Improving Clinical trials with Machine Learning: Discovering Governing Equations in Medicine & Beyond

Mihaela van der Schaar

John Humphrey Plummer Professor of Machine Learning, Artificial Intelligence and Medicine, University of Cambridge //
Director, Cambridge Center for AI in Medicine // Turing Faculty Fellow, The Alan Turing Institute

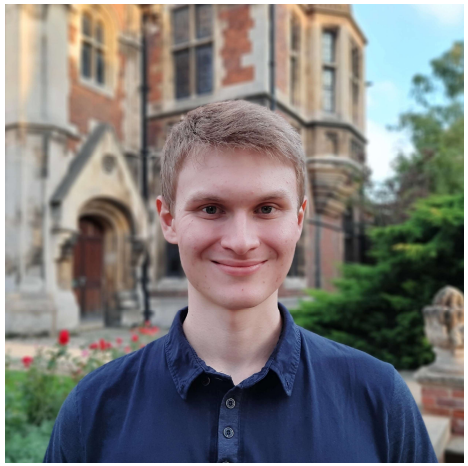
van_der_Schaar \ LAB
vanderschaar-lab.com

UNIVERSITY OF CAMBRIDGE

mv472@cam.ac.uk
@MihaelaVDS
linkedin.com/in/mihaela-van-der-schaar/

Acknowledgements

Zhaozhi Qian



Krzysztof Kacprzyk

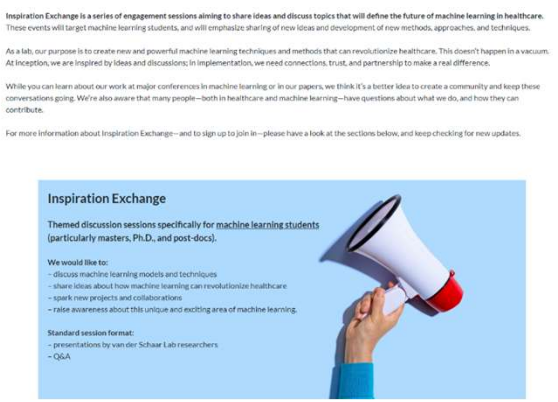
Alex Chan

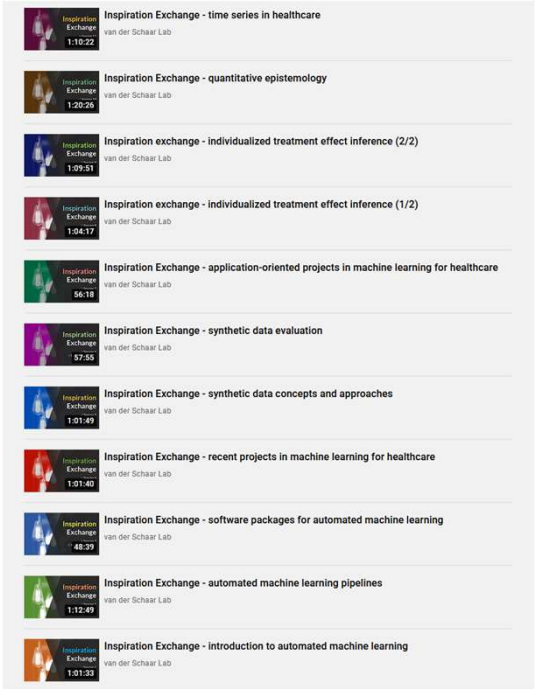
Alihan Huyuk

Engagement sessions: Inspiration Exchange

Online engagement sessions for ML researchers in healthcare; themed presentations & Q&A

https://www.vanderschaar-lab.com/
→ Engagement sessions
→ Inspiration Exchange





Subscribe & join us!

van_der_Schaar \ LAB

vanderschaar-lab.com

UNIVERSITY OF CAMBRIDGE

Discovery

Most of the ML community: Drug discovery

Our focus: Improving clinical trials with ML

Problem addressed:
Clinical trials - difficult, expensive and ripe for disruption

Situation today

Typical size of Phase II/III RCT: 100-1000s of patients
- of which up to half are controls
- and of which ~20% will drop out[1]

Typical all-in cost per patient: >$40,000[2]

Typical cost per trial: $4-40M+
- Most are in the $12-33M range[2]

Typical time per trial: 1-2 years
- of which as much as 30% is recruitment
- to which recruitment drives delays in >80% of trials[3]

Average PoS per Phase II/III trial: 50-60%[4]

Total Phase II/III trials commenced per year: ~>1,800[5]
- Of which around half are in oncology

Opportunity

Fundamentally, the way we run clinical trials has not changed in decades despite major tech advances

What if we could innovate, for example, to:
- Reduce required patient numbers by **20%** through replacing controls and preventing drop-outs
- Reduce cost-per-patient by **20%** through improved recruitment, monitoring and operations
- Improve PoS through better recruitment, prediction, trial design and analytics?

Cost saving per trial: ~$4-12M (for most)

Yielding an annual market opportunity: $7-22Bn+
- of which the innovator(s) could capture an appreciable fraction

Plus months' time saving per trial and upside on PoS

1. Tufts CSDD report 2020; higher dropout rate in CNS and oncology
2. Moore et al., JAMA Int Med, 2018 3. Huang et al., Cont Clin Trials, 2018 4. Takebe et al., Clin Transl Sci, 2018 5. Estimated based on Bloomberg data on identifiable trials sponsored by 15 large pharma companies in 2020

Challenges are felt throughout the clinical development journey

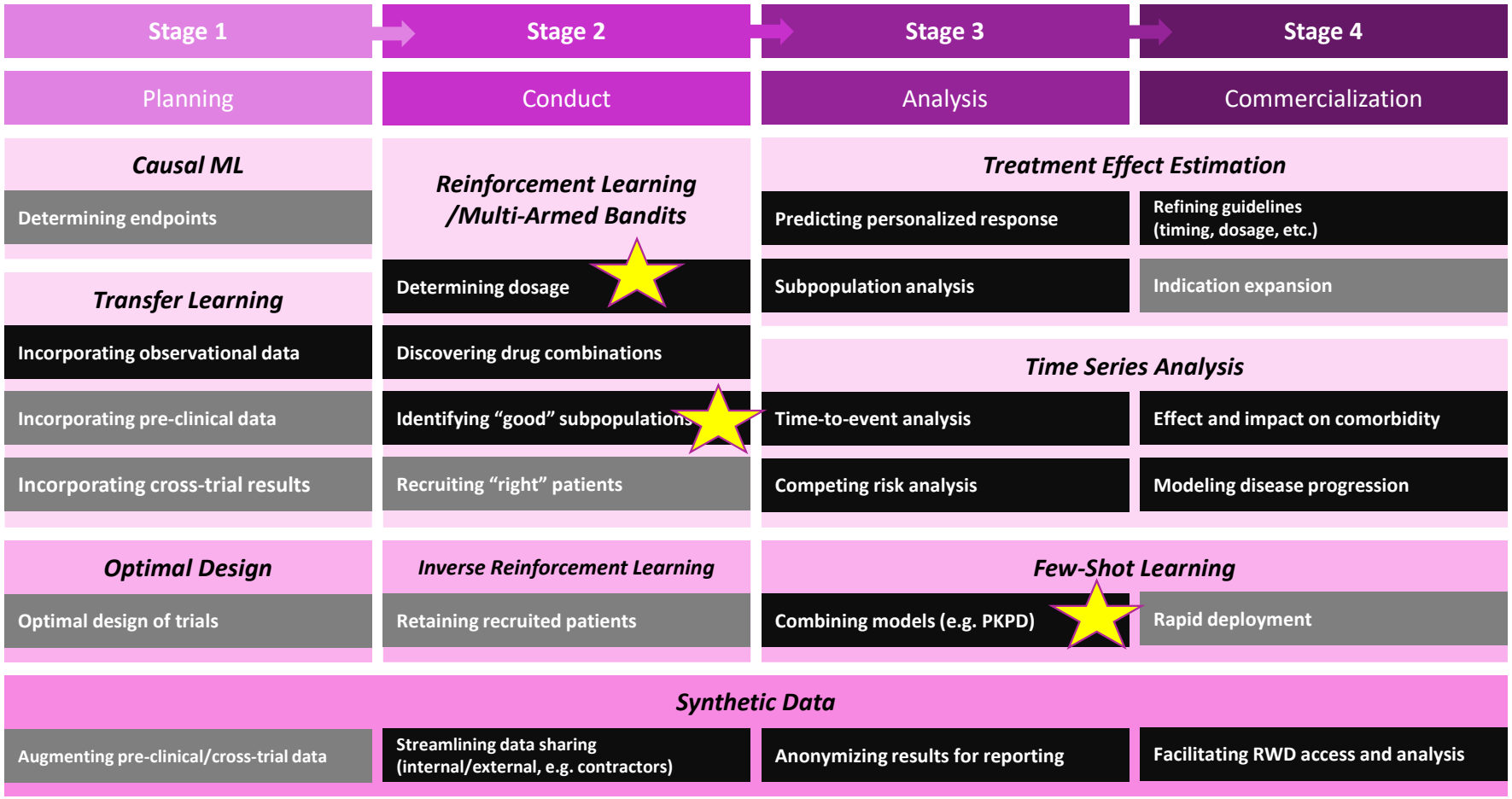
Stage 1	Stage 2	Stage 3	Stage 4
Planning	Conduct	Analysis	Commercialization
Determining endpoints	Determining dosage	Predicting personalized response	Refining guidelines (timing, dosage, etc.)
Incorporating observational data	Discovering drug combinations	Subpopulation analysis	Indication expansion
Incorporating pre-clinical data	Identifying “good” subpopulations	Time-to-event analysis	Effect and impact on comorbidity
Incorporating cross-trial results	Recruiting “right” patients	Competing risk analysis	Modeling disease progression
Optimal design of trials	Retaining recruited patients	Combining models (e.g. PKPD)	Rapid deployment
Augmenting pre-clinical/cross-trial data	Streamlining data sharing (internal/external, e.g. contractors)	Anonymizing results for reporting	Facilitating RWD access and analysis

Many are perceived as poorly addressed today

ML and related techniques can address many of them

Stage 1	Stage 2	Stage 3	Stage 4
Planning	Conduct	Analysis	Commercialization
Causal ML	***Reinforcement Learning /Multi-Armed Bandits***	***Treatment Effect Estimation***	
Determining endpoints		Predicting personalized response	Refining guidelines (timing, dosage, etc.)
Transfer Learning	Determining dosage	Subpopulation analysis	Indication expansion
Incorporating observational data	Discovering drug combinations	***Time Series Analysis***	
Incorporating pre-clinical data	Identifying "good" subpopulations	Time-to-event analysis	Effect and impact on comorbidity
Incorporating cross-trial results	Recruiting "right" patients	Competing risk analysis	Modeling disease progression
Optimal Design	***Inverse Reinforcement Learning***	***Few-Shot Learning***	
Optimal design of trials	Retaining recruited patients	Combining models (e.g. PKPD)	Rapid deployment
Synthetic Data			
Augmenting pre-clinical/cross-trial data	Streamlining data sharing (internal/external, e.g. contractors)	Anonymizing results for reporting	Facilitating RWD access and analysis

Some of our ML work to date

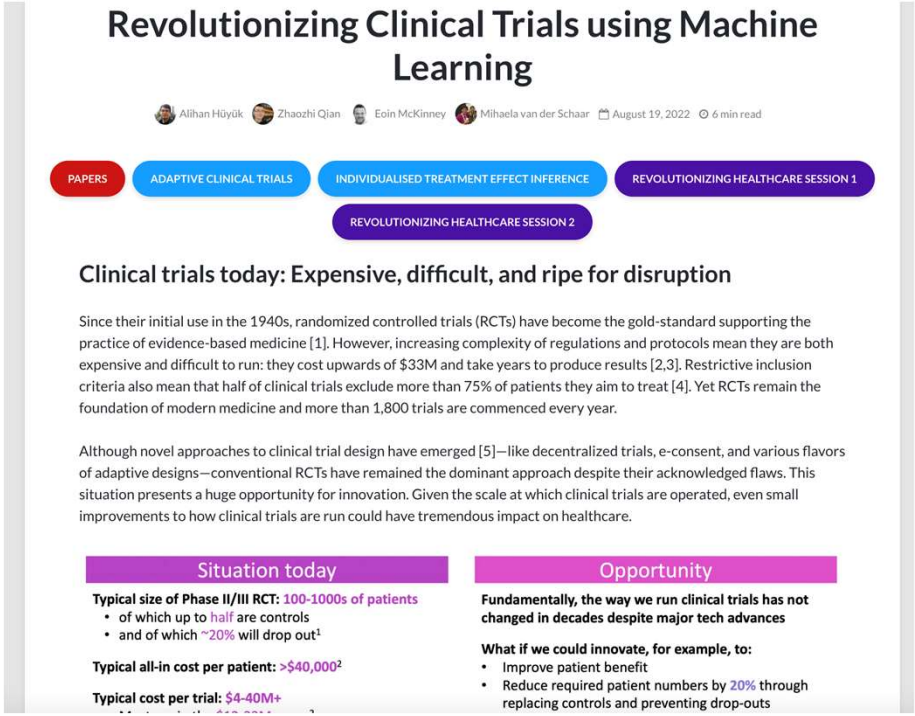


Revolutionizing clinical trials using machine learning

Comprehensive ‘Big Idea’ piece on clinical trials and the potential impact of ML/AI.

Explores key challenges of clinical trials and explaining the opportunities ML brings to the table.

vanderschaar-lab.com/
→ Big ideas
→ Revolutionizing Clinical Trials using Machine Learning



Revolutionizing Clinical Trials using Machine Learning

Alihan Hüyük · Zhaozhi Qian · Eoin McKinney · Mihaela van der Schaar · August 19, 2022 · 6 min read

PAPERS | ADAPTIVE CLINICAL TRIALS | INDIVIDUALISED TREATMENT EFFECT INFERENCE | REVOLUTIONIZING HEALTHCARE SESSION 1 | REVOLUTIONIZING HEALTHCARE SESSION 2

Clinical trials today: Expensive, difficult, and ripe for disruption

Since their initial use in the 1940s, randomized controlled trials (RCTs) have become the gold-standard supporting the practice of evidence-based medicine [1]. However, increasing complexity of regulations and protocols mean they are both expensive and difficult to run: they cost upwards of $33M and take years to produce results [2,3]. Restrictive inclusion criteria also mean that half of clinical trials exclude more than 75% of patients they aim to treat [4]. Yet RCTs remain the foundation of modern medicine and more than 1,800 trials are commenced every year.

Although novel approaches to clinical trial design have emerged [5]—like decentralized trials, e-consent, and various flavors of adaptive designs—conventional RCTs have remained the dominant approach despite their acknowledged flaws. This situation presents a huge opportunity for innovation. Given the scale at which clinical trials are operated, even small improvements to how clinical trials are run could have tremendous impact on healthcare.

Situation today

Typical size of Phase II/III RCT: 100-1000s of patients
- of which up to half are controls
- and of which ~20% will drop out[1]

Typical all-in cost per patient: >$40,000[2]

Typical cost per trial: $4-40M+

Opportunity

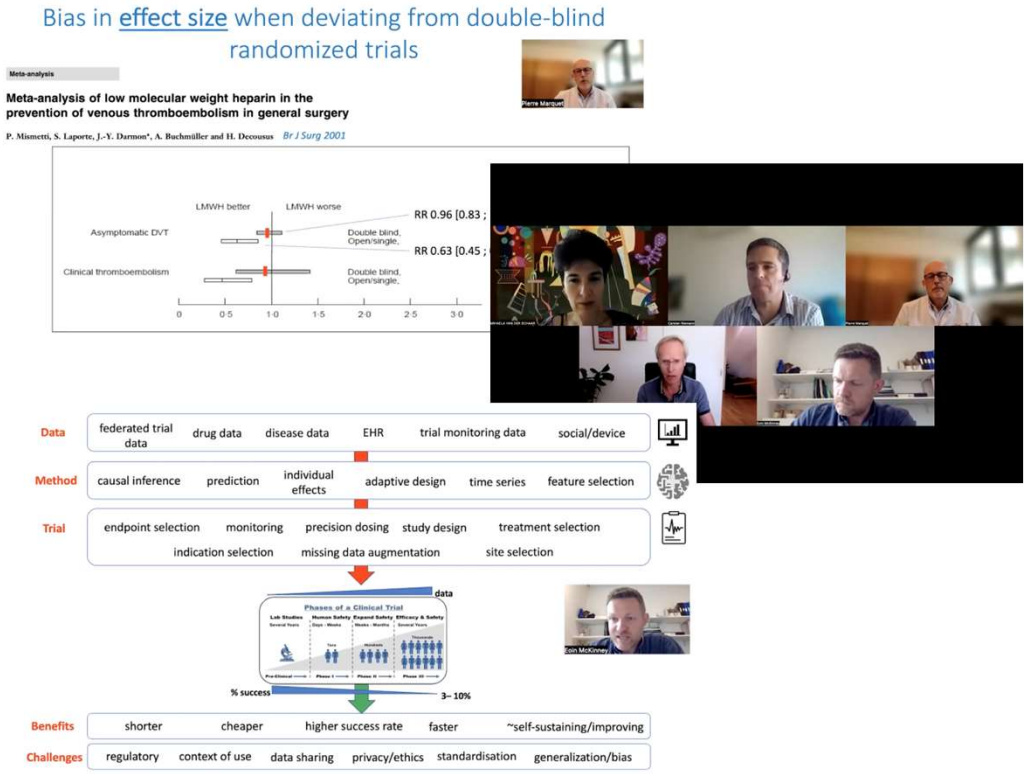
Fundamentally, the way we run clinical trials has not changed in decades despite major tech advances

What if we could innovate, for example, to:
- Improve patient benefit
- Reduce required patient numbers by 20% through replacing controls and preventing drop-outs

Revolutionizing Healthcare – Next-generation clinical trials

Clinician-focused Revolutionizing Healthcare roundtables (May & July 2022)

youtube.com/vanderSchaarLab/
→ Revolutionizing Healthcare

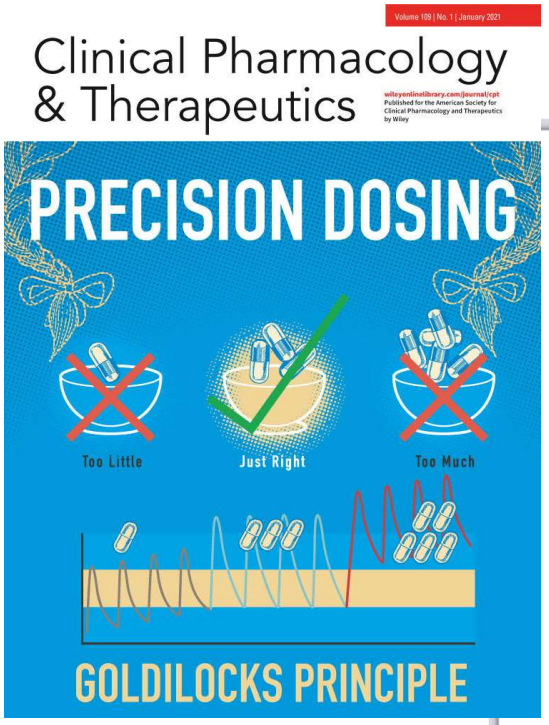


van_der_Schaar \ LAB

vanderschaar-lab.com

UNIVERSITY OF CAMBRIDGE

Precision Dosing using ML



TUTORIAL

From Real-World Patient Data to Individualized Treatment Effects Using Machine Learning: Current and Future Methods to Address Underlying Challenges

Ioana Bica[1,2,*], Ahmed M. Alaa[3], Craig Lambert[4] and Mihaela van der Schaar[2,3,5]

Clinical decision making needs to be supported by evidence that treatments are beneficial to individual patients. Although randomized control trials (RCTs) are the gold standard for testing and introducing new drugs, due to the focus on specific questions with respect to establishing efficacy and safety vs. standard treatment, they do not provide a full characterization of the heterogeneity in the final intended treatment population. Conversely, real-world

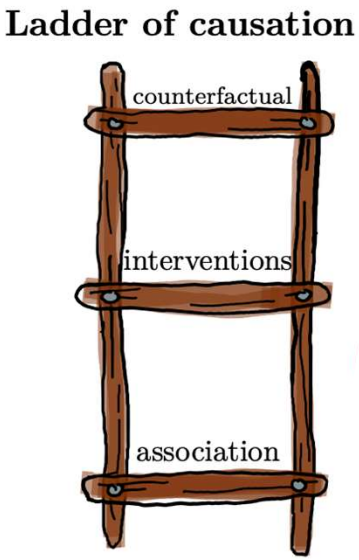
Discovery

Focus today: Discovery of governing equations using ML – the science of medicine

- Impact physiology (e.g. tumor growth)
- Impact pharmacology (e.g. precision dosing)
- Etc.

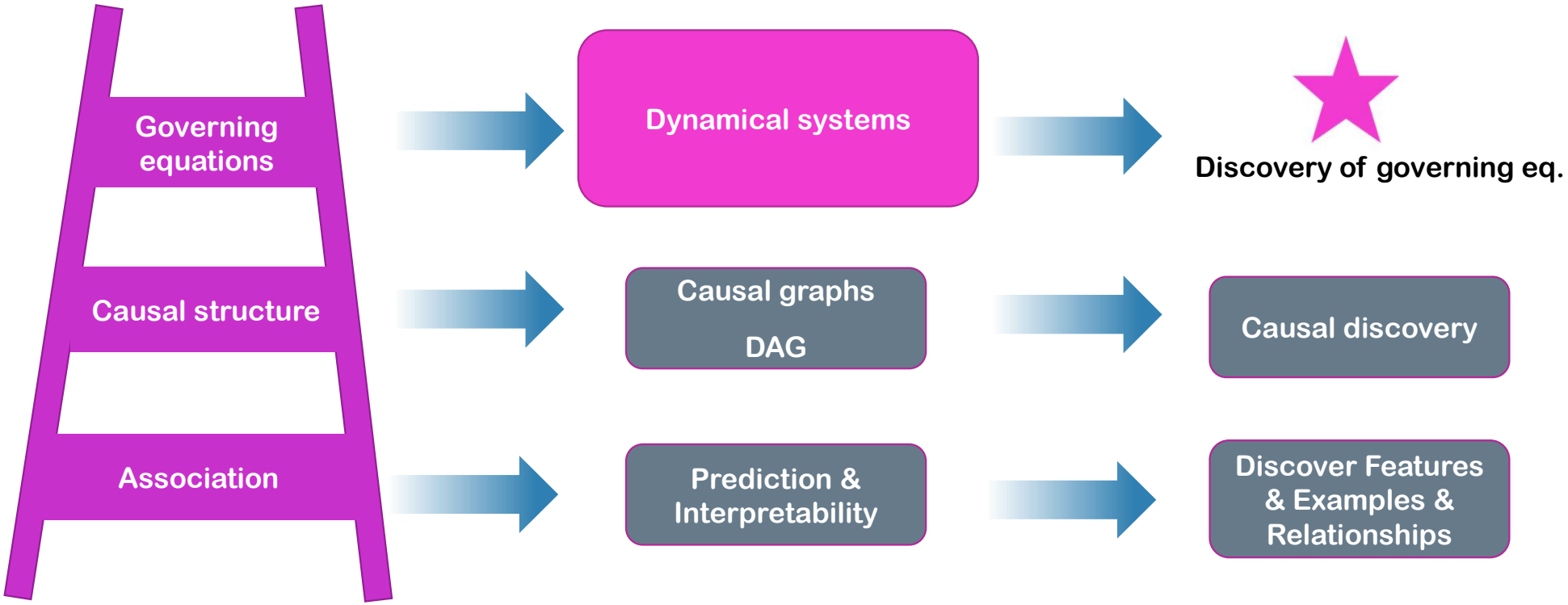
We need to go beyond causal structure

Causal models – structure + structural equations
Most research today: focus on structure



Pearl's ladder of causation

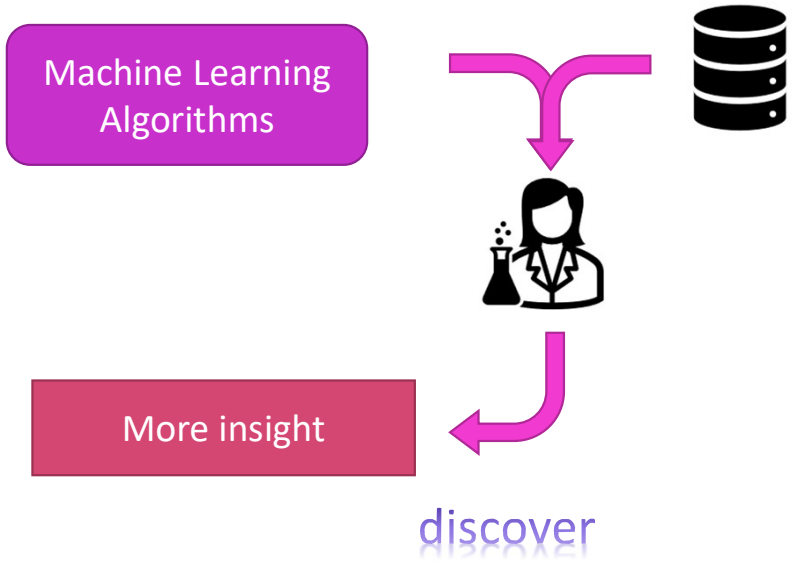
The “Discovery” Ladder



Discover closed-form governing equations from data

Closed-form equation are:

- Conscise
- Transparent
- Interpretable to human experts
- Amenable to further analysis (e.g., identifying stable equilibira)



Discovery of governing equations using ML

	Explicit function	Implicit function	Ordinary differential equation	Partial differential equation
Typical form	$y = f(x)$	$f(x, y) = c$	$\frac{dx}{dt} = f(x, t)$	$\frac{\partial u}{\partial t} = f(u, x)$

Discovery of risk equations using ML

	Explicit function
Typical form	$y = f(x)$

Risk equations

NHS Predict Breast Cancer equations

- if ER+

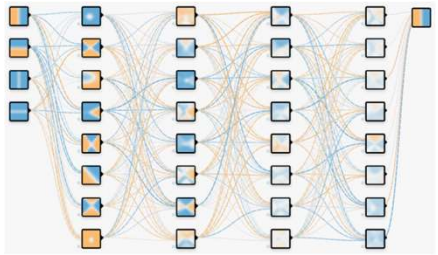
$$H_c(t) = \exp[0.7424402 - 7.527762/\sqrt{t} - 1.812513 * \log(t)/\sqrt{t}]$$

- if ER-

$$H_c(t) = \exp[-1.156036 + 0.4707332/t^2 - 3.51355/t].$$

Turning black boxes into white boxes using symbolic metamodels [Alaa & vdS, NeurIPS 2019] [Crabbe, Zhang, vdS, NeurIPS 2020]

Black-box ML model



$f(\mathbf{x})$

Symbolic Metamodeling

$$g(\mathbf{x}) = G(\mathbf{x}; \theta^*)$$

$$\theta^* = \arg\min_{\theta \in \Theta} \ell(f(\mathbf{x}), G(\mathbf{x}; \theta))$$

Explicit function

$$\alpha_1 X_1 + \alpha_2 X_2^2 + \alpha_3 X_1 X_2$$
$$\alpha_4 X_3^3 + \alpha_5 \log(X_4)$$

$g(\mathbf{x})$

Metamodels

Operates on a trained machine learning model and outputs a symbolic formula describing the model's prediction surface

Discovery of governing equations using ML

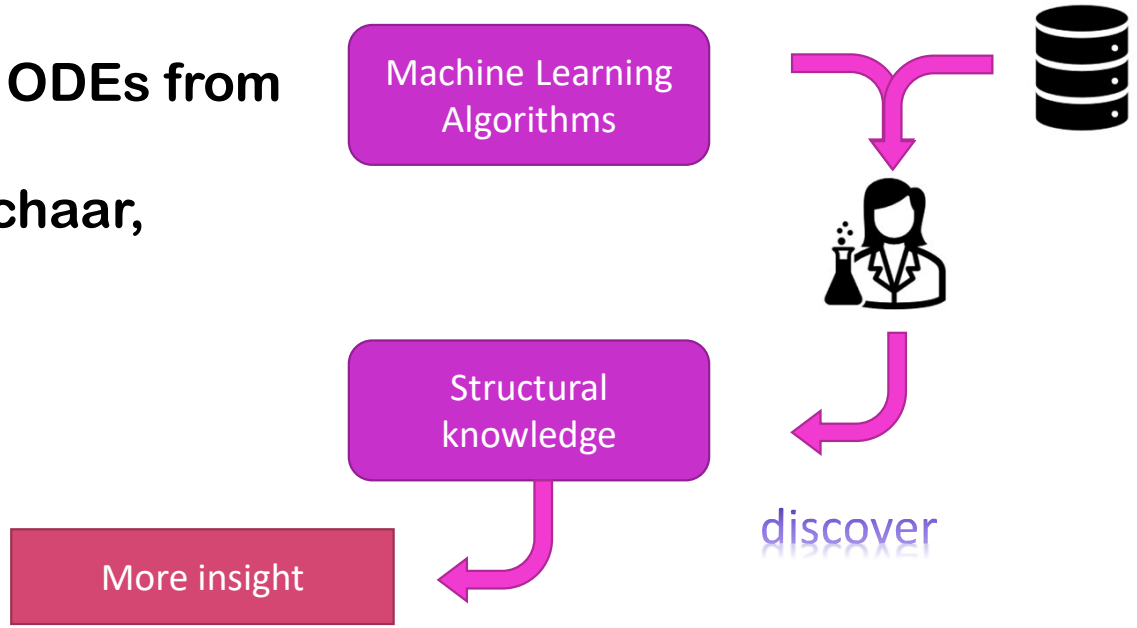
	Explicit function	Implicit function	Ordinary differential equation
Typical form	$y = f(x)$	$f(x, y) = c$	$\frac{dx}{dt} = f(x, t)$

A much harder problem

$x(t)$ ⟷ ODE ⟷ $x'(t)$

Discover closed-form ordinary differential equations (ODEs) from observed trajectories - *D-CODE*

D-CODE: Discovering Closed-form ODEs from Observed Trajectories
Z. Qian, K. Kacprzyk, M. van der Schaar,
ICLR 2022



Unique challenges in discovering ODEs

1. The time derivative is not observed
   - Only observe the states over time
   - Conventional *symbolic regression* methods are not applicable

2. It is difficult to estimate the time derivative
   - States are observed sporadically with noise
   - Naïve two-step symbolic regression is likely to fail

3. Difficulty in directly solving the initial value problem of ODE
   - The true initial condition is unknown & difficult to infer
   - Sensitive to initial condition
   - Computationally challenging

Problem formulation

Dataset

$$\begin{Bmatrix} \mathbf{y}_1(t) \\ \vdots \\ \mathbf{y}_D(t) \end{Bmatrix}$$

$t \in \{t_1, t_2, \dots, T\}$

$\mathbf{y}_i(t) \in \mathbb{R}^J$

Goal: Discover

System of J ODEs

$$\begin{Bmatrix} f_1(\boldsymbol{x}) = \dot{x}_1 \\ \vdots \\ f_J(\boldsymbol{x}) = \dot{x}_J \end{Bmatrix}$$

$x_j: [0, T] \to \mathbb{R}$

$\boldsymbol{x}(t) = [x_1, \dots, x_J]^T$

$f_j: \mathbb{R}^J \to \mathbb{R}$

D-CODE: Discovering Closed-Form ODEs [Qian, Kacprzyk, vdS, ICLR 2022]

Variational formulation of ordinary differential equations

$$\dot{x}_j(t) = f_j(\boldsymbol{x}(t)),\ \forall j = 1, \ldots, J,\ \forall t \in [0, T] \tag{1}$$

Characterize an ODE without referring to the derivative!

D-CODE: motivation

Variational formulation of ordinary differential equations

$$\dot{x}_j(t) = f_j(\boldsymbol{x}(t)),\ \forall j = 1, \ldots, J,\ \forall t \in [0, T] \tag{1}$$

Definition 1. Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, continuous functions $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, $f : \mathbb{R}^J \to \mathbb{R}$, and $g \in \mathcal{C}^1[0, T]$, where $\mathcal{C}^1$ is the set of continuously differentiable functions. We define the functionals

$$C_j(f, \boldsymbol{x}, g) := \int_0^T f\big(\boldsymbol{x}(t)\big) g(t) dt + \int_0^T x_j(t) \dot{g}(t) dt; \quad \forall j \in \{1, 2, \ldots, J\}$$

function f
trajectory $\boldsymbol{x}$
testing function g

D-CODE: motivation

Variational formulation of ordinary differential equations

$$\dot{x}_j(t) = f_j(\boldsymbol{x}(t)),\ \forall j = 1, \ldots, J,\ \forall t \in [0, T] \tag{1}$$

Definition 1. Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, continuous functions $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, $f : \mathbb{R}^J \to \mathbb{R}$, and $g \in \mathcal{C}^1[0, T]$, where $\mathcal{C}^1$ is the set of continuously differentiable functions. We define the functionals

$$C_j(f, \boldsymbol{x}, g) := \int_0^T f\big(\boldsymbol{x}(t)\big) g(t) dt + \int_0^T x_j(t) \dot{g}(t) dt; \quad \forall j \in \{1, 2, \ldots, J\}$$

D-CODE: motivation

Variational formulation of ordinary differential equations

$$\dot{x}_j(t) = f_j(\boldsymbol{x}(t)),\ \forall j = 1, \ldots, J,\ \forall t \in [0, T] \tag{1}$$

Definition 1. Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, continuous functions $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, $f : \mathbb{R}^J \to \mathbb{R}$, and $g \in \mathcal{C}^1[0, T]$, where $\mathcal{C}^1$ is the set of continuously differentiable functions. We define the functionals

$$C_j(f, \boldsymbol{x}, g) := \int_0^T f\big(\boldsymbol{x}(t)\big) g(t) dt + \int_0^T x_j(t) \dot{g}(t) dt; \quad \forall j \in \{1, 2, \ldots, J\}$$

D-CODE: motivation

Variational formulation of ordinary differential equations

$$\dot{x}_j(t) = f_j(\boldsymbol{x}(t)),\ \forall j = 1, \ldots, J,\ \forall t \in [0, T] \tag{1}$$

Definition 1. Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, continuous functions $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, $f : \mathbb{R}^J \to \mathbb{R}$, and $g \in \mathcal{C}^1[0, T]$, where $\mathcal{C}^1$ is the set of continuously differentiable functions. We define the functionals

$$C_j(f, \boldsymbol{x}, g) := \int_0^T f\big(\boldsymbol{x}(t)\big) g(t) dt + \int_0^T x_j(t) \dot{g}(t) dt; \quad \forall j \in \{1, 2, \ldots, J\}$$

Hackbusch, W. (2017)
Variational Formulation

D-CODE: motivation

Variational formulation of ordinary differential equations

$$\dot{x}_j(t) = f_j(\boldsymbol{x}(t)),\ \forall j = 1, \ldots, J,\ \forall t \in [0, T] \tag{1}$$

Definition 1. Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, continuous functions $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, $f : \mathbb{R}^J \to \mathbb{R}$, and $g \in \mathcal{C}^1[0, T]$, where $\mathcal{C}^1$ is the set of continuously differentiable functions. We define the functionals

$$C_j(f, \boldsymbol{x}, g) := \int_0^T f\big(\boldsymbol{x}(t)\big) g(t) dt + \int_0^T x_j(t) \dot{g}(t) dt; \quad \forall j \in \{1, 2, \ldots, J\}$$

Proposition 1. *(Hackbusch, 2017) Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, a continuously differentiable function $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, and continuous functions $f_j : \mathbb{R}^J \to \mathbb{R}$ for $j = 1, \ldots, J$. Then $\boldsymbol{x}$ is the solution to the system of ODEs in Equation 1 if and only if*

$$C_j(f_j, \boldsymbol{x}, g) = 0,\ \forall j \in \{1, \ldots, J\},\ \forall g \in \mathcal{C}^1[0, T],\ g(0) = g(T) = 0$$

Hackbusch, W. (2017)
Variational Formulation

D-CODE: motivation

Variational formulation of ordinary differential equations

$$\dot{x}_j(t) = f_j(\boldsymbol{x}(t)),\ \forall j = 1, \ldots, J,\ \forall t \in [0, T] \tag{1}$$

Definition 1. Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, continuous functions $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, $f : \mathbb{R}^J \to \mathbb{R}$, and $g \in \mathcal{C}^1[0, T]$, where $\mathcal{C}^1$ is the set of continuously differentiable functions. We define the functionals

$$C_j(f, \boldsymbol{x}, g) := \int_0^T f\big(\boldsymbol{x}(t)\big) g(t) dt + \int_0^T x_j(t) \dot{g}(t) dt; \quad \forall j \in \{1, 2, \ldots, J\}$$

Proposition 1. *(Hackbusch, 2017) Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, a continuously differentiable function $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, and continuous functions $f_j : \mathbb{R}^J \to \mathbb{R}$ for $j = 1, \ldots, J$. Then $\boldsymbol{x}$ is the solution to the system of ODEs in Equation 1 if and only if*

$$C_j(f_j, \boldsymbol{x}, g) = 0,\ \forall j \in \{1, \ldots, J\},\ \forall g \in \mathcal{C}^1[0, T],\ g(0) = g(T) = 0$$

Hackbusch, W. (2017)
Variational Formulation

D-CODE: motivation

Variational formulation of ordinary differential equations

Proposition 1. *(Hackbusch, 2017) Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, a continuously differentiable function $\boldsymbol{x} : [0,T] \to \mathbb{R}^J$, and continuous functions $f_j : \mathbb{R}^J \to \mathbb{R}$ for $j = 1, \ldots, J$. Then $\boldsymbol{x}$ is the solution to the system of ODEs in Equation 1 if and only if*

$$C_j(f_j, \boldsymbol{x}, g) = 0, \ \forall j \in \{1, \ldots, J\}, \ \forall g \in \mathcal{C}^1[0,T], \ g(0) = g(T) = 0$$

Proposition 1. *(Hackbusch, 2017) Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, a continuously differentiable function $\boldsymbol{x} : [0,T] \to \mathbb{R}^J$, and continuous functions $f_j : \mathbb{R}^J \to \mathbb{R}$ for $j = 1, \ldots, J$. Then $\boldsymbol{x}$ is the solution to the system of ODEs in Equation 1 if and only if*

Insight: Minimising value of this functional corresponds to finding better approximations of the true ODE.

D-CODE: theory

$$d_{\boldsymbol{x}}(f, f^*) := ||f \circ \boldsymbol{x} - f^* \circ \boldsymbol{x}||_2 = ||(f - f^*) \circ \boldsymbol{x}||_2$$

Theorem 1. *Consider $J \in \mathbb{N}^+$, $j \in \{1, \dots, J\}$, $T \in \mathbb{R}^+$. Let $f^* : \mathbb{R}^J \to \mathbb{R}$ be a continuous function, and let $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$ be a continuously differentiable function satisfying $\dot{x}_j(t) = f^*(\boldsymbol{x}(t))$. Consider a sequence of functions $(\widehat{\boldsymbol{x}}_k)$, where $\widehat{\boldsymbol{x}}_k : [0, T] \to \mathbb{R}^J$ is a continuously differentiable function. If $(\widehat{\boldsymbol{x}}_k)$ converges to $\boldsymbol{x}$ in L^2 norm. Then for any Lipschitz continuous function f*

$$\lim_{S \to \infty} \lim_{k \to \infty} \sum_{s=1}^{S} C_j(f, \widehat{\boldsymbol{x}}_k, g_s)^2 = d_{\boldsymbol{x}}(f, f^*)^2, \tag{7}$$

where $\{g_1, g_2, \dots\}$ is a Hilbert (orthonormal) basis for $L^2[0, T]$ such that $\forall i$, $g_i(0) = g_i(T) = 0$ and $g_i \in \mathcal{C}^1[0, T]$.

D-CODE: theory

$$d_{\boldsymbol{x}}(f, f^*) := ||f \circ \boldsymbol{x} - f^* \circ \boldsymbol{x}||_2 = ||(f - f^*) \circ \boldsymbol{x}||_2$$

Theorem 1. *Consider $J \in \mathbb{N}^+$, $j \in \{1, \dots, J\}$, $T \in \mathbb{R}^+$. Let $f^* : \mathbb{R}^J \to \mathbb{R}$ be a continuous function, and let $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$ be a continuously differentiable function satisfying $\dot{x}_j(t) = f^*(\boldsymbol{x}(t))$. Consider a sequence of functions $(\widehat{\boldsymbol{x}}_k)$, where $\widehat{\boldsymbol{x}}_k : [0, T] \to \mathbb{R}^J$ is a continuously differentiable function. If $(\widehat{\boldsymbol{x}}_k)$ converges to $\boldsymbol{x}$ in L^2 norm. Then for any Lipschitz continuous function f*

$$\lim_{S \to \infty} \lim_{k \to \infty} \sum_{s=1}^{S} C_j(f, \widehat{\boldsymbol{x}}_k, g_s)^2 = d_{\boldsymbol{x}}(f, f^*)^2, \tag{7}$$

where $\{g_1, g_2, \dots\}$ is a Hilbert (orthonormal) basis for $L^2[0, T]$ such that $\forall i$, $g_i(0) = g_i(T) = 0$ and $g_i \in \mathcal{C}^1[0, T]$.

Theoretical Justification for Our Algorithm

D-CODE: theory

$$d_{\boldsymbol{x}}(f, f^*) := ||f \circ \boldsymbol{x} - f^* \circ \boldsymbol{x}||_2 = ||(f - f^*) \circ \boldsymbol{x}||_2$$

Theorem 1. *Consider $J \in \mathbb{N}^+$, $j \in \{1, \dots, J\}$, $T \in \mathbb{R}^+$. Let $f^* : \mathbb{R}^J \to \mathbb{R}$ be a continuous function, and let $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$ be a continuously differentiable function satisfying $\dot{x}_j(t) = f^*(\boldsymbol{x}(t))$. Consider a sequence of functions $(\widehat{\boldsymbol{x}}_k)$, where $\widehat{\boldsymbol{x}}_k : [0, T] \to \mathbb{R}^J$ is a continuously differentiable function. If $(\widehat{\boldsymbol{x}}_k)$ converges to $\boldsymbol{x}$ in L^2 norm. Then for any Lipschitz continuous function f*

$$\lim_{S \to \infty} \lim_{k \to \infty} \sum_{s=1}^{S} C_j(f, \widehat{\boldsymbol{x}}_k, g_s)^2 = d_{\boldsymbol{x}}(f, f^*)^2, \tag{7}$$

where $\{g_1, g_2, \dots\}$ is a Hilbert (orthonormal) basis for $L^2[0, T]$ such that $\forall i$, $g_i(0) = g_i(T) = 0$ and $g_i \in \mathcal{C}^1[0, T]$.

Natural choice

$$g_s(t) = \sqrt{2/T}\sin(s\pi t/T)$$

D-CODE: algorithm

Preprocessing

Optimization

D-CODE: algorithm

Preprocessing

$$\left\{\begin{array}{c}\mathbf{y}_1(t)\\ \cdot\\ \cdot\\ \cdot\\ \mathbf{y}_D(t)\end{array}\right\}$$

$t \in \{t_1, t_2, \ldots, T\}$

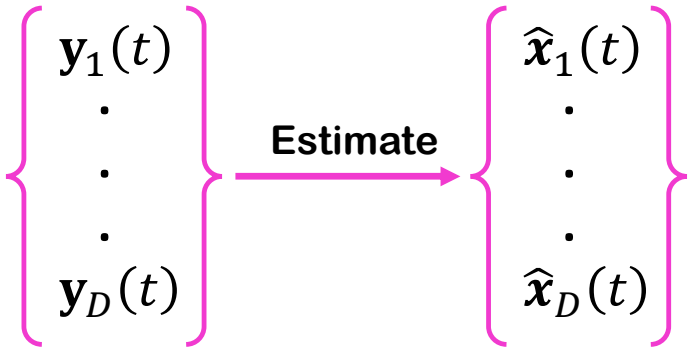
$\mathbf{y}_i(t) \in \mathbb{R}^J$

Optimization

Denoise & Interpolate

D-CODE: algorithm

Preprocessing



$t \in \{t_1, t_2, \ldots, T\}$
$\mathbf{y}_i(t) \in \mathbb{R}^J$

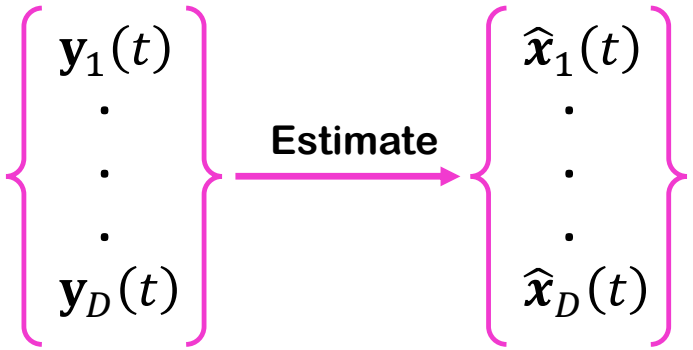
$t \in [0, T]$
$\boldsymbol{x}_i(t) \in \mathbb{R}^J$

Optimization

We estimate trajectories, *not* derivatives!

D-CODE: algorithm

Preprocessing



$t \in \{t_1, t_2, \ldots, T\}$
$\mathbf{y}_i(t) \in \mathbb{R}^J$

$t \in [0, T]$
$\boldsymbol{x}_i(t) \in \mathbb{R}^J$

Optimization

$$C_j(f, \boldsymbol{x}, g) := \int_0^T f\big(\boldsymbol{x}(t)\big) g(t) dt + \int_0^T x_j(t) \dot{g}(t) dt$$

$$\widehat{f}_j = \arg\min_f \sum_{i=1}^{N} \sum_{s=1}^{S} C_j(f, \widehat{\boldsymbol{x}}_i, g_s)^2$$

System of J ODEs

D-CODE: algorithm

Preprocessing

$$\begin{Bmatrix} \mathbf{y}_1(t) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{y}_D(t) \end{Bmatrix} \xrightarrow{\text{Estimate}} \begin{Bmatrix} \hat{\boldsymbol{x}}_1(t) \\ \cdot \\ \cdot \\ \cdot \\ \hat{\boldsymbol{x}}_D(t) \end{Bmatrix}$$

$t \in \{t_1, t_2, \ldots, T\}$
$\mathbf{y}_i(t) \in \mathbb{R}^J$

$t \in [0, T]$
$\boldsymbol{x}_i(t) \in \mathbb{R}^J$

Optimization

$$C_j(f, \boldsymbol{x}, g) := \int_0^T f\big(\boldsymbol{x}(t)\big) g(t) dt + \int_0^T x_j(t) \dot{g}(t) dt$$

$$\widehat{f}_j = \arg\min_f \sum_{i=1}^{N} \sum_{s=1}^{S} C_j(f, \widehat{\boldsymbol{x}}_i, g_s)^2$$

Prespecified testing functions

$$g_s(t) = \sqrt{2/T} \sin(s\pi t/T)$$

Symbolic regression

D-CODE: experiments

Dynamical systems:
- Gompertz model
- Generalized logistic model
- Glycolytic oscillator
- Lorenz system

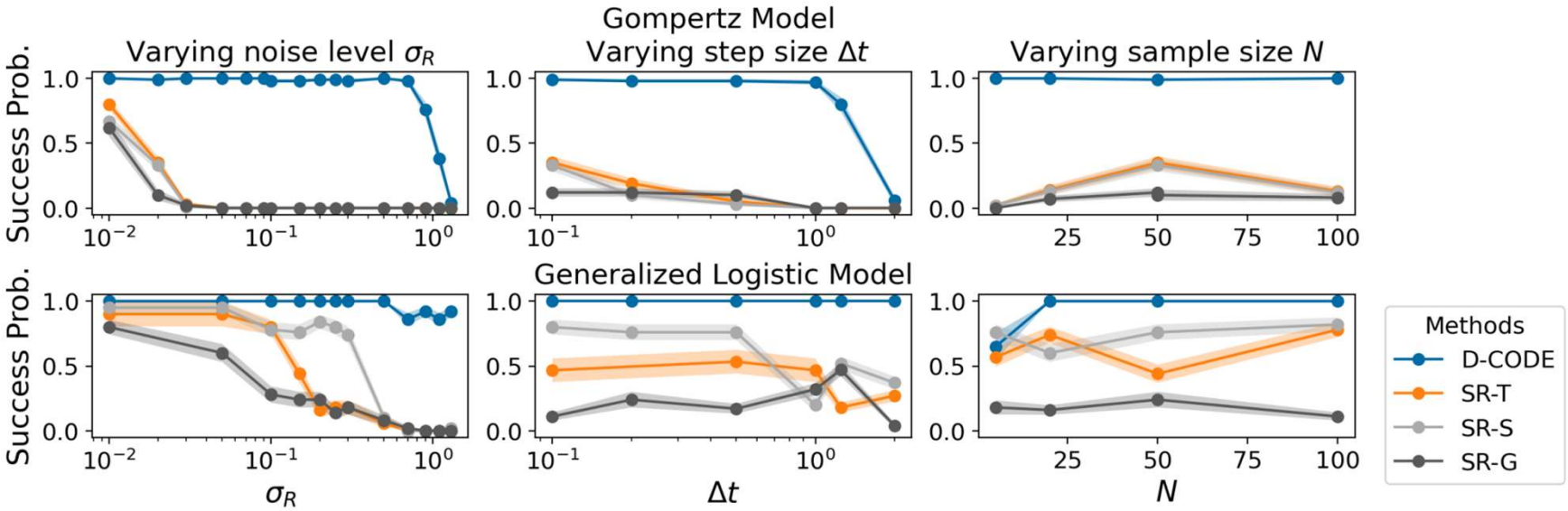
Benchmarks:
Two-step symbolic regression with
a) total variation regularized differentiation (SR-T)
b) spline-smoothed differentiation (SR-S)
c) Gaussian process smoothed differentiation (SR-G)

D-CODE: Experiments

$$\dot{x}(t) = -\theta_1 x(t) \cdot \log\left(\theta_2 x(t)\right) \quad \text{Gompertz Model}$$

$$\dot{x}(t) = \theta_1 x(t) \cdot \left(1 - x(t)^{\theta_2}\right) \quad \text{Generalized Logistic Model}$$

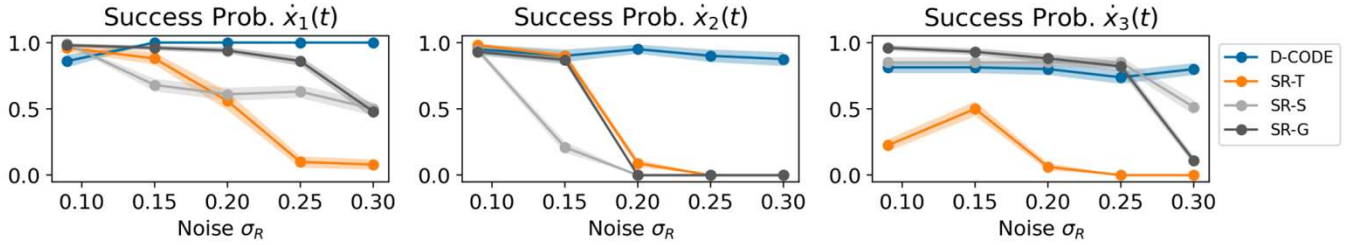
asymmetric growth with saturation

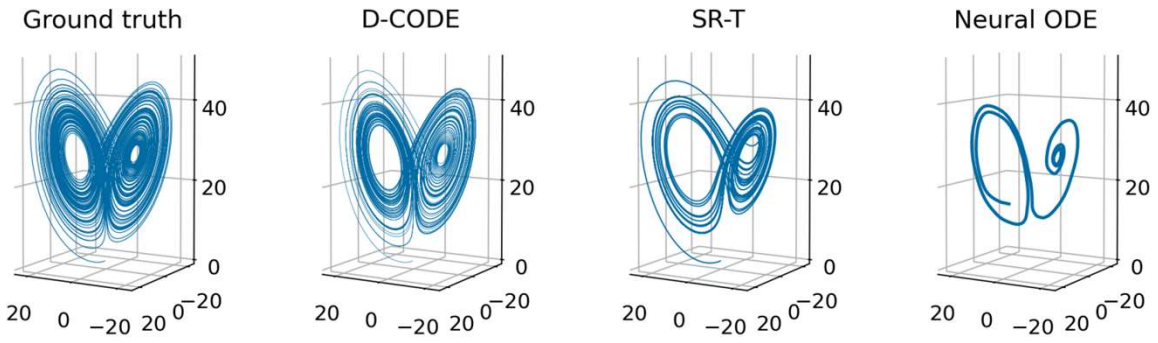


D-CODE: Experiments

Chaotic Lorenz system. The Lorenz system is a model system for chaotic dynamics, defined as:

$$\dot{x}_1(t) = \theta_1\big(x_2(t) - x_1(t)\big);\quad \dot{x}_2(t) = x_1(t)\big(\theta_2 - x_3(t)\big) - x_2(t);\quad \dot{x}_3(t) = x_1(t)x_2(t) - \theta_3 x_3(t)$$

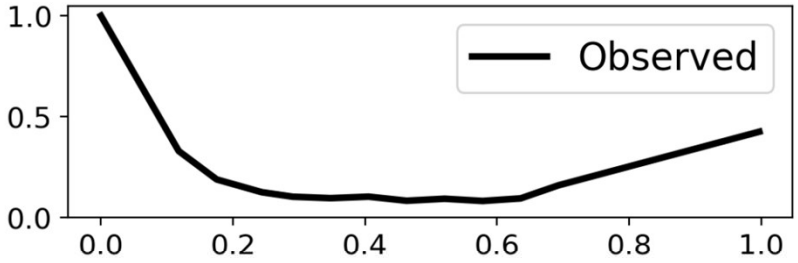




chaotic &
non-periodic systems

D-CODE in action

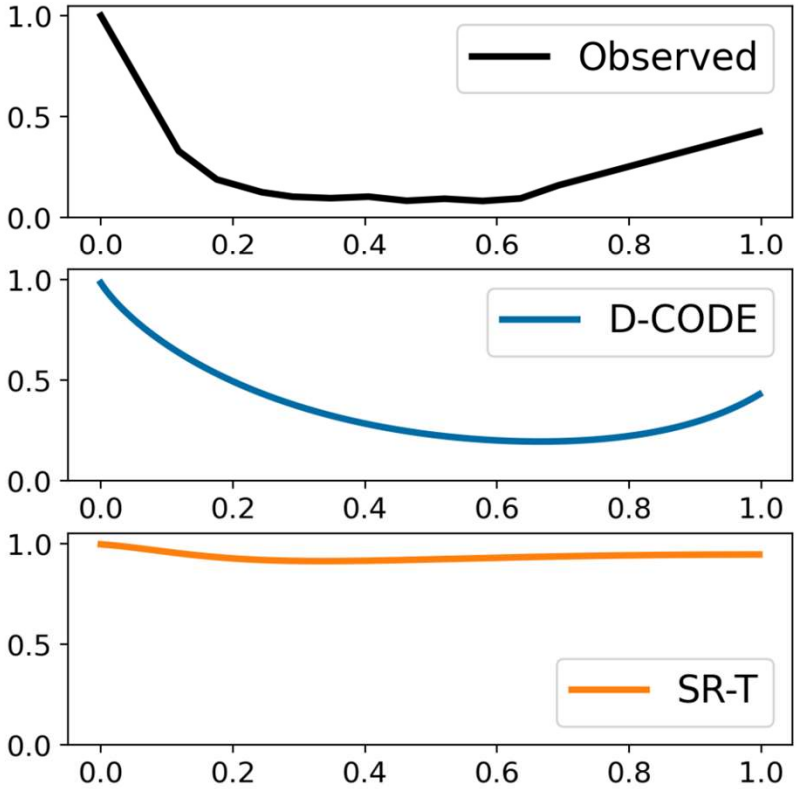
Discover temporal effects of chemotherapy on tumor volume



Dataset: 8 clinical trials on cancer patients

D-CODE in action

Discover temporal effects of chemotherapy on tumor volume



Dataset: 8 clinical trials on cancer patients

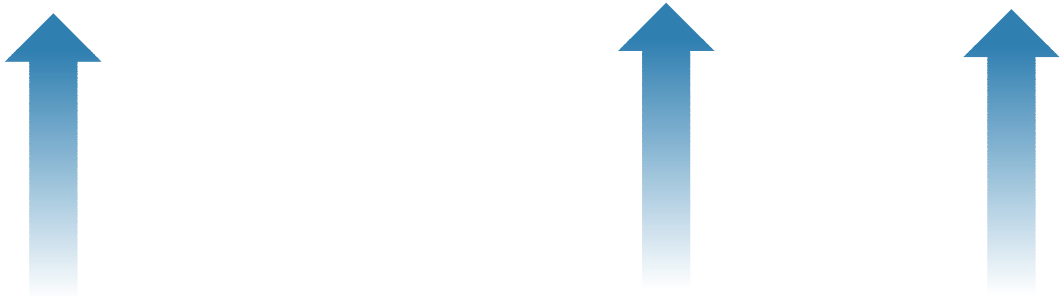
The following two ODEs are discovered by D-CODE and SR-T.

$$\dot{x}(t) = 4.48t^2x(t) + \log(t); \quad \text{D-CODE}$$

$$\dot{x}(t) = 4x(t)\log\big(tx(t)\big)\log\big(tx(t)+2t\big); \quad \text{SR-T}$$

Discovery of governing equations using ML

	Explicit function	Implicit function	Ordinary differential equation	Partial differential equation
Typical form	$y = f(x)$	$f(x, y) = c$	$\frac{dx}{dt} = f(x, t)$	$\frac{\partial u}{\partial t} = f(u, x)$



Symbolic Metamodels [NeurIPS '19, '20]

D-Code [ICLR '22]

D-CIPHER [NeurIPS '22 AI4Science Workshop]

Problem solved?

Simple ODEs – Expert models

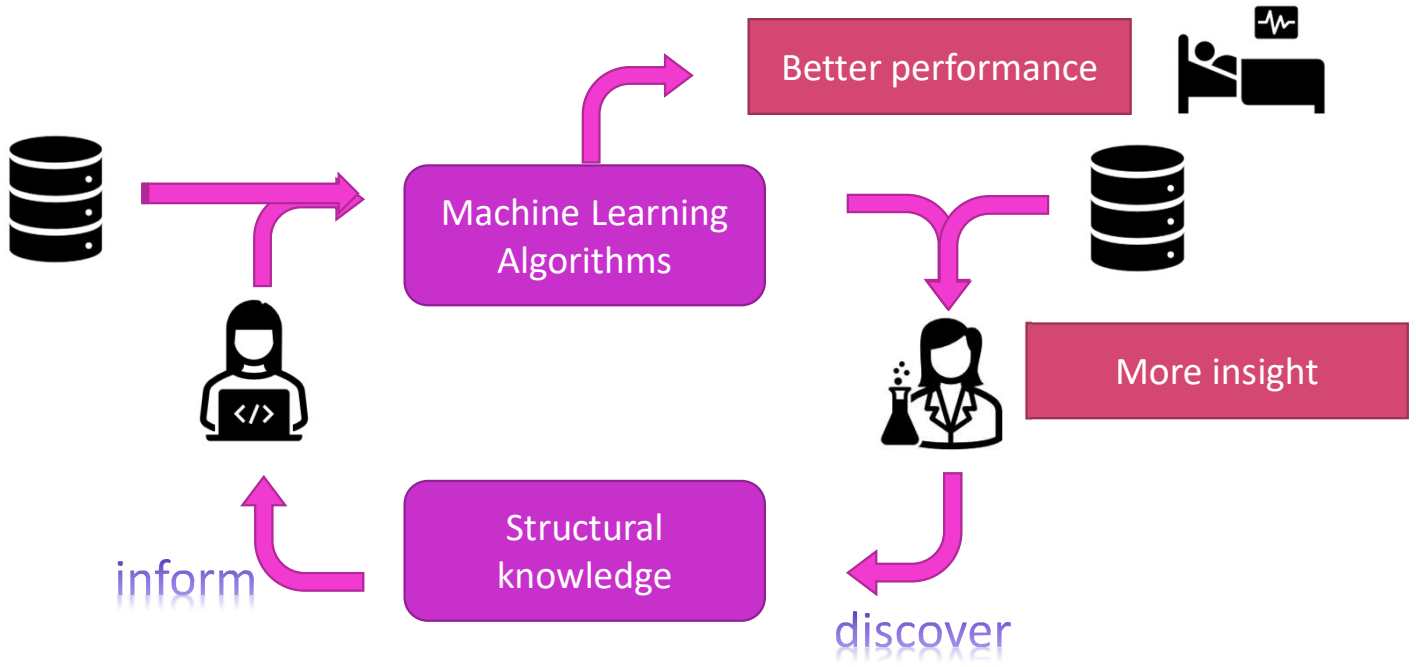
e.g. discovered by ML (D-CODE)

e.g. human-discovered equations - pharmacological models, physiological model etc.

Remaining challenges:

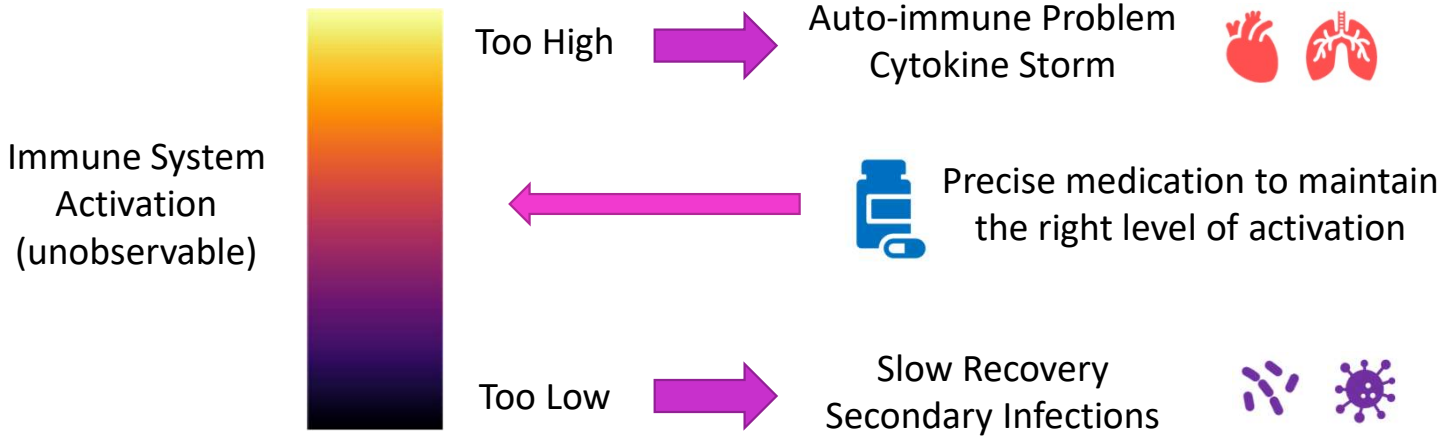
- Complex dynamics and high dimensionality
- Partially observable
- Incorrect
- Incomplete

We proposed ML-enabled Discovery Framework



Impact: repurposing dexamethasone for COVID-19

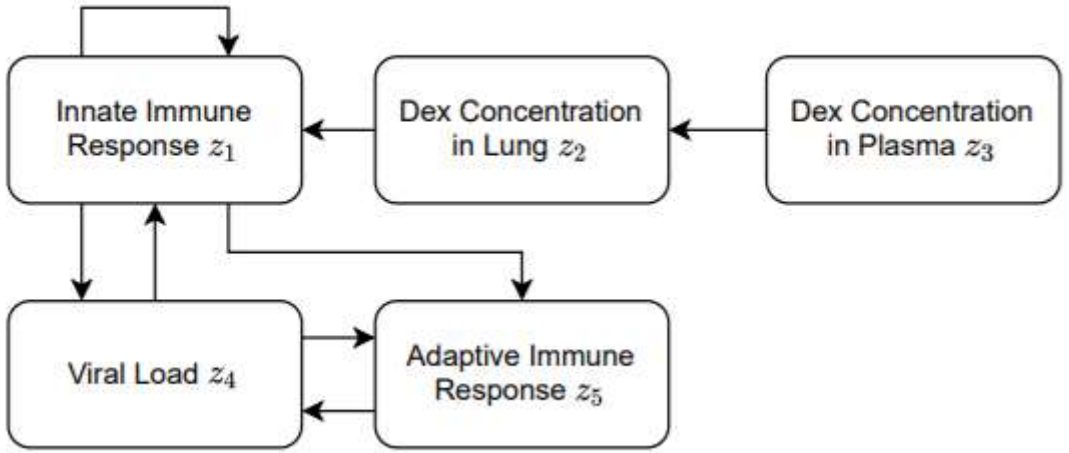
- **Dexamethasone:** the first approved drug for COVID-19 treatment in the UK
- **Well-documented immunosuppressive effect:** Previously used for severe allergies, asthma, COPD
- **Repurposing to COVID-19: a precision dosing problem**
  - Average treatment effect of 6mg flat rate is verified by clinical trials
  - But the clinical practice is much more complex…



Bridge the gap between research lab and clinic

- **Observable clinical variables**
- **PKPD models**: well-studied in the lab
- **Expert variables**: not easily or routinely measured in the clinic

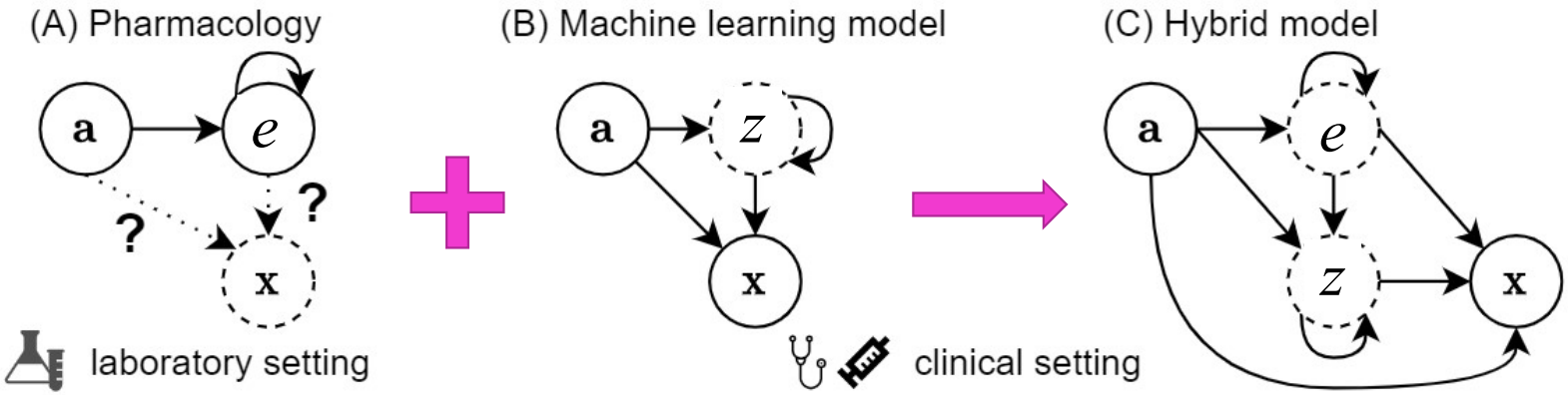
How to use these PKPD models to empower clinicians?



ML solution

Integrating Expert ODEs into Neural ODEs: Pharmacology and Disease Progression
[Qian, Zame, Fleuren, Elbers, vdS, NeurIPS 2021]

Proposed solution: Latent Hybridization Model (LHM)



Expert Model

System of Neural ODE

Latent Hybridization Model

Latent Hybridization Model (LHM)

LHM – advantages

- expert variables and model provide additional insights to users (clinicians)
- provides links between the expert variables and the real-world (clinical) measurements
- underlying model significantly improves sample efficiency

Latent Hybridization Model (LHM)

Simple ODE Model + System of Neural ODE → Latent Hybridization Model

- **LHM**
  - Expert model
  - Latent variables learned by ML
  - Observational time-series data

$$\dot{\mathbf{z}}^e(t) = f^e(\mathbf{z}^e(t), \mathbf{a}(t); \theta^e)$$
$$\dot{\mathbf{z}}^m(t) = f^m(\mathbf{z}^m(t), \mathbf{z}^e(t), \mathbf{a}(t);\ \theta^m)$$
$$\mathbf{x}(t) = g(\mathbf{z}^e(t), \mathbf{z}^m(t), \mathbf{a}(t);\ \gamma)$$

Latent Hybridization Model (LHM)

- **LHM**
  - Expert model
  - Latent variables learned by ML
  - Observational time-series data

$$\dot{\mathbf{z}}^e(t) = f^e(\mathbf{z}^e(t), \mathbf{a}(t); \theta^e)$$
$$\dot{\mathbf{z}}^m(t) = f^m(\mathbf{z}^m(t), \mathbf{z}^e(t), \mathbf{a}(t);\ \theta^m)$$
$$\mathbf{x}(t) = g(\mathbf{z}^e(t), \mathbf{z}^m(t), \mathbf{a}(t);\ \gamma)$$

In LHM, we use observational data to learn

- **the evolution of the unobservable latent variables**

Latent Hybridization Model (LHM)

- **LHM**
  - Expert model
  - Latent variables learned by ML
  - Observational time-series data

$$\dot{\mathbf{z}}^e(t) = f^e(\mathbf{z}^e(t), \mathbf{a}(t); \theta^e)$$
$$\dot{\mathbf{z}}^m(t) = f^m(\mathbf{z}^m(t), \mathbf{z}^e(t), \mathbf{a}(t);\ \theta^m)$$
$$\mathbf{x}(t) = g(\mathbf{z}^e(t), \mathbf{z}^m(t), \mathbf{a}(t);\ \gamma)$$

In LHM, we use observational data to learn

- **the relationship between measurements and all latent variables**

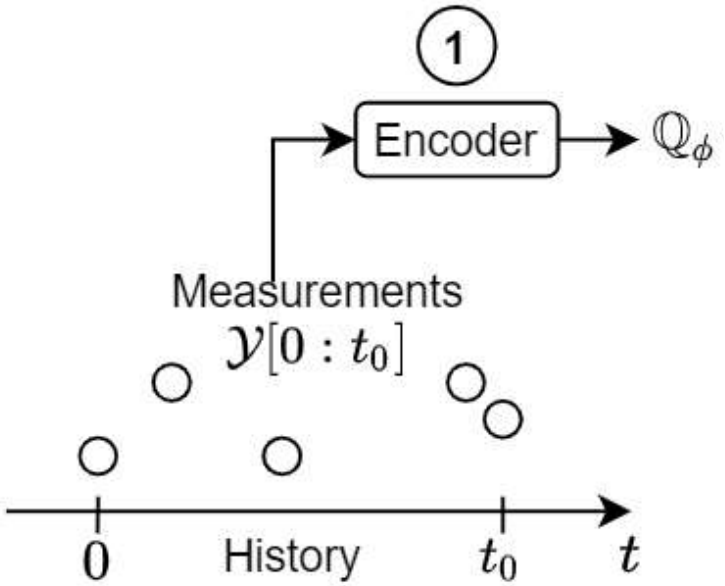
Latent Hybridization Model (LHM)

- **LHM**
  - **Expert model**
  - **Latent variables learned by ML**
  - **Observational time-series data**

$$\dot{\mathbf{z}}^e(t) = f^e(\mathbf{z}^e(t), \mathbf{a}(t); \theta^e)$$
$$\dot{\mathbf{z}}^m(t) = f^m(\mathbf{z}^m(t), \mathbf{z}^e(t), \mathbf{a}(t);\ \theta^m)$$
$$\mathbf{x}(t) = g(\mathbf{z}^e(t), \mathbf{z}^m(t), \mathbf{a}(t);\ \gamma)$$

Learn from data: Unknown coefficients
Estimate from data:
- Initial state of the patient zi(0)
- Variation in initial states reflects heterogeneity of patient population

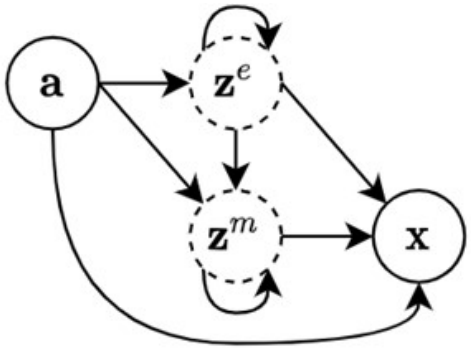
LHM: Learning procedure

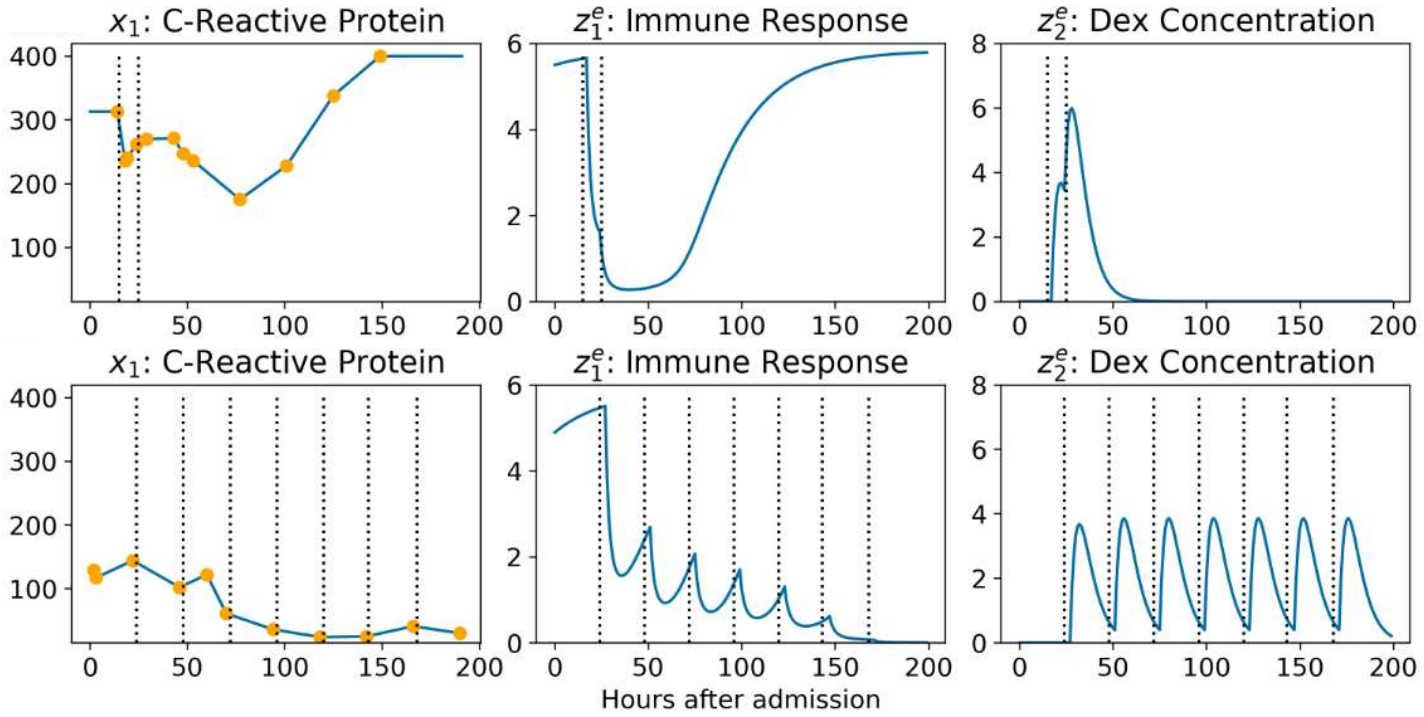


Amortized Variational Inference

Use LHM to provide clinical decision support

Our case study: use of dexamethasone for COVID-19 patients in the ICU





Use LHM to provide clinical decision support

Table 1: Prediction accuracy (RMSE) on COVID-19 intensive care data under different training sample sizes N. Prediction horizon $H = 24$ hours. The standard deviations are shown in the brackets.

Method $\backslash N_0$	100	250	500	1000
Expert	0.718 (0.71)	0.704 (0.02)	0.702 (0.02)	0.713 (0.01)
Residual	0.958 (0.63)	1.003 (0.03)	0.717 (0.05)	0.635 (0.04)
Ensemble	0.707 (0.60)	0.657 (0.05)	0.628 (0.05)	0.599 (0.05)
NODE	0.662 (0.65)	0.659 (0.02)	0.644 (0.05)	0.650 (0.04)
ODE2VAE	0.674 (0.62)	0.666 (0.02)	0.643 (0.02)	0.619 (0.02)
GRU-ODE	0.722 (0.60)	0.673 (0.05)	0.623 (0.05)	0.601 (0.05)
Time LSTM	0.706 (0.63)	0.649 (0.03)	0.600 (0.03)	0.631 (0.02)
LHM	**0.633 (0.51)**	**0.605 (0.02)**	**0.529 (0.02)**	**0.511 (0.02)**

Problem solved?

ODEs (Neural ODEs) are fundamentally inadequate to model systems with more general temporal dynamics such as long-range dependencies or discontinuities

In medicine/science, there are many types of differential equations (DEs)

E.g. Delay Differential Equation (DDE) and Integro-Differential Equation (IDE) – a natural way of capturing the impact of history

Clinical DE examples

Table 1. Families of DEs captured by Neural Laplace.

Model	Equation
ODE	$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}(t))$
DDE	$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}(t), \mathbf{x}(t-\tau)), \tau \in \mathbb{R}^+$
IDE	$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}(t)) + \int_0^t \mathbf{h}(\tau, \mathbf{x}(\tau))d\tau$
Forced ODE	$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}(t), \mathbf{u}(t))$
Stiff ODE	$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}(t)), \exists\, i, j, \dot{x}_i \gg \dot{x}_j$

- **ODE**: PK/PD (pharmacokinetic/pharmacodynamic) models (Koch et al., 2014).
- **DDE**: Delayed PK/PD models (Koch et al., 2014), Cardiac Tissue models (Moreira Gomes et al., 2019).
- **IDE**: Epidemic models (El-Doma et al., 1987).
- **Forced ODE**: Forced oscillation in bio-engineering (Oostveen et al., 2003).
- **Stiff ODE**: Healthcare analytics (Rackauckas et al., 2022).

Clinical DE examples

Table 1. Families of DEs captured by Neural Laplace.

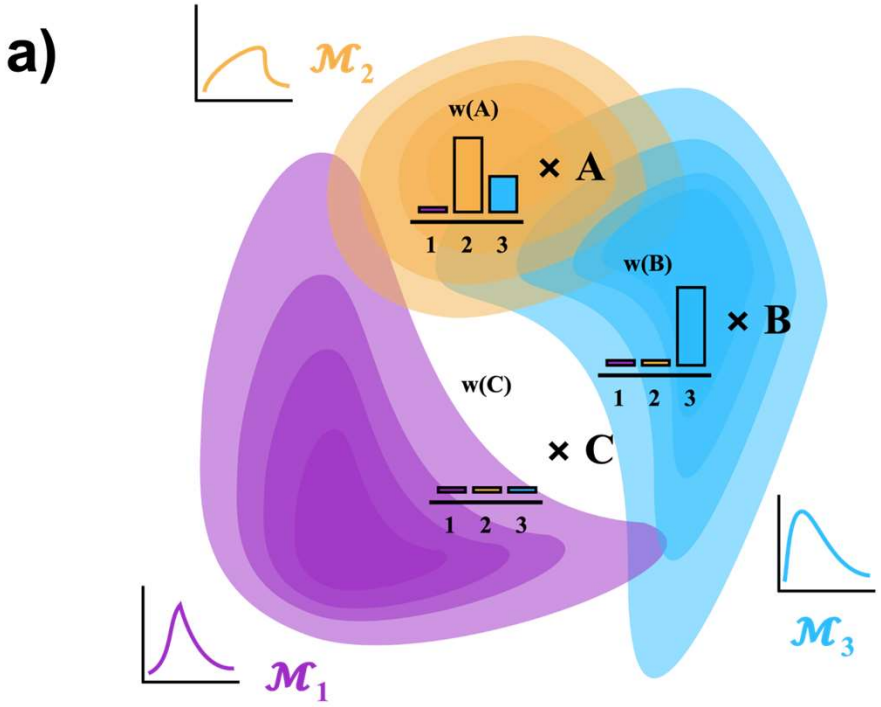
Model	Equation
ODE	$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}(t))$
DDE	$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}(t), \mathbf{x}(t-\tau)), \tau \in \mathbb{R}^+$
IDE	$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}(t)) + \int_0^t \mathbf{h}(\tau, \mathbf{x}(\tau))d\tau$
Forced ODE	$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}(t), \mathbf{u}(t))$
Stiff ODE	$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}(t)), \exists\, i, j, \dot{x}_i \gg \dot{x}_j$

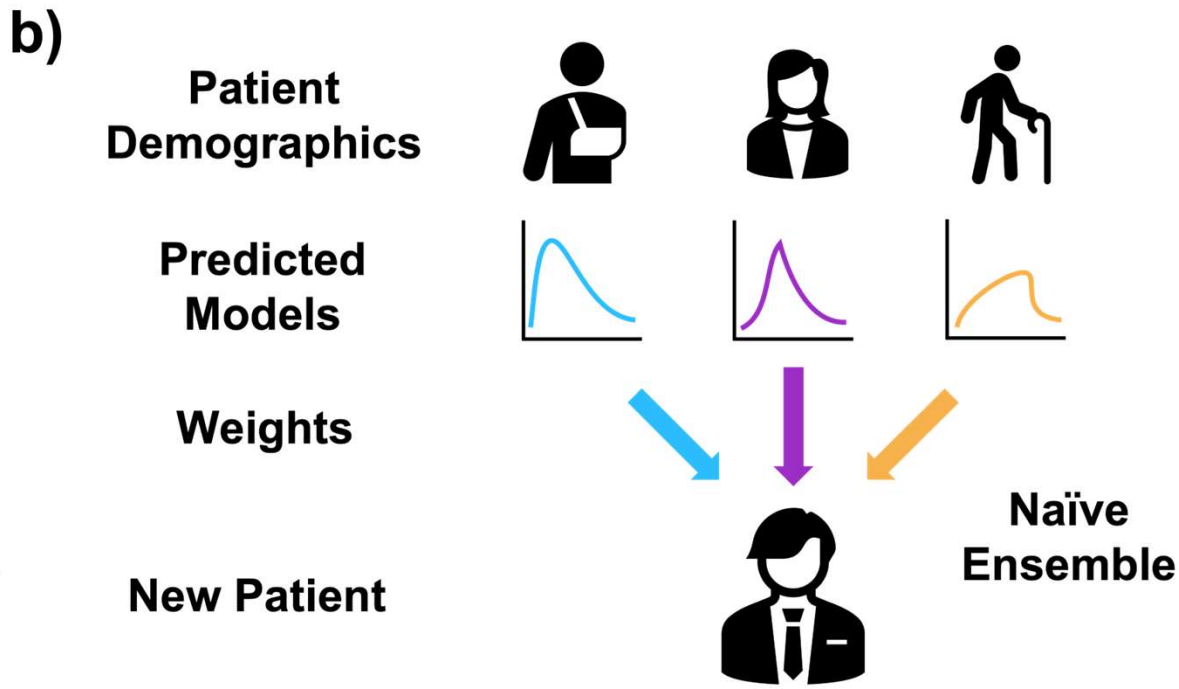
A unified approach to capture many types of DE

Neural Laplace: models broad range of DEs in Laplace domain
[Holt, Qian, vdS, ICML 2022]

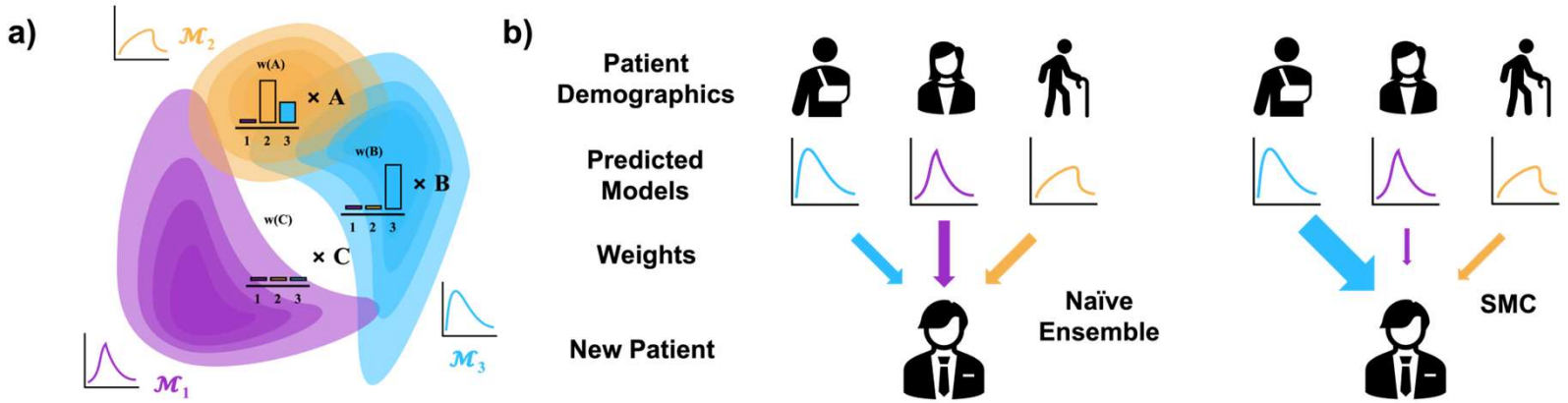
Does not require the user to specify the class of DE a priori
Appropriate class of DE determined implicitly, in a data-driven way.
Significantly extends flexibility and modeling capabilities of Neural ODEs

Other ways to hybridize/combine expert-models?





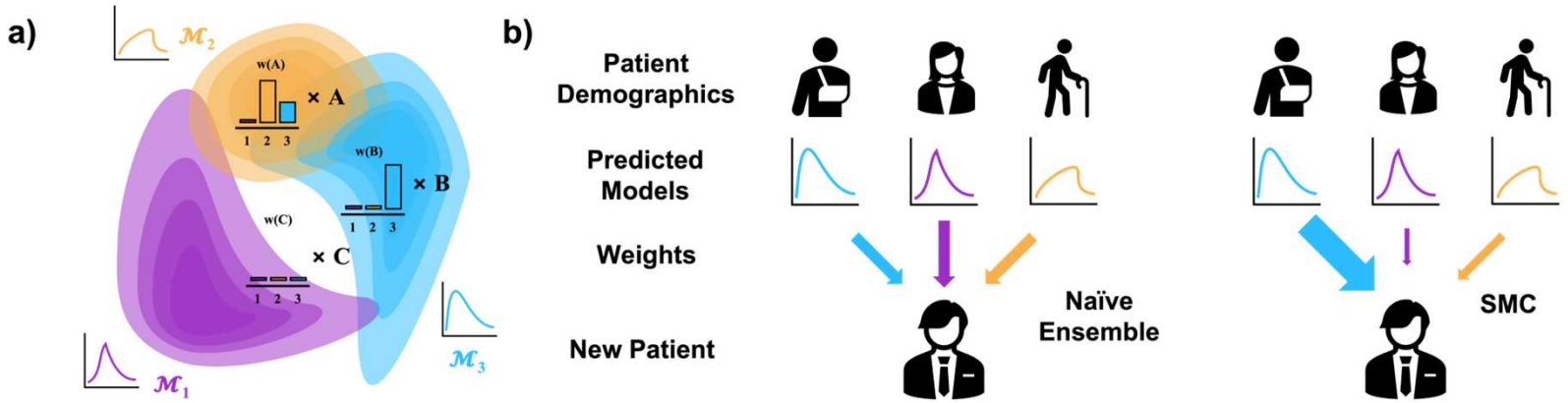
Synthetic Model Combination



A. Chan, vdS, "Synthetic Model Combination: An Instance-wise Approach to Unsupervised Ensemble Learning", NeurIPS 2022

- Novel representation learning for handling sparse high-dimensional domains
- Uses ideas from synthetic control

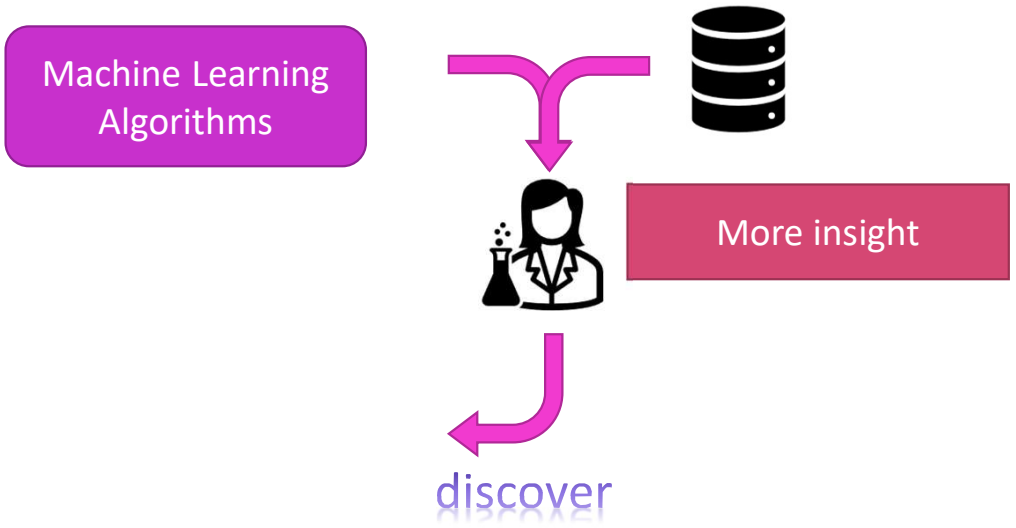
Synthetic Model Combination



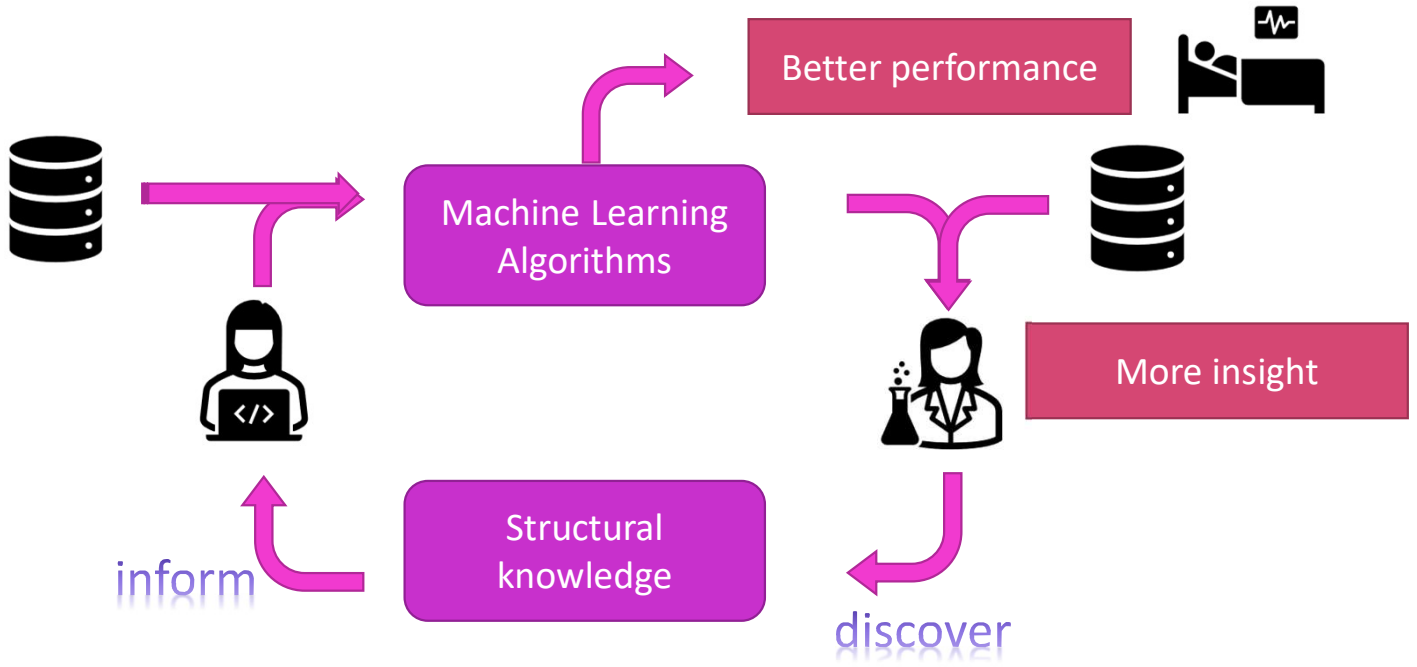
A. Chan, R. Peck, M. Gibbs, vdS, "Synthetic Model Combination: A new machine learning method for pharmacometrics ensembling", Clinical Pharmacology 2023

- Demonstrated use for precision dosing of Vancomycin

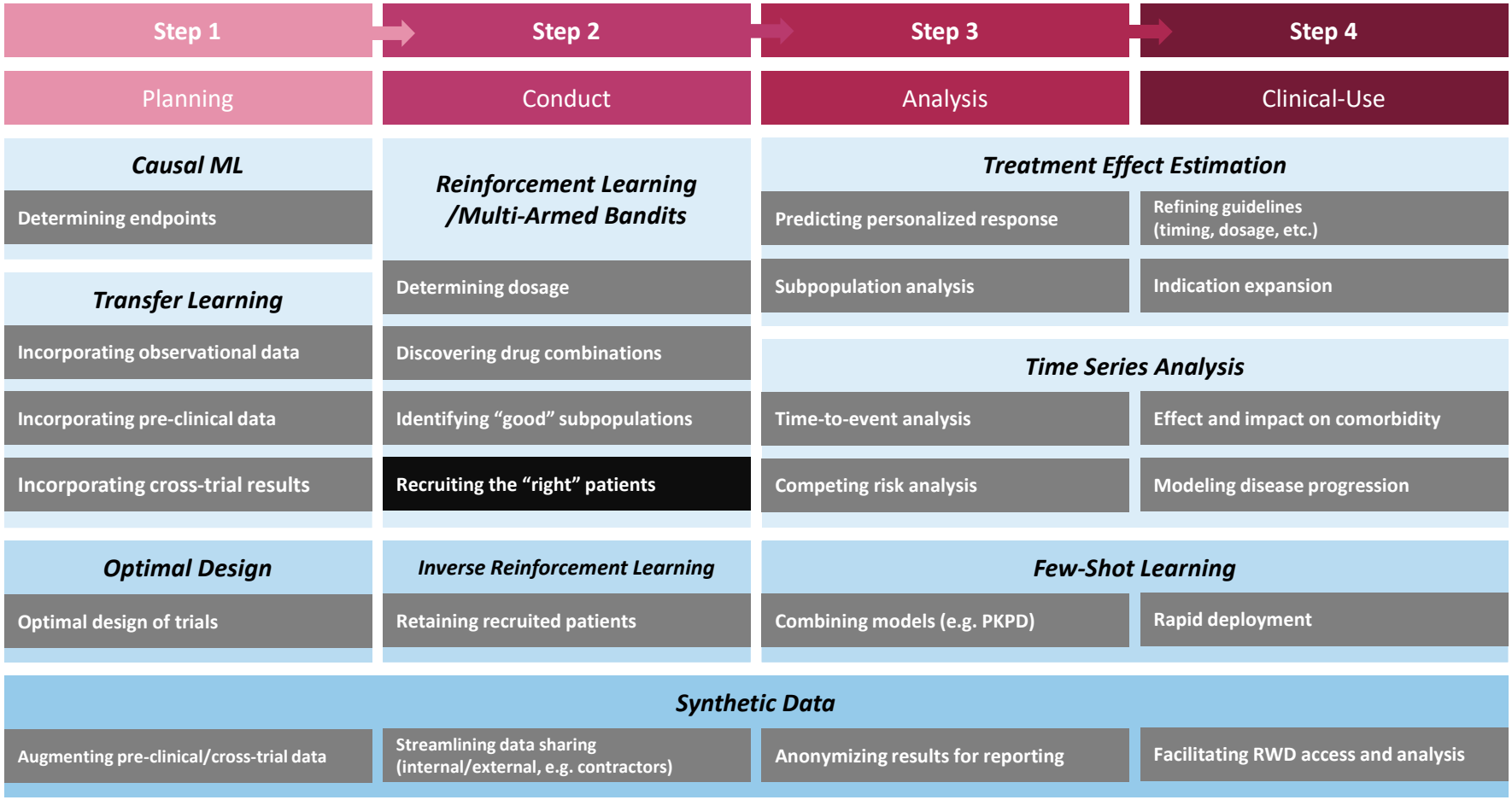
Summary: We proposed ML-enabled Discovery Framework



Summary: We proposed ML-enabled Discovery Framework



Machine learning for clinical trials



Randomized controlled trials

- Gold standard for showing efficacy
- **Problem:** Hard to adapt their design
  Usually targets only one population

Motivating Scenario

- Treatment is effective for a subpopulation
  but ineffective for the overall population
- RCT targeting the overall population → Treatment being denied for the subpopulation

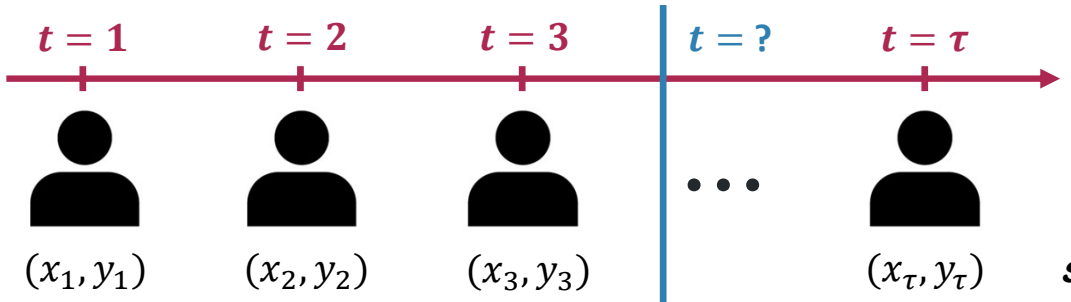
Adaptive experiment designs

Experiment Design	*When?*	*Which?*
RCT	Never	Only the initial population

A. Huyuk, Z. Qian, vdS, When to make & break commitments? ICLR 2023

Optimal Commitment Problem

- A new type of optimal stopping/switching problem
- **Setup:** Experiment with design ψ is launched:



1) Continual costs: $-C_\psi$
2) Uncertain reward: $+R_\psi$ *iff success*
3) Not possible to modify ψ!

$$a_t \in \begin{cases} \psi & \textit{(continue the trial)} \\ \psi' & \textit{(switch to an alternative trial)} \\ \emptyset & \textit{(stop)} \end{cases}$$

OCP as a reinforcement learning problem

- **State** $s_t = (\text{active design } \psi_t, \text{ mean outcomes } \{\theta_x\}_{x \in X})$
- **Transition function** $\mathcal{T}$:

$$\psi_{t+1} = \text{action } a_t \in \{\emptyset, \psi, \psi', \psi'' \ldots\}$$

- **Observations** $\omega_t = (\text{population } x_t, \text{ outcome } y_t)$
- **Observation function** $\mathcal{O}$:

$$x_t \sim \psi_t, \quad y_t \sim \mathcal{N}(\theta_{x_t}, \sigma^2)$$

- **Reward function:**

$$r_t = -C_{\psi_t} + R_{\psi_t} \cdot \underbrace{\mathbb{I}\{\psi_t = \psi_{t-1} = \cdots = \psi_{t-\tau+1}\}}_{\text{commitment until } \tau\text{-samples collected}} \cdot \underbrace{\rho(x_{t-\tau:t}, y_{t-\tau,t})}_{\text{success criteria}}$$

Warm-up: When to break a single commitment?

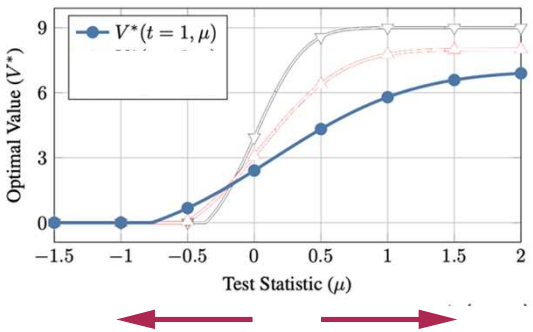
- **One population/design** $a_t \in \{$ **stop** $\emptyset$, **continue** $\psi\}$
- **Observations:** $y_t \sim \mathcal{N}(\theta, \sigma^2 = 1)$
  $\mu_t = (\sum_{t' \le t} y_{t'})/t$
- **Reward function:** $r_t = -C + R \cdot \mathbb{I}\{t = \tau\} \cdot \mathbb{I}\{\mu_\tau > 0\}$

commitment until τ-samples collected

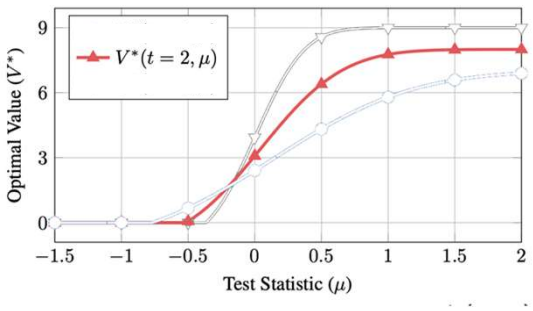
success criteria

Value function is non-convex

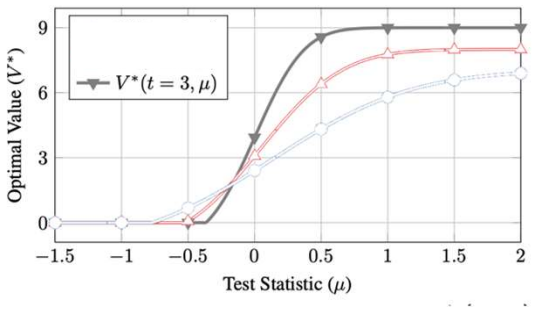
$t = 1$



$t = 2$



$t = 3$

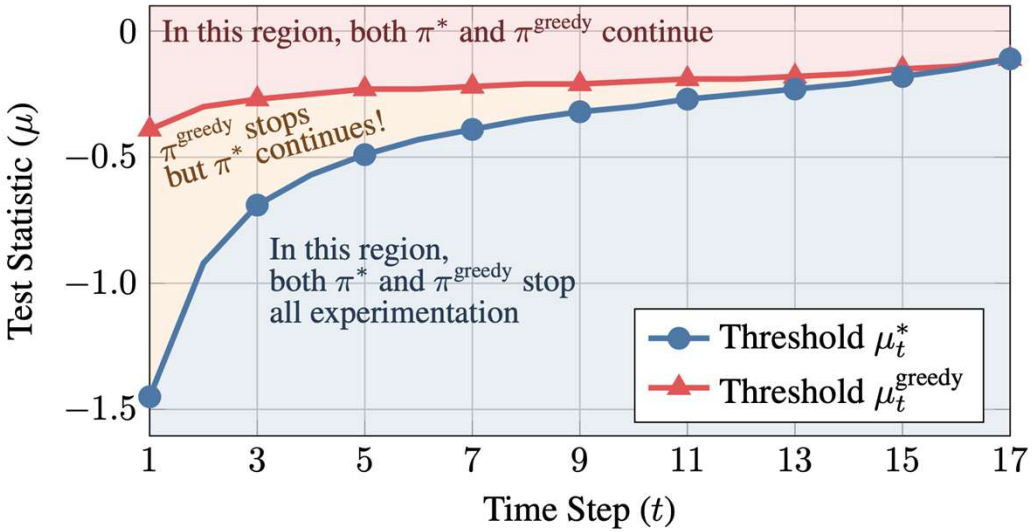


less likely to succeed ($\lim V^* = 0$) **more likely to succeed** ($\lim V^* = R - 2C$)

- **POMDP solvers that rely on convex function approximators are not feasible!**
- **There is a *threshold* μ_t^* for stopping**

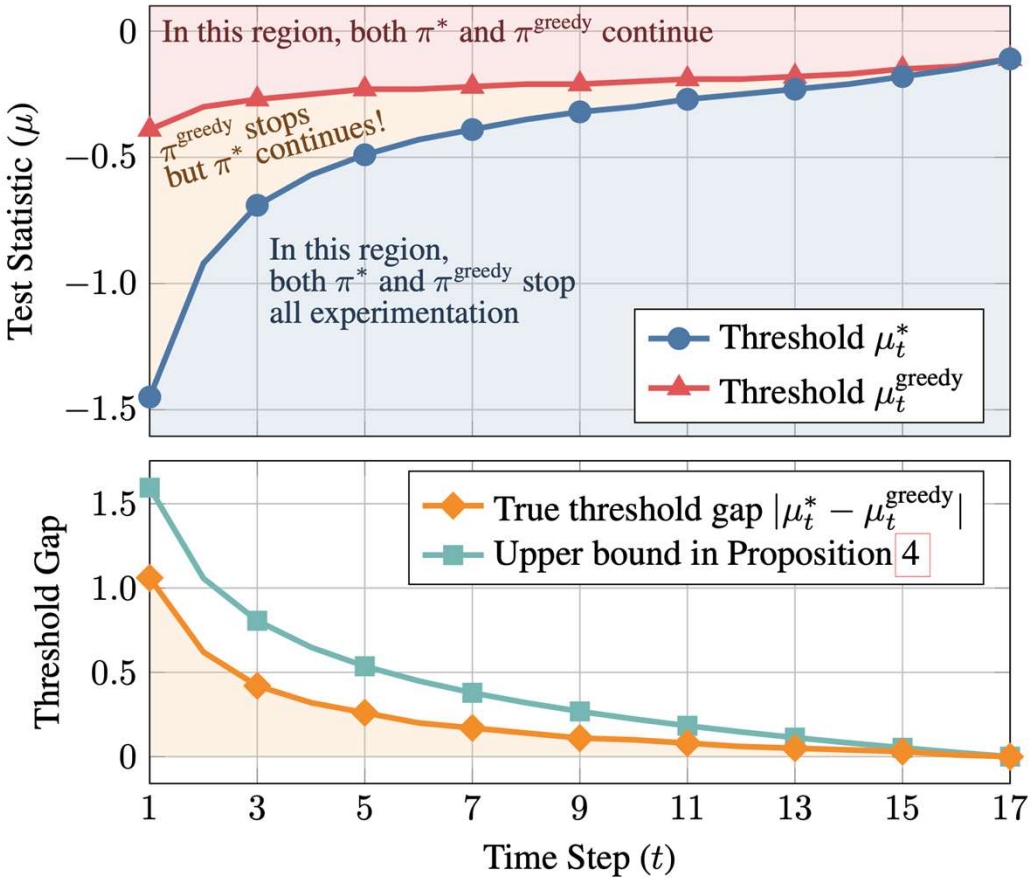
Optimal solution is optimistic

- **Greedy approach: Stop iff** $\mathrm{R} \cdot \mathbb{P}\{\mu_\tau > 0 | \mu_t\} - C \cdot (\tau - t) < 0$
  (***equivalently,*** $\mu_t < \mu_t^{\text{greedy}}$)



(Optimal solution has a lower threshold)

Optimal solution is *increasingly less* optimistic



(The gap between thresholds decreases)

Bayes-OCP

Algorithm 1

Initialize μ_x and σ_x^2 for all $x \in \mathcal{X}$
$X \leftarrow \mathcal{X}$, $\mathcal{D}_0 \leftarrow \varnothing$
Start experiment $\psi = (\mathcal{X}, \tau, \rho)$
loop:
    Observe x_t, y_t; $\mathcal{D}_t \leftarrow \mathcal{D}_{t-1} \cup \{x_t, y_t\}$
    $1/\sigma_{x_t}^2 \leftarrow 1/\sigma_{x_t}^2 + 1$
    $\mu_{x_t} \leftarrow \mu_{x_t} + (y_t - \mu_{x_t})\sigma_{x_t}^2$

(i) Identify a candidate subpopulation X' to replace X:

    $X' \leftarrow \varnothing$
    while $X \setminus X' \supset \varnothing$:
        $x^* \leftarrow \text{argmax}_{x \in X \setminus X'}$
            $\mathbb{E}_{\theta_x \sim \mathcal{N}(\mu_x, \sigma_x^2)}[\mathcal{G}^{(0)}(X' \cup \{x\}; \{\theta_x\})]$
        if $\mathbb{E}_{\theta_x \sim \mathcal{N}(\mu_x, \sigma_x^2)}[\mathcal{G}^{(0)}(X' \cup \{x^*\}; \{\theta_x\})]$
                $> \mathbb{E}_{\theta_x \sim \mathcal{N}(\mu_x, \sigma_x^2)}[\mathcal{G}^{(0)}(X'; \{\theta_x\})]$:
            $X' \leftarrow X' \cup \{x^*\}$
        else: break

(ii) Decide whether to actually replace X with X':

    if $\mathbb{P}_{\theta_x \sim \mathcal{N}(\mu_x, \sigma_x^2)}\{\mathcal{G}^{(0)}(X'; \{\theta_x\})$
            $> \mathcal{G}(X, \mathcal{D}_t; \{\theta_x\})\} > \beta$:
        $X \leftarrow X'$, $\mathcal{D}_0 \leftarrow \varnothing$
        Start a new experiment $\psi = (X, \tau, \rho)$

Bayesian posterior

Identifying a candidate experiment
- **(combinatorial search)**

Comparing the candidate and ongoing experiments

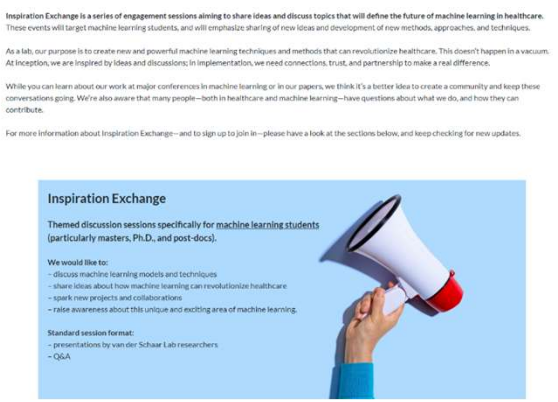
ML and related techniques can address many of them

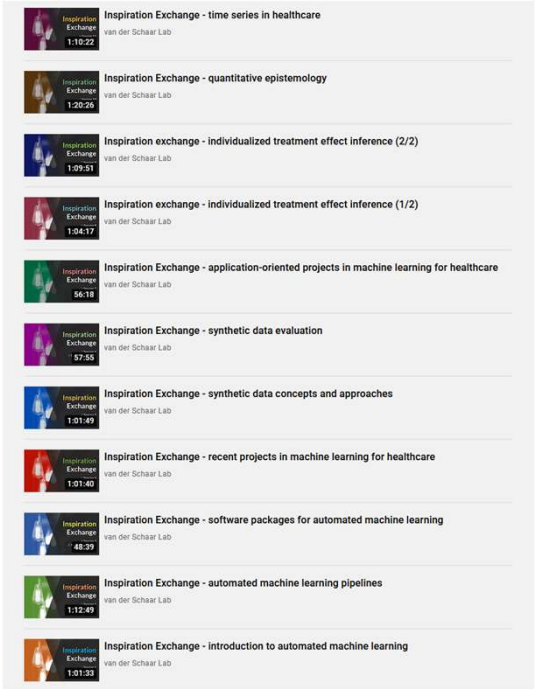
Stage 1	Stage 2	Stage 3	Stage 4
Planning	Conduct	Analysis	Commercialization
Causal ML	***Reinforcement Learning /Multi-Armed Bandits***	***Treatment Effect Estimation***	
Determining endpoints		Predicting personalized response	Refining guidelines (timing, dosage, etc.)
Transfer Learning	Determining dosage	Subpopulation analysis	Indication expansion
Incorporating observational data	Discovering drug combinations	***Time Series Analysis***	
Incorporating pre-clinical data	Identifying "good" subpopulations	Time-to-event analysis	Effect and impact on comorbidity
Incorporating cross-trial results	Recruiting "right" patients	Competing risk analysis	Modeling disease progression
Optimal Design	***Inverse Reinforcement Learning***	***Few-Shot Learning***	
Optimal design of trials	Retaining recruited patients	Combining models (e.g. PKPD)	Rapid deployment
Synthetic Data			
Augmenting pre-clinical/cross-trial data	Streamlining data sharing (internal/external, e.g. contractors)	Anonymizing results for reporting	Facilitating RWD access and analysis

Engagement sessions: Inspiration Exchange

Online engagement sessions for ML researchers in healthcare; themed presentations & Q&A

https://www.vanderschaar-lab.com/
→ Engagement sessions
→ Inspiration Exchange





Subscribe & join us!

van_der_Schaar \ LAB

vanderschaar-lab.com

UNIVERSITY OF CAMBRIDGE