

# **Learning Engines for Healthcare: Using Deep Learning to Transform Clinical Practice and Discovery**

**Mihaela van der Schaar**

John Humphrey Plummer Professor of Machine Learning,  
Artificial Intelligence and Medicine

University of Cambridge

Alan Turing Institute

University of California Los Angeles

**<http://www.vanderschaar-lab.com>**

# My research

Develop cutting-edge machine learning and AI theory, methods, algorithms and systems to **deliver precision medicine at the patient-level**

- 1) **understand** the basis of health and disease
- 2) **support** clinical decisions for the patient at hand
- 3) **inform and improve** clinical pathways, better utilize resources & reduce costs
- 4) **transform** public health and policy

## Opportunity:

- Develop unique deep learning methods inspired by unique challenges posed by medicine/healthcare
- Be part of a vibrant machine learning for healthcare community
- Make medical discoveries
- Improve healthcare delivery

# The challenge of medicine - personalisation



- Pathogenesis
- Inherited, exposure and lifestyle risk
- Demography
- Co-morbidities
- Treatments and interventions
- etc

**Goal: deliver decision support direct to individual clinicians and patients**

# Cancer - a useful exemplar

- Common, costly and important group of disorders
- Complex affecting multiple clinical systems
- Varied aetiologies, presentations, management and long-term outcomes
- Care delivered through multiple organisations,
- Leads in personalised medicine therapeutics & genotype-phenotype correlation

ML-AIM Predictor (Beta)

ML-AIM PREDICTOR BETA

BREAST CANCER   COLON CANCER   LUNG CANCER   PROSTATE CANCER   HOW IT WORKS   CREDITS

# 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

ML-AIM Predictor (Beta)

ML-AIM PREDICTOR BETA

BREAST CANCER COLON CANCER LUNG CANCER PROSTATE CANCER HOW IT WORKS CREDITS

## Breast Cancer

**Input Diagnosis Information**

Age at Diagnosis	Tumor Size
Enter value...	Enter value...
ER Status	HER2 Status
Select one...	Select one...
Cancer Stage	Nodes Involved
Select one...	Enter value...
Tumor Grade	Detected by Screening
Select one...	Select one...

**Input Pathology Information**

Vascular Invasion	Distance to Resection
Enter value...	Enter value...
PT	PN
Enter value...	Enter value...
PR	Ki67
Enter value...	Enter value...
B5B	E-cadherin
Enter value...	Enter value...

Pathology Report

Grade 3, her2 positive,

Select File

Prediction Importance

Prediction Horizons →

Y

Y

ML-AIM Predictor (Beta)

ML-AIM PREDICTOR BETA

BREAST CANCER COLON CANCER LUNG CANCER PROSTATE CANCER HOW IT WORKS CREDITS

## Breast Cancer

Input Diagnosis Information

Age at Diagnosis      Tumor Size  
Enter value...      Enter value...

ER Status      HER2 Status  
Select one...      Select one...

Cancer Stage      Nodes Involved  
Select one...      Enter value...

Tumor Grade      Detected by  
Select one...      Select one...

Input Pathology Information

Vascular Invasion      Distance to Resection  
Enter value...      Enter value...

PT      PN  
Enter value...      Enter value...

PR      Ki67  
Enter value...      Enter value...

or, Upload Pathology Report

Distance margin: 3mm, grade 3, her2 positive, ki67 negative, ....

Drag-and-Drop or [Select File](#)

Select a patient

Time since Initial Diagnosis (Months)

Mortality Risk over Time

Load or Create New Patient...

Create New Patient...

Patient 1

Patient 2

Patient 3

Patient 4

ML-AIM Predictor (Beta)

BREAST CANCER COLON CANCER LUNG CANCER PROSTATE CANCER HOW IT WORKS CREDITS

### Input Diagnosis Information

Age at Diagnosis	Tumor Size
60	41

ER Status	HER2 Status
Positive	Negative

Cancer Stage	Nodes Involved
Stage 2	4

Tumor Grade	Detected by Screening
Grade 3	Yes

Time since Initial Diagnosis (Months)

0 6 12 18 24 30 36 42 48 54

### Input Diagnosis Information

Age at Diagnosis	Tumor Size
60	41

ER Status	HER2 Status
Positive	Negative

Cancer Stage	Nodes Involved
Stage 2	4

Tumor Grade	Detected by Screening
Grade 3	Yes

### Mortality Risk over Time

Historical One-Year Risk    Estimated Forward Risk

Y-axis: Mortality Risk (0.0 to 1.0). X-axis: Time since Initial Diagnosis (Months) (0 to 54). The chart shows two lines: a solid blue line for Historical One-Year Risk and a dashed orange line with dots for Estimated Forward Risk. Both lines start at approximately 0.05 at month 0 and remain relatively flat until month 12, then rise to about 0.15 by month 54.

### Individualized Feature Importance

← Prediction Horizons →

The chart displays feature importance for two variables: Age and Tumor Size. The importance for Age is shown in light orange, and for Tumor Size in dark orange. A vertical bar indicates a prediction horizon of 1 unit. The legend indicates that the left side represents the 'Prediction Horizons'.

Feature	Importance
Age	1
Tumor Size	0.8

ML-AIM Predictor (Beta)

BREAST CANCER   COLON CANCER   LUNG CANCER   PROSTATE CANCER   HOW IT WORKS   CREDITS

## Breast Cancer

**Input Pathology Information**

Vascular Invasion	Distance to Resection
N	7
PT	PN
X	pn0
PR	Ki67
N	X
B5B	E-cadherin
X	X

or, Upload Pathology Report

marking axillary tail. At approximately 12 o'clock position needle tract with surrounding fat is identified. There is some fibrosis and this area extends up to around 60mm and lies 15mm from the deep margin, 20mm from superior and 60mm from inferior margin. Part C - labelled ""Left sentinel node #2"". A lymph node with surrounding fat measuring

Drag-and-Drop or [Select File](#)

or, Upload Pathology Report

marking axillary tail. At approximately 12 o'clock position needle tract with surrounding fat is identified. There is some fibrosis and this area extends up to around 60mm and lies 15mm from the deep margin, 20mm from superior and 60mm from inferior margin. Part C - labelled ""Left sentinel node #2"". A lymph node with surrounding fat measuring

Drag-and-Drop or [Select File](#)

Patient 3

Individualized Feature Importance  
← Prediction Horizons →

Feature	Historical One-Year Risk (Blue)	Estimated Forward Risk (Orange)
Age	1.0	1.0
Tumor Size	0.8	0.8

**Upload a pathology report**

ML-AIM Predictor (Beta)

**ML-AIM PREDICTOR BETA**

Input Diagnosis Information      Input Pathology Information      BREAST CANCER      COLON CANCER      LUNG CANCER      PROSTATE CANCER      HOW IT WORKS      CREDITS

or, Upload Pathology Report

marking axillary tail. At approximately 12 o'clock position needle tract with surrounding fat is identified. There is some fibrosis and this area extends up to around 60mm and lies 15mm from the deep margin, 20mm from superior and 60mm from inferior margin. Part C - labelled "Left sentinel node #2". A lymph node with surrounding fat measuring

Drag-and-Drop or [Select File](#)

Time since Initial Diagnosis (Months)

Mortality Risk over Time

Historical One-Year Risk    Estimated Forward Risk

Estimated Probability

Individualized Feature Importance

Patient 3

← Prediction Horizons →

Age  
Tumor Size  
ER Status  
HER2 Status  
Stage  
Nodes Involved  
Grade  
Detected by Screening

1 yr    2 yr    3 yr    4 yr

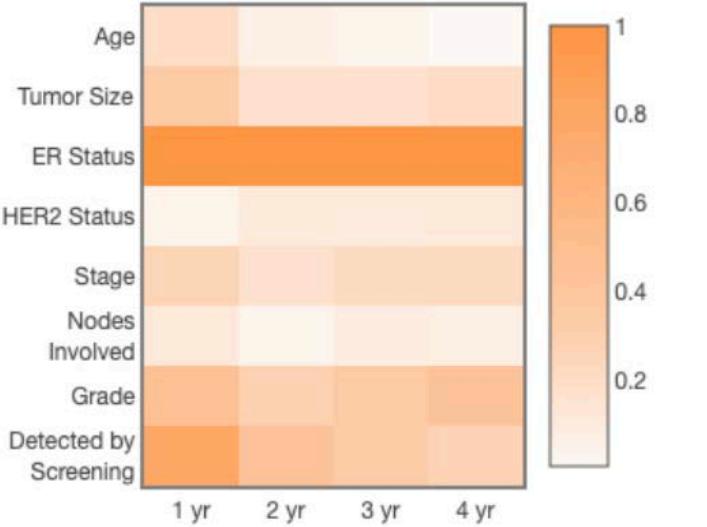
# AutoPrognosis



ML-AIM PRED

### Individualized Feature Importance

← Prediction Horizons →



# INVASE

Age at Diagnosis

Tumor Size

60

41

ER Status

HER2 Status

Positive

Negative

Cancer Stage

Nodes Involved

Stage 2

4

Tumor Grade

Detected by Screening

Grade 3

Yes

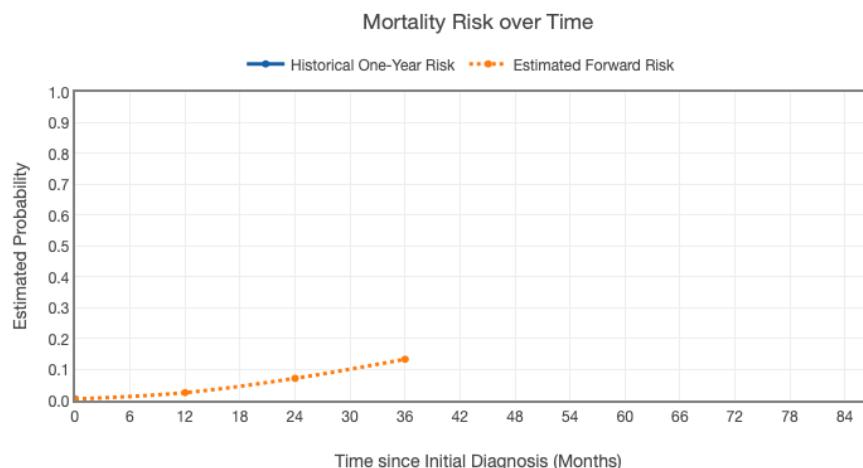
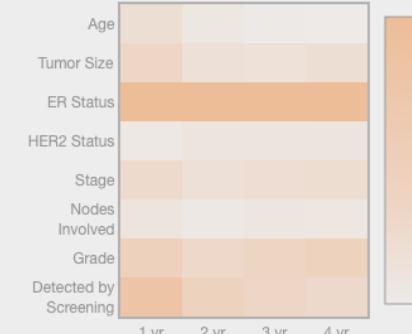
Diagnosis (Months)

48 54 60 66 72 78 84

Patient 3

### Individualized Feature Importance

← Prediction Horizons →

[PROSTATE CANCER](#) [HOW IT WORKS](#) [CREDITS](#)



## Input Diagnosis Information

Age at Diagnosis	Tumor Size
60	41
ER Status	HER2 Status
Positive	Negative
Cancer Stage	Nodes Involved
Stage 2	4
Tumor Grade	Detected by Screening
Grade 3	Yes

Vascular  
N  
P  
X  
PR  
N  
B5B  
XTime since Initial Diagnosis (Months)  
0 6 12 18 24 30 36 42 48 54 60 66 72 78 84

Patient 3

## Age at Diagnosis

## Tumor Size

60

41

## ER Status

## HER2 Status

Negative

Negative

## Cancer Stage

## Nodes Involved

Stage 2

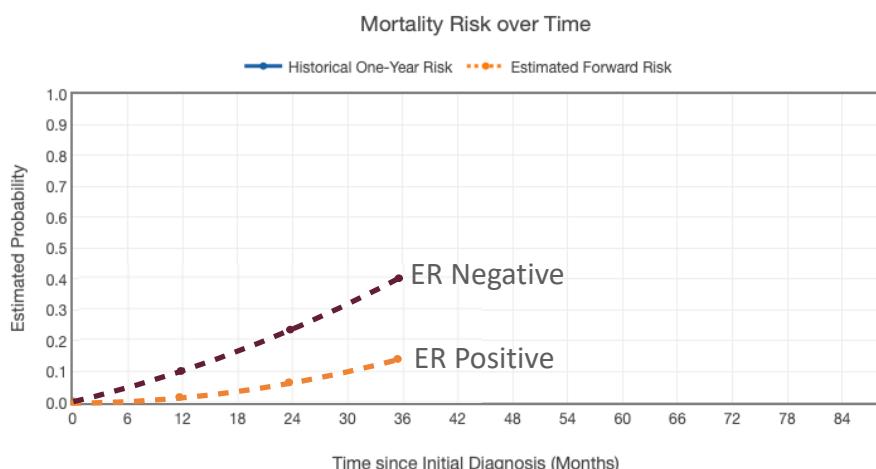
4

## Tumor Grade

## Detected by Screening

Grade 3

Yes



## Individualized Feature Importance

← Prediction Horizons →





## Time since Initial Diagnosis (Months)

0

6

12

18

24

30

36

42

48

54

60

66

72

78

84

Estimated Probability

1.0  
0.9  
0.8  
0.7  
0.6  
0.5  
0.4  
0.3  
0.2  
0.1  
0.0

Time since Initial Diagnosis (Months)

Events (Occurrences)

50  
45  
40  
35  
30  
25  
20  
15  
10  
5

Time since Initial Diagnosis (Months)

Age

Tumor Size

ER Status

HER2 Status

Stage

Nodes

Involved

Grade

Detected by Screening

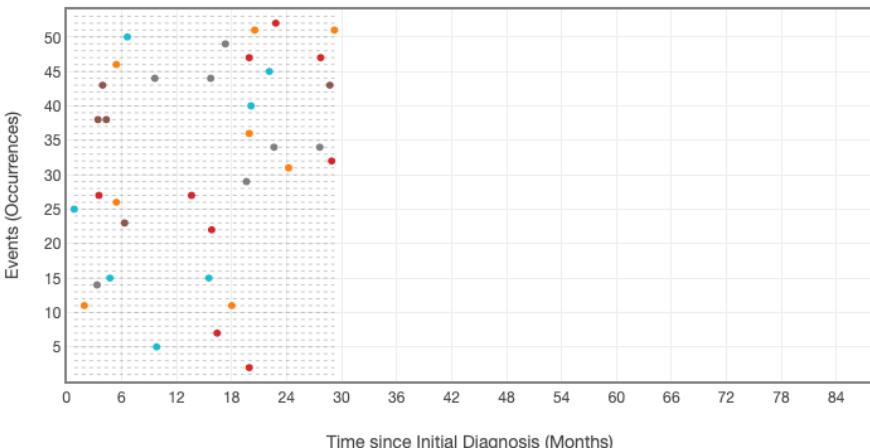
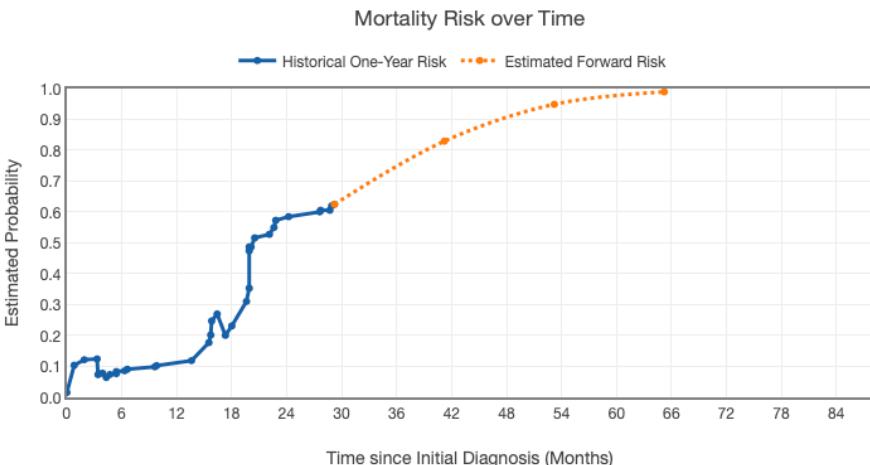
1 yr  
2 yr  
3 yr  
4 yr1  
0.8  
0.6  
0.4  
0.2

- Event 14 - MRI, COSD
- Event 15 - Other tumor diagnosis
- Event 16 - CT from AV\_TREATMENT
- Event 17 - RT from AV\_TREATMENT
- Event 18 - Surgery from AV\_TREATMENT
- Event 19 - Major Surgery from AV\_TREATM
- Event 20 - SACT cycle event
- Event 21 - SACT drug event
- Event 22 - Hormone from AV\_TREATMENT
- Event 23 - Brachy from AV\_TREATMENT
- Event 24 - Biopsy
- Event 25 - Endoscopy from COSD
- Event 26 - COSD Pathology
- Event 27 - PTD start point

ADD NEW EVENT



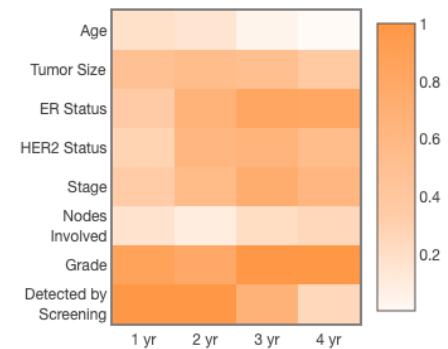
PASS



Patient 3

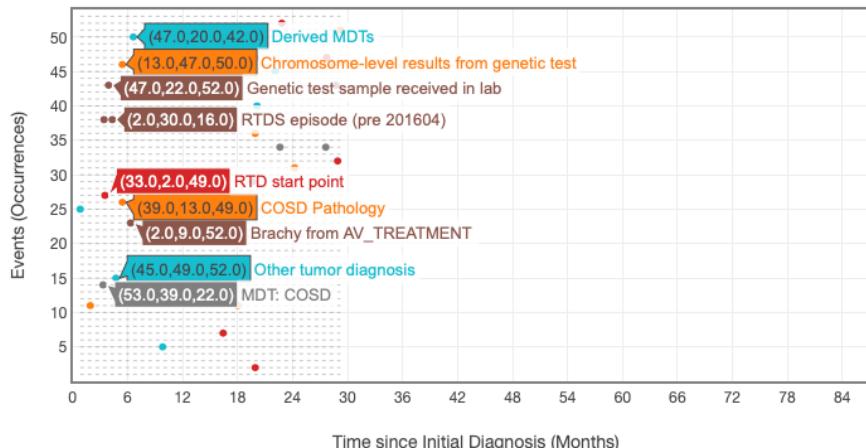
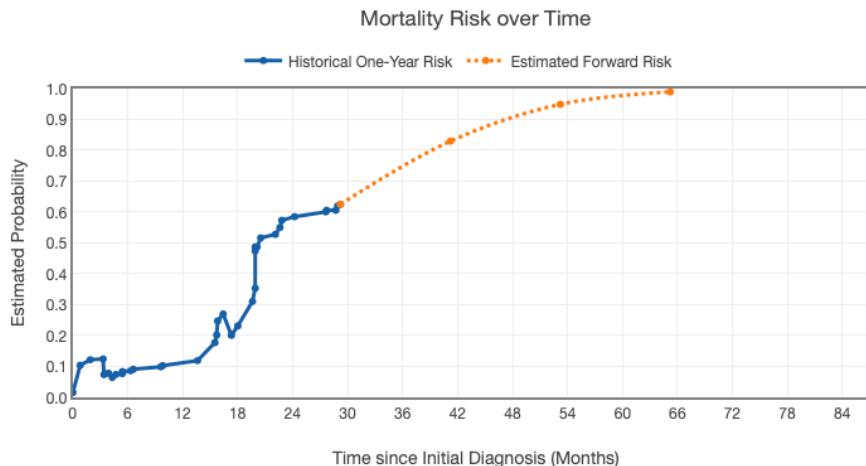
## Individualized Feature Importance

← Prediction Horizons →



- Event 14 - IMRT\_COSD
- Event 15 - Other tumor diagnosis
- Event 16 - CT from AV\_TREATMENT
- Event 17 - RT from AV\_TREATMENT
- Event 18 - Surgery from AV\_TREATMENT
- Event 19 - Major Surgery from AV\_TREATMENT
- Event 20 - SACT cycle event
- Event 21 - SACT drug event
- Event 22 - Hormone from AV\_TREATMENT
- Event 23 - Brachy from AV\_TREATMENT
- Event 24 - Biopsy
- Event 25 - Endoscopy from COSD
- Event 26 - COSD Pathology
- Event 27 - RTD start point

ADD NEW EVENT



Patient 3

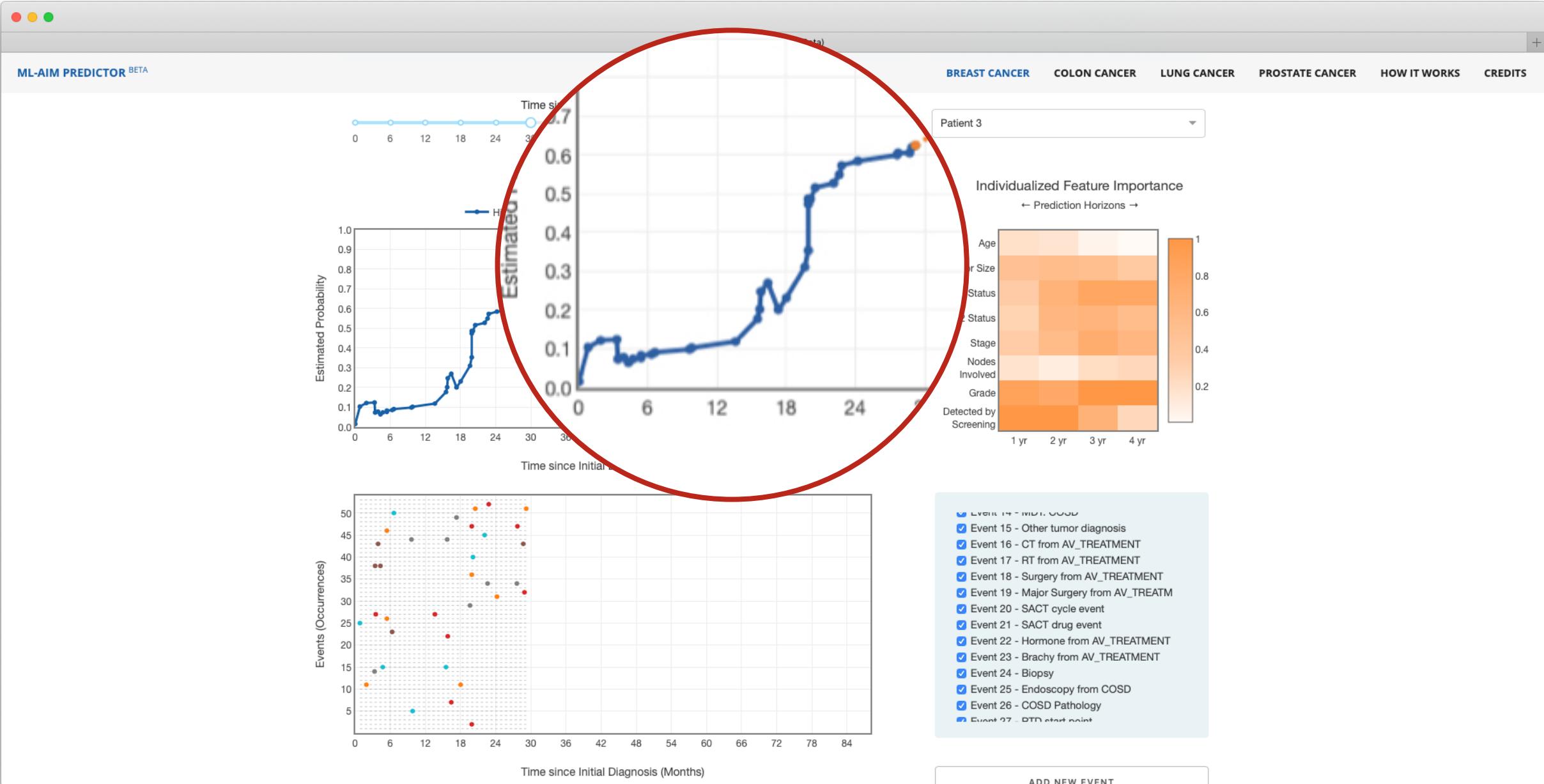
### Individualized Feature Importance

← Prediction Horizons →



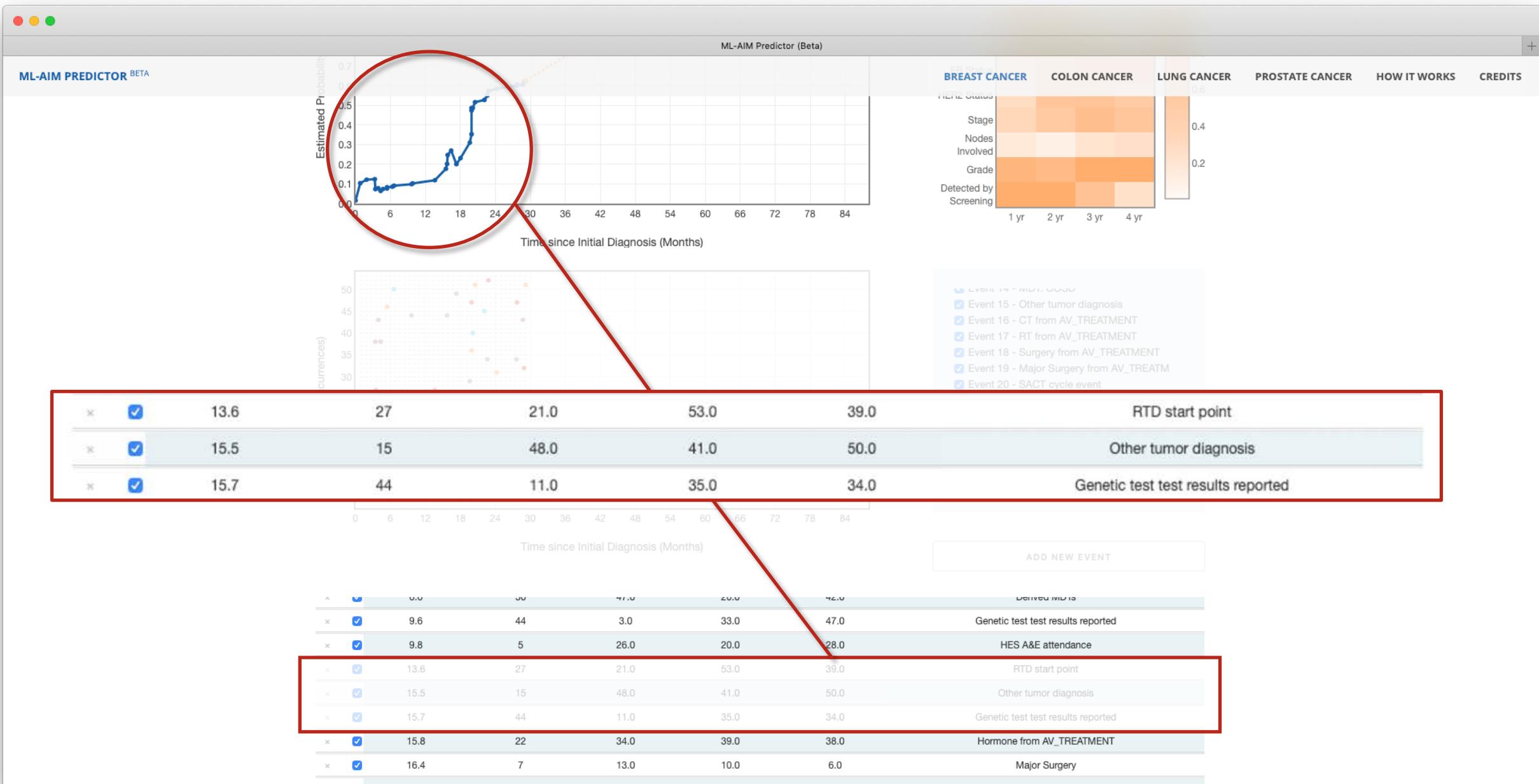
- Event 14 - MDT: COSD
- Event 15 - Other tumor diagnosis
- Event 16 - CT from AV\_TREATMENT
- Event 17 - RT from AV\_TREATMENT
- Event 18 - Surgery from AV\_TREATMENT
- Event 19 - Major Surgery from AV\_TREATMENT
- Event 20 - SACT cycle event
- Event 21 - SACT drug event
- Event 22 - Hormone from AV\_TREATMENT
- Event 23 - Brachy from AV\_TREATMENT
- Event 24 - Biopsy
- Event 25 - Endoscopy from COSD
- Event 26 - COSD Pathology
- Event 27 - RTD start point

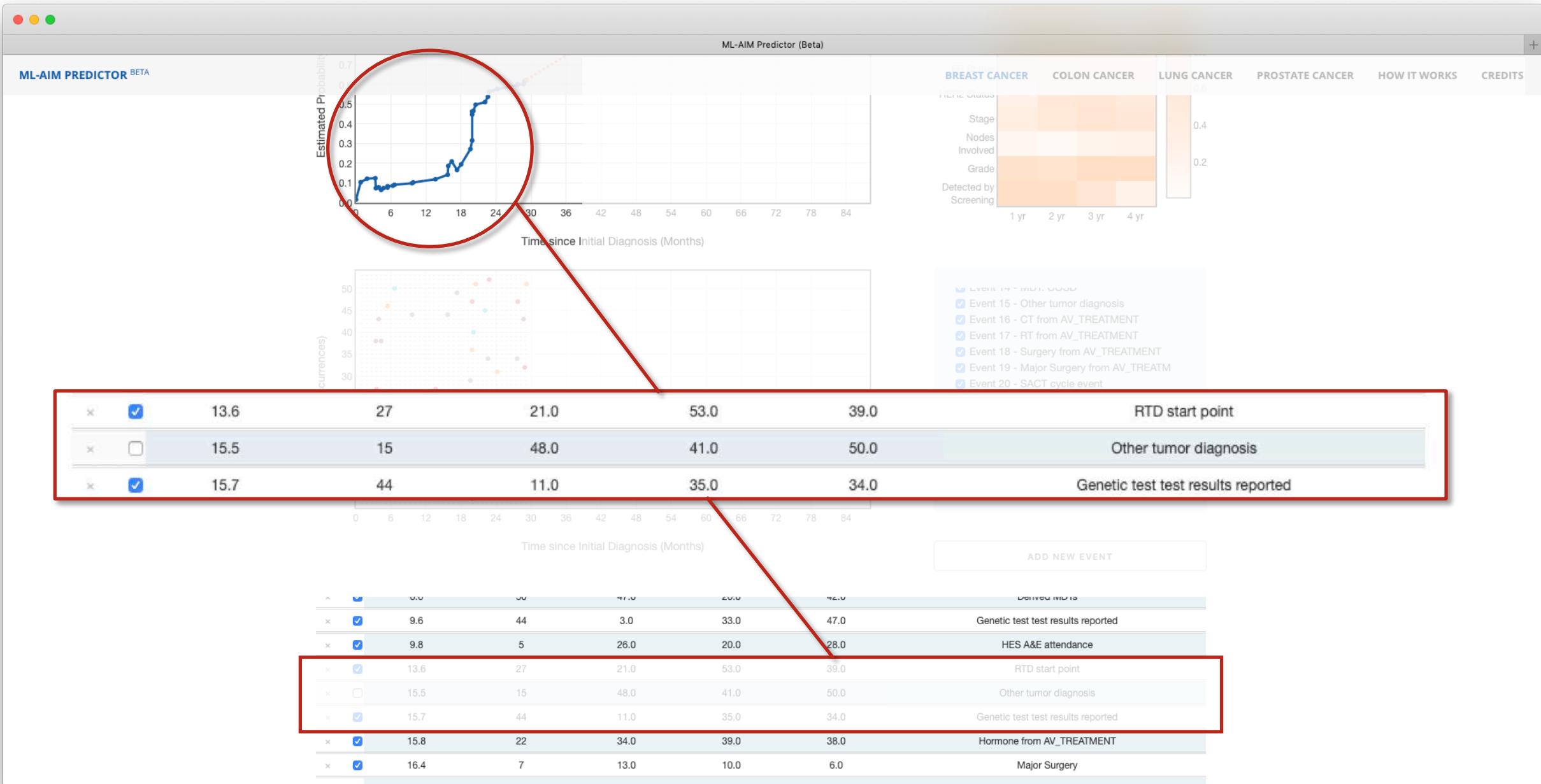
ADD NEW EVENT

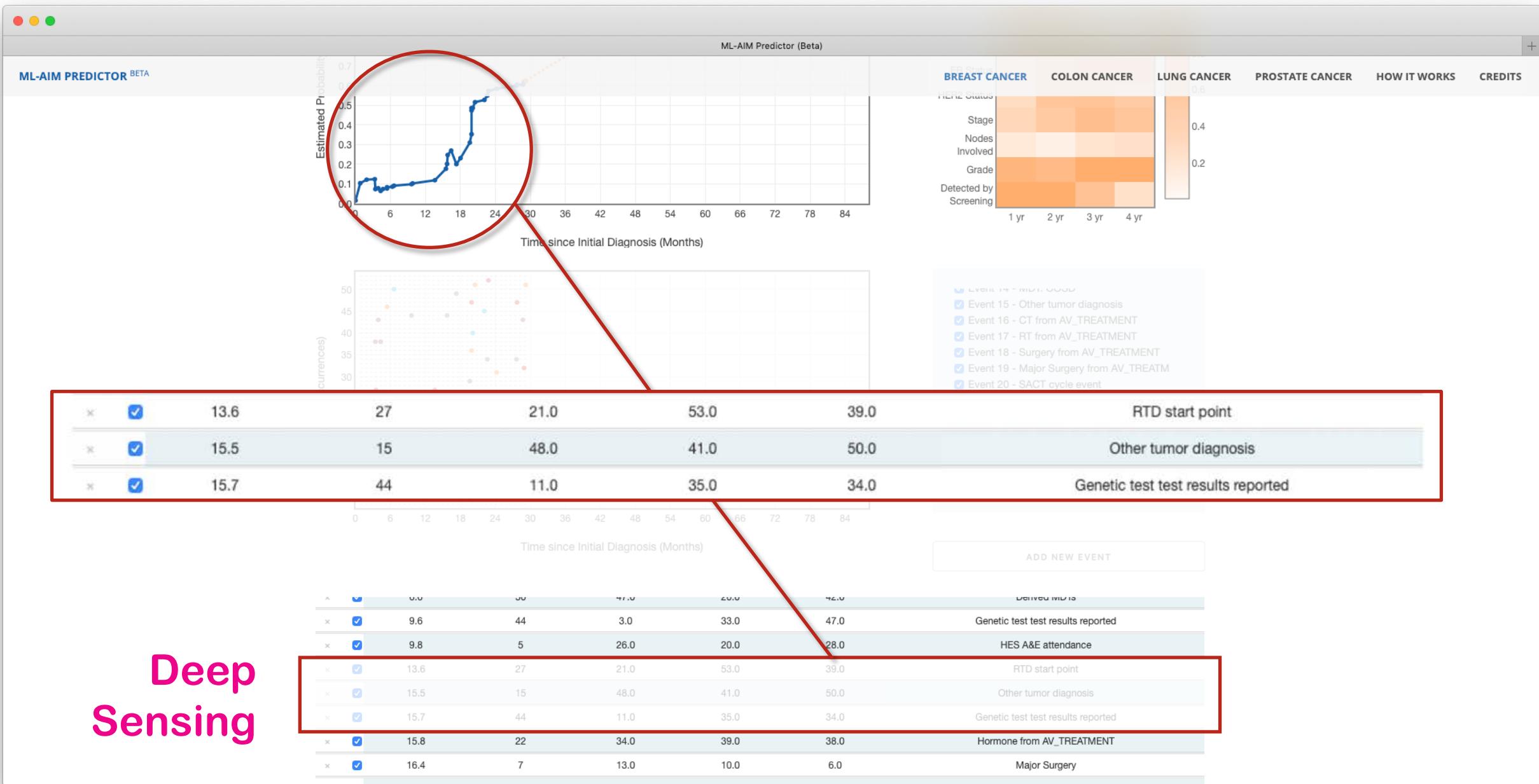












Deep  
Sensing

ML-AIM Predictor (Beta)

**ML-AIM PREDICTOR BETA**

	Age	Tumor Size	ER Status	HER2 Status	Stage 1	Stage 2	Stage 3	Stage 4	Nodes Involved	Grade 1	Grade 2	Grade 3	Detected by Screening
x	14.7	31	42.0	25.0	19.0								
x	16.1	42	18.0	26.0	7.0								
x	18.6	42	1.0	23.0	17.0								
x	18.7	46	7.0	26.0	51.0								
x	20.0	18	29.0	23.0	39.0								
x	21.0	45	29.0	40.0	35.0								
v	21.1	33	14.0	11.0	24.0								

BREAST CANCER   Endoscopy for HCC   COLON CANCER   LUNG CANCER   PROSTATE CANCER   HOW IT WORKS   CREDITS

Risk of Recurrence vs. Treatment Options

Treatment Option	One-Year Risk (Population-based)	One-Year Risk (Individualized)	Treatment Propensity Score
No Treatment	50%	35%	35%
Radiotherapy	32%	21%	47%
Chemotherapy	26%	7%	13%
Chemo + Radiotherapy	21%	13%	31%

Top 3 Similar Patients

Patient ID	Age	Tumor Size	ER Status	HER2 Status	Stage 1	Stage 2	Stage 3	Stage 4	Nodes Involved	Grade 1	Grade 2	Grade 3	Detected by Screening
1	48	19	1	1	0	1	0	0	4	1	0	0	0
32	53	17	1	1	0	1	0	0	4	0	1	0	0
27	45	25	0	1	0	1	0	0	3	0	1	0	0

ML-AIM Predictor (Beta)

**ML-AIM PREDICTOR BETA**

	14.7	31	42.0	25.0	19.0	BREAST CANCER	Endoscopy for HCC	COLON CANCER	LUNG CANCER	PROSTATE CANCER	HOW IT WORKS	CREDITS
x	16.1	42	18.0	26.0	7.0		Genetic test sample analysis requested					
x	18.6	42	1.0	23.0	17.0		Genetic test sample analysis requested					
x	18.7	46	7.0	26.0	51.0		Chromosome-level results from genetic test					
x	20.0	18	29.0	23.0	39.0		Surgery from AV_TREATMENT					
x	21.0	45	29.0	40.0	35.0		Gene-level results from genetic test					
v	21.1	33	14.0	11.0	24.0		Path sample taken					

Risk of Recurrence vs. Treatment Options

Treatment Option	One-Year Risk (Population-based)	One-Year Risk (Individualized)	Treatment Propensity Score
No Treatment	49%	35%	35%
Radiotherapy	31%	21%	47%
Chemotherapy	26%	7%	13%
Chemo + Radiotherapy	21%	13%	31%

Top 3 Similar Patients

Patient ID	Age	Tumor Size	ER Status	HER2 Status	Stage 1	Stage 2	Stage 3	Stage 4	Nodes Involved	Grade 1	Grade 2	Grade 3	Detected by Screening
1	48	19	1	1	0	1	0	0	4	1	0	0	0
32	53	17	1	1	0	1	0	0	4	0	1	0	0
27	45	25	0	1	0	1	0	0	3	0	1	0	0

ML-AIM Predictor (Beta)

**ML-AIM PREDICTOR BETA**

	14.7	31	42.0	25.0	19.0	BREAST CANCER	Endoscopy for HCC	COLON CANCER	LUNG CANCER	PROSTATE CANCER	HOW IT WORKS	CREDITS
x	16.1	42	18.0	26.0	7.0		Genetic test sample analysis requested					
x	18.6	42	1.0	23.0	17.0		Genetic test sample analysis requested					
x	18.7	46	7.0	26.0	51.0		Chromosome-level results from genetic test					
x	20.0	18	29.0	23.0	39.0		Surgery from AV_TREATMENT					
x	21.0	45	29.0	40.0	35.0		Gene-level results from genetic test					
v	21.1	33	14.0	11.0	24.0		Path sample taken					

Risk of Recurrence vs. Treatment Options

No Treatment    Radiotherapy    Chemotherapy    Chemo + Radiotherapy

Treatment Category	No Treatment (%)	Radiotherapy (%)	Chemotherapy (%)	Chemo + Radiotherapy (%)
One-Year Risk (Population-based)	50%	32%	26%	21%
One-Year Risk (Individualized)	35%	21%	7%	13%
Treatment Propensity Score	35%	47%	13%	31%

Top 3 Similar Patients

Patient ID	Age	Tumor Size	ER Status	HER2 Status	Stage 1	Stage 2	Stage 3	Stage 4	Nodes Involved	Grade 1	Grade 2	Grade 3	Detected by Screening
1	48	19	1	1	0	1	0	0	4	1	0	0	0
32	53	17	1	1	0	1	0	0	4	0	1	0	0
27	45	25	0	1	0	1	0	0	3	0	1	0	0

ML-AIM Predictor (Beta)

**ML-AIM PREDICTOR BETA**

	14.7	31	42.0	25.0	19.0	BREAST CANCER	Endoscopy for HCC	COLON CANCER	LUNG CANCER	PROSTATE CANCER	HOW IT WORKS	CREDITS
x	16.1	42	18.0	26.0	7.0		Genetic test sample analysis requested					
x	18.6	42	1.0	23.0	17.0		Genetic test sample analysis requested					
x	18.7	46	7.0	26.0	51.0		Chromosome-level results from genetic test					
x	20.0	18	29.0	23.0	39.0		Surgery from AV_TREATMENT					
x	21.0	45	29.0	40.0	35.0		Gene-level results from genetic test					
v	21.1	33	14.0	11.0	24.0		Path sample taken					

Risk of Recurrence vs. Treatment Options

Treatment Category	No Treatment (%)	Radiotherapy (%)	Chemotherapy (%)	Chemo + Radiotherapy (%)
One-Year Risk (Population-based)	50%	32%	27%	22%
One-Year Risk (Individualized)	35%	21%	7%	13%
Treatment Propensity Score	35%	47%	13%	31%

Top 3 Similar Patients

Patient ID	Age	Tumor Size	ER Status	HER2 Status	Stage 1	Stage 2	Stage 3	Stage 4	Nodes Involved	Grade 1	Grade 2	Grade 3	Detected by Screening
1	48	19	1	1	0	1	0	0	4	1	0	0	0
32	53	17	1	1	0	1	0	0	4	0	1	0	0
27	45	25	0	1	0	1	0	0	3	0	1	0	0

- GANITE
- NSGP
- Counterfactual Recurrent Nets

### Top 3 Similar Patients

Patient ID	Age	Tumor Size	ER Status	HER2 Status	Stage 1	Stage 2	Stage 3	Stage 4	Nodes Involved	Grade 1	Grade 2	Grade 3	Detected by Screening
1	48	19	1	1	0	1	0	0	4	1	0	0	0
32	53	17	1	1	0	1	0	0	4	0	1	0	0
27	45	25	0	1	0	1	0	0	3	0	1	0	0

### Top 3 Similar Patients

Patient ID	Age	Tumor Size	ER Status	HER2 Status	Stage 1	Stage 2	Stage 3	Stage 4	Nodes Involved	Grade 1	Grade 2	Grade 3	Detected by Screening
1	48	19	1	1	0	1	0	0	4	1	0	0	0
32	53	17	1	1	0	1	0	0	4	0	1	0	0
27	45	25	0	1	0	1	0	0	3	0	1	0	0

# Deep Predictive Clustering

This website does not provide medical or professional services advice. The content on this website is intended to be a general information resource in regard to the subject matter covered, is provided solely on an "as is" and "as available" basis without warranty of any kind, neither express nor implied, and is not intended or implied to be a substitute for medical advice offered by a physician. If you desire or need such services or advice, regarding the information contained on this website or otherwise, you should consult a professional healthcare provider. Your use of the website or any content on the website is at your own risk. We specifically disclaim any liability, whether based in contract, tort, strict liability or otherwise, for any direct, indirect, incidental, consequential, or special damages arising out of or in any way connected with access to or use of the website or the content on the website including, but not limited to, the risk calculator.



Copyright © 2019 • ML-AIM Group  
University of Cambridge  
All Rights Reserved

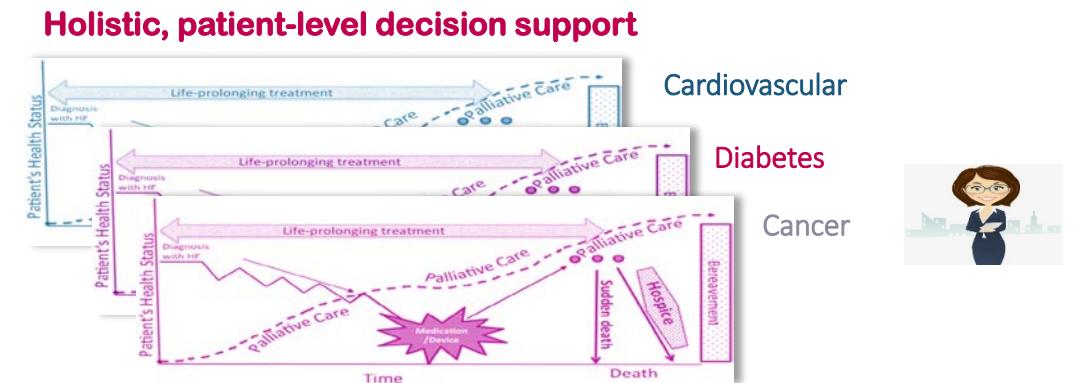
Part 1:

**Building Clinical Decision Support Systems**

enabling delivery of precision medicine at the patient-level

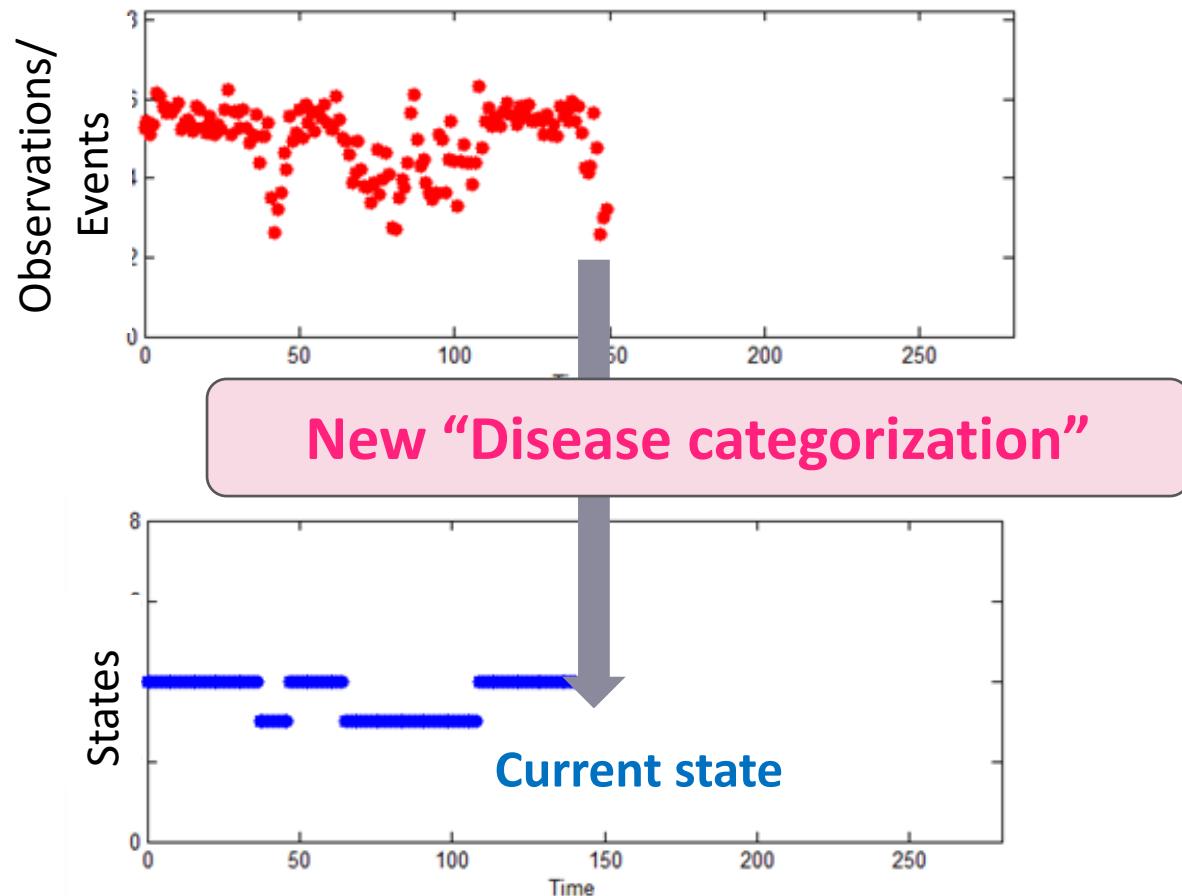
# Goal: Learn and forecast patient-level trajectories

- Understand, infer and forecast patient-level trajectory (diagnosis, evolution of bio-markers and subsequent outcomes, treatment effects, etc.)
- Revise patient-level trajectory as care continues
- Feed back to stakeholders extracted patient-level intelligence to deliver personalized care



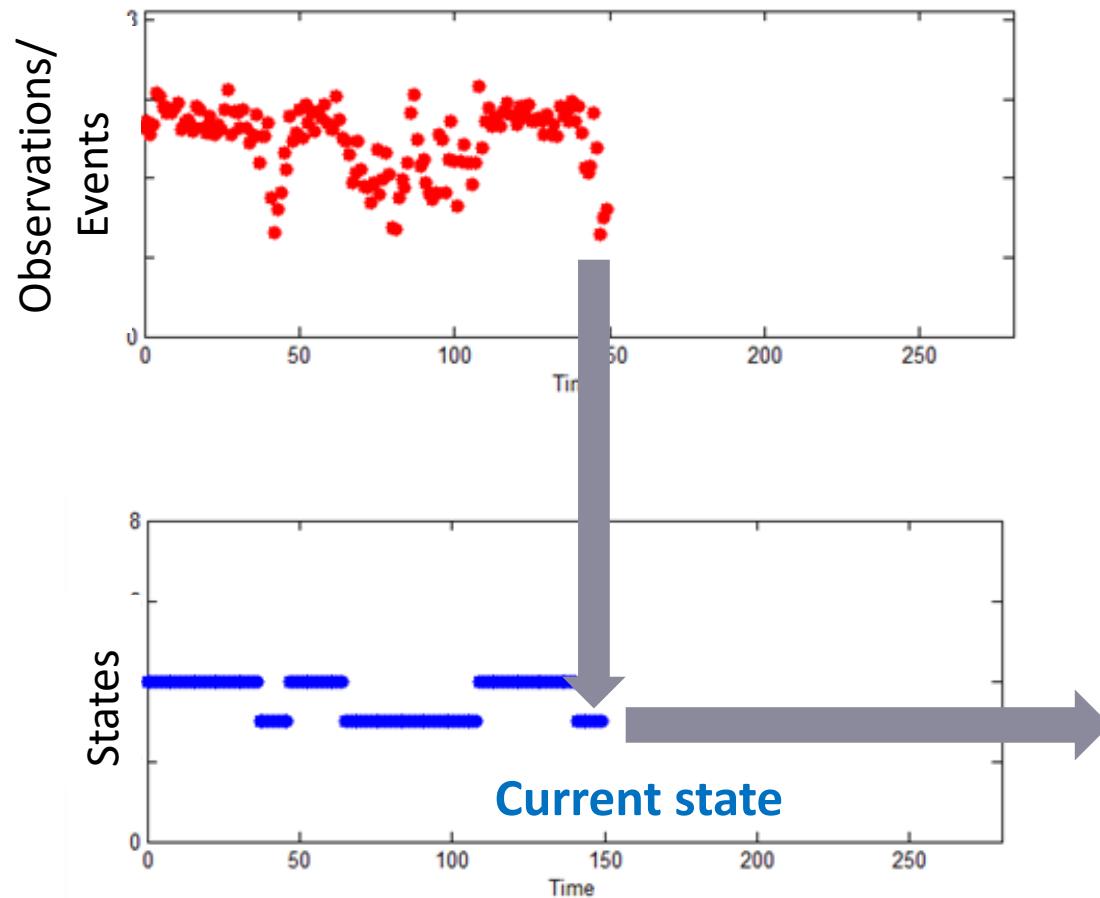
# How to think about health and disease? How to track disease?

Use ML to infer present and future health states of the **current patient** on the basis of observations about him/her



# How to think about health and disease? How to track disease?

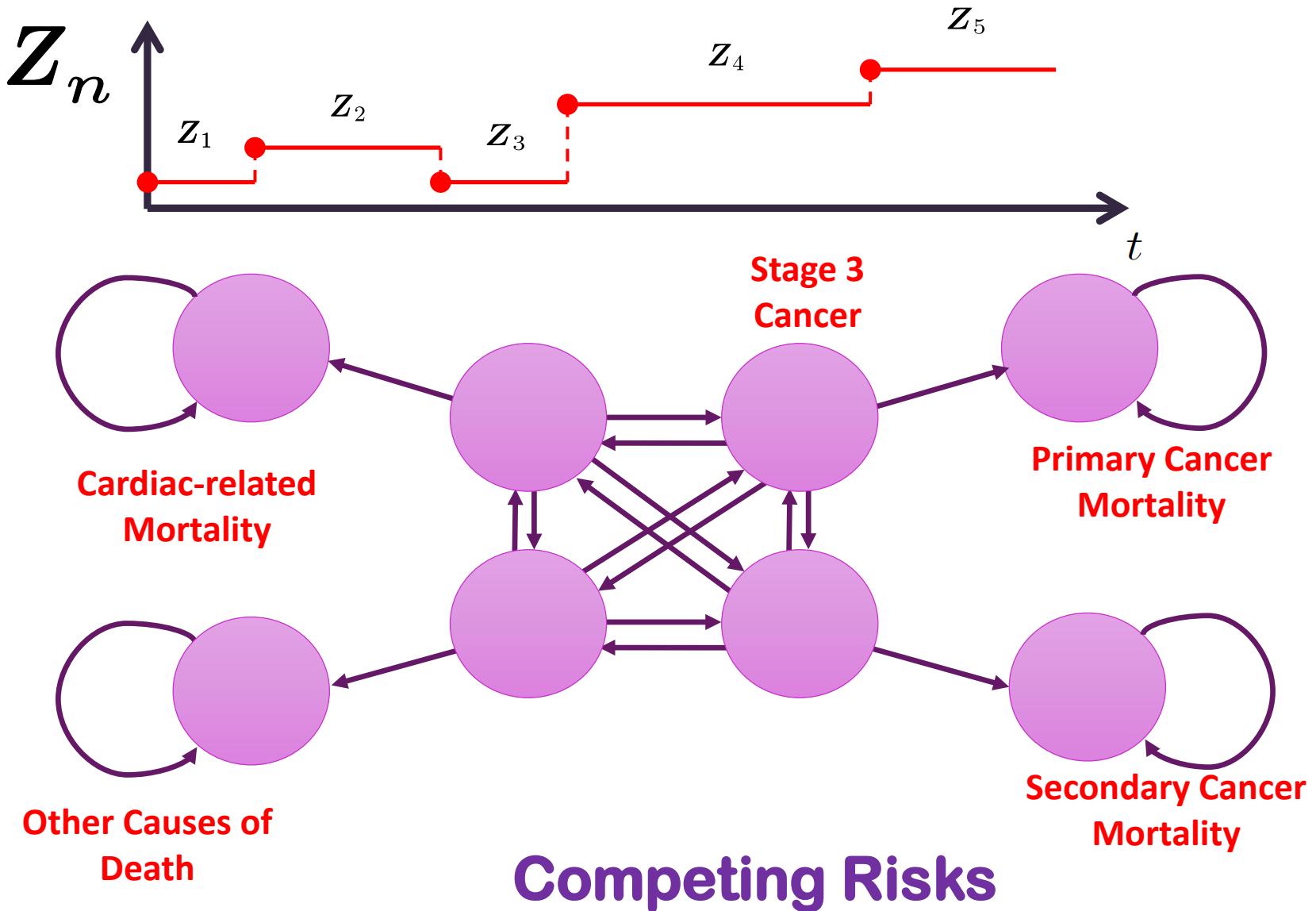
Use ML to infer present and future health states of the **current patient** on the basis of observations about him/her



Forecast Trajectory =  
Probability to be in a certain state  
at a time  $T$  in the future

**ACT**

# Define, infer and forecast health/disease state and state transitions



# **Why has AI/ML not been used so far for in medicine for decision support and discovery?**

## **Inadequate, simplistic models**

- **Unable to capture the complexity of medicine**
- **One-size-fits-all**
- **Uninterpretable**
- **Not easy to act upon**

# Current Disease Progression Models – simplistic and wrong

Markov Models

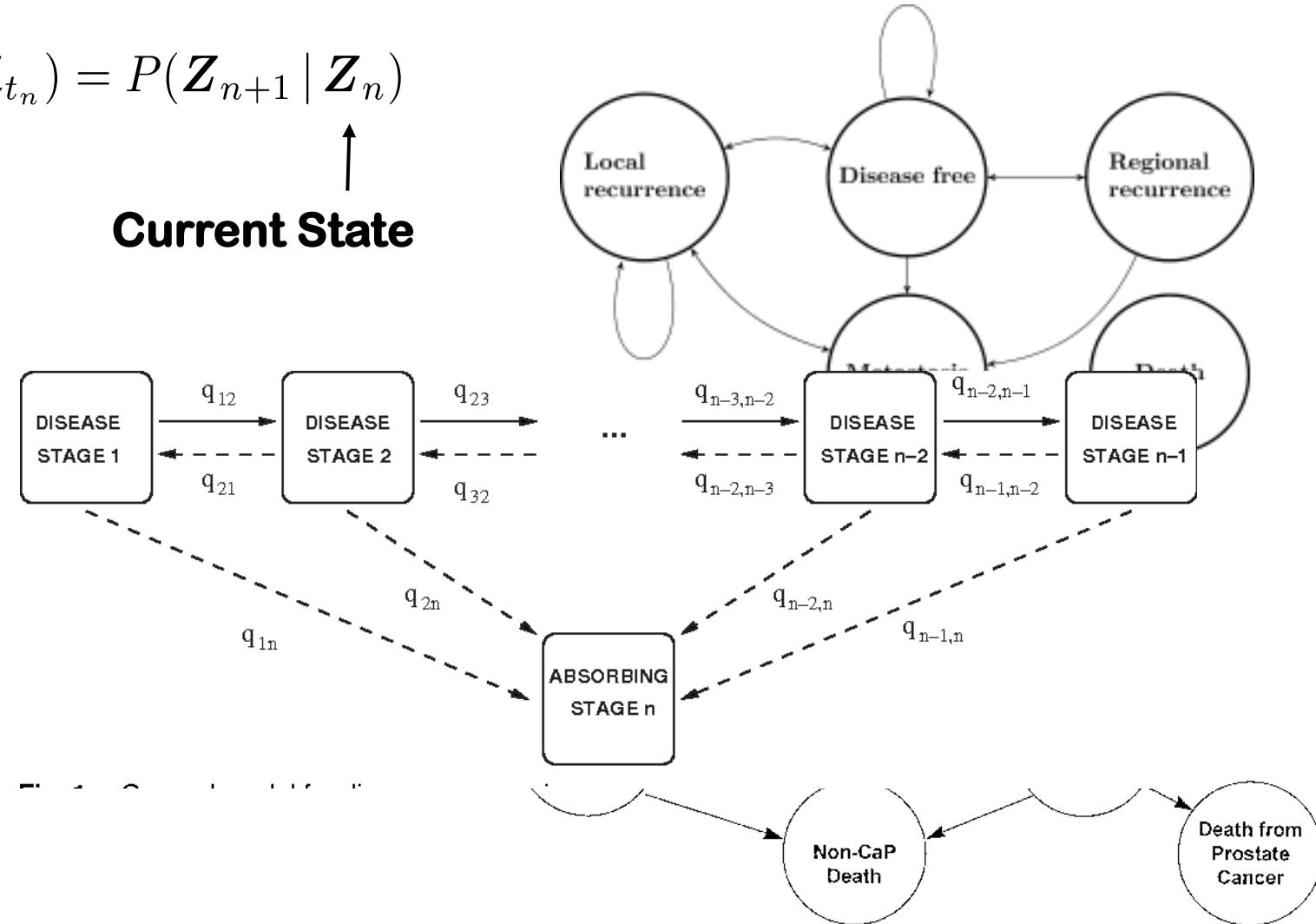
HMMs

Deep Markov Models

Etc.

$$P(Z_{n+1} | \mathcal{H}_{t_n}) = P(Z_{n+1} | Z_n)$$

↑  
Current State



Easy to understand  
and compute.....

But **WRONG**

## **Markov models?**

$$P(Z_{n+1} | \mathcal{H}_{t_n}) = P(Z_{n+1} | Z_n)$$

### **History matters!**

**Ignore history**

- Previous states
- Order of states
- Duration in a state

### **One size fits all!**

**Only capture population-level transitions across progression stages**  
**Ignores individual clinical trajectories**

# Do existing Deep Learning methods provide suitable solutions?

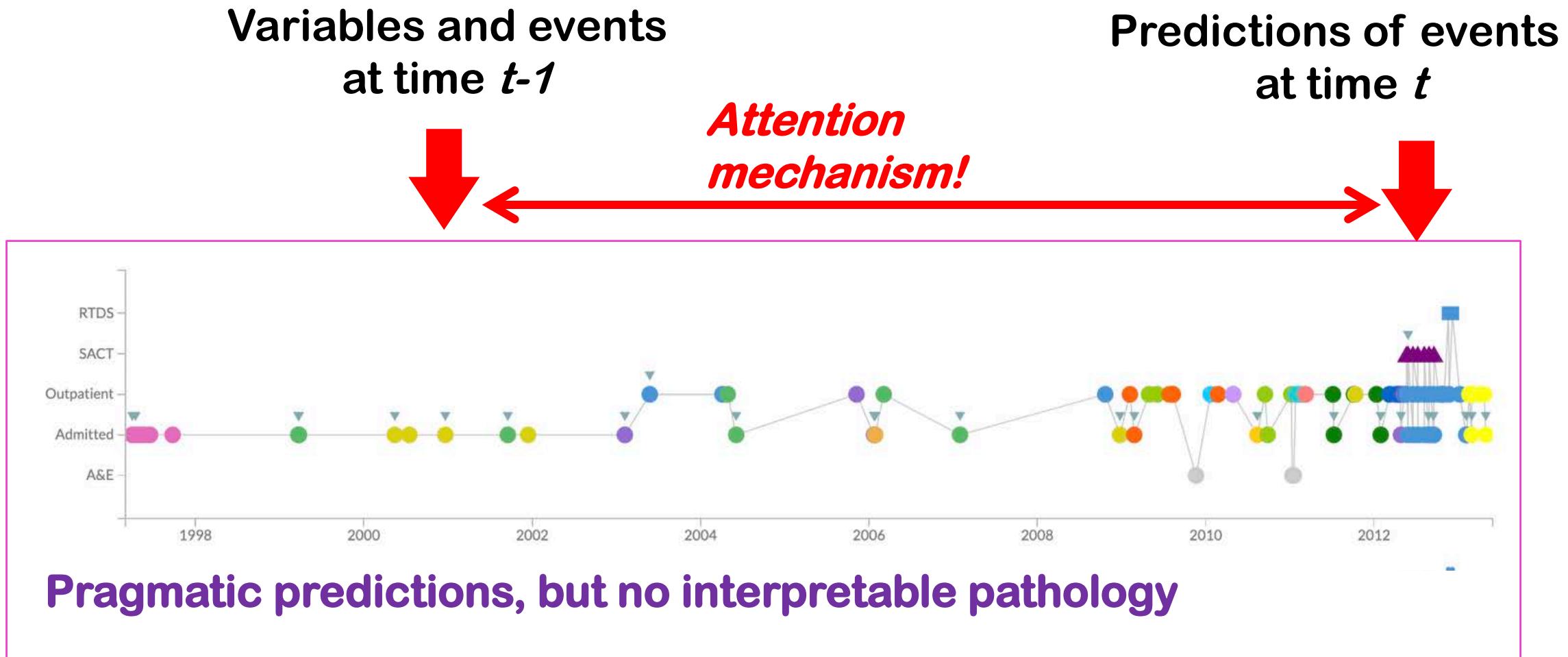
Modeling  $P(Z_{n+1} | \mathcal{H}_{t_n})$  using deep learning methods - recurrent neural network (RNN)?

# Current Disease Progression Models based on Deep Learning

[E. Choi, 2017][Lim and van der Schaar, ML4HC 2018, NeurIPS 2018]

## RNNs with attention mechanisms:

identify important variables for future predictions based on patient's history

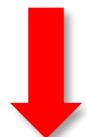


# Do existing Deep Learning methods provide suitable solutions?

**Modeling  $P(Z_{n+1} | \mathcal{H}_{t_n})$  using deep learning methods - recurrent neural network (RNN)?**



RNN considers the timing and order of events, but no notion of states



Not interpretable!

Cannot use or extract clinical knowledge!

Not able to answer important questions about early diagnosis/progression

# Models for Health and Disease Trajectories - Requirements

**Clinically actionable models for patient-level trajectory needed!**

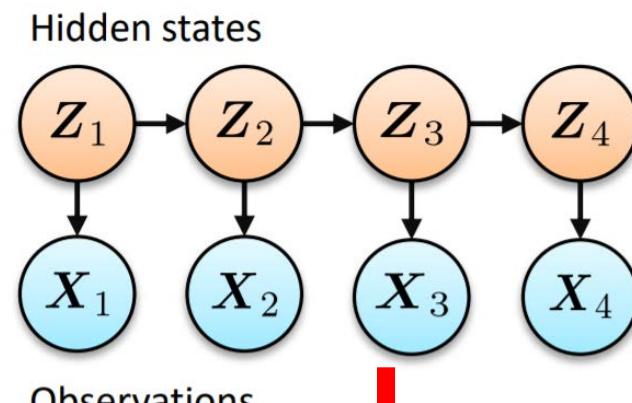
- Learn from complex data, including event times and order
- Learn from clinical annotations, codes, expertise etc.
- History matters! Non-stationary models needed
- Learn holistically! Multiple morbidities
- Heterogeneous patients – personalization matters
- Interpretable models

# First Clinically actionable model for patient-level trajectory!

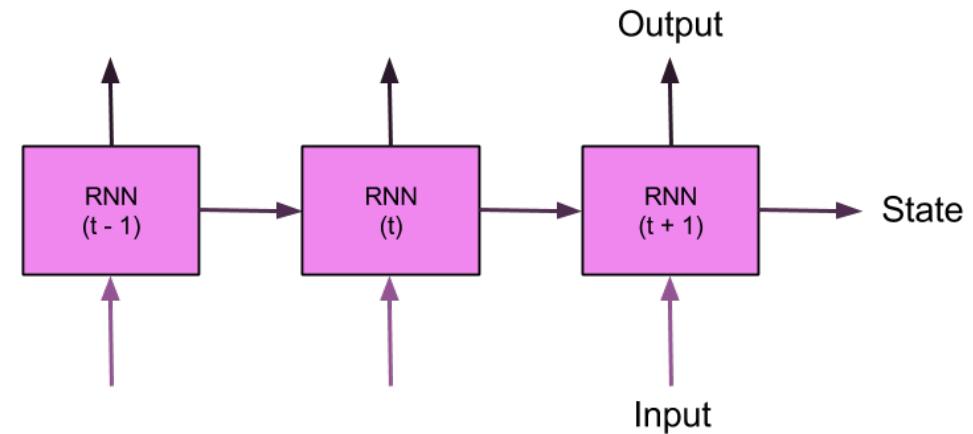
PASS [Alaa & van der Schaar, 2018]

Main idea: a general and versatile deep probabilistic model capturing complex, non-stationary representations for patient-level trajectories

Maintain probabilistic structure of HMMs



But use RNNs to model state dynamics



$$P(\{Z_m\}_m, \{X_m\}_m | Y, \{t_m\}_m) = \prod_{m'=1}^m P(X_{m'} | Z_{m'}) \cdot P(Z_{m'} | \mathcal{F}_{t_{m'-1}})$$

Emission

Transition

# PASS: Attentional State-Space!

- Model complex disease dynamics

$$f(Z_{n+1} | t_{n+1}, \mathcal{H}_{t_n})$$



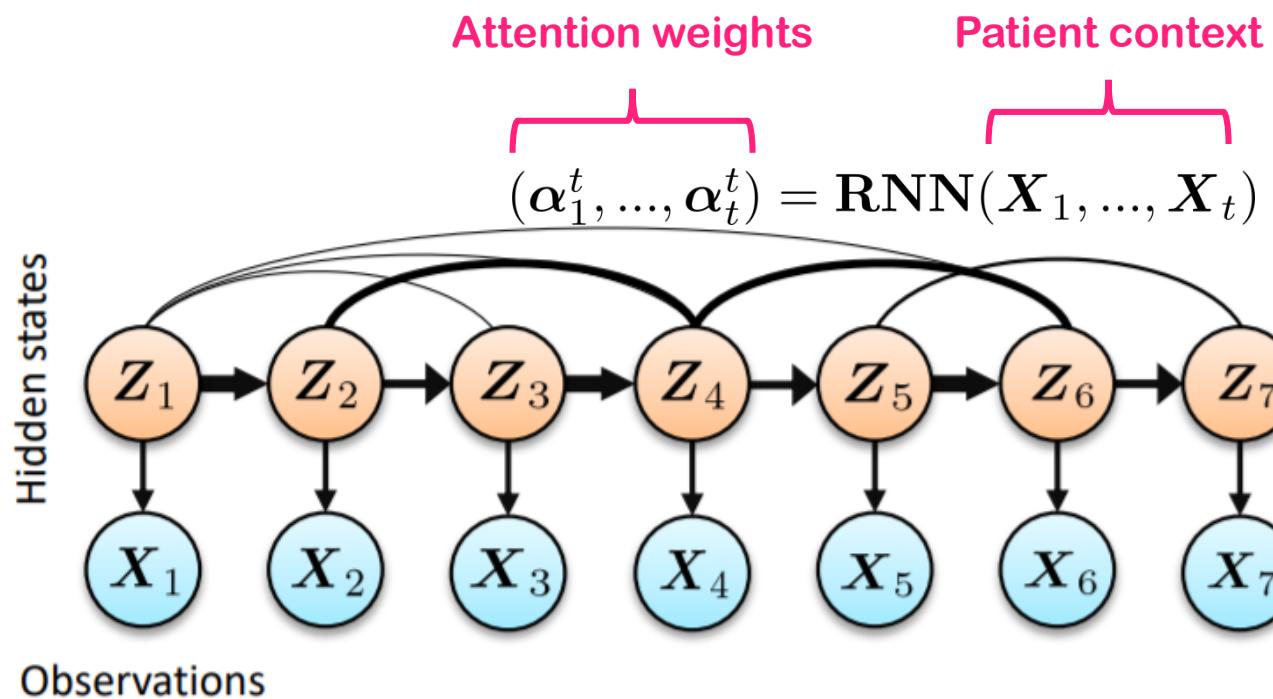
- Key idea: model complex dynamics on history as non-stationary interpretable state dynamics

$$f(Z_{n+1} | t_{n+1}, \mathcal{H}_{t_n}) = f(Z_{n+1} | g(\sum_{m=1}^n \alpha_m(t_{n+1}, \mathcal{H}_{t_n}) \cdot Z_m))$$



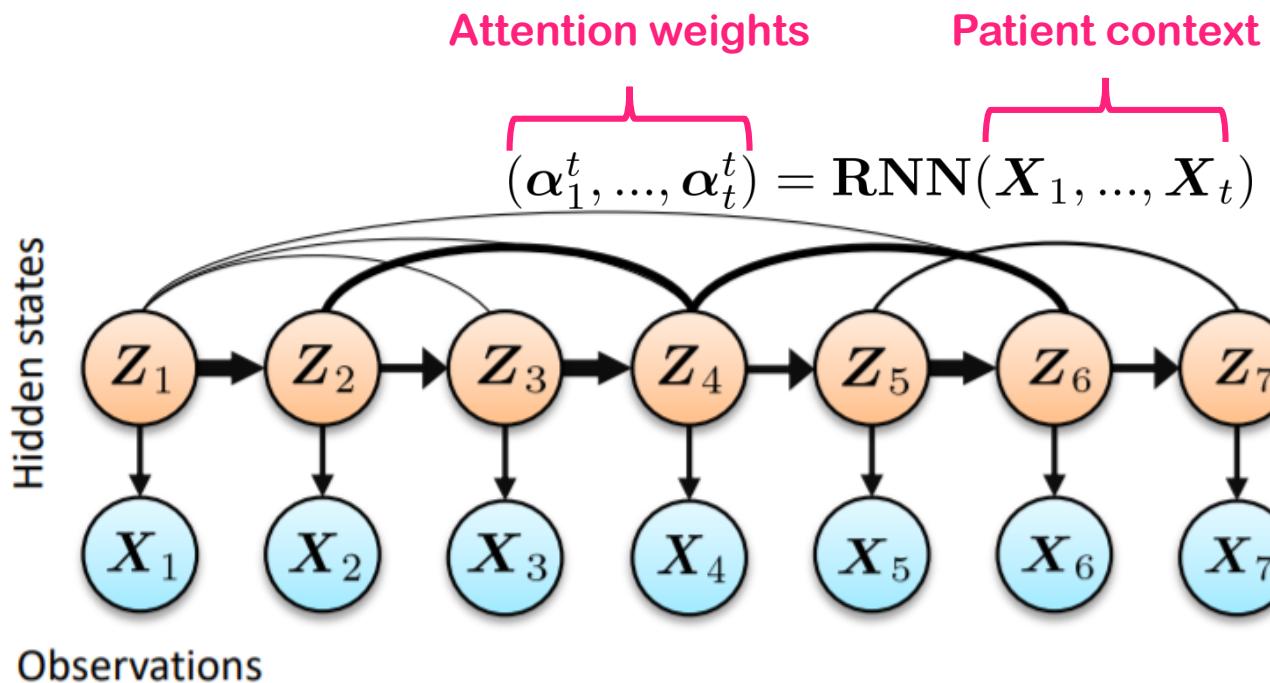
# PASS: Going Beyond Markov

- Attention weights determine the influences of past state realizations on future state transitions



# PASS: Overcomes shortcoming of Markov Models

Attention weights create a "soft" version of a non-stationary, variable-order Markov model where **underlying dynamics of a patient change over time based on an individual's clinical context!**

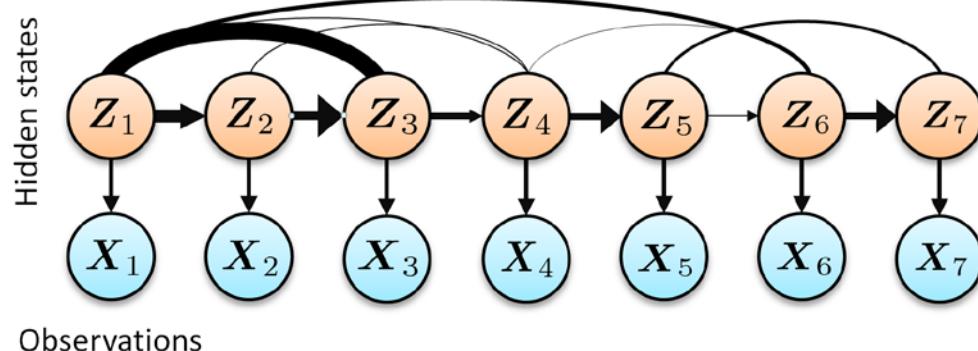
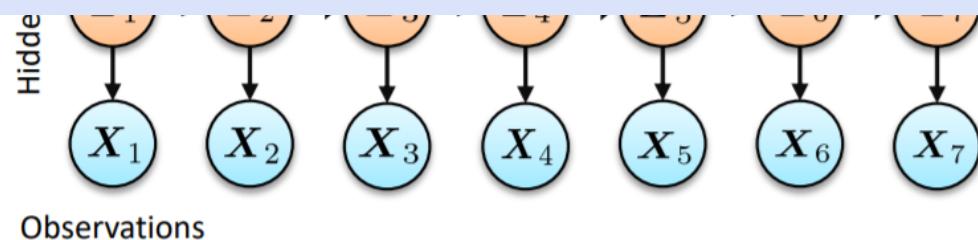


PASS “memory” is shaped by patient’s current context (clinical events, treatments, etc.)

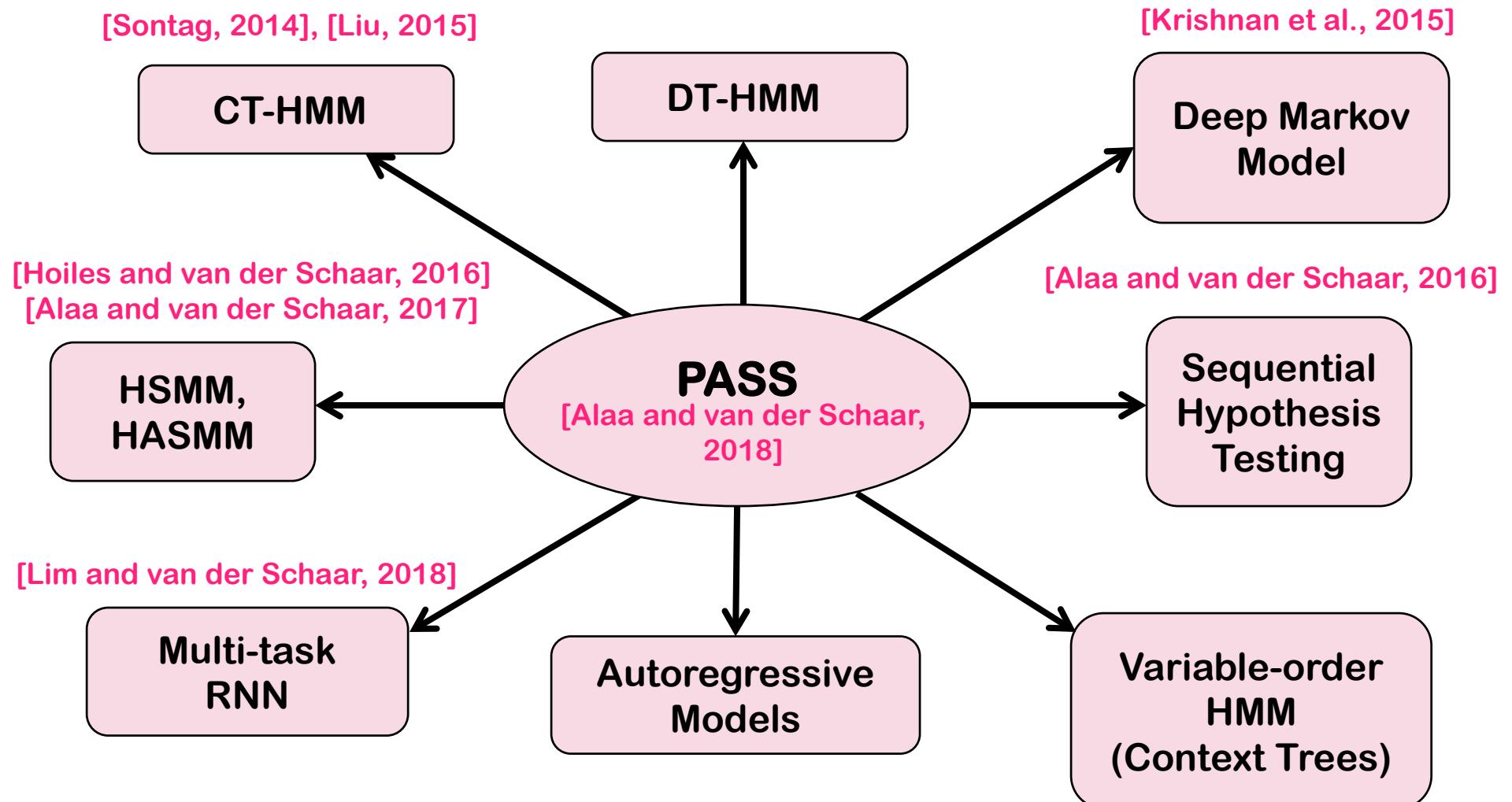
# PASS: Beyond One-size-fits-all using Contextual Attention

- Attentive state-space model: Individualized dynamics  
Individualization – two fold: Static + Dynamic Context

Attention weights explain causative and associative relationships between hidden disease states and past clinical events for *that* patient!

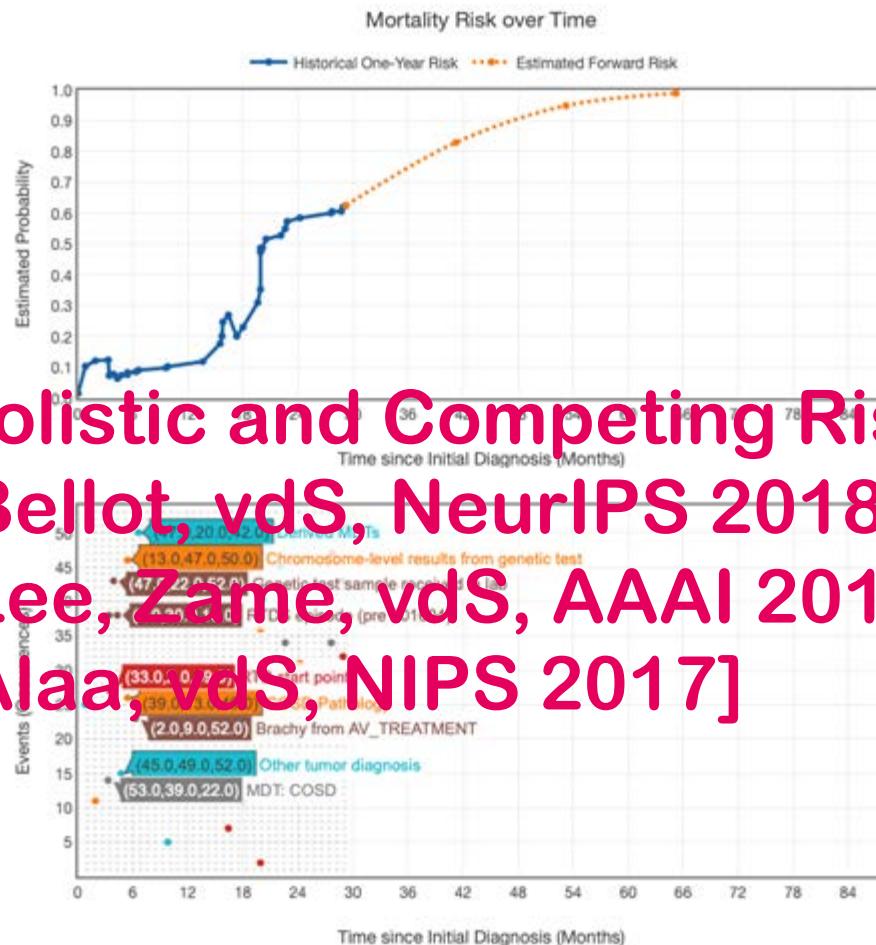


# PASS: A General, Versatile and Clinically Actionable Model

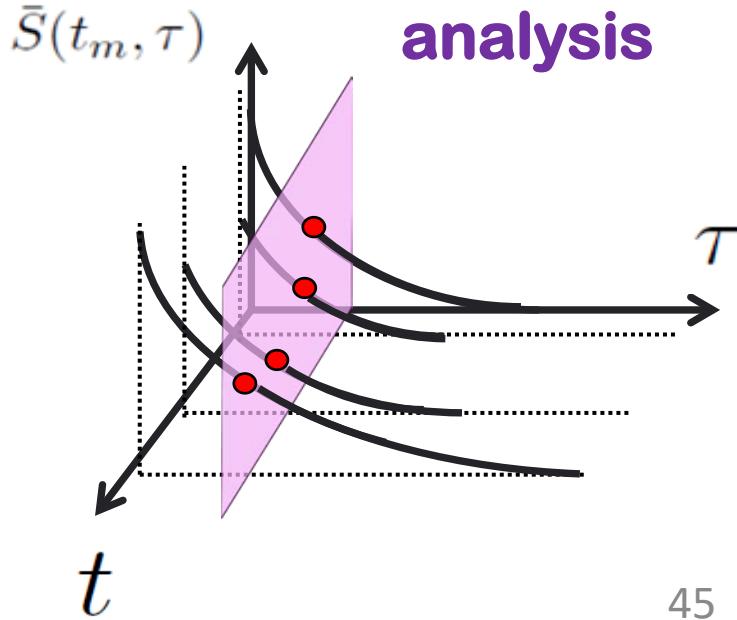


# PASS: A tool for decision support and discovery

**Dynamic and personalized forecasting** of the health and disease trajectory of a patient as data is gathered over time

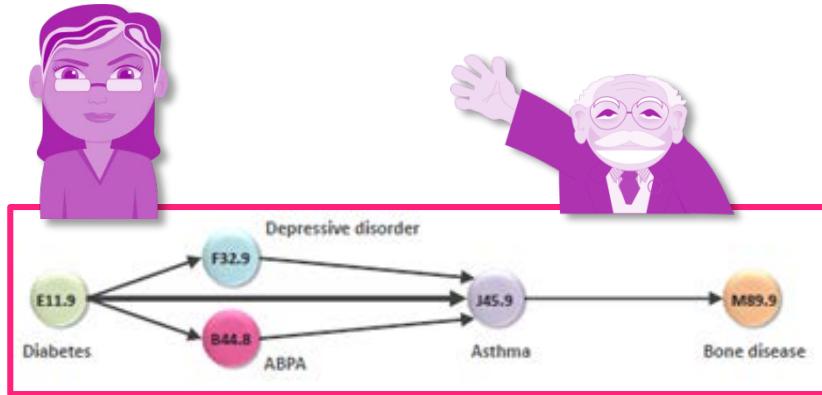


# Dynamic Time-to-event analysis

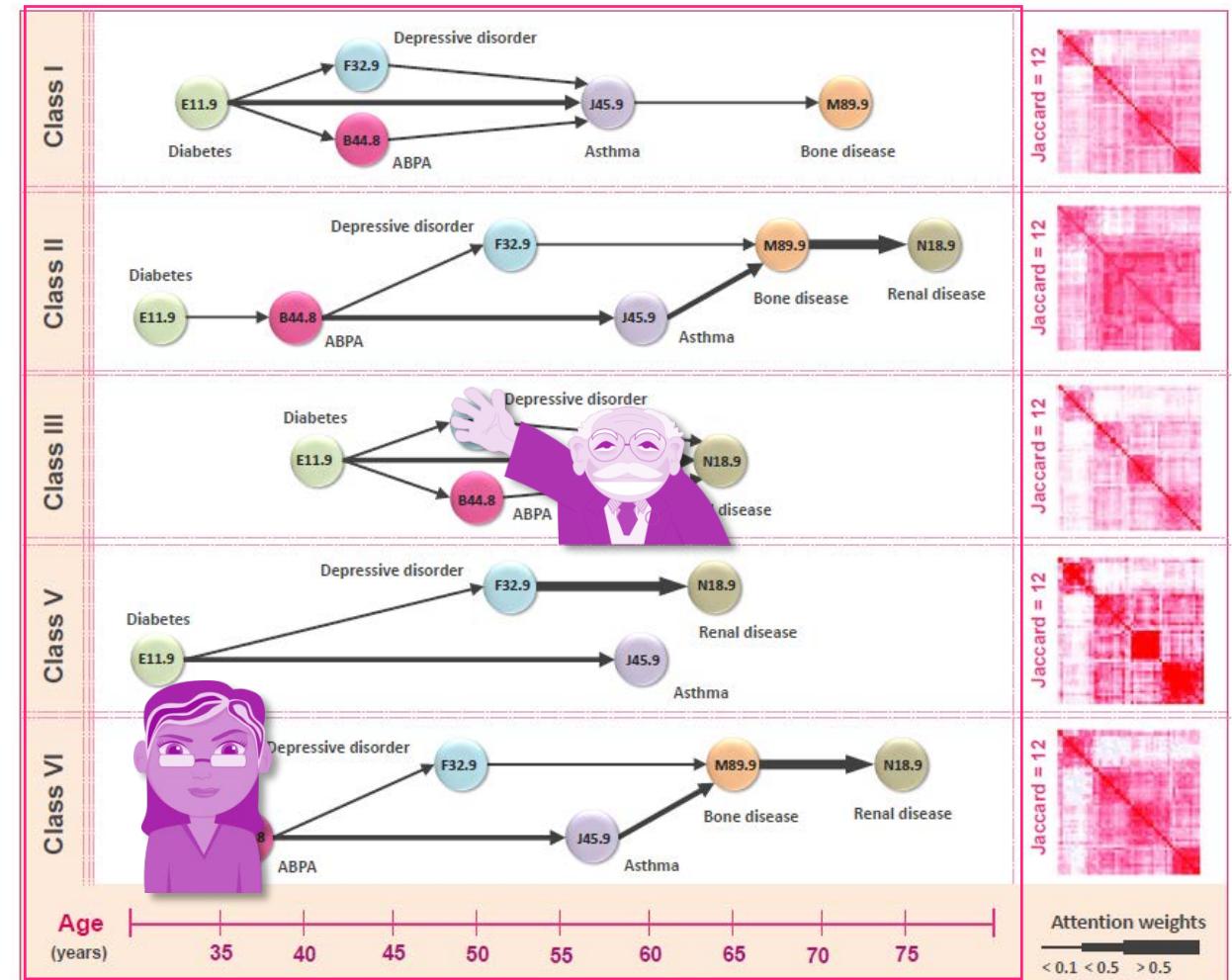


# Morbidity networks: Personalized

Population-level  
morbidity network

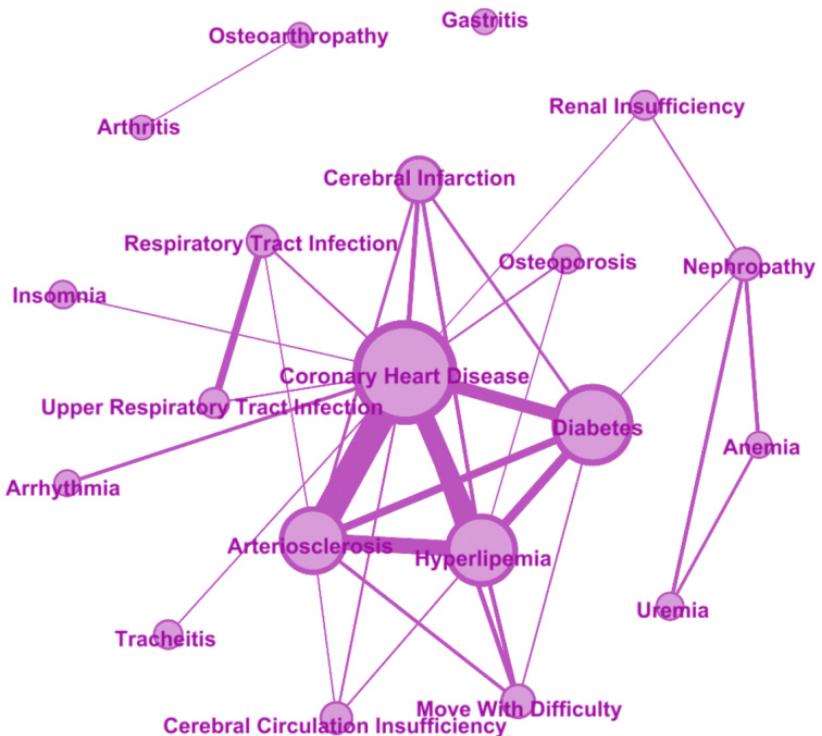


Personalized morbidity networks



# Morbidity networks: Dynamic

- **Dynamic Morbidity Maps** – Inferred based on attention weights
- **Causal associations:** how much attention is paid to diagnosis of morbidity A when predicting morbidity B

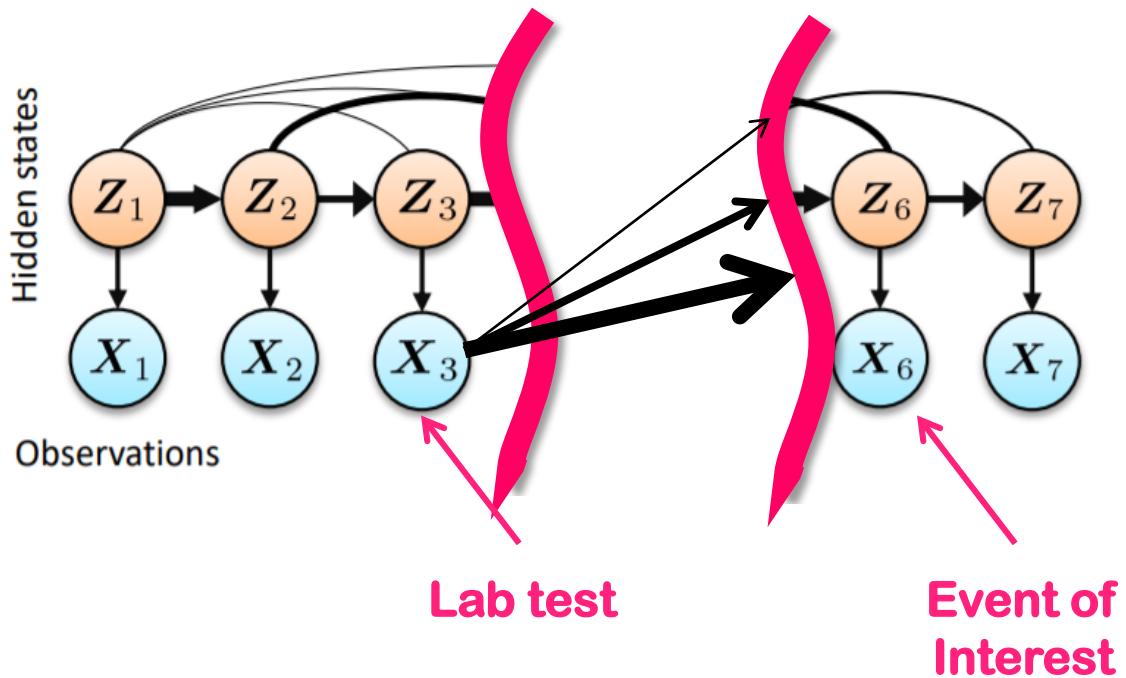


# Personalized Screening/Monitoring: Who to Screen? When to Screen? What to Screen?

Deep Sensing [Yoon, Zame, vdS, ICLR 2018]

Disease Atlas [Lim, vdS, ML4HC 2018]

Which Modality of Screening? [Alaa, Moon, Hsu, vdS, TMM 2016]



# Deep Sensing: Active Sensing using Multi-directional Recurrent Neural Networks [ICLR 2018]

- **Motivation:**

- Monitoring and screening (sensing) is costly
- Trade-off between value of information and cost of sensing
- Sensing should be an active choice

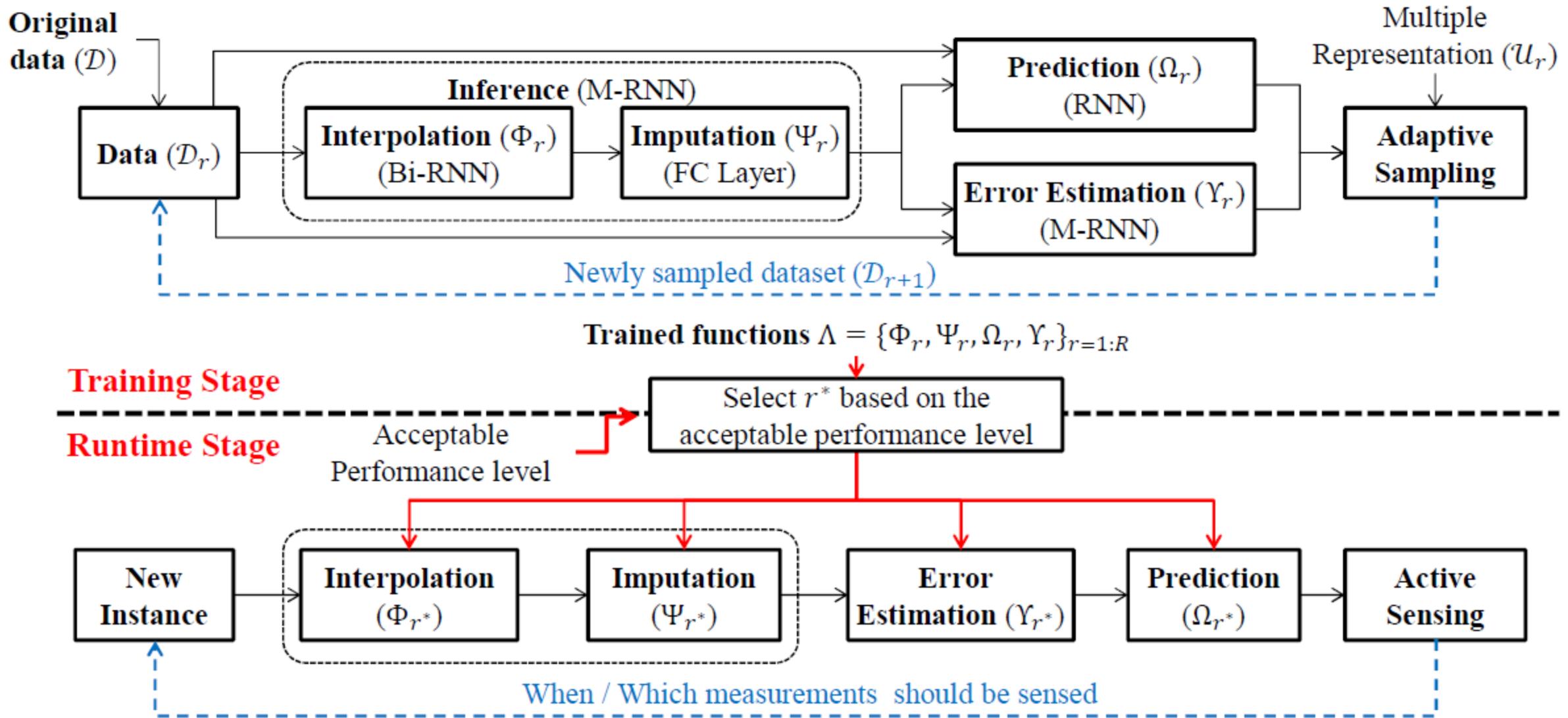
- **Challenges:**

- Value of information is unknown & dynamically changing – needs to be learned!

- **Ideas:**

- A neural network must learn – at training time – how to issue predictions at various cost-performance points.
- To do this, it creates multiple representations at various performance levels associated with different measurement rates (costs).
- Each representation is learned and constructed recursively and adaptively learned by deliberately introducing missing data

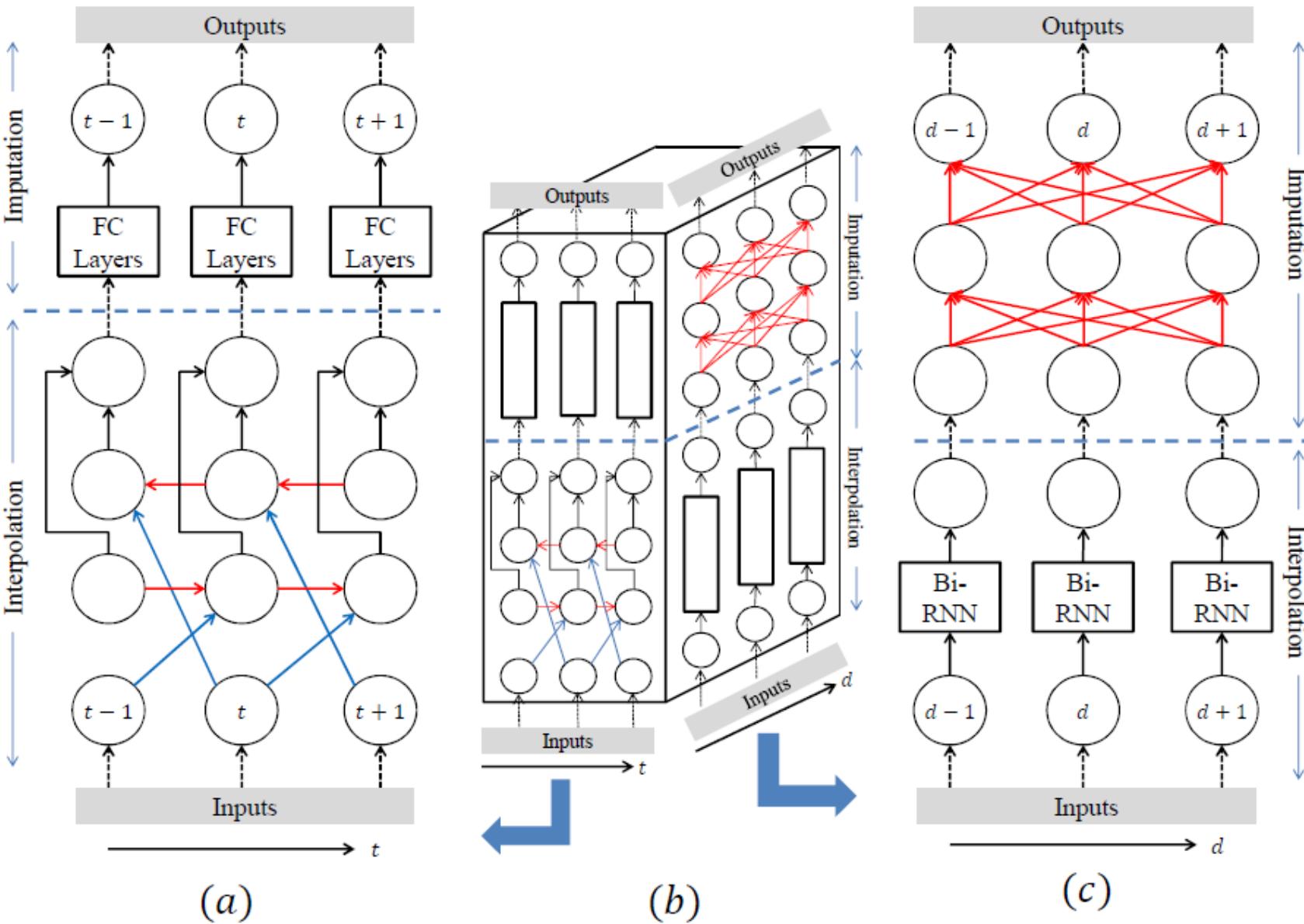
# Deep Sensing Architecture [ICLR 2018]



# M-RNN

An M-RNN differs from a bi-directional RNN:

- it sequentially operates across streams in addition to within streams
- the timing of inputs into the hidden layers is both lagged in the forward direction and advanced in the backward direction



Part 2:

## Personalized medicine needs to go beyond risk predictions- Individualized Treatment Recommendations

# Individualized Treatment Recommendations

Bob



Diagnosed with  
Disease X

Which treatment is best for Bob?

- Problem:  
Estimate the effect of a **treatment/intervention** on an **individual**

# RCTs do **not** support Personalized Medicine

**Randomized Control Trials:**  
**Average Treatment Effects**

Population-level



**Non-representative patients**

**Small sample sizes**

**Time consuming**

**Enormous costs**

**Adaptive Clinical Trials**

**[Atan, Zame, vdS, AISTATS 2019]**

# Delivering Personalized (Individualized) Treatments

**Randomized Control Trials:**  
**Average Treatment Effects**

Population-level



**Non-representative patients**  
**Small sample sizes**  
**Time consuming**  
**Enormous costs**

**Machine Learning:**  
**Individualized Treatment Effects**

Patient-centric



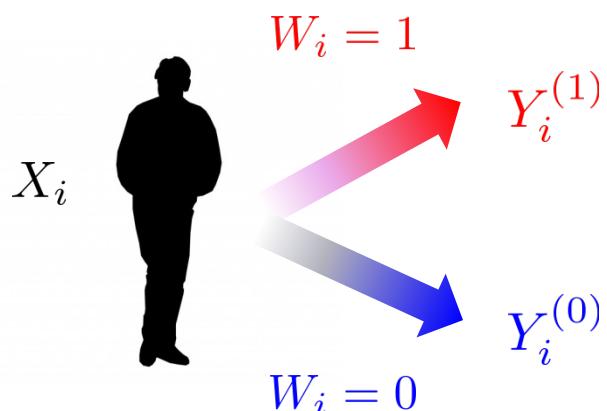
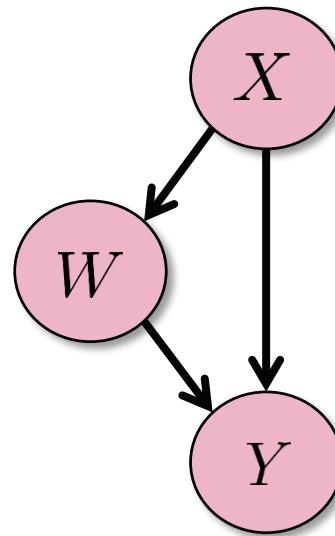
**Real-world observational data**  
**Scalable & adaptive implementation**  
**Fast deployment**  
**Cost-effective**

[Atan, vdS, 2015, 2018]  
[Alaa, vdS, 2017, 2018, 2019]  
[Yoon, Jordon, vdS, 2017]  
[Lim, Alaa, vdS, 2018]  
[Bica, Alaa, vdS, 2019]

# Potential outcomes framework [Neyman, 1923]

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



Factual outcomes

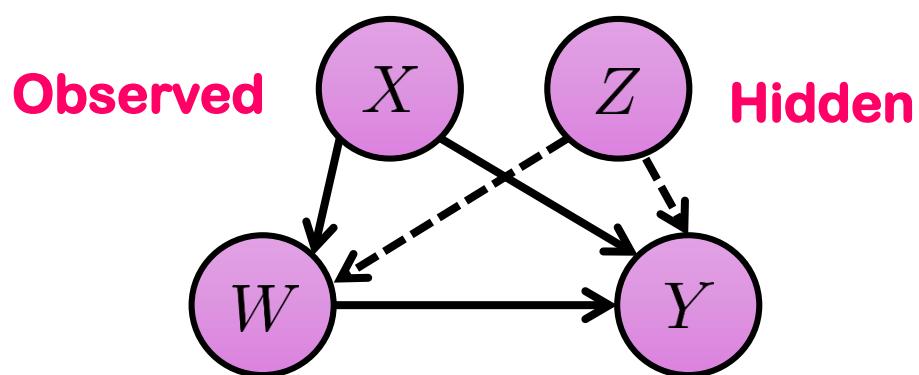
$$Y_i = W_i Y_i^{(1)} + (1 - W_i) Y_i^{(0)}$$

Causal effects

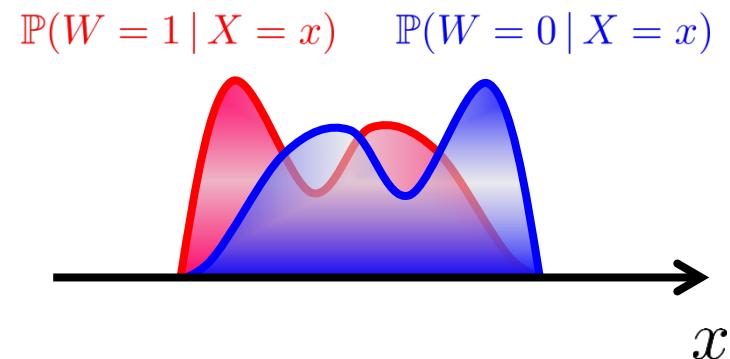
$$T(x) = \mathbb{E} \left[ Y_i^{(1)} - Y_i^{(0)} \mid X_i = x \right]$$

# Assumptions

No unmeasured  
confounders (Ignorability)



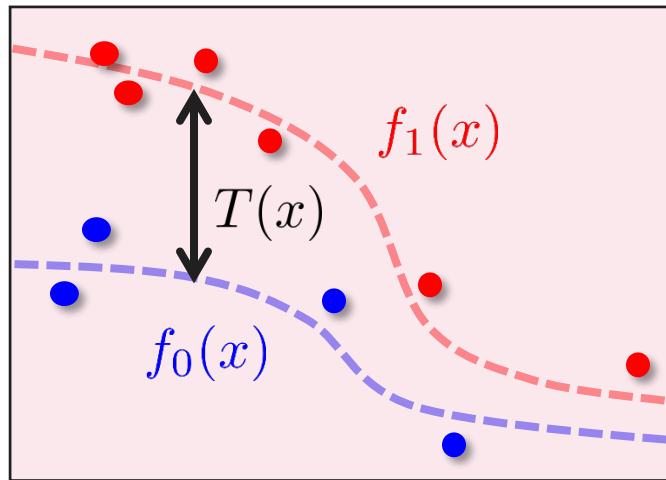
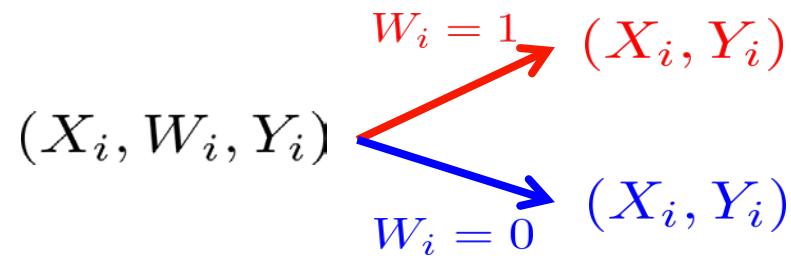
Common support



Our work on hidden confounders  
[Lee, Mastronarde, vdS, 2018]  
[Bica, Alaa, vdS, 2019]

# Estimating individualized treatment effects

- Observational data



- Treatment response surfaces

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

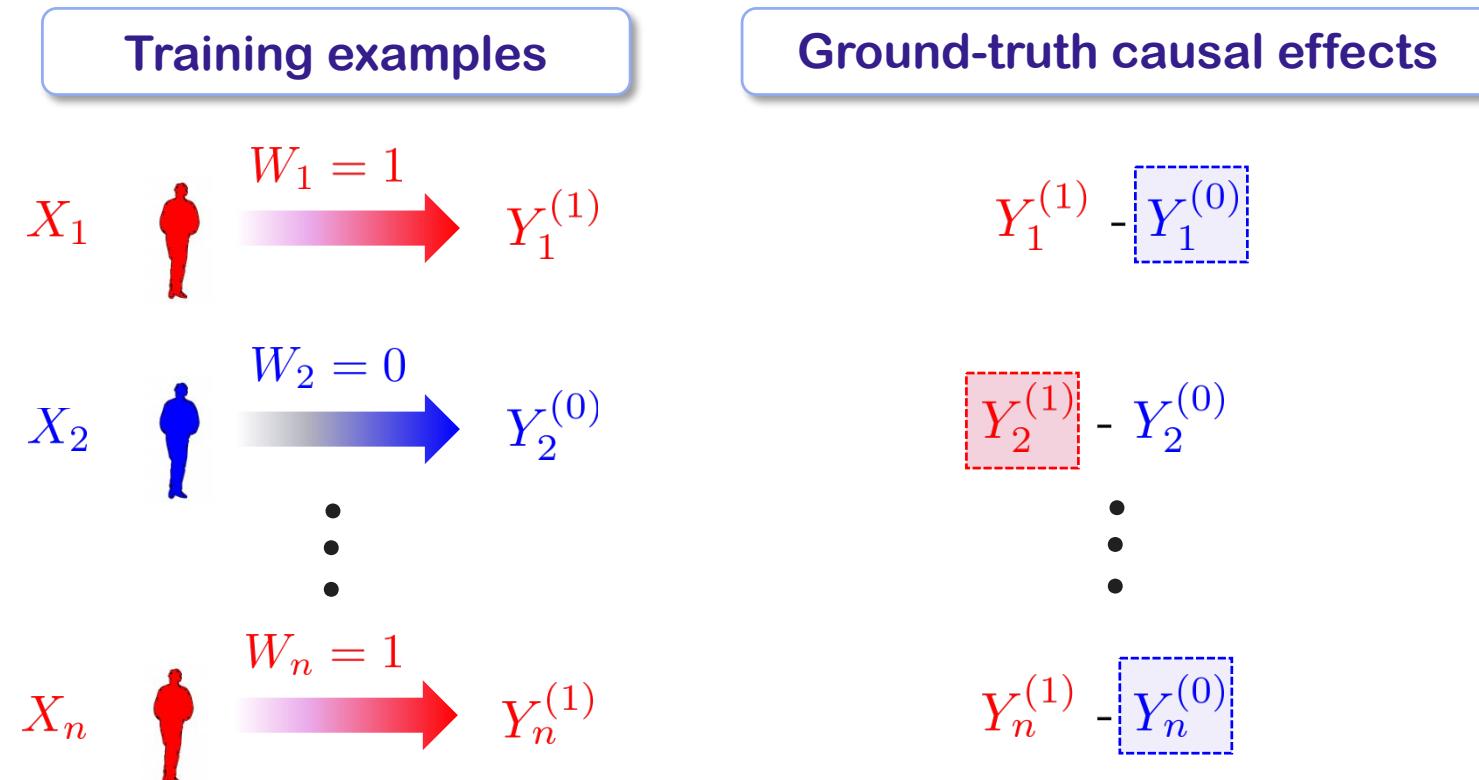
$$f_0(x) = \mathbb{E}[Y^{(0)} | X = x]$$

- Estimate causal effects: individualized treatment effects

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

# Beyond supervised learning...

- Fundamental challenge of causal inference:  
we never observe **counterfactual** outcomes

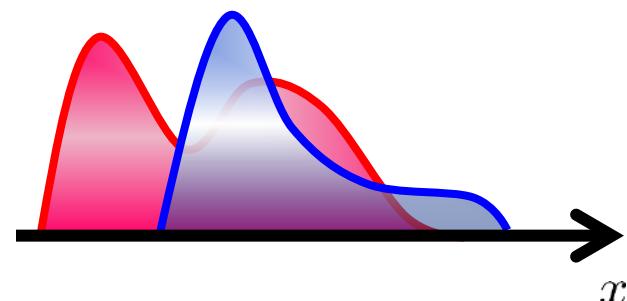


# Causal modeling $\neq$ predictive modeling

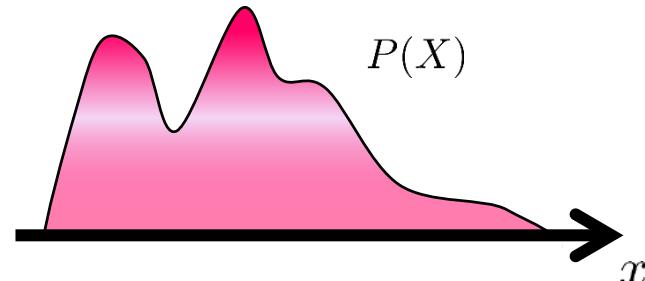
1- Need to model interventions  $(X_i, W_i, Y_i)$

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

$$P(X | W = 1) \quad P(X | W = 0)$$



Training distribution



Testing distribution

# Many recent works on individualized treatment effects (ITEs)

- Bayesian Additive Regression Trees (BART) [Chipman et. al, 2010], [J. Hill, 2011]
- Causal Forests [Wager & Athey, 2016]
- Nearest Neighbor Matching (kNN) [Crump et al., 2008]
- Balancing Neural Networks [Johansson, Shalit and Sontag, 2016]
- Causal MARS [Powers, Qian, Jung, Schuler, N. Shah, T. Hastie, R. Tibshirani, 2017 ]
- Targeted Maximum Likelihood Estimator (TMLE) [Gruber & van der Laan, 2011]
- Counterfactual regression [Johansson, Shalit and Sontag, 2016]
- CMGP [Alaa & van der Schaar, 2017]
- GANITE [Yoon, Jordon & van der Schaar, 2018]

No theory, ad-hoc models

# A first theory for causal inference - individualized treatment effects

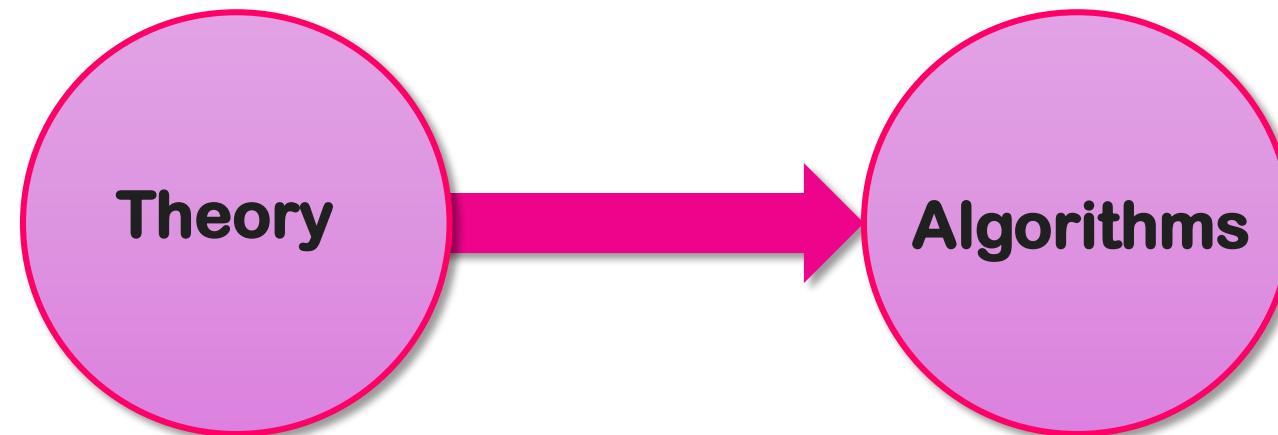
[Alaa, vdS, JSTSP 2017][ICML 2018]

**What is possible?**

(Fundamental limits)

**How can it be achieved?**

(Practical implementation)



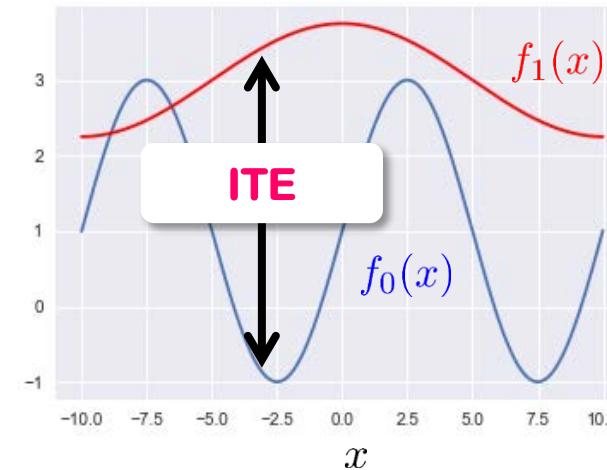
# Bayesian nonparametric ITE estimation

- True ITE model

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

- ITE estimation

- Prior over response functions:  $f_0, f_1 \sim \Pi$
- Point estimator  $\hat{T}(\cdot)$  induced by Bayesian posterior  $d\Pi_n(T \mid \mathcal{D})$
- Precision of estimating heterogeneous effects  $\text{PEHE}(\hat{T}) \triangleq \mathbb{E} \| \hat{T} - T \|_{L^2(\mathbb{P})}^2$



# Minimax Rate for ITE Estimation

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

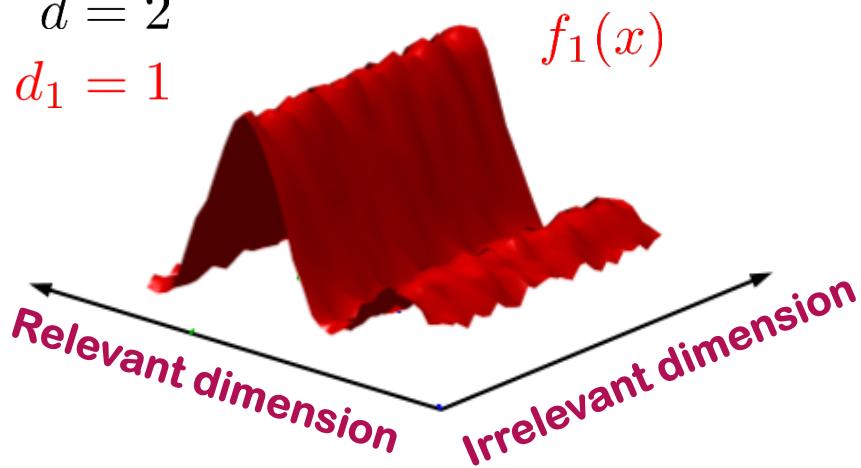
Sparsity  $d$

$f_0(x) \rightarrow d_0$  relevant dimensions

$f_1(x) \rightarrow d_1$  relevant dimensions

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

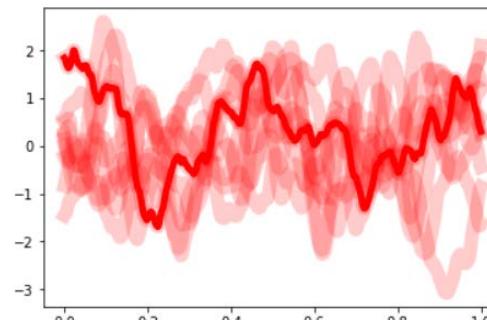
$$\begin{aligned} d &= 2 \\ d_1 &= 1 \end{aligned}$$



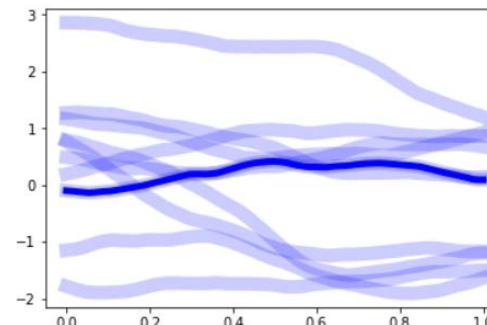
Smoothness  $\alpha$

$f_0(x) \rightarrow$  Hölder ball  $H^{\alpha_0}$

$f_1(x) \rightarrow$  Hölder ball  $H^{\alpha_1}$



$\alpha_1 \downarrow\downarrow$   
Rough  
functions



$\alpha_0 \uparrow\uparrow$   
Smooth  
functions

# Minimax Rate for ITE Estimation

- **Theorem 1**

**The minimax rate for ITE estimation is given by:**

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

# Should we care about selection bias?

- Assume that  $\alpha_0 = \alpha_1$  and  $d_0 = d_1$

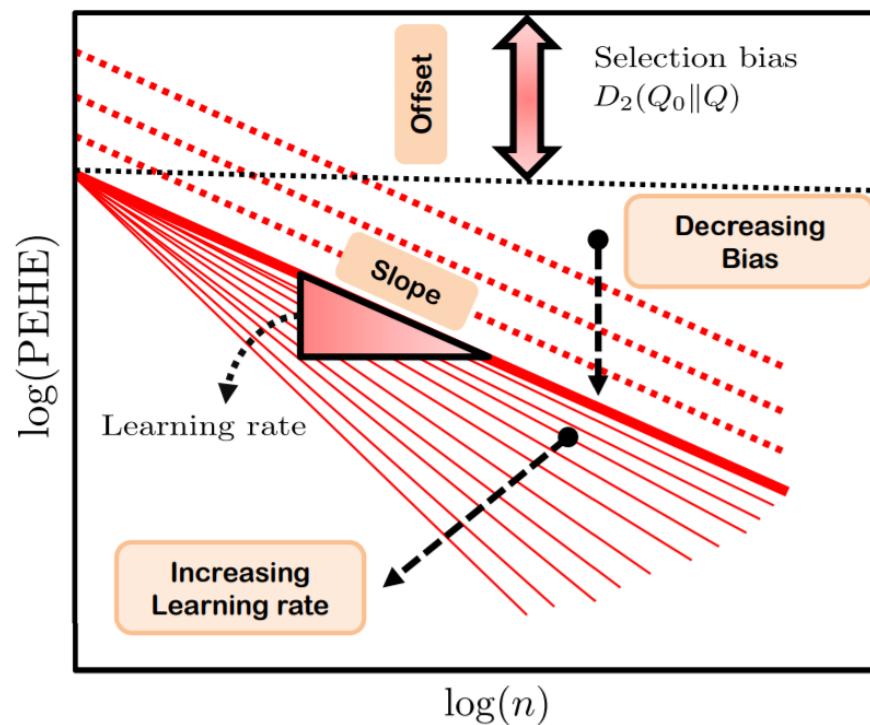
Minimax-optimal estimator

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

Slope

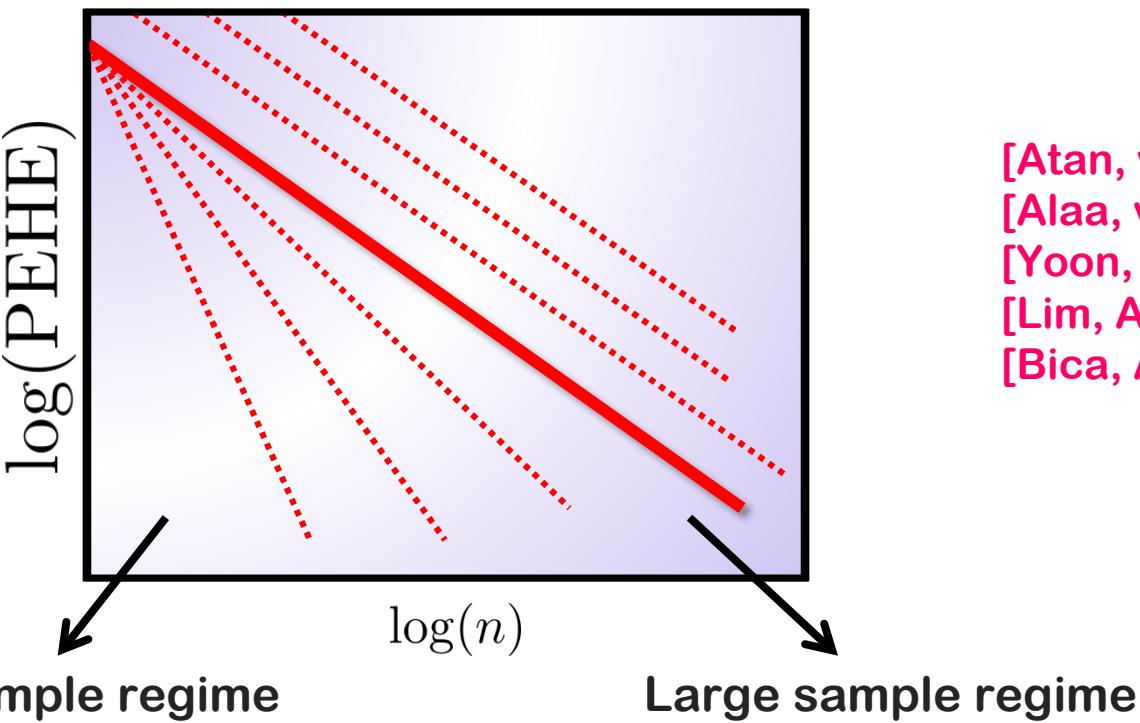
Rényi  
Divergence

Offset



# Theory guides model design

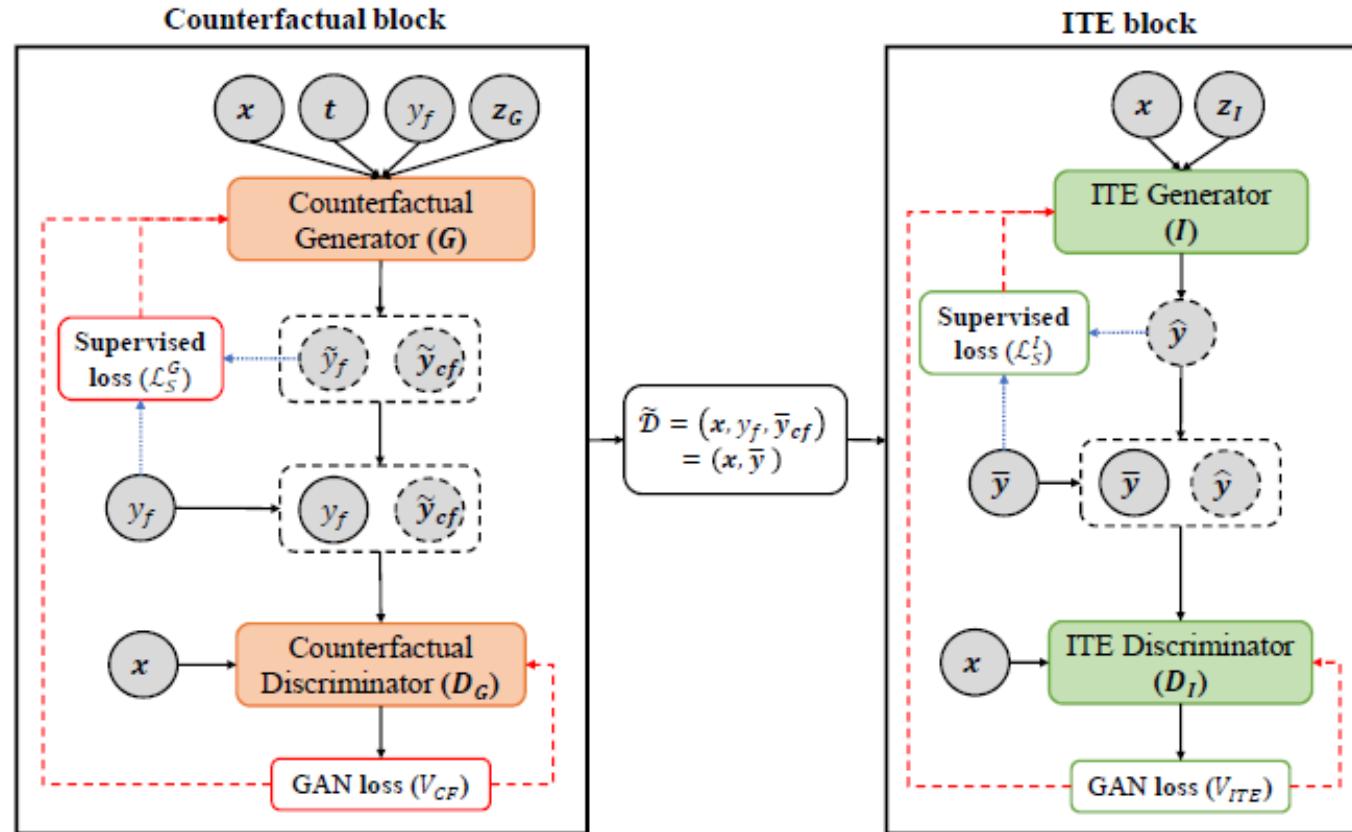
- We want models that do well in both small and large sample regimes



- Handling selection bias
- Sharing training data between response surfaces
- Flexible model and hyperparameter tuning

[Atan, vdS, 2015, 2018]  
[Alaa, vdS, 2017, 2018, 2019]  
[Yoon, Jordon, vdS, 2018]  
[Lim, Alaa, vdS, 2018]  
[Bica, Alaa, vdS, 2019]

# GANITE: Estimation of Individualized Treatment Effects using Generative Adversarial Nets [ICLR 2018]



**Intuition:**

If generated counterfactuals follow the underlying distribution, it should not be possible to discriminate real and generated outcomes.

## Our Approach:

- Develop a **new conditional GAN** framework to **generate** counterfactual outcomes conditioned on real outcomes
- Train a **new type of GAN** to generate the full potential outcome distribution conditioned only on the features

# GANITE: Counterfactual imputation block

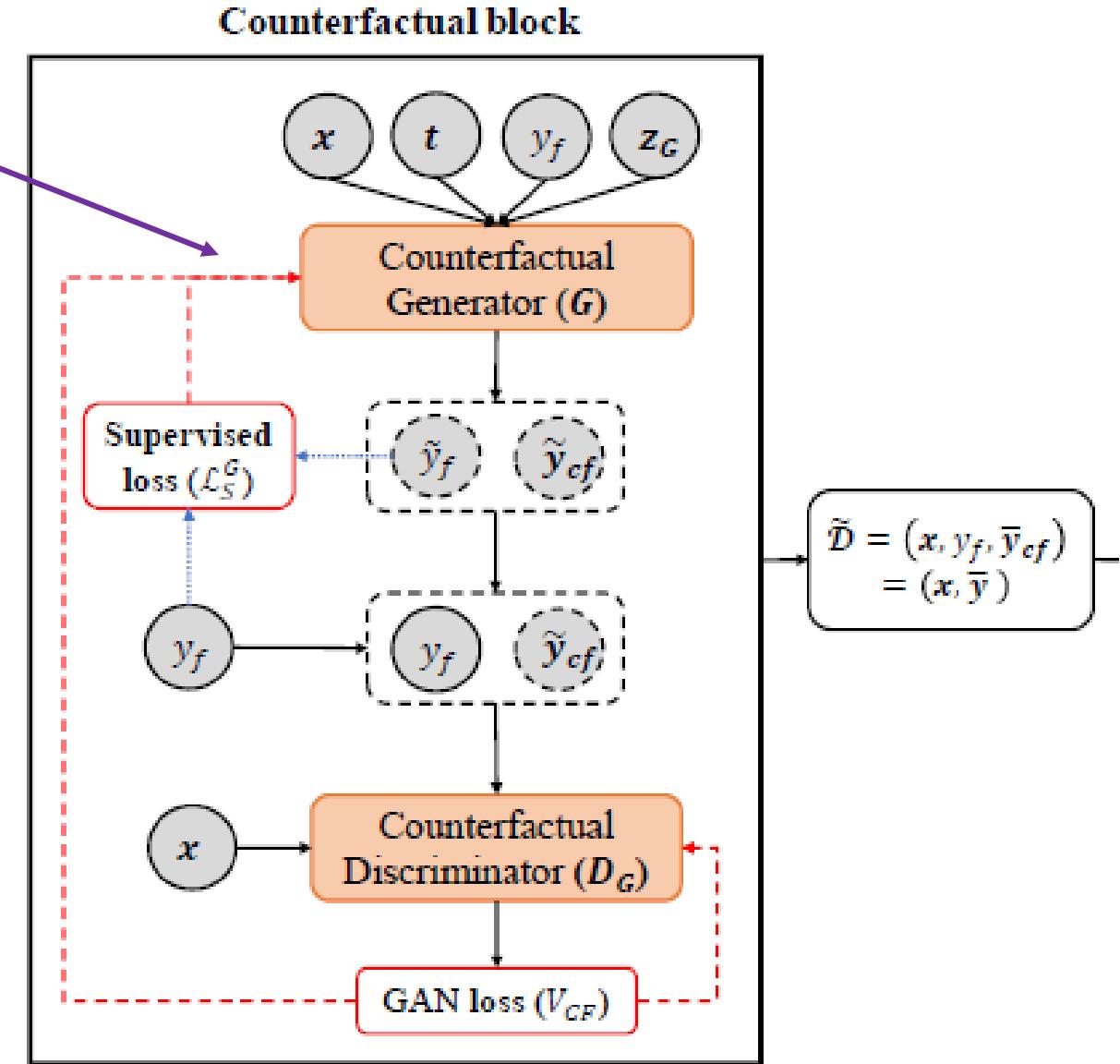
## Counterfactual Generator

**Inputs:**  $X, Y_f, T$ , and noise  $Z$

**Outputs:** Factuals  $\hat{Y}_f$  and counterfactuals  $\hat{Y}_{cf}$

## Supervised loss:

Forces generated factual outcome to be close to the actually observed factual outcome



# GANITE: Counterfactual imputation block

## Counterfactual Generator

**Inputs:**  $X, Y_f, T$ , and noise  $Z$

**Outputs:** Factuals  $\hat{Y}_f$  and counterfactuals  $\hat{Y}_{cf}$

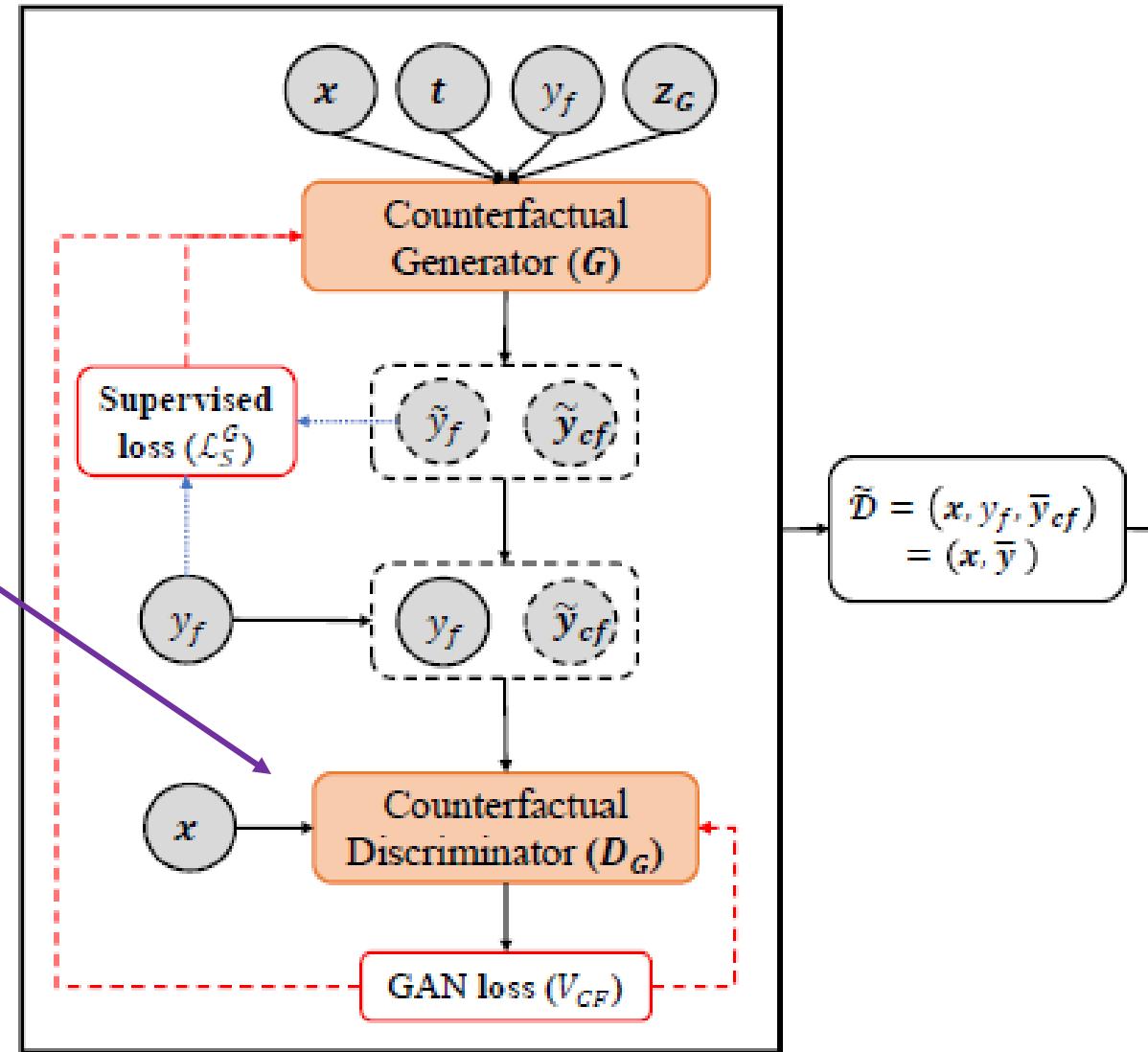
## Counterfactual Discriminator

**Inputs:**  $X, Y_f, \hat{Y}_{cf}$

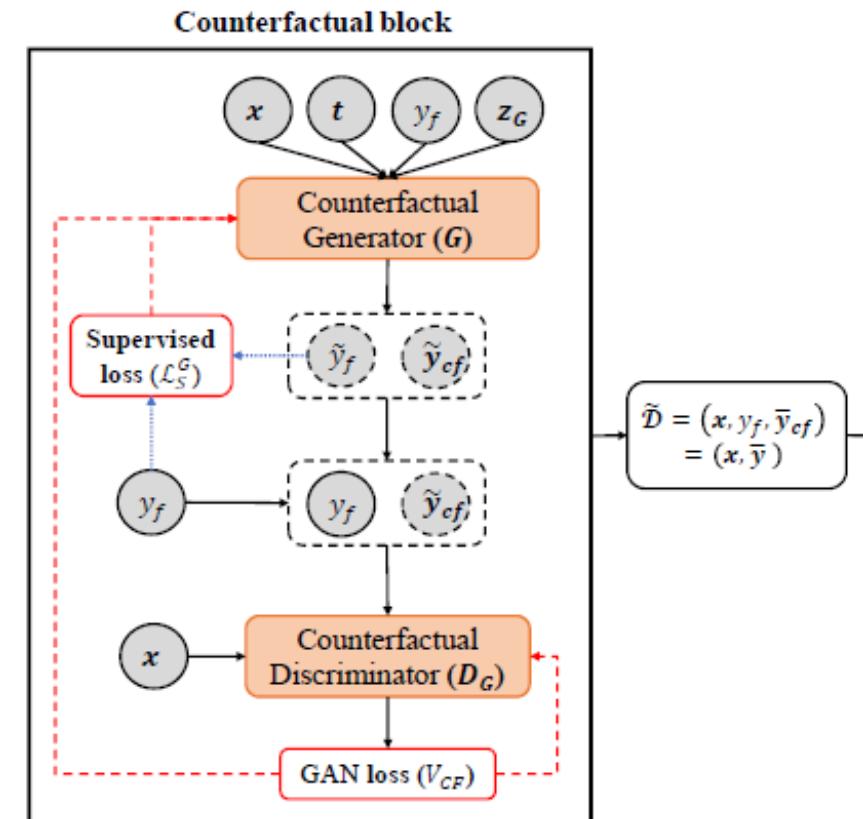
**Outputs:** Treatment prediction,  $\hat{T}$

Discriminator attempts to **distinguish** which outcome was the **true outcome**

## Counterfactual block



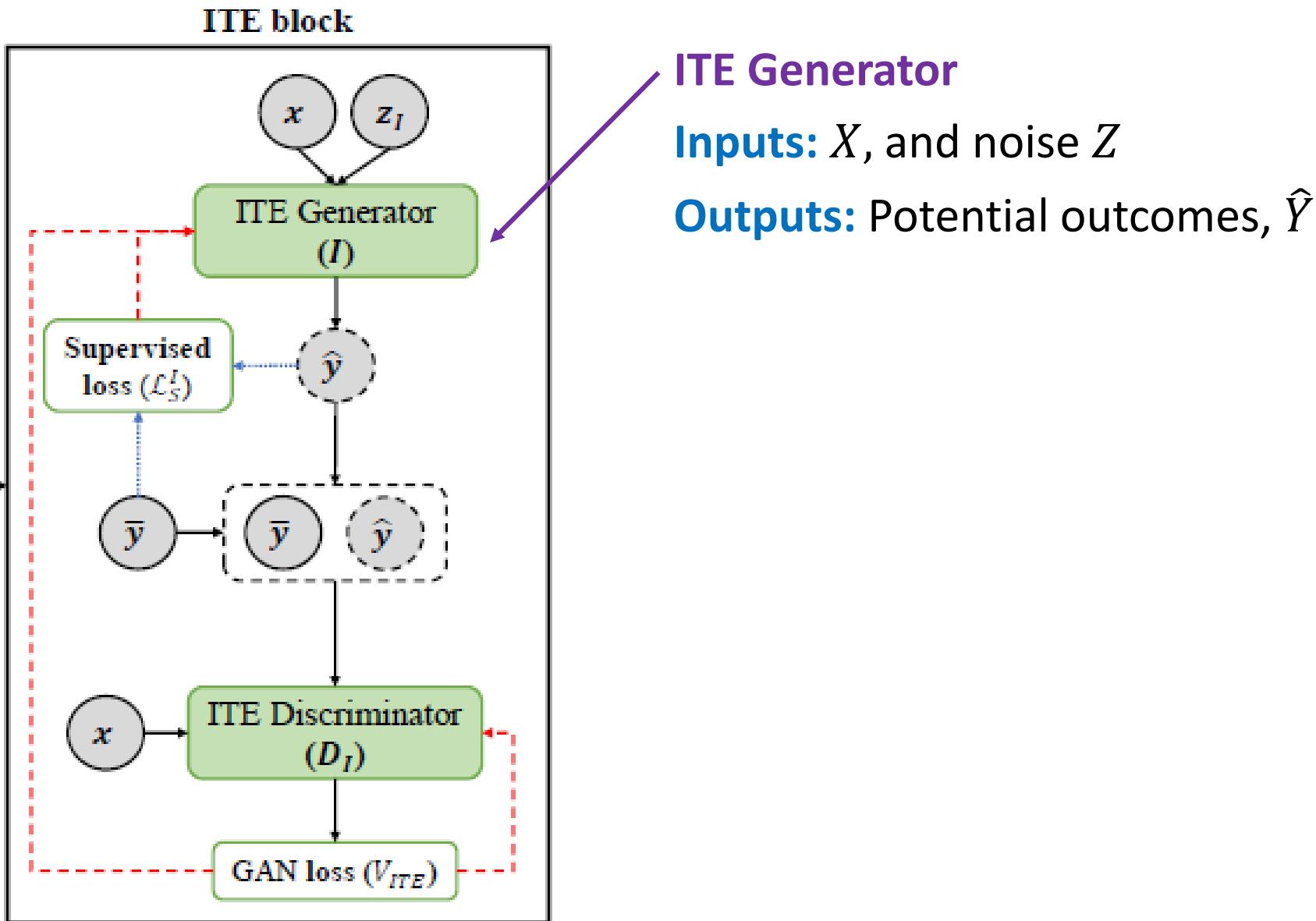
# GANITE: Differences to standard GAN



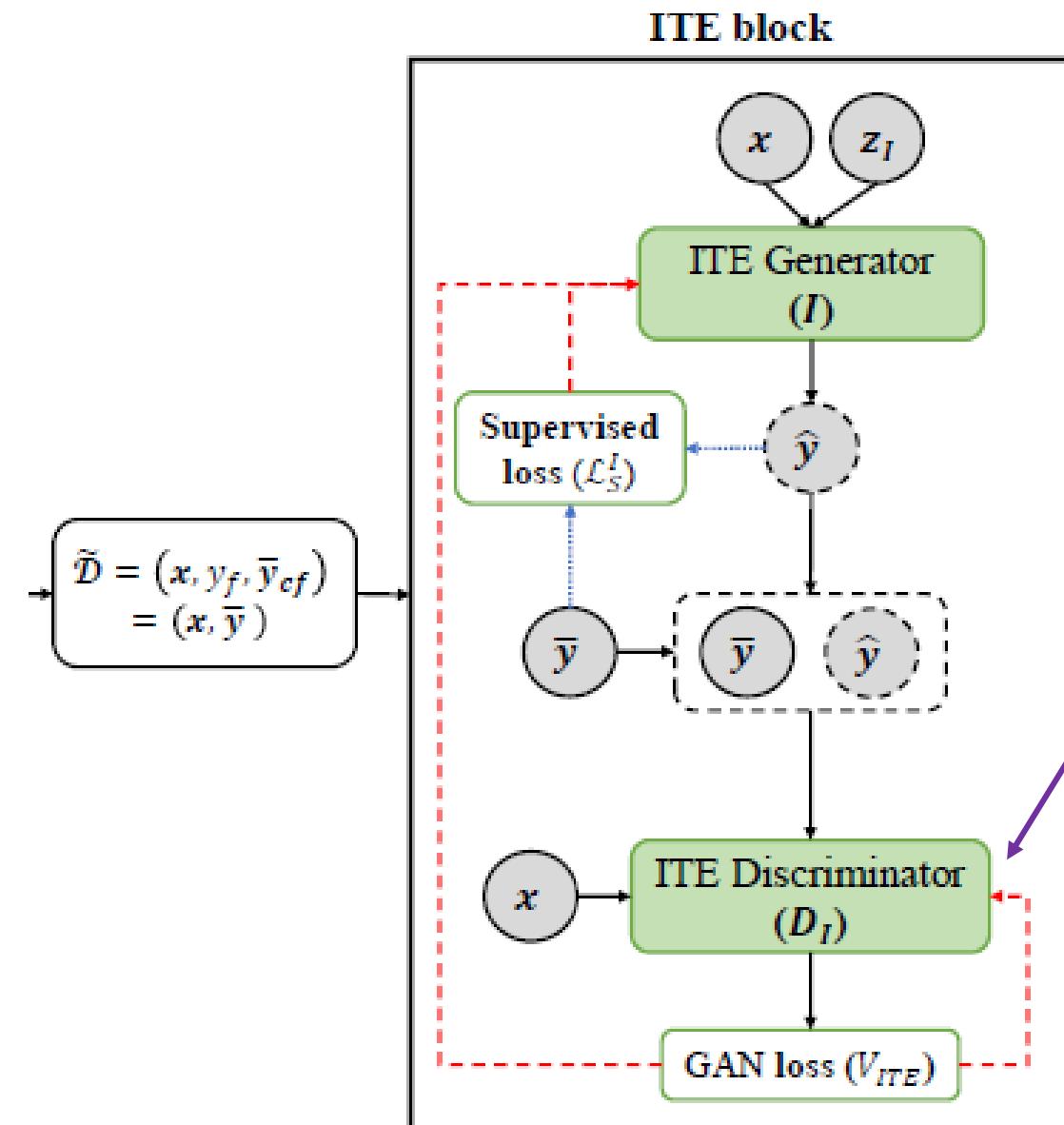
In standard GAN, the discriminator is distinguishing between 2 distributions and any given input comes from exactly 1 of the distributions.

In GANITE, the discriminator is trying to untangle 2 distributions that have been mixed together to form a new joint distribution. The input to the discriminator always consists of this mixture.

# GANITE: Components



# GANITE: Components



## ITE Generator

**Inputs:**  $X$ , and noise  $Z$

**Outputs:** Potential outcomes,  $\hat{Y}$

## ITE Discriminator

**Inputs:**  $X, \hat{Y}$  or  $Y$

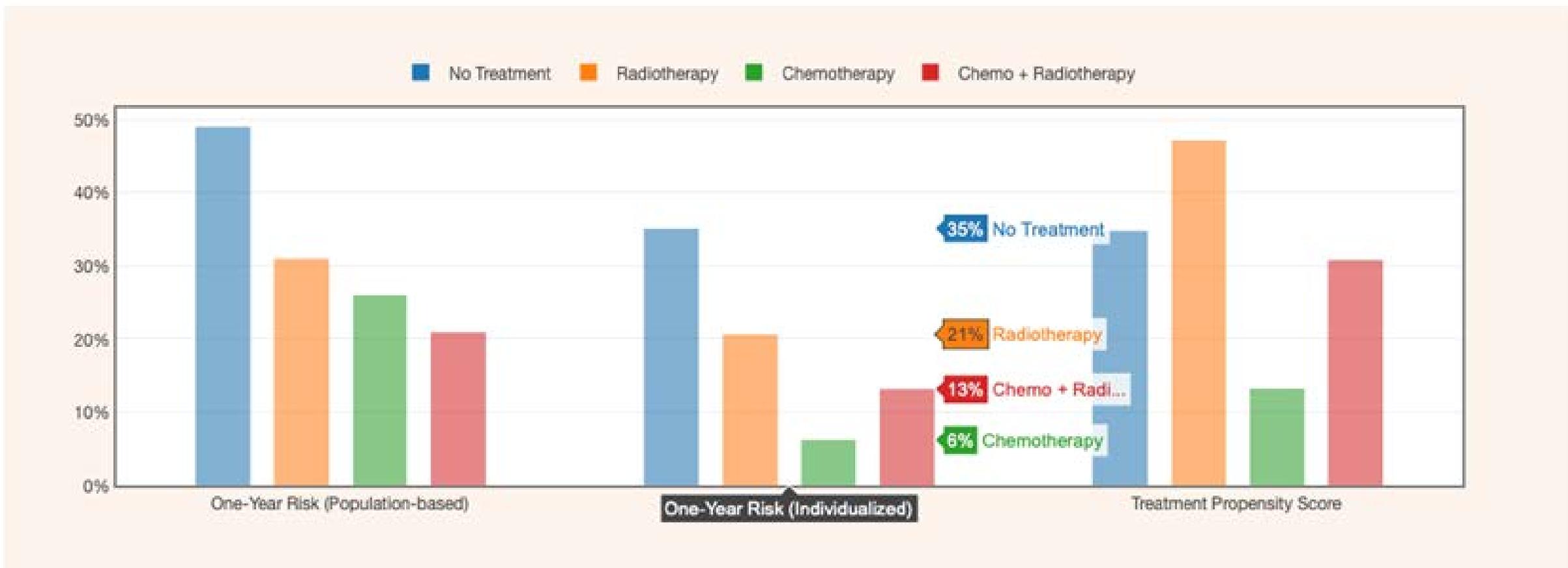
**Outputs:** Real/fake prediction

Discriminator attempts to **distinguish** whether outcomes came from counterfactual generator or from ITE generator

# Multiple Treatments: GANITE [Yoon, Jordon, vdS, ICLR 2018]

## Estimation of Individualized Treatment Effects using Generative Adversarial Nets

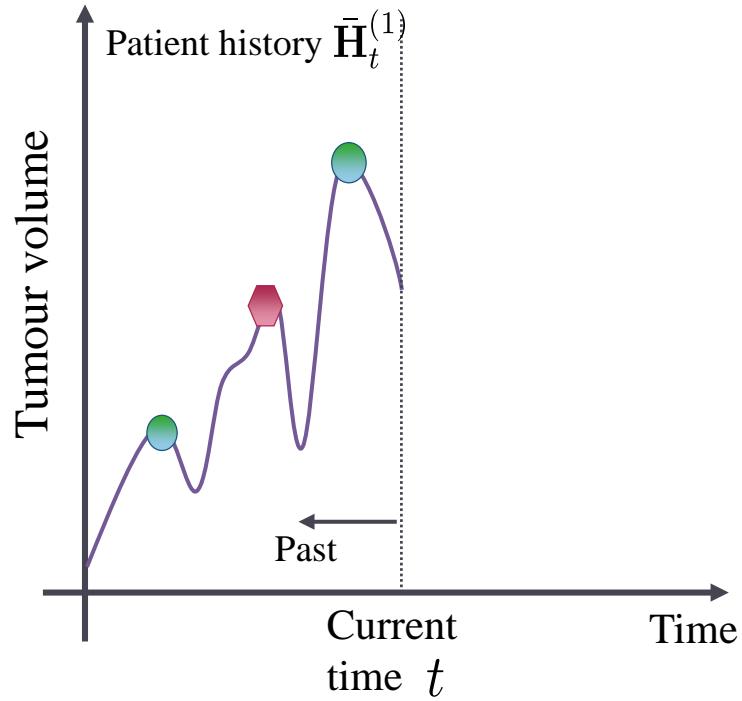
Risk of Recurrence vs. Treatment Options



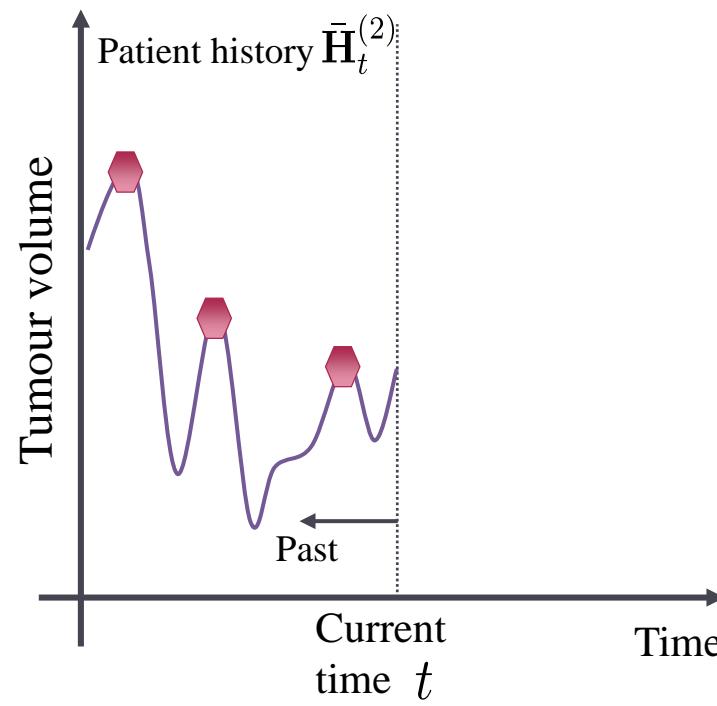
# Individualized Treatment Effects over Time

[Lim, Alaa, vdS, NeurIPS 2018][Bica, Alaa, vdS, 2019]

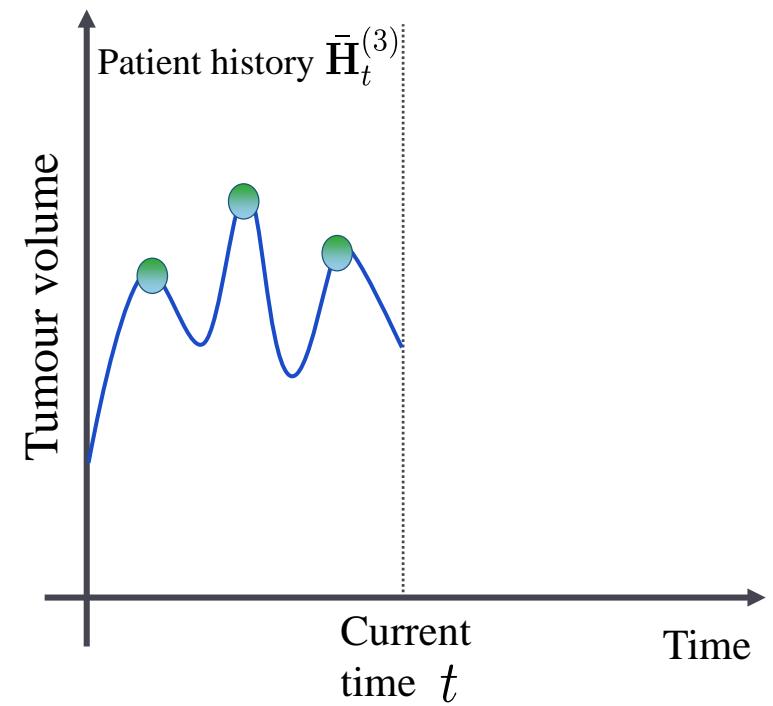
● Chemotherapy      ♦ Radiotherapy



(a) Decide treatment plan



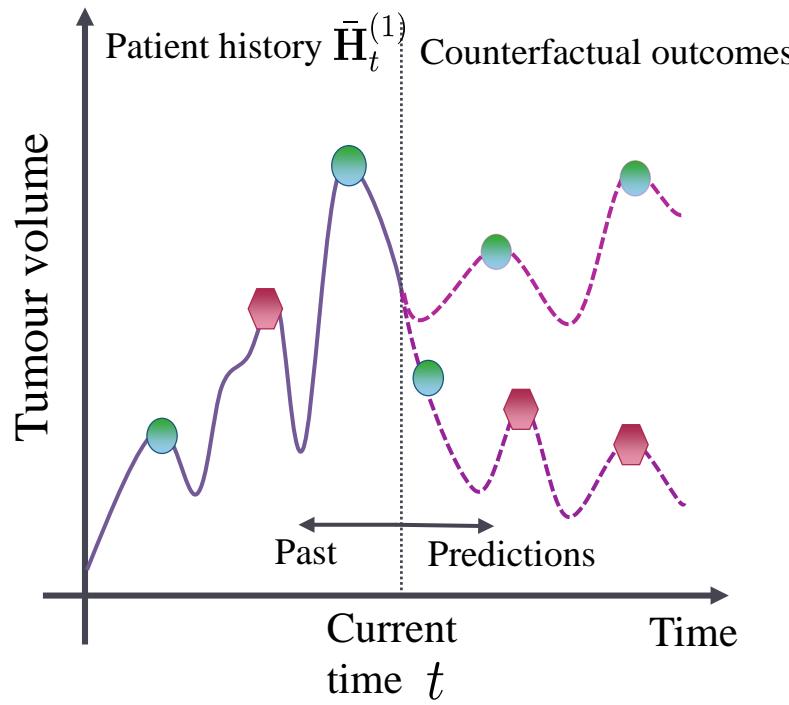
(b) Decide optimal time of treatment



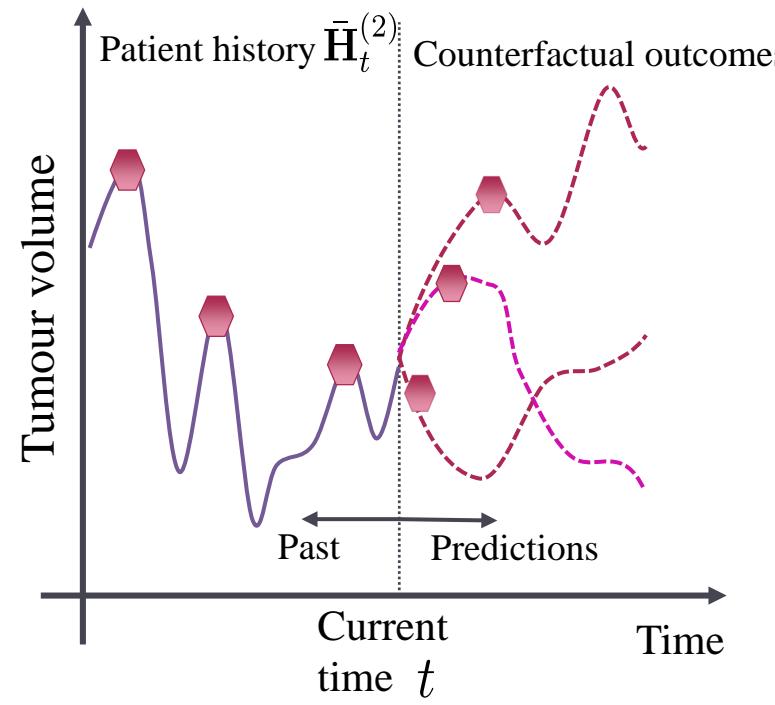
(c) Decide when to stop treatment

# When to treat? How to treat? When to stop?

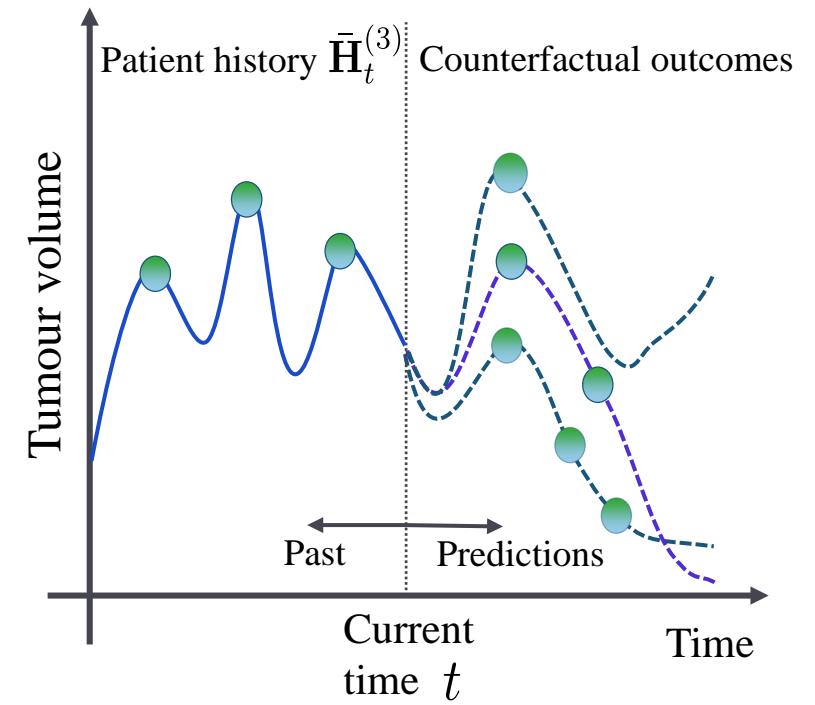
● Chemotherapy      ♦ Radiotherapy



(a) Decide treatment plan



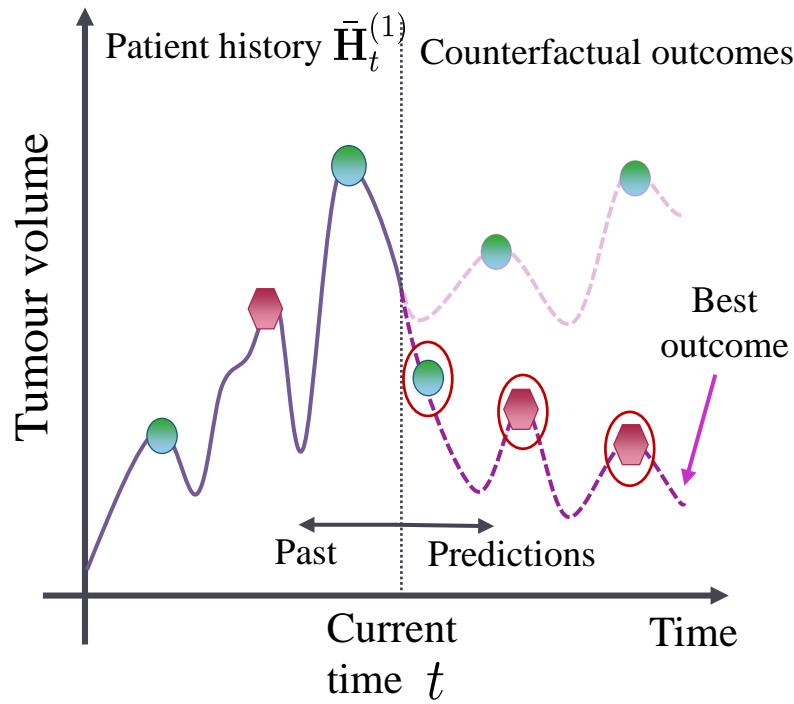
(b) Decide optimal time of treatment



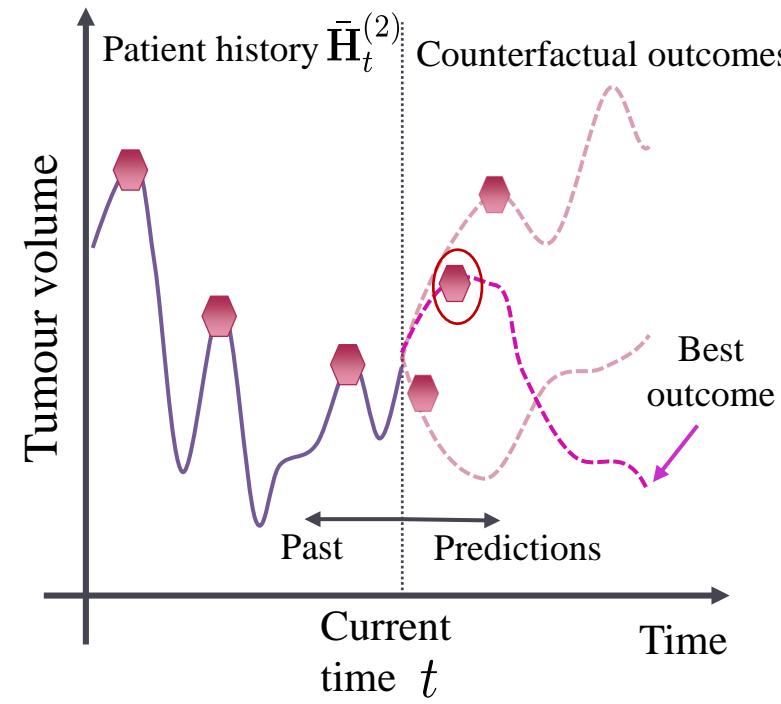
(c) Decide when to stop treatment

# When to treat? How to treat? When to stop?

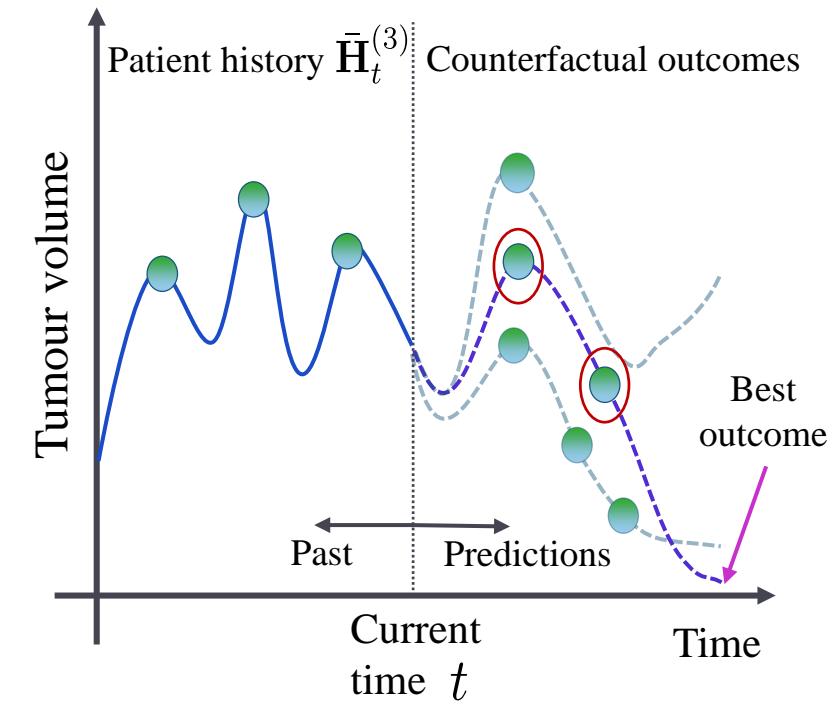
● Chemotherapy      ♦ Radiotherapy



(a) Decide treatment plan



(b) Decide optimal time of treatment

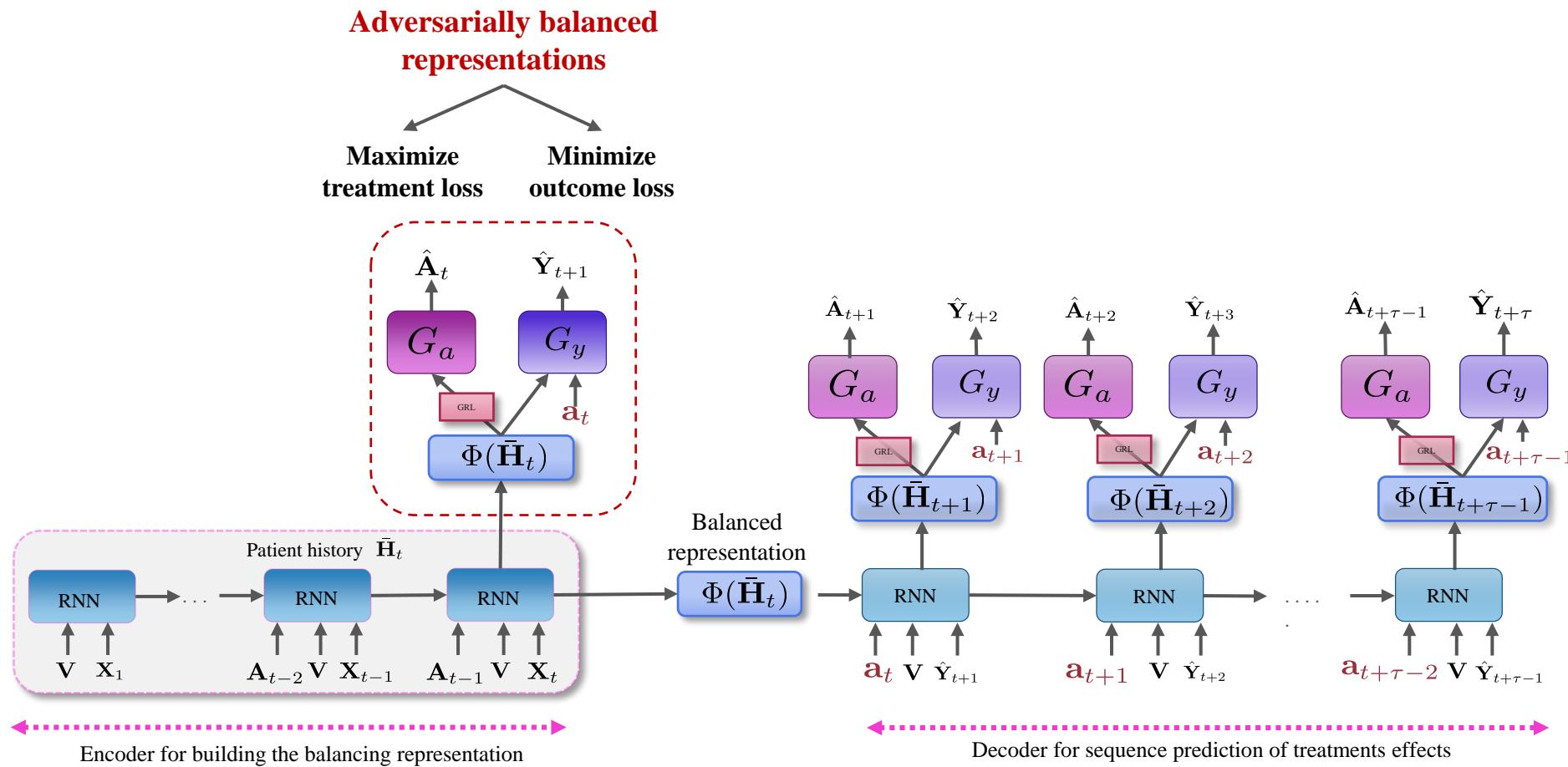


(c) Decide when to stop treatment

# Counterfactual Recurrent Network [Bica, Alaa, vdS, 2019]

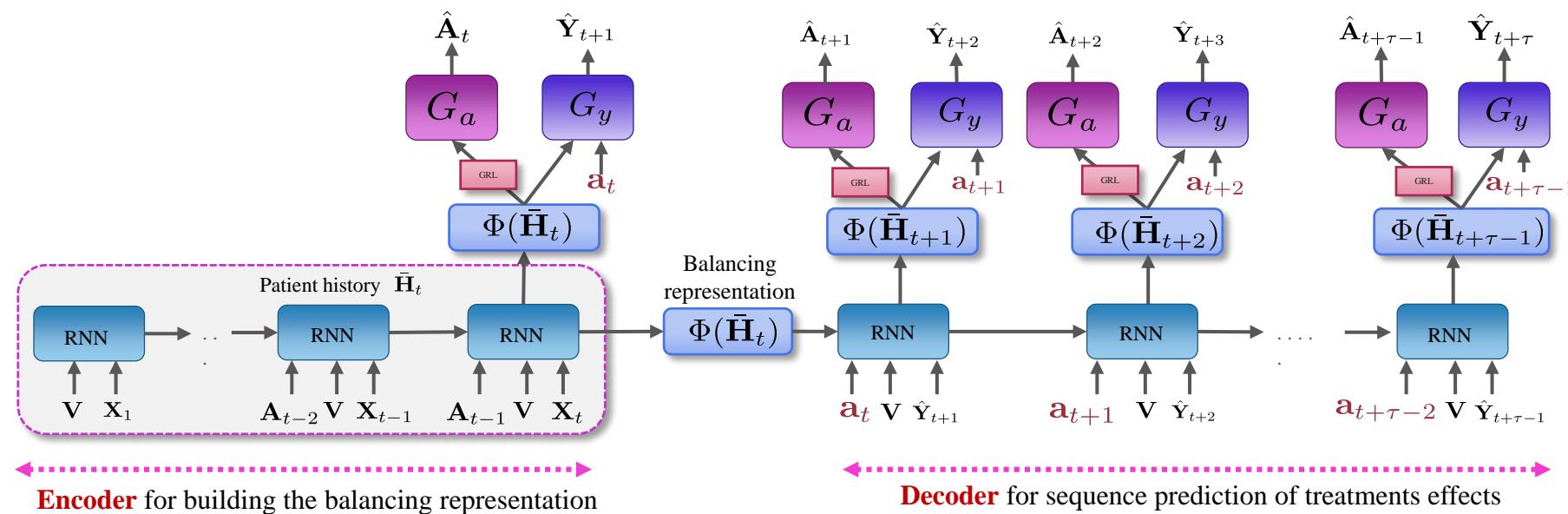
<https://arxiv.org/abs/1902.00450>

- Builds treatment invariant representations using domain adversarial training.

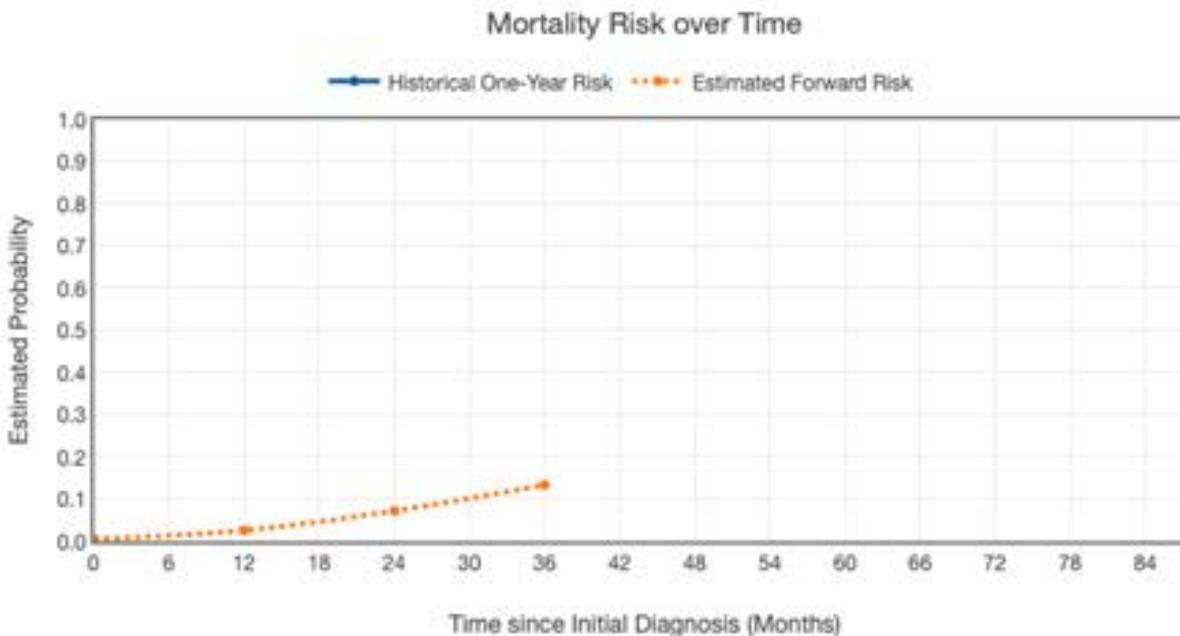


# Counterfactual Recurrent Network [Bica, Alaa, vdS, 2019]

- Builds treatment invariant representations using domain adversarial training.
- Predicts counterfactuals using a novel sequence-to-sequence architecture



# Part 3: Automate the process of designing Clinical Predictive Analytics at Scale



BREAST CANCER    COLON CANCER    LUNG CANCER    PROSTATE CANCER

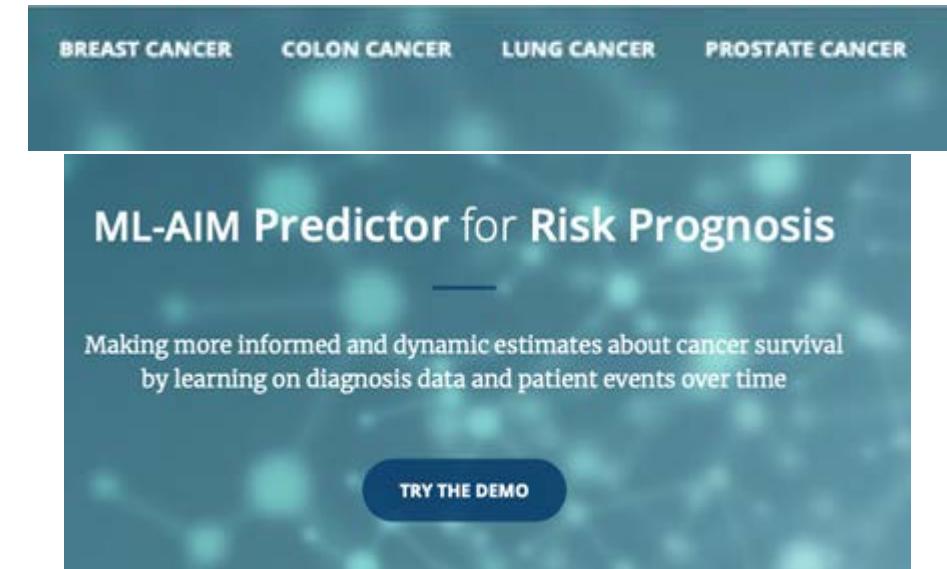
ML-AIM Predictor for Risk Prognosis

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

TRY THE DEMO

# Part 3: Automate the process of designing Clinical Predictive Analytics at Scale

Different types of cancer  
Cardiovascular disease  
Diabetes  
Cystic Fibrosis  
Alzheimer's  
etc.



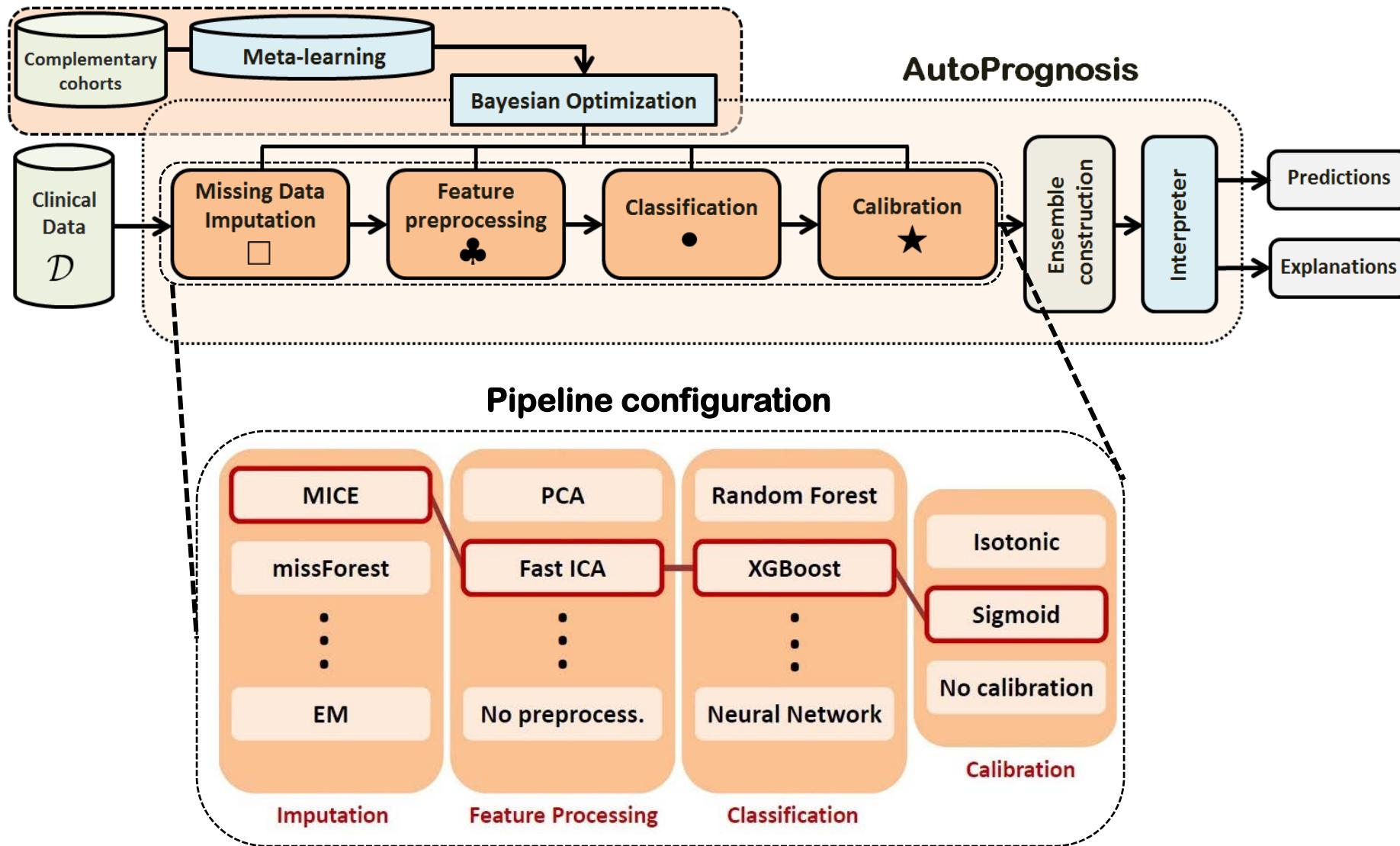
Many diseases, many variables, various needs!  
All is changing!

Can't craft a model for each disease!

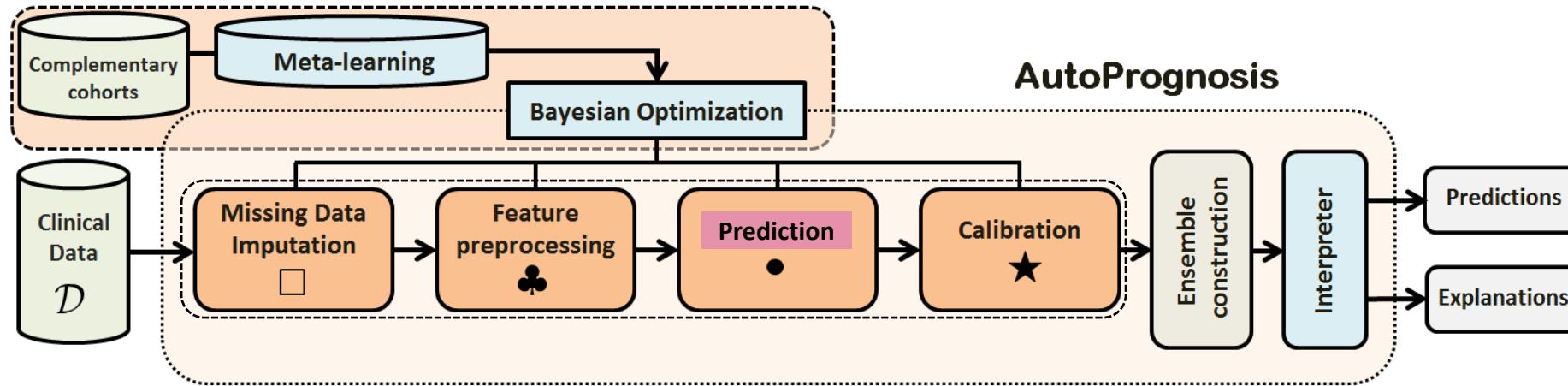
Make Machine Learning DO the Crafting

# AutoPrognosis: Automated ML for clinical analytics

[Alaa & vdS, ICML 2018, Scientific Reports 2018, Plos 2019]:  
A tool for crafting Clinical Scores



# AutoPrognosis pipelines include our algorithms!

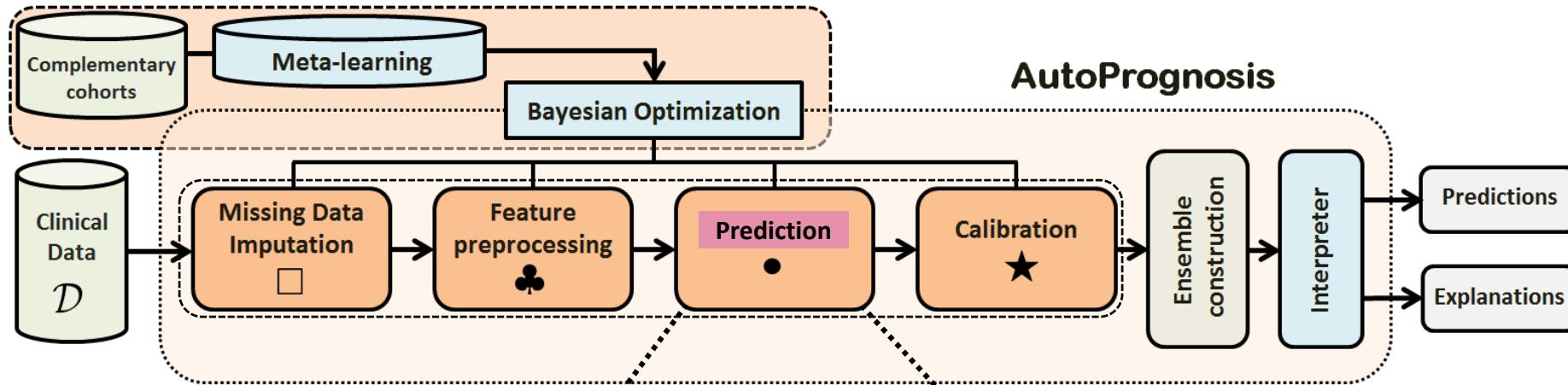


**Missing data:** GAIN [ICML 2018], Deep Sensing [ICLR 2018]

**Feature selection:** Knock-off GAN [ICLR 2019], INVASE [ICLR 2019]

**Prediction – time-to-event analysis:** DeepHit [AAAI 2018]

# Automated ML for clinical analytics (beyond predictions)



ICML 2018  
Scientific Reports  
Plos One

Survival Models

Competing Risks

Temporal Models

Causal Models

Lee, Alaa, Zame, vdS, AISTATS 2019

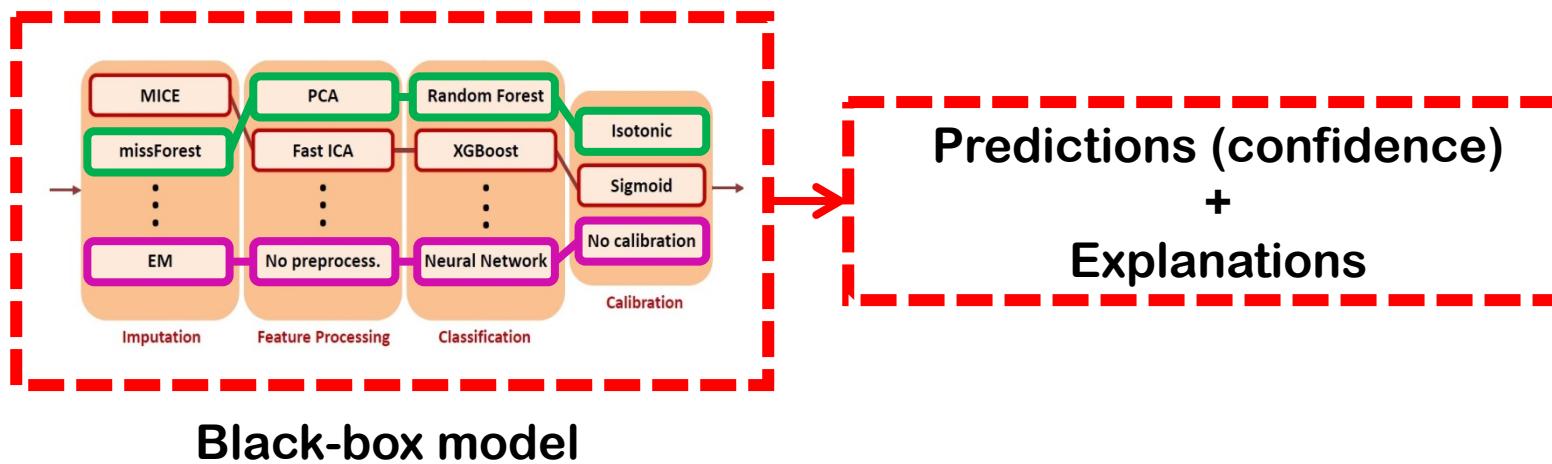
Alaa, vdS, NIPS 2017  
Bellot, vdS, AISTATS 2018

In submission

Alaa, vdS, ICML 2019

# Not only **black-box** predictions, also **interpretations**

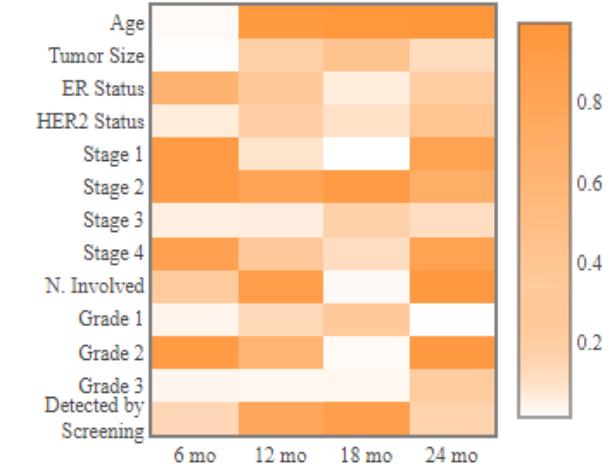
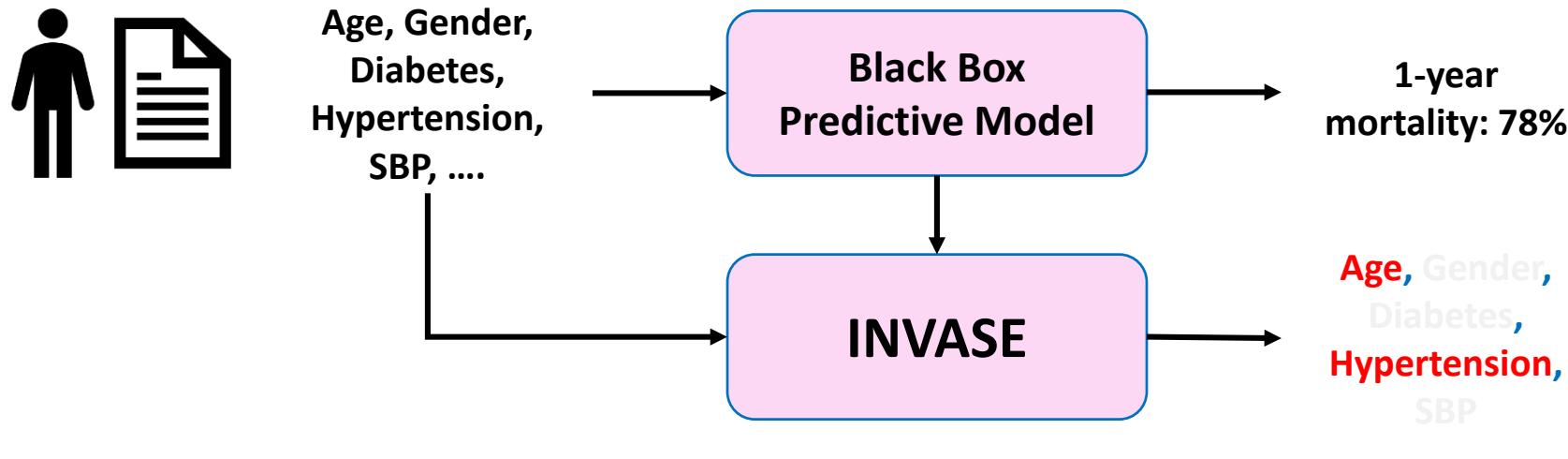
- Essential for trustworthiness, transparency etc.



- INVASE: Instance-wise Variable Selection using Deep Learning [Yoon, Jordon, vdS, ICLR 2019]
- Clinician-AI interaction using Reinforcement Learning [Lahav, vdS, NeurIPS workshop 2018]
- Metamodeling [Alaa, vdS, 2019]

# INVASE: Instance-wise Variable Selection using Deep Learning

[Yoon, Jordon, vdS, ICLR 2019]



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

Sir William Osler (1892)

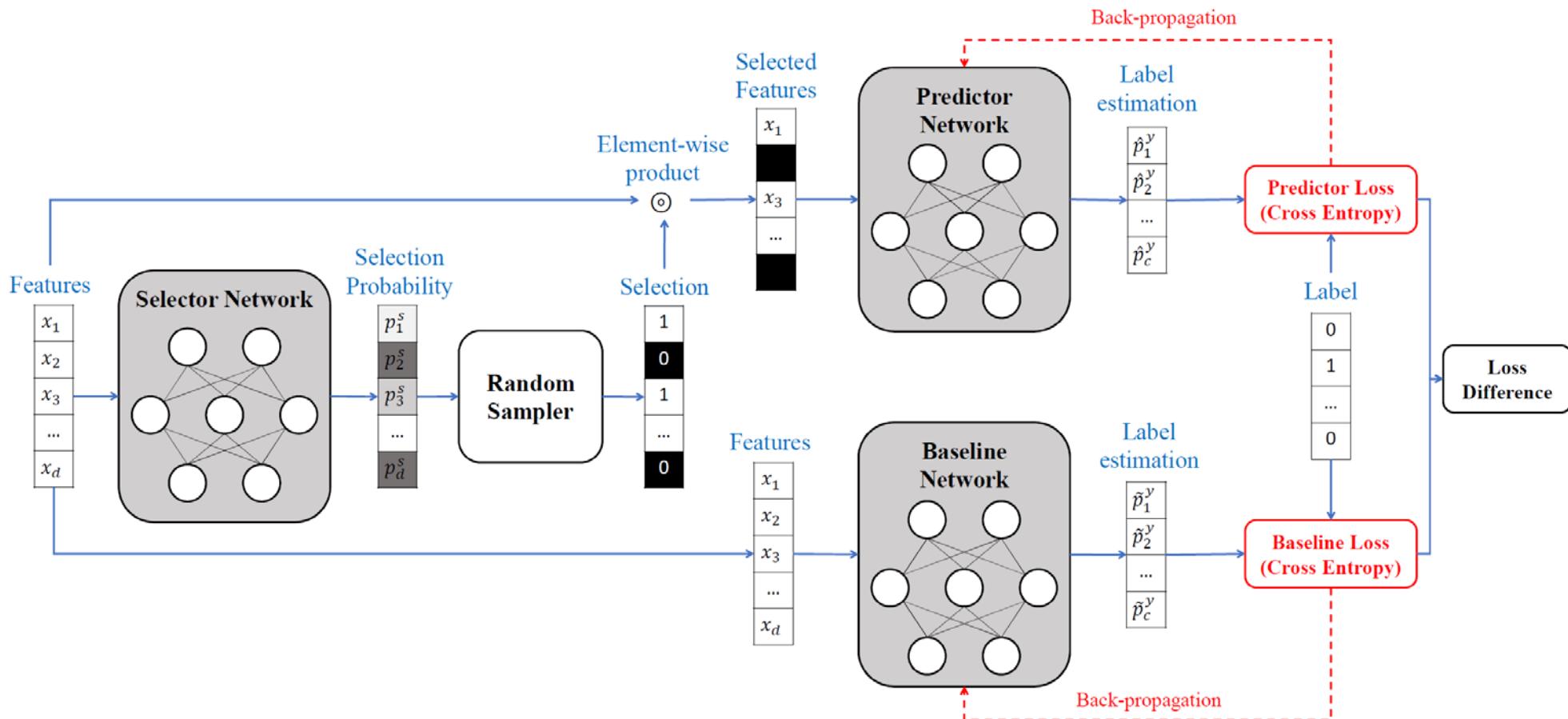
Understanding which features are relevant for a prediction outcome for each instance is critical for interpretability and also improves predictions.

# Relationship to existing works – Instance-wise feature selection

	Key ideas	Experiments shown	Global/ Instance-wise	Model agnostic	# of relevant features
L2X [4]	Mutual Information maximization with Gumbel-softmax	Interpretation	Instance-wise	Yes	Should be given
LIME [24]	Locally linear approximation	Interpretation	Instance-wise	Yes	Should be given
Shapley [18]	Shapley value estimation to quantify feature importance	Feature selection	Instance-wise	Yes	Should be given
DeepLIFT [27]	Decompose the output of NN on a reference input	Interpretation	Instance-wise	No	Should be given
Saliency [29]	Backpropagation from the output of the NN to the input	Interpretation	Instance-wise	No	Should be given
Tree SHAP [19]	Shapley value estimation only for tree-ensemble models	Interpretation	Instance-wise	No	Should be given
Pixel-wise [1]	Measuring the effects on the output using input perturbation	Interpretation	Instance-wise	No	Should be given
<b>INVASE (Ours)</b>	Minimize KL divergence using deep NN influenced by actor-critic models	Feature selection Interpretation Prediction	Instance-wise	Yes	Not needed

**INVASE discovers *different numbers* of relevant features for each instance**

# INVASE – Actor Critic framework



- Instances are fed into the **selector network (actor)** which outputs a vector of selection probabilities.
- The **predictor network (critic)** receives the selected features and makes prediction and provides **feedback to the actor**.

## INVASE – Problem formulation

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

- **Objective: minimize  $S(x)$**
- **Constraints:**

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

- $x$ : Features for *a given realization*
- $S: \mathcal{X} \rightarrow \{0,1\}^d$ : Selector function,  $S(x)$ : Selected features
- $Y$ : Labels (Outcomes)
  - **come from a dataset: the problem is of selecting predictive features**
  - **come from a predictive model: the problem is of explaining the model's predictions**

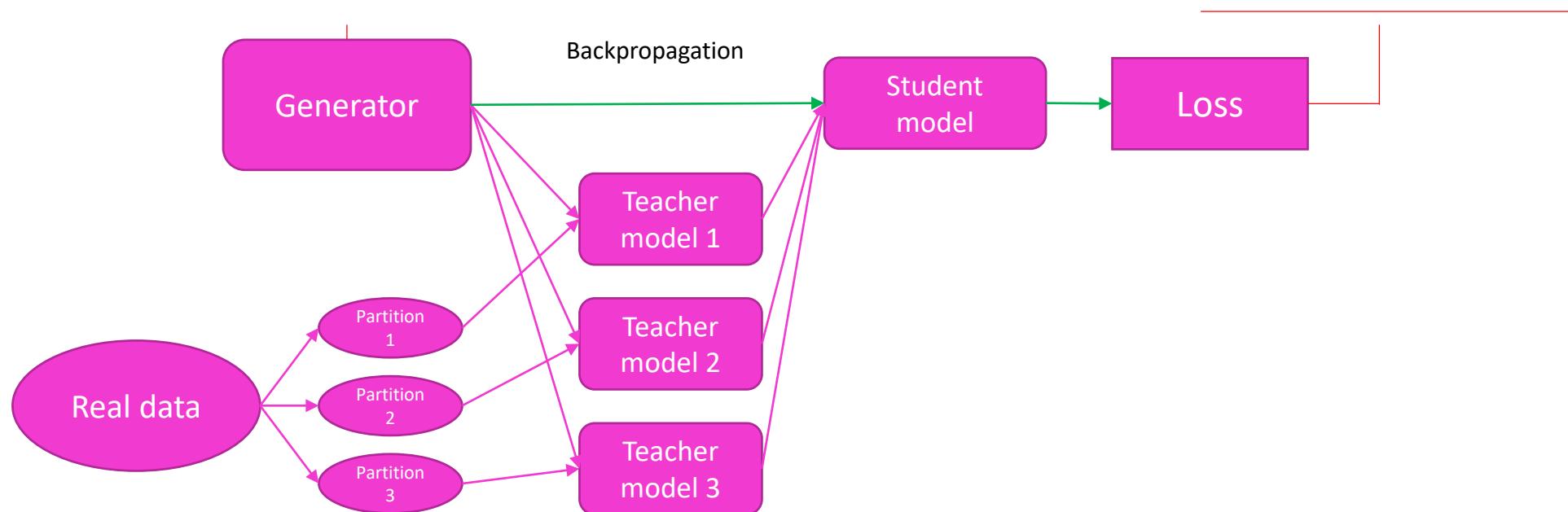
# How to access medical data?

## PATE-GAN: Generating Synthetic Data with Differential Privacy Guarantees [ICLR 2019]

**Idea:** Use a student PATE model as discriminator for training the generator

We extend the iterative process of the GAN framework to include a step in each iteration in which (several) teacher models are used to update a student model.

**PATE framework** guarantees that the discriminator is differentially private, while the **post-processing theorem** guarantees the generator is differentially private





**ML-AIM**

Research Laboratory led by Prof. Mihaela van der Schaar

Machine Learning and Artificial Intelligence for Medicine

Details about our algorithms:

<http://www.vanderschaar-lab.com>

Details about our software:

<http://www.vanderschaar-lab.com>



# ML-AIM

Research Laboratory led by Prof. Mihaela van der Schaar

Machine Learning and Artificial Intelligence for Medicine

## PhD students:

Ahmed Alaa

Changhee Lee

Alexis Bellot

Trent Kyono

Ioana Bica

Jinsung Yoon

Dan Jarrett

Yao Zhang

James Jordon

We are recruiting!!!