A nationally-implemented AI solution for Covid-19

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

Email: mv472@cam.ac.uk
Website: http://www.vanderschaar-lab.com/

Agenda

Using AI/ML to fight Covid-19 at:

- Global Level - Policy Impact Predictor
- National Level - Resource Planning
- Patient Level - Interpretable and explainable AI/ML

Joining forces to transform healthcare!

Why is AI/ML for healthcare is different?

AI/ML has accomplished wonders …. on well-posed problems where the notion of a "solution" is well-defined and solutions are verifiable

Healthcare is different – problems are not well-posed and notion of a "solution" is often not well-defined and solutions are hard to verify

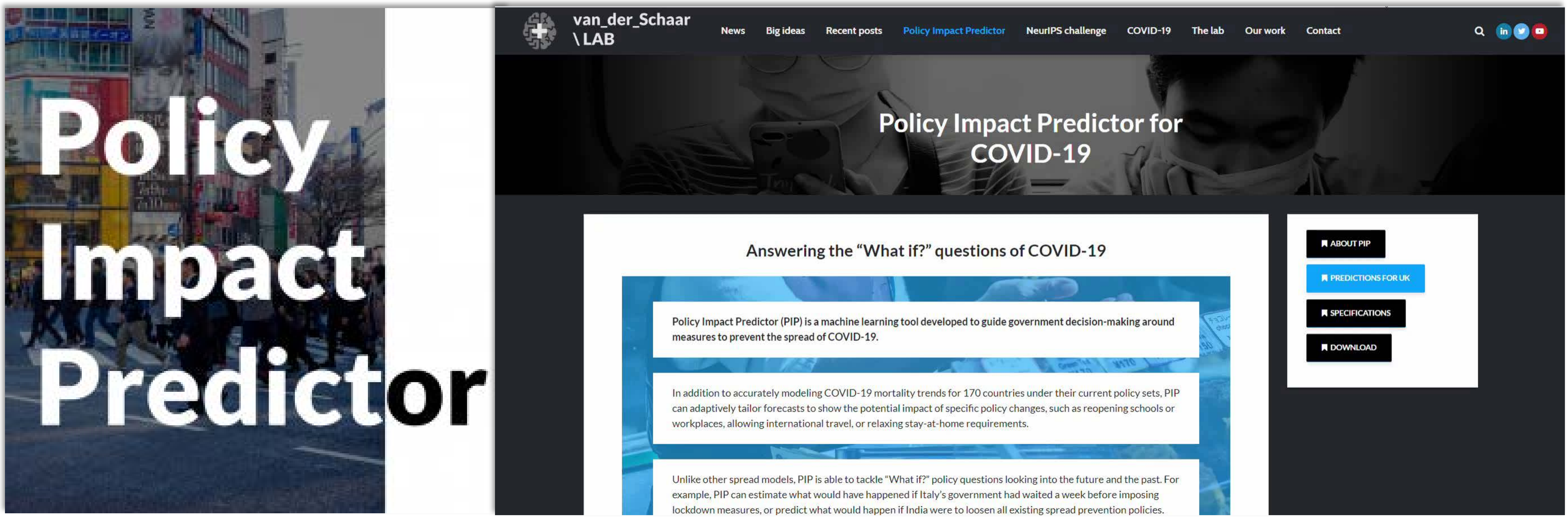
This presents enormous challenges – and also enormous opportunities

Goal: Augment human decision making

- clinicians, medical researchers and policy makers

New ML <u>models and techniques</u>

Covid-19 at the Global Level



https://www.vanderschaar-lab.com/policy-impact-predictor-for-covid-19/

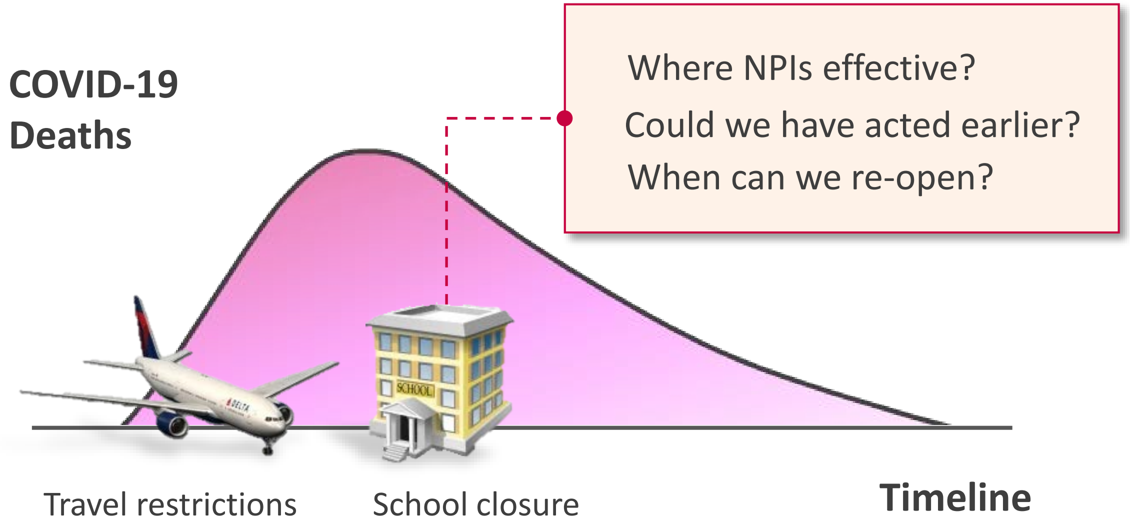
Non-pharmaceutical interventions to contain COVID-19

- **The problem:**

  Estimating the *causal effects of NPIs* applied over time on COVID-19 deaths to conduct *counterfactual scenario analysis*.

- **Why is it important?**

  Inform governments and policy-makers on what NPIs to apply over the next months.



- **What is new?**

  Learning *heterogeneous* NPI effects using *global* data from different countries.

- **Potential impact…**

  Reduction of future COVID-19 deaths in various countries around the world.

Current Modeling Efforts: CDC National Forecasts*

IHME

University of Washington

Combination of a mechanistic disease transmission model and a curve-fitting approach.

Imperial

Imperial College London

Ensembles of mechanistic transmission models, fit to different parameter assumptions.

LANL

Los Alamos National Laboratory

Statistical dynamical growth model accounting for population susceptibility.

DELPHI

Massachusetts Institute of Technology

SEIR model.

YYG

Independent

SEIS mechanistic model

CDC-ensemble

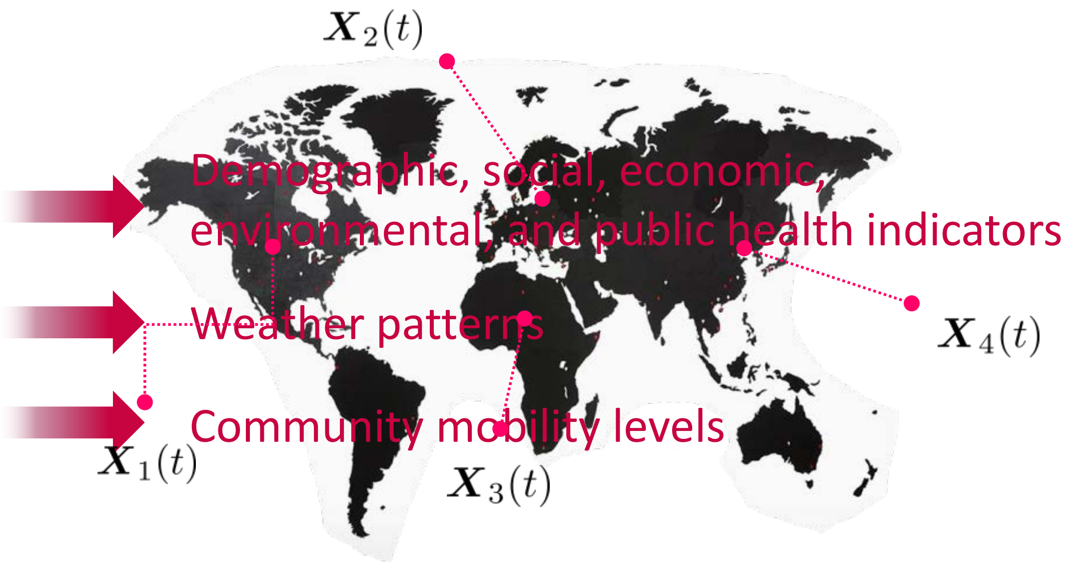
Centre for Disease Control and Prevention

An ensemble of 21 individual forecasts.

*https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html

Problem Formulation

- COVID-19 deaths over time: $Y_i(t) \in \mathbb{N} \cup \{0\}$ for country $i \in \{1, \ldots, N\}$
- Country-level features $\boldsymbol{X}_i(t) \in \mathbb{R}^d$
  - Country meta-data $\boldsymbol{X}_i^s$
  - Exogenous time-varying features $\boldsymbol{X}_i^{ex}(t)$
  - Endogenous time-varying features $\boldsymbol{X}_i^{en}(t)$



- NPI indicators: $\mathcal{P}_i[1:t] \triangleq \{p_i(1), \ldots, p_i(t)\}$
  - $p_i(t) \in \{0,1\}^K$

  K policy indicators

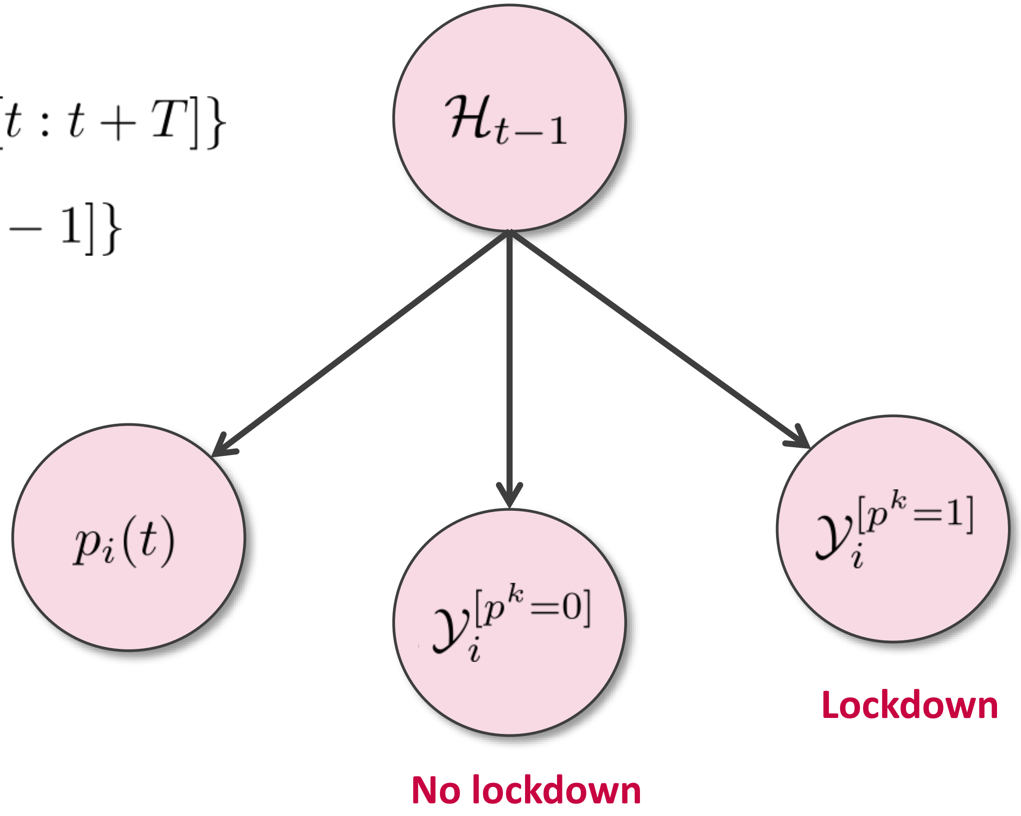
Dataset

$$\mathcal{D}_{N,t} \triangleq \{\boldsymbol{X}_i, \mathcal{Y}_i[1:t], \mathcal{P}_i[1:t]\}_{i=1}^N$$

Problem Formulation

- **Potential outcomes framework**
  - Interventions: $p_i(t) = [p_i^1(t), \ldots, p_i^K(t)]$
  - Potential outcomes: $\{\mathcal{Y}_i^{[p^k=1]}[t : t+T], \mathcal{Y}_i^{[p^k=0]}[t : t+T]\}$
  - Confounders: $\mathcal{H}_{t-1} \triangleq \{\boldsymbol{X}_i, \mathcal{Y}_i[1 : t-1], \mathcal{P}_i[1 : t-1]\}$
- **Assumptions**
  - No unobserved confounders

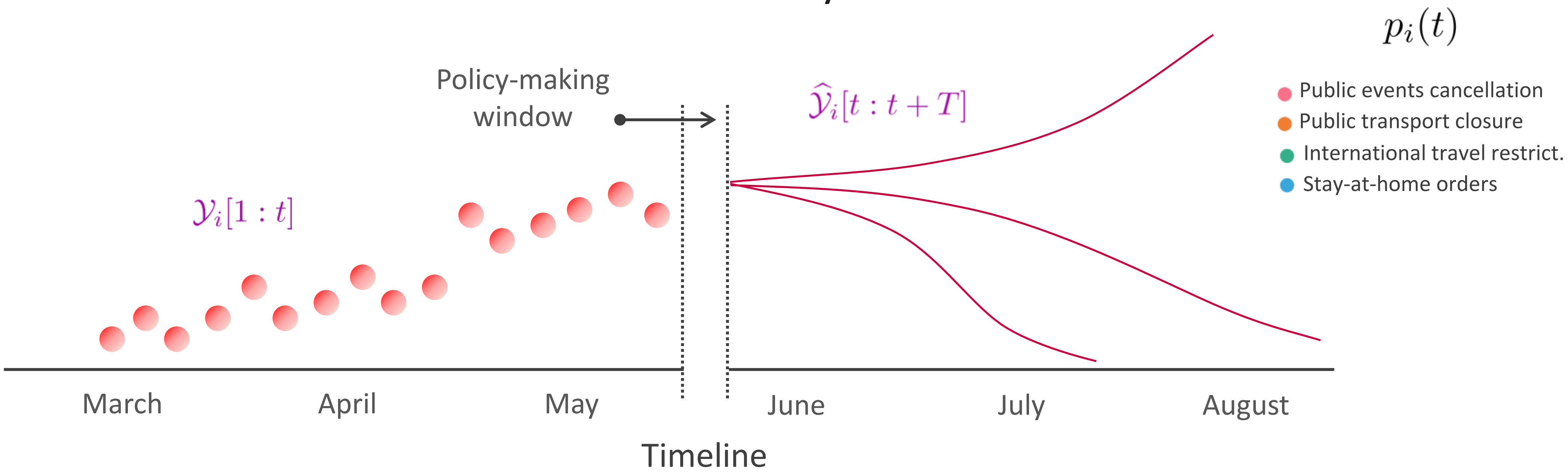
$$\mathcal{Y}_i^{[p^k]}[t : t+T] \perp p_i(t) \,|\, \mathcal{H}_{t-1}$$



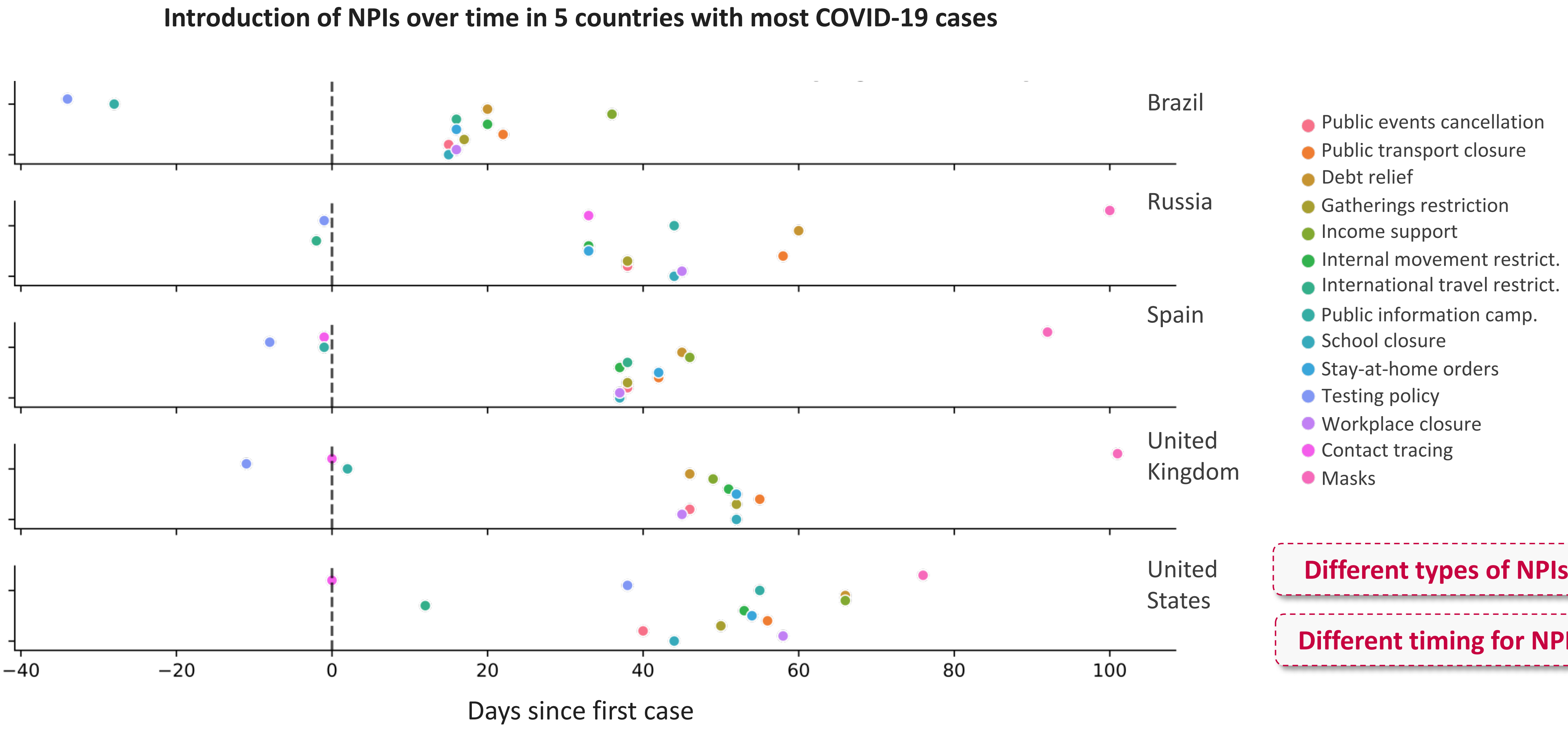
COVID-19 Counterfactual Scenario Analysis

- For a given country $i \in \{1, \ldots, N\}$, forecast the trajectory of COVID-19 deaths within a future time horizon of T days under a given set of future NPIs

$$\widehat{\mathcal{Y}}_i[t:t+T] = \mathbb{E}\Big[\mathcal{Y}_i[t:t+T] \,\Big|\, \underbrace{p_i(t),}_{\textbf{Policy decision}}\, \mathcal{H}_{t-1}\Big]$$



Global Modeling: Exploit Heterogeneity of NPIs



Different types of NPIs

Different timing for NPIs

Modeling COVID-19 Death Curves

- **Model the trajectory of COVID-19 deaths within country *i* as a Gaussian process**

A function of time

$$f_{L,i} \sim \mathcal{GP}(D_{\theta_i}(t), K_{\theta_i}(t, t'))$$

Mean function

Kernel function

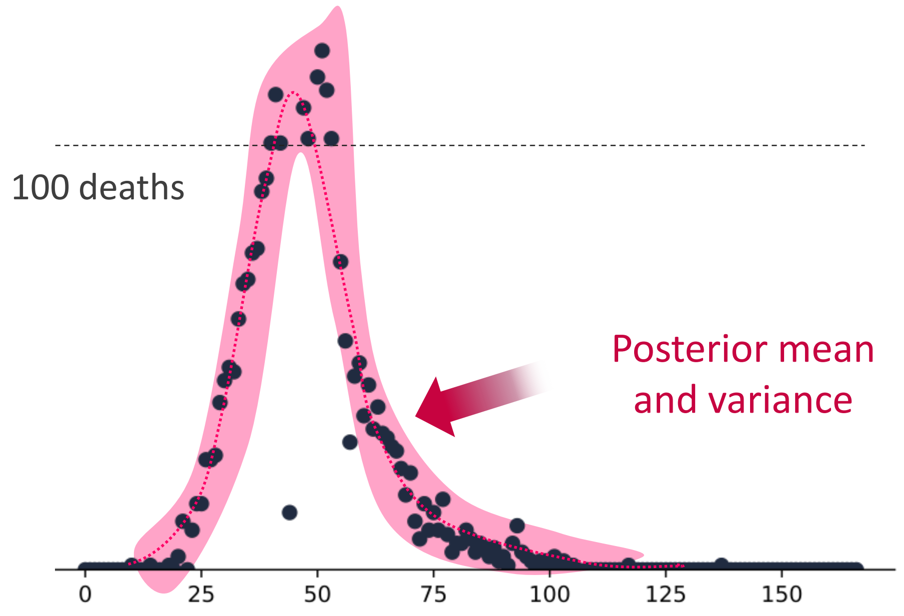
- Mean function = compartmental model

$D_{\theta_i}(t)$ → SEIR model

- Kernel function

$$K_{\theta_i}(t, t') = \exp\left(\frac{-1}{2} \|t - t'\|^2\right)$$

Daily COVID-19 deaths in China



Modeling COVID-19 Death Curves

- Model the trajectory of COVID-19 deaths within country *i* as a Gaussian process

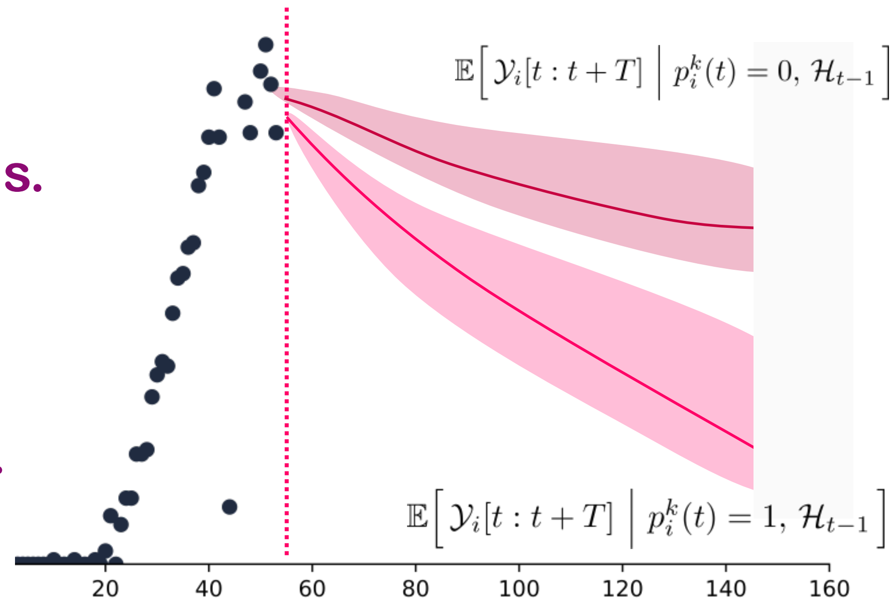
A function of time $$f_{L,i} \sim \mathcal{GP}(D_{\theta_i}(t), K_{\theta_i}(t,t'))$$

- Mean function = compartmental model

$D_{\theta_i}(t)$ → SEIR model

- Death forecasts = posterior trajectory of deaths.
- Prior = rigorous mechanistic model.
- Posterior = data-driven.
- Posterior variance = Uncertainty in projections.

Daily COVID-19 deaths in China



Compartmental Prior

- **Mean function =**
  Susceptible, Exposed, Infectious, Recovered (SEIR) model

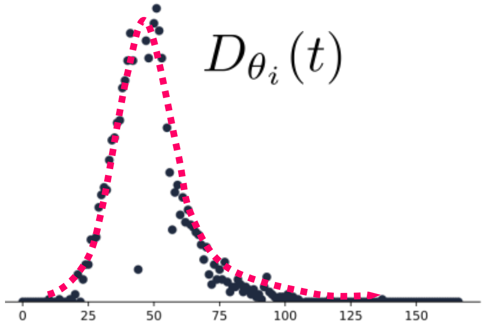
$$f_{L,i} \sim \mathcal{GP}(\boxed{D_{\theta_i}(t)}, K_{\theta_i}(t,t')),$$

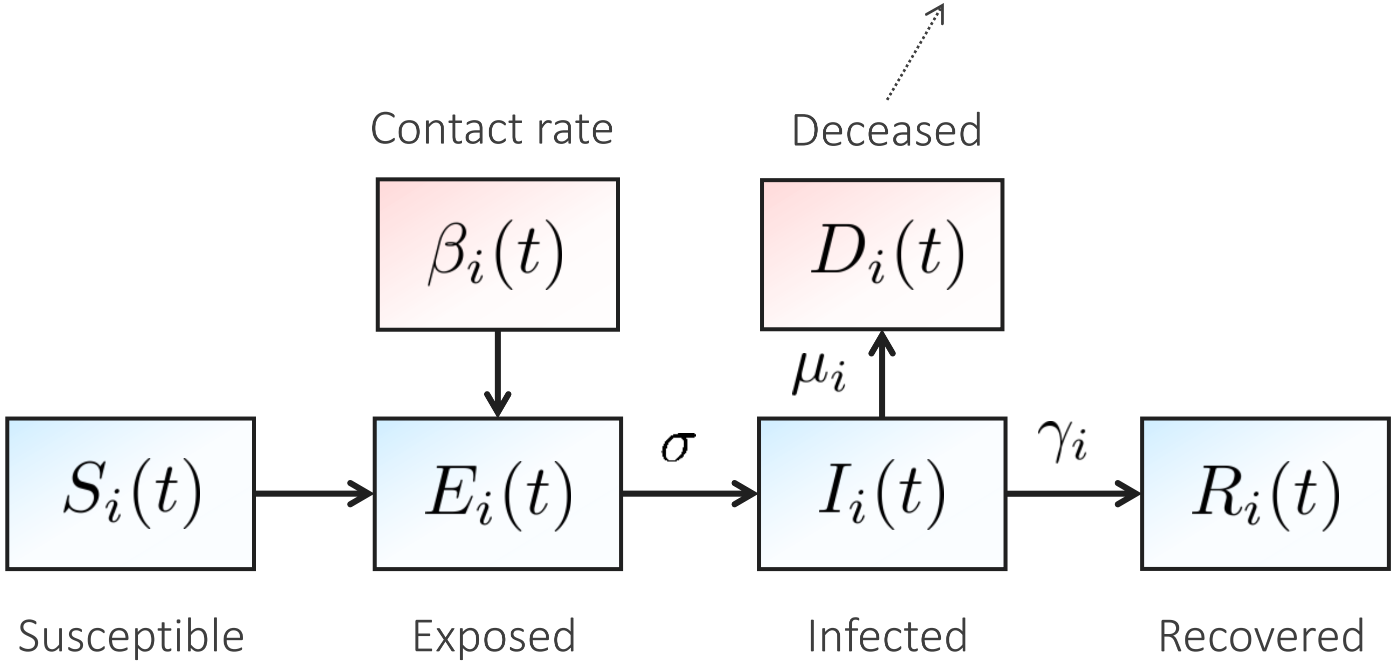
$$\frac{dS_i}{dt} = \mu_i(n_i - S_i) - \frac{\beta_i S_i I_i}{n_i},$$

$$\frac{dE_i}{dt} = \frac{\beta_i S_i I_i}{n_i} - (\mu_i + \sigma)\, E_i,$$

$$\frac{dI_i}{dt} = \sigma\, E_i - (\gamma_i + \mu_i)\, I_i$$

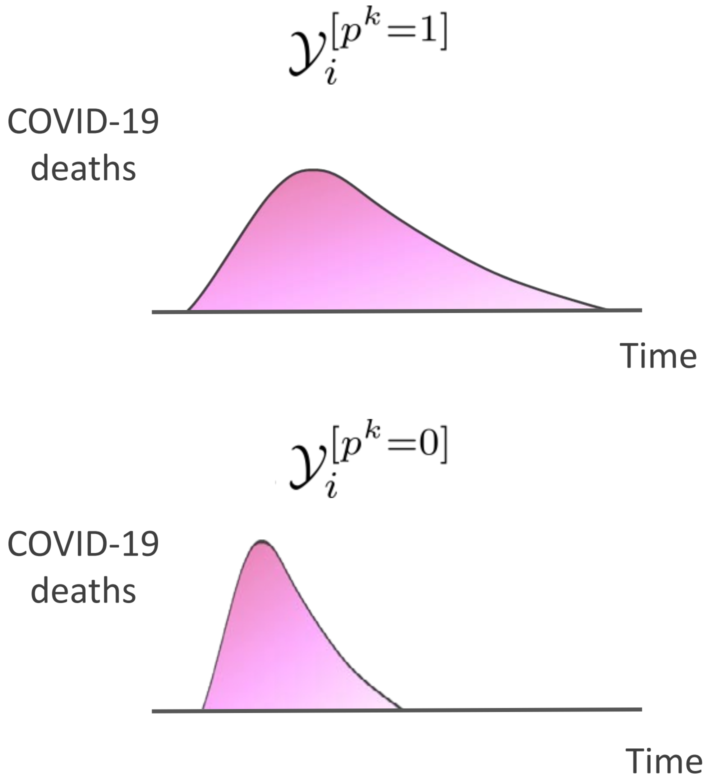
$$\boxed{\frac{dD_i}{dt} = \mu_i\, I_i}$$

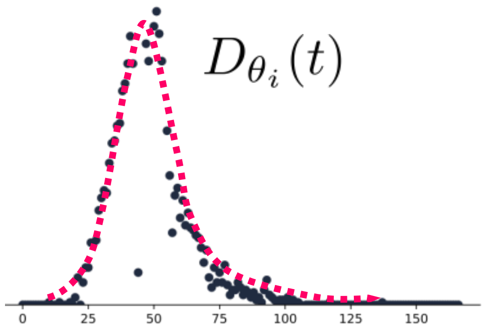


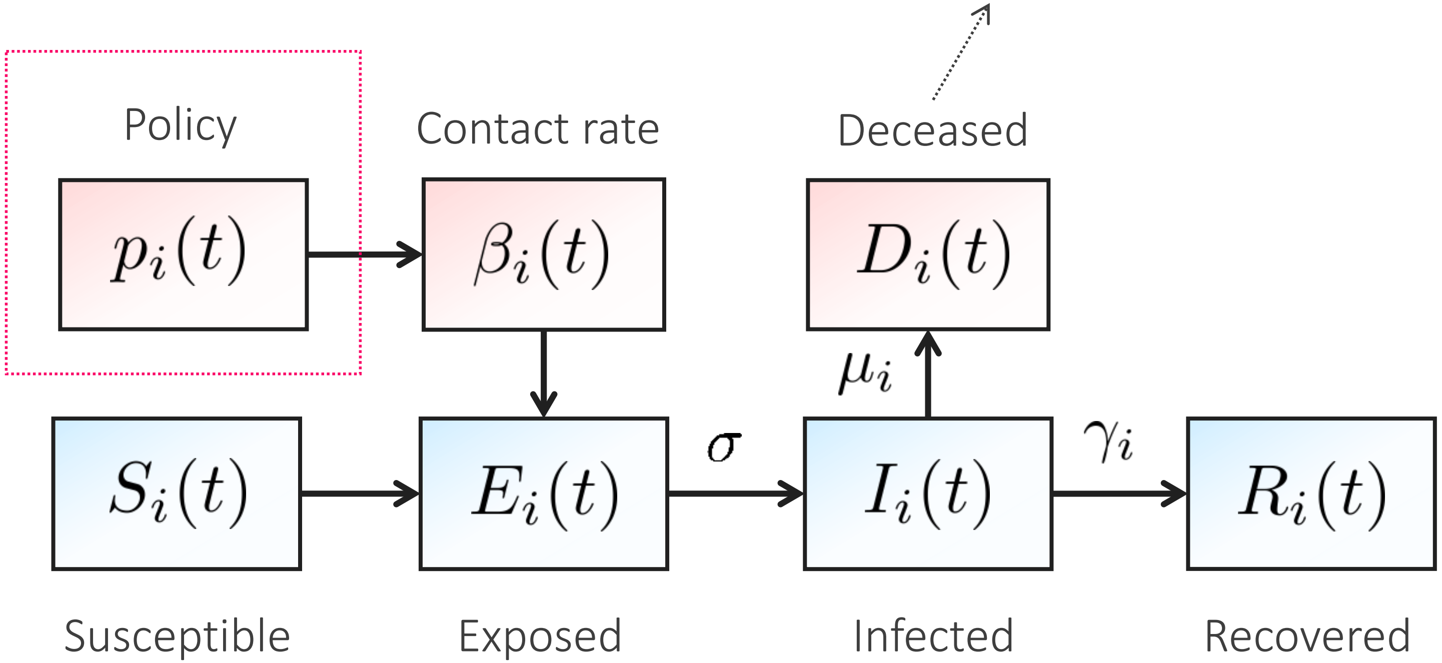


Modeling Potential Outcomes

- Reproduction number $R_{0,i}(t) = \frac{\sigma}{\mu_i+\sigma} \cdot \frac{\beta_i(t)}{\mu_i+\gamma_i}$ Contact rate
- NPI effects on potential outcomes = “Flattening” the curve.







Joint Model for all Countries

- **Hierarchical Bayesian model**
  - Multi-layered model

Upper-layer Gaussian process

$$f_U \sim \mathcal{GP}(m_\alpha(\boldsymbol{X}, p), K_\alpha((\boldsymbol{X}, p), (\boldsymbol{X}', p')))$$

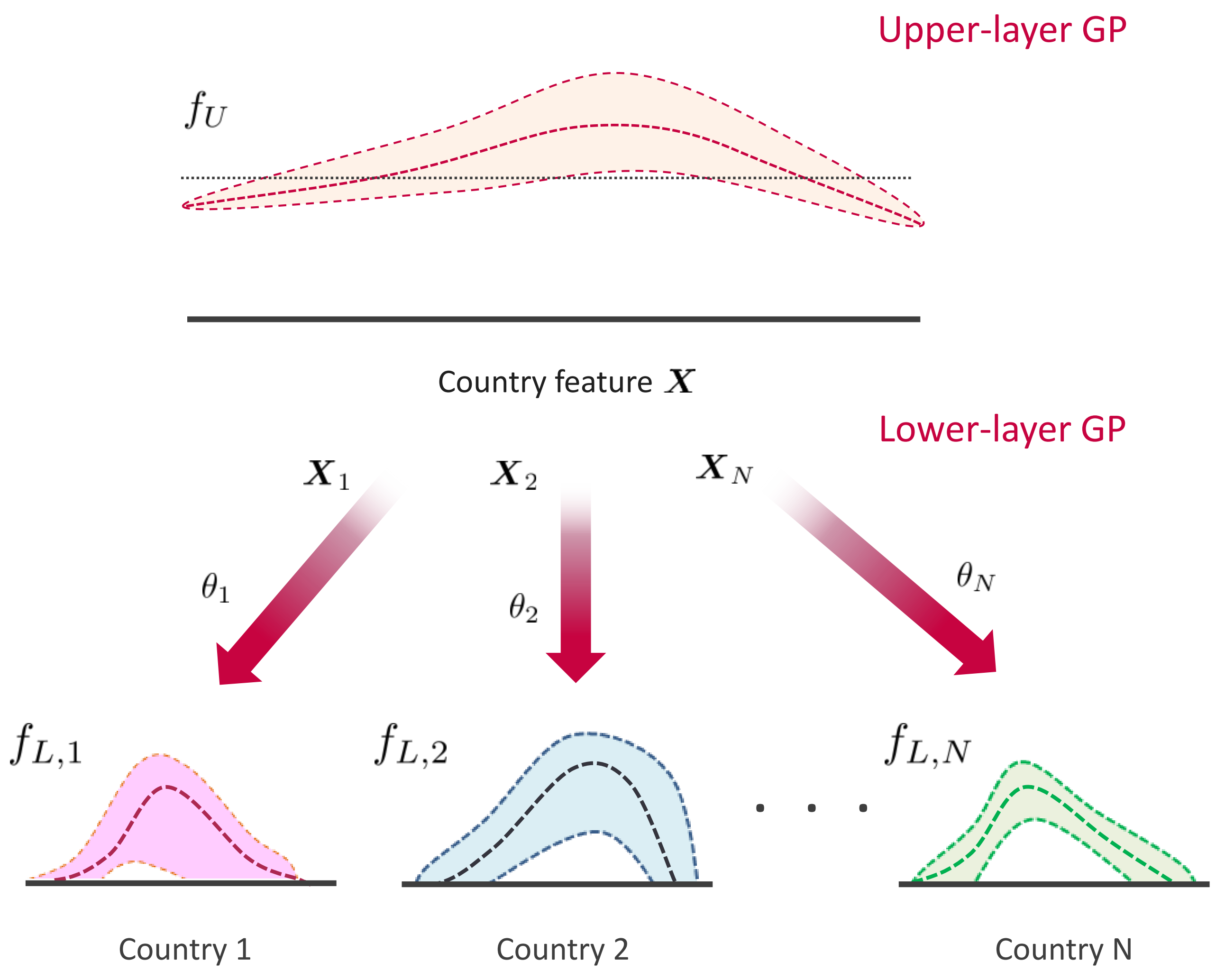
NPI variables and country-level features,
shared among all countries

Lower-layer Gaussian process

$$f_{L,i} \sim \mathcal{GP}(D_{\theta_i}(t), K_{\theta_i}(t, t')),$$

$$\theta_i = v(f_U(\boldsymbol{X}_i, p_i))$$

Specific to a country, models deaths over time



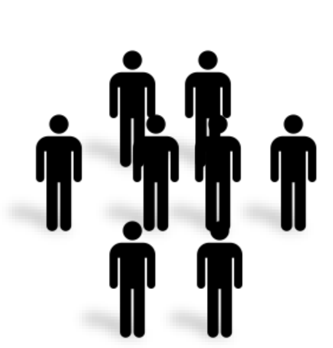
Model: Impact of Non-pharmaceutical Interventions

- Reproduction number $R_{0,i}(t) = \frac{\sigma}{\mu_i + \sigma} \cdot \frac{\beta_i(t)}{\mu_i + \gamma_i}$ ← Baseline R0 depends on country's features

- NPIs effects depend on the country's features…

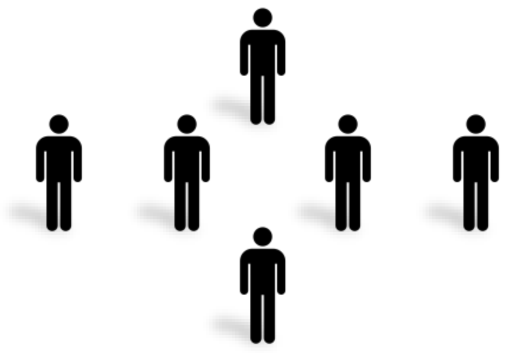
$$\beta_i(t) = v(f_U(\boldsymbol{X}_i, p_i(t))) = 2\,\bar{\beta}\,\mathrm{Sigmoid}(f_U(\boldsymbol{X}_i, p_i(t)))$$ ← Contact rate

Loose NPIs

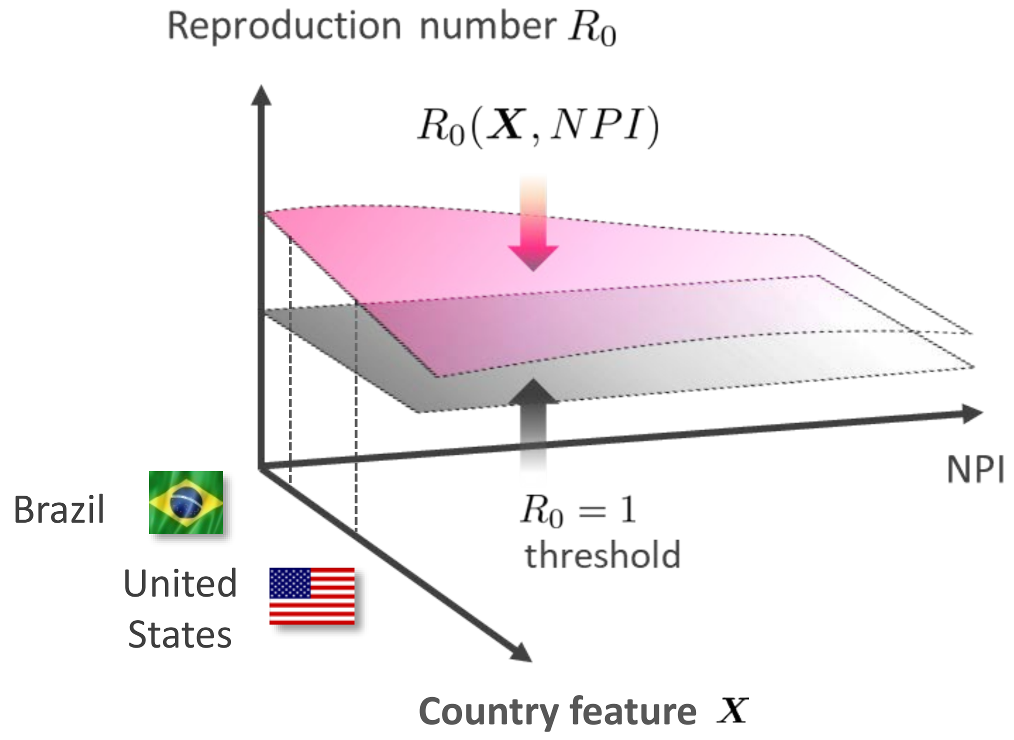


$\beta_i(t)$ ↑↑

Strict NPIs



$\beta_i(t)$ ↓↓



A Dataset for Global COVID-19 Trajectories

- **Country-level COVID-19 data from multiple sources was collated for 170 countries.**
  - COVID-19 cases, deaths, tests, non-pharmaceutical interventions, excess mortality, mobility statistics, weather patterns, country meta-data.

COVID-19 cases, deaths and tests	Non-pharmaceutical interventions	Excess mortality	Mobility statistics	Weather patterns	Country-level Meta-data
European center for disease prevention and control (ECDC)	Oxford Blavatnik school of government ACAPS Gov. measures	Human mortality Database EuroStat The Economist	Google community mobility reports Apple maps	UK Met office	World Bank reports

Non-pharmaceutical interventions (NPI)

- **The Oxford Government Response Tracker provides 13 NPIs on an ordinal scale for each country, reflecting both the intensity of enforcement of the intervention.**

Public events cancellation	Internal movement restrictions	Stay-at-home orders
Public transport closure	International travel restrictions	Testing policy
Gatherings restrictions	Public information campaigns	Workplace closure
Income support	School closure	Contact tracing
	Mandating on mask usage	

Country-level Meta-data

- **We collected 35 economic, social, demographic, environmental and public health indicators from published World Bank reports**

Economic	GDP per capita, GNI per capita, Income share held by lowest 20%
Social and demographic	Population, Life expectancy, Birth rate, Death rate, Infant mortality rate, Land Area, % People with basic hand-washing facilities including soap and water, Smoking prevalence, Prevalence of undernourishment, Prevalence of overweight, Urban population, Population density, Population ages 65 and above, Access to electricity (% of population), UHC service coverage index, Total alcohol consumption per capita, Air transport (passengers carried)
Environmental	Forest Area, PM2.5 air pollution (mean annual exposure in micrograms per cubic meter)
Public health	Immunization for measles, % deaths by communicable diseases, Current health expenditure, Current health expenditure per capita, Diabetes prevalence, Immunization for DPT, Immunization for HepB3, Incidence of HIV, Incidence of malaria, Incidence of tuberculosis, % deaths by CVD/cancer/diabetes/CRD , % deaths due to household and ambient air pollution, % deaths due to unsafe water/unsafe sanitation/lack of hygiene, Physicians (per 1,000 people)

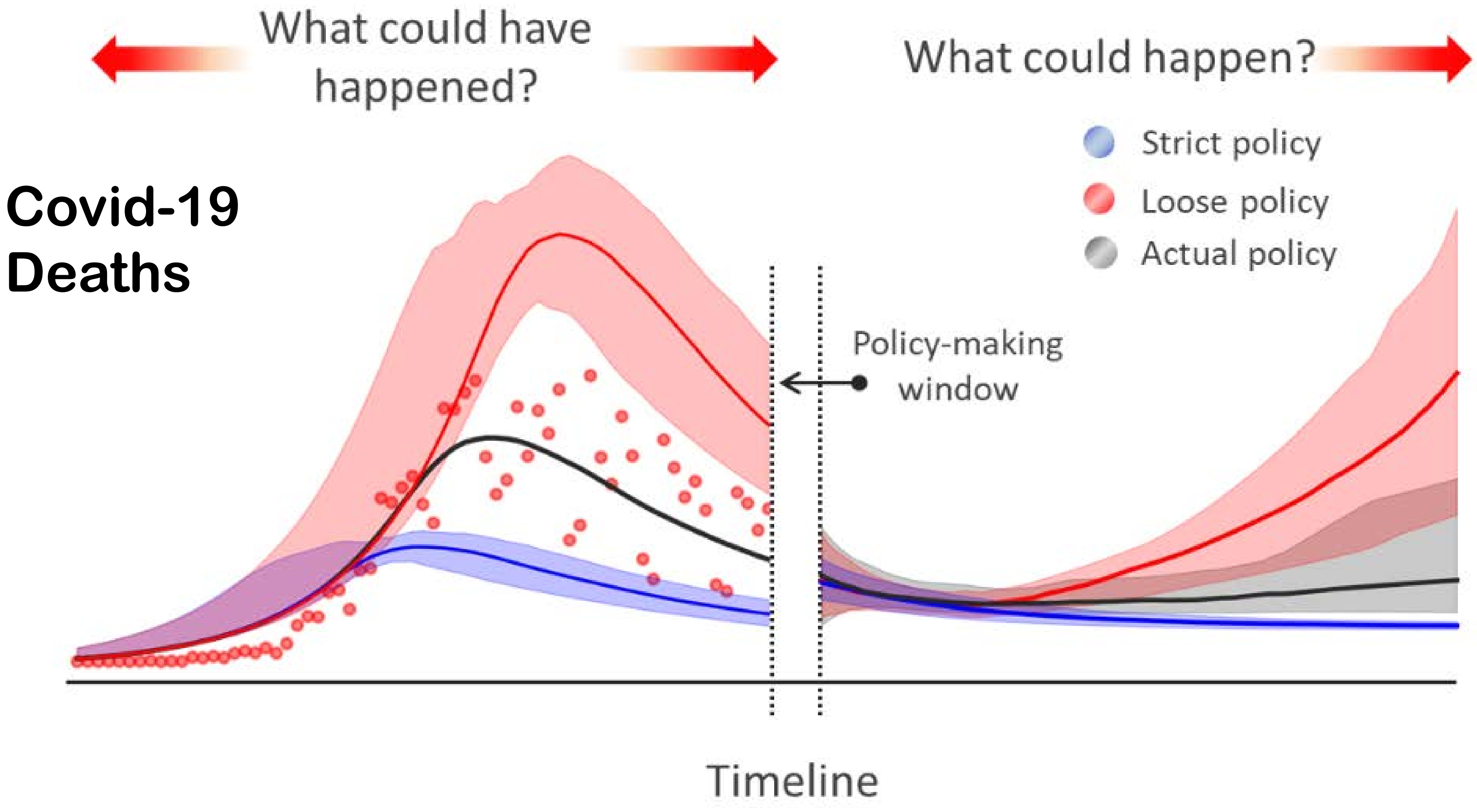
Results: Accuracy of US Projections

Model	March 28 forecasts (before the peak)		April 11 forecasts (During the peak)		April 25 forecasts (After the peak)	
	7 days	14 days	7 days	14 days	7 days	14 days
YYG	___	___	-6,470	-10,528	-662	-1,458
Imperial		___	**-1,757**	___	**14**	___
LANL	___	___	-6,010	**-3,161**	-2,018	-2,989
MIT-DELPHI	___	___	___	___	-2,054	**549**
Gompertz curve	___	___	2,174	4,689	-2,728	-7,062
Vanilla SEIR	2,723	4,822	-11,328	-24,189	-9,696	-21,314
IHME	-1,999	-2,289	-6,134	-10,129	-3,623	-9,999
CDC-ensemble	___	___	-2,739	-8,188	-4,244	-5,091
PIP model (US only)	**-642**	-4,380	-3,182	-8,260	-560	-881
PIP model (global)	-867	**-1,396**	-1,906	-4,518	-439	611

Results: Accuracy of Global Projections

Model	March 28 forecasts				April 11 forecasts			
	IHME	YYG	Imperial	PIP model	IHME	YYG	Imperial	PIP model
United Kingdom	-981	-3,479	-182	**-131**	658	-3,433	**-29**	761
Italy	-1,082	451	1,804	**294**	-2,591	**732**	1,600	901
Germany	-420	244	-417	**104**	-661	628	-288	**179**
Spain	1,104	**167**	-499	317	102	76	712	**50**
Brazil	___	-283	___	**-105**	___	-768	___	**-298**
Sweden	311	107	-102	**24**	693	256	-35	**-7**
France	-501	803	-2,415	**-79**	-1,412	1,601	-1,974	**-485**
Netherlands	512	172	265	**-21**	360	363	228	**-47**
Iran	___	40	___	**9**	___	79	___	**9**
Mexico	___	-82	___	**-56**	___	-518	___	**-315**
Japan	___	___	___	**-3**	___	___	___	**-74**
South Africa	___	___	___	**-8**	___	___	___	**-34**

Counterfactual NPI Scenario Analysis using PIP



Covid-19 at the National Level - Resource Planning

NHS Digital

Search

NHS Digital › News and events

News

Trials begin of machine learning system to help hospitals plan and manage COVID-19 treatment resources developed by NHS Digital and University of Cambridge

Trials have begun of a system that will use machine learning to help predict the upcoming demand for intensive care (ICU) beds and ventilators needed to treat patients with COVID-19 at individual hospitals and across regions in England.

Date: 20 April 2020

Adjutorium: AutoPrognosis for Covid-19

Our goal: Provide evidence that reliably assists the difficult decisions clinicians and managers have to make to save lives

Use depersonalized data

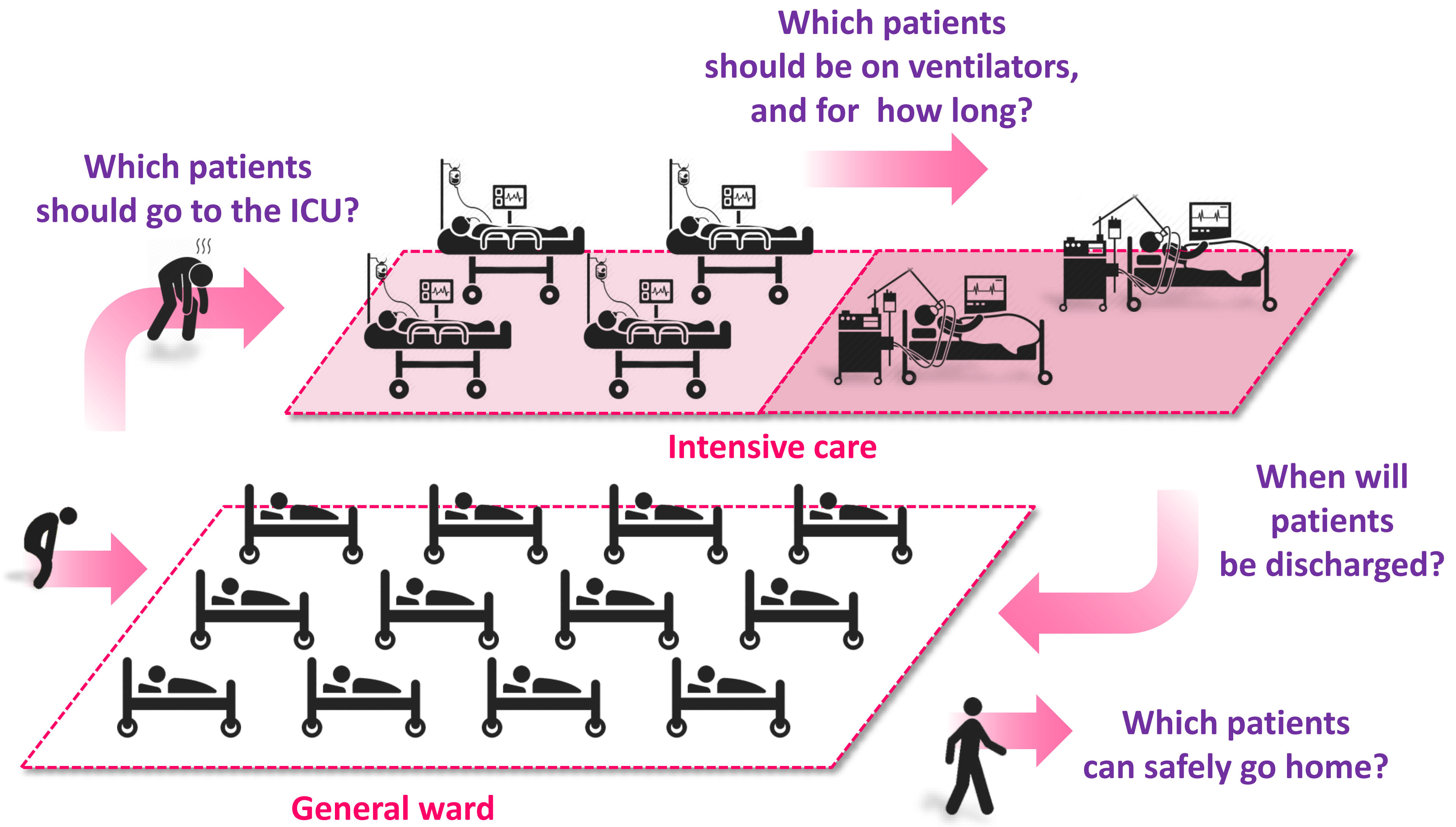
- demographic info, comorbidities, hospitalization details, outcomes

to:

- **forecast personalized risk for each patient**
- **forecast personalized patient benefit from resources**
- **forecast which treatments are needed by each patient and when**
- **forecast which resources are needed by each patient and when**
- **forecast future resource requirements at the hospital level**

www.vanderschaar-lab.com/covid-19/

Decisions that healthcare professionals need to make



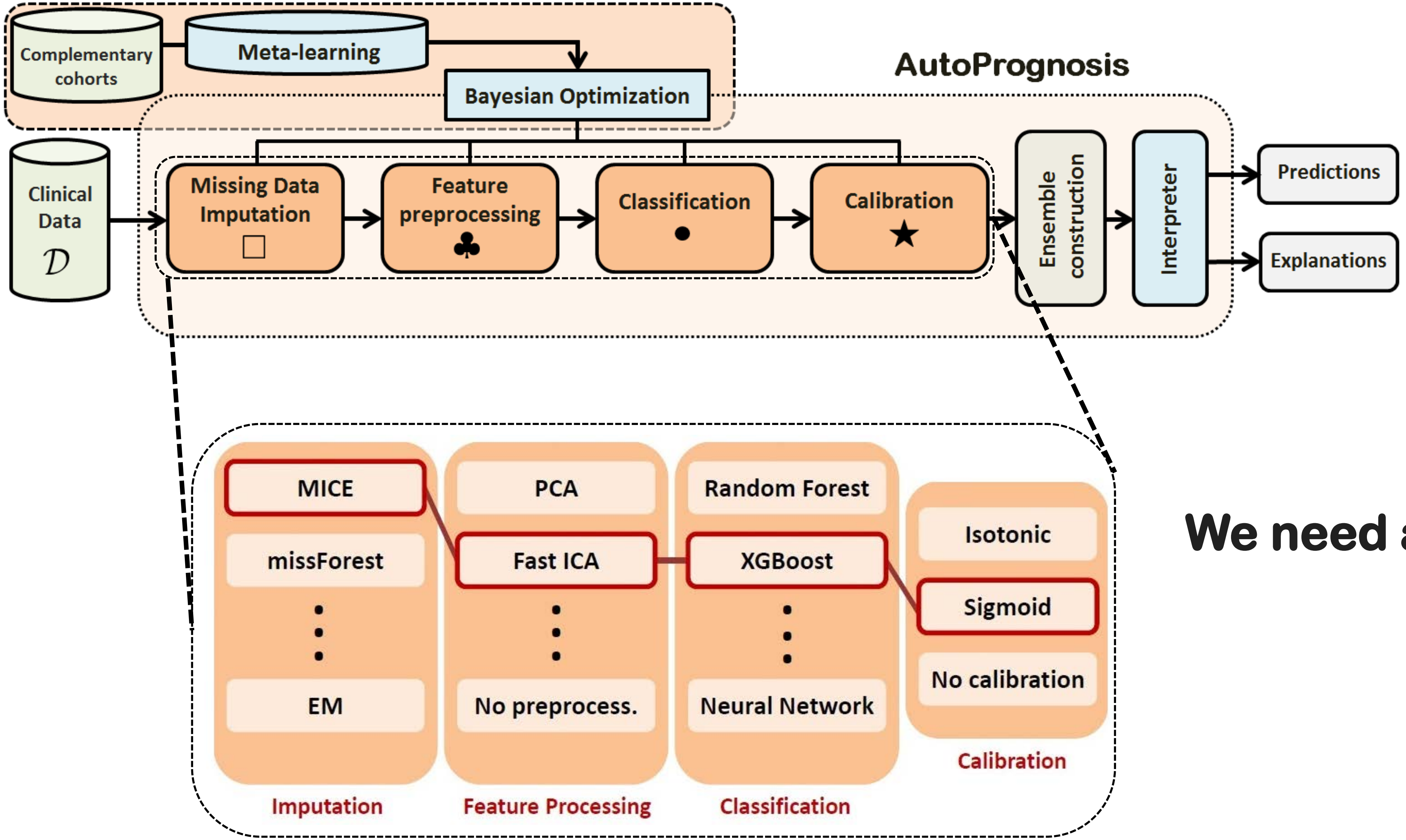
How?

AutoML

Many diseases, many variables, various needs! All is changing! → Can't craft a model for each disease! → Make Machine Learning DO the Crafting

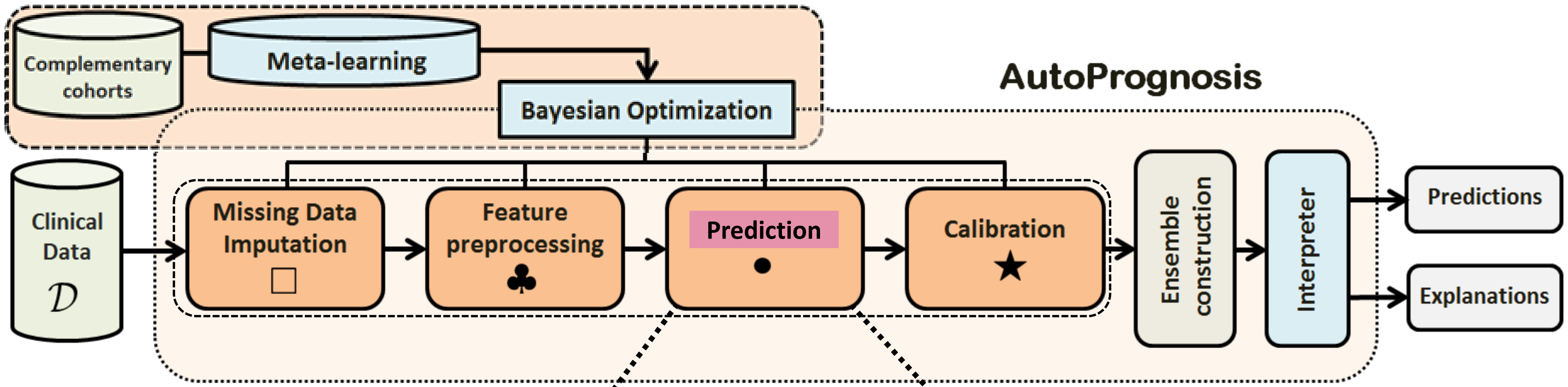
- **Previous AutoML? Auto-WEKA and Auto-Sklearn**
  - Limited performance gains
  - Meta-learning
  - Simplistic handling of missing data
  - Do not capture uncertainty
  - Limited to classification problems **(survival, competing risks, time-series etc.)**

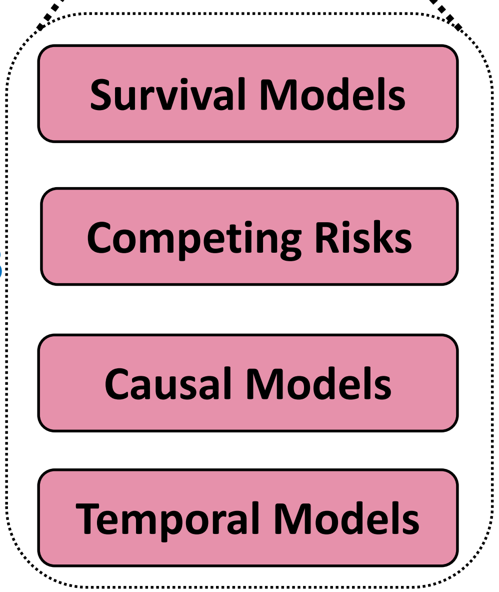
AutoPrognosis [Alaa & vdS, ICML 2018]:
A tool for crafting Clinical Scores



We need an entire pipeline!

Automated ML for clinical analytics





Lee, Alaa, Zame, vdS, AISTATS 2019

Alaa, vdS, NIPS 2017
Lee, Yoon, vdS, TBME 2019

Alaa, vdS, ICML 2019

Zhang, Jarett, vdS, AISTATS 2020

AutoPrognosis
ICML 2018
Cystic Fibrosis - Scientific Reports - 2018
UK Biobank - Plos One 2018
Breast Cancer – 2019
Covid - 2020

Clairvoyance [Jarrett, Yoon, Bica, Ercole, vdS, 2020]

AutoPrognosis at work
Covid-19: Who needs ventilation?

- **AUC-ROC accuracy for predicting whether a patient will need ventilation based on info available at hospital admission**

Model	AUC-ROC
AP: all features	**0.771 ± 0.002**
AP: age + specific comorbidities	0.761 ± 0.001
AP: age + no. of comorbidities	0.720 ± 0.003
Cox Regression: all features	0.690 ± 0.002
Charlson Comorbidity Index	0.618 ± 0.002

Demonstrator

Covid-19 at the Patient Level

djutorium

LEARN MORE | GO BACK TO WEBSITE | SEND FEEDBACK

Patient Information

Age at diagnosis

65

Sex

M F

Comorbidities

× Diabetes × Hypertension ×

Previous Treatments

× Immunosuppression Therapies × Chemotherapy ×

Reason for Admission

COVID-19 Flu RSV

Pregnancy

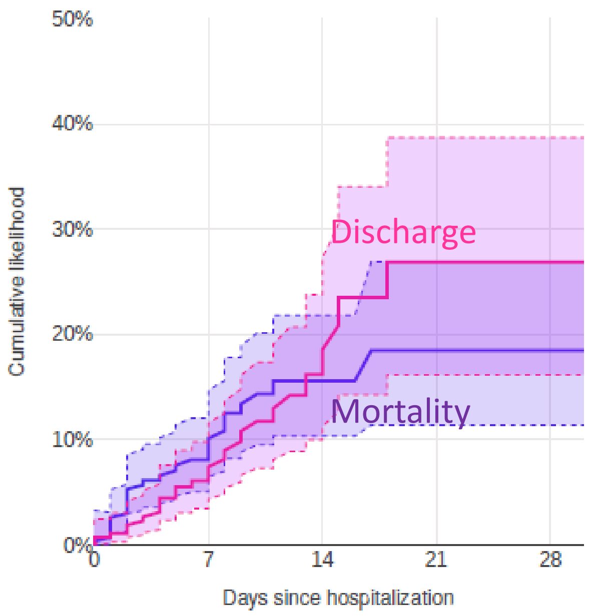
No Yes

Obesity

No Yes

Prediction Results

Adjutorium is a system designed to assist decision-making by offering predictions based on existing data. It can support and inform healthcare professionals, but is not intended to replace their own decision-making processes.



Ventilator benefit

This table shows the expected survival chance for 14 days after diagnosis based on the patient information.

Treatment	Additional Benefit	Overall Survival %
Non-invasive Ventilator	20%	84%
Invasive Ventilator	23%	87%

Build the ecosystem!

How can we turn ML models into actionable intelligence?

We need

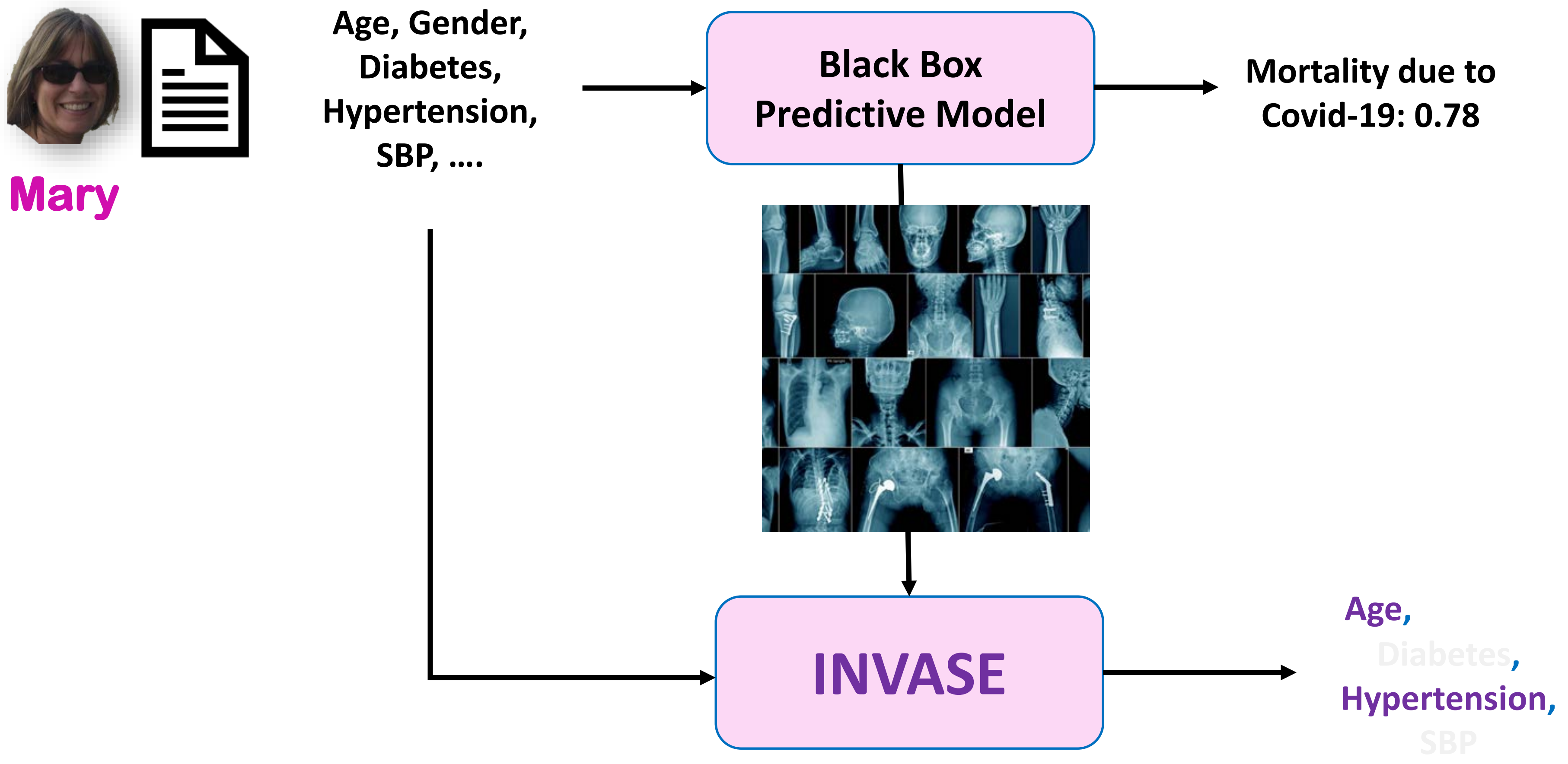
- ***Risk score understanding***: Users need to understand, quantify and manage risk
- ***Transparency***: Users need to comprehend how the model makes predictions
- ***Avoid implicit bias***: Users need to be able to check whether the model does not learn biases
- ***Discovery***: Users need to distil insights and new knowledge from the learned model
- ***Know what we do not know:*** Users need to have a quantification of the model's prediction uncertainty

Many kinds of interpretations exist…

- Current methods are tailored to ***one*** type of interpretation
  Uncovering one of the following
  - What features are globally important, i.e. for the entire population?
  - What features are locally important, i.e. for this patient?
  - Feature interaction
  - Model non-linearity

- Desiderata
  - Model-independent: general, not tailored to specific models
  - Post-hoc: should not interfere with model training which may introduce bias and compromise accuracy

Which features of an *individual* are relevant for a prediction?



[Yoon, Jordon, vdS, ICLR 2019]

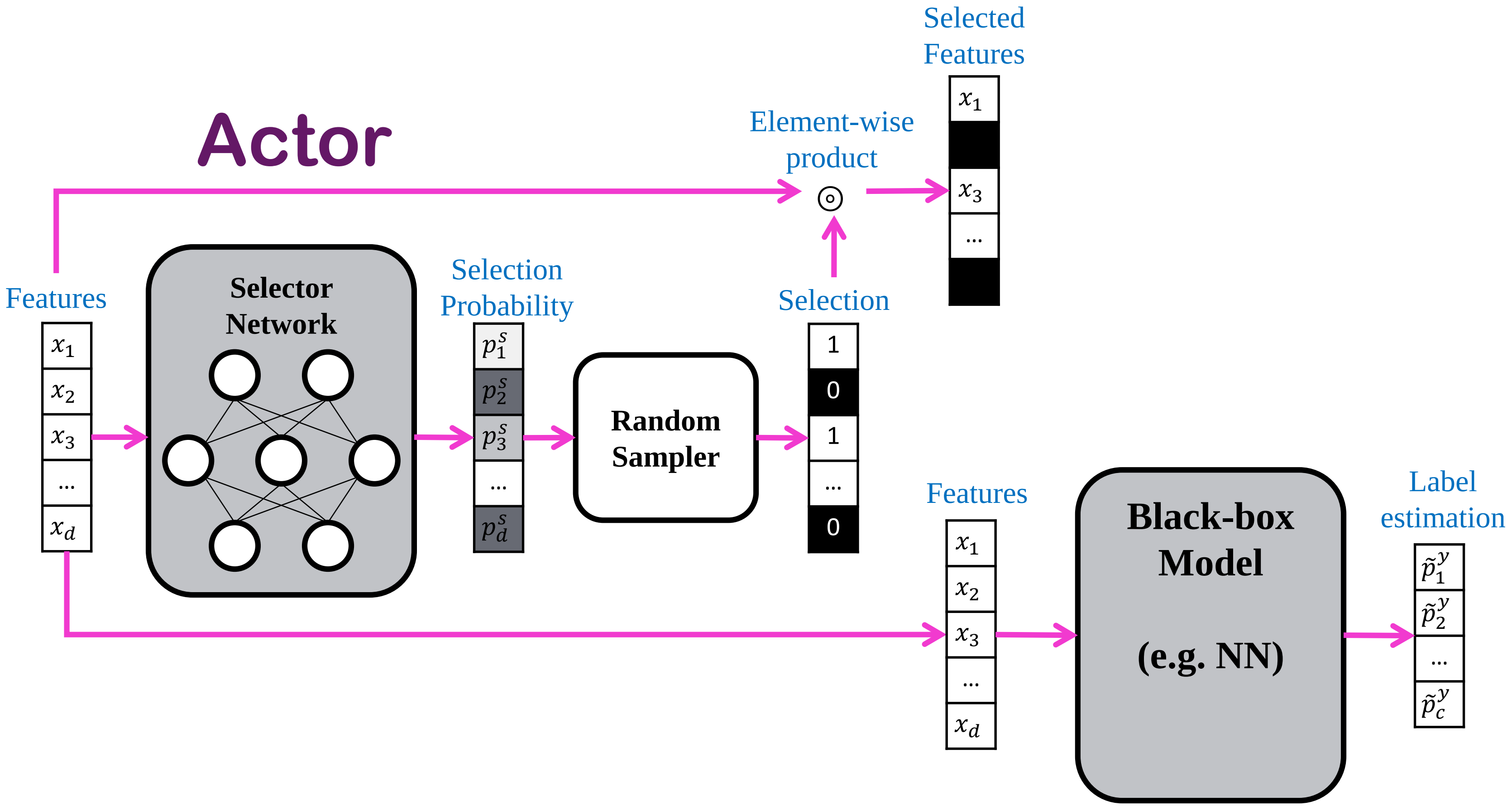
INVASE [Yoon, Jordon, vdS, ICLR 2019]

- How can we learn individualized feature importance?
- Key idea: **Use Reinforcement Learning (RL)**
  - Make observations
  - Select “actions” on the basis of these observations
  - Determine “rewards” for these actions

  - Ultimately learn a policy which selects the best actions
    - i.e. actions that maximize rewards given observations

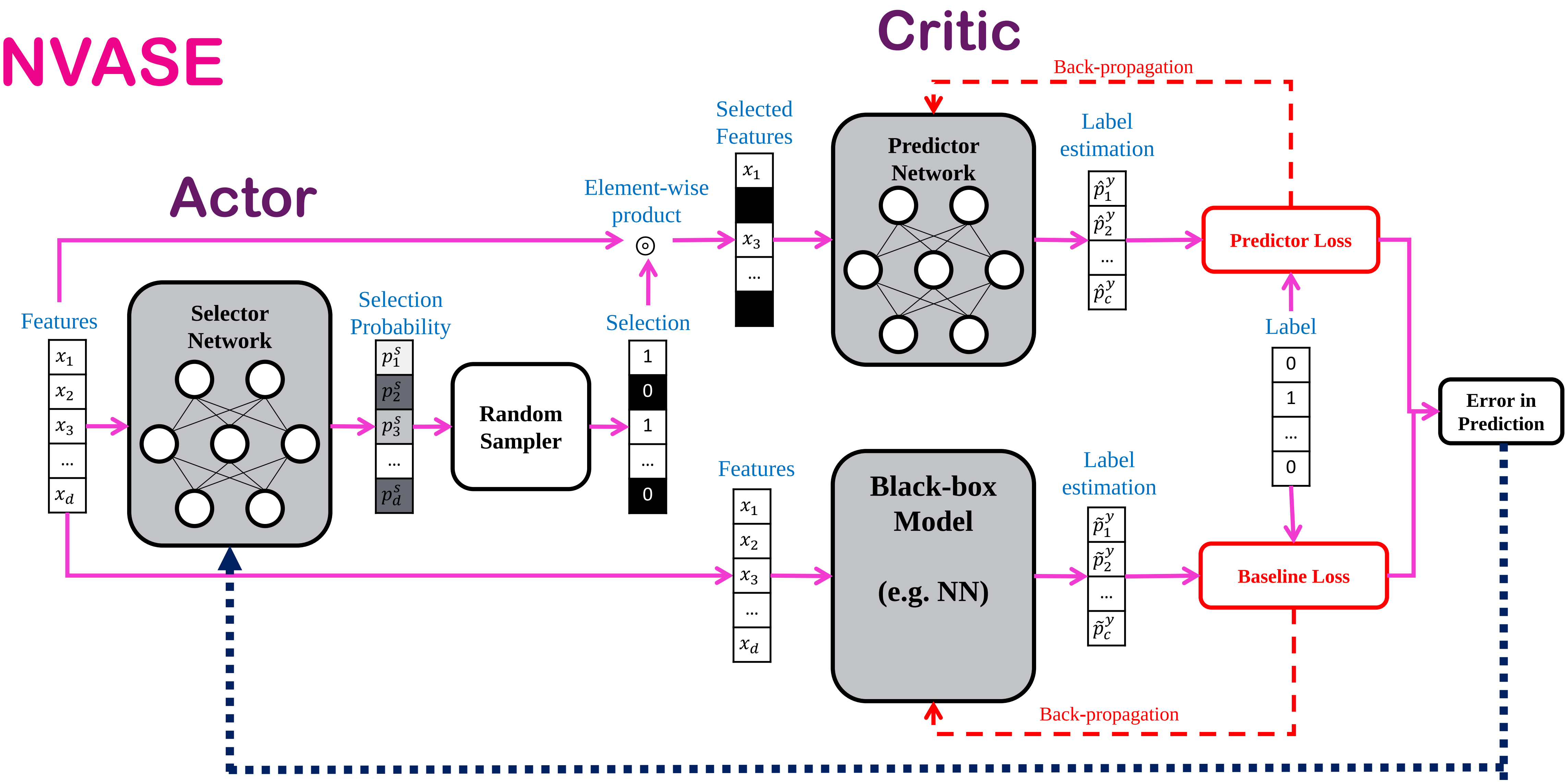
- We use the Actor-Critic approach to RL

INVASE



- **Selector network (actor)** takes instances and outputs vector of selection probabilities.

INVASE



- **Predictor network (critic)** **receives the selected features, makes predictions and provides feedback to the actor.**

INVASE: Instance-wise feature importance for prediction

Find **selector function S** that minimizes **features selected $S(x)$** while satisfying equality constraints on the **conditional distribution of the predictions.**

- **Objective: minimize** $S(\boldsymbol{x})$
- **Constraints:**

$$\left(Y \middle| \boldsymbol{X}^{(S(\boldsymbol{x}))} = \boldsymbol{x}^{(S(\boldsymbol{x}))}\right) \stackrel{\mathrm{d}}{=} (Y|\boldsymbol{X} = \boldsymbol{x})$$

- x: Features for *a given realization*
- $S: \mathcal{X} \to \{0,1\}^d$: Selector function, $S(\boldsymbol{x})$: Selected features
- Y: Predictions made by black-box model

INVASE: Instance-wise feature importance for prediction

Find **selector function S** that minimizes **features selected $S(x)$** while satisfying equality constraints on the **conditional distribution of the predictions.**

- **Objective: minimize** $S(\boldsymbol{x})$
- **Constraints:**

$$\left(Y \middle| \boldsymbol{X}^{(S(\boldsymbol{x}))} = \boldsymbol{x}^{(S(\boldsymbol{x}))}\right) \stackrel{\mathrm{d}}{=} (Y|\boldsymbol{X} = \boldsymbol{x})$$

- **Lagrangian optimization:**

$$\mathcal{L}(S) = \mathbb{E}[KL\left(Y \middle| \boldsymbol{X}^{(S(x))} = x^{(S(x))}\right)||(Y|\boldsymbol{X} = x) + \lambda||S(x)||]$$

- **Challenging problem:**
  - Output space of the selector function is large - its size increases exponentially with the dimension of the feature space!
  - We do not have access to the densities required – need to be learned

[Yoon, Jordon, vdS, ICLR 2019]

Limitations of past methods for model interpretability

Method	Feature importance	Individualized feature importance	Model-independent	Identifying the set of relevant features for each instance
LASSO [Tibshirani, 1996]	✓		✓	
Knock-off [Candes et al, 2016]	✓		✓	
L2X [Chen et al, 2018]	✓	✓	✓	
LIME [Ribeiro et al, 2016]	✓	✓	✓	
SHAPE [Lundberg et al, 2017]	✓	✓	✓	
DeepLIFT [Shrikumar et al, 2017]	✓	✓		
Saliency [Simonyan et al, 2013]	✓	✓		
TreeSHAP [Lundberg et al, 2018]	✓	✓		
Pixel-wise [Batch et al, 2015]	✓	✓		
INVASE [Yoon, Jordon and van der Schaar, 2019]	✓	✓	✓	✓

INVASE discovers *the number* of relevant features for each instance

Are we done?

- **NO!**
- **Need to ALSO understand what the model discovered: feature/statistical interactions, model non-linearity, etc.**

Method	Feature importance	Individualized feature importance	Feature interaction	Model-independent	Post-hoc
LIME [Ribeiro et al, 2016]	✓			✓	✓
SHAPE [Lundberg et al, 2017]	✓		✓	✓	
DeepLIFT [Shrikumar et al, 2017]	✓				
INVASE [Yoon, Jordon and van der Schaar, 2019]	✓	✓		✓	✓
L2X [Chen et al, 2018]	✓	✓			
GAM [Lou et al, 2013]	✓		✓		
NIT [Tsang et al, 2018]	✓		✓		

What we are aiming for?

- Understand what the model discovered:
  feature importance, instance-wise feature importance, feature/statistical interactions, model non-linearity, etc.
- Produce a transparent risk equation describing the model for approval in practice guidelines
- Enable model explainability, not only interpretability

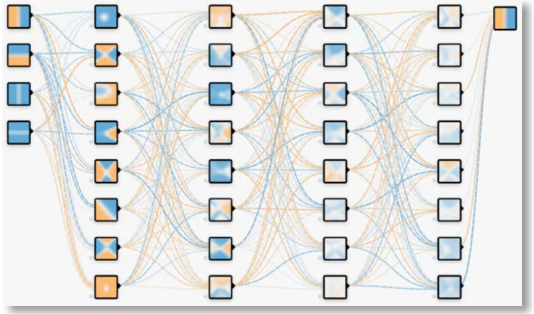
Can we have it all??

YES!

Demystifying Black-box Models with Symbolic Metamodels

[A. Alaa & vdS, NeurIPS 2019]

Black-box ML model



$f(\mathbf{x})$

Symbolic Metamodeling

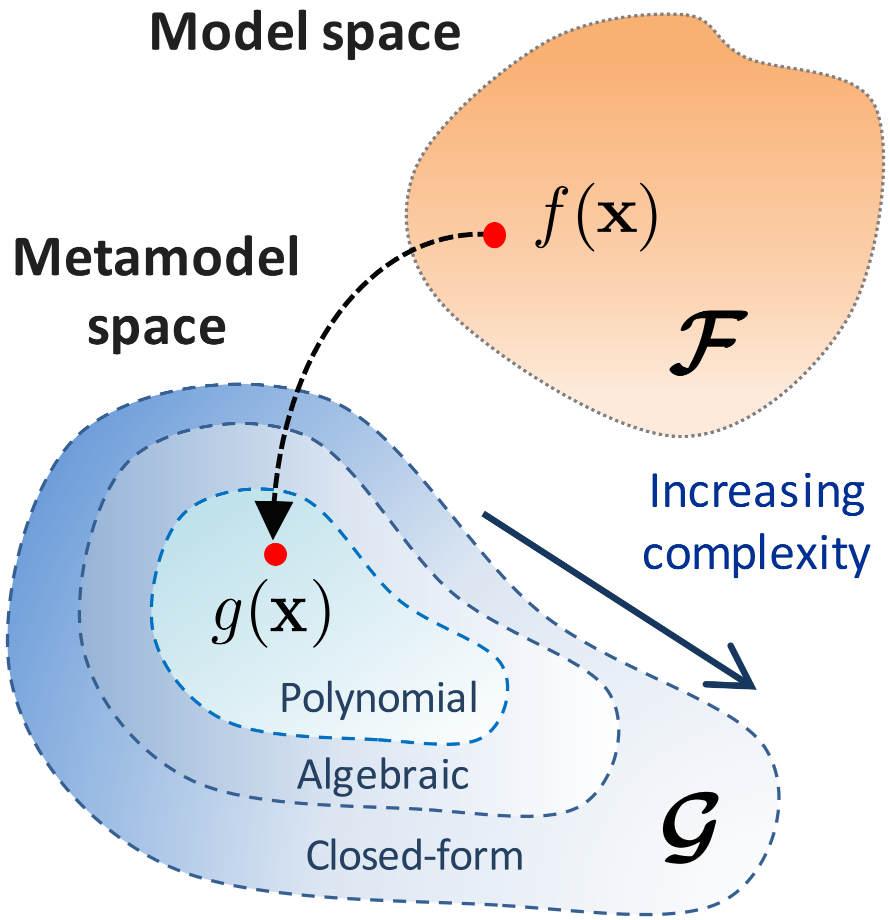
White-box (transparent) model

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

$g(\mathbf{x})$

- **Metamodel = a model of a model.**
- **A symbolic metamodel outputs a transparent function *describing* the predictions of the black box model**
- **Metamodeling needs only query access to trained black-box model.**

Symbolic Metamodeling



Black-box ML model

Model space (uninterpretable)

White-box model

Metamodel space

Metamodel space can be *chosen* by the user!

How are we going to achieve this?

- **Kolmogorov-Arnold Theorem** [Kolmogorov et al, 1961]

  Every multivariate continuous function can be written as a finite composition of univariate continuous functions

$$g(\mathbf{x}) = \sum_{q=0}^{r} g_q \left(\sum_{p=1}^{n} g_{q,p}(x_p) \right)$$

- **The symbolic metamodeling problem**

Metamodel representation

$$g(\mathbf{x}; \theta) = \sum_{q=0}^{2n} G \left(\sum_{p=1}^{n} G(x_p; \theta_{q,p}); \theta_q \right)$$

Metamodel optimization

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

What basic functions?

- **Meijer G-functions** [C. S. Meijer, 1936]

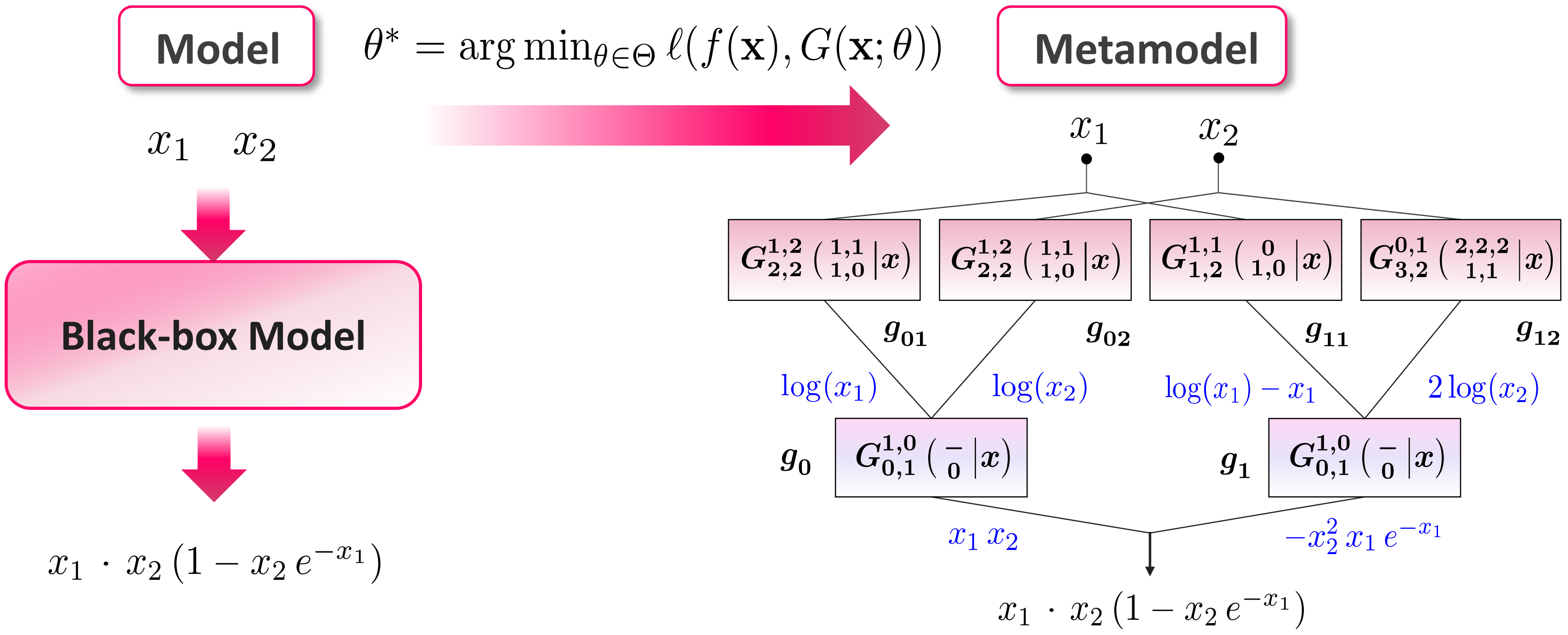
$$G_{p,q}^{m,n}\left(\begin{matrix} a_1,\ldots,a_p \\ b_1,\ldots,b_q \end{matrix} \middle| x\right) = \frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^{m}\Gamma(b_j-s)\prod_{j=1}^{n}\Gamma(1-a_j+s)}{\prod_{j=m+1}^{q}\Gamma(1-b_j+s)\prod_{j=n+1}^{p}\Gamma(a_j-s)}\, x^s\, ds$$

- **Very general class of functions**
- **Parameter selection yields many familiar functions**

G-function	Equivalent function	G-function	Equivalent function
$G_{0,1}^{1,0}\left(\begin{matrix} - \\ 0 \end{matrix} \middle\| -x\right)$	e^x	$G_{2,2}^{1,2}\left(\begin{matrix} \frac{1}{2},1 \\ \frac{1}{2},0 \end{matrix} \middle\| x^2\right)$	$2\arctan(x)$
$G_{2,2}^{1,2}\left(\begin{matrix} 1,1 \\ 1,0 \end{matrix} \middle\| x\right)$	$\log(1+x)$	$G_{1,2}^{2,0}\left(\begin{matrix} 1 \\ \alpha,0 \end{matrix} \middle\| x\right)$	$\Gamma(\alpha,x)$
$G_{0,2}^{1,0}\left(\begin{matrix} - \\ 0,\frac{1}{2} \end{matrix} \middle\| \frac{x^2}{4}\right)$	$\frac{1}{\sqrt{\pi}}\cos(x)$	$G_{1,2}^{2,0}\left(\begin{matrix} 1 \\ 0,\frac{1}{2} \end{matrix} \middle\| x^2\right)$	$\sqrt{\pi}\,\mathrm{erfc}(x)$
$G_{0,2}^{1,0}\left(\begin{matrix} - \\ \frac{1}{2},0 \end{matrix} \middle\| \frac{x^2}{4}\right)$	$\frac{1}{\sqrt{\pi}}\sin(x)$	$G_{0,2}^{1,0}\left(\begin{matrix} - \\ \frac{a}{2},\frac{-a}{2} \end{matrix} \middle\| \frac{x^2}{4}\right)$	$J_a(x)$

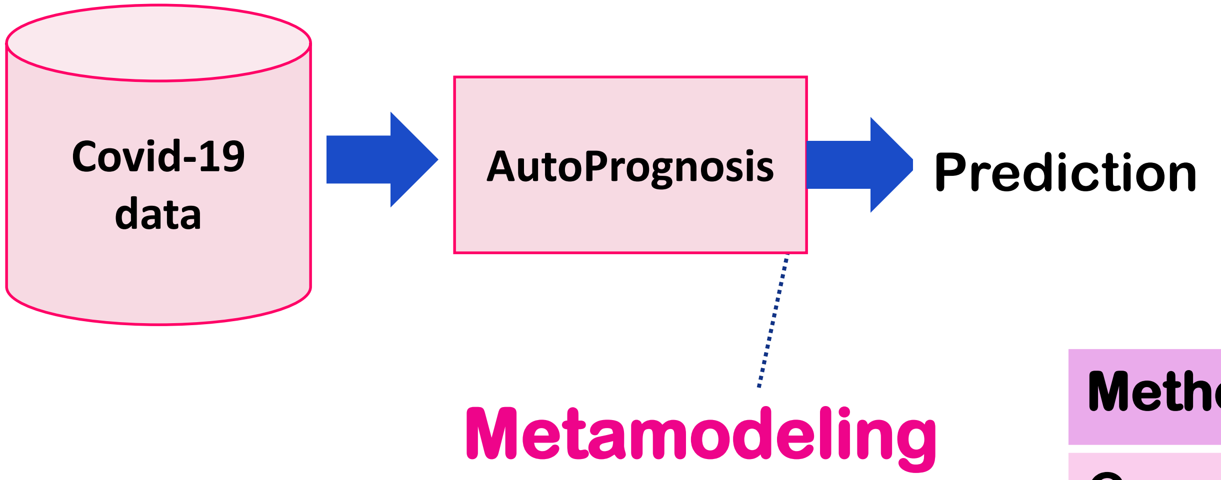
Building a symbolic metamodel

- **Metamodel construction is "analogous" to a 2-layer neural network**



Parameters of a Meijer-G function can be learned by gradient descent!
This can be done very fast!

Interpretability using symbolic metamodeling <u>in practice</u>



Method	AUC-ROC
Cox	0.690 ± 0.002
AutoPrognosis	0.771 ± 0.002

Example: Use Metamodels for Individual-level feature importance

Age
Gender
diabetes
BMI

→ ML Model → Mortaility risk

Original Metamodel

$$g(\mathbf{x}) \quad \begin{array}{l} \alpha_0\,\mathrm{Age} + \alpha_1\,\mathrm{BMI}^2 + \alpha_2\mathrm{Age}\cdot\mathrm{BMI} + \alpha_3\mathrm{Age}\cdot\mathrm{Gender} \\ \alpha_4\mathrm{Gender}\cdot(1 + \alpha_5\,\mathrm{Diabetes}) + \alpha_6\log(\mathrm{Age}\cdot\mathrm{Diabetes} + 1) \end{array}$$

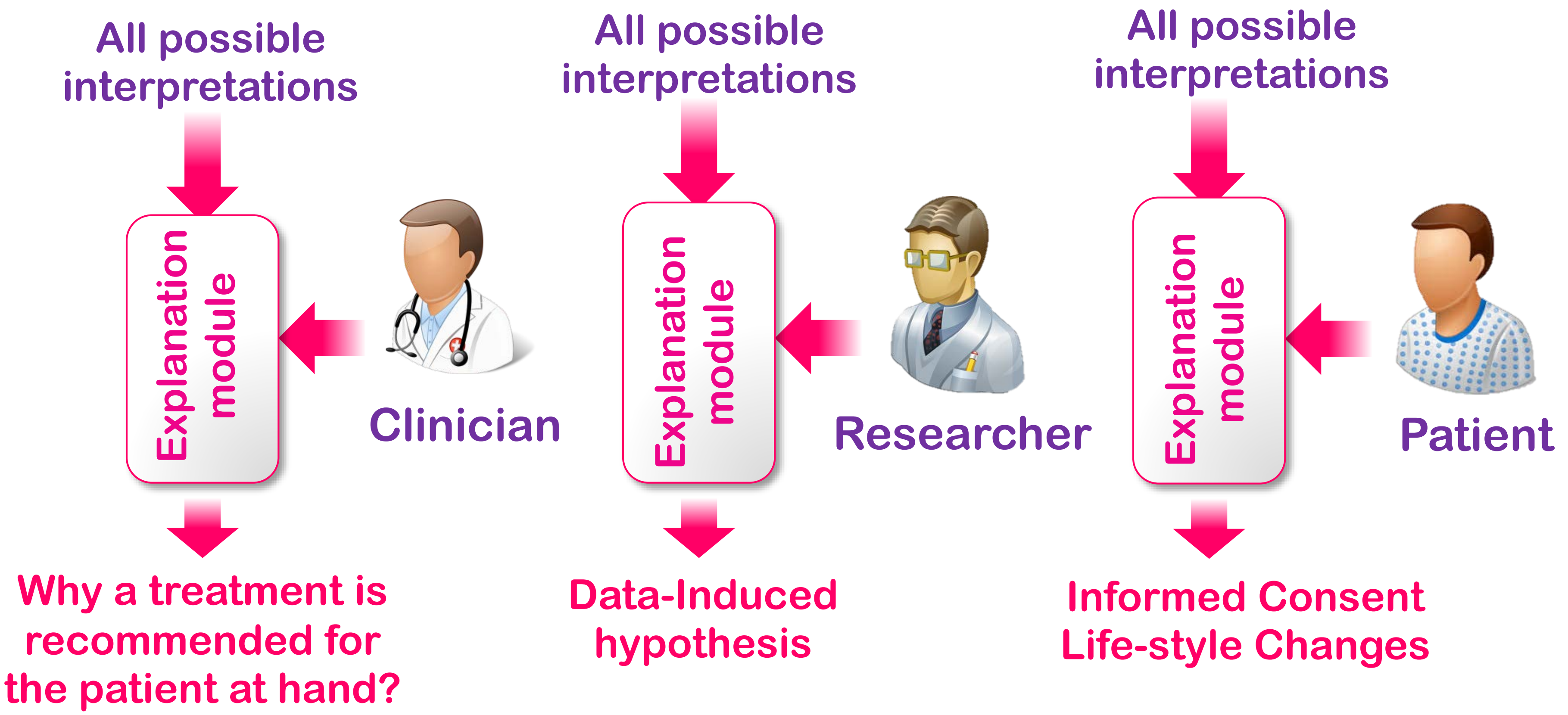
Individual-level feature importance

$$\frac{\partial\, g(\boldsymbol{x})}{\partial\,\mathrm{Age}} = \alpha_0 + \alpha_2\,\mathrm{BMI} + \alpha_3\,\mathrm{Gender} + \frac{\alpha_6\,\mathrm{Diabetes}}{\mathrm{Age}{+}1}$$

Beyond current feature importance

Explainability =User-dependent Interpretability

- Different users seek different forms of “understanding”…



Metamodels: How to use them?

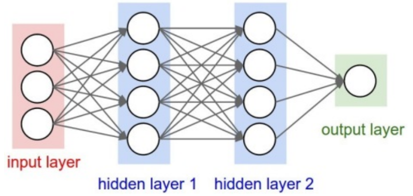
- Different forms of interpretations can be extracted from a Metamodel’s forward and backward views!

Forward use

Input= features
Output=risk

- Treatment justification
- Hypothesis induction
- Variable interactions
- Variable importance

Black-box model



Symbolic Metamodel

Metamodel

g(Age, ER, HER2, Tumor size, Nodes)

Backward use

Input= reduced risk
Output=features

- Modifiable variables
- Dosage recommendation
- Policy design

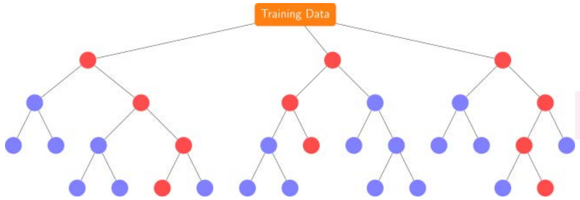
How can researchers use Metamodels?

- Regardless of the model f(x), g(x) is always a symbolic expression
- Unified format for many different types of black-box models: identify common discoveries by comparing their Metamodels

Neural network

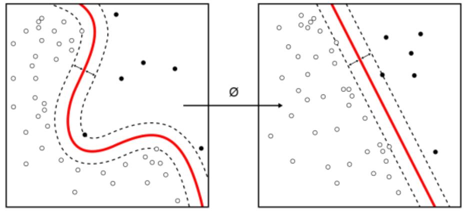
$$\alpha_1 X_1 + \alpha_2 X_2^2 + \alpha_3 X_1 \, X_2$$
$$\alpha_4 X_3^3 + \alpha_5 \log(X_4)$$

Random forest



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

SVM

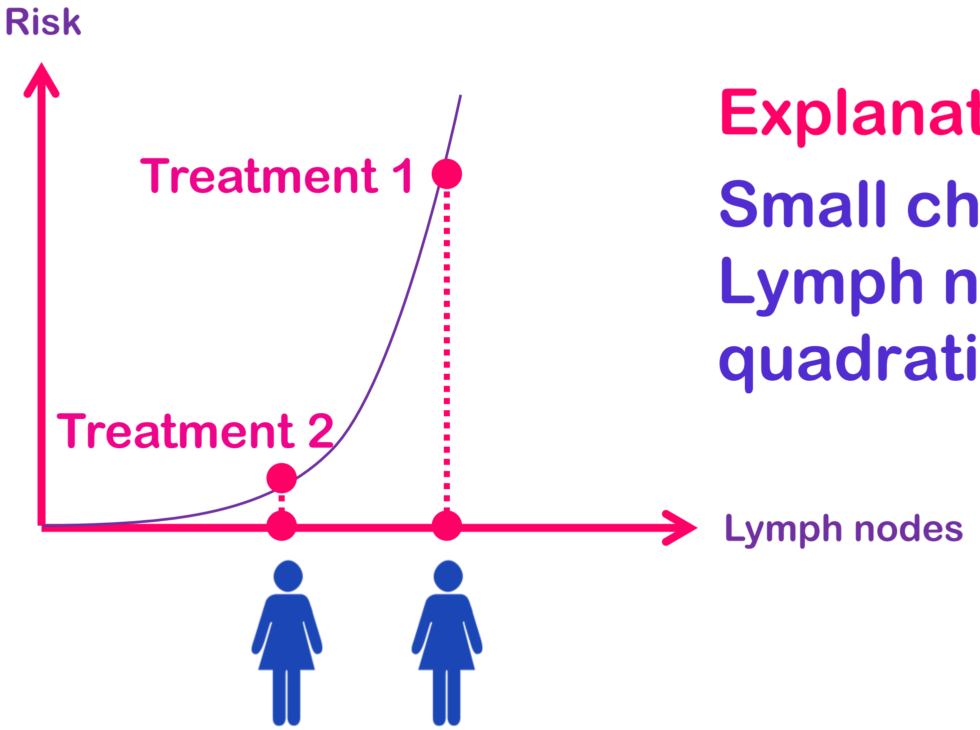


$$\alpha_1 X_1 + \alpha_2 X_2^2 + \alpha_3 X_1 \, X_2$$
$$\alpha_4 X_3^3 + \alpha_5 \log(X_4)$$

Report common discoveries

How can clinicians use Metamodels?

- Understand why how predictions or treatment recommendations are being made by the ML-model
- Example: Two patients with apparently similar features get different treatment recommendations!



Explanation:

Small changes in number of Lymph node causes a quadratic increase in risk

How can patients use Metamodels?

- Patients can be informed how to alter behavior to lower risk.

Can be set through the inverse Metamodel equation

$$\text{BMI Reduction} = g^{-1}(\boldsymbol{Family\ history}, \boldsymbol{Genetics}, \boldsymbol{Diabetes} |\ \boldsymbol{Risk} = \boldsymbol{X}\ \%)$$

In addition to interpretability & explainability…. trustworthiness is key

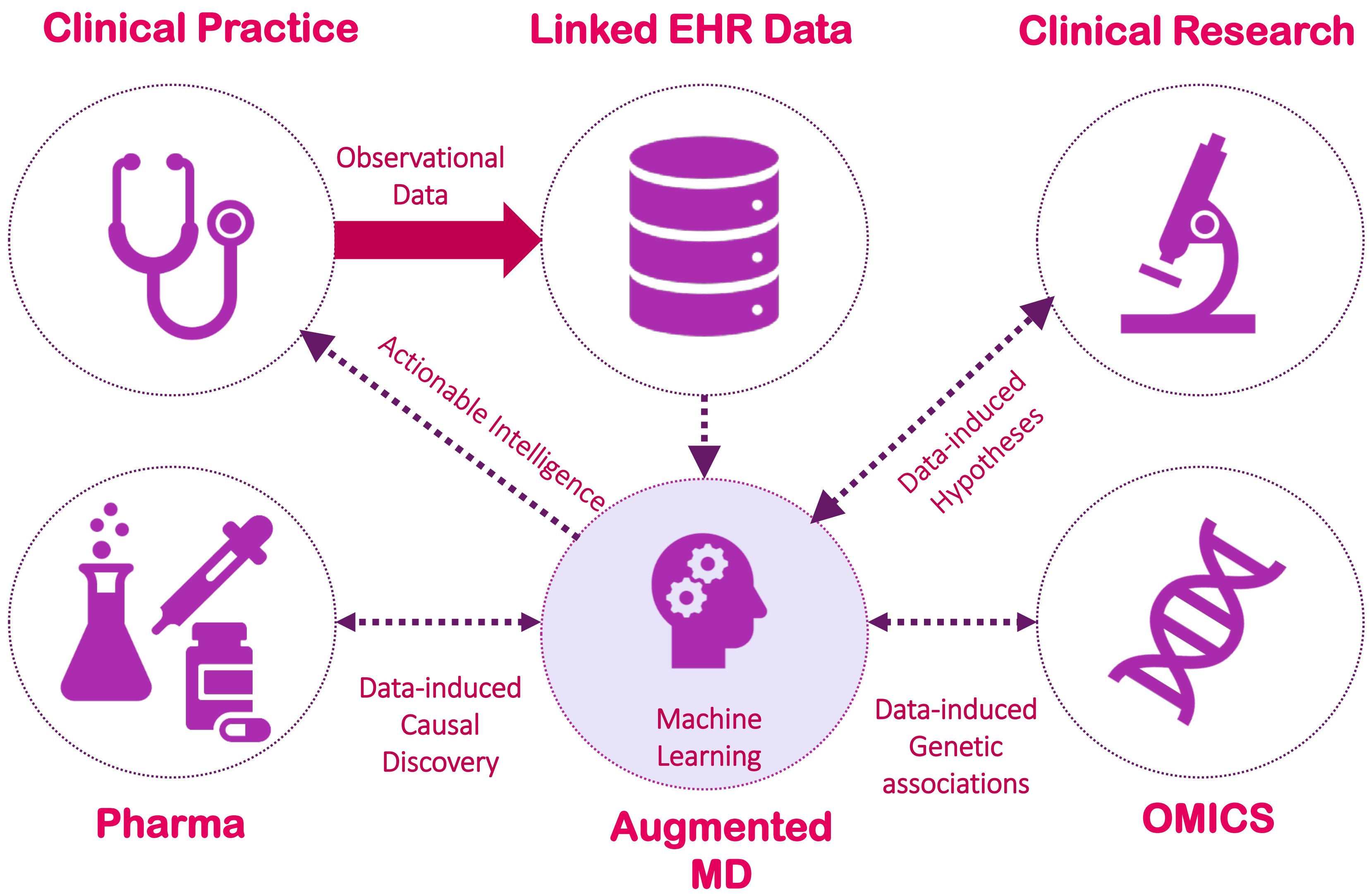
Our approach: Post-hoc methodology with frequentist coverage guarantees

	Method	Post-hoc vs Built-in	Coverage
	Bayesian neural nets (Ritter et al., 2018)	Built-in	No guarantees
	Probabilistic backprop. (Blundell et al., 2015)	Built-in	No guarantees
	Monte Carlo dropout (Gal & Ghahramani, 2016)	Built-in	No guarantees
	Deep Ensembles (Lakshminarayanan et al., 2017)	Built-in	No guarantees
*	**Discriminative Jackknife (Alaa and vdS,ICML2020)**	**Post-hoc**	**1-α**

- Does not interfere with model training or compromise accuracy!

Alaa and vdS,ICML2020 - Frequentist Uncertainty in Recurrent Neural Networks via *Blockwise* Influence Functions

Machine Learning & Healthcare: Vision



Transforming healthcare using ML

- New understanding of diseases (Think Covid-19!)
- New understanding of relationships among diseases (multiple morbidities)
- New understanding of effects of interventions/treatments

- New ways of screening and monitoring
- New ways of preventing disease
- New ways of diagnosing and staging disease
- New ways of treating disease
  - Impact on clinical trials
- New ways of allocating resources
- New pathways of care

Join us!

Acknowledgements

Dr. Ahmed Alaa
Dr. Hyunsuk Lee
Jinsung Yoon
Alexis Bellot
Changhee Lee
Yao Zhang
Ioana Bica
Zhaozhi Qian
Trent Kyono
James Jordon
Dan Jarrett
Alihan Huyuk

Email: mv472@cam.ac.uk
Website: http://www.vanderschaar-lab.com/

http://www.vanderschaar-lab.com

We are hiring brilliant students and post-docs