The reinvention of medicine with AI

...and what that means for the life sciences

Ziad Obermeyer
UC Berkeley, NBER, Dandelion Health

Nick Gossen
Dandelion Health

Also implicated:
Esther Duflo (MIT), Markus Lingman (Region Halland),
Sendhil Mullainathan (MIT), Alex Schubert (UC Berkeley), Jenny Wang (MIT)

Disclosures

Ziad Obermeyer

Chief Science Officer -- Dandelion Health

Nick Gossen

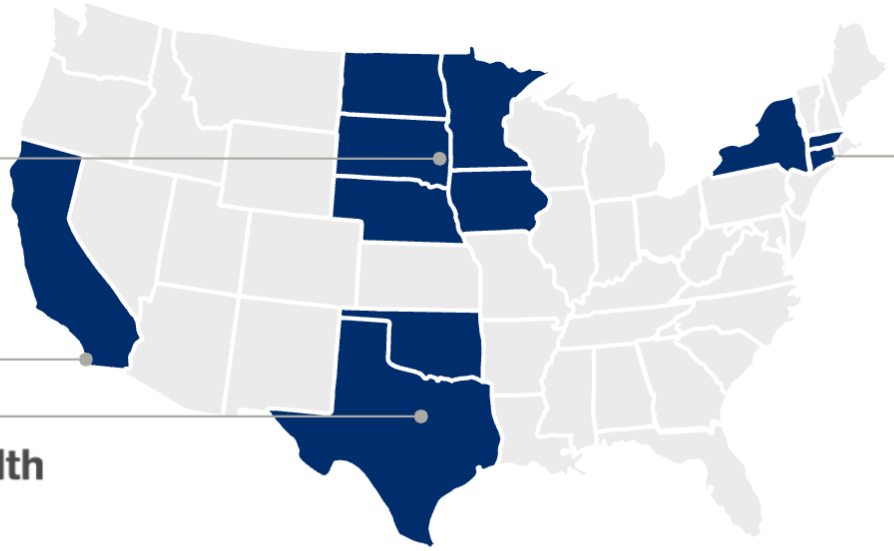
Chief Strategy Officer -- Dandelion Health

Dandelion Health

SANFORD HEALTH
Midwest / Rural

SHARP
West / Urban

Texas Health Resources®
Southwest / Urban



Montefiore
North East / Urban

73
hospitals

15M
patients

Primary, secondary & tertiary care

Diversity needed for regulatory compliance

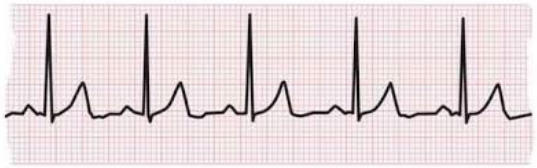
Demographic	Dandelion	U.S. Avg
White	**54%**	58%
Black	**12%**	14%
Hispanic	**25%**	20%
AANHPI	**5%**	6%
Native American	**3%**	3%
Rural	**10%**	20%

Dandelion

Images of doctors at work

Images, waveforms, etc.: A blind spot for research

- They are **everywhere** in the hospital
- But they are **hard to fit** into this:



?

$$y = X\beta + \epsilon$$

- So we **ignore them** and focus on data that fits
  - Claims
  - Structured health records and outcomes data
  - Molecular biomarkers, genes
  - …

Today's agenda

- ML allows direct engagement with high-dimensional data
  - This will transform medical practice
  - And reinvent the science underlying medicine

- Why should you care? ML will also change
  1. How we understand and study medicine
     - The health system, life sciences
     - Biomarker discovery in medicine
  2. How we design clinical trials

Three examples

- **ECG to predict sudden cardiac death**
- ECG to measure hERG channel inhibition
- CT to measure liver steatosis

Suboptimal allocation of medical technology—two views

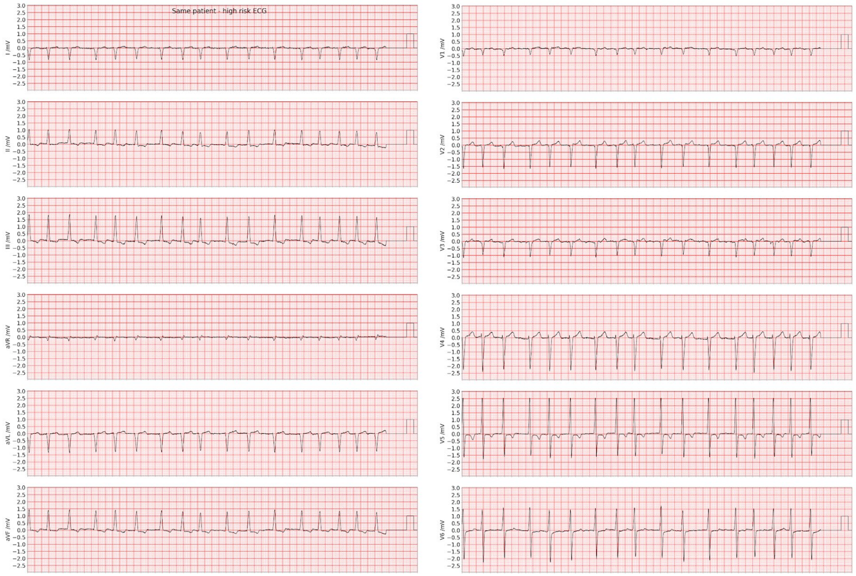
- Defibrillators ($25-30k)
  - Great for some patients
- Over-use: 2/3 never fire
  - No arrhythmia develops
  - Misaligned incentives?
- Under-use: 300-450,000/yr drop dead without warning in US alone
  - Can incentives explain this?
- Over- vs. under-: beside the point

Prediction problem:
Who will need this?

What we do

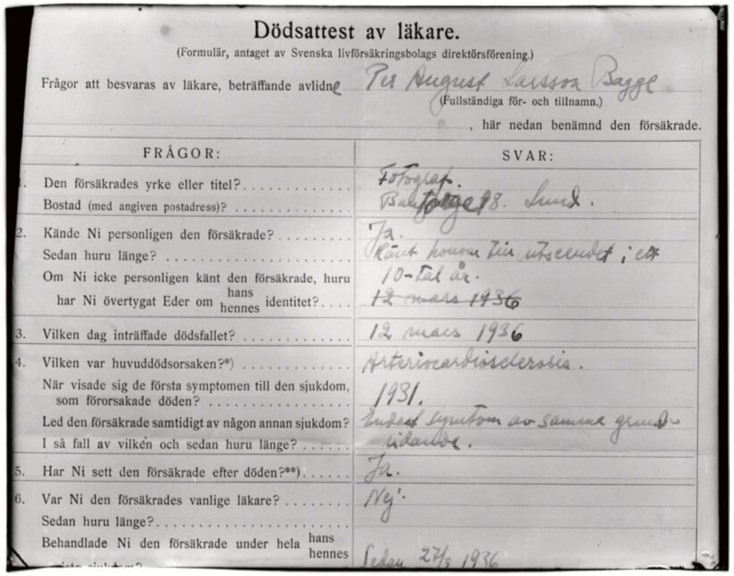
Input: ECG waveform

- All 441,614 ECGs in Region Halland (2010-16)
  - Results: 30% hold-out set



Output: Death certificate

- 100% linkage for standard epidemiological definition
- Also: Full EHR data



Dödsattest av läkare.

(Formulär, antaget av Svenska livförsäkringsbolags direktörsförening.)

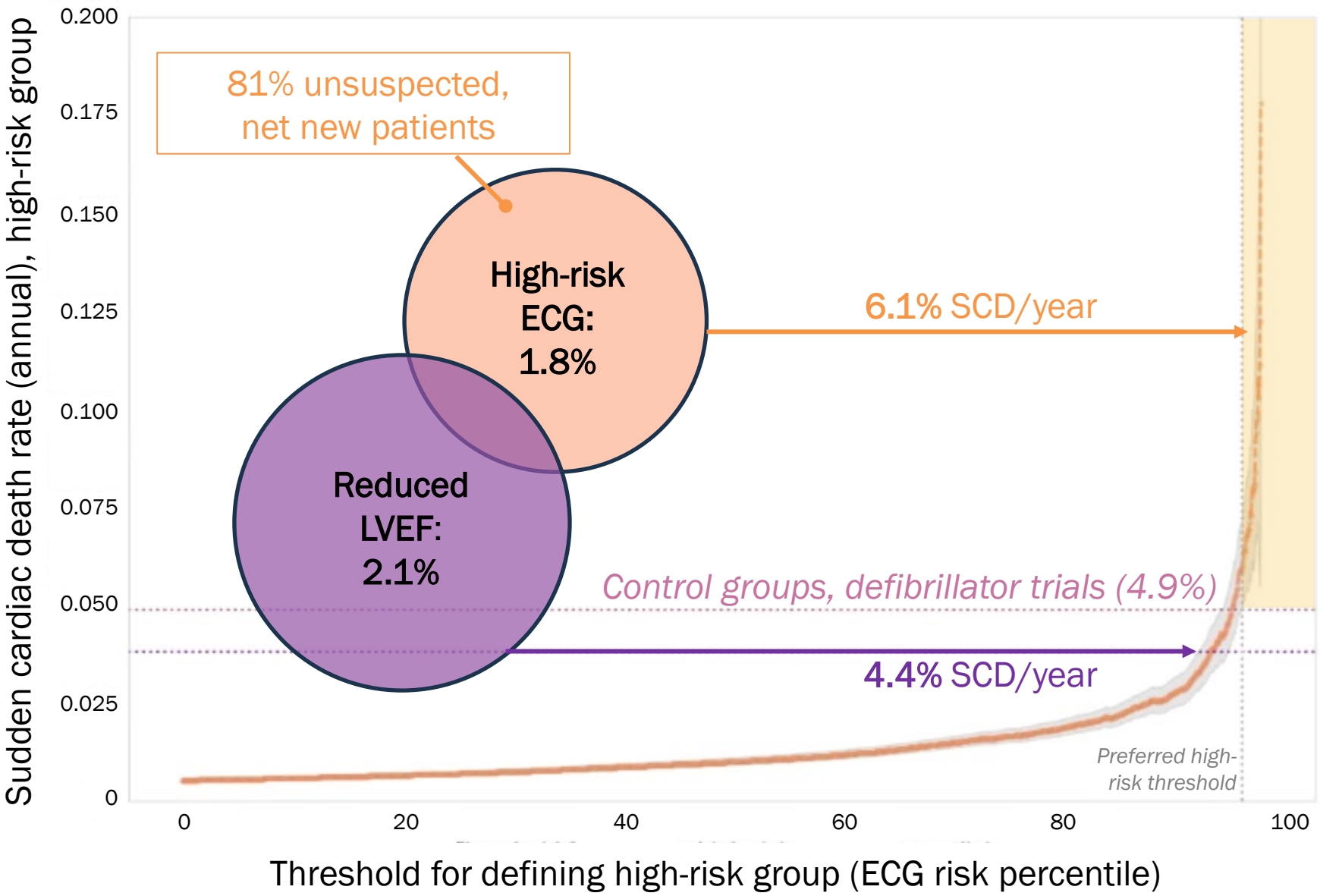
Frågor att besvaras av läkare, beträffande avlidne Per August Larsson Bagge

(Fullständiga för- och tillnamn.)

, här nedan benämnd den försäkrade.

FRÅGOR:	SVAR:
1. Den försäkrades yrke eller titel?	Fotograf.
Bostad (med angiven postadress)?	Bantorget 8. Lund.
2. Kände Ni personligen den försäkrade?	Ja.
Sedan huru länge?	Känt honom till utseendet; ett 10-tal år.
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?*)	Arteriocardiosclerosis.
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?	Endast symptom av samma grundsjukdom.
5. Har Ni sett den försäkrade efter döden?**)	Ja.
6. Var Ni den försäkrades vanlige läkare?	Nej.
Sedan huru länge?	
Behandlade Ni den försäkrade under hela hans / hennes	Sedan 27/2 1936

Sudden cardiac death rate vs. ECG-predicted risk



No definition of 'sudden cardiac death' is perfect

- Death certificates: a pretty good measure
  - But this can include non-arrhythmic sudden deaths
  - A problem if we want to predict preventable deaths

- Three empirical tests, out of sample: Can our model
  1. Predict arrythmias that cause sudden death?
     - California: High-risk group has 19% VT/VF per year
  2. Identify adjudicated cause of cardiac arrest in registries?
     - Taiwan: Predictions specific to arrhythmic causes
  3. Find people who seem to benefit from defibrillators?
     - Sweden: High-risk + ICD die 53.7% less vs. predicted

What we're doing next

- Multi-site prospective study (Sweden, US, Taiwan)
- Run algorithm on all stored ECGs in large health systems
  - Identify high-risk patients
  - Randomly sample (along with some low-risk controls)
- Wearable ECG for 2-4 weeks
  - Look for incident malignant arrhythmias (VF/VT)
  - Flag positives for follow up

What can you do with this (1/3)? Allocate defibrillators better

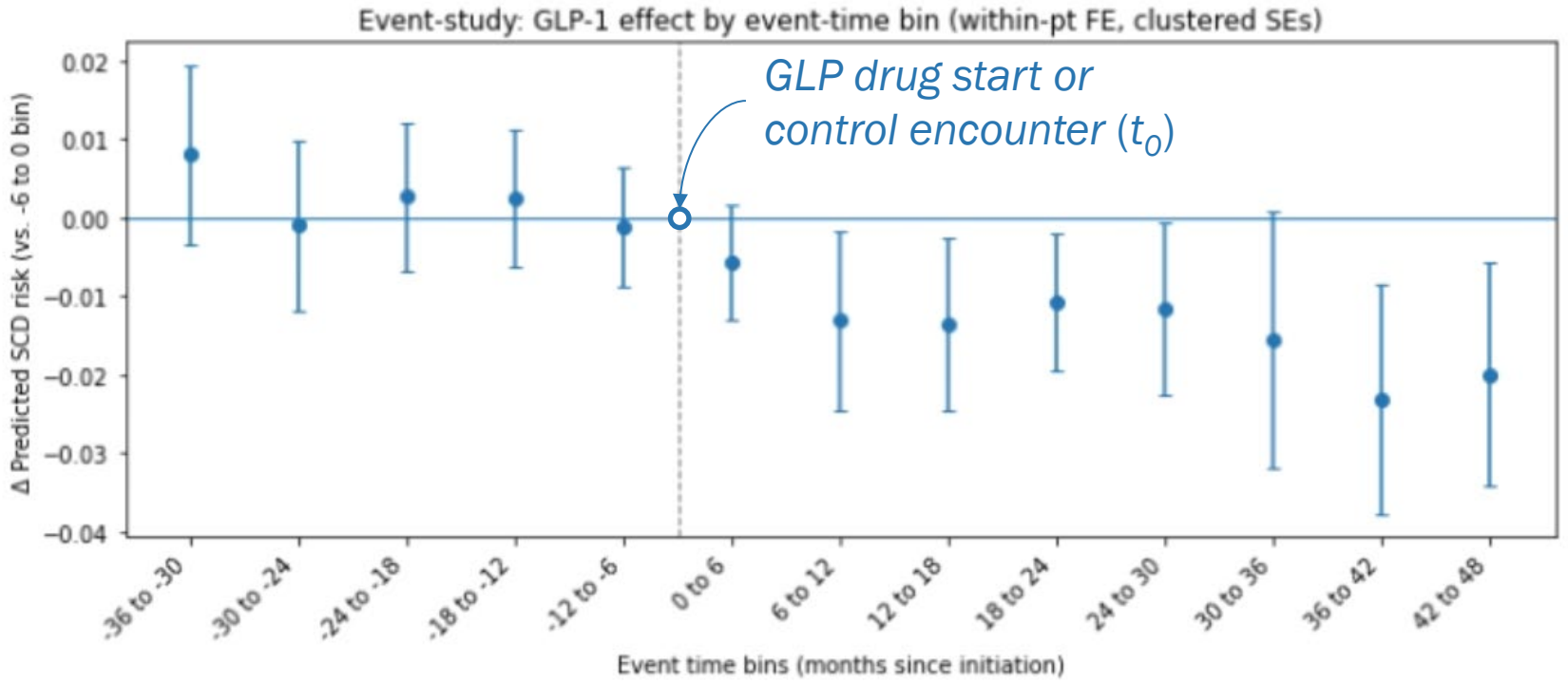
	(1)	(2)	(3)	(4)	(5)	(6)
	Sudden cardiac death			All-cause mortality		
	ECG model	*LVEF*	*Both*	*ECG model*	*LVEF*	*Both*
Risk indicator variables						
ECG high-risk	0.0541***		0.0492***	0.2367***		0.2261***
	(0.002)		(0.002)	(0.006)		(0.006)
LVEF high-risk		0.0375***	0.0312***		0.0951***	0.0664***
		(0.002)	(0.002)		(0.005)	(0.005)
Defibrillator present	0.0054***	0.0030*	0.0027*	-0.0277***	-0.0361***	-0.0386***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.004)
Defibrillator x high-risk interactions						
Defibrillator x ECG high-risk	-0.0343***		-0.0323***	-0.1308***		-0.1354***
	(0.004)		(0.004)	(0.011)		(0.011)
Defibrillator x LVEF high-risk		-0.0186***	-0.0139***		-0.0251**	-0.0044
		(0.003)	(0.003)		(0.009)	(0.009)
Baseline outcome rate	0.0044***	0.0045***	0.0038***	0.0503***	0.0525***	0.0600***
	(0.0003)	(0.0003)	(0.0003)	(0.0007)	(0.0007)	(0.0021)

High-risk + defibrillator:
54% lower SCD rate

n=94,736 for all regressions. All models control for age and sex. Standard errors clustered by patient.

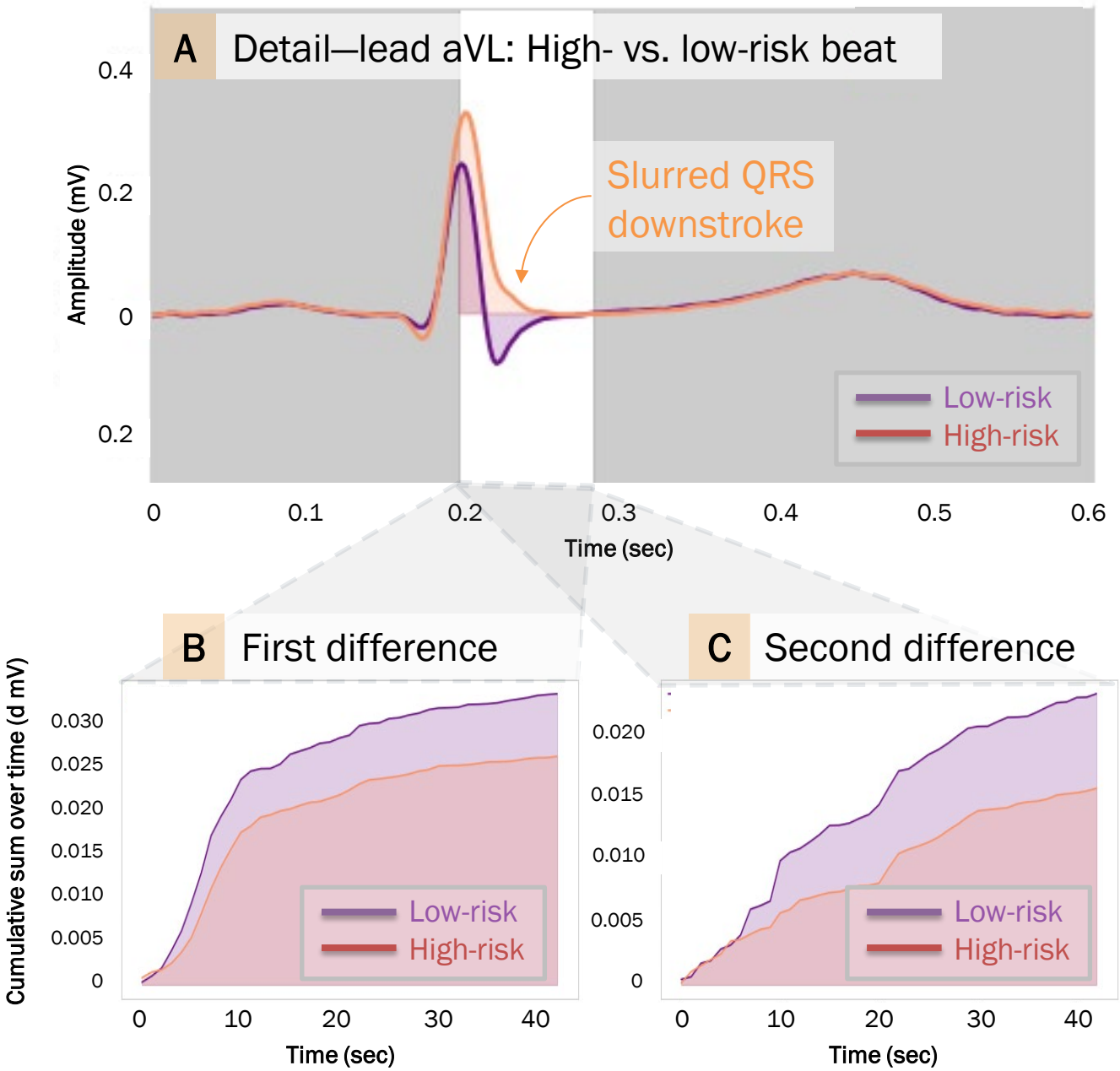
What can you do with this (2/3)? Discover more drugs, earlier

- Predictive model turns ECGs into time series of risk
  - Frequently measured, continuous, responsive



- Correlation of $\left(\Delta\widehat{SCD}, \Delta Weight\right) = -0.055$
  - Reductions in sudden death via some other mechanism

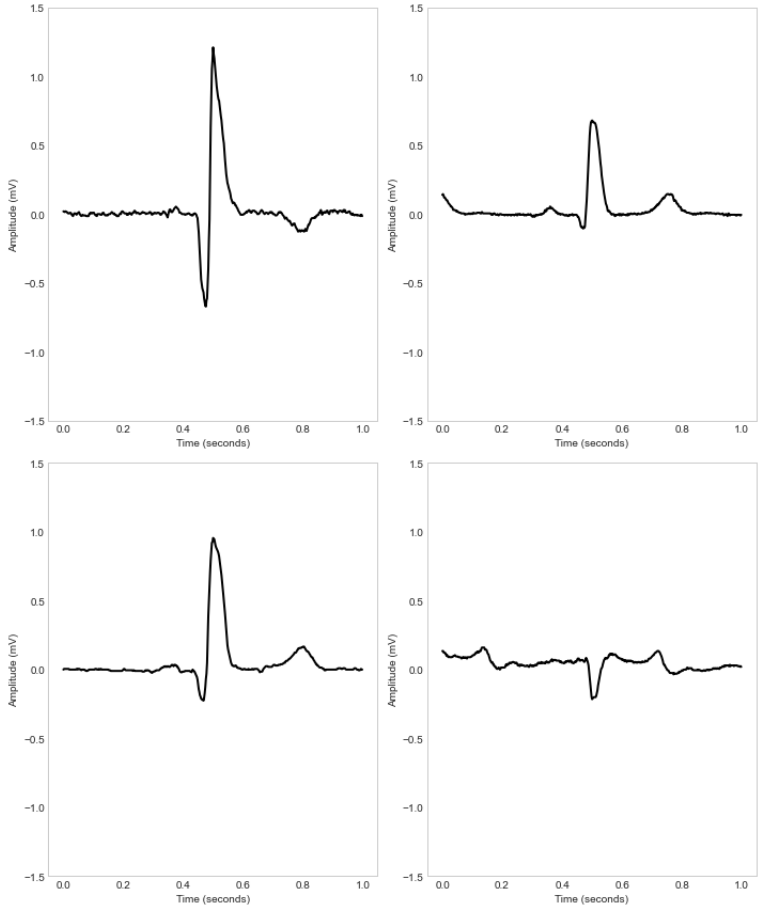
What can you do with this (3/3)? **Generate new hypotheses**



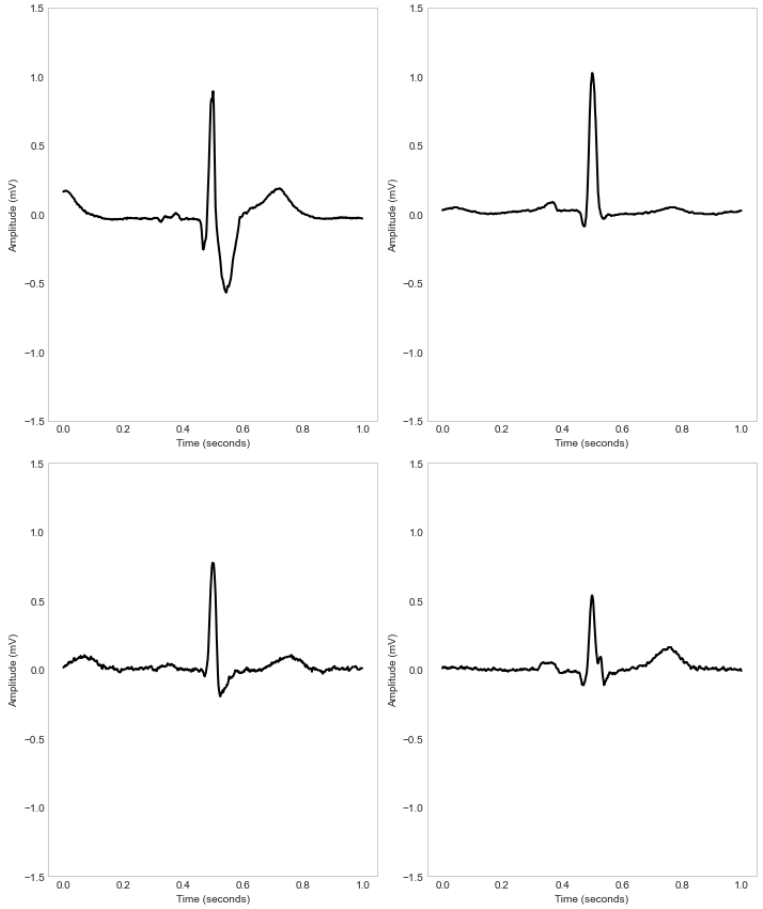
- Qualitative insight: signal ‘peters out’
- Quantitative features: 1st and 2nd diffs
- New features predict sudden death, VF/VT
  - In Sweden, Taiwan, California

This is pretty obvious (ex post)—but not previously described

A Low 1st and 2nd difference, aVL

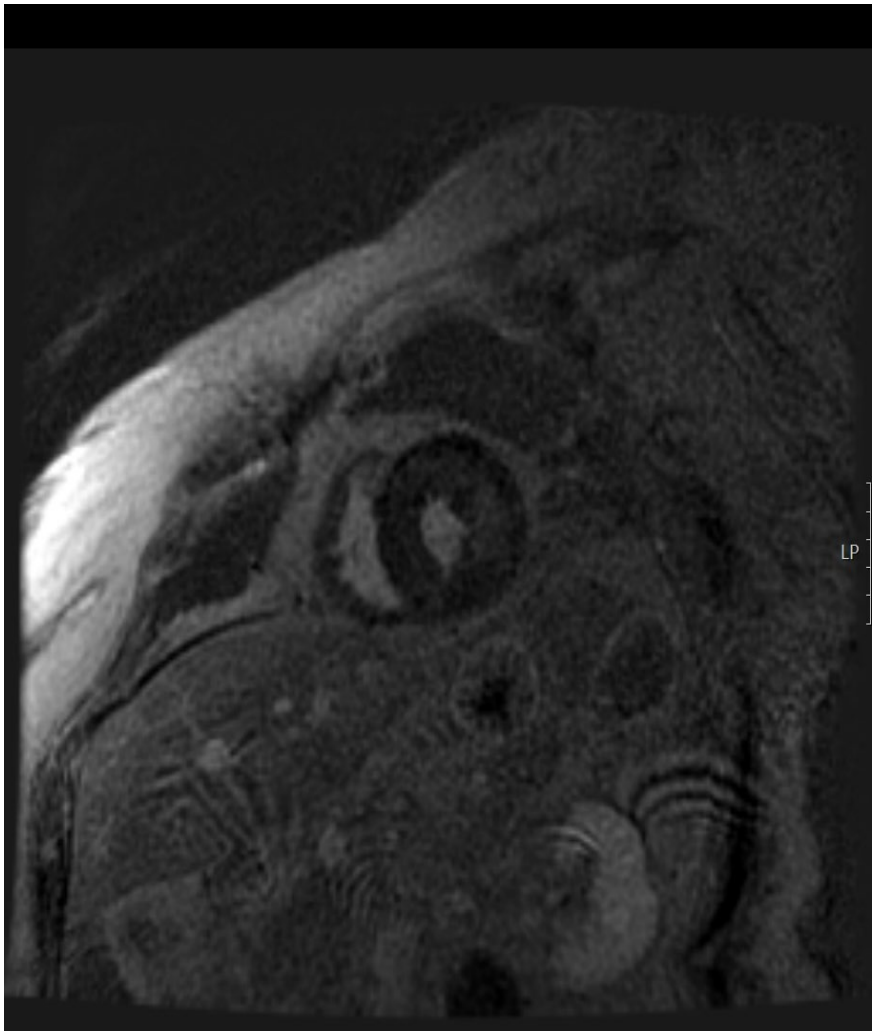


B Age- and sex-matched controls



What could produce both ECG **waveform** and **sudden death**?

Hypothesis generation



Cardiac fibrosis

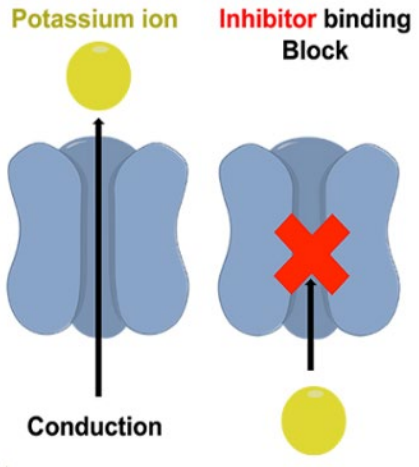
- Cardiac MRI: mild, diffuse LGE only in high-risk (33%)
  - ‘Scatter’ and arrhythmia
- Hard to study: need endomyocardial biopsy
  - Linkage (Lund) ongoing
  - UKB genetics, proteomics
- Perhaps why it’s under-appreciated in the clinic
  - Cardiologist reaction: meh

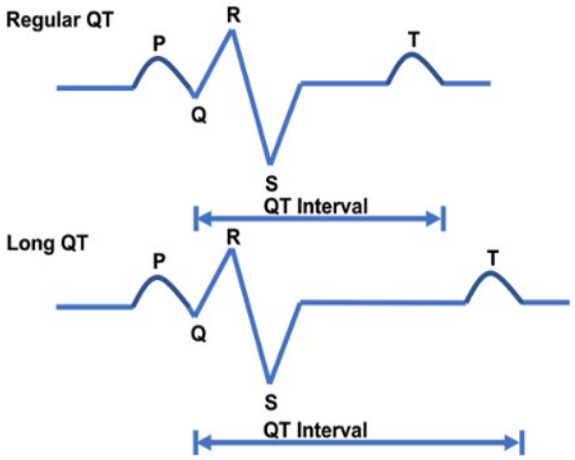
Three examples

- ECG to predict sudden cardiac death
- **ECG to measure hERG channel inhibition**
- CT to measure liver steatosis

Context

- hERG inhibition slows cardiac action potential repolarization and can increase arrhythmia risk
- Off-target hERG inhibition is a critical problem during drug development
- Responsible for 30% of drugs withdrawn 1990-2006





QT interval is a well-established biomarker for arrhythmia risk

https://www.nanoimagingservices.com/about/blog/characterization-of-the-herg-anti-target
Cavalluzzi, M.M. et al. (2020) Drug Disc Today 25(2), 344-366

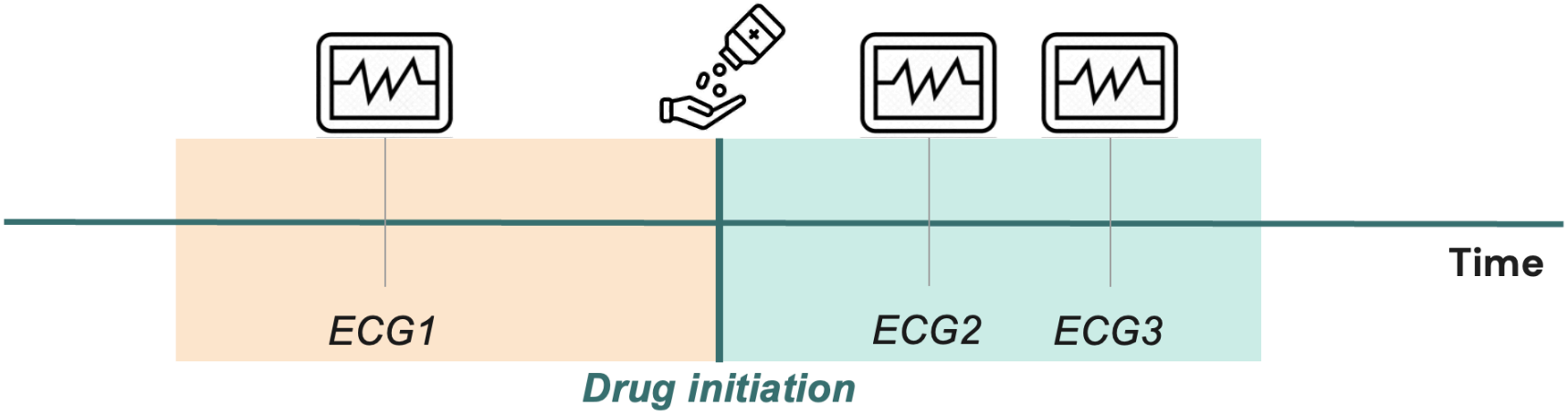
Two datasets to train and validate an ECG-based hERG predictor

Used for model training

	Medical administration data	Long term exposure data
Dandelion Dataset	**Medical administration data** (N=442,723) Granular **timestamps** for drug intake **Serial ECG data** pre- and post treatment	**Drug order data** (N=590,770) Focus on hERG-blocking drugs with **prolonged exposure** **Arrhythmia outcomes** via ICD codes

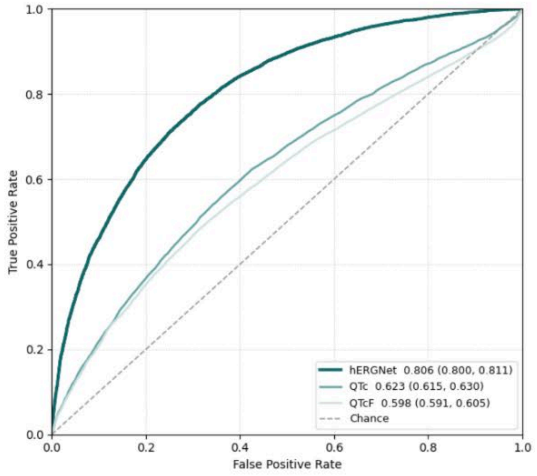
What we do

- Observe same patients ECGs prior and post drug initiation to isolate hERG effect



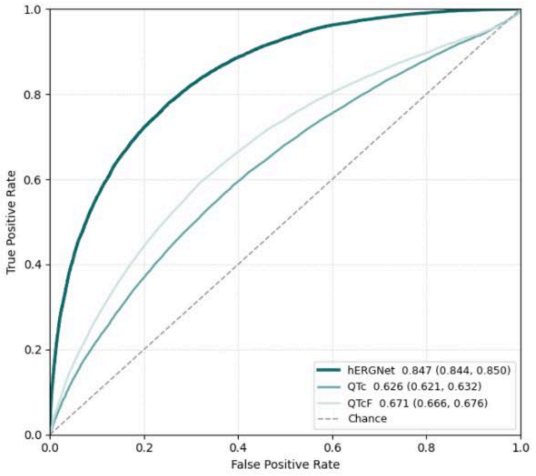
Good performance across validation settings…

Post treatment
hERG vs. Placebo drugs

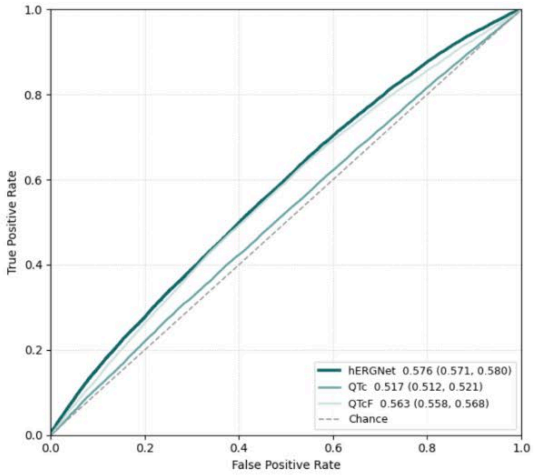


Discriminate ECGs taken before vs. after treatment

hERG drugs



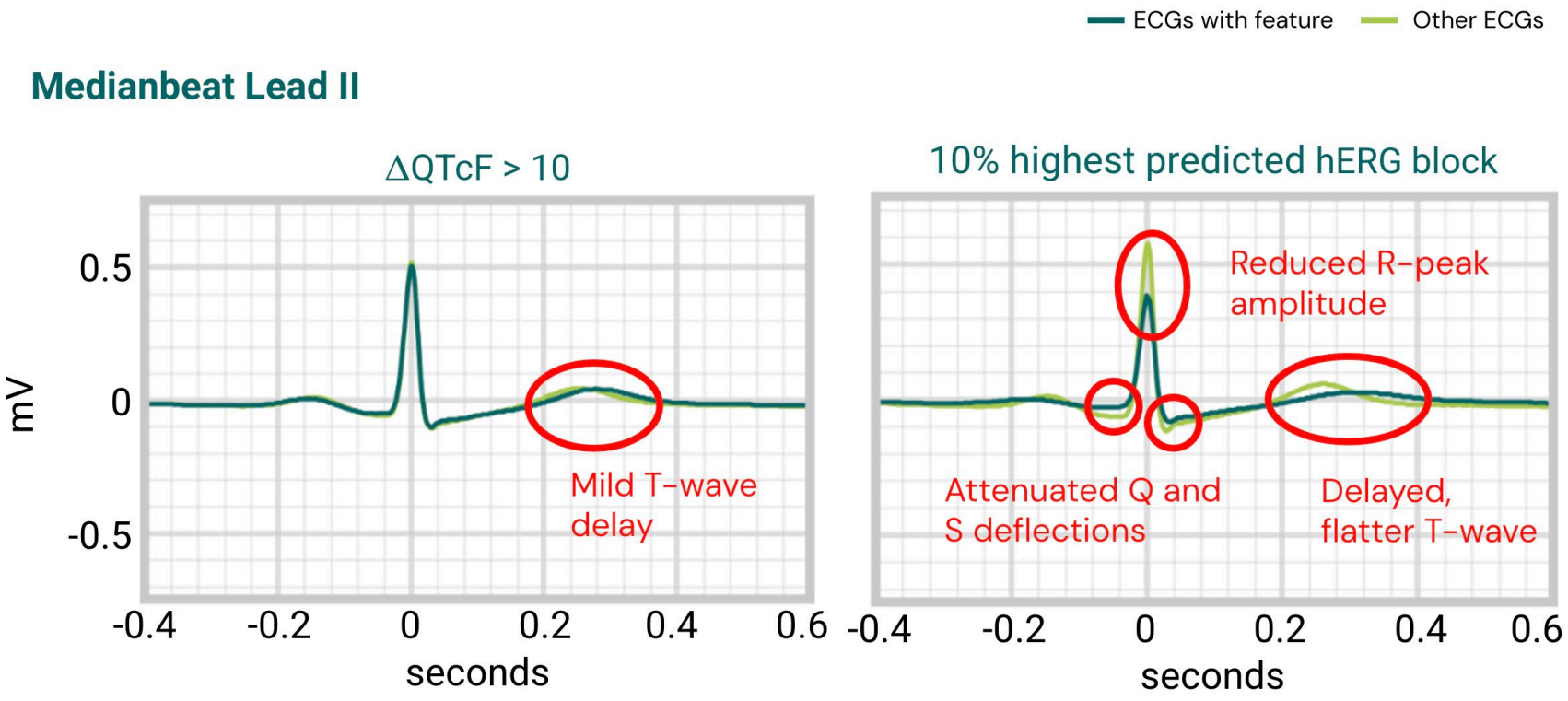
Placebo drugs



- …and generalizes to known hERG inhibitors which were not seen during training

The model identifies a distinct signature for hERG block

- Predicted hERG block shows ECG morphology changes beyond QTc prolongation



External validation using RCT data

- FDA Critical Path Initiative RCT: Randomized, double-blind, placebo-controlled crossover trial with the ECG dataset released publicly

- Strong ground truth:
  - 22 healthy volunteers, each receiving 4 drugs + placebo in a 5-period crossover
  - ECGs collected pre-dose + 15 post-dose time points, with paired plasma concentration (PK) measurements to confirm exposure

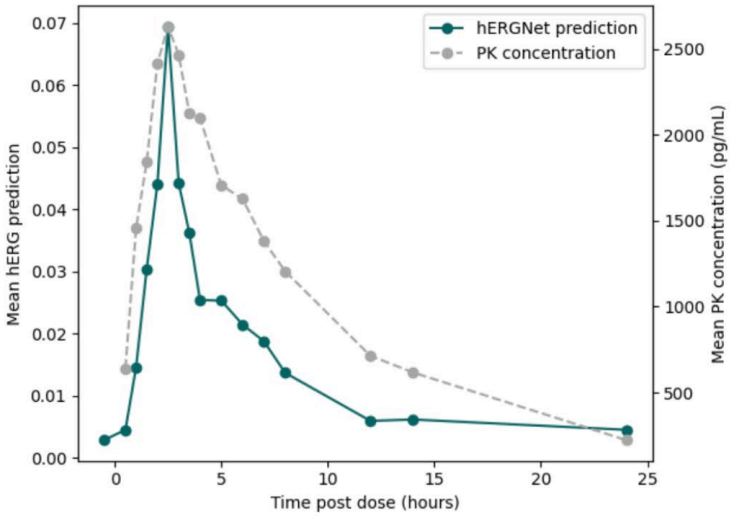
Johannesen, L., Vicente, J., Mason, J. W., Sanabria, C., Waite-Labott, K., Hong, M., ... & Strauss, D. G. (2014). Differentiating drug-induced multichannel block on the electrocardiogram: randomized study of dofetilide, quinidine, ranolazine, and verapamil. Clinical Pharmacology & Therapeutics, 96(5), 549-558.

Predictions capture strength of hERG blockade, not just signature

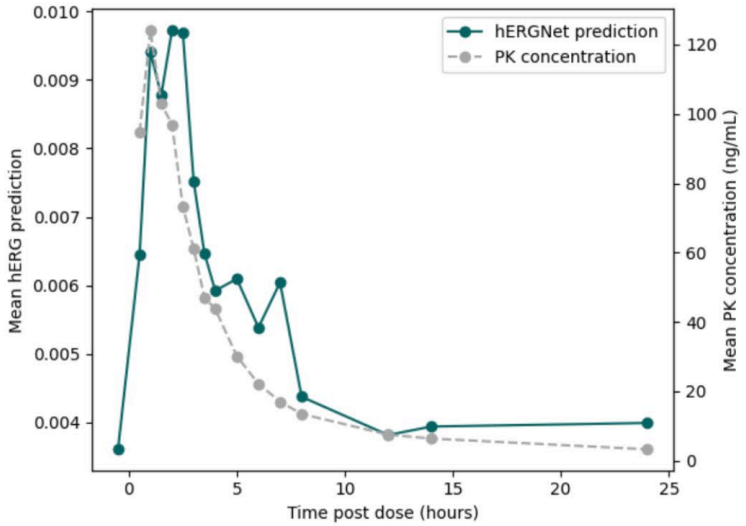
- Model predictions track drug plasma concentrations

External validation in trial data: AI ECG predictions follow plasma concentration time-courses.

Dofetilide

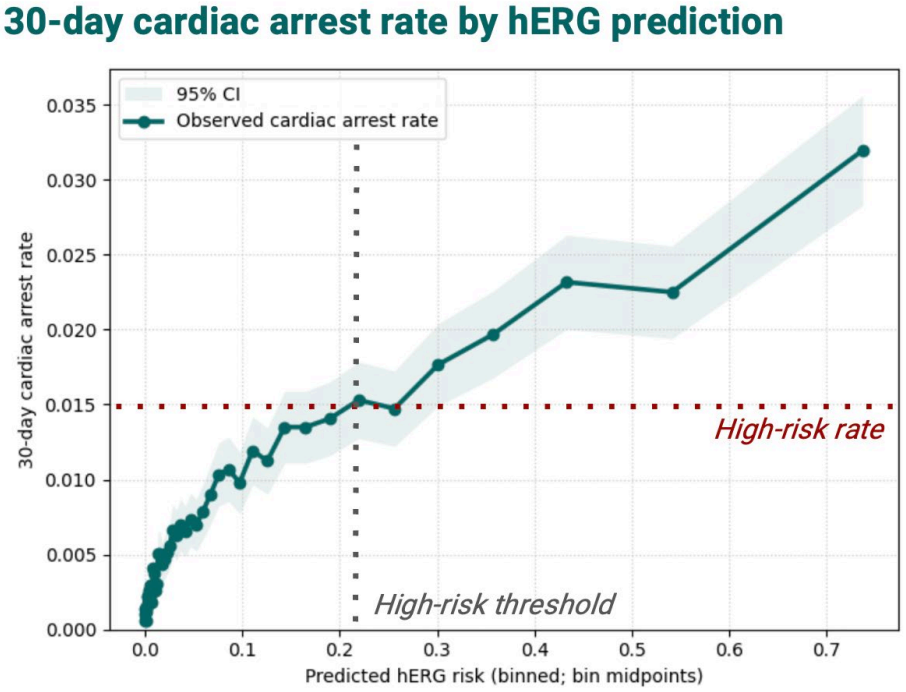


Verapamil

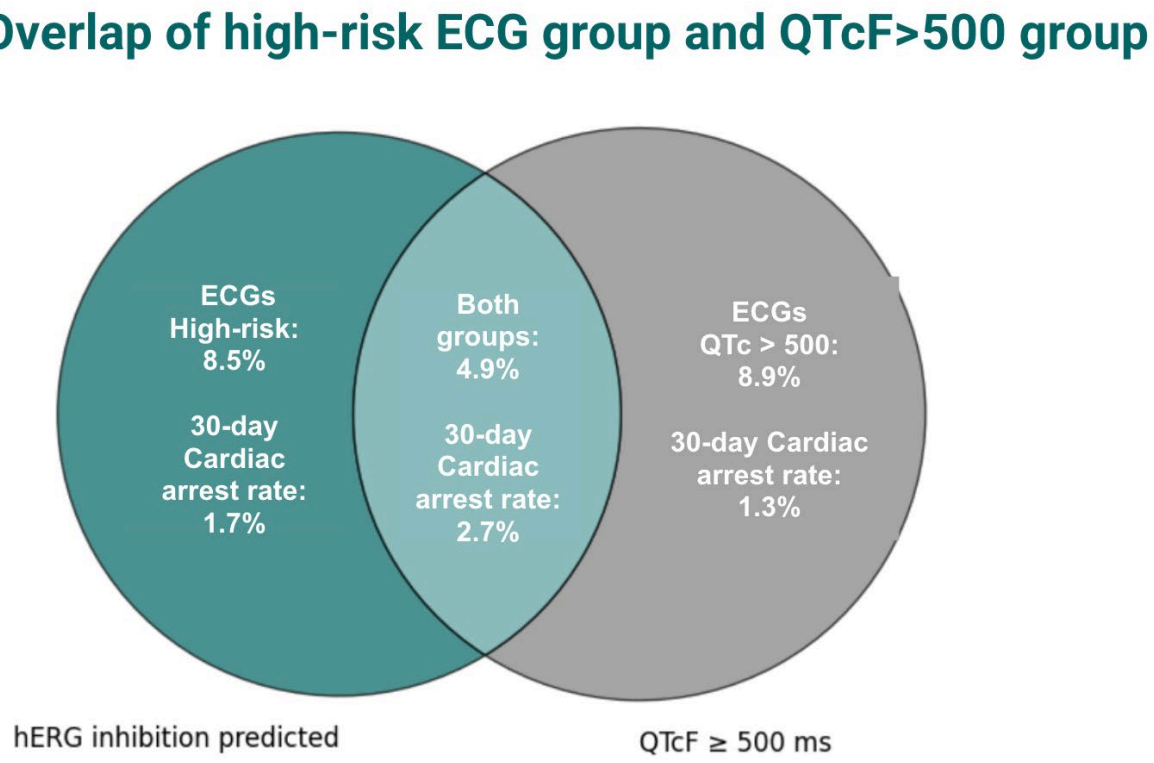


Predicted hERG inhibition is connected to future MI risk

- Higher hERG prediction relates to increased 30-day cardiac arrest risk
- Allows us to define a model-based high-risk group



High-risk cohort complements existing measures



- 63% of ECGs with high hERG prediction are unsuspected based on QTc > 500

Three examples

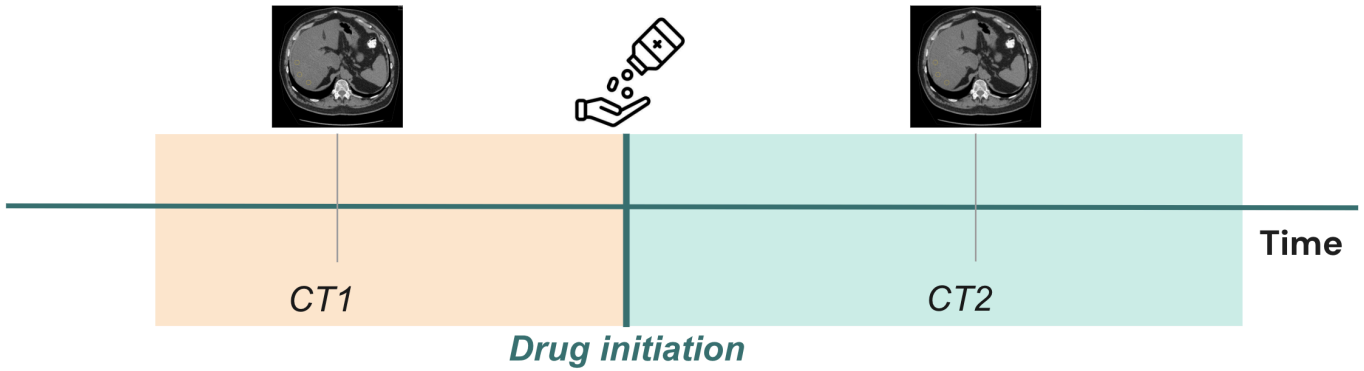
- ECG to predict sudden cardiac death
- ECG to measure hERG channel inhibition
- **CT to measure liver steatosis**

Do GLP-1s impact liver health earlier in disease progression?

- GLP-1s help with MASH, but what about MAFLD?
  - August 2025: Semaglutide approved for MASH w/ moderate-to-advanced fibrosis
  - Several trials are underway with no fibrosis requirement
- The broader challenge: we want interventions that prevent progression rather than treating advance illness, but these clinical trials are long and expensive
- RWD might help, but gold standard (MRI-PDFF) is rarely performed in clinical practice

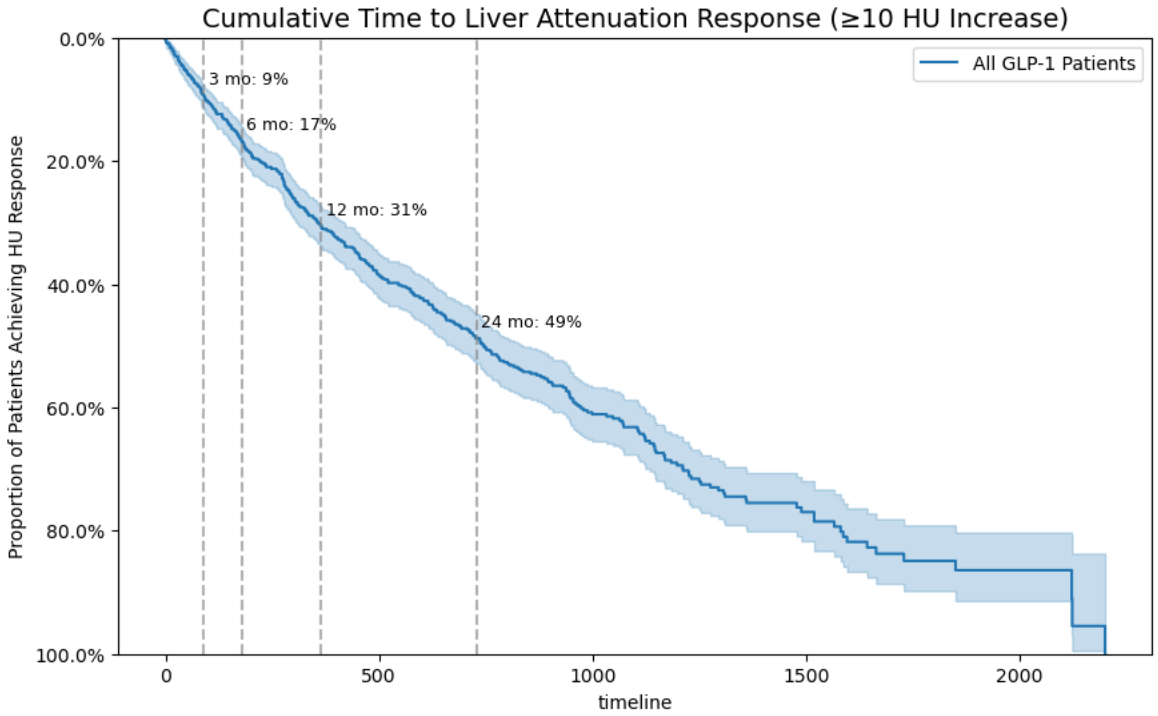
What we do

- Use an AI algorithm (Nanox HealthFLD) to measure liver steatosis using abdominal CTs take for another reason
- Observe same patients abdominal CT prior and post GLP-1 initiation (n = 951)
- Estimate time-to-event for clinically meaningful reduction in liver steatosis



What we find

- 46% of patients in cohort see a clinically meaningful reduction in hepatics steatosis (>10 HU); 54% do not
- 49% of responders show response within 2 years



Summary

- Images, waveforms: Rich data sources, central to medicine
  - At the same time: humans can't fully process them

- Pre-market use cases for ML-derived biomarkers
  - Enrich clinical trials with high-risk patients
  - Enable clinical trials focused on primary prevention

- Post-market use cases for ML-derived biomarkers
  - Rigorous use of RWD to generate evidence for indication expansion
  - Screening of patients for under-diagnosed conditions