculus: von Mises expansion

- A “distributional” analog of Taylor expansion [Fernholz, 1983]

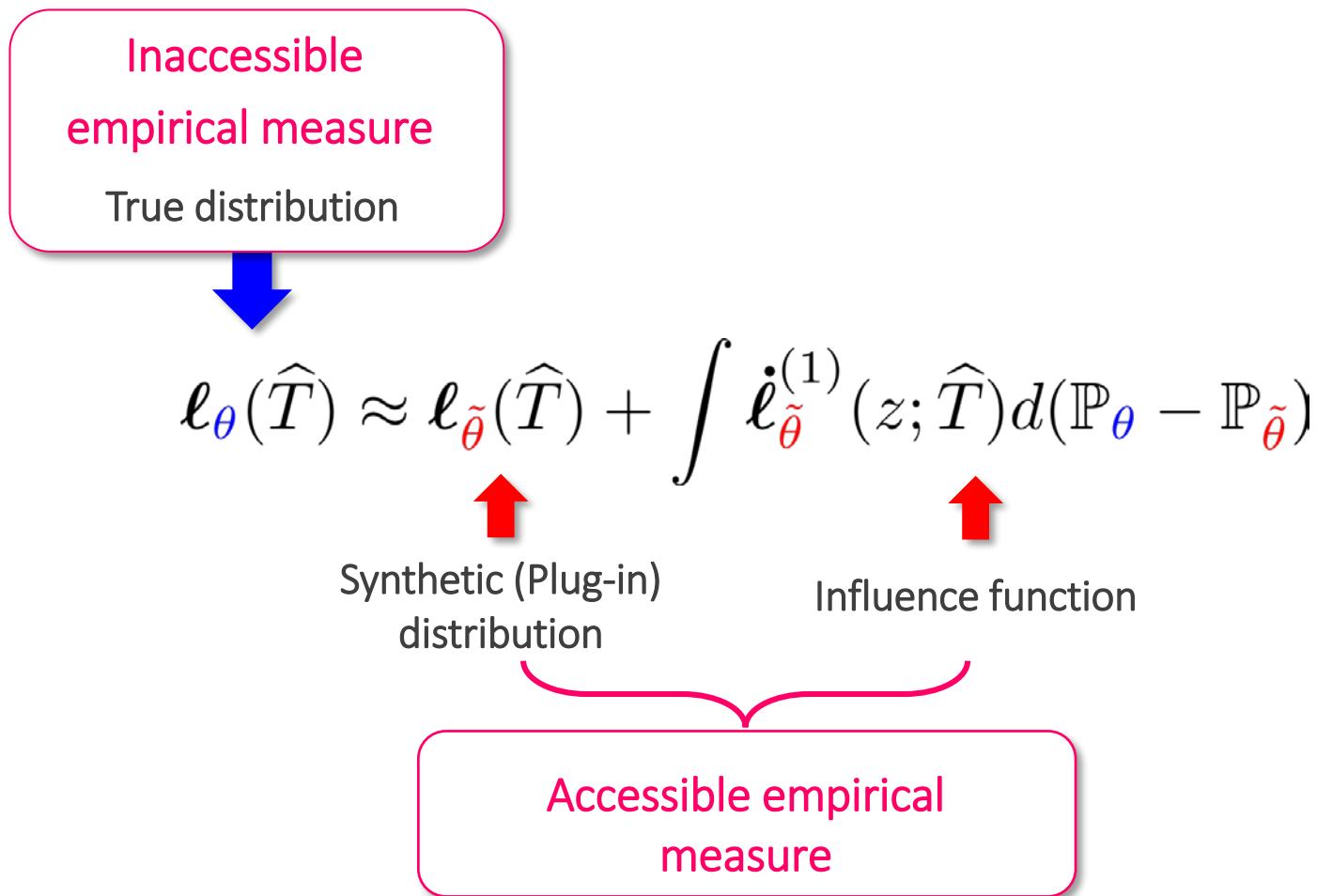
$$\begin{aligned}\ell_{\theta_1}(\hat{T}) &= \ell_{\theta_0}(\hat{T}) + \int \dot{\ell}_{\theta_0}^{(1)}(z; \hat{T}) d(\mathbb{P}_{\theta_1} - \mathbb{P}_{\theta_0}) \\ &\quad + \frac{1}{2!} \int \dot{\ell}_{\theta_0}^{(2)}(z; \hat{T}) d(\mathbb{P}_{\theta_1} - \mathbb{P}_{\theta_0})^2 + \dots\end{aligned}$$

- **Influence Functions \leftrightarrow Derivatives**

We can predict performance of a causal inference model using the **influence functions (IF)** of its loss on a “similar” synthetic dataset.

Estimating a causal model's performance

- First-order “Taylor approximation”



How to estimate a model's performance?

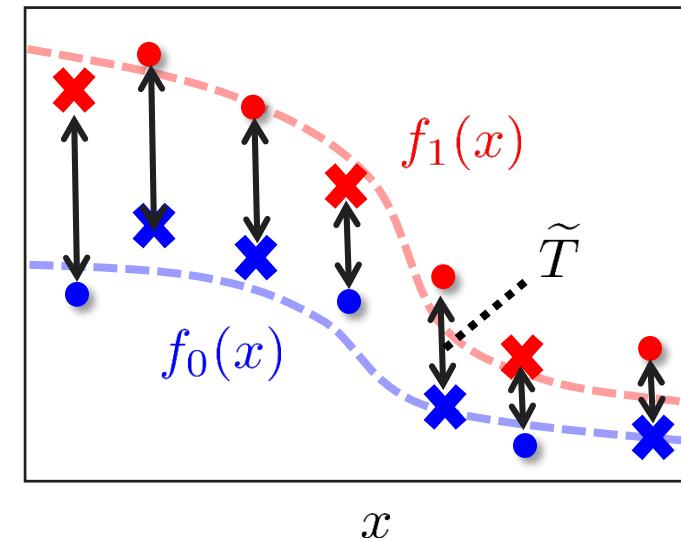
- Synthesize counterfactuals!

Step 1: Plug-in estimation

- Plug-in model \tilde{T}
- Plug-in PEHE loss $\ell_{\tilde{\theta}}(\hat{T})$

Step 2: Bias correction

$$\ell_{\theta}(\hat{T}) = \ell_{\tilde{\theta}}(\hat{T}) + \int \dot{\ell}_{\tilde{\theta}}^{(1)}(z; \hat{T}) d\mathbb{P}_{\theta}$$



AutoML for causal inference! [Alaa&vdS, ICML 2019]

- Average performance on the **77** benchmark datasets.
- No absolute single winner on all datasets.
- Our IF-based selection is better than any single model.

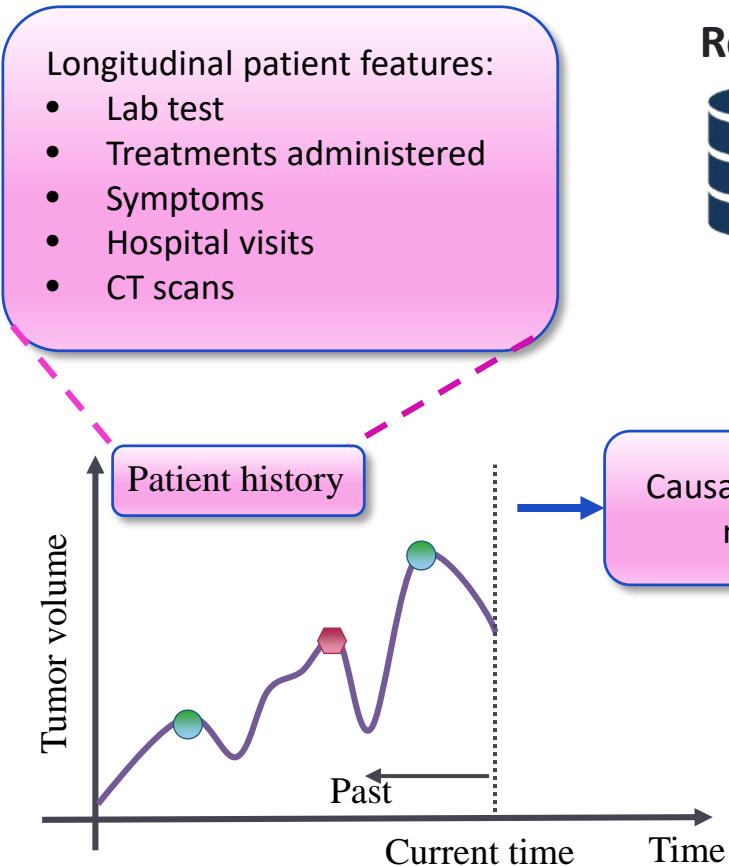
Method	% Winner
BNN	3%
CMGP	12%
NSGP	17%
TARNet	8%
CFR Wass.	9%
CFR MMD	12%
GANITE	7%
SITE	7%
BART	15%
C. Forest	7%
AutoML IF-based	72%

Individualized Treatment effects over time

Breast cancer patient



Diagnosis (baseline) information



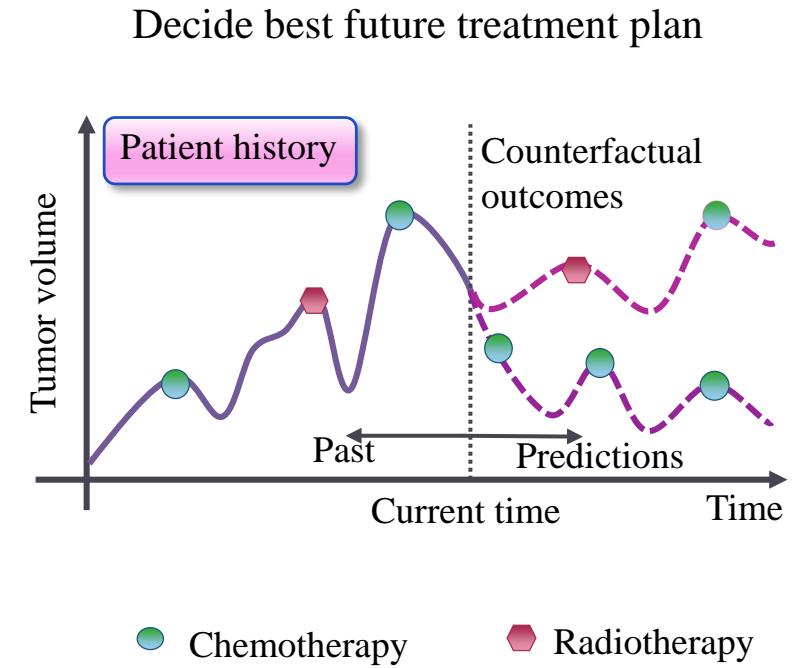
Electronic Health
Records



Train

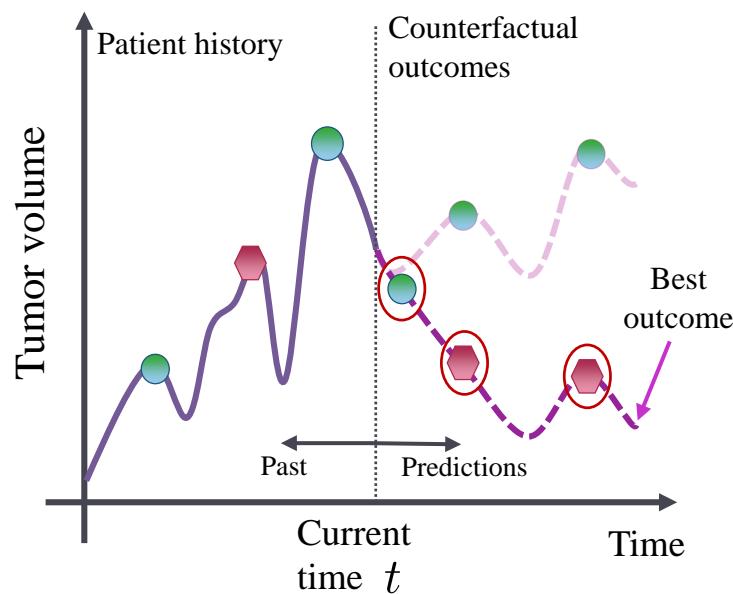
Causal inference
model

Estimate
counterfactual
trajectories



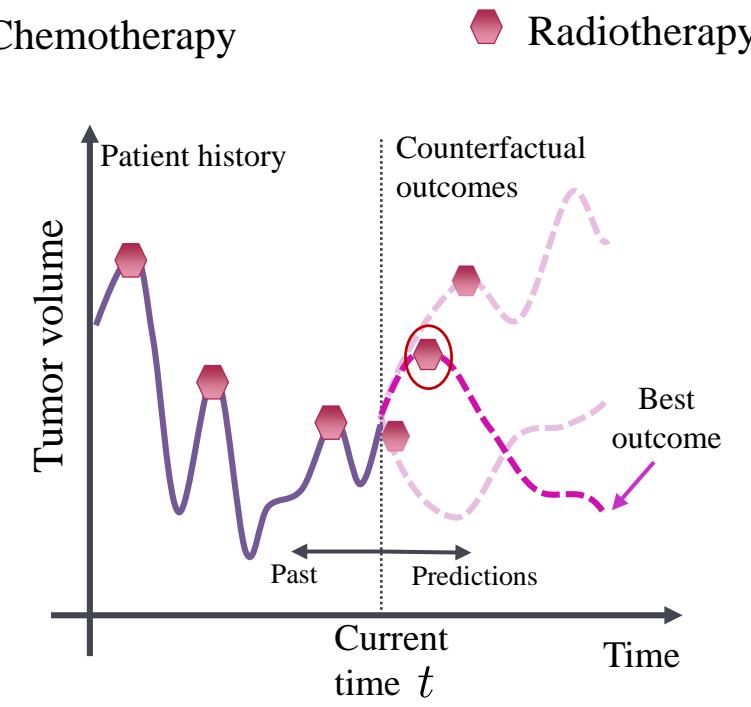
Individualized Treatment effects over time

How to treat?



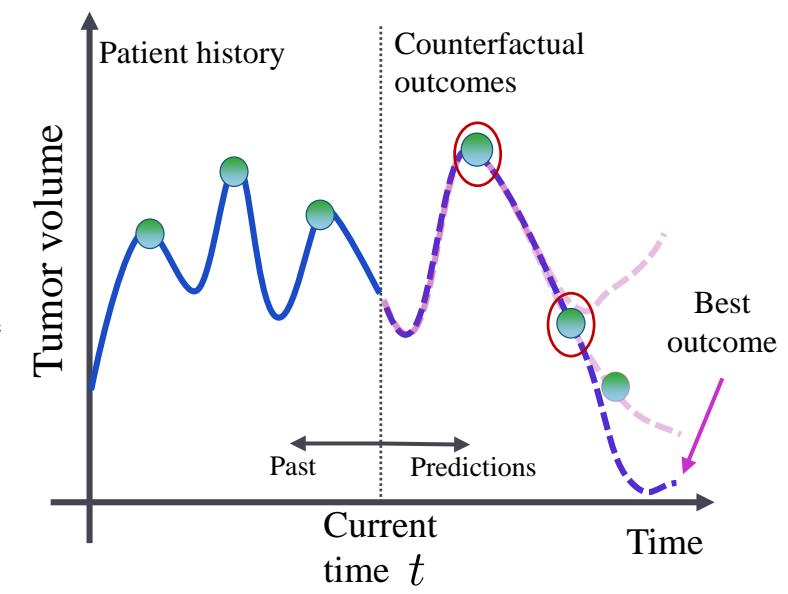
(a) Decide treatment plan

When to give treatment?



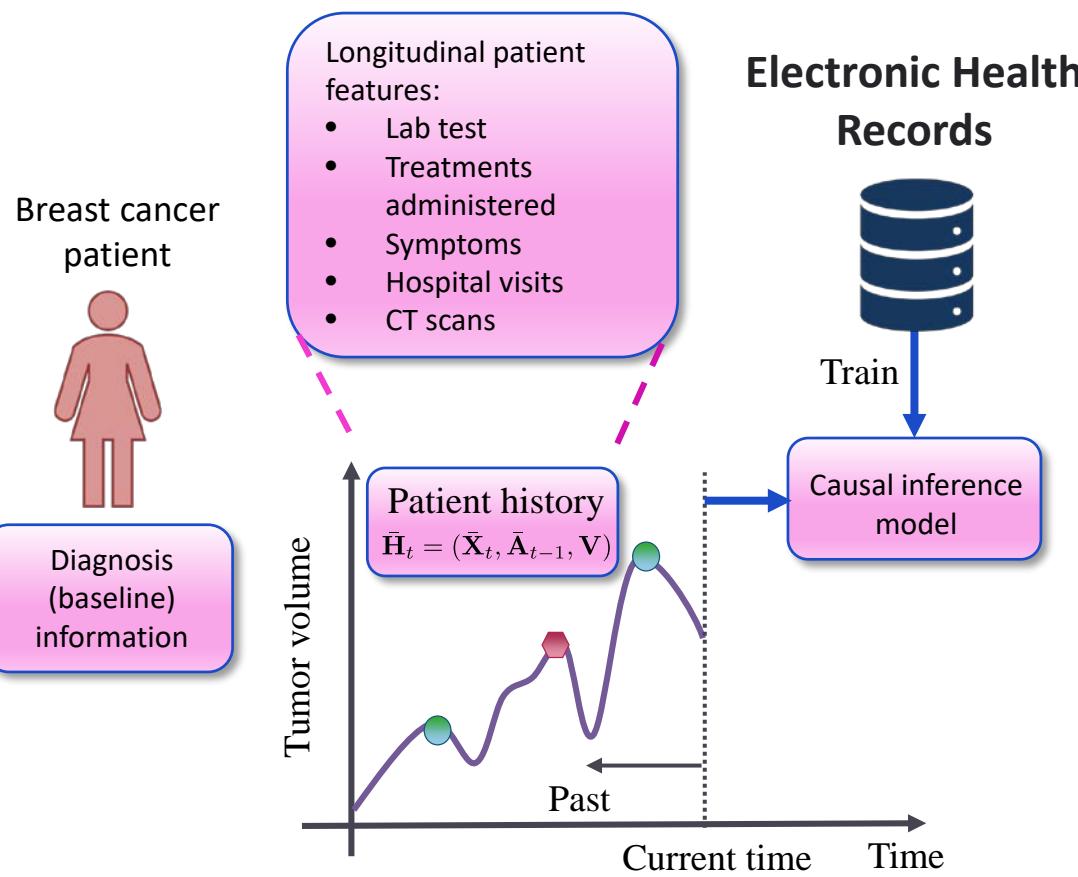
(b) Decide optimal time of treatment

When to stop treatment?



(c) Decide when to stop treatment

Causal effect inference based on longitudinal patient observational data



Longitudinal patient observational data

- Time-dependent patient features: $\bar{X}_t = (\mathbf{X}_1, \dots, \mathbf{X}_t)$
- Time-dependent treatments: $\bar{\mathbf{A}}_t = (\mathbf{A}_1, \dots, \mathbf{A}_t)$ where $\mathbf{A}_t \in \{A_1, \dots, A_K\}$

- Static patient features: V

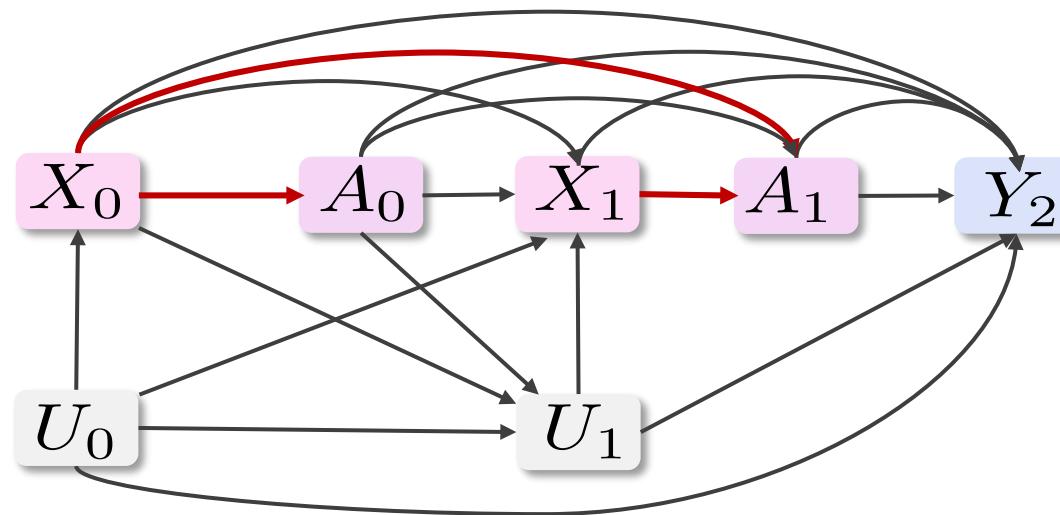
Patient history: $\bar{H}_t = (\bar{X}_t, \bar{A}_{t-1}, V)$

Observed (factual) outcome for treatment A_t given patient history $\bar{H}_t : Y_{t+1}$

Challenges in using longitudinal observational data for estimating individualized outcomes

The patient history $\bar{H}_t = (\bar{X}_t, \bar{A}_{t-1}, V)$ contains time-dependent confounders which bias the treatment assignment A_t in the observational dataset.

Patient covariates - affected by past treatments which then influence future treatments and outcomes

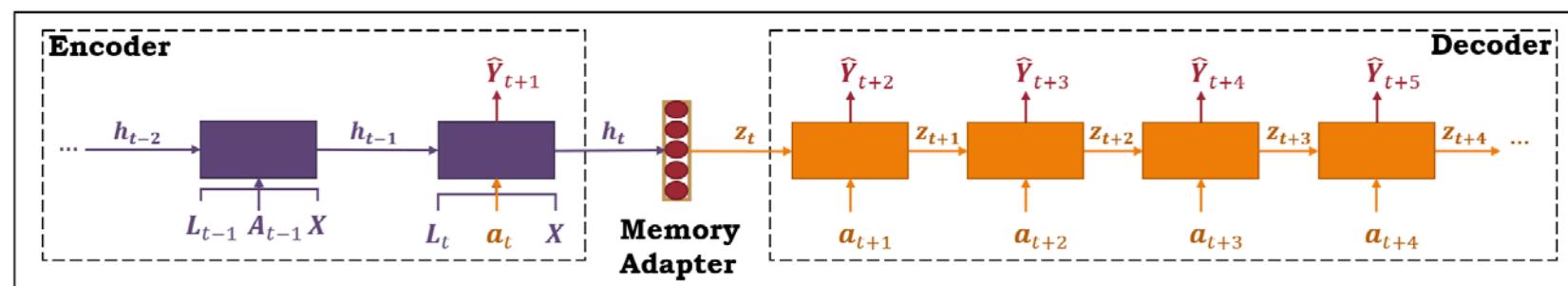
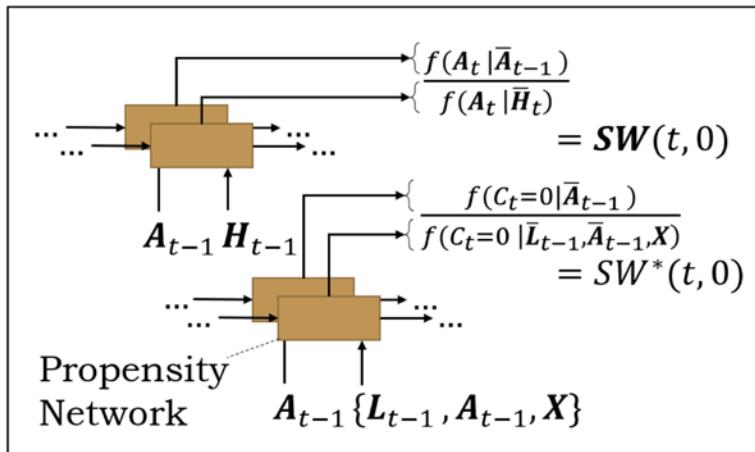


Bias from time-dependent confounders.

Handling time-dependent confounding bias

Inverse probability of treatment weighting

- Marginal Structural Models [Robins, Hernan, Brumback, Epidemiology 2000]
- Recurrent Marginal Structural Networks [Lim, Alaa, van der Schaar, NeurIPS 2018]



$$e(i, t, \tau) = \mathbf{SW}_i(t, \tau - 1) \times \tilde{\mathbf{SW}}_i^*(t, \tau - 1) \times \|\mathbf{Y}_{t+\tau, i} - g(\tau, a(t, \tau - 1), \bar{\mathbf{H}}_t)\|^2$$

$$\mathbf{SW}(t, \tau) = \prod_{n=t}^{t+\tau} \frac{f(\mathbf{A}_n | \bar{\mathbf{A}}_{n-1})}{f(\mathbf{A}_n | \bar{\mathbf{H}}_n)} = \prod_{n=t}^{t+\tau} \frac{\prod_{k=1}^{\Omega_a} f(A_n(k) | \bar{\mathbf{A}}_{n-1})}{\prod_{k=1}^{\Omega_a} f(A_n(k) | \bar{\mathbf{H}}_n)}$$

Handling time-dependent confounding bias

Inverse probability of treatment weighting

- Marginal structural models [Robins, Hernan, Brumback, Epidemiology 2000]
- Recurrent Marginal Structural Networks [Lim, Alaa, van der Schaar, NeurIPS 2018]

Numerically unstable

High variance

Representation Learning

- Counterfactual Recurrent Network [Bica, Alaa, Jordon, van der Schaar, ICLR 2020]

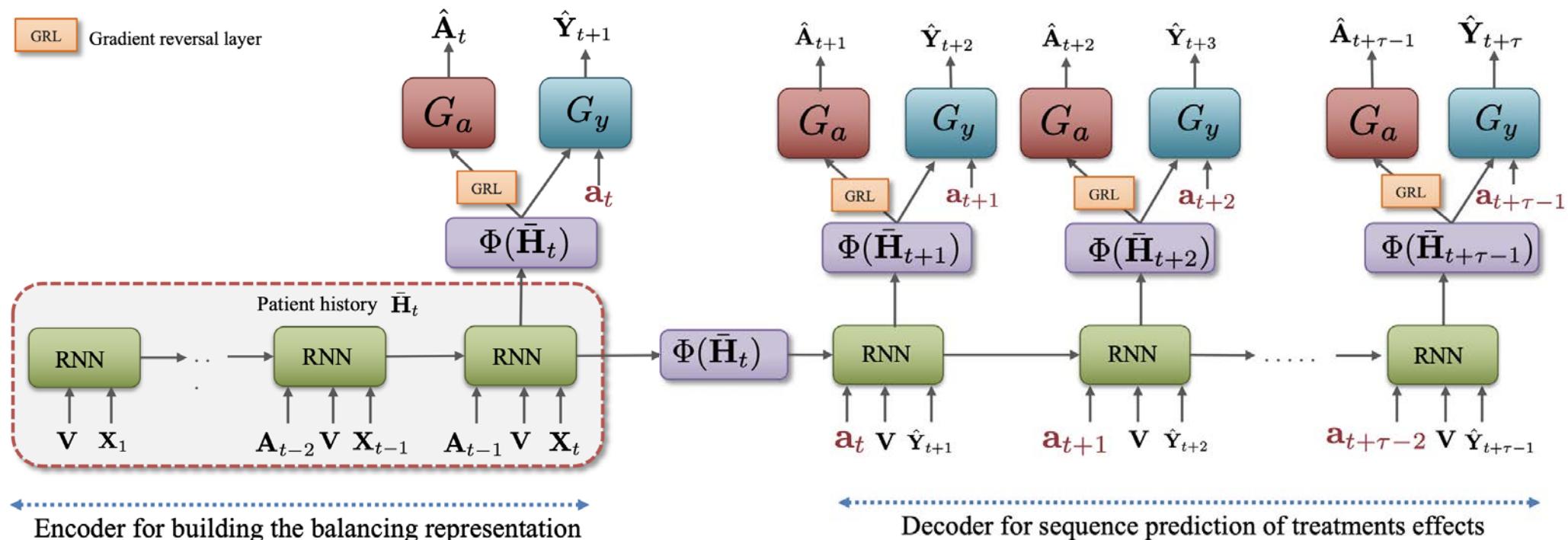
$$P(\Phi(\bar{\mathbf{H}}_t) \mid \mathbf{A}_t = A_1) = \dots = P(\Phi(\bar{\mathbf{H}}_t) \mid \mathbf{A}_t = A_K)$$

Balanced representations/
Treatment invariant representations

Counterfactual Recurrent Network

[Bica, Alaa, Jordon & van der Schaar, ICLR 2020]

- Builds treatment invariant representations using domain adversarial training [Ganin et al., 2016].
- Estimates counterfactual trajectories using sequence-to-sequence architecture.



Experiments using model of tumour growth

- Tumour volume $t + 1$ days after diagnosis

$$V(t+1) = \underbrace{\left(1 + \rho \log\left(\frac{K}{V(t)}\right)\right)}_{\text{Tumor growth}} - \underbrace{\beta_c C(t)}_{\text{Chemotherapy}} - \underbrace{\left(\alpha_r d(t) + \beta_r d(t)^2\right)}_{\text{Radiotherapy}} + \underbrace{e_t}_{\text{Noise}} V(t)$$

- Chemotherapy and radiotherapy treatment assignments

$$p_c(t) = \sigma \left(\frac{\gamma_c}{D_{\max}} (\bar{D}(t) - \delta_c) \right) \quad p_r(t) = \sigma \left(\frac{\gamma_r}{D_{\max}} (\bar{D}(t) - \delta_r) \right)$$

- Parameters γ_c and γ_r control the amount of time-dependent confounding.

Results: Treatment plans – how and when to treat?

Chemotherapy application

$$\begin{array}{l|l} \mathbf{Y}_{t+\tau} & \mathbf{a}_t = A_1, \mathbf{a}_{t+1} = A_0, \dots \mathbf{a}_{t+\tau-1} = A_0, \bar{\mathbf{H}}_t \\ \mathbf{Y}_{t+\tau} & \mathbf{a}_t = A_0, \mathbf{a}_{t+1} = A_1, \dots \mathbf{a}_{t+\tau-1} = A_0, \bar{\mathbf{H}}_t \\ \dots & \\ \mathbf{Y}_{t+\tau} & \mathbf{a}_t = A_0, \mathbf{a}_{t+1} = A_0, \dots \mathbf{a}_{t+\tau-1} = A_1, \bar{\mathbf{H}}_t \end{array}$$

Radiotherapy application

$$\begin{array}{l|l} \mathbf{Y}_{t+\tau} & \mathbf{a}_t = A_2, \mathbf{a}_{t+1} = A_0, \dots \mathbf{a}_{t+\tau-1} = A_0, \bar{\mathbf{H}}_t \\ \mathbf{Y}_{t+\tau} & \mathbf{a}_t = A_0, \mathbf{a}_{t+1} = A_2, \dots \mathbf{a}_{t+\tau-1} = A_0, \bar{\mathbf{H}}_t \\ \dots & \\ \mathbf{Y}_{t+\tau} & \mathbf{a}_t = A_0, \mathbf{a}_{t+1} = A_0, \dots \mathbf{a}_{t+\tau-1} = A_2, \bar{\mathbf{H}}_t \end{array}$$

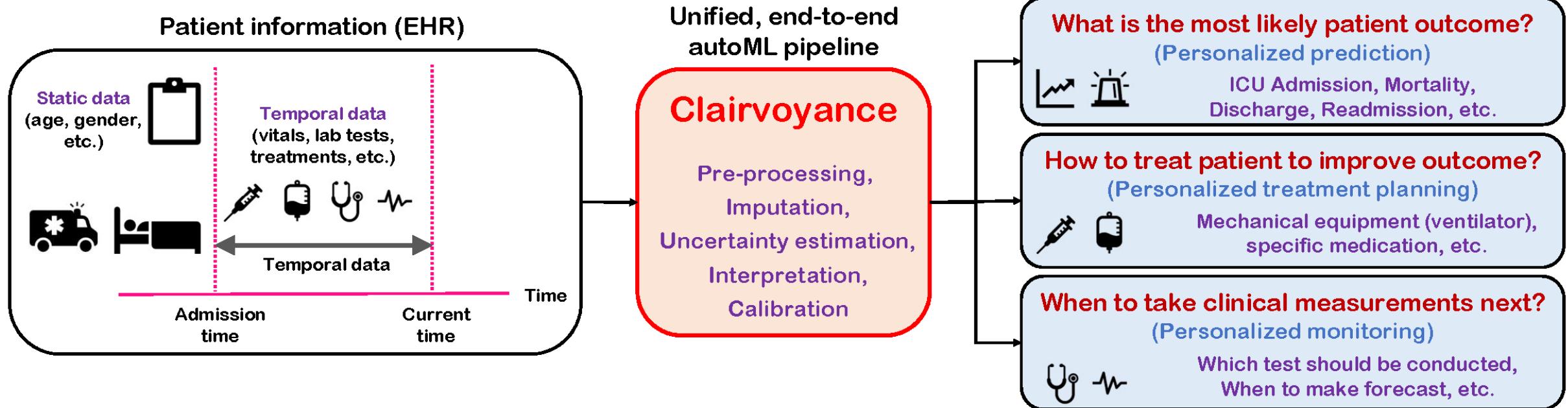
A_0 = no treatment

A_1 = chemotherapy

A_2 = radiotherapy

	$\gamma_c = 5, \gamma_r = 5$	$\gamma_c = 5, \gamma_r = 0$			$\gamma_c = 0, \gamma_r = 5$					
	τ	CRN	RMSN	MSM	CRN	RMSN	MSM	CRN	RMSN	MSM
Treatment Accuracy	3	83.1%	75.3%	73.9%	83.2%	78.6%	77.1%	92.9%	87.3%	74.9%
	4	82.5%	74.1%	68.5%	81.3%	77.7%	73.9%	85.7%	83.8%	74.1%
	5	73.5%	72.7%	63.2%	78.3%	77.2%	72.3%	83.8%	82.1%	72.8%
	6	69.4%	66.7%	62.7%	79.5%	76.3%	71.8%	78.6%	69.7%	64.5%
	7	71.2%	68.8%	62.4%	72.7%	71.8%	71.6%	71.9%	69.3%	61.2%
Treatment Timing Accuracy	3	79.6%	78.1%	67.6%	80.5%	76.8%	77.5%	79.8%	75.7%	60.6%
	4	73.9%	70.3%	63.1%	79.0%	77.2%	73.4%	75.4%	71.4%	58.2%
	5	69.8%	68.6%	62.4%	78.3%	73.3%	63.6%	66.9%	31.3%	29.5%
	6	66.9%	66.2%	62.6%	73.5%	72.1%	63.9%	65.8%	24.2%	15.5%
	7	64.5%	63.6%	62.2%	70.6%	57.4%	44.2%	63.9%	25.6%	12.5%

Clairvoyance: A unified, end-to-end pipeline for clinical-decision support [ICLR 2021]



Focus on REPRODUCIBILITY!

ITE Estimation can complement RCTs

1. Before an RCT – determine improved patient recruitment
2. During an RCT – adaptive recruitment
3. After an RCT – post-hoc analysis to identify subgroups
4. In clinical practice – personalized therapeutics based on observational data



van_der_Schaar
\ LAB

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Personalized therapeutics: Adaptive Clinical Trials

Next-generation Randomized Control Trials



[Atan, Zame, vdS, AISTATS 2019]

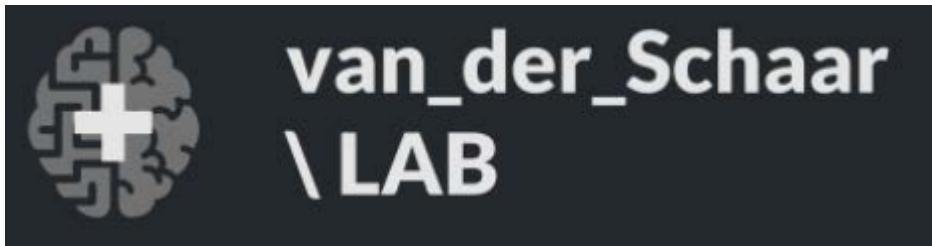
[Shen, Villar, vdS, ICML 2020]

[Lee, Zhang, Shen, Zame, vdS, NeurIPS 2020]

[Bica, Jordon, Alaa, vdS, NeurIPS 2020]

[Lee, Shen, Zame, vdS, AISTATS 2021]

Comprehensive Tutorial



**Tutorial series: individualized
treatment effect inference**

<https://www.vanderschaar-lab.com/video-tutorials-individualized-treatment-effect-inference>

<https://www.vanderschaar-lab.com/publications/causal-inference>

2. Empowering healthcare professionals

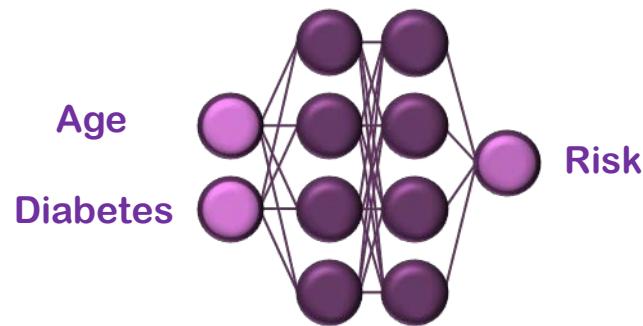
- Personalised ML assistants to support clinicians
 - Reflective practice
 - Augment the decision-making of clinicians
- ML must be interpretable, explainable, trustworthy

Interpretable, explainable, trustworthy analytics

Understand

why a prediction is made by the model

Interpretability



Interpretation 1

$$Risk \approx \beta_0 \text{Age} + \beta_1 \text{Diabetes}$$

Interpretation 2

Feature importance: β_0, β_1

what can we learn from the model

Explainability

All possible interpretations



User context



Interpretation 2

Feature importance: β_0, β_1

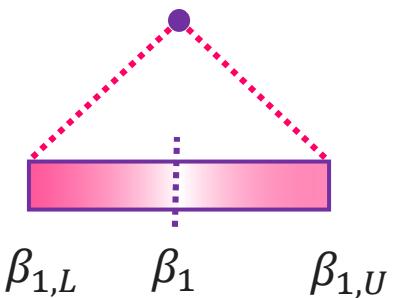
how trustworthy is the model's prediction

Trustworthiness

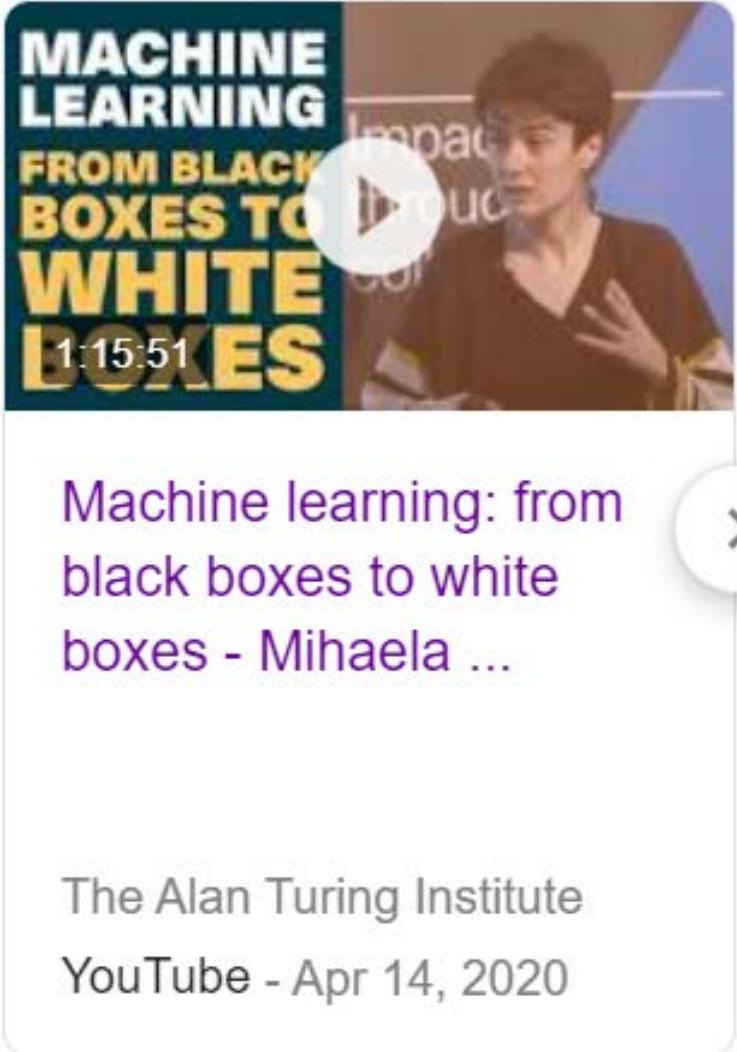
$$Risk \approx \beta_0 \text{Age} + \beta_1 \text{Diabetes}$$



Confidence interval



Machine learning: from black boxes to white boxes

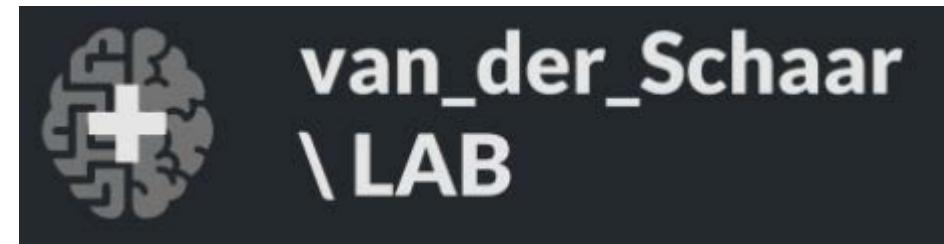


The thumbnail shows a woman speaking at a conference. The title text 'MACHINE LEARNING FROM BLACK BOXES TO WHITE BOXES' is overlaid on the left, with a play button icon in the center. Below the thumbnail, the video title and channel information are visible.

Machine learning: from
black boxes to white
boxes - Mihaela ...

The Alan Turing Institute
YouTube - Apr 14, 2020

<https://www.youtube.com/watch?v=EVI5iMpX1cg>



Making machine learning
interpretable: a dialog with
clinicians

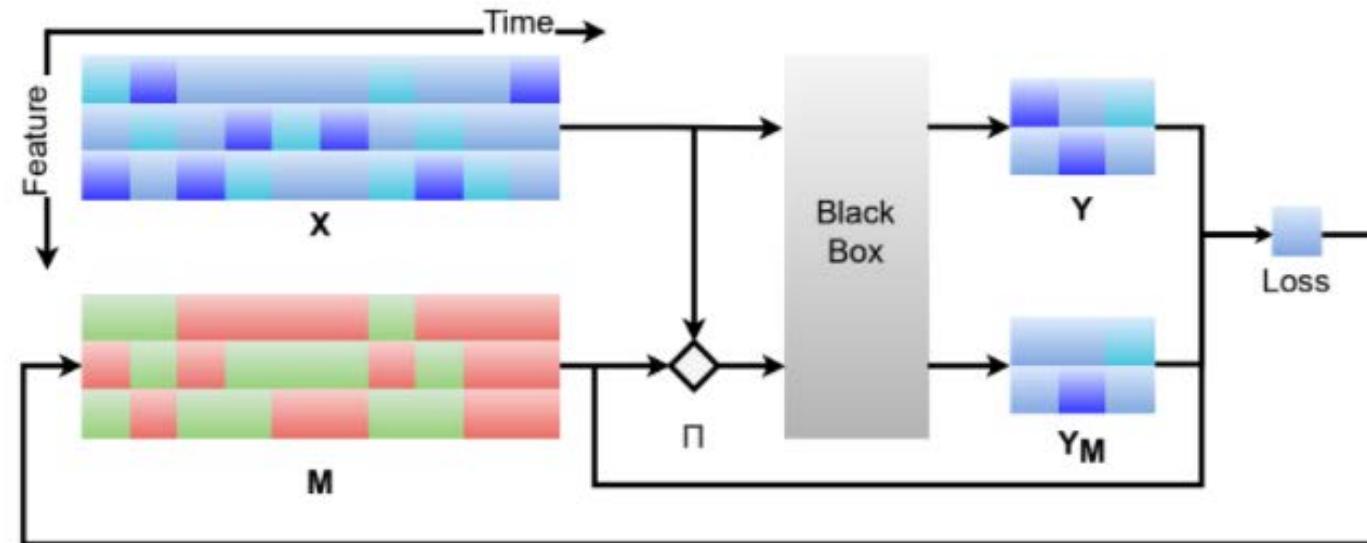
Several modes/types of ML interpretability

1. Where is the ML model looking at?
Explanatory patient features
2. For which similar/dissimilar patients has the ML model led to the same prediction?
Similarity classification
3. Which rules/laws has the ML model unraveled?
Hypothesis/law discovery
4. Which transparent risk equation has the ML model learned?
Transparent risk equations

ML interpretability: type 1

Explanatory patient features

Explaining Time Series Predictions With Dynamic Masks
[Crabbe, vdS, ICML 2021]



ML interpretability: type 4

Transparent risk equations:

provide transparent risk equations for the ML model's predictions

- *Unlike regressions, non-linear interactions*



van_der_Schaar
\ LAB

vanderschaar-lab.com

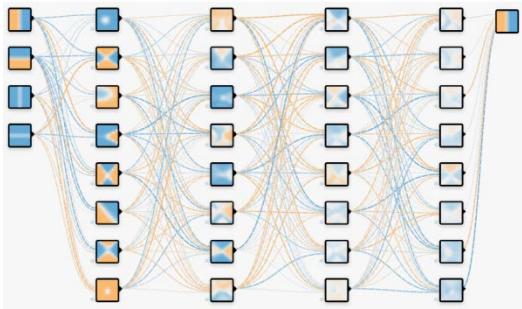


UNIVERSITY OF
CAMBRIDGE

Demystify black-box models using symbolic metamodels

[Alaa & vdS, NeurIPS 2019][Crabbe, Zhang, vdS, NeurIPS 2020]

Black-box ML model



$$f(\mathbf{x})$$

Symbolic Metamodelling

White-box (transparent) model

$$\begin{aligned} \alpha_1 X_1 + \alpha_2 X_2^2 + \alpha_3 X_1 X_2 \\ \alpha_4 X_3^3 + \alpha_5 \log(X_4) \end{aligned}$$

$$g(\mathbf{x})$$

- **Metamodel = a model of a model.**
- **A symbolic metamodel outputs a transparent risk function describing the predictions of the black box model**



van_der_Schaar
\\ LAB

vanderschaar-lab.com

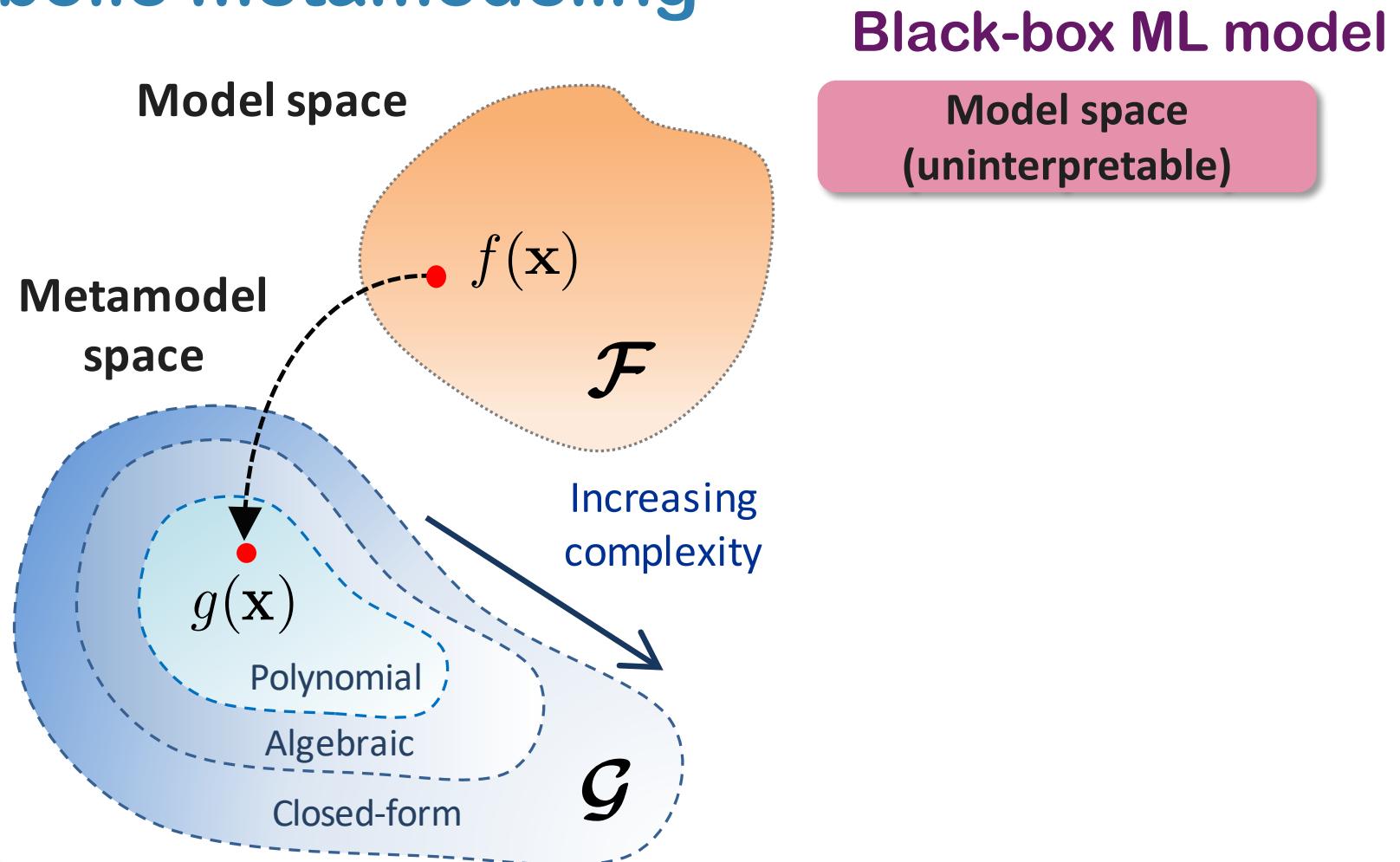


UNIVERSITY OF
CAMBRIDGE

Building transparent risk equations of black-box ML methods using symbolic metamodeling

White-box model
Metamodel
Transparent Risk Equations

$$\exp\left(\frac{\text{Age}}{5} - \log\left(\frac{\text{Tumor size}}{100}\right) + \frac{1}{10}\log(\text{Nodes})\right) \times \\ \exp\left(\frac{\text{ER} \cdot \text{Nodes}}{20} + \frac{\text{ER} \cdot \text{Tumor size}}{23}\right)$$



Black-box ML model

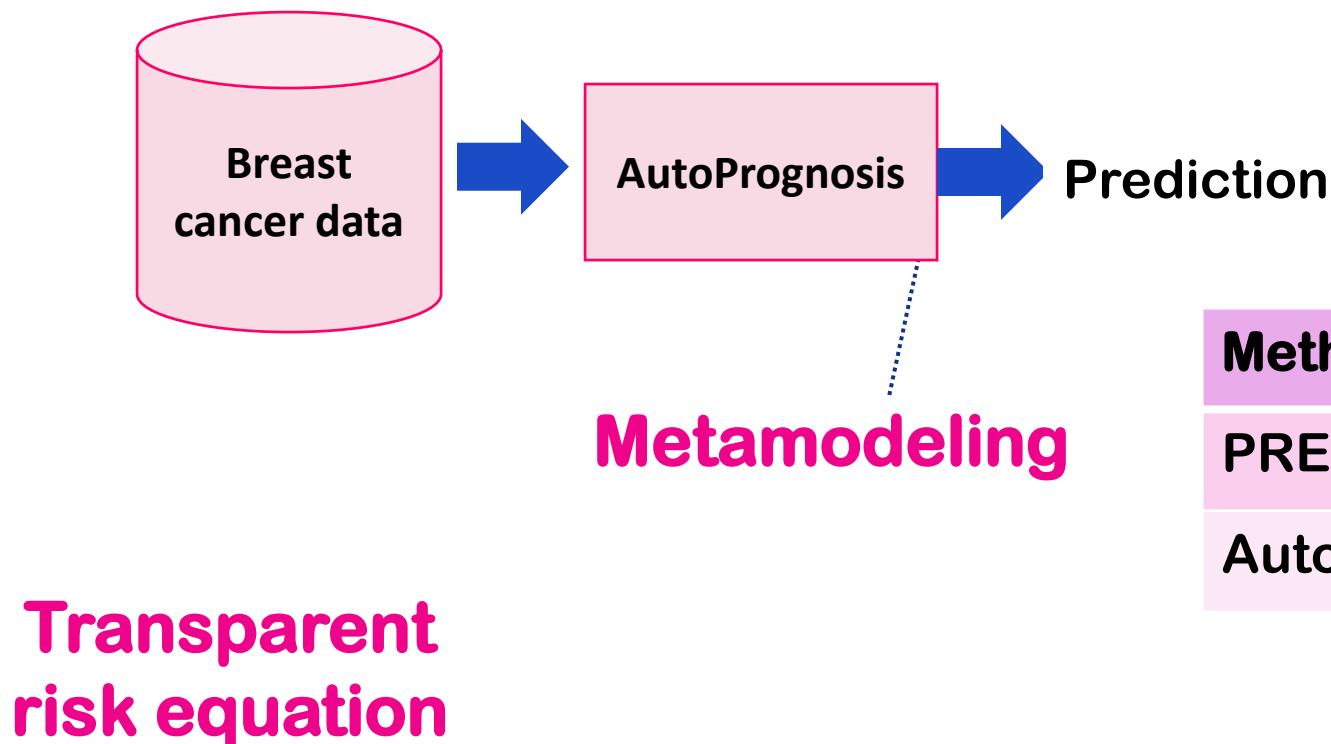
Model space
(uninterpretable)

Metamodel space can be *chosen* by the user!

Interpretability using symbolic metamodeling in practice

[Alaa, Gurdasani, Harris, Rashbass & vdS, Nature MI 2021]

Example: Predicting breast cancer risk survival (5 years)

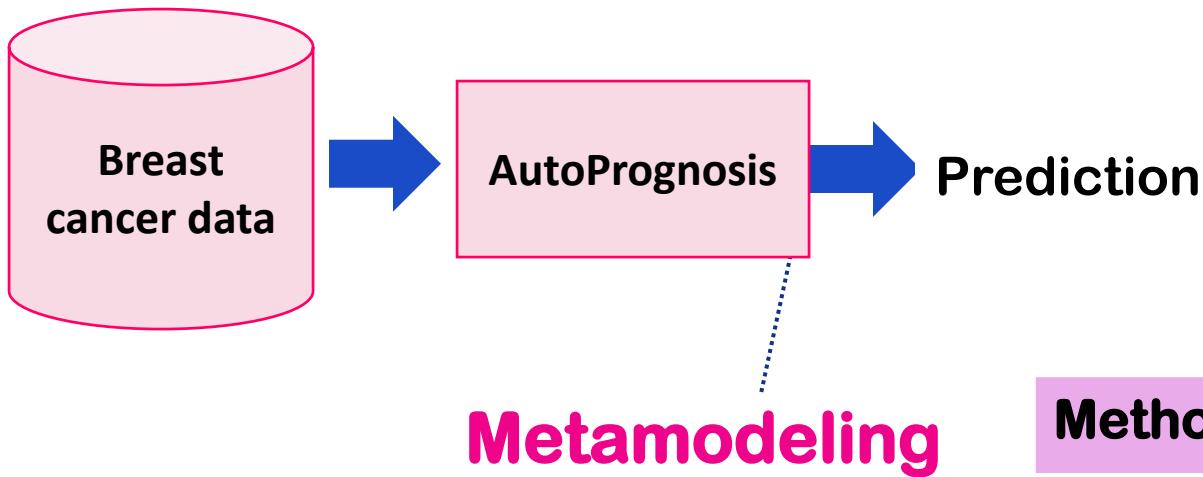


Method	AUC-ROC
PREDICT	0.75 ± 0.0033
AutoPrognosis	0.84 ± 0.0032

Interpretability using symbolic metamodeling in practice

[Alaa, Gurdasani, Harris, Rashbass & vdS, Nature MI 2021]

Example: Predicting breast cancer risk survival (5 years)



$f(\text{Age}, \text{ER}, \text{HER2}, \text{Tumor size}, \text{Grade}, \text{Nodes}, \text{Screening})$

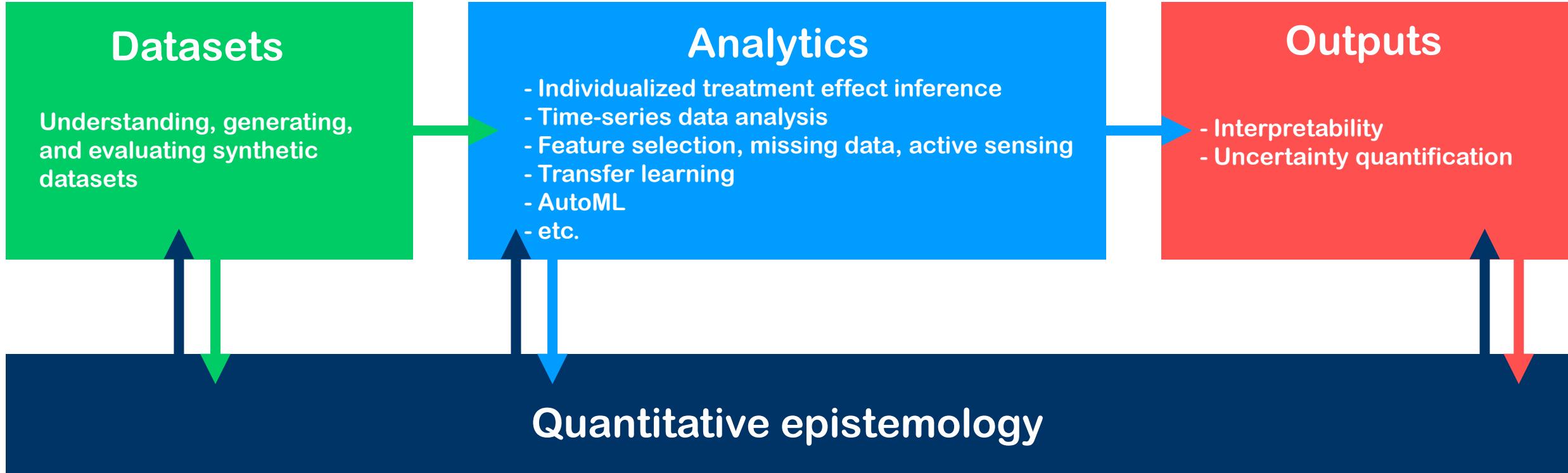
$$\exp\left(\frac{\text{Age}}{5} - \log\left(\frac{\text{Tumor size}}{100}\right) + \frac{1}{10}\log(\text{Nodes})\right) \times \\ \exp\left(\frac{\text{ER} \cdot \text{Nodes}}{20} + \frac{\text{ER} \cdot \text{Tumor size}}{23}\right)$$

Method	AUC-ROC
PREDICT	0.75 ± 0.0033
AutoPrognosis	0.84 ± 0.0032
Metamodel	0.83 ± 0.0020

2. Empowering healthcare professionals

- Personalised ML assistants to support clinicians
 - Reflective practice
 - Augment the decision-making of clinicians
- ML must be interpretable, explainable, trustworthy
- More is needed....

Our group's research agenda: New ML aimed at revolutionizing healthcare



A new field of research:
conceiving a human-machine partnership

Explaining the name...

Quantitative epistemology

Refers to things that can be measured

The study of knowledge



van_der_Schaar
\ LAB

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Inverse decision modeling (understanding humans)



- Understanding, explaining & auditing decisions
- Giving quantitative accounts of past behavior
- Identifying “suboptimal” behavior
- Analyzing variation in practice
- Improving policies

Conventional decision-making analysis (replacing humans/guiding humans)

- Optimal control
- Reinforcement learning
- Apprenticeship learning
- Imitating behavior

Quantitative Epistemology (partnering with & empowering humans)



- *Help humans acquire better information*
- *Direct humans towards the right information*
- *Help humans evaluate and integrate diverse sources of information and turn them into decisions*
- *Learn various knowledge representations that humans use*
- *Identify each individual's internal knowledge models and make the best use of that knowledge*
- *Representations to use when interacting with humans*
- *Aid human communication*
- *Help humans learn*
- ...



The Standard ML Agenda

A standard ML scenario: no human agency

Screening and presentation



Diagnosis



Prognosis



Treatment



Follow-up



Predictions and recommendations

"should be screened early"

"stage 1 breast cancer, 1 cm, 2 positive nodes"

"10-year survival likelihood of 90%"

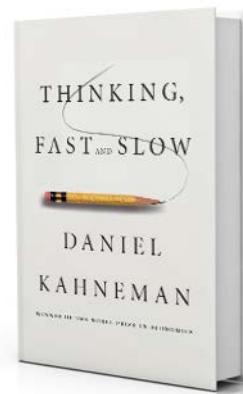
"low survival benefit from chemotherapy"

"mammograms every 6 months"



Decisions

Learn how humans make decisions;
incorporate this into the design of more human-like AI/ML;
REPLACE & OUTPERFORM HUMANS



van_der_Schaar
\ LAB

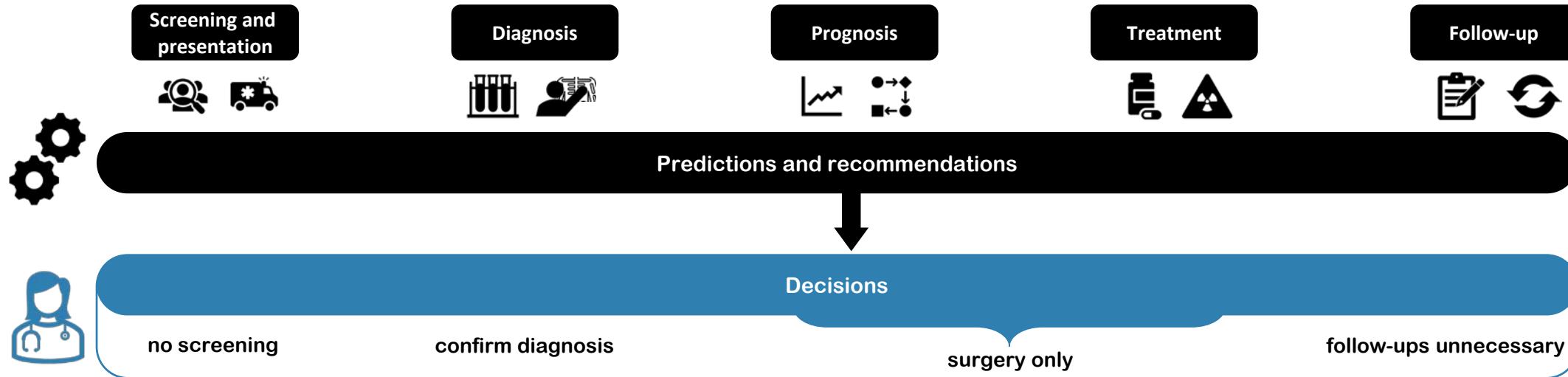
vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

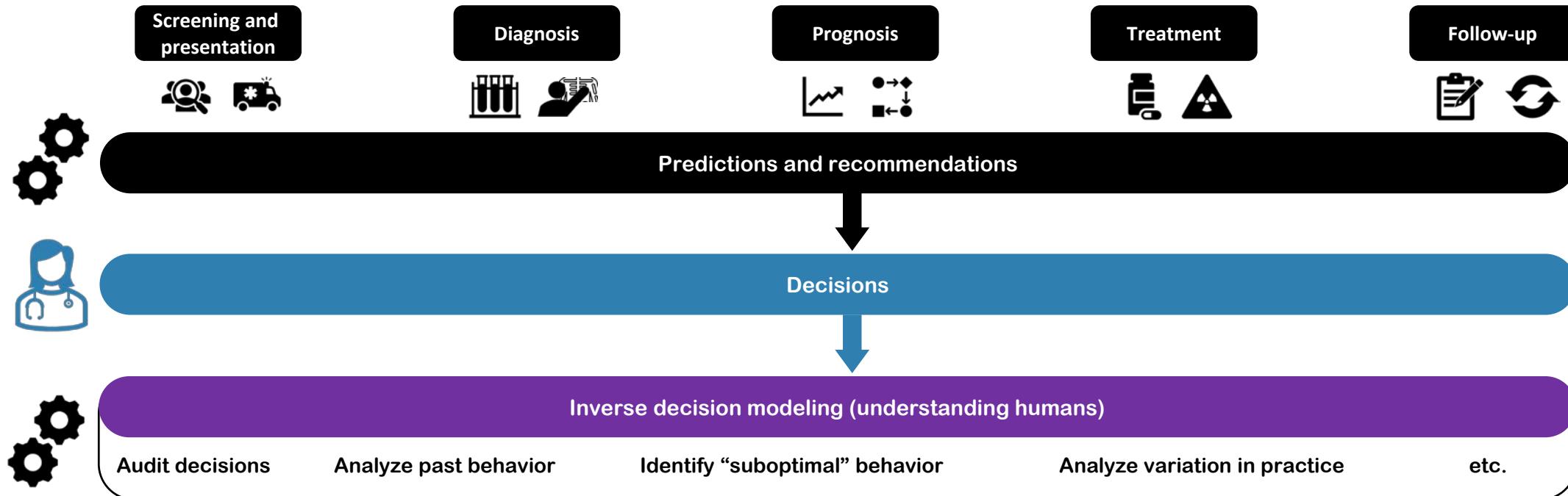
Standard Decision Support

AI/ML predictions and recommendations guiding human decision-making



Inverse decision modeling (vdS-Lab)

Surface-level analysis of/insight into human decision-making



van_der_Schaar
LAB

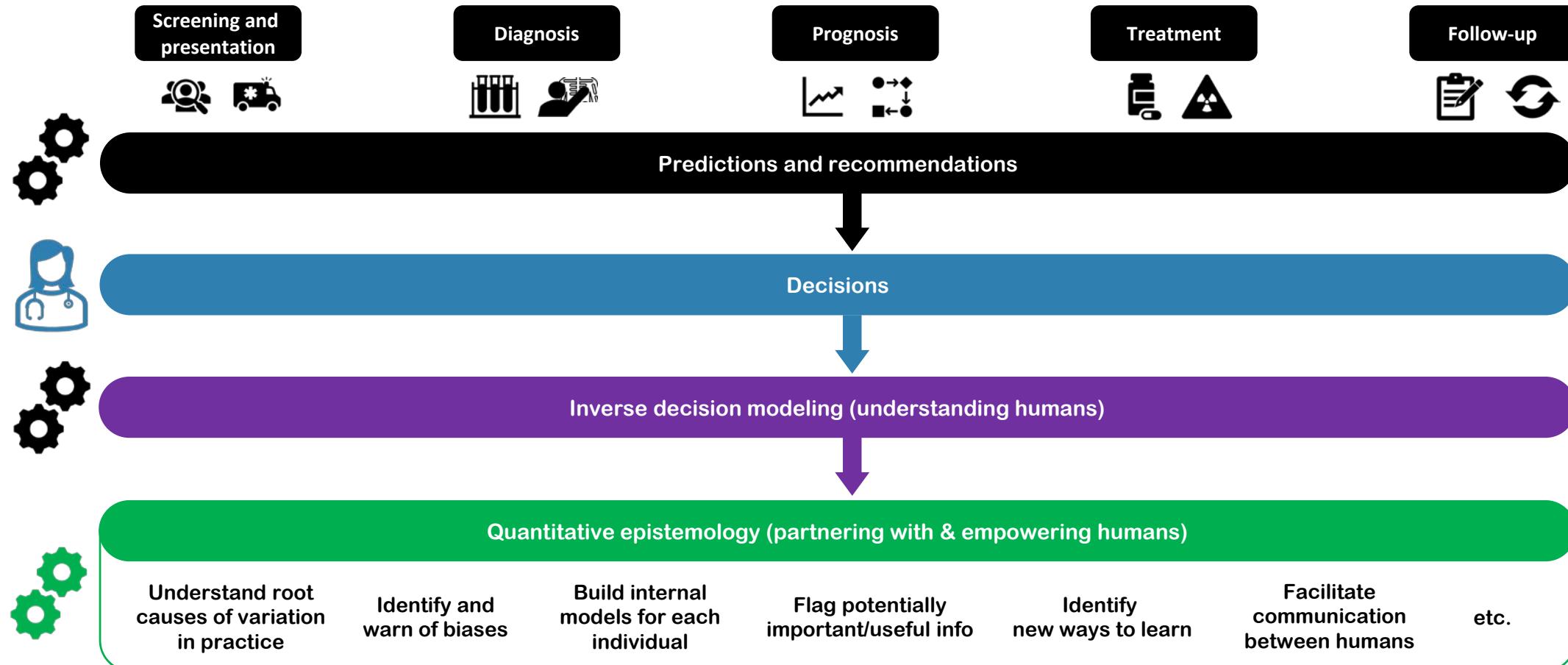
vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Quantitative Epistemology (vdS-Lab)

Extracting actionable meaning from analysis of decision-making...



van_der_Schaar
LAB

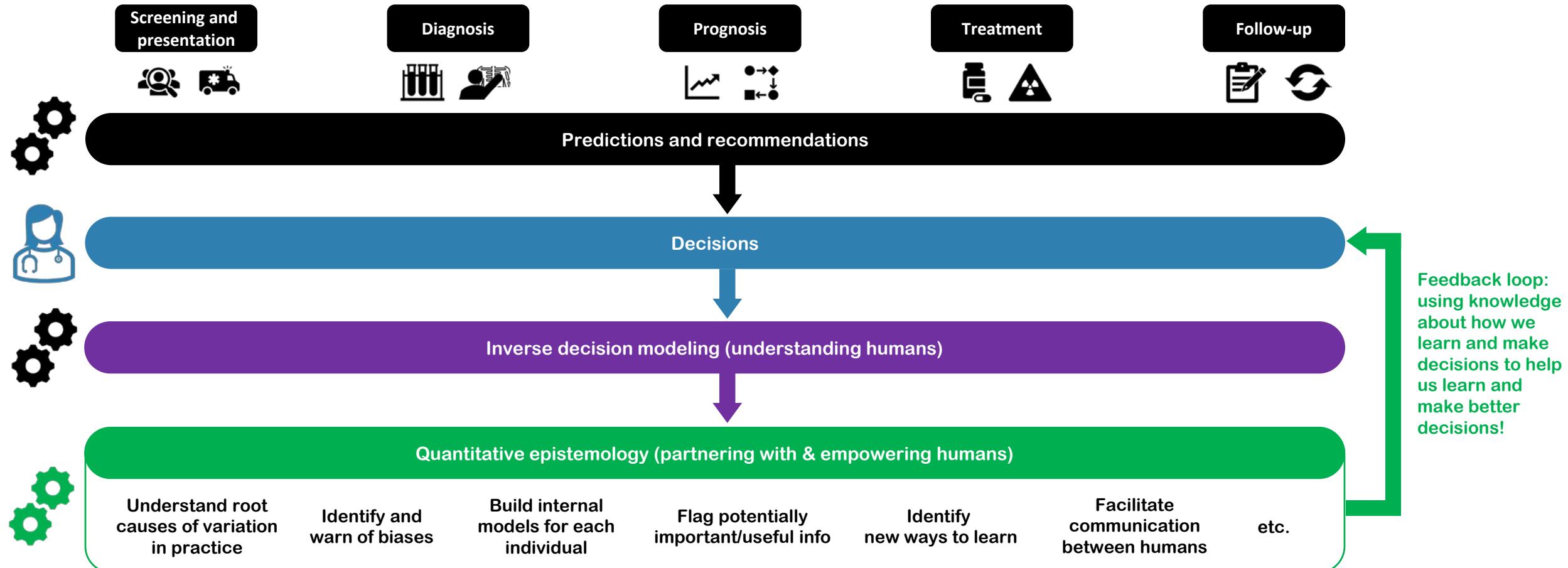
vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

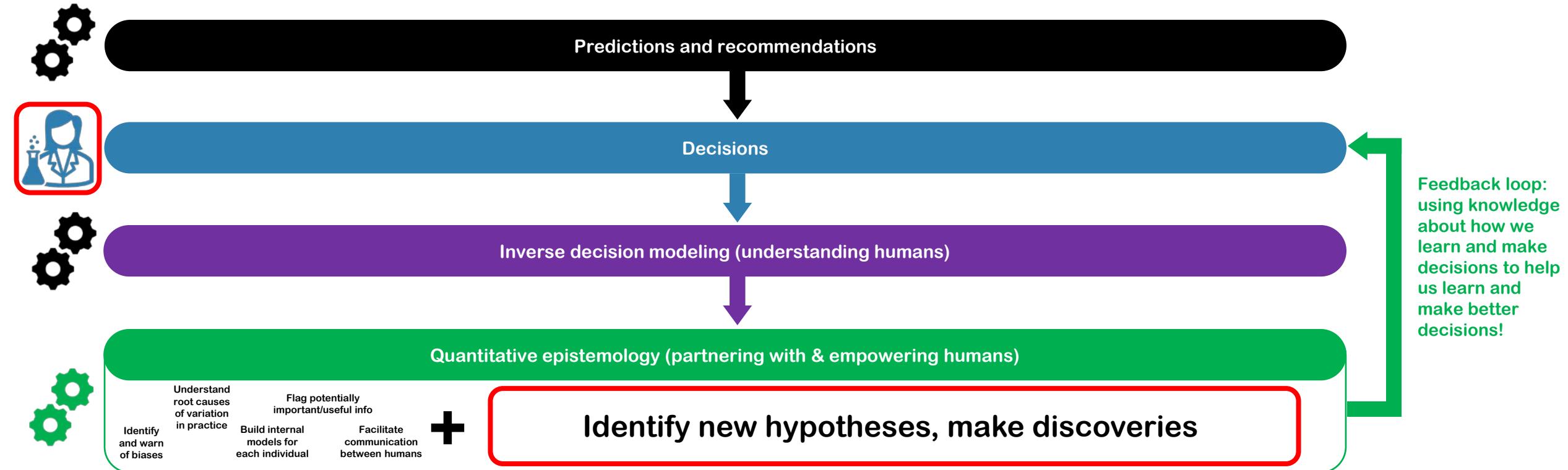
Quantitative Epistemology (vdS-Lab)

... creating an empowering loop that maximizes human agency and helps us make better decisions



Quantitative Epistemology (vdS-Lab)

For the researcher: new hypotheses and discoveries!



van_der_Schaar
LAB

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

A paradigm shift

	Decision support
Who?	Decision-makers
What purpose?	<ul style="list-style-type: none">▪ Detect one-time mistakes▪ Predict outcomes
When?	Before making a decision



van_der_Schaar
\\ LAB

vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

A paradigm shift

	Decision support	Quantitative Epistemology						
Who?	Decision-makers	Decision-makers	Stakeholders	Supervisors /managers	Policy-makers	Trainers /teachers	Researchers	
What purpose?	<ul style="list-style-type: none"> ▪ Detect one-time mistakes ▪ Predict outcomes 	<ul style="list-style-type: none"> ▪ Detect one-time mistakes & systematic errors ▪ Validate intended decisions ▪ Become more aware ▪ Avoid cognitive biases ▪ Recommend training 	<ul style="list-style-type: none"> ▪ Understand reasoning behind a decision ▪ Ensure the right decisions is made 	<ul style="list-style-type: none"> ▪ Detect systematic errors ▪ Analyze variation in practice ▪ Monitor behavior ▪ Recommend training 	<ul style="list-style-type: none"> ▪ Assess current practices ▪ Measure the impact of new policies 	<ul style="list-style-type: none"> ▪ Find best ways to present info and knowledge ▪ Assess how much is learned ▪ Personalized & continuous learning 	<ul style="list-style-type: none"> ▪ Analyze behavioral data ▪ Form hypotheses ▪ Contribute to decision-making theory 	
When?	Before making a decision	Before making a decision & Continuously	After a decision is made & Continuously	Continuously	Before and after introducing new policies	In a feedback loop	During scientific process, after data collection, before experimentation	



Quantitative Epistemology: Our work so far

IDM framework (ICML'21)					
Agent = human		Planner	Normative params.	Descriptive params.	
Method	Goal / motivating question				
IAS (ICML'20)	How “timely” is agent decision making?	Timely active sensing	Deadline, cost of acquisition	Importance of accuracy, speed, efficiency	
AVRIL (ICLR'21)	What reward function does the agent optimize?	RL planner	-	Reward function	
CIRL (ICLR'21)	How important are various counterfactuals in making decisions?	Counterfactual RL planner	Counterfactuals	Importance weights	
INTERPOLE (ICLR'21)	What are the subjective beliefs of the agent?	Policies based on decision boundaries	Interpretable state space	Decision dynamics & decision boundaries	
IBRC (ICML'21)	How optimal is agent behavior relative to an “ideal” reward function?	Bounded rational planner	“Ideal” reward function	Flexibility, optimism, adaptivity	
ICB (submitted to NeurIPS'21)	How does agent’s behavior evolve over time?	Contextual bandit strategies	-	Time-varying beliefs over reward functions	



Replacing & Outperforming humans

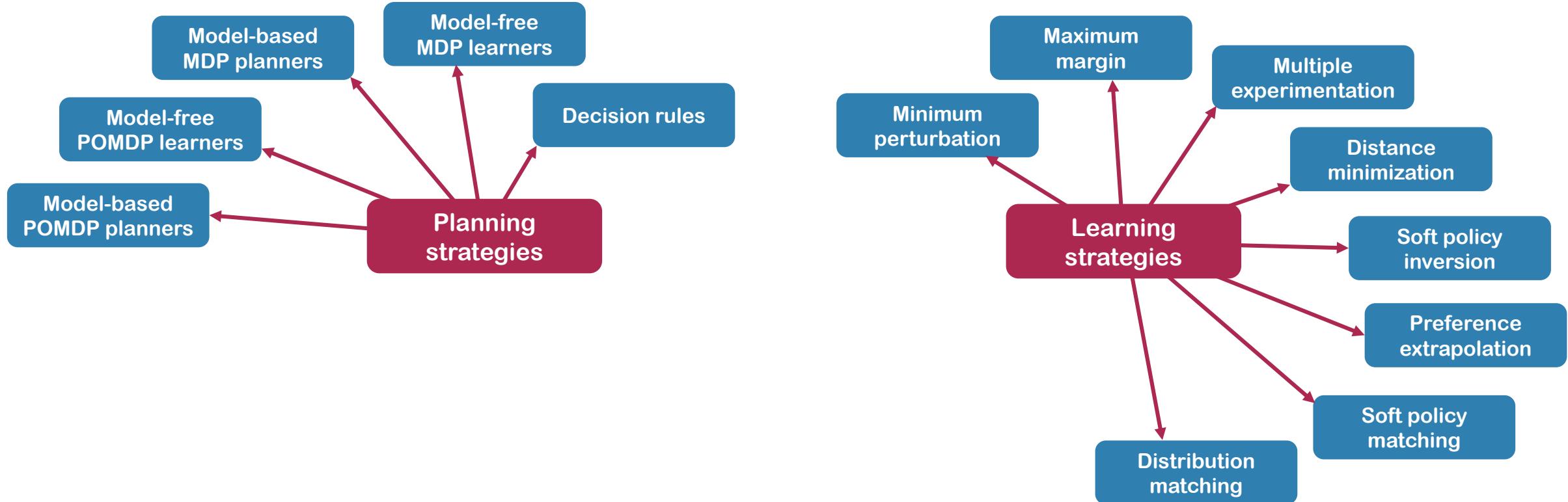
Previous works	Partially controllable	Partially observable	Purposeful behavior	Subjective dynamics	Action stochasticity
Behavioral cloning	✓	✓	✗	✗	✓
Subjective behavioral cloning	✓	✓	✗	✓	✓
Deterministic distribution matching	✓	✗	✗	✗	✗
Stochastic distribution matching	✓	✗	✗	✗	✓
Deterministic IRL	✓	✗	✓	✗	✗
Stochastic IRL	✓	✗	✓	✗	✓
Subjective IRL	✓	✗	✓	✓	✓
Risk sensitive IRL	✓	✗	✓	✓	✗
Deterministic partially-observable IRL	✓	✓	✓	✗	✗
Stochastic partially-observable IRL	✓	✓	✓	✗	✓
Subjective partially-observable IRL	✓	✓	✓	✓	✓
Maximum entropy IRL	✓	✗	✓	✗	✓
Subjective maximum entropy IRL	✓	✗	✓	✓	✓



Inverse decision model	Partially controllable	Partially observable	Purposeful behavior	Subjective dynamics	Action stochasticity	Knowledge uncertainty	Decision complexity	Specification complexity	Recognition complexity
Behavioral cloning	✓	✓	✗	✗	✓	✗	✗	✗	✗
Subjective behavioral cloning	✓	✓	✗	✓	✓	✗	✗	✗	✗
Deterministic distribution matching	✓	✗	✗	✗	✗	✗	✗	✗	✗
Stochastic distribution matching	✓	✗	✗	✗	✓	✗	✗	✗	✗
Deterministic IRL	✓	✗	✓	✗	✗	✗	✗	✗	✗
Stochastic IRL	✓	✗	✓	✗	✓	✗	✗	✗	✗
Subjective IRL	✓	✗	✓	✓	✓	✗	✗	✗	✗
Risk sensitive IRL	✓	✗	✓	✓	✗	✓	✗	✗	✗
Deterministic partially-observable IRL	✓	✓	✓	✗	✗	✗	✗	✗	✗
Stochastic partially-observable IRL	✓	✓	✓	✗	✓	✗	✗	✗	✗
Subjective partially-observable IRL	✓	✓	✓	✓	✓	✗	✗	✗	✗
Maximum entropy IRL	✓	✗	✓	✗	✓	✗	✓	✗	✗
Subjective maximum entropy IRL	✓	✗	✓	✓	✓	✗	✓	✗	✗
Inverse bounded rational control	✓	✓	✓	✓	✓	✓	✓	✓	✓



Quantitative Epistemology: New ML needed



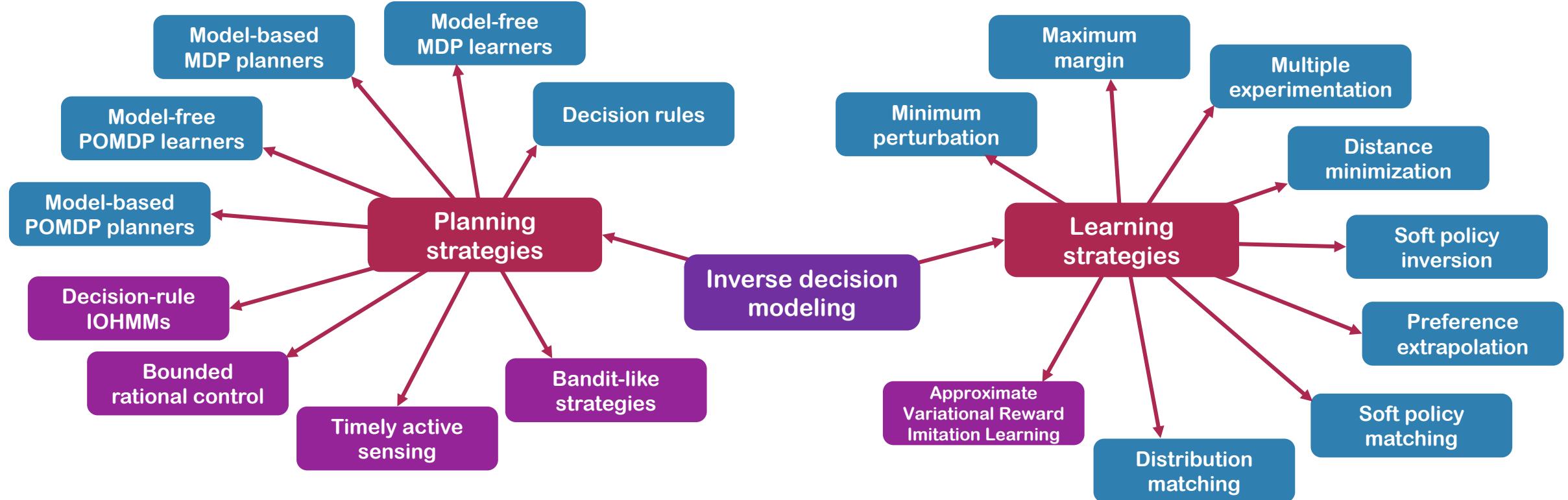
van_der_Schaar
\ LAB

vanderschaar-lab.com

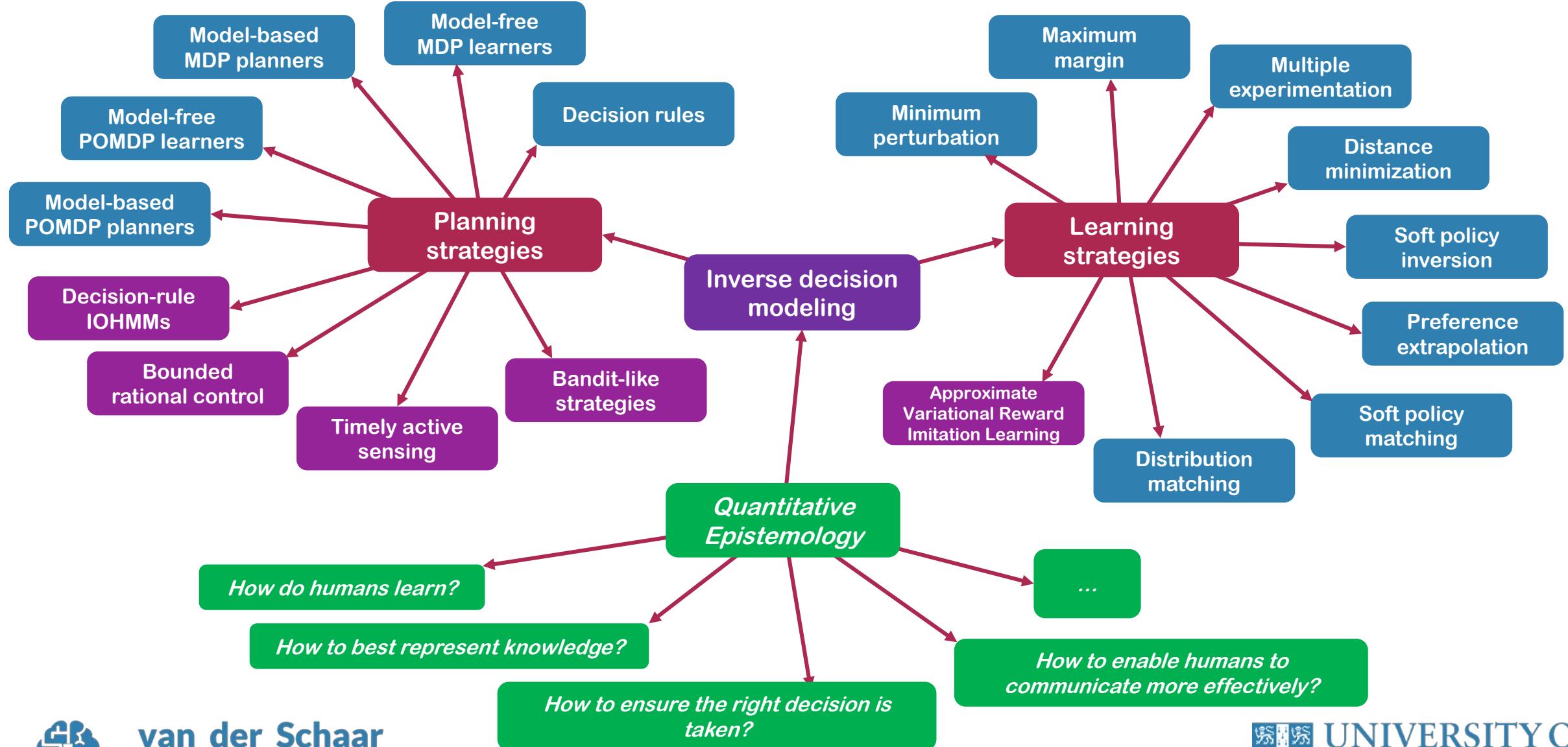


UNIVERSITY OF
CAMBRIDGE

Quantitative Epistemology: New ML needed



Quantitative Epistemology: New ML needed



Quantitative epistemology

- A new human-machine partnership
- A new field of multi-disciplinary research
- Partnering with humans to empower them, not to replace them!



van_der_Schaar
\ LAB

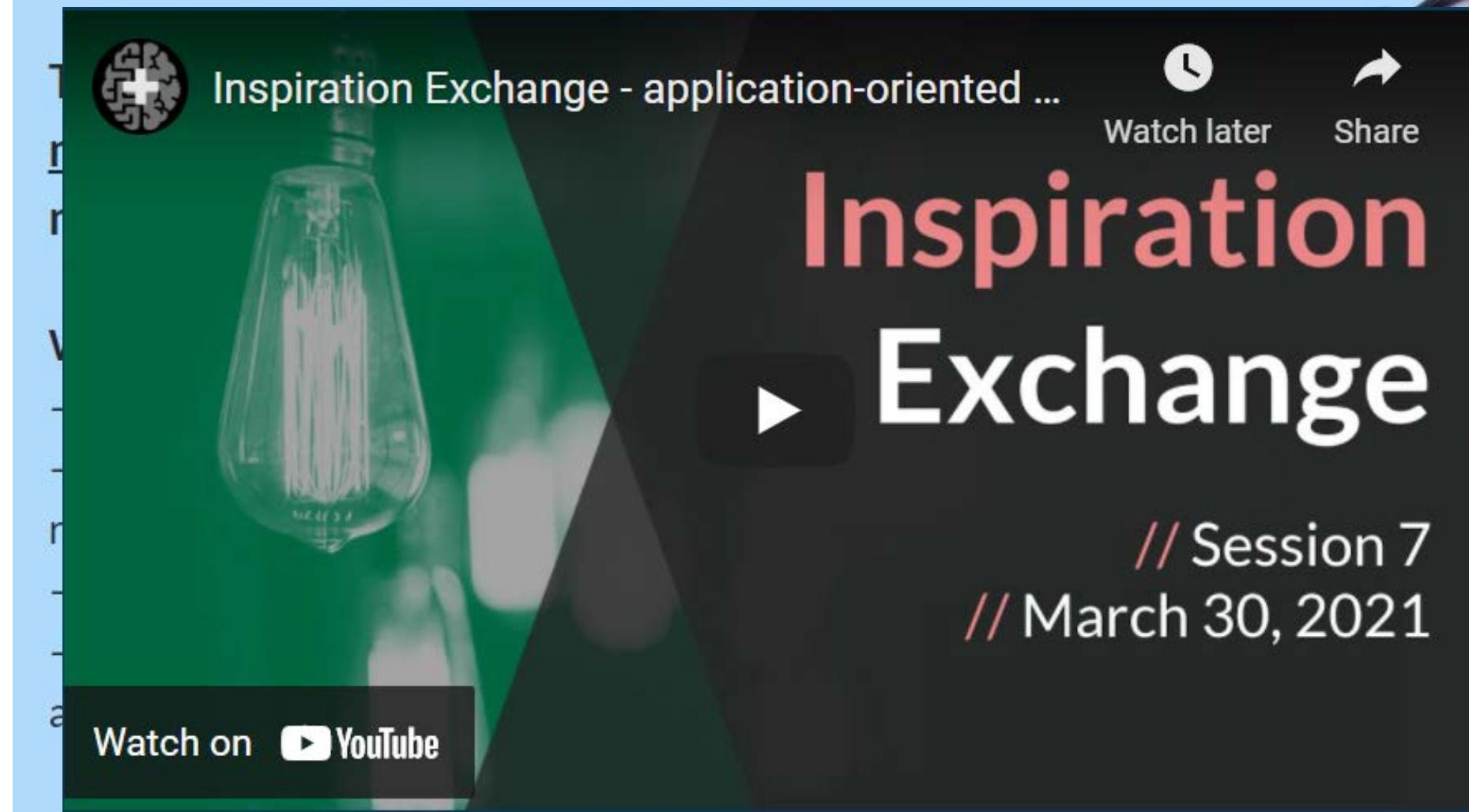
vanderschaar-lab.com



UNIVERSITY OF
CAMBRIDGE

Engagement
sessions

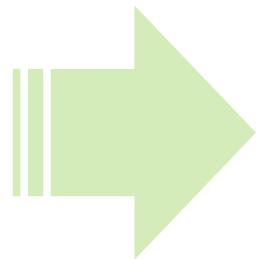
Inspiration Exchange



Join us!

Next session today: Quantitative Epistemology

	Patient-oriented	Profession-oriented
Individual	<p>Bespoke medicine</p> <ul style="list-style-type: none"> • Risk scores • Competing risks • Screening and monitoring • Diagnostic support • Longitudinal disease trajectories • Treatment effects 	<p>Empowering healthcare professionals</p> <ul style="list-style-type: none"> • Personalised ML assistants to support clinicians • Interpretable, explainable, trustworthy • Multi-disciplinary clinical contributions
At scale	<p>Population health and public health policy</p> <ul style="list-style-type: none"> • Discover & disentangle public risks and risk factors • Population risk assessment → personalized risk • Data-driven guidelines, protocols, standards • Cross-country learning and interventions 	<p>Systems, pathways and processes</p> <ul style="list-style-type: none"> • Improving healthcare pathways • Integrating and curating data sources • Integrating a multitude of analytics into delivery systems • Cooperation, interaction and learning



**Catalyze
a revolution
in healthcare**

Cambridge Centre for AI in Medicine's Inaugural Event

FREE WEBINAR 22 JANUARY 2021
2pm - 6pm GMT (9am - 1pm EST)

Speakers include...



Prof. Mihaela
van der Schaar
Director
Cambridge Centre for
AI in Medicine



Prof. Andres
Floto
Co-Director
Cambridge Centre for
AI in Medicine



Dr Sarah
Teichmann
Head of Cellular Genetics
Wellcome Sanger
Institute



Dr Jim
Weatherall
Vice President, Data
Science & AI, R&D
AstraZeneca



Dr Tony
Wood
Senior VP, Medicinal
Science & Technology
GSK

New Horizons in Machine Learning for Medicine

Join us at the Cambridge Centre for AI in Medicine as we reveal the transformations we are driving at the critical interface of machine learning, medicine and bioscience.

Our speakers are drawn from the frontiers of machine learning, science, clinical research, pharmaceutical R&D and the NHS.

PLUS: the brightest stars among the CCAIM faculty's PhD students offer bite-sized presentations on their latest high-impact research.

- **10+ high-profile speakers**
- **Two audience Q&A sessions**
- **PhD student showcases**
- **22 January, from 2pm**

To **REGISTER** your interest and receive
an agenda & Zoom link ahead of the event,
visit Eventbrite at bit.ly/CCAIMEVENT



UNIVERSITY OF
CAMBRIDGE



Cambridge Centre
for AI in Medicine