MATHEMATICAL FRONTIERS

The National Academies of SCIENCES ENGINEERING MEDICINE

nas.edu/MathFrontiers

Board on Mathematical Sciences & Analytics

MATHEMATICAL FRONTIERS

2019 Monthly Webinar Series, 2-3pm ET

February 12: *Machine Learning for Materials Science**

March 12: *Mathematics of Privacy**

April 9: *Mathematics of Gravitational Waves**

May 14: *Algebraic Geometry**

June 11: *Mathematics of Transportation**

July 9: *Cryptography & Cybersecurity**

August 13: *Machine Learning in Medicine*

September 10: *Logic and Foundations*

October 8: *Mathematics of Quantum Physics*

November 12: *Quantum Encryption*

December 10: *Machine Learning for Text*

Made possible by support for BMSA from the
National Science Foundation
Division of Mathematical Sciences
and the
Department of Energy
Advanced Scientific Computing Research

** Webinar posted*

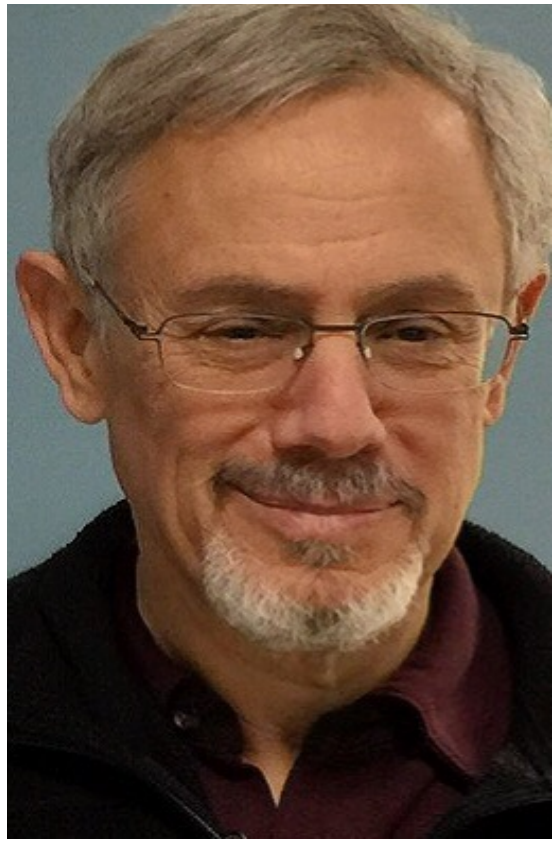
View webinar videos and learn more about BMSA at www.nas.edu/MathFrontiers

MATHEMATICAL FRONTIERS

Machine Learning in Medicine

**Mihaela van der Schaar,
University of Cambridge,
Alan Turing Institute, and UCLA**

**Juan Gutierrez,
University of Georgia**



**Mark Green,
UCLA (moderator)**

**Mihaela van der Schaar,
University of Cambridge,
Alan Turing Institute, and UCLA**

John Humphrey Plummer Professor of Machine Learning, Artificial Intelligence and Medicine at University of Cambridge
Turing Faculty Fellow at the Alan Turing Institute in London
Chancellor's Professor at UCLA

Transforming medicine through AI-enabled healthcare pathways

My research

Develop **cutting-edge machine learning and AI** theory, methods, algorithms and systems to **deliver effective personalized healthcare**

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

Genomic Data is Big Data

Genome, Transcriptome and Proteome

&

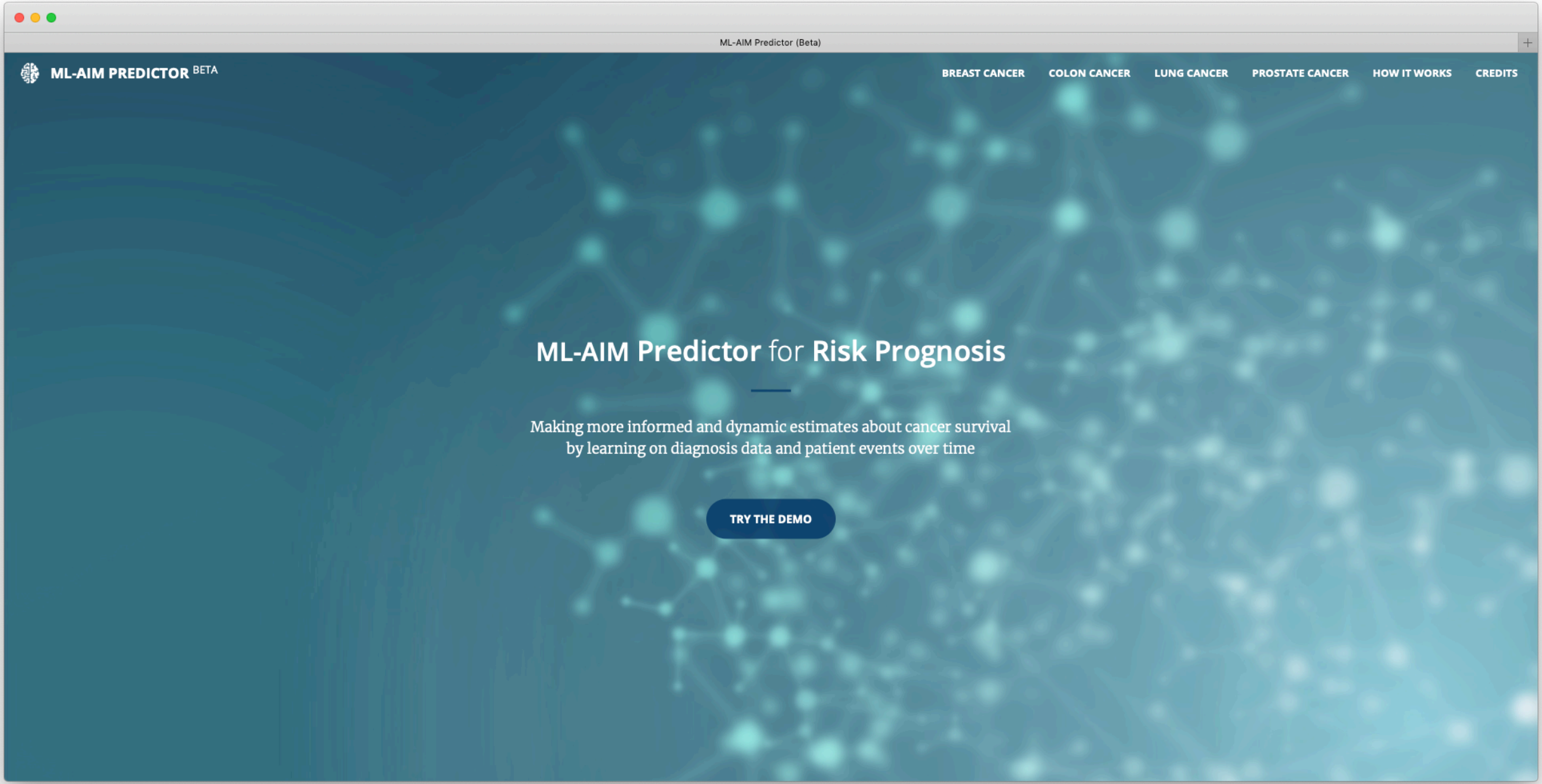
Clinical Data is Complex Data

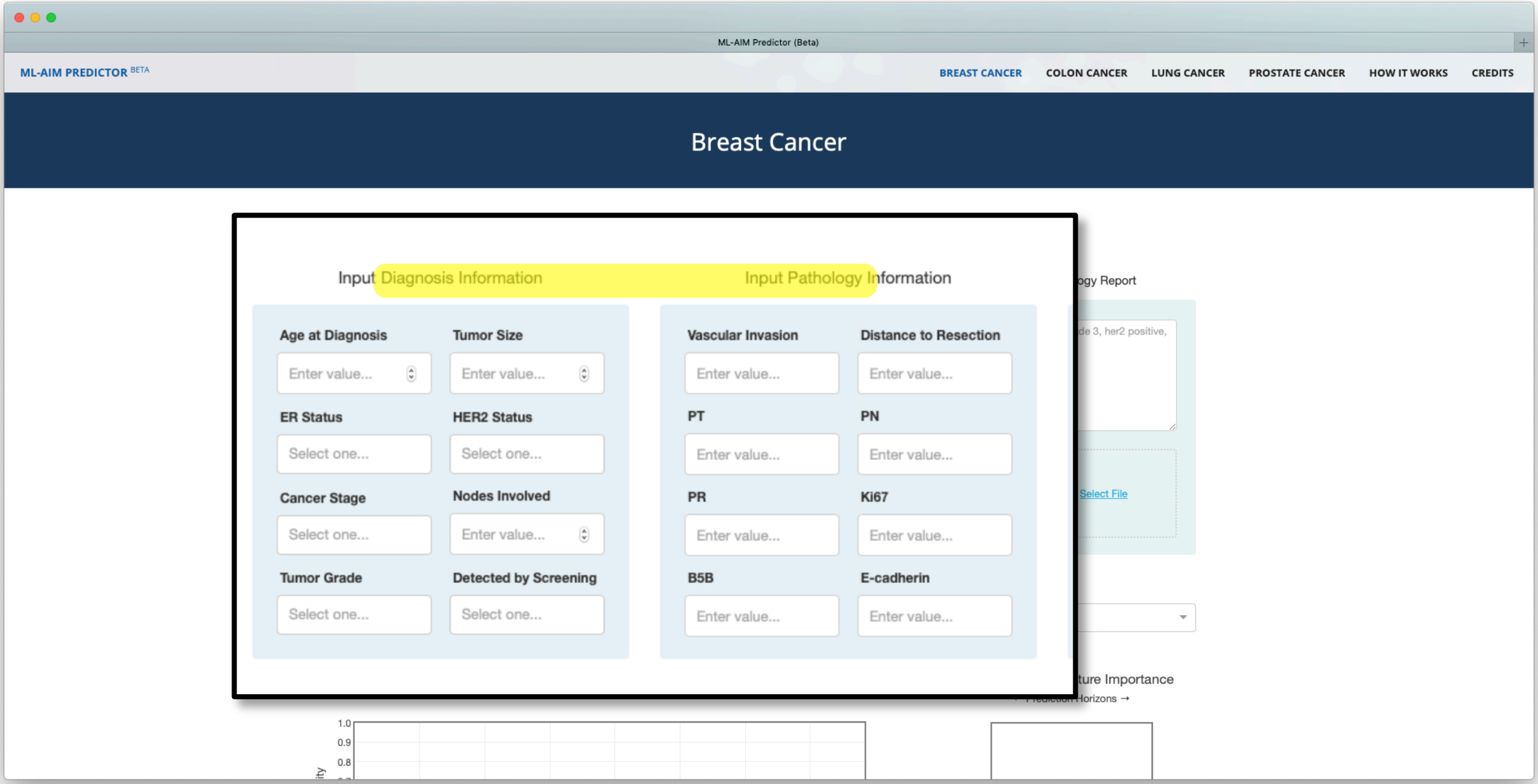
"Clinome", Patient Experience, Risk Factors etc

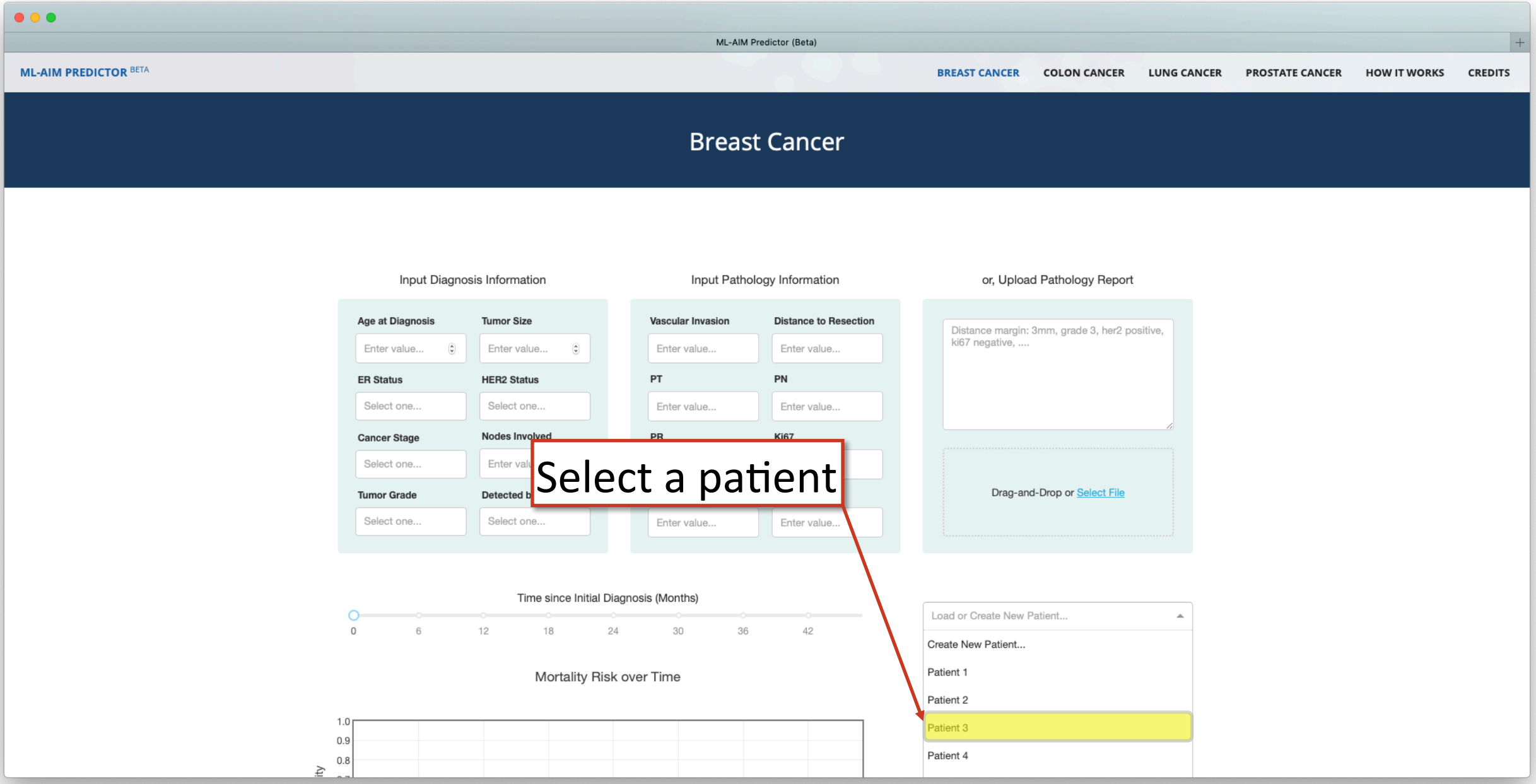
Goal: deliver decision support for the patient at hand

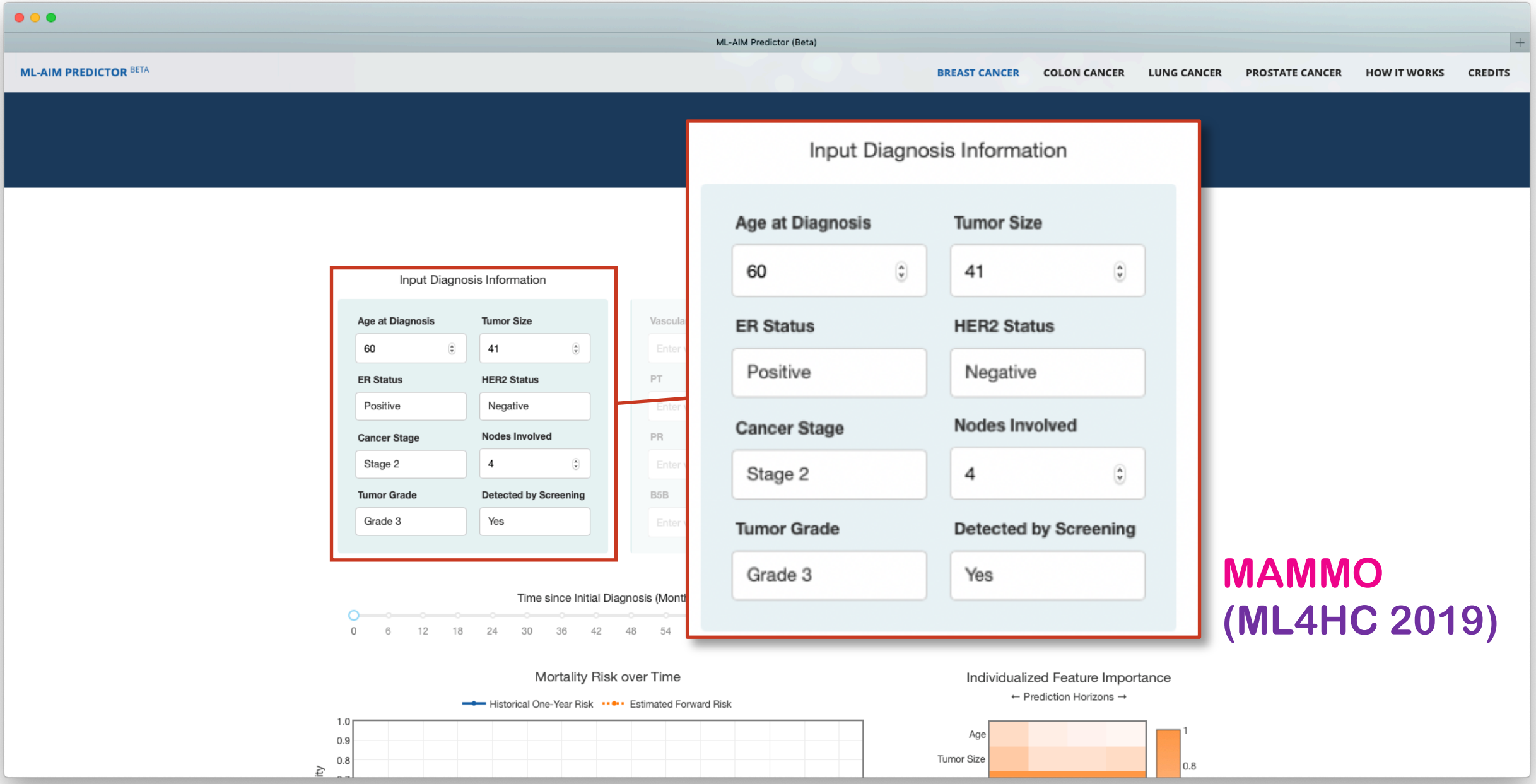
Cancer - a useful exemplar

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

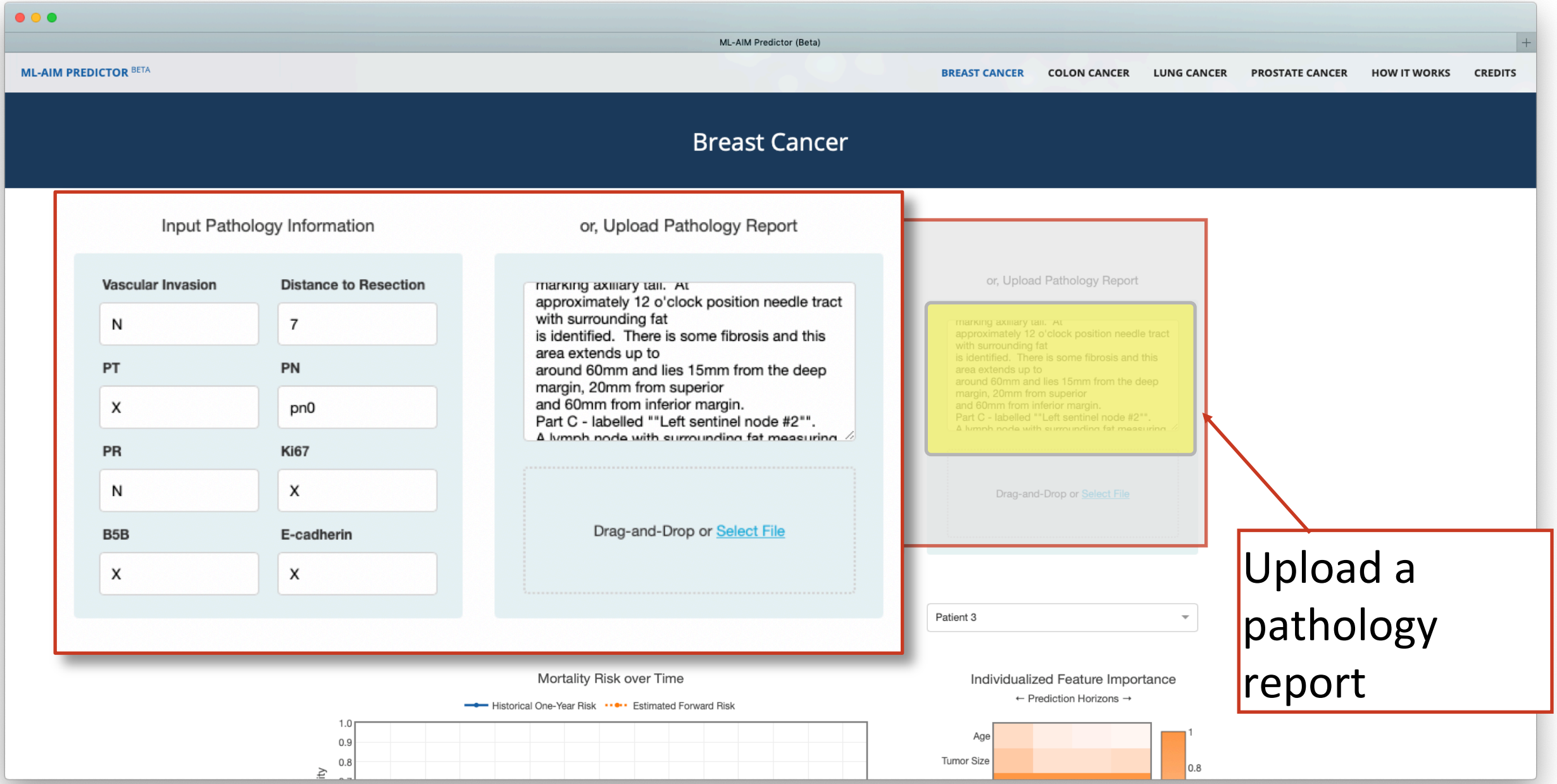

ML-AIM Predictor (Beta)
ML-AIM PREDICTOR BETA
BREAST CANCER
COLON CANCER
LUNG CANCER
PROSTATE CANCER
HOW IT WORKS
CREDITS
ML-AIM Predictor for Risk Prognosis
Making more informed and dynamic estimates about cancer survival
by learning on diagnosis data and patient events over time
TRY THE DEMO


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
Input Pathology Information
Age at Diagnosis
Enter value...
Tumor Size
Enter value...
ER Status
Select one...
HER2 Status
Select one...
Cancer Stage
Select one...
Nodes Involved
Enter value...
Tumor Grade
Select one...
Detected by Screening
Select one...
Vascular Invasion
Enter value...
Distance to Resection
Enter value...
PT
Enter value...
PN
Enter value...
PR
Enter value...
Ki67
Enter value...
B5B
Enter value...
E-cadherin
Enter value...
ogy Report
de 3, her2 positive,
Select File
ture Importance
Prediction Horizons →
1.0
0.9
0.8

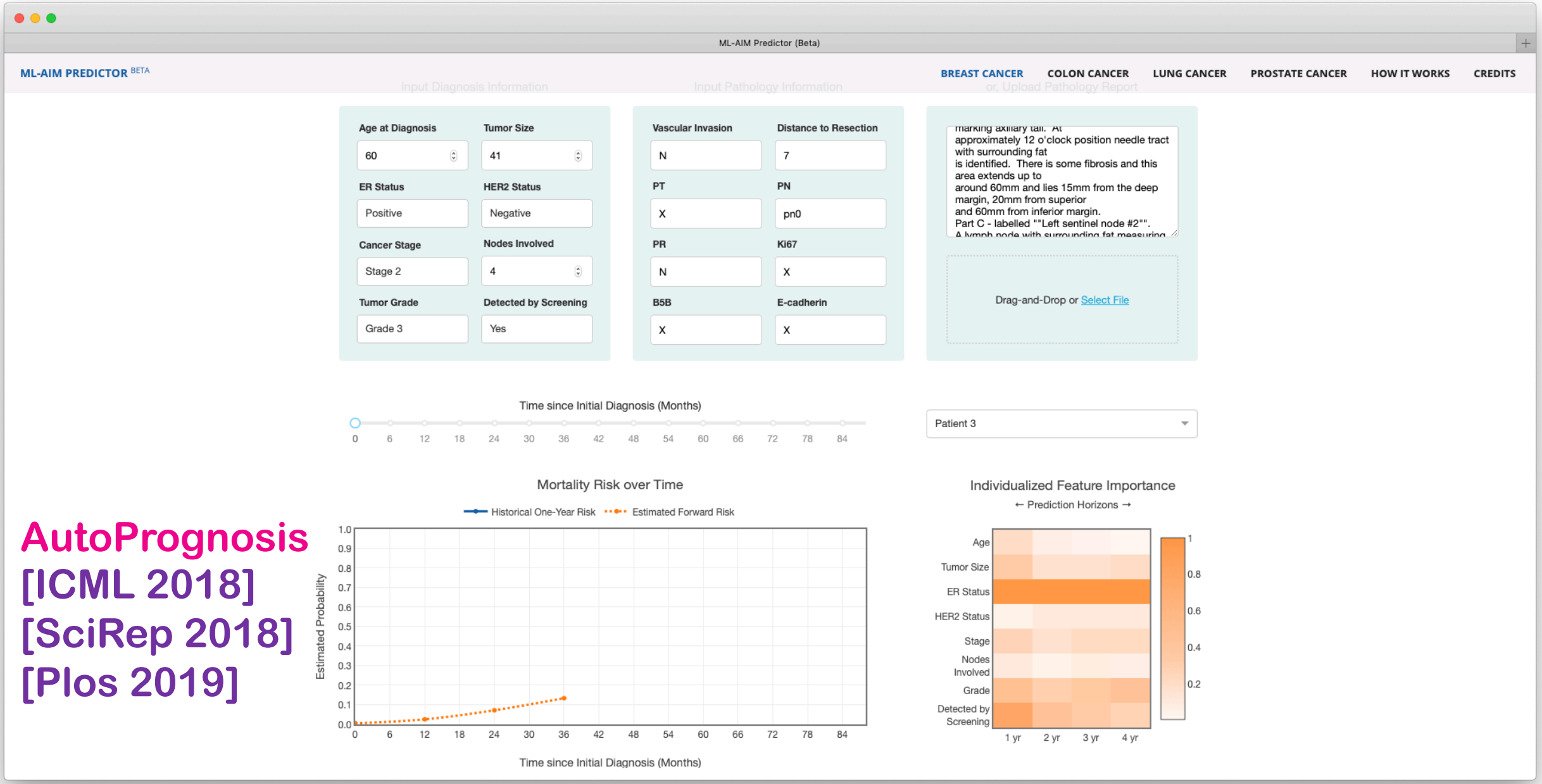

ML-AIM Predictor (Beta)
ML-AIM PREDICTOR BETA
BREAST CANCER
COLON CANCER
LUNG CANCER
PROSTATE CANCER
HOW IT WORKS
CREDITS
Breast Cancer
Input Diagnosis Information
Age at Diagnosis
Tumor Size
Enter value...
ER Status
HER2 Status
Select one...
Cancer Stage
Nodes Involved
Tumor Grade
Detected b
Input Pathology Information
Vascular Invasion
Distance to Resection
PT
PN
PR
Ki67
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)
0
6
12
18
24
30
36
42
Mortality Risk over Time
1.0
0.9
0.8
Load or Create New Patient...
Create New Patient...
Patient 1
Patient 2
Patient 3
Patient 4



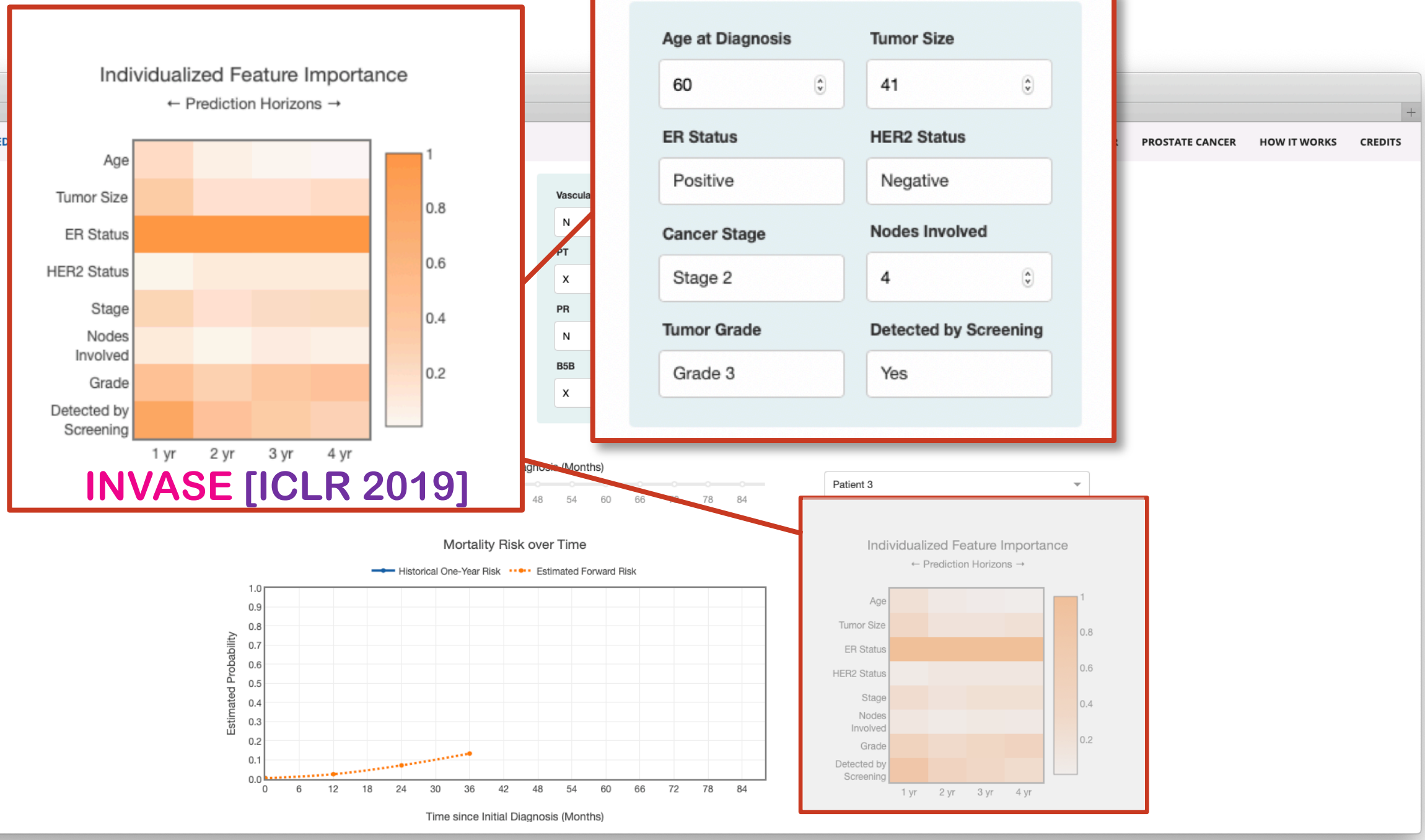
MAMMO (ML4HC 2019)

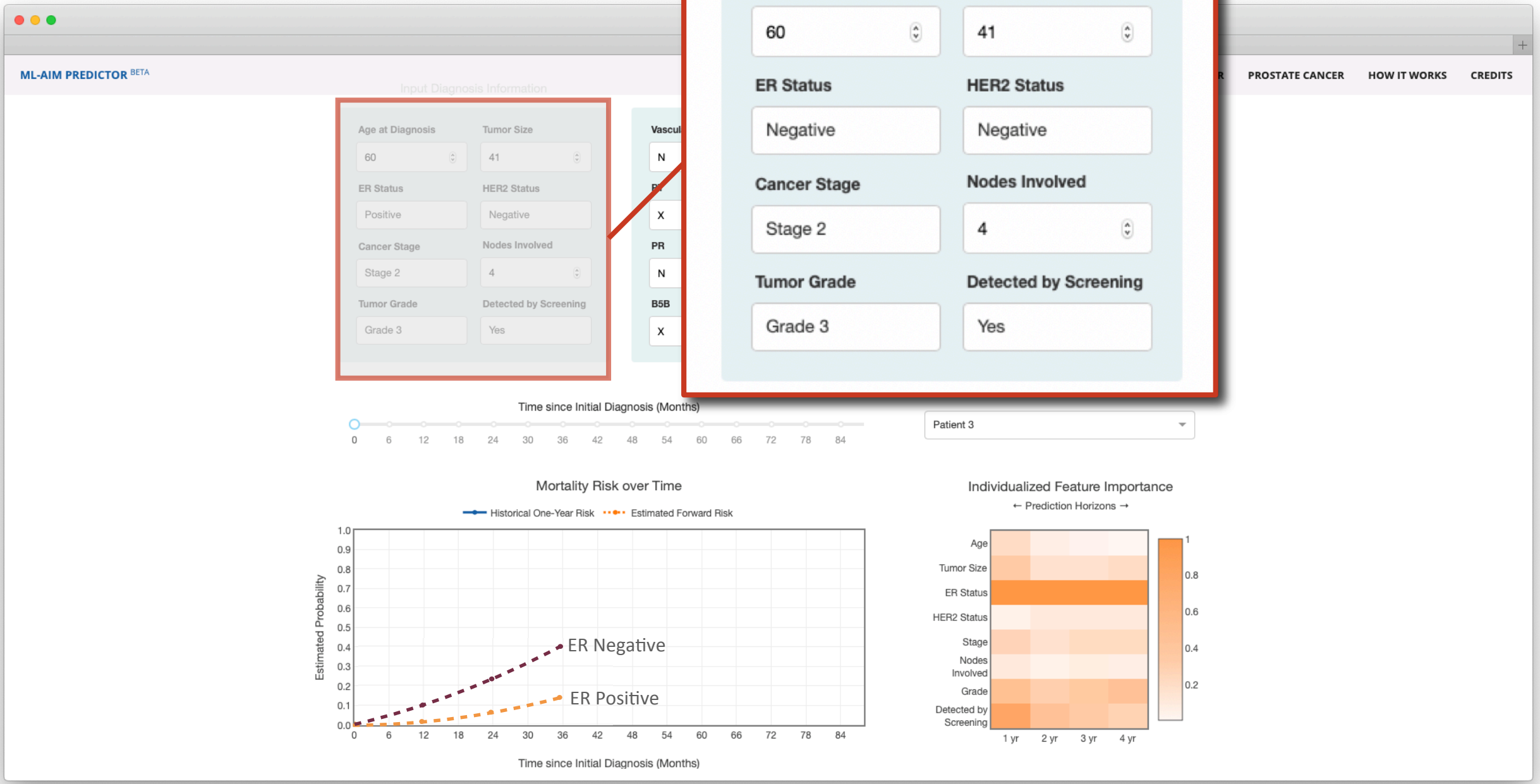


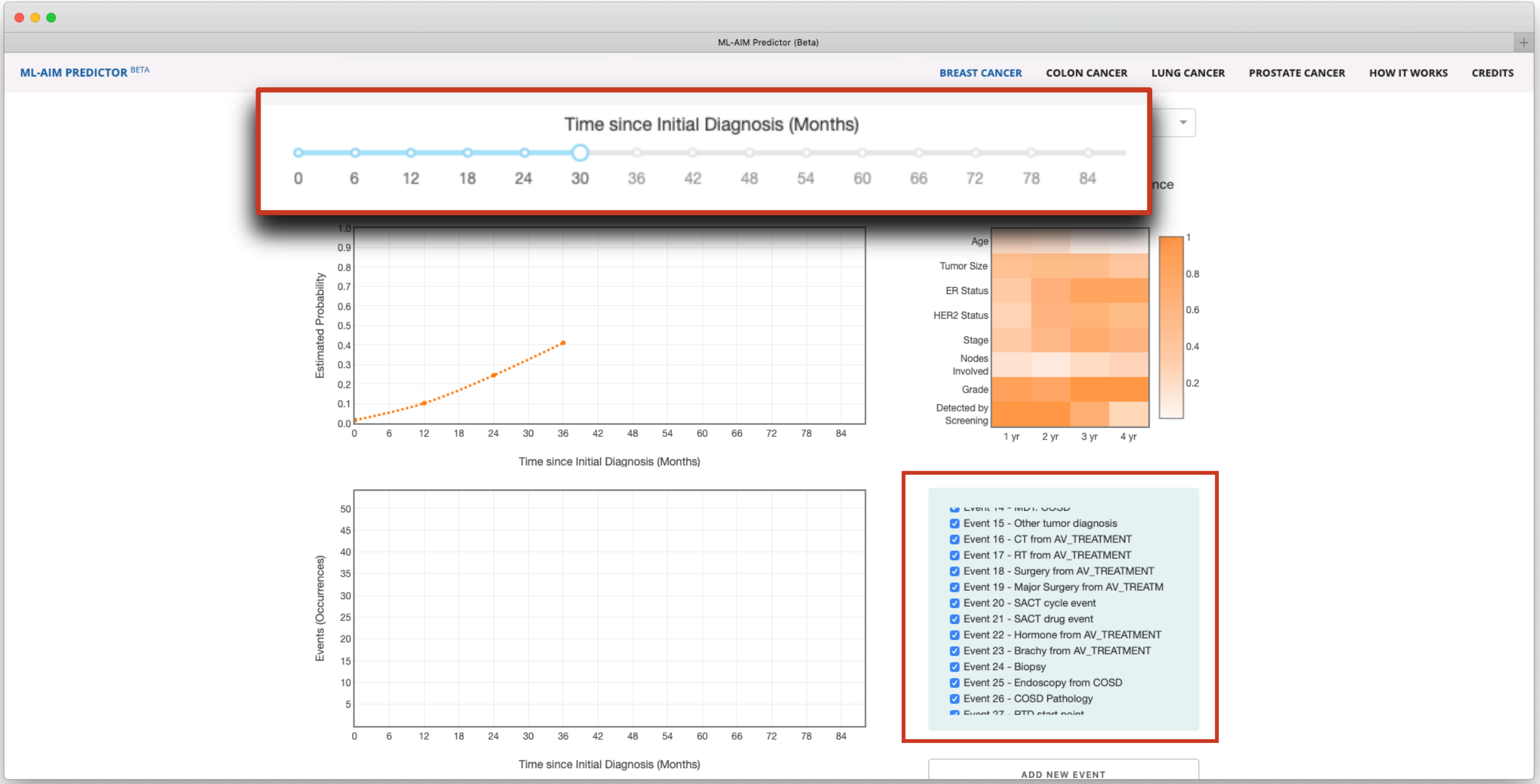
Upload a pathology report



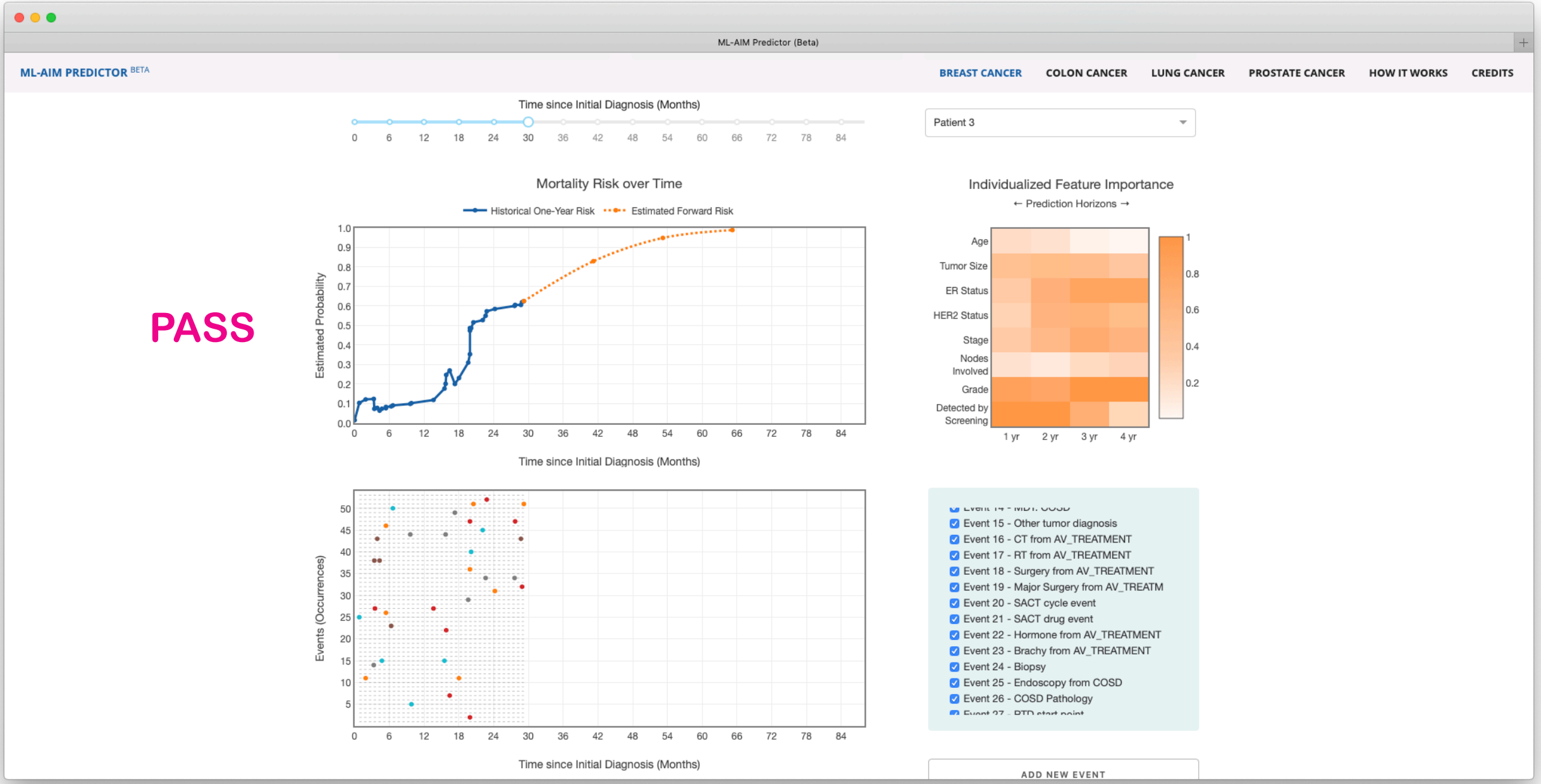
AutoPrognosis
[ICML 2018]
[SciRep 2018]
[Plos 2019]


Individualized Feature Importance
← Prediction Horizons →
Age
Tumor Size
ER Status
HER2 Status
Stage
Nodes Involved
Grade
Detected by Screening
1 yr
2 yr
3 yr
4 yr
1
0.8
0.6
0.4
0.2
INVASE [ICLR 2019]
Age at Diagnosis
60
Tumor Size
41
ER Status
Positive
HER2 Status
Negative
Cancer Stage
Stage 2
Nodes Involved
4
Tumor Grade
Grade 3
Detected by Screening
Yes
ML-AIM PRED
PROSTATE CANCER
HOW IT WORKS
CREDITS
Patient 3
Mortality Risk over Time
Historical One-Year Risk
Estimated Forward Risk
Estimated Probability
Time since Initial Diagnosis (Months)
Individualized Feature Importance
← Prediction Horizons →


ML-AIM PREDICTOR BETA
PROSTATE CANCER
HOW IT WORKS
CREDITS
Input Diagnosis Information
Age at Diagnosis
60
Tumor Size
41
ER Status
Positive
HER2 Status
Negative
Cancer Stage
Stage 2
Nodes Involved
4
Tumor Grade
Grade 3
Detected by Screening
Yes
Time since Initial Diagnosis (Months)
0 6 12 18 24 30 36 42 48 54 60 66 72 78 84
Patient 3
Mortality Risk over Time
Historical One-Year Risk
Estimated Forward Risk
Estimated Probability
1.0 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0
ER Negative
ER Positive
0 6 12 18 24 30 36 42 48 54 60 66 72 78 84
Time since Initial Diagnosis (Months)
Individualized Feature Importance
← Prediction Horizons →
Age
Tumor Size
ER Status
HER2 Status
Stage
Nodes Involved
Grade
Detected by Screening
1 yr 2 yr 3 yr 4 yr
1 0.8 0.6 0.4 0.2


ML-AIM Predictor (Beta)
ML-AIM PREDICTOR BETA
BREAST CANCER
COLON CANCER
LUNG CANCER
PROSTATE CANCER
HOW IT WORKS
CREDITS
Time since Initial Diagnosis (Months)
0 6 12 18 24 30 36 42 48 54 60 66 72 78 84
Estimated Probability
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 yr
Events (Occurrences)
Time since Initial Diagnosis (Months)
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
ADD NEW EVENT

PASS


ML-AIM Predictor (Beta)
ML-AIM PREDICTOR BETA
BREAST CANCER
COLON CANCER
LUNG CANCER
PROSTATE CANCER
HOW IT WORKS
CREDITS
Time since Initial Diagnosis (Months)
0 6 12 18 24 30 36 42 48 54 60 66 72 78 84
Patient 3
Mortality Risk over Time
Historical One-Year Risk
Estimated Forward Risk
Estimated Probability
1.0 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0
0 6 12 18 24 30 36 42 48 54 60 66 72 78 84
Time since Initial Diagnosis (Months)
Individualized Feature Importance
← Prediction Horizons →
Age
Tumor Size
ER Status
HER2 Status
Stage
Nodes Involved
Grade
Detected by Screening
1 yr 2 yr 3 yr 4 yr
1 0.8 0.6 0.4 0.2
Events (Occurrences)
50 45 40 35 30 25 20 15 10 5
0 6 12 18 24 30 36 42 48 54 60 66 72 78 84
Time since Initial Diagnosis (Months)
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
ADD NEW EVENT

ML-AIM Predictor (Beta)

ML-AIM PREDICTOR BETA

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

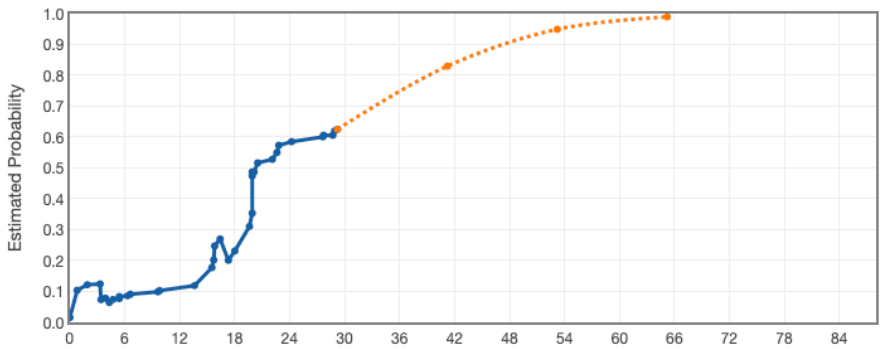
Time since Initial Diagnosis (Months)

0 6 12 18 24 30 36 42 48 54 60 66 72 78 84

Patient 3

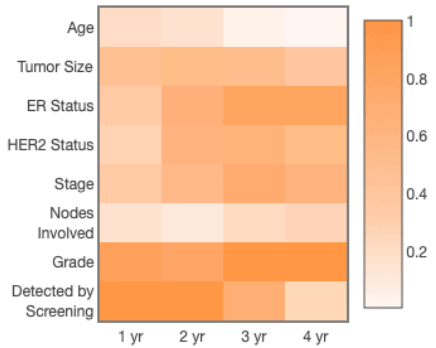
Mortality Risk over Time

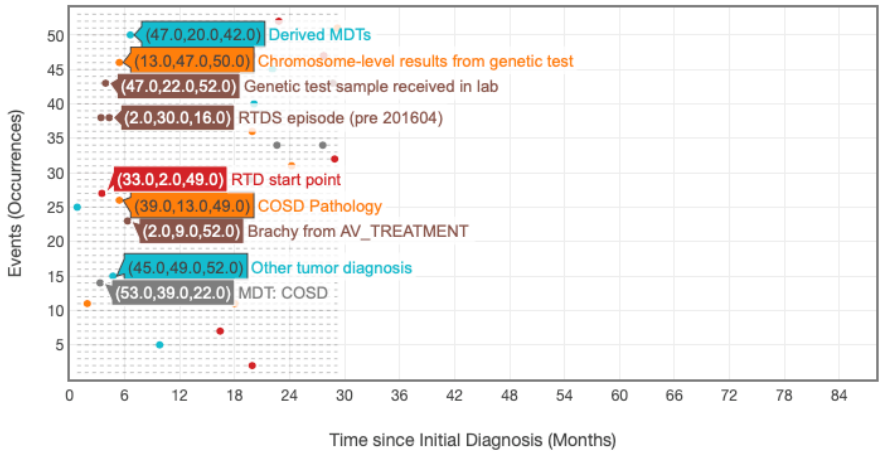
Historical One-Year Risk — Estimated Forward Risk



Individualized Feature Importance

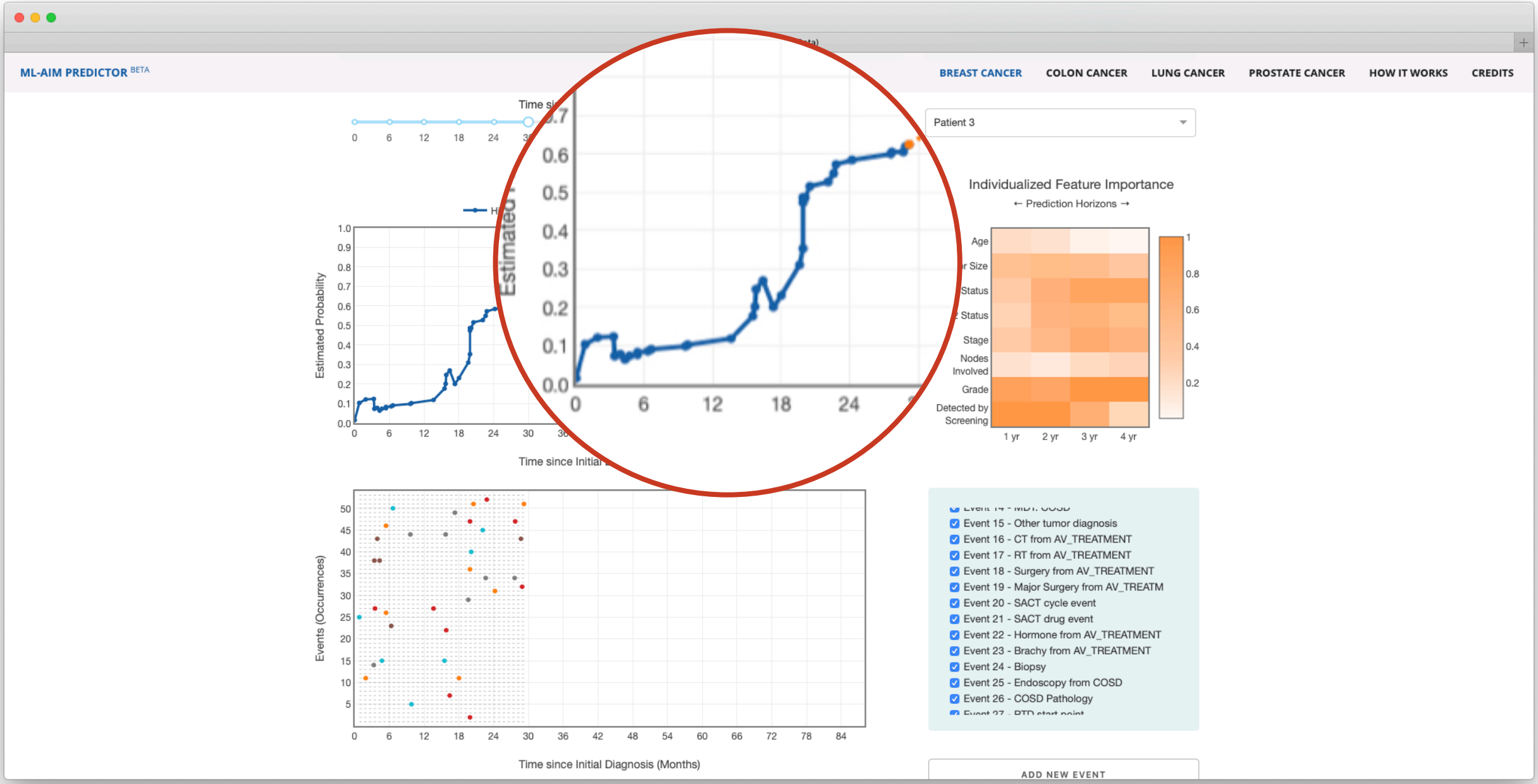
← Prediction Horizons →





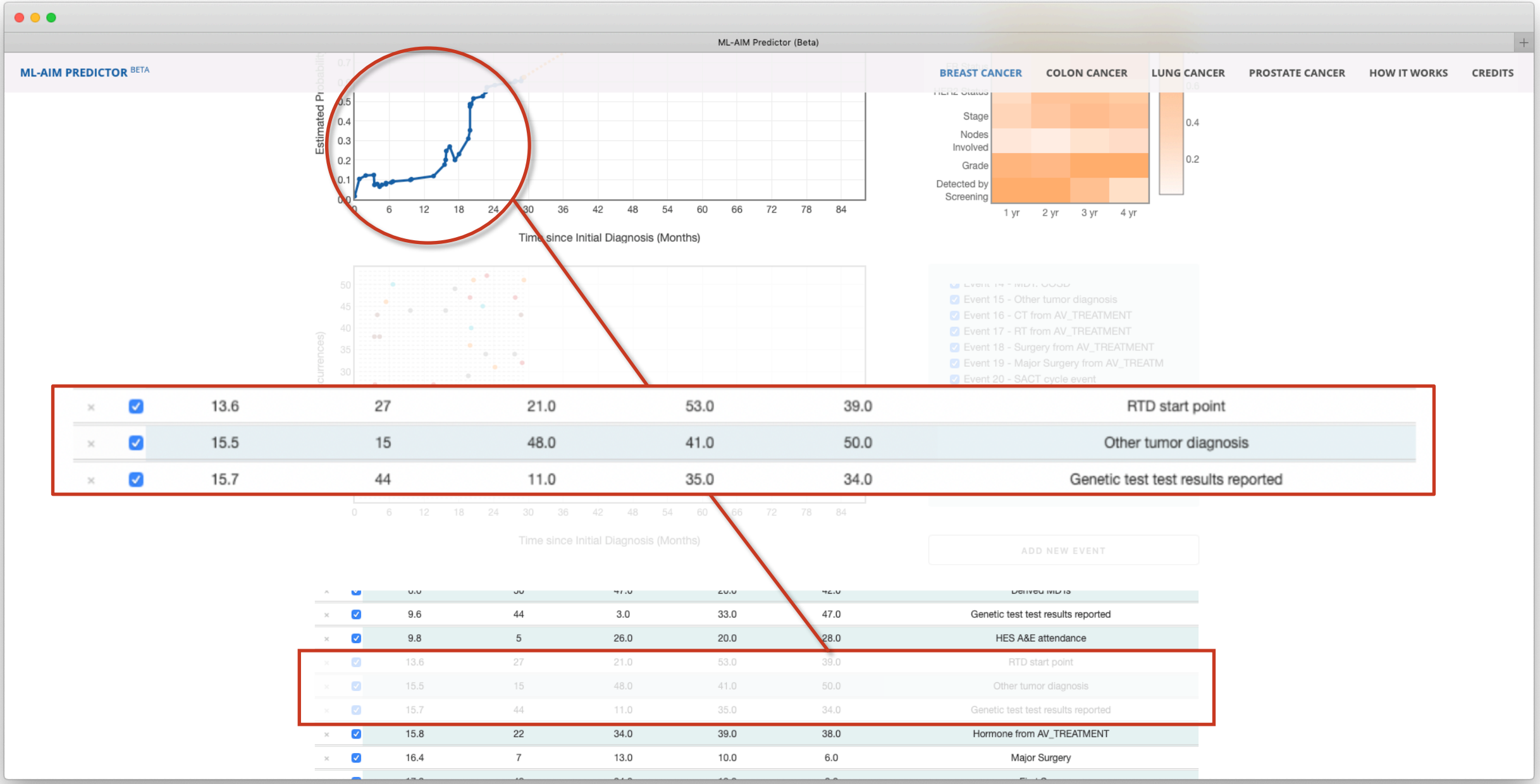
- Event 15 - Other tumor diagnosis
- Event 16 - CT from AV_TREATMENT
- Event 17 - RT from AV_TREATMENT
- Event 18 - Surgery from AV_TREATMENT
- Event 19 - Major Surgery from AV_TREATM
- Event 20 - SACT cycle event
- Event 21 - SACT drug event
- Event 22 - Hormone from AV_TREATMENT
- Event 23 - Brachy from AV_TREATMENT
- Event 24 - Biopsy
- Event 25 - Endoscopy from COSD
- Event 26 - COSD Pathology

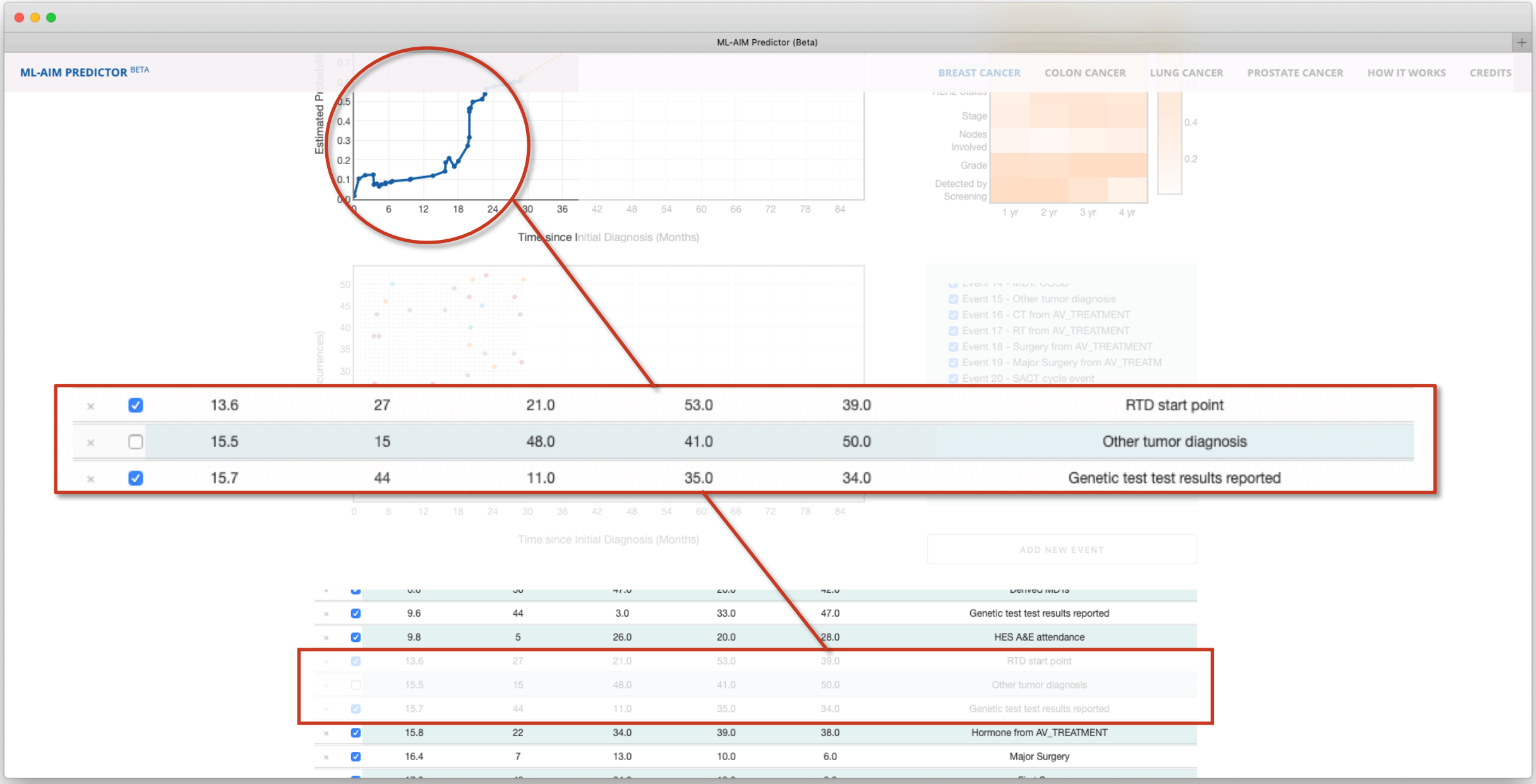
ADD NEW EVENT

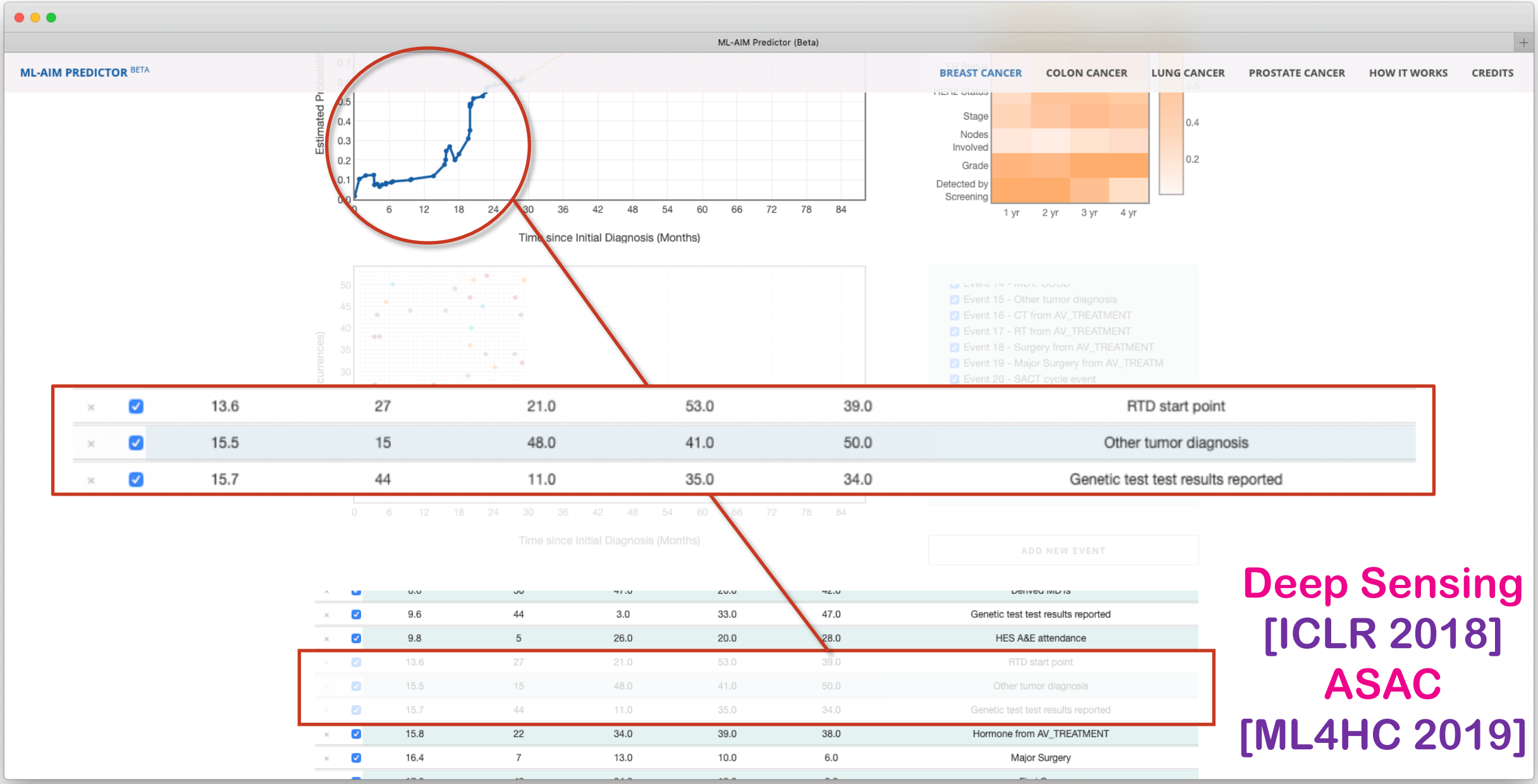

ML-AIM PREDICTOR BETA
BREAST CANCER
COLON CANCER
LUNG CANCER
PROSTATE CANCER
HOW IT WORKS
CREDITS
Patient 3
Individualized Feature Importance
← Prediction Horizons →
Age
Size
Status
Status
Stage
Nodes Involved
Grade
Detected by Screening
1 yr
2 yr
3 yr
4 yr
Estimated Probability
Time since Initial
Events (Occurrences)
Time since Initial Diagnosis (Months)
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
ADD NEW EVENT


ML-AIM PREDICTOR BETA
BREAST CANCER
COLON CANCER
LUNG CANCER
PROSTATE CANCER
HOW IT WORKS
CREDITS
Patient 3
Individualized Feature Importance
← Prediction Horizons →
Age
Status
Stage
Nodes Involved
Grade
Detected by Screening
1 yr
2 yr
3 yr
4 yr
Estimated Probability
0.12 Historical One-Year Risk
Time since Initial Diagnosis (Months)
Events (Occurrences)
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
ADD NEW EVENT

Month	Event Type	Property 1	Property 2	Property 3	Event Description
0.8	25	45.0	27.0	20.0	Endoscopy from COSD
1.9	11	44.0	47.0	28.0	Cancer Treatment Period Start
3.3	14	53.0	39.0	22.0	MDT: COSD
3.4	38	36.0	13.0	42.0	RTDS episode (pre 201604)
3.5	27	33.0	2.0	49.0	RTD start point
3.9	43	47.0	22.0	52.0	Genetic test sample received in lab


ML-AIM Predictor (Beta)
ML-AIM PREDICTOR BETA
BREAST CANCER
COLON CANCER
LUNG CANCER
PROSTATE CANCER
HOW IT WORKS
CREDITS
Estimated Probability
Time since Initial Diagnosis (Months)
Stage
Nodes Involved
Grade
Detected by Screening
1 yr
2 yr
3 yr
4 yr
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
ADD NEW EVENT
13.6 27 21.0 53.0 39.0 RTD start point
15.5 15 48.0 41.0 50.0 Other tumor diagnosis
15.7 44 11.0 35.0 34.0 Genetic test test results reported
9.6 44 3.0 33.0 47.0 Genetic test test results reported
9.8 5 26.0 20.0 28.0 HES A&E attendance
15.8 22 34.0 39.0 38.0 Hormone from AV_TREATMENT
16.4 7 13.0 10.0 6.0 Major Surgery


ML-AIM Predictor (Beta)
ML-AIM PREDICTOR BETA
BREAST CANCER
COLON CANCER
LUNG CANCER
PROSTATE CANCER
HOW IT WORKS
CREDITS
Time since Initial Diagnosis (Months)
Stage
Nodes Involved
Grade
Detected by Screening
1 yr
2 yr
3 yr
4 yr
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
ADD NEW EVENT
13.6 27 21.0 53.0 39.0 RTD start point
15.5 15 48.0 41.0 50.0 Other tumor diagnosis
15.7 44 11.0 35.0 34.0 Genetic test test results reported
9.6 44 3.0 33.0 47.0 Genetic test test results reported
9.8 5 26.0 20.0 28.0 HES A&E attendance
15.8 22 34.0 39.0 38.0 Hormone from AV_TREATMENT
16.4 7 13.0 10.0 6.0 Major Surgery



Deep Sensing
[ICLR 2018]
ASAC
[ML4HC 2019]

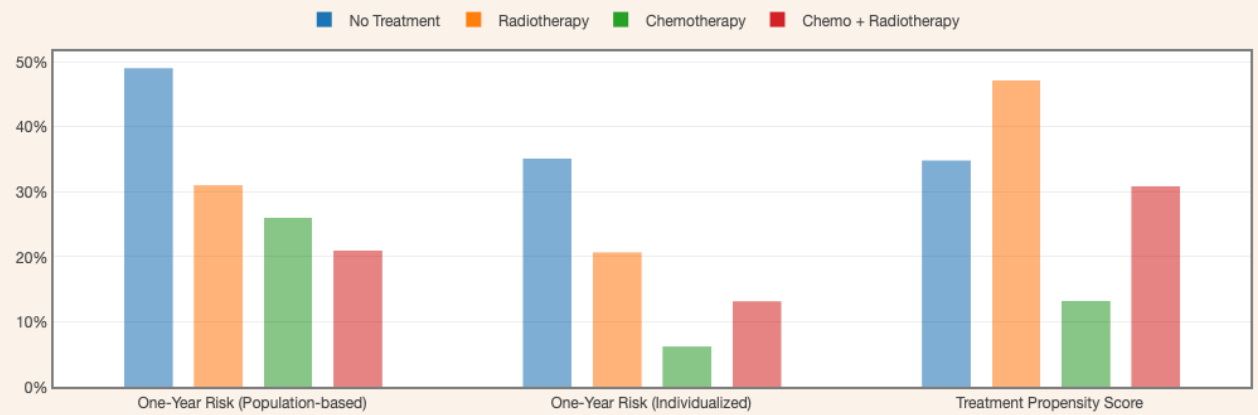
ML-AIM Predictor (Beta)

ML-AIM PREDICTOR BETA

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

×	☑	18.6	42	1.0	23.0	17.0	Genetic test sample analysis requested
×	☑	18.7	46	7.0	26.0	51.0	Chromosome-level results from genetic test
×	☑	20.0	18	29.0	23.0	39.0	Surgery from AV_TREATMENT
×	☑	21.0	45	29.0	40.0	35.0	Gene-level results from genetic test

Risk of Recurrence vs. Treatment Options



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

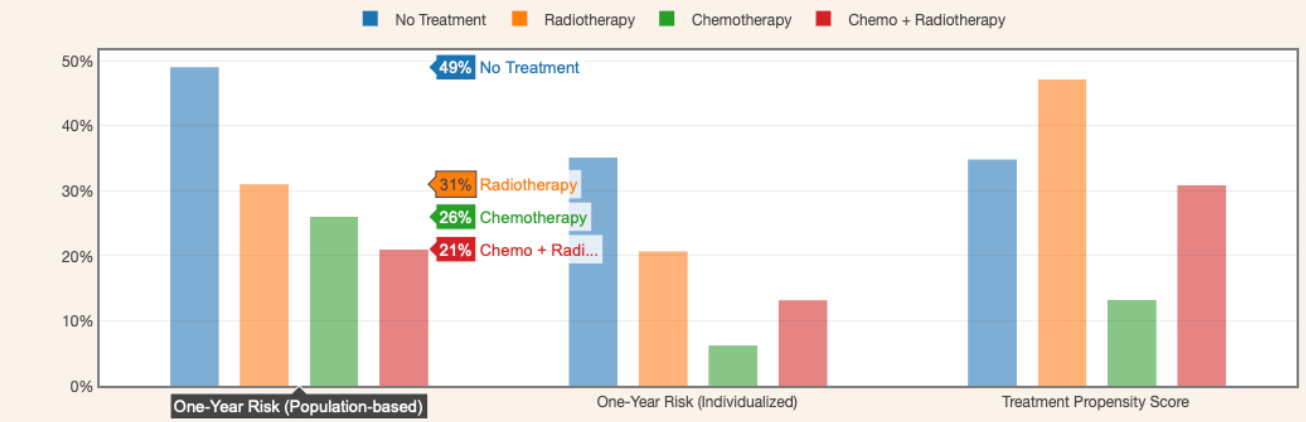
ML-AIM Predictor (Beta)

ML-AIM PREDICTOR BETA

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

×	☑	18.6	42	1.0	23.0	17.0	Genetic test sample analysis requested
×	☑	18.7	46	7.0	26.0	51.0	Chromosome-level results from genetic test
×	☑	20.0	18	29.0	23.0	39.0	Surgery from AV_TREATMENT
×	☑	21.0	45	29.0	40.0	35.0	Gene-level results from genetic test

Risk of Recurrence vs. Treatment Options



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

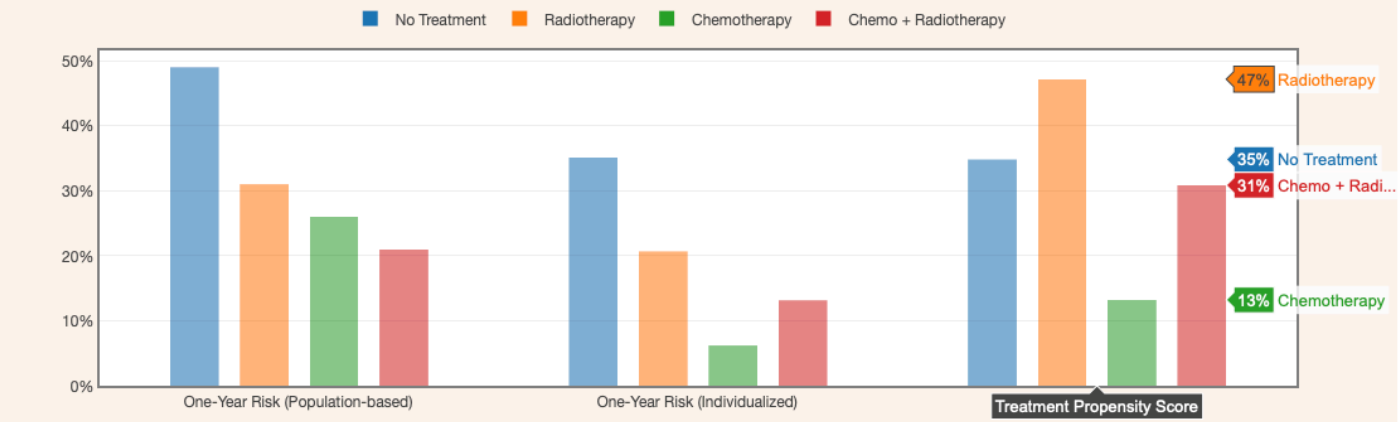
ML-AIM Predictor (Beta)

ML-AIM PREDICTOR BETA

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

18.6	42	1.0	23.0	17.0	Genetic test sample analysis requested
18.7	46	7.0	26.0	51.0	Chromosome-level results from genetic test
20.0	18	29.0	23.0	39.0	Surgery from AV_TREATMENT
21.0	45	29.0	40.0	35.0	Gene-level results from genetic test

Risk of Recurrence vs. Treatment Options



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

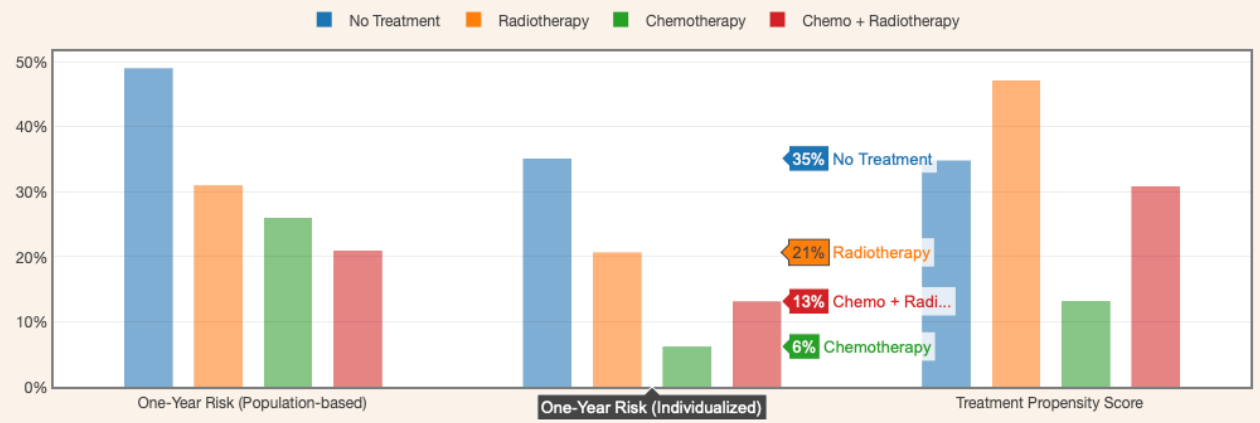
ML-AIM Predictor (Beta)

ML-AIM PREDICTOR BETA

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

18.6	42	1.0	23.0	17.0	Genetic test sample analysis requested
18.7	46	7.0	26.0	51.0	Chromosome-level results from genetic test
20.0	18	29.0	23.0	39.0	Surgery from AV_TREATMENT
21.0	45	29.0	40.0	35.0	Gene-level results from genetic test

Risk of Recurrence vs. Treatment Options



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

NSGP [NIPS 2017]
GANITE [ICLR 2018]
CMGP [ICML 2018]
Counterfactual Recurrent Nets [NIPS 2018]

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

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

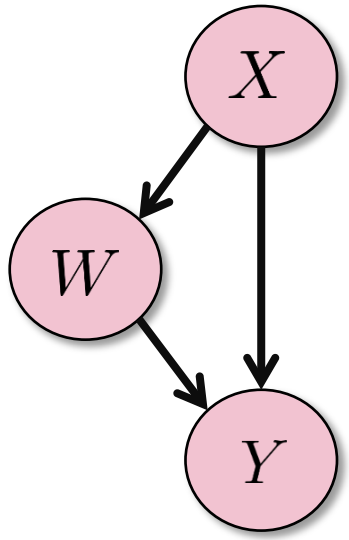
Real-world observational data
Scalable & adaptive implementation
Fast deployment
Cost-effective

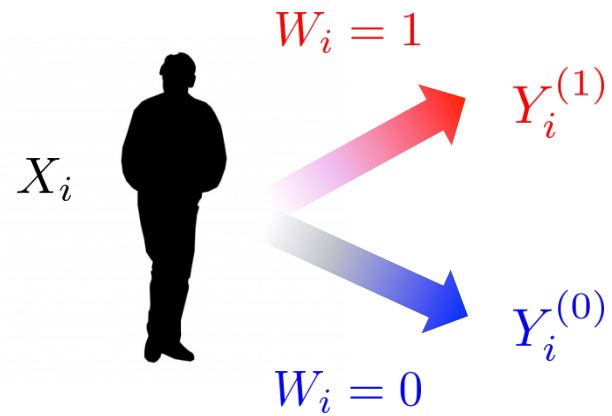
[Atan, vdS, 2015, 2018]
[Alaa, vdS, 2017, 2018, 2019]
[Yoon, Jordon, vdS, 2017]
[Lim, Alaa, vdS, 2018]
[Bica, Alaa, vdS, 2019]

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

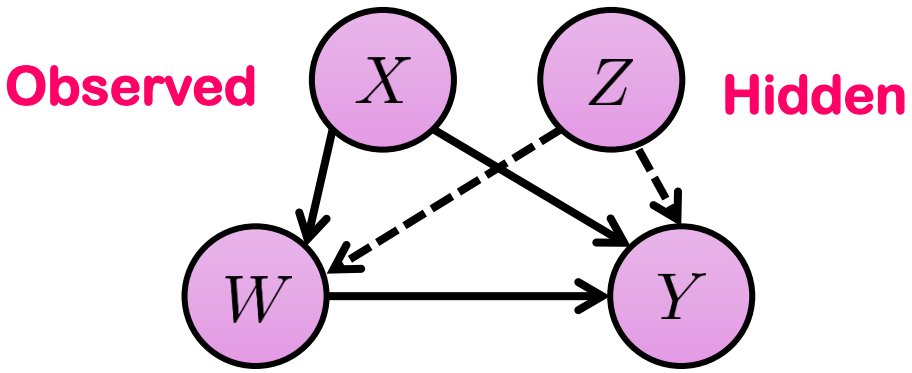
$$Y_i = W_i\, Y_i^{(1)} + (1 - W_i)\, Y_i^{(0)}$$

Causal effects

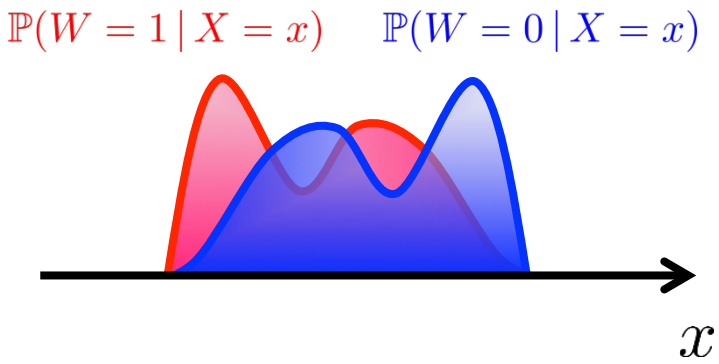
$$T(x) = \mathbb{E}\left[Y_i^{(1)} - Y_i^{(0)} \,\middle|\, X_i = x \right]$$

Assumptions

No unmeasured confounders (Ignorability)



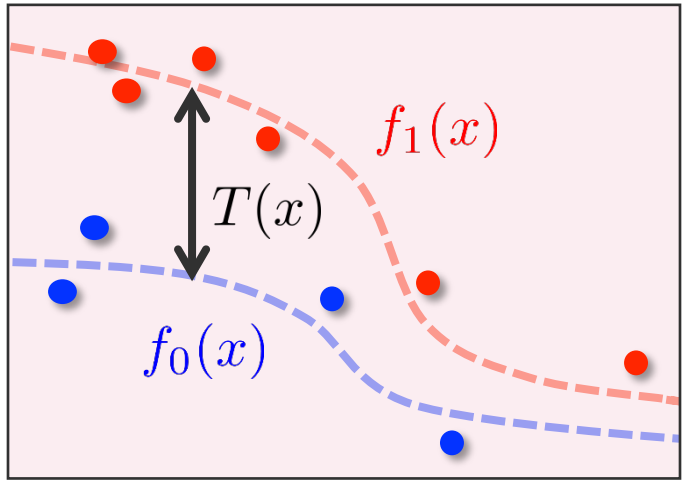
Common support



Estimating individualized treatment effects

- **Observational data**

$$(X_i, W_i, Y_i) \begin{cases} W_i = 1 \rightarrow (X_i, Y_i) \\ W_i = 0 \rightarrow (X_i, Y_i) \end{cases}$$



- **Treatment response surfaces**

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

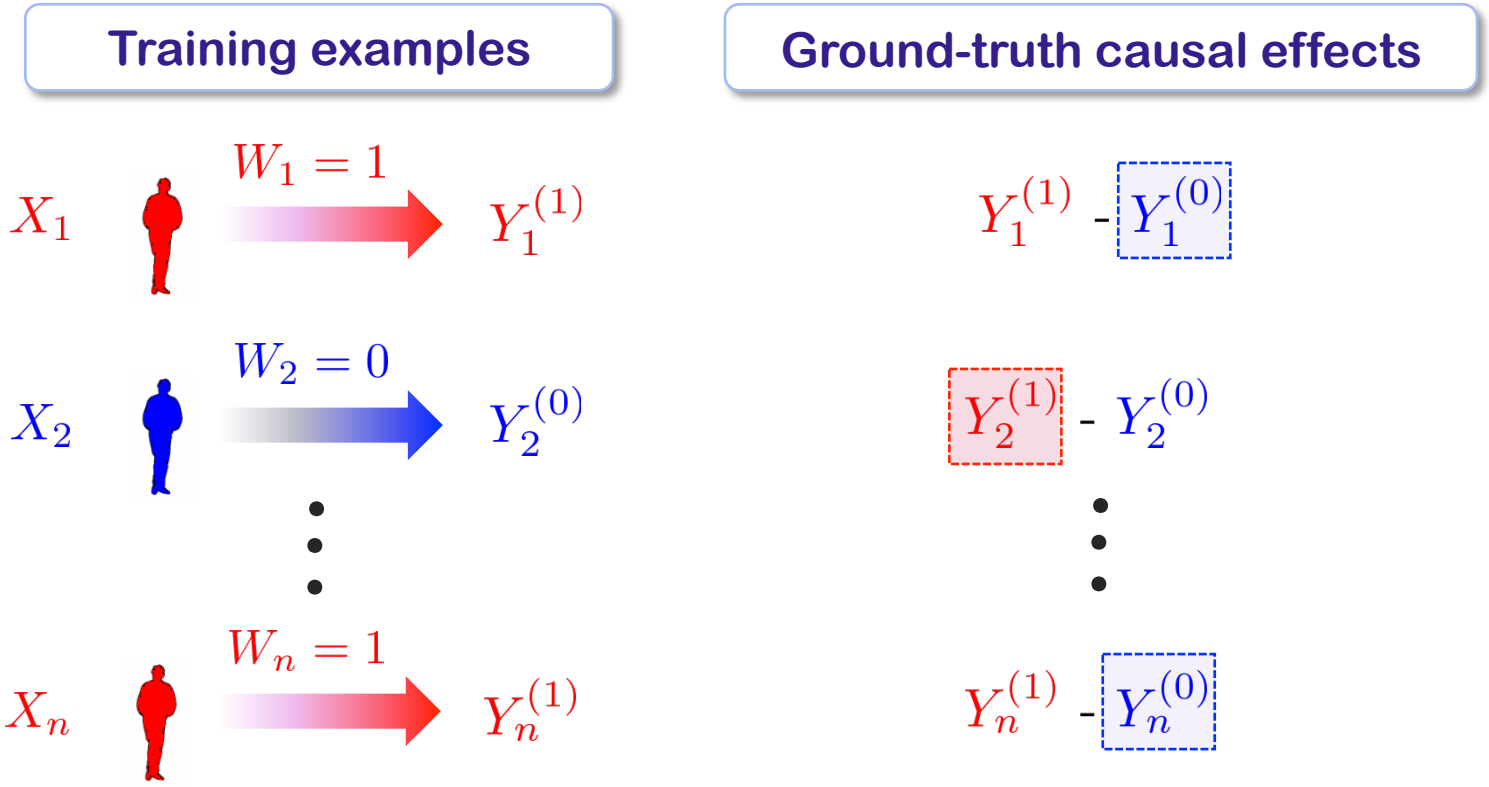
$$f_0(x) = \mathbb{E}[Y^{(0)} \mid X = x]$$

- **Estimate causal effects: individualized treatment effects**

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

Beyond supervised learning…

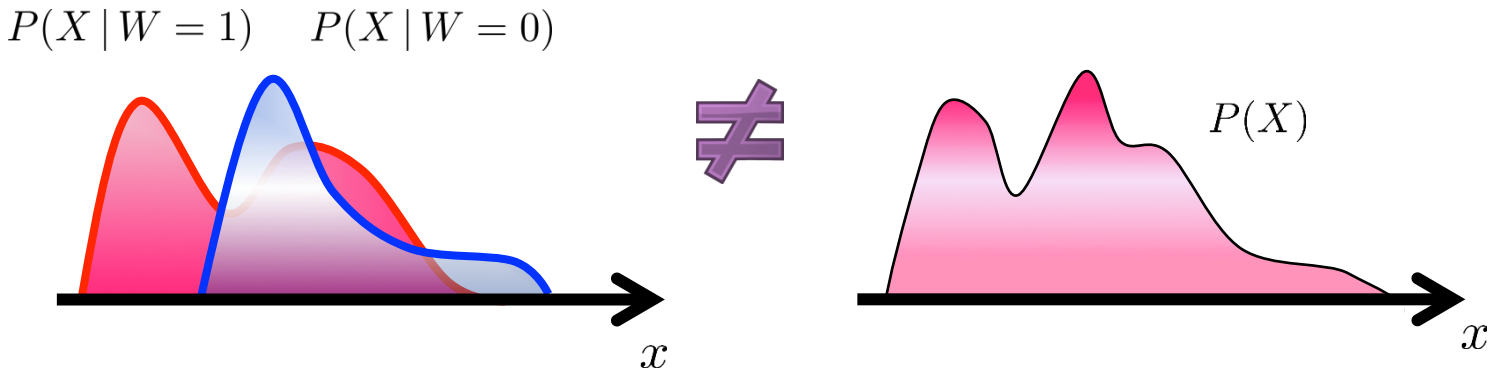
- **Fundamental challenge of causal inference:**
  we never observe counterfactual outcomes



Causal modeling ≠ predictive modeling

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

2- Selection bias → covariate shift:
training distribution ≠ testing distribution



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

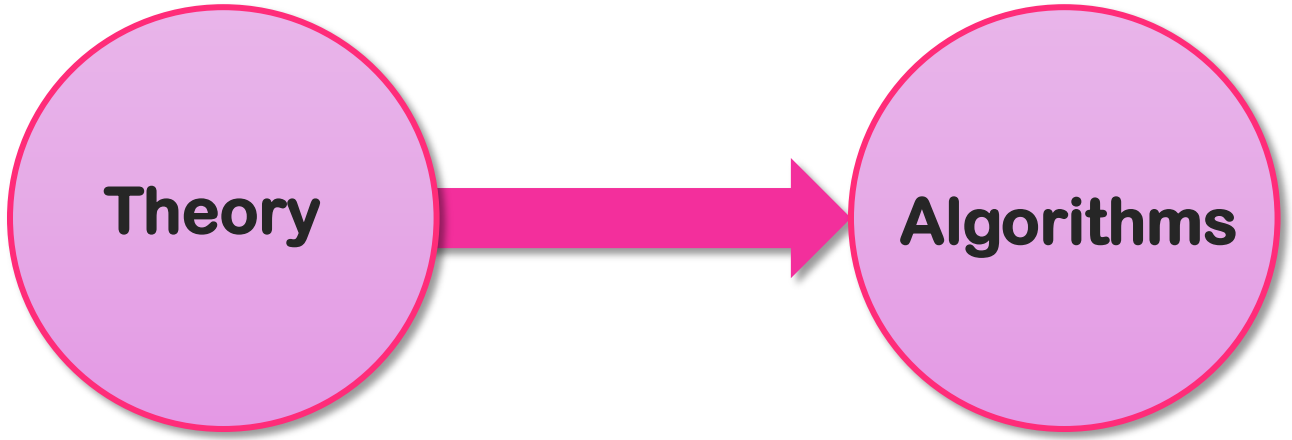
[Alaa, vdS, JSTSP 2017][ICML 2018]

What is possible?

(Fundamental limits)

How can it be achieved?

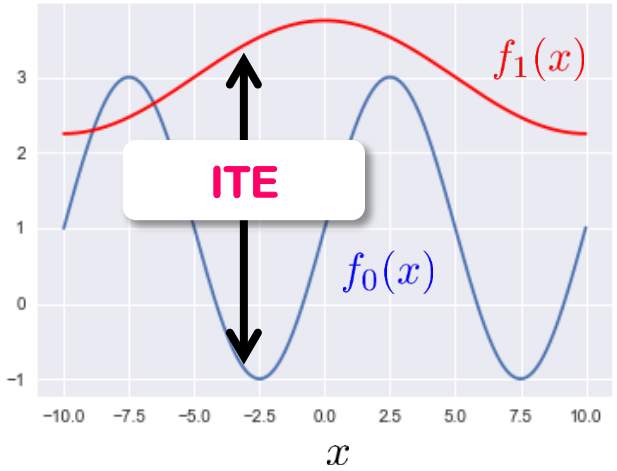
(Practical implementation)



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** $\hat{T}(.)$ **induced by Bayesian posterior** $d\Pi_n(T \mid \mathcal{D})$
  - **Precision of estimating heterogeneous effects** $\mathrm{PEHE}(\hat{T}) \triangleq \mathbb{E} \parallel \hat{T} - T \parallel^2_{L^2(\mathbb{P})}$

Minimax Rate for ITE Estimation

- **Depends on the "complexity" of $f_0(x)$ and $f_1(x)$ …**

Sparsity d

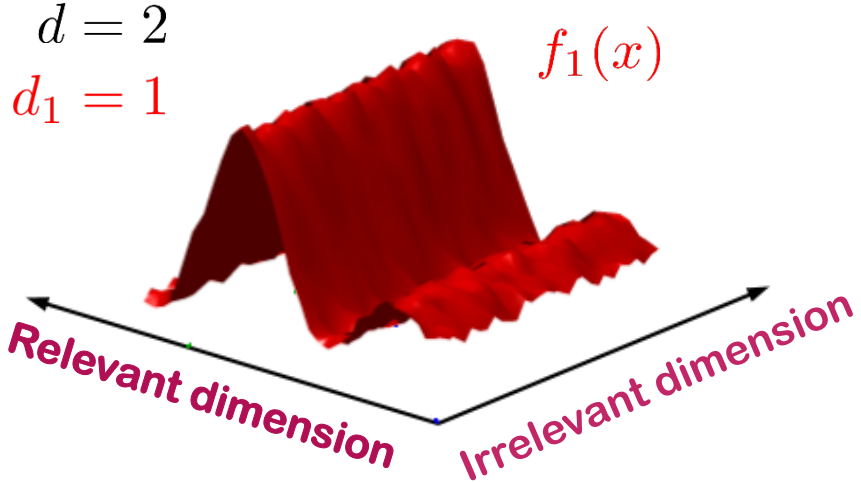
$f_0(x)$ → d_0 **relevant dimensions**

$f_1(x)$ → d_1 **relevant dimensions**

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

$d = 2$

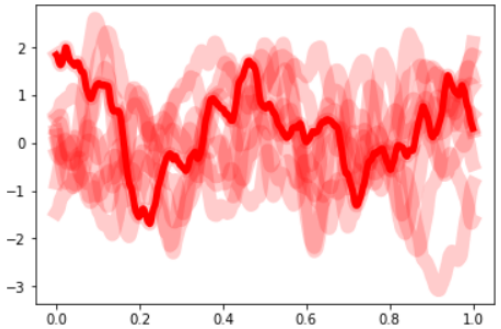
$d_1 = 1$



Smoothness α

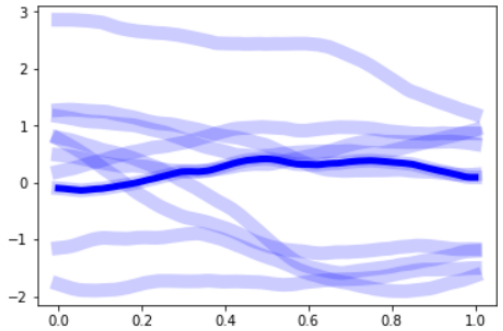
$f_0(x)$ → **Hölder ball** H^{α_0}

$f_1(x)$ → **Hölder ball** H^{α_1}



$\alpha_1 \downarrow\downarrow$

Rough functions



$\alpha_0 \uparrow\uparrow$

Smooth functions

Minimax Rate for ITE Estimation

- **Theorem 1**

  The minimax rate for ITE estimation is given by:

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

Should we care about selection bias?

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

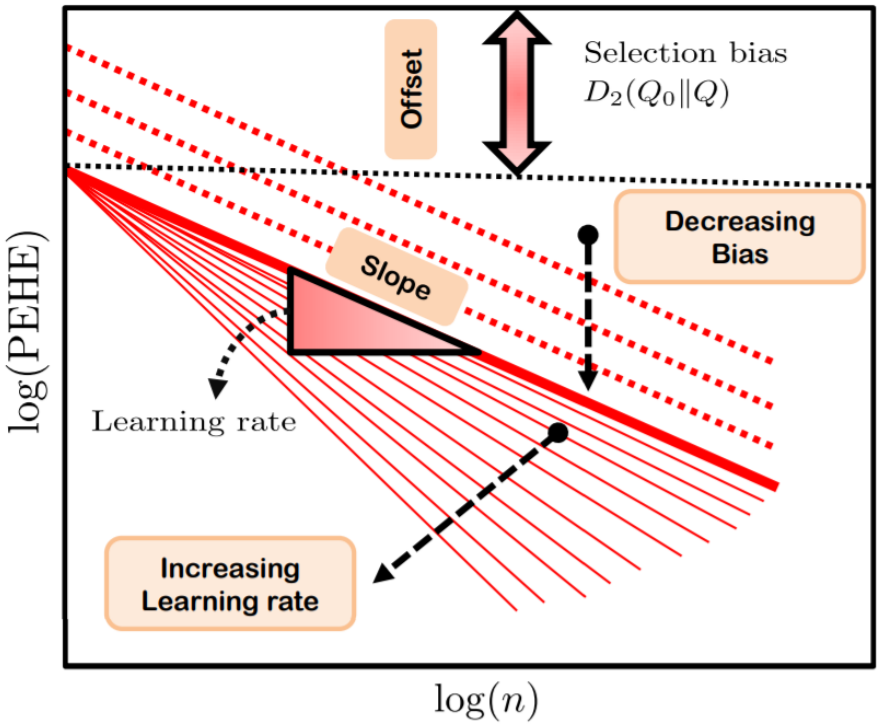
Minimax-optimal estimator

Rényi Divergence

Offset

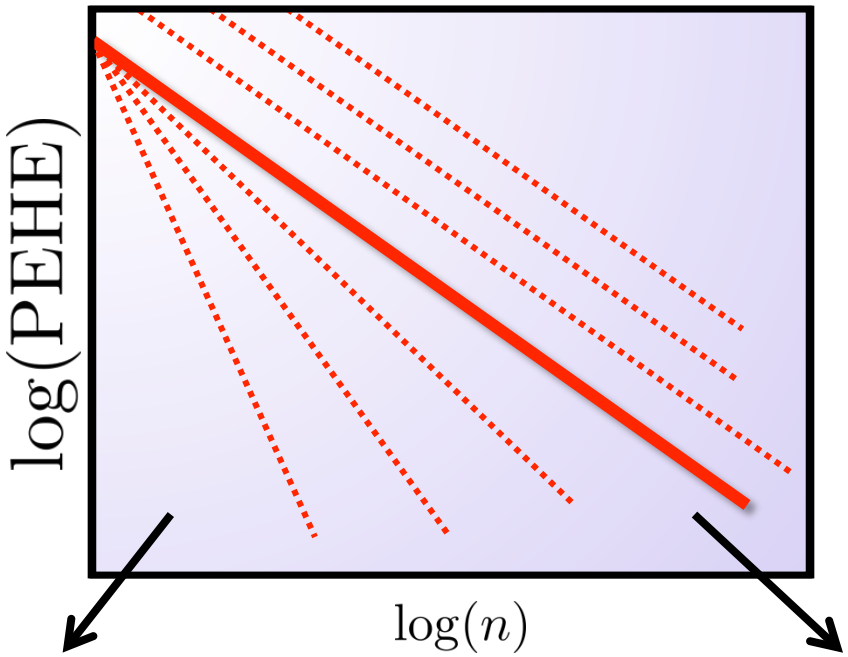
$$\log(\mathrm{PEHE}(\hat{T})) \approx D_2(\mathbb{P}(X \mid W=0) \| \mathbb{P}(X)) + D_2(\mathbb{P}(X \mid W=1) \| \mathbb{P}(X)) + \log(C) - \frac{2\alpha_0}{2\alpha_0 + d_0} \log(n).$$

Slope



Theory guides model design

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



Small sample regime

- Handling selection bias
- Sharing training data between response surfaces

Large sample regime

- Flexible model and hyperparameter tuning

ITE Estimation using Multi-task Gaussian processes
[Alaa and vdS, NIPS 2017, ICML 2018]

- **Multi-task Gaussian Process** [Bonilla et al., 2008].

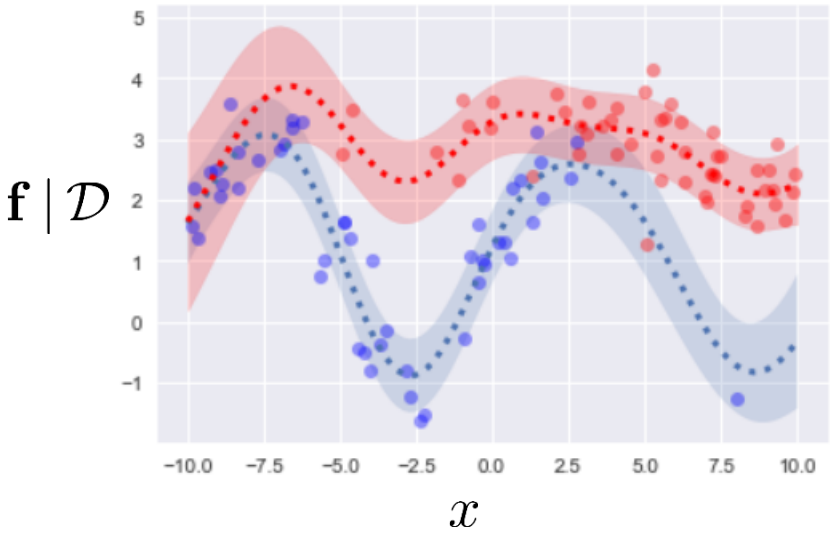
$$f_0, f_1 \sim \mathcal{GP}(0, \mathbf{K}_{\beta_0,\beta_1})$$

Matern kernel = Prior over on vvRKHS $H^{\beta_0} \times H^{\beta_1}$

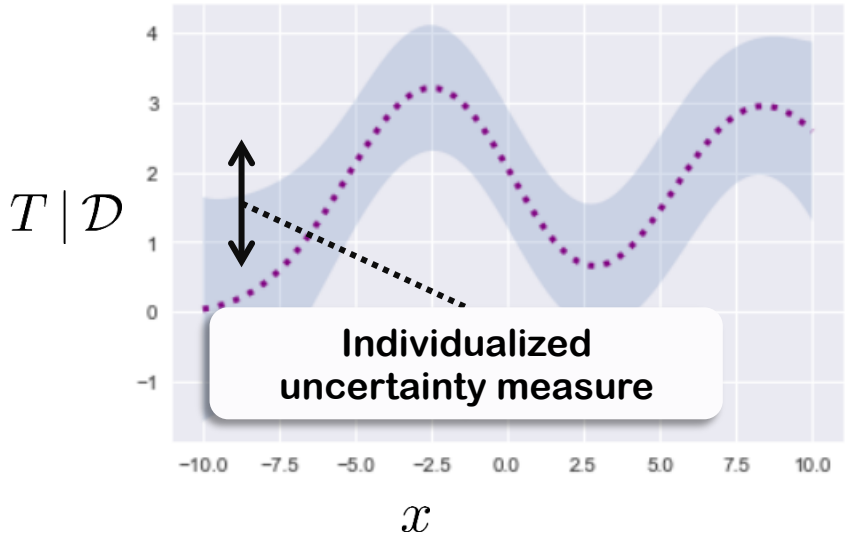
$$\mathbf{K}_\theta(x, x') = \mathbf{A}_0\, k_{\beta_0}(x, x') + \mathbf{A}_1\, k_{\beta_1}(x, x')$$

Shared representations!

Posterior potential outcomes distribution



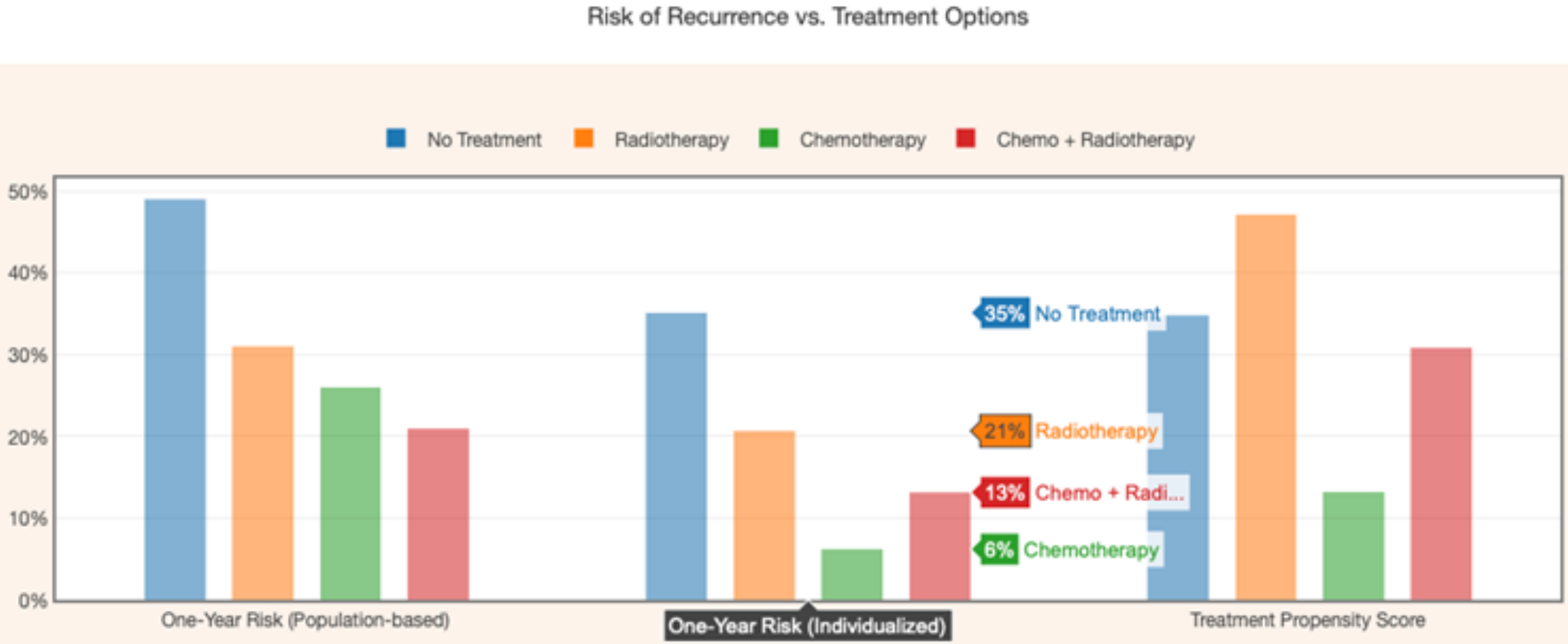
Posterior ITE distribution



Automated Feature Relevance Determination !

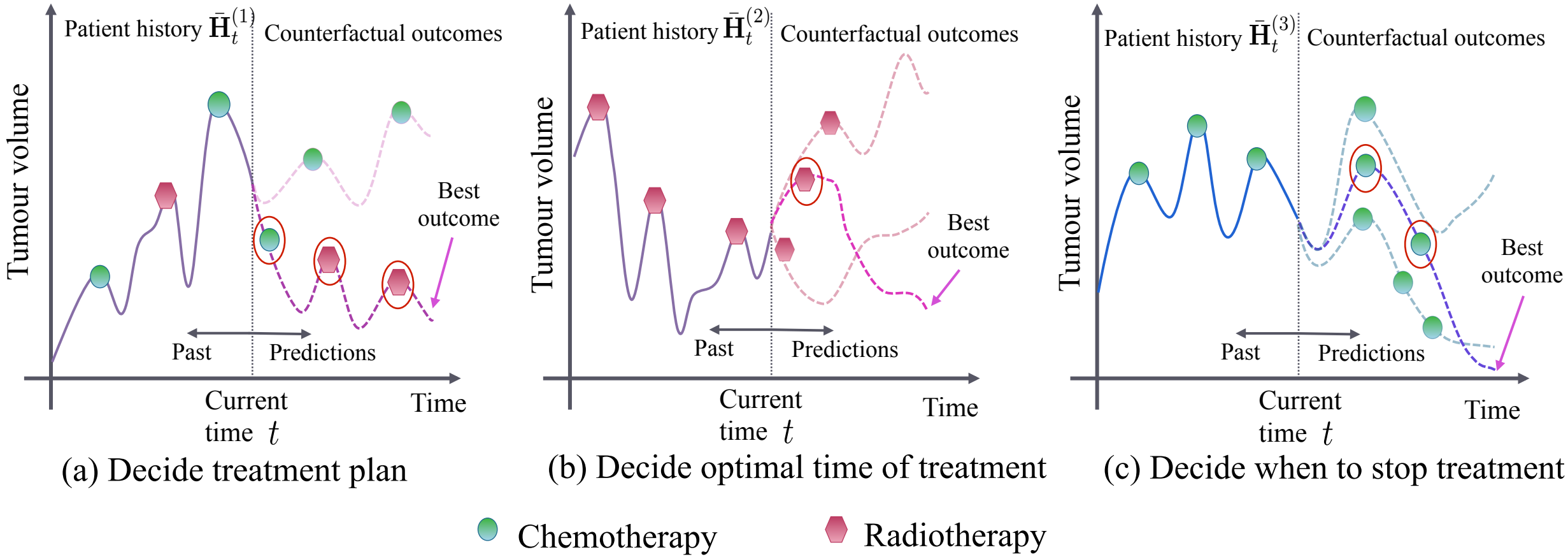
Multiple Treatments: GANITE [Yoon, Jordon, vdS, ICLR 2018]

Estimation of Individualized Treatment Effects using Generative Adversarial Nets



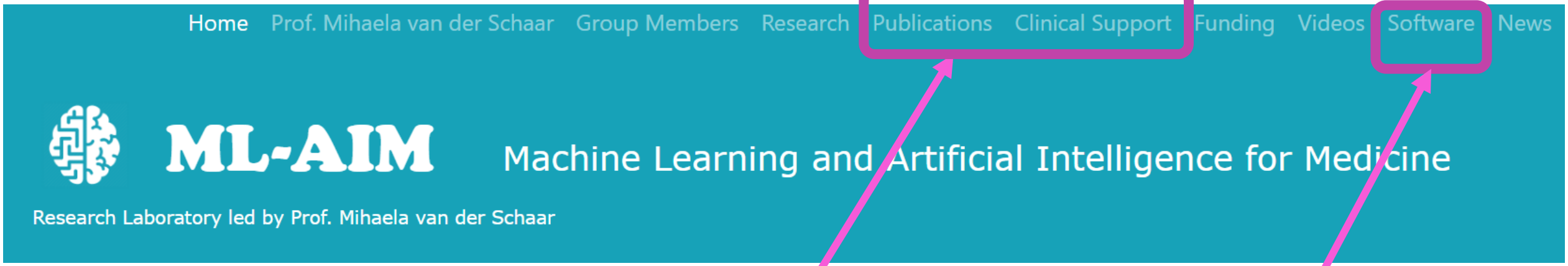
Individualized Treatment Effects over Time
[Lim, Alaa, vdS, NeurIPS 2018][Bica, Alaa, vdS, 2019]

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



(a) Decide treatment plan (b) Decide optimal time of treatment (c) Decide when to stop treatment

Chemotherapy Radiotherapy



Details about our algorithms:

http://www.vanderschaar-lab.com

Details about our software:

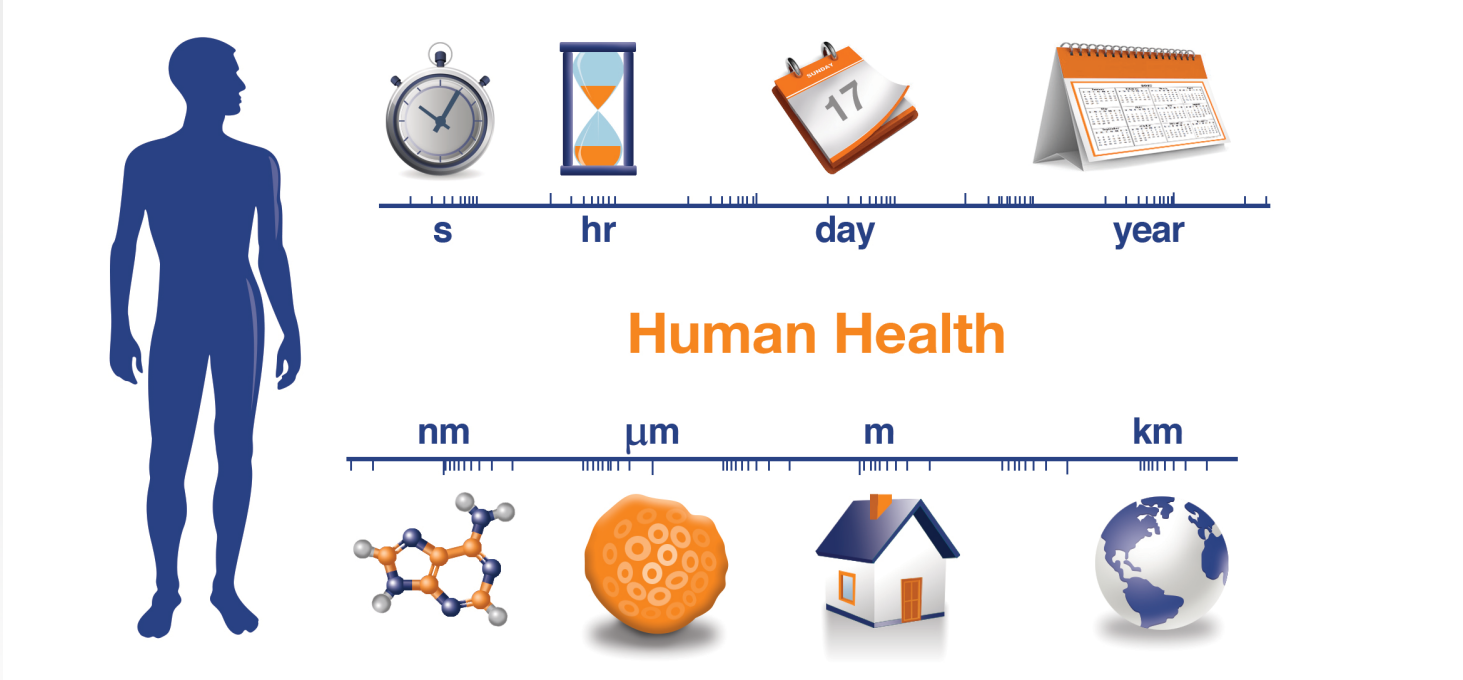
http://www.vanderschaar-lab.com

Juan Gutierrez,
University of Texas
at San Antonio

Professor and Chair of Mathematics

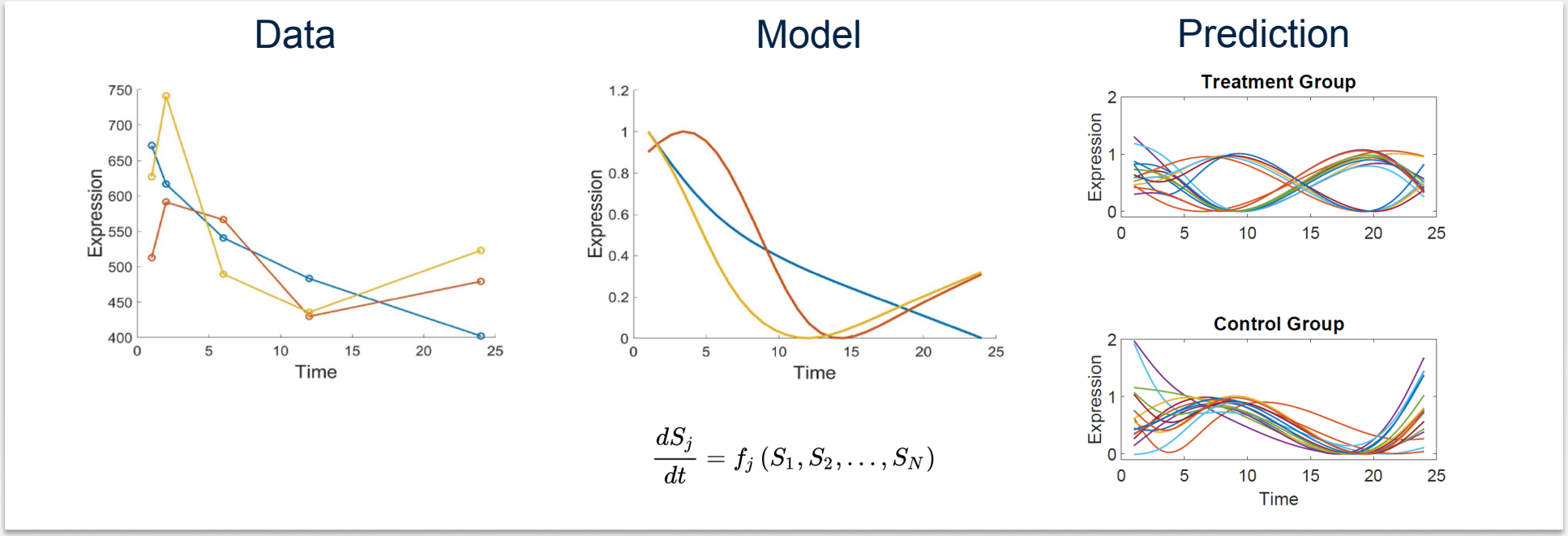
Machine Learning in Biomedical Sciences

Human Health Spans Multiple Temporal and Spatial Scales



The increased availability of data is changing how we approach this comprehensive understanding.

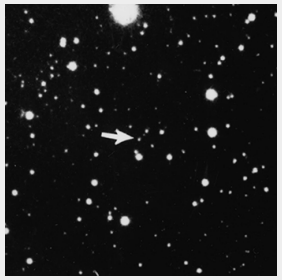
Data → Model → Prediction



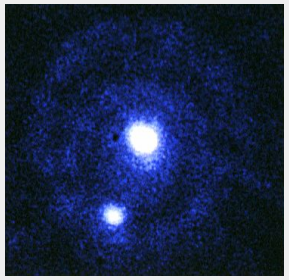
The canon in mathematical biology is to progress iteratively through data, models and predictions (not necessarily in that order).

This paradigm does not always work.

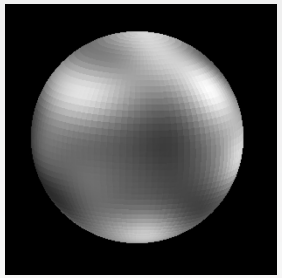
The questions have evolved in mathematical biology
Pluto: An analogy



18 Feb 1930
4.7 billion miles



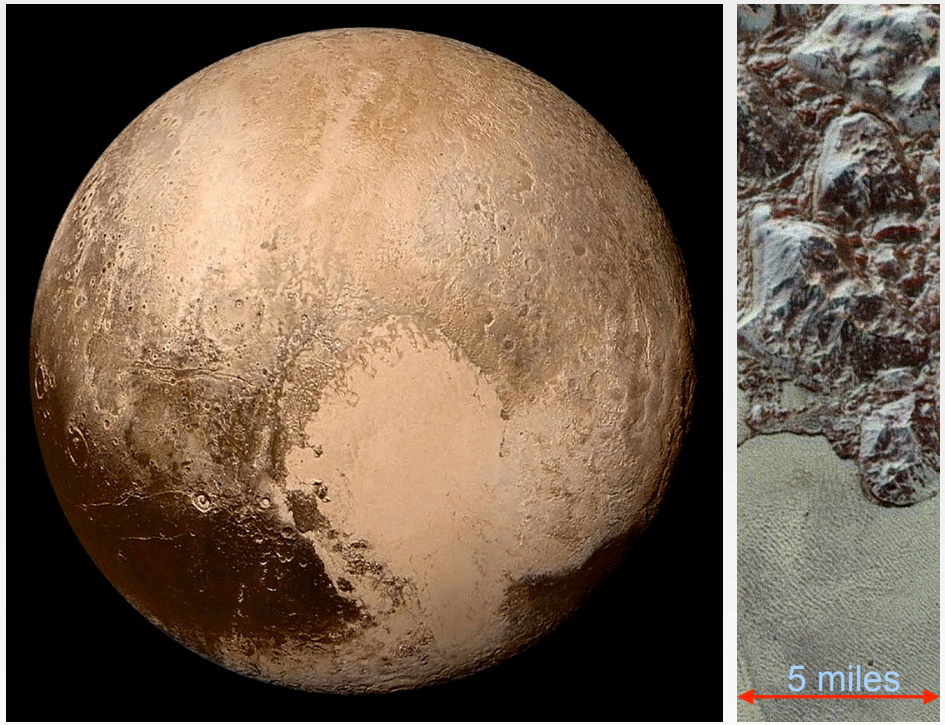
1994, HST/FOC



1996, HST/FOC

July 1, 2015 NHS

HST: Hubble Space Telescope
FOC: Faint Object Camera
NHS: New Horizons Spacecraft



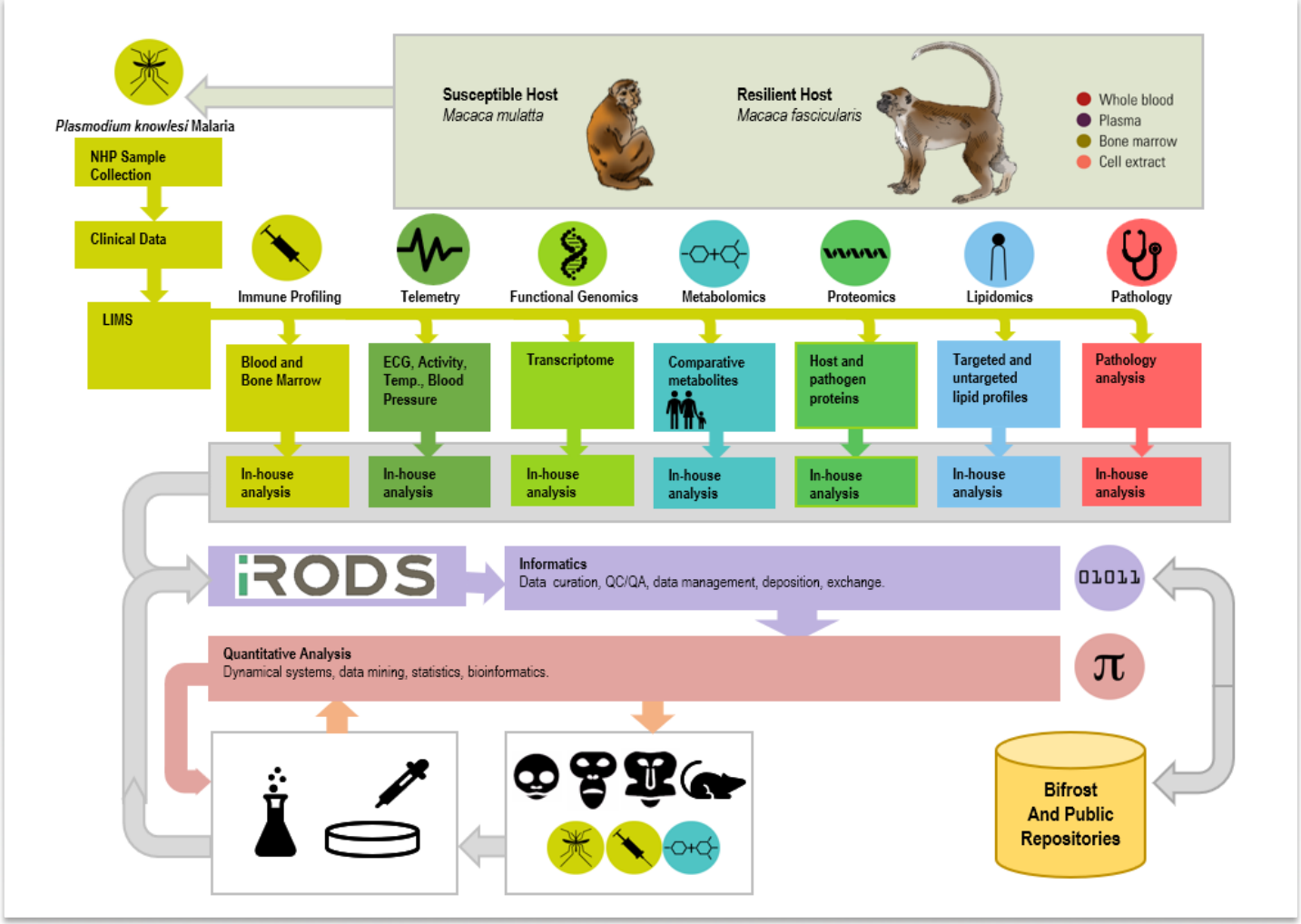
July 14, 2015, New Horizons Spacecraft

We went from "*where in the sky is this planet*" to "*name this mountain*".
The difference is ***availability of data***.

Given the brevity of this presentation, no technical details are discussed. Only the high-order organization of a case study are presented next.

AN EXAMPLE OF THE NEED OF MACHINE LEARNING: EARLY DETECTION OF DISEASE

A "New Horizons" for Biomedical Research: MaHPIC/HAMMER



US National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services contract #HHSN272201200031C, 2012-2017, which supported the *Malaria Host-Pathogen Interaction Center* (**MaHPIC**).

Defense Advanced Research Projects Agency (DARPA) and the US Army Research Office through the program *Technologies for Host Resilience - Host Acute Models of Malaria to study Experimental Resilience* (**THoR's HAMMER**), DARPA contract #W911NF-16-C-0008, 2016-2019

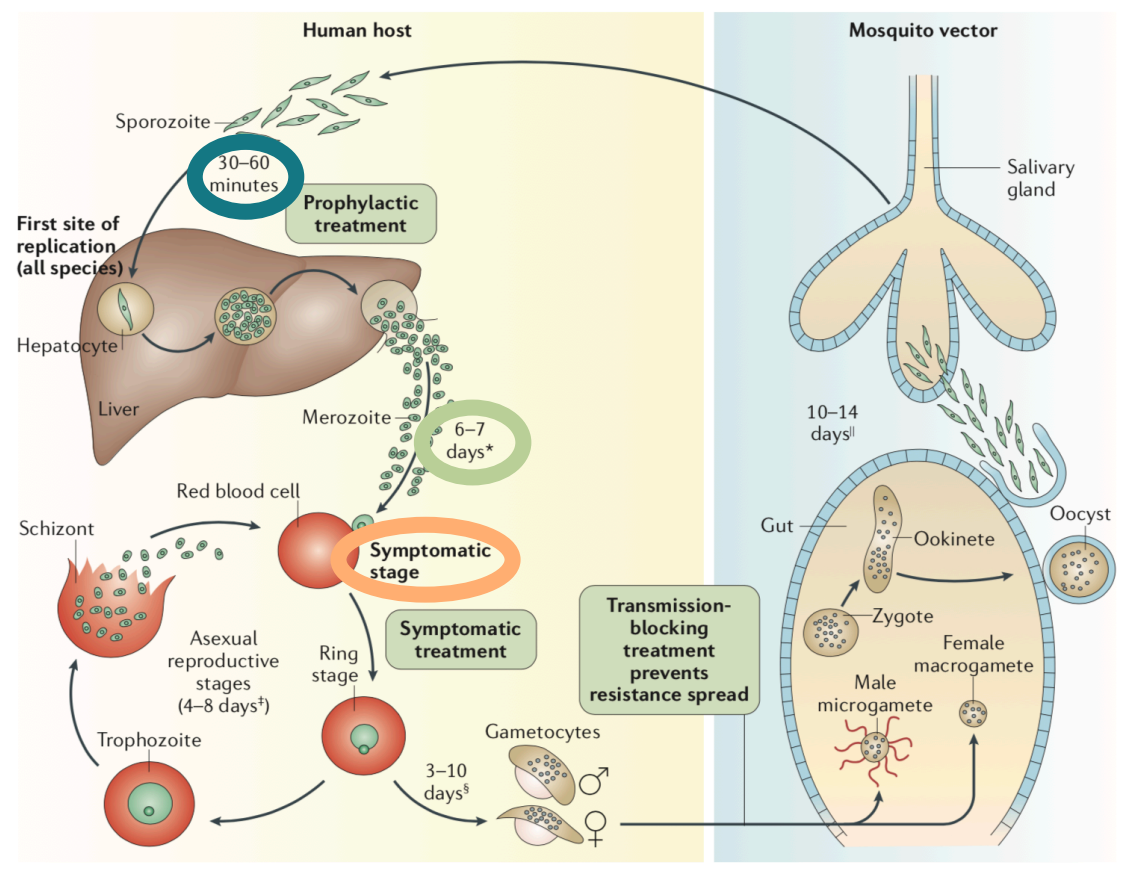
Terabytes of heterogenous biomedical data!!!

Two interesting questions can be addressed now that we have more data

- What causes severity of disease in some individuals, and mild or manageable disease in others?
  - Some individuals affected with malaria show no symptoms.
- Is it possible to detect disease before the onset of symptoms?
  - Once symptoms develop, severity might be impossible to stop.
  - Within 24 hours of the onset of symptoms, a severe malaria patient could die.

Can mathematics help answer these questions?

Malaria Parasite Life Cycle



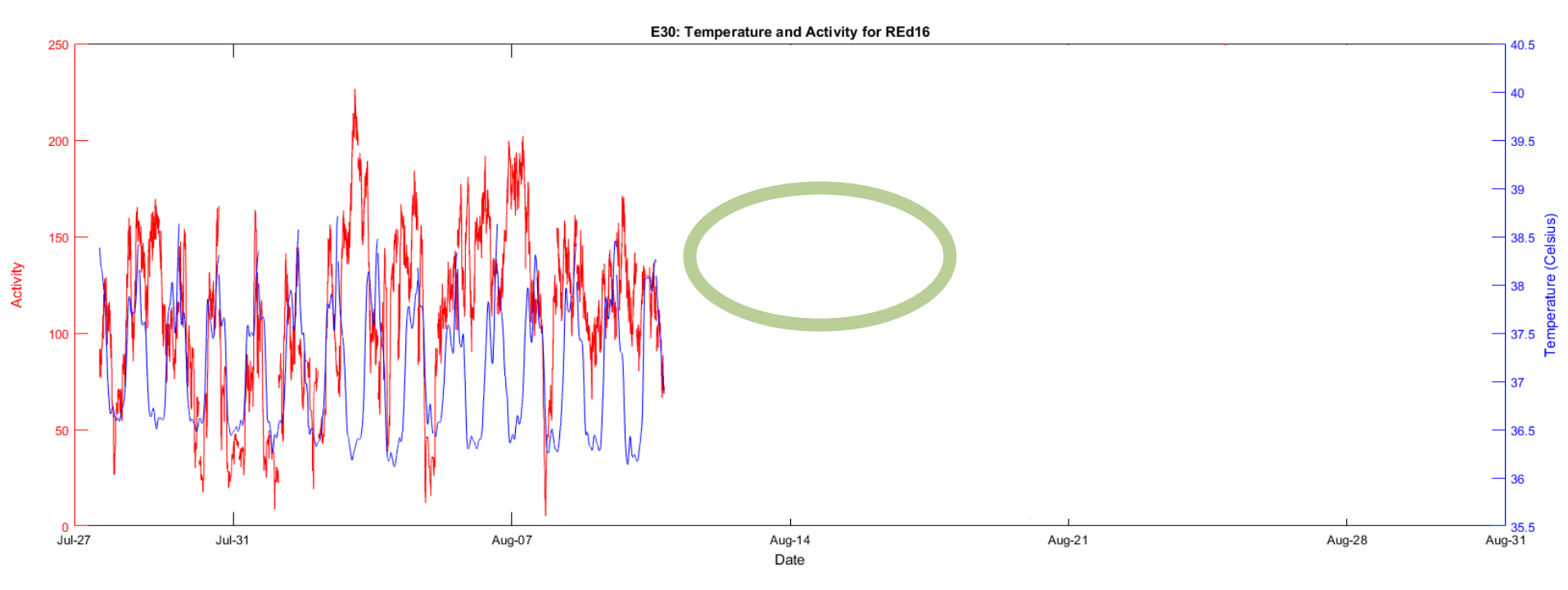
Margaret A. Phillips1, Jeremy N. Burrows2, Christine Manyando3, Rob Hooft van Huijsduijnen2, Wesley C. Van Voorhis4 and Timothy N. C. Wells2, Malaria Primer

Malaria killed 435,000 people worldwide in 2017, mostly kids under the age of 5 in developing countries. 219 million reported cases.

Currently, there is no early diagnostic test to confirm the presence of *Plasmodium* parasites in the liver.

Once symptoms erupt, complications can occur within 24 hours.

Focus on question 2: Early Detection of Disease



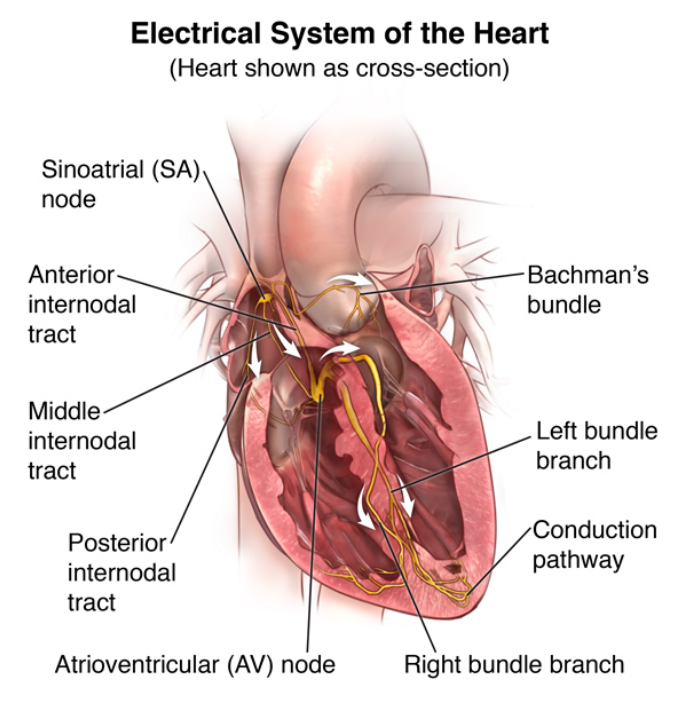
Early Detection of Disease

ED2

- With the entire time series, we could easily identify the different stages using: wavelets, Fourier analysis, etc.
- But in practice the entire time series is **not** available.
- The goal: detect infection as early as possible with as little data as possible.
- Hefty goal: Detect disease with 10 seconds of ECG data

Apple Watch Series 4
(Not an endorsement. Shown as an example)

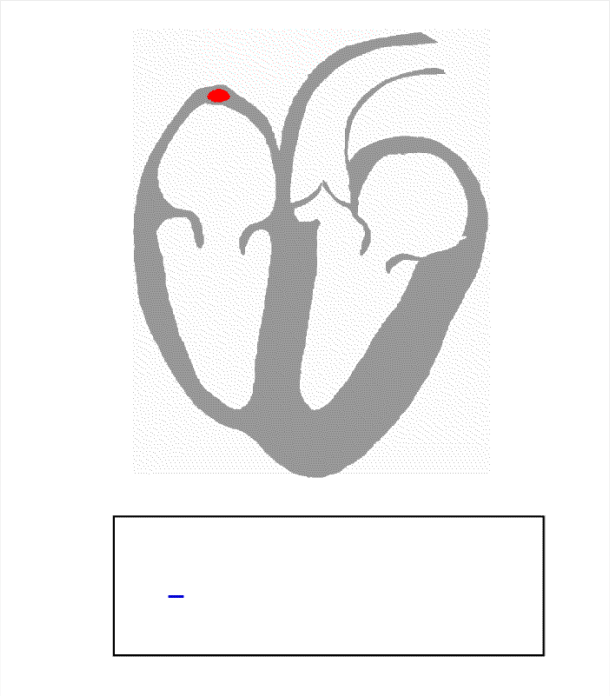
The heart is a pump made of muscle and regulated by electricity



- An electrical impulse is generated in the **SA**.
- It travels to the **AV**, where signal is slowed down briefly, then continue via the bundle of His into the ventricles.
- The right and left atria (upper chambers) contract first for a short time before the left and right ventricles (lower chambers)

Image source: http://www.hopkinsmedicine.org/healthlibrary/test_procedures/cardiovascular/signal-averaged_electrocardiogram_92,P07984/
Accessed on 1/31/2016

The sequence of depolarization results in a characteristic electrical pattern



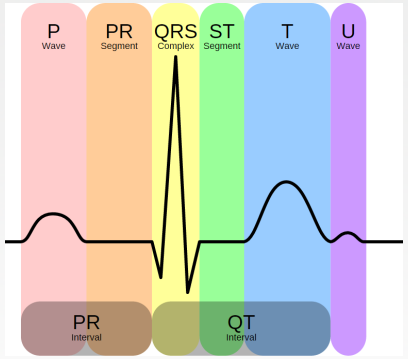
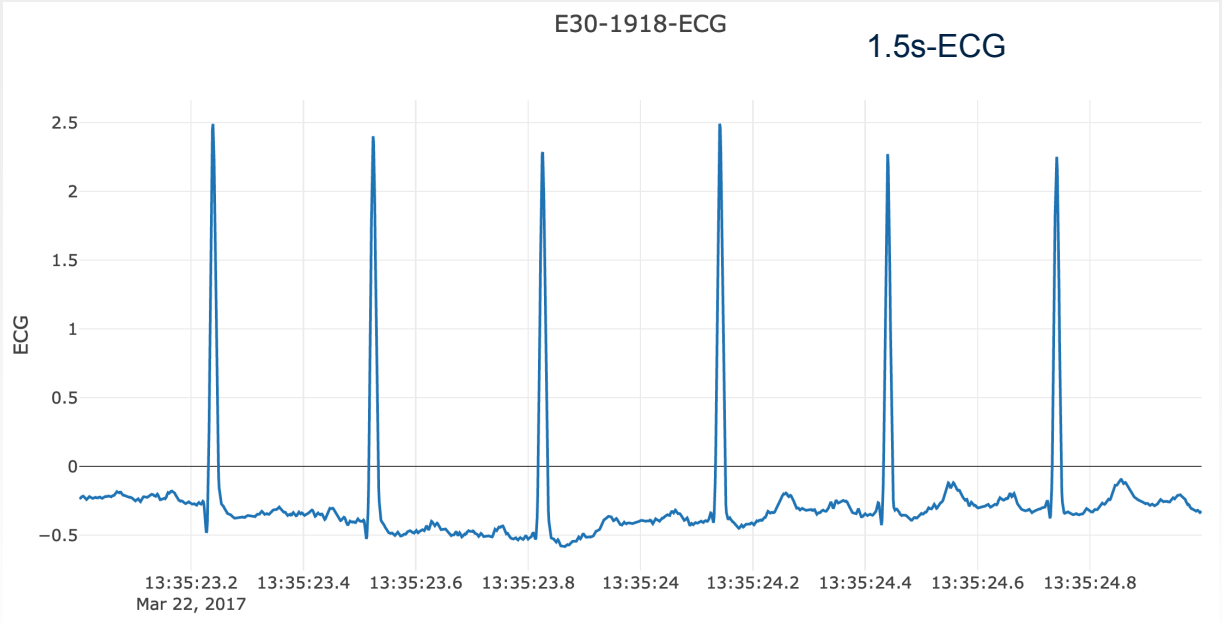


Image source: https://commons.wikimedia.org/wiki/File:ECG_Principle_fast.gif
Accessed on 1/31/2016

- Depolarization begins in the **right atrium** at the sinoatrial node (SA).
- It propagates to the **left atrium** first via the Bachman's bundle, and then interatrial septum (IS), anterior IS, and the coronary sinus.
- Then it propagates to the **atrioventricular node**, resulting in a large electrical discharge.
- It follows the path of the **Bundle of His**, and into the **Purkenje fibers**.

ECG Pre-processing

- Four days before inoculation and four days after inoculation.
- Each day was segmented into hourly intervals of data.
- Hour intervals were segmented into 10-second intervals.
- 13 subjects * 8 days * 24 hours * 60 minutes * 6 10-second segments
  = 898,560 observations. Each observation is a time series with 10,000 points



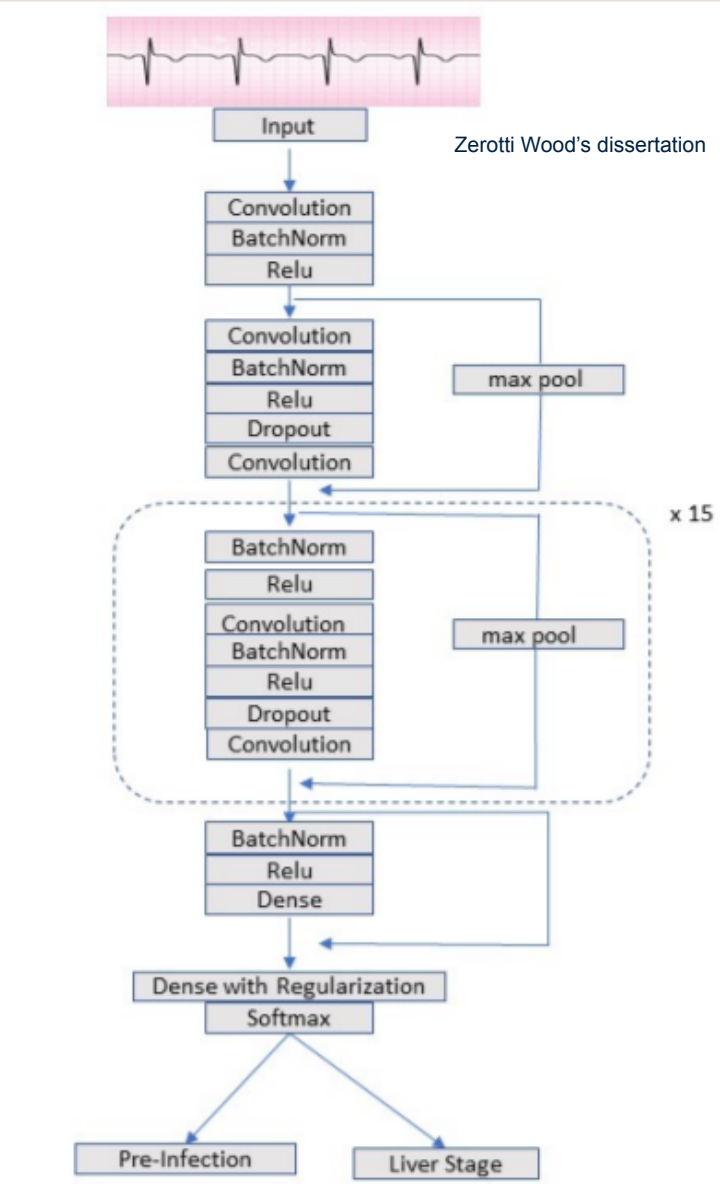
Tao Sheng's dissertation

The Canonical Methods Fail

- In trying to identify what 10-second segments of ECG data correspond to pre-infection of liver-stage, traditional methods were unable to surpass 70% accuracy (COSINOR models, wavelets, Fourier analysis).
- This opened the door to explore machine learning. Could a neural network detect these states with greater accuracy?

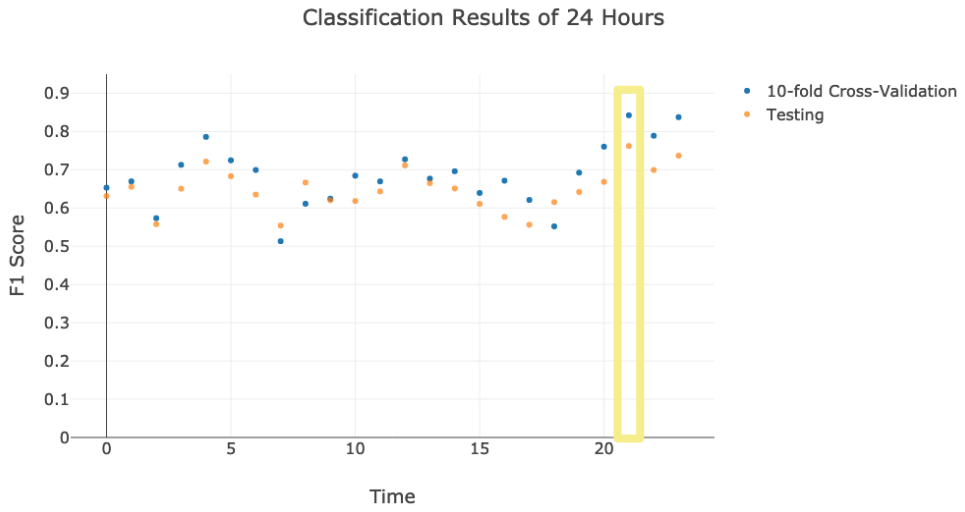
ECG Processing

- A random 90% of these observations were used to train a neural network.
- A random 10% was used as the development.
- Each hour was analyzed. Hours with lower activity (as measured with accelerometers) had better predictive power, as expected.
- 13 subjects * 8 days * 60 minutes * 6 10-second segments
  = 18,720 observations. Each observation is a time series with 10,000 points



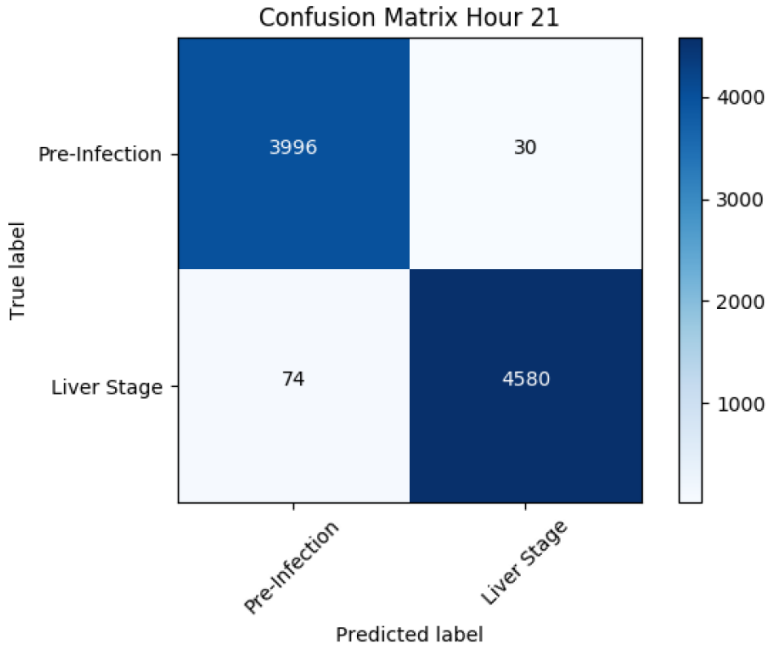
Early Detection of Disease via Telemetry

Using machine learning, we detected *Plasmodium* infection based on electrocardiogram (ECG) signals with over 98% accuracy during the liver stage of the disease, importantly, ***before*** the onset of symptoms that are caused by the blood-stage of the disease that is characteristic of malaria.



The later hours with less activity yield the best classification

Tao Sheng's dissertation



	Precision	Recall	F_1 score	Support
Pre-Infection	.982	.993	.987	4026
Liver Stage	.993	.984	.989	4654

Zerotti Wood's dissertation

Machine Learning: Improved Cognition for Humans

- ML allows detection of phenomena that remains undetectable through other methods.
- It requires large data sets.
- ML present challenges that open opportunities for foundational mathematical work.

Challenges

- We looked at a particular study in a specific context, but there are traits that seem to be universal when we consider machine learning applied to medicine.
- Data collection and management is an often overlooked and a deep source of complications.
- There are foundational questions that deserve attention:
  - What are (necessary and sufficient) conditions for consistency of a learning process?
  - How to accelerate the rate of convergence of the learning process?
  - Why does it work

Challenges

- Machine learning adds value to quantitative studies, particularly when we understand poorly a phenomenon, or when a well understood phenomenon becomes too complex to model.
- There is a delicate balance between knowledge and understanding. In this case, a neural network detected a phenomenon we poorly understand. It opens opportunities for basic research.
- The expertise required to undertake such studies is broad and takes a long time to develop. But this poses a conundrum: How to empower the next generation of scientists to execute multi-scale studies?

FUTURE DIRECTION: Early Detection of Disease

Early **D**etection of **D**isease

ED2

The ability to detect disease before symptoms occur could revolutionize health care and public health.

Malarial liver-stage infections could be detected with only 10 seconds of ECG data.

Other diseases may have unique signatures as well.

ED2 could be the next big thing. It would be a historical win, comparable to the Internet and GPS.

FUTURE DIRECTION: Early Detection of Disease

1

Apple Watch Series 4

ECG, accelerometers, and other telemetry measures, have reached the consumer market. There is a complex ecosystem of capabilities distributed across many manufacturers.

2

The roadblock for broad adoption of these findings is the inability to reconcile telemetry data of hundreds of thousands of subjects with medical records to train artificial intelligence classifiers.

3

In the US, *only the US Armed Forces have the capability to create this market* (large personnel + VA).

The outcome would be a low-cost mechanism to detect physiological changes and promote early interventions.

JUAN.GUTIERREZ3@UTSA.EDU

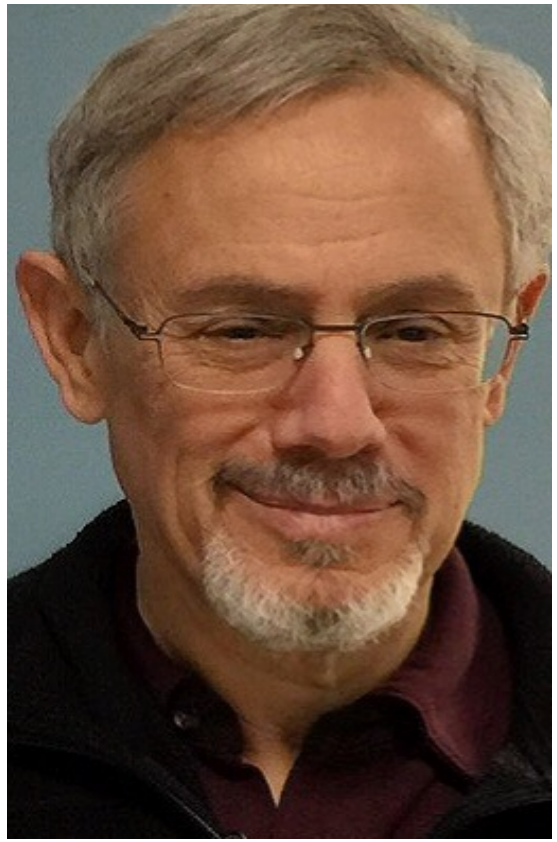
THANKS FOR YOUR ATTENTION!

MATHEMATICAL FRONTIERS

Machine Learning in Medicine

**Mihaela van der Schaar,
University of Cambridge,
Alan Turing Institute, and UCLA**

**Juan Gutierrez,
University of Georgia**



**Mark Green,
UCLA (moderator)**

View webinar videos and learn more about BMSA at www.nas.edu/MathFrontiers

MATHEMATICAL FRONTIERS

2019 Monthly Webinar Series, 2-3pm ET

February 12: *Machine Learning for Materials Science**

March 12: *Mathematics of Privacy**

April 9: *Mathematics of Gravitational Waves**

May 14: *Algebraic Geometry**

June 11: *Mathematics of Transportation**

July 9: *Cryptography & Cybersecurity**

August 13: *Machine Learning in Medicine*

September 10: *Logic and Foundations*

October 8: *Mathematics of Quantum Physics*

November 12: *Quantum Encryption*

December 10: *Machine Learning for Text*

Made possible by support for BMSA from the
National Science Foundation
Division of Mathematical Sciences
and the
Department of Energy
Advanced Scientific Computing Research

** Webinar posted*

View webinar videos and learn more about BMSA at www.nas.edu/MathFrontiers