Where machine learning fits into health and economics

Ziad Obermeyer
UC Berkeley and NBER

Today: ML in health

- The ML playbook so far: Automation of human judgment
  - Reduce cost, eliminate noise
- But automation seems like an unambitious goal
  - Also: replicates all the problems in human judgment
- Today: Some more interesting uses of ML
  - Along the way: questions this opens up
    - …beyond automation of human labor
  - The econ toolkit has a huge role to play here

Testing for heart attack: A microcosm of a broken system

- Over-use: up to 90% of tests are wasted
  - Exposing patients to costs, risks, with no benefit

- Assumption: Test value depends on result (ex post)
  - Positive tests have net benefit:[1] treating heart attack
  - Negative tests have only costs:[2] financial, health risks

- If we knew risk, we'd make better decisions (ex ante)
  - High-risk patients: Test, unlock treatment benefits
  - Low-risk patients: Don't test, avoid risks and costs

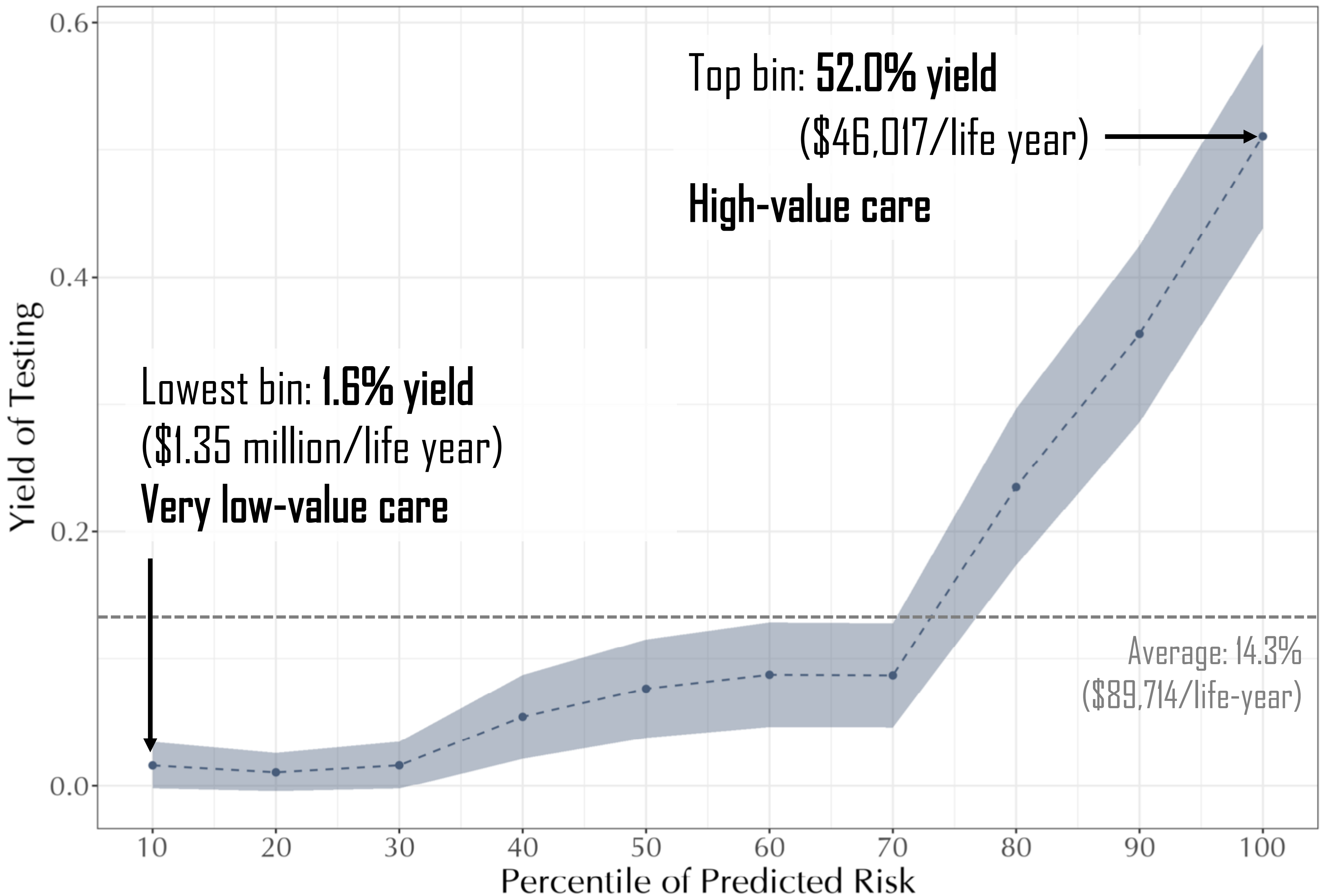
[1] As risk $\longrightarrow$ 1 the test becomes less valuable, but mechanically the test is still required to know where to put the stent
[2] This assumes there is no intrinsic value of 'knowing' heart attack is not present

Machine learning solves this kind of prediction problem

- Form explicit predictions on heart attack (blockage) risk
  - In tested ER patients: predict test outcome Y with X
  - Find potential errors: patients with mismatched $\hat{Y}$ vs. T
- But algorithm ≠ arbiter of truth: We don't assume it's right
  - Physician has information advantage based on Z
  - Many signals for risk, treatment benefit unobserved
- So actual errors are identified using health outcomes
  - In tested: Test results—is patient having heart attack?
  - In untested: Detective work—was heart attack missed?

Mullainathan & Obermeyer, *QJE* 2022

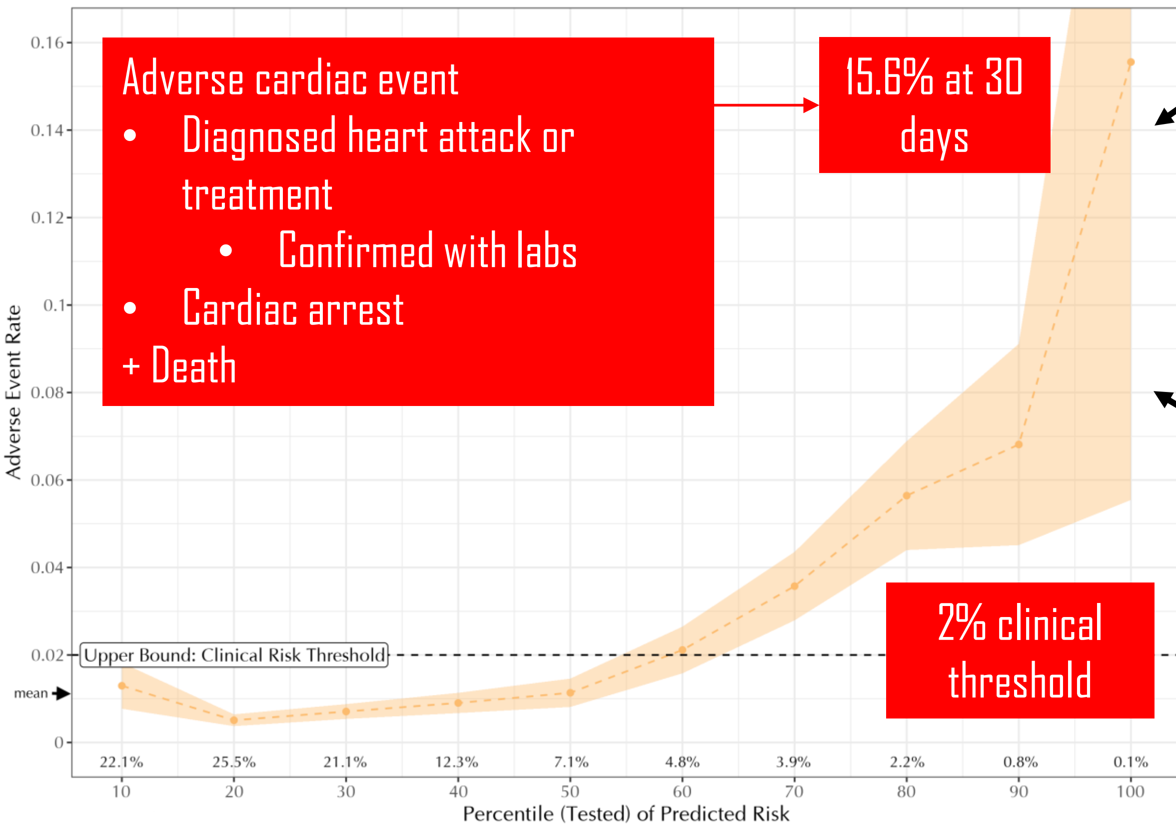
Tested: Over-testing low-risk → low yield



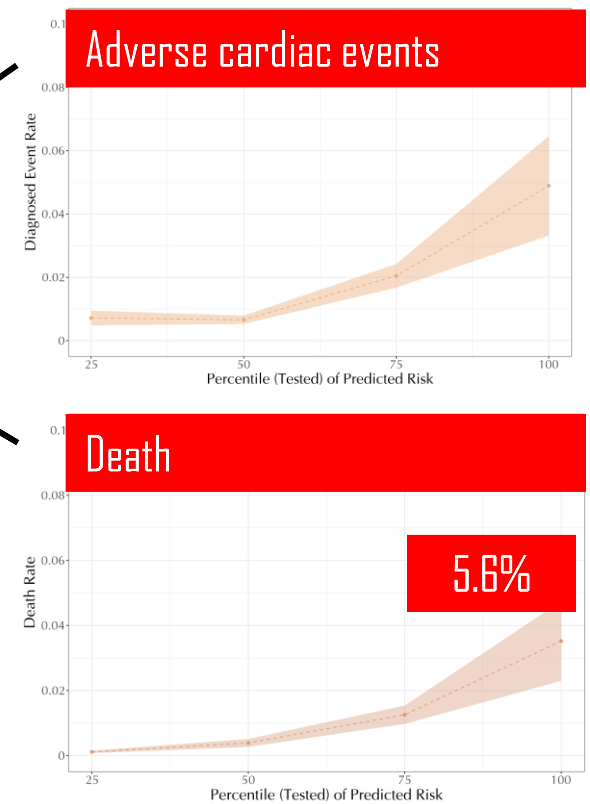
Mullainathan & Obermeyer, *QJE* 2022

Untested: Under-testing high-risk ⟶ high adverse event rate *excluded: frail, life-limiting illness, diagnosed heart problem in ER

Total Adverse Event Rate



Components



Mullainathan & Obermeyer, *QJE* 2022

More direct evidence of under-testing

	Diagnosed Event (31-365)	Death (31-365)	Death (0-365)
	(1)	(2)	(3)
Panel (a): Average Effect			
Predicted Risk	0.05^{***}	0.15^{***}	0.25^{***}
	(0.005)	(0.01)	(0.01)
Shift Test Rate	0.02	0.005	0.005
	(0.01)	(0.01)	(0.02)
Observations	123,289	123,289	123,289
Panel (b): Heterogeneous Effect By Risk			
Predicted Risk	0.06^{***}	0.17^{***}	0.27^{***}
	(0.01)	(0.01)	(0.01)
Shift Test Rate	0.04^{**}	0.04^{**}	0.04^{*}
	(0.02)	(0.02)	(0.02)
Predicted Risk × Shift Test Rate	-0.25^{*}	-0.49^{***}	-0.43^{**}
	(0.15)	(0.17)	(0.20)
Observations	123,289	123,289	123,289

No effect of testing <u>on average</u>

'Flat of the curve' health care

Large effect in <u>high-risk only</u>

Move to high-test: 2.5 p.p. (32%) lower mortality

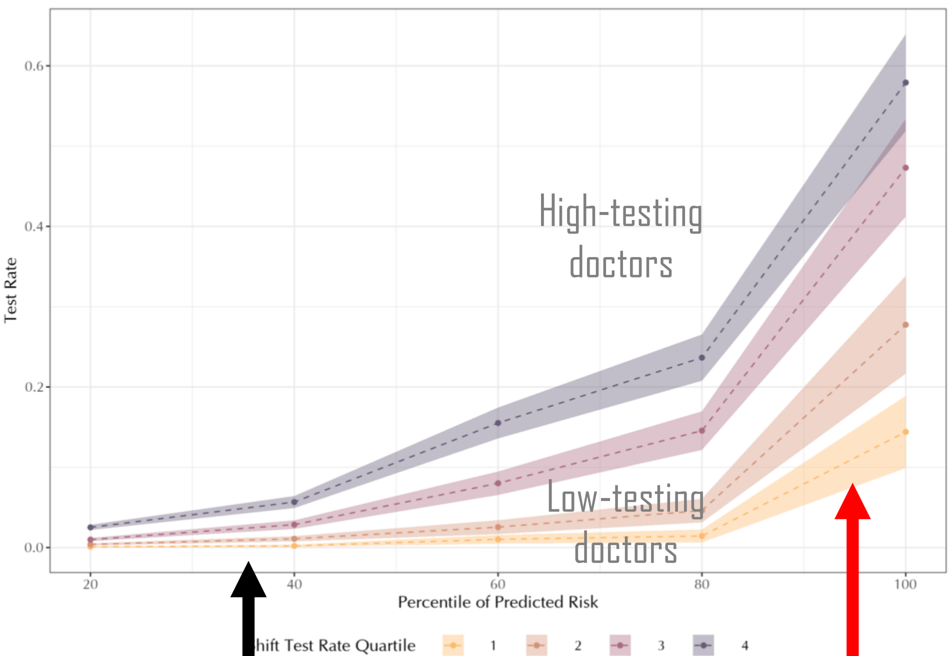
Where are physicians going wrong?

- Evidence of both over- and under-testing
  - ML would cut 62% of existing tests… and add 16% new
- We often look to incentives—but can't explain under-testing

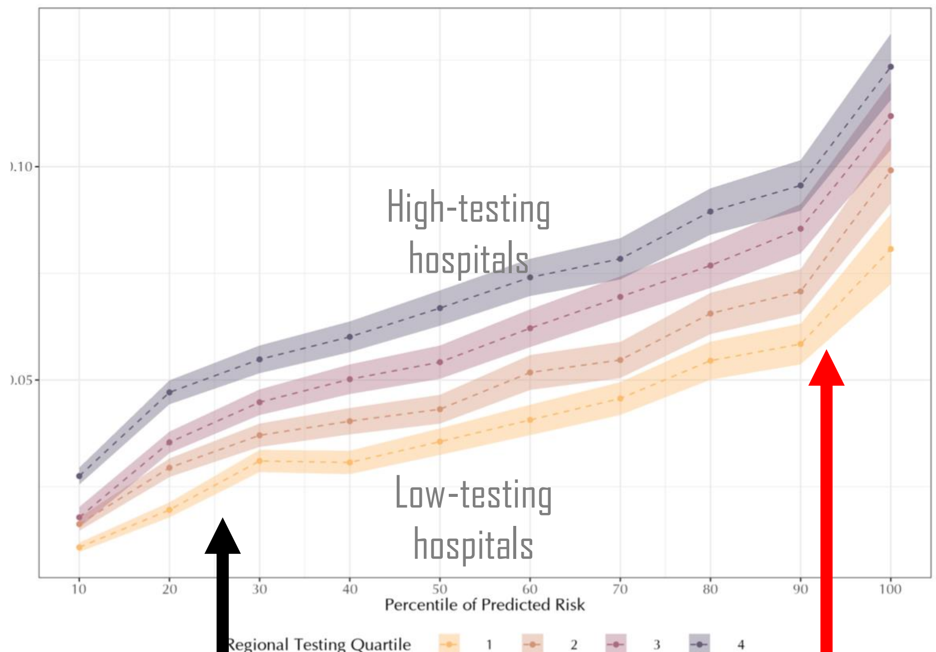
Mullainathan & Obermeyer, *QJE* 2022

Policy implication: Incentives can backfire

(a) Hospital Sample



(b) National Medicare Sample



- Low-testing **doctors** cut wasteful tests
  - And also valuable tests

- Low-testing **hospitals** cut wasteful tests
  - And also valuable tests

Mullainathan & Obermeyer, *QJE* 2022

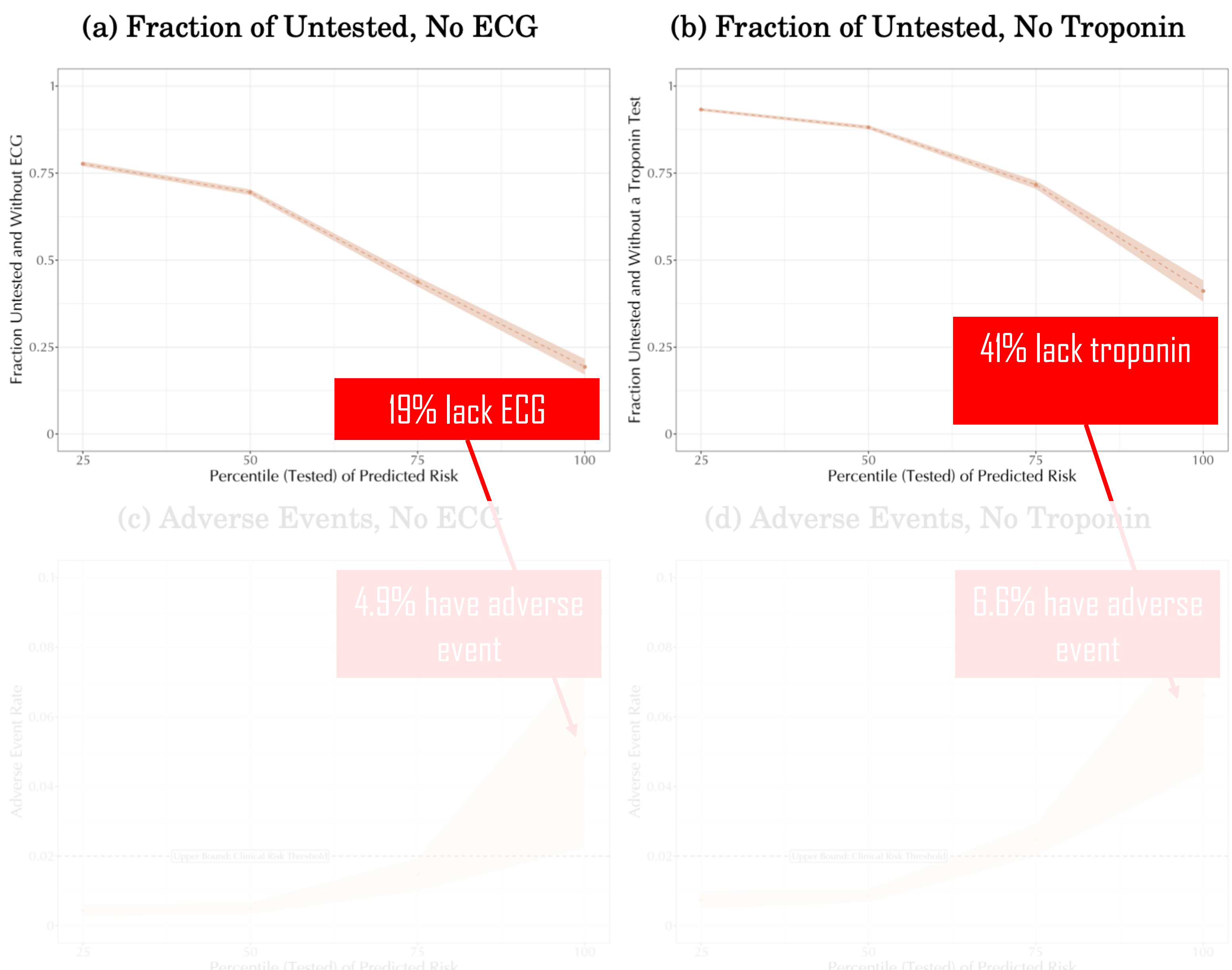
Some core econ points (that CS needs)

- Predictions fit into some cost-benefit framework
  - Not just some abstract loss measure
- Predictions get at marginal not average risk
  - No need to "choose wisely" about entire classes of tests
- Predictions validated with quasi-experiment
  - That acknowledge selective testing, treatment
- Predictions have policy implications
  - Incentives alone are insufficient

Some open questions

- How do predictions change doctor-planner dynamics?
  - See Agarwal, Gans, Goldfarb (2022)
  - Also: doctor-patient, patient-insurer, …
- What is optimal human–ML combination
  - …given that there must be Z's?

Untested, unsuspected patients: Short-term adverse events

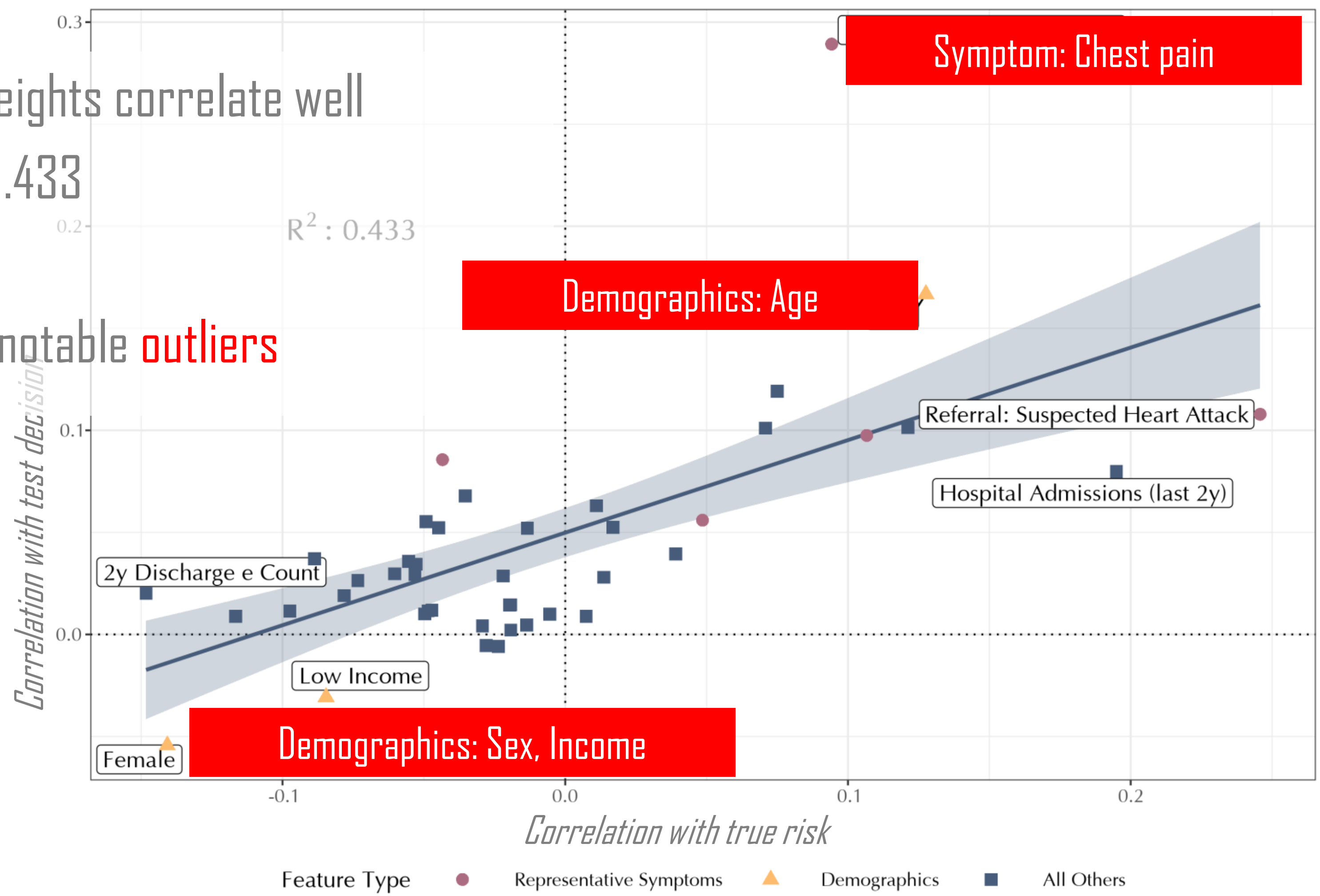


Some open questions

- How do predictions change doctor-planner dynamics?
  - See Agarwal, Gans, Goldfarb (2022)
  - Also changes doctor-patient games
- What is optimal human–ML combination
  - …given that there must be Z's?
- How does ML do better than doctors
  - …using data collected by doctors?

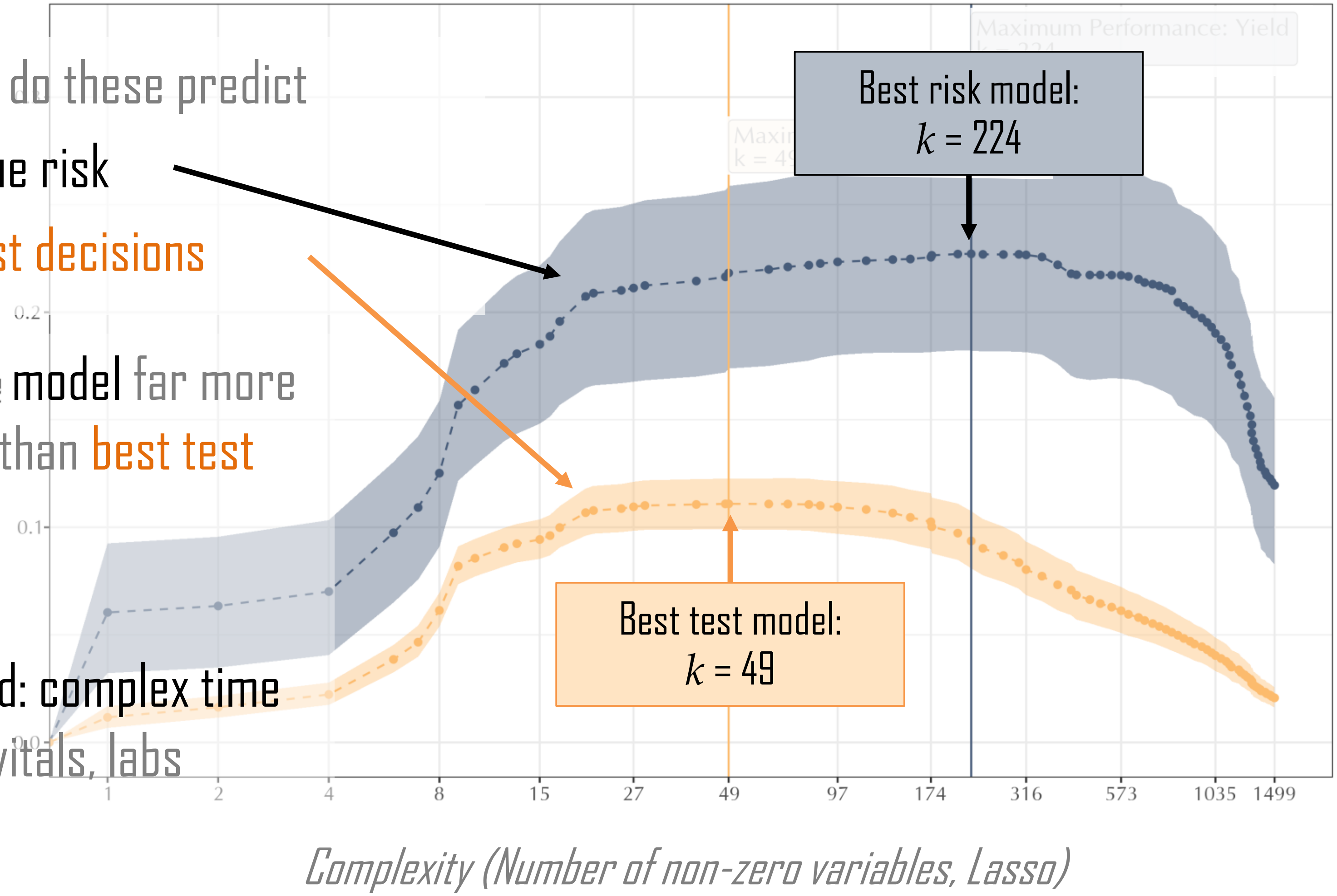
Physicians mis-weight individual variables

- Take **important variables** for ML model
  - Correlation with test decision vs. correlation with true risk
- Overall, weights correlate well
  - $R^2 = 0.433$
- But some notable outliers



Doctors are bounded

- Estimate best-fit risk models of varying complexity
  - Lasso complexity measure: number of non-zero variables
- How well do these predict
  1. True risk
  2. Test decisions
- Best risk model far more complex than best test model
- Neglected: complex time series—vitals, labs



Some interesting implications of this

- Humans seem to regularize (Camerer 2019)
  - Make (pretty) good use of a small set of variables

- A different conclusion from Dawes, Faust, & Meehl (1989)
  - Where people use too complex a model
  - And a statistical model does better by being simpler

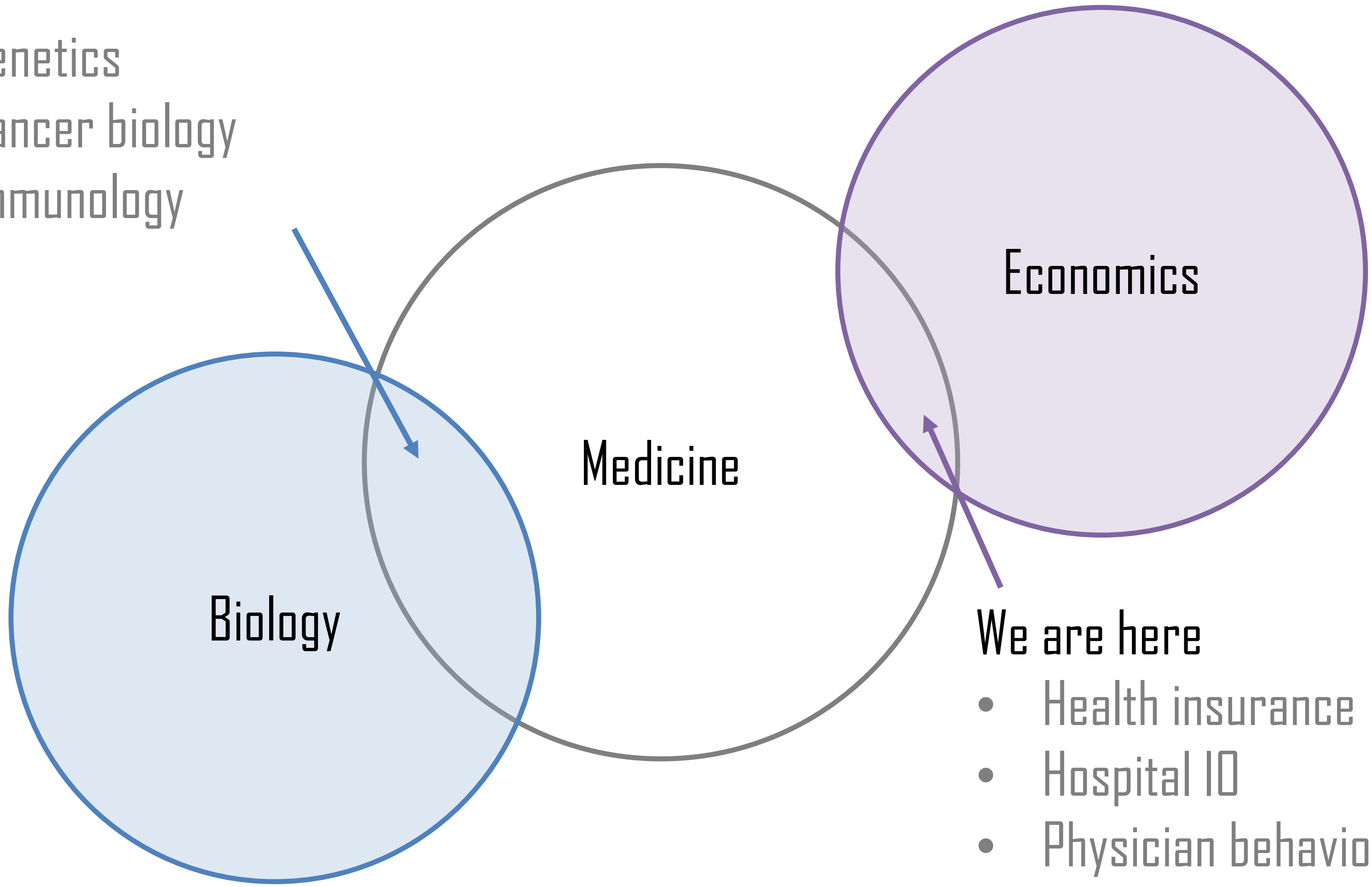
- Here we find physicians use too simple a model
  - A statistical model does better by being more complex
  - Maybe because phenomenon being modeled is complex
    - The 'illusion of sparsity' (Giannone et al. 2021)

Summary: ML, economics, and health (1/2)

- ML as an object of study for economists
  - Many of these tools go very wrong: racial bias, etc.
  - Applied micro toolkit sorely needed
- ML as a new tool to answer core health economics questions
  - Resource allocation, optimal policy
  - Frictions and administrative burden (Sahni et al. 2023)
  - Adverse selection, targeting, etc.
- ML as a source of huge economic value
  - Products: diagnostics, predictive trials, drug+device, …
  - Markets: drugs, consumers, hospitals, insurers, gov't, …

Medicine intersects with many other fields

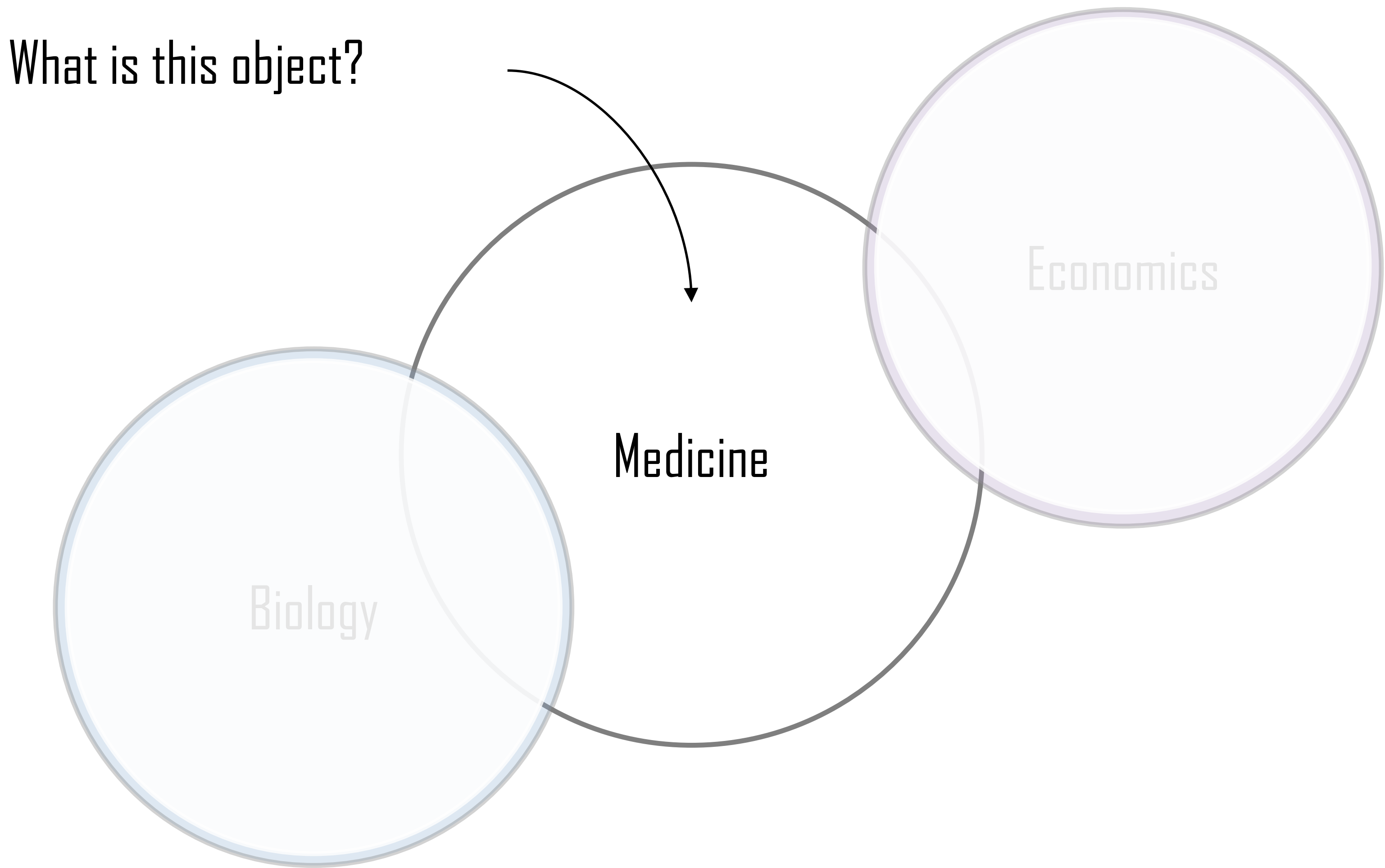
- Genetics
- Cancer biology
- Immunology
- …



We are here

- Health insurance
- Hospital IO
- Physician behavior
- …

Medicine intersects with many other fields



Medicine: A lot of white space

A domain with many facts…

- E.g., depression criteria: X

Little interest or pleasure in doing things
Feeling down, depressed, or hopeless
Trouble falling or staying asleep, or sleeping too much
Feeling tired or having little energy
Poor appetite or overeating
Feeling bad about yourself—or that you are a failure or have let yourself or your family down
Trouble concentrating on things, such as reading the newspaper or watching television
Moving or speaking so slowly that other people could have noticed; or the opposite—being so fidgety or restless that you have been moving around a lot more than usual
Thoughts that you would be better off dead or of hurting yourself in some way

…but very few theories

- E.g., beliefs: $\boldsymbol{\pi}$, effort: $\boldsymbol{\lambda}$, …

Depression for Economists
Jonathan de Quidt and Johannes Haushofer
NBER Working Paper No. 22973
December 2016
JEL No. D03,I1,I15,I3

ABSTRACT

Major depressive disorder (MDD) is one of the most prevalent mental illnesses worldwide. Existing evidence suggests that it has both economic causes and consequences, such as unemployment. However, depression has not received significant attention in the economics literature. In this paper, we present a simple model which predicts the core symptoms of depression from economic primitives, i.e. beliefs. Specifically, we show that when exogenous shocks cause an agent to have pessimistic beliefs about the returns to her effort, this agent will exhibit depressive symptoms such undereating or overeating, insomnia or hypersomnia, and a decrease in labor supply. When these effects are strong enough, they can generate a poverty trap. We present descriptive evidence that illustrates the predicted relationships.

- Why do we need theory?
  - Is treating X useful?
  - Counterfactuals

A medical mystery

- Every year in US alone 300-450,000 drop dead—no warning
- What makes this even more tragic
  - We have the cure
- We're just very bad at getting the cure into the right patients
  1. False negatives: Many deaths without ICD
  2. False positives: 30-40% of ICDs never fire

Useful to predict who will need this

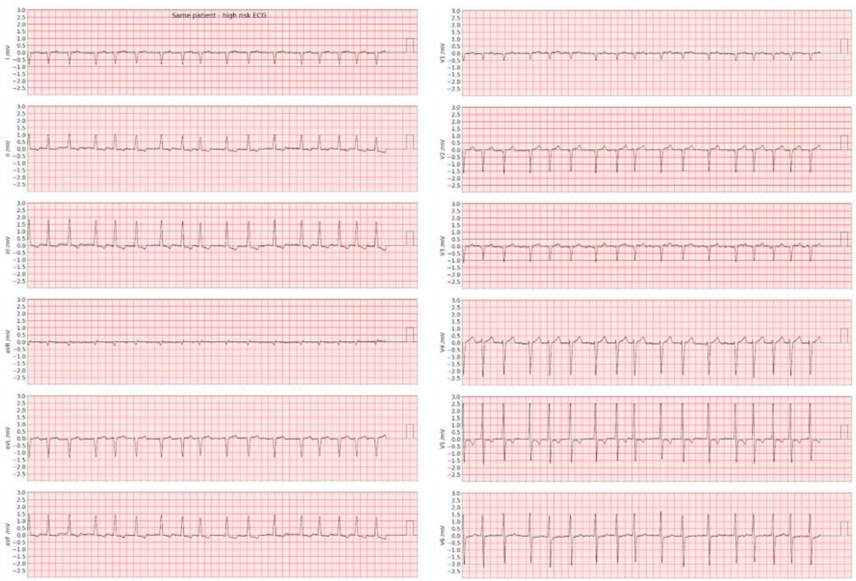
What we do

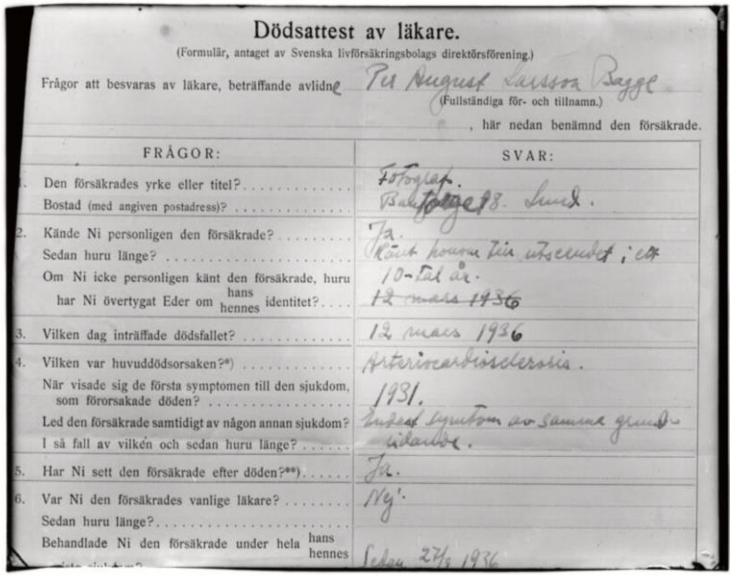
Input: ECG waveform

- All 401,765 ECGs (2014-18)
- From 119,724 patients

Output: Death certificate

- 100% linkage to SCD label
- Full EHR data





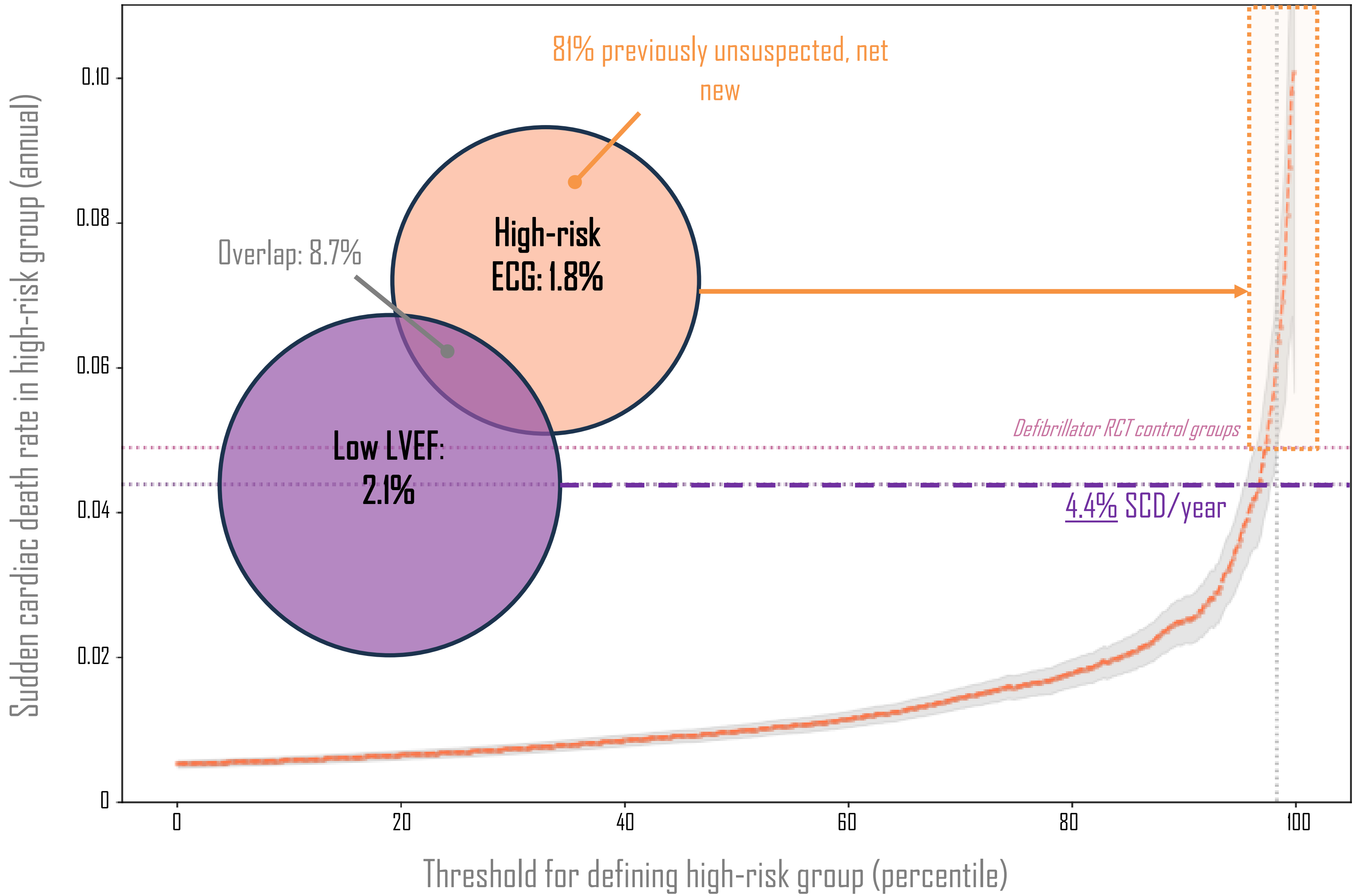
Dödsattest av läkare.

(Formulär, antaget av Svenska livförsäkringsbolags direktörsförening.)

Frågor att besvaras av läkare, beträffande avlidne Per August Larsson Bagge
(Fullständiga för- och tillnamn.)
, här nedan benämnd den försäkrade.

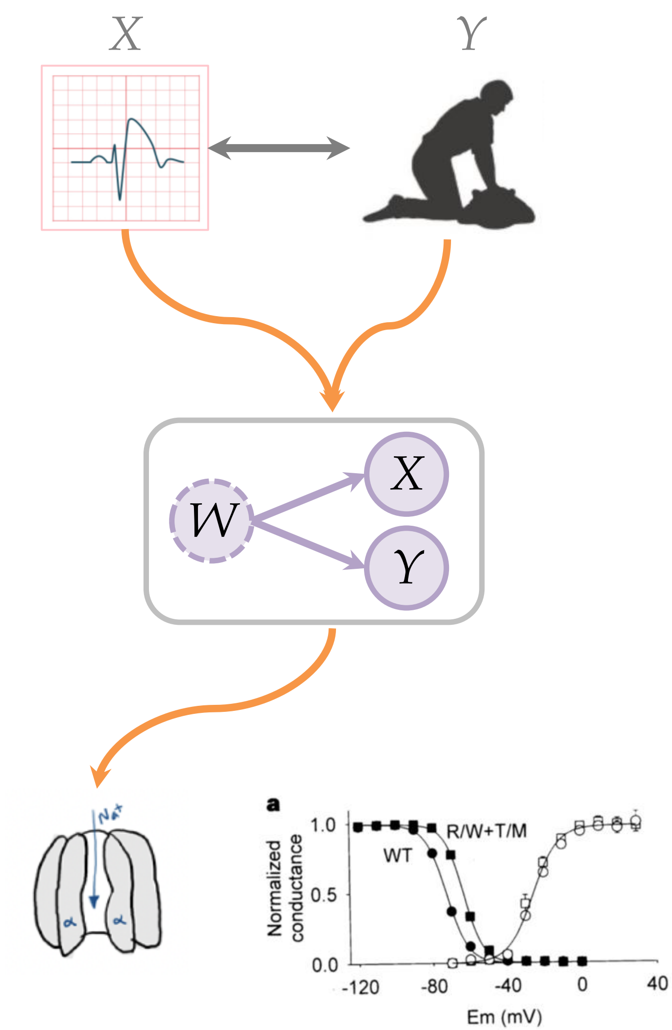
FRÅGOR:	SVAR:
1. Den försäkrades yrke eller titel?	Fotograf.
Bostad (med angiven postadress)?	
2. Kände Ni personligen den försäkrade?	Ja.
Sedan huru länge?	
Om Ni icke personligen känt den försäkrade, huru har Ni övertygat Eder om hans / hennes identitet?	12 mars 1936
3. Vilken dag inträffade dödsfallet?	12 mars 1936
4. Vilken var huvuddödsorsaken?*)	
När visade sig de första symptomen till den sjukdom, som förorsakade döden?	1931.
Led den försäkrade samtidigt av någon annan sjukdom?	
I så fall av vilken och sedan huru länge?	
5. Har Ni sett den försäkrade efter döden?**)	Ja.
6. Var Ni den försäkrades vanlige läkare?	
Sedan huru länge?	
Behandlade Ni den försäkrade under hela hans / hennes	

Sudden cardiac death rate vs. ECG-predicted risk



Such facts are fundamental to human discovery process

1. Notice curious **fact**
   - Correlation: $X \leftrightarrow Y$
   - Not hypothesis driven
2. **Reason** about cause
   - What could produce both X, Y
3. **Test** hypotheses
   - Collect new data, with counterfactuals

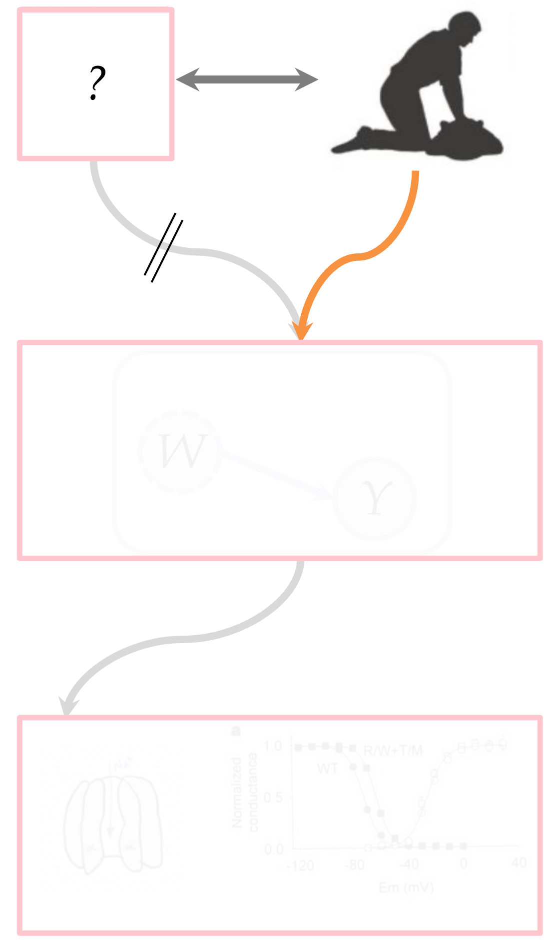


This pathway has dried up

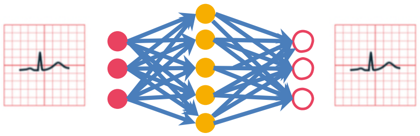
- Why? Low-hanging fruit is picked
  - And today's doctors don't have much time for curiosity
- Today: All in on bench to bedside
  - Model disease biology in the lab
  - Translate understanding into diagnostics, drugs
  - Hugely successful for some problems
    - Targeted cancer therapies, mRNA, CRISPR, …
  - Less so for complex, poorly understood problems
- Can ML reboot the "bedside to bench" pathway?

Key problem: ML for science

1. Very robust correlation
   - But **no curious** X
2. Can't reason about cause
   - **No bridge**: from Y to patient physiology via X
3. **No hypotheses** to test

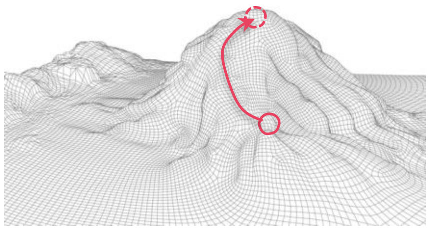


A way to visualize what the model is 'seeing'



- Train a **generative** model
  - Encode patients' ECGs

- Use predictive model to calculate **risk gradient** around ECG_i



- "**Morph**" ECG_i along risk gradient
  - Generate **counterfactual ECG**
  - …Repeat

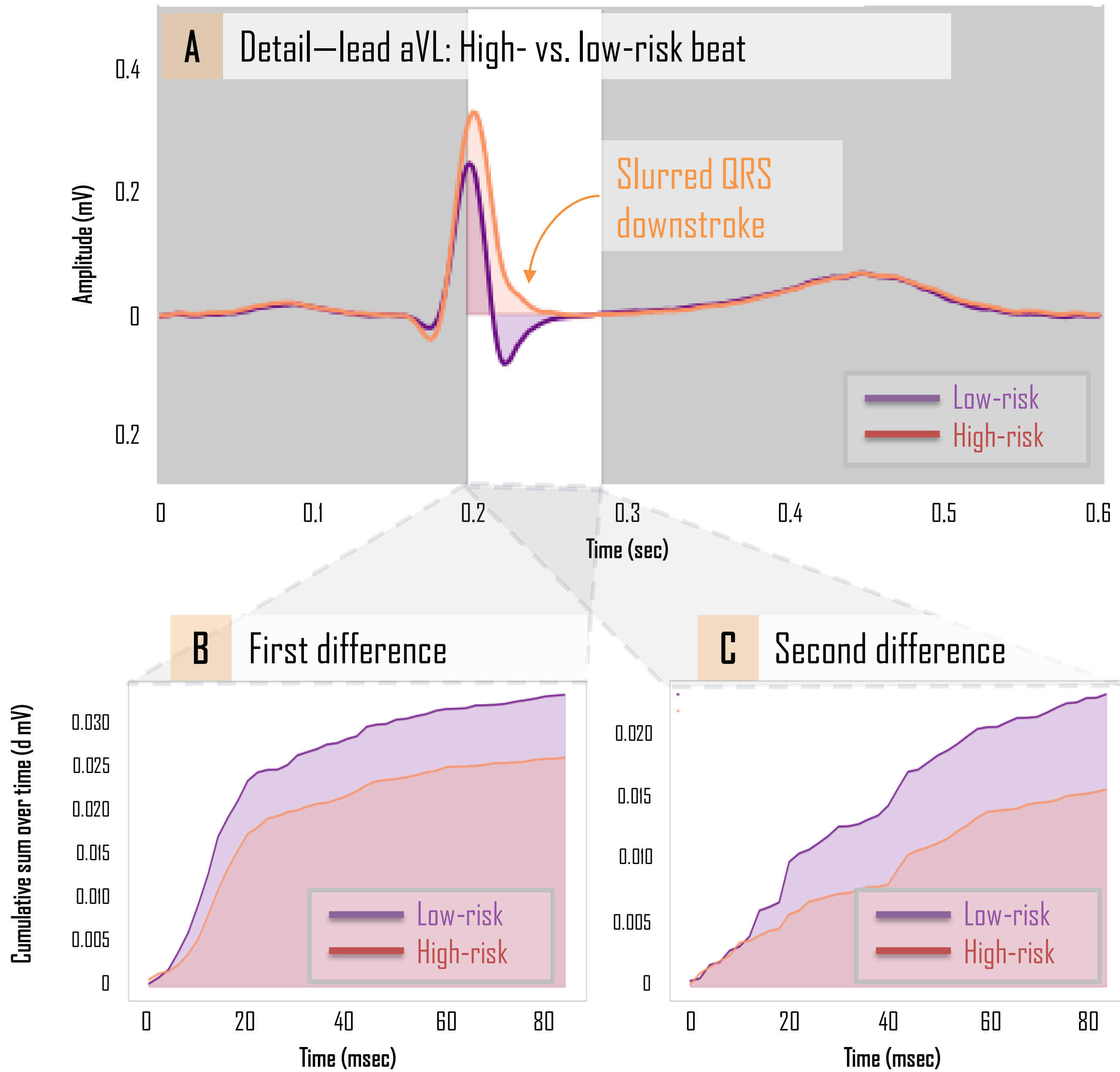
Result: A representative morph

- This allows 2 things to happen

1. Focus on **one observation**: reduces dimensionality

2. Get model ‘discovery’ into biological **space accessible to human theory**: ECGs and hearts

An intriguing feature of high-risk morphs

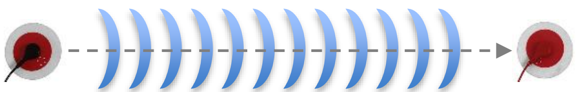


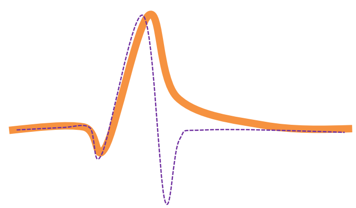
- **Qualitative** insight: signal 'peters out'
  - Easy to see
  - (...now)
- **Quantitative** features: 1st and 2nd diffs
- New features **predict sudden death, VF/VT**
  - In Sweden, Taiwan, California

One hypothesis to link X, Y

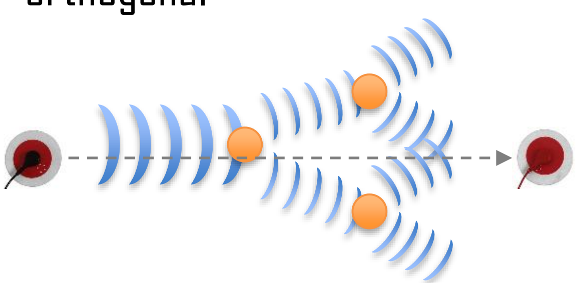
1. Hypothesis generation

Low risk: wavefront and recording vectors match



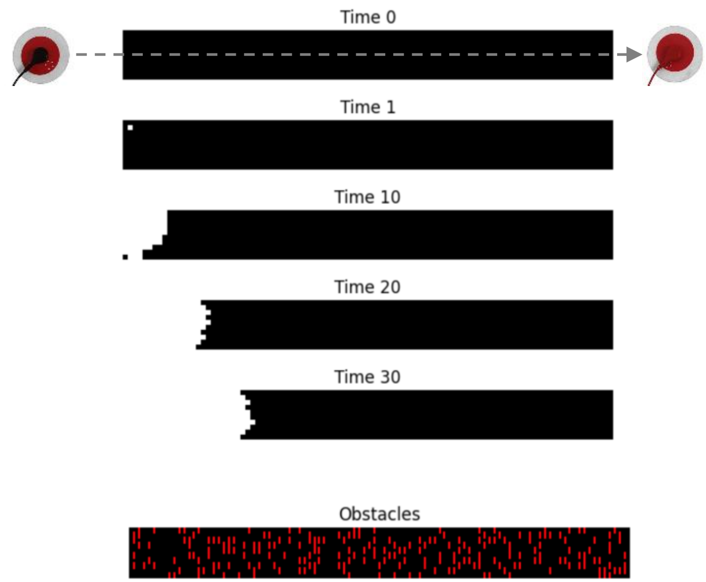


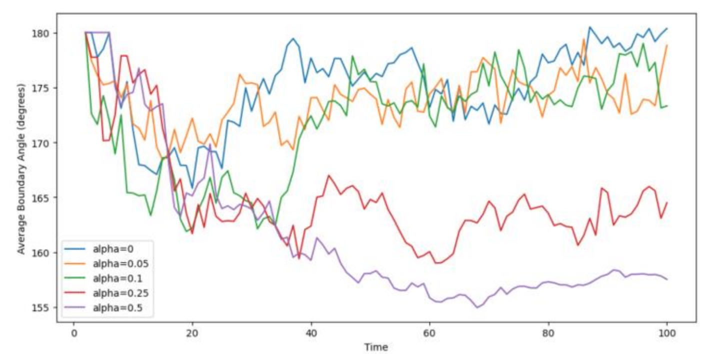
High risk: wave vector gets more orthogonal



What could do this? scatter

2. Simple simulation





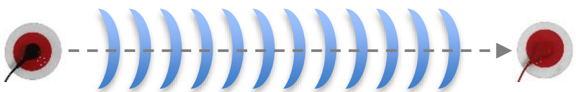
More obstacles

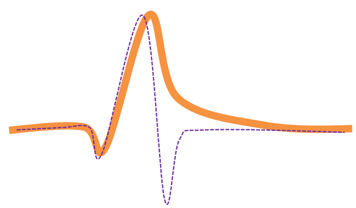
More orthogonal

One hypothesis to link X, Y

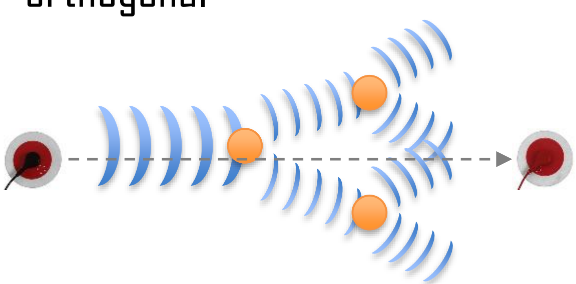
1. Hypothesis generation

Low risk: wavefront and recording vectors match

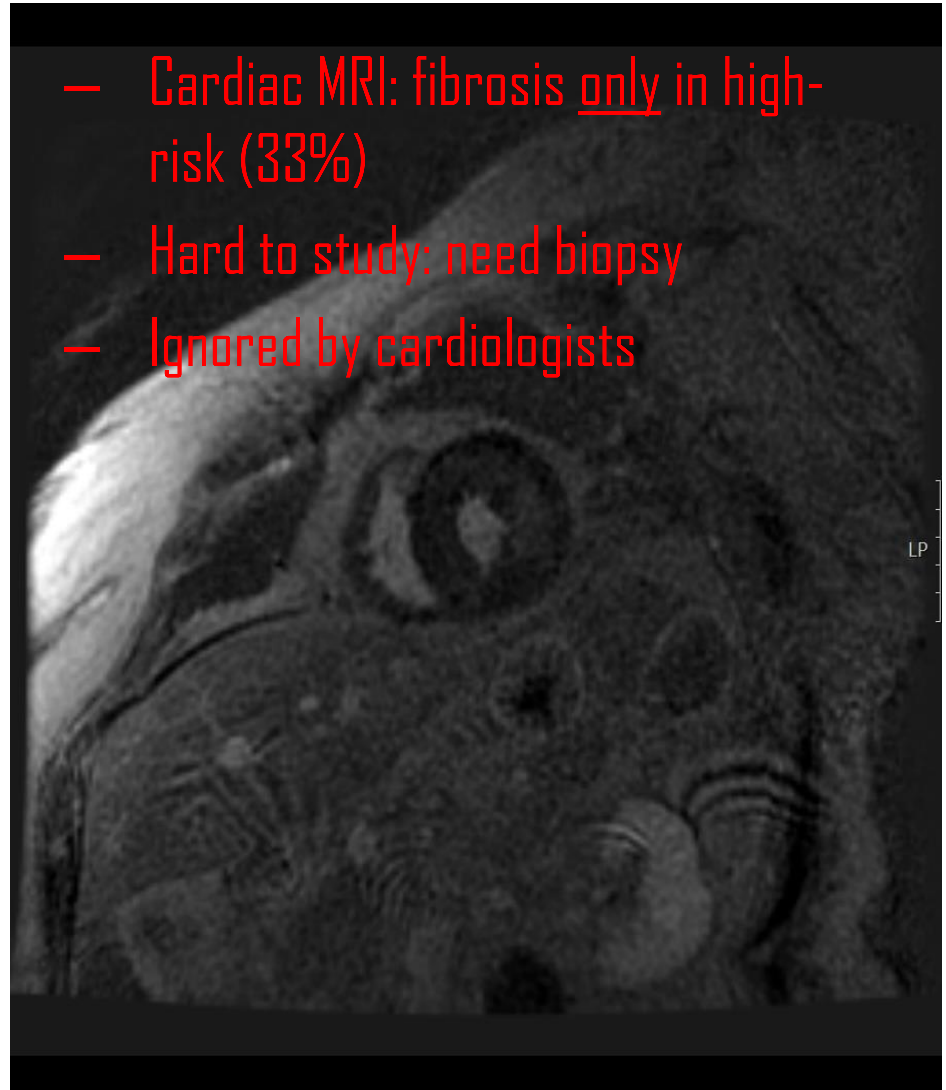




High risk: wave vector gets more orthogonal



What could do this? scatter

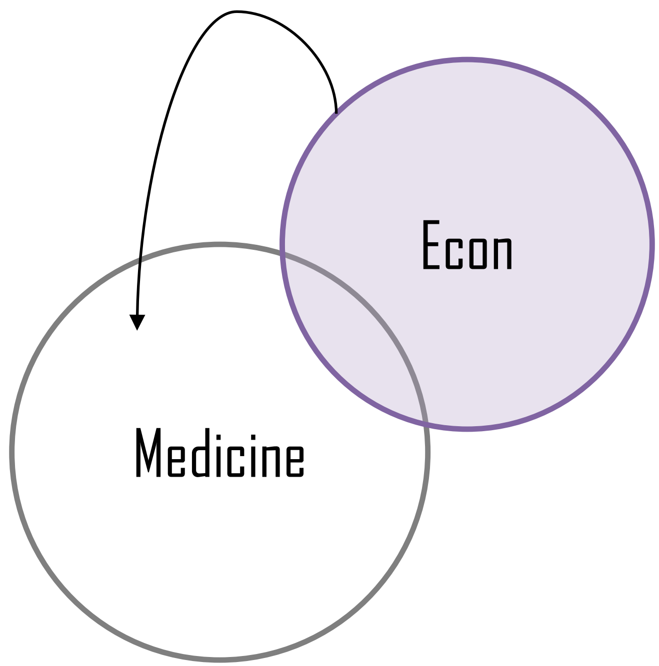


Summary: ML, economics, and health (2/2)

- ML is an engine for generating new facts about the world
  - Finds signal in rich medical data that humans miss
- This makes ML a powerful new tool for scientific discovery
  - Discoveries often start with surprising facts
- Tying facts into theory: open problem
  - Many things we care about are not in the dataset
    - E.g., shocks for cardiac arrest
  - Need theories for new treatments, new data collection
  - But not something ML can learn

Summary: ML, economics, and health (2/2)

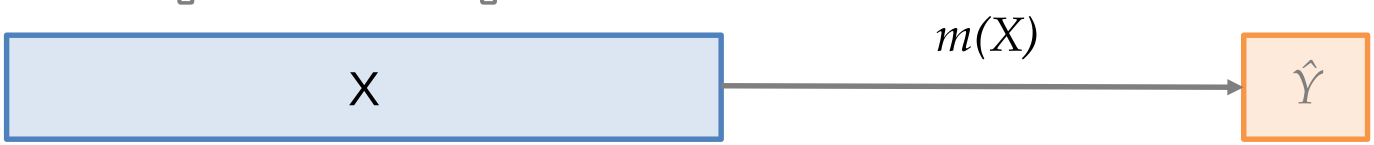
- Why now?
  - Core medical data now accessible
  - This has been a huge gap to date
- Why you?
  - Economists are A+ at abstraction
  - Investments in learning some medicine will pay off
  - Reminiscent of early behavioral economics



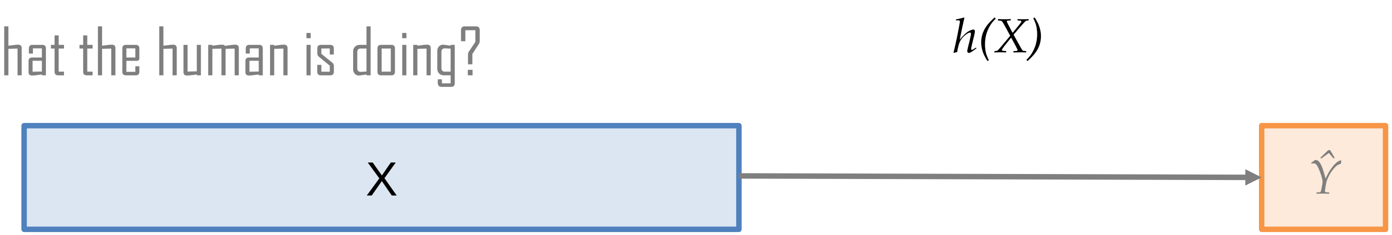
Machine learning solves this kind of prediction problem

- Form **explicit predictions** on heart attack (blockage) risk
  - In tested ER patients: predict test outcome Y with X
  - Find <u>potential</u> errors: patients with mismatched $\hat{Y}$ vs. T

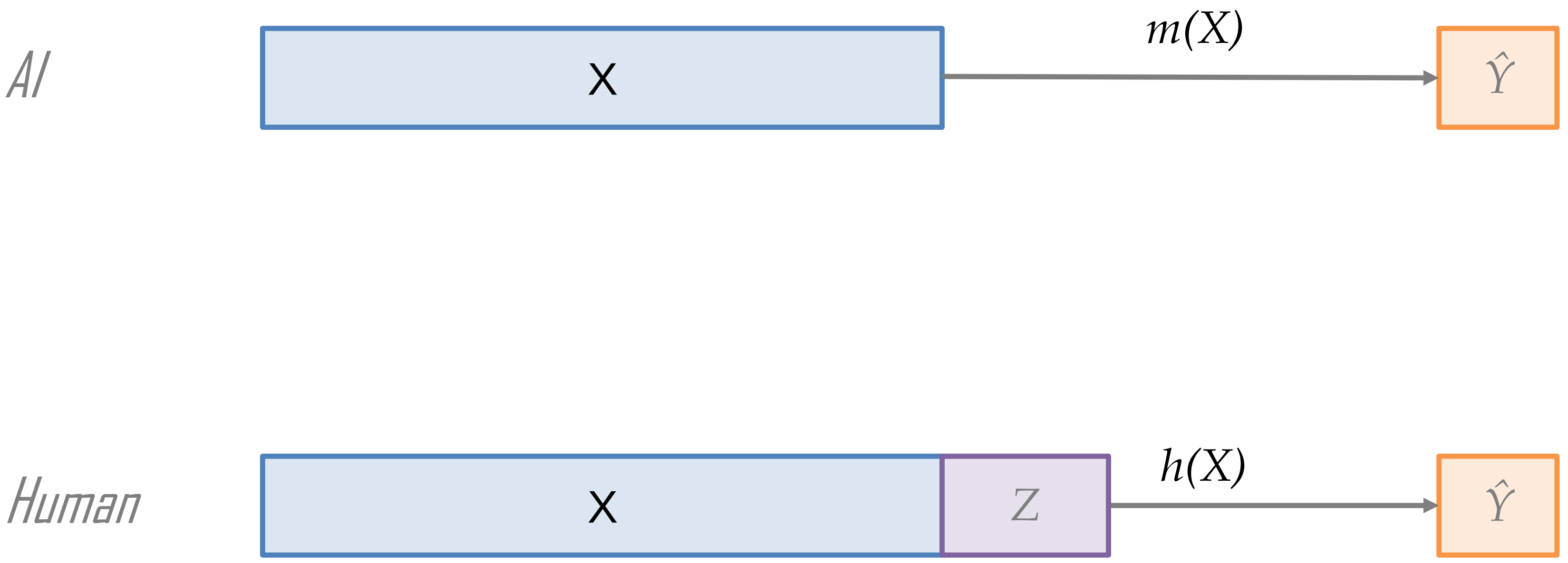
- What the algorithm is doing



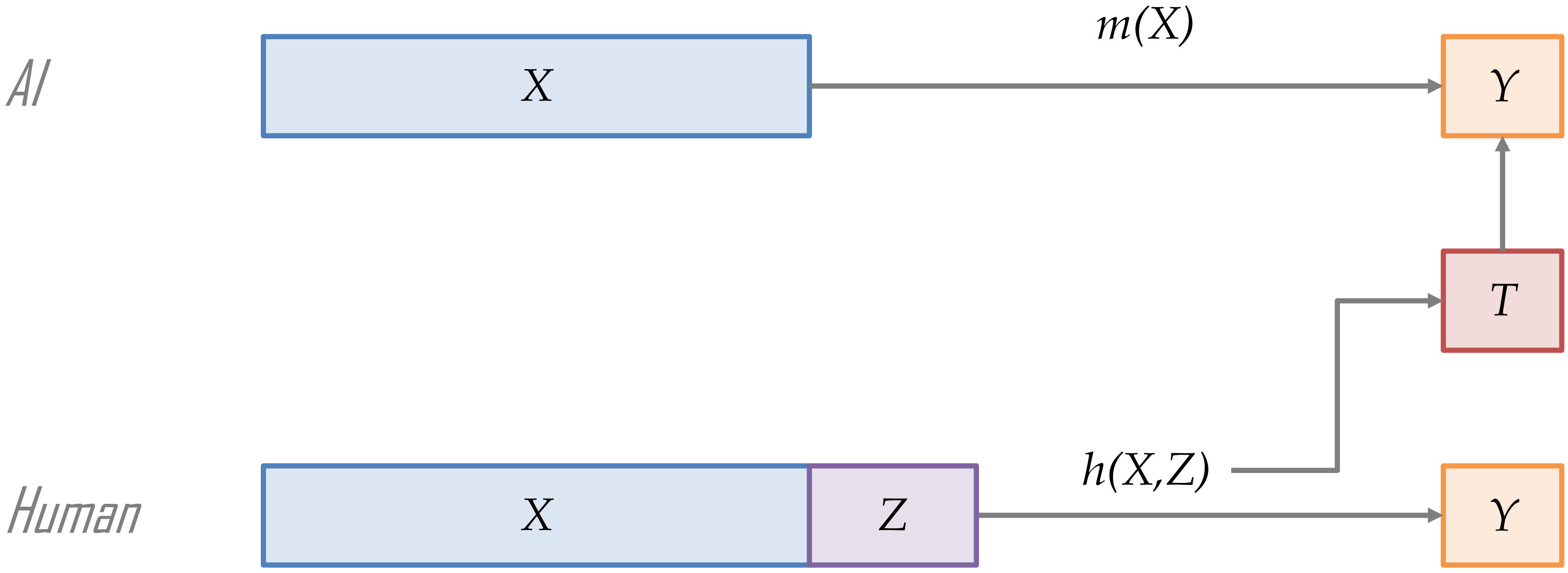
- What the human is doing?



Mullainathan & Obermeyer, *QJE* 2022



AI
X
$m(X)$
$\hat{Y}$
Human
X
Z
$h(X)$
$\hat{Y}$



AI
X
m(X)
Y
T
Human
X
Z
h(X,Z)
Y