Advanced Cardiovascular Risk Detection For the Critical Decades

OCTOBER 30, 2025

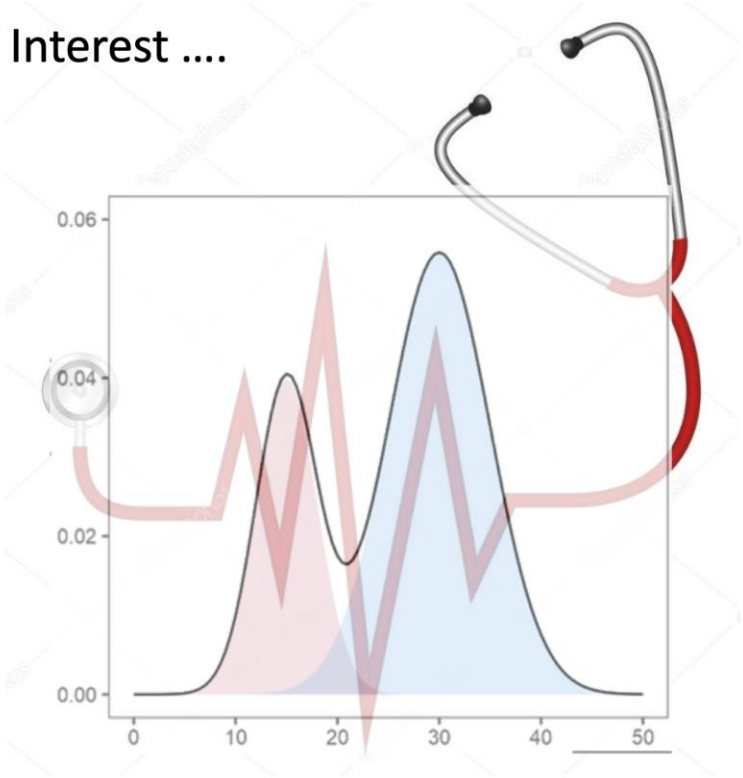
Bayes through the Life Course
Sarah Urbut, MD PhD

HEART HOUSE ROUNDTABLES

AMERICAN COLLEGE of CARDIOLOGY®

Disclosures
No disclosures to report

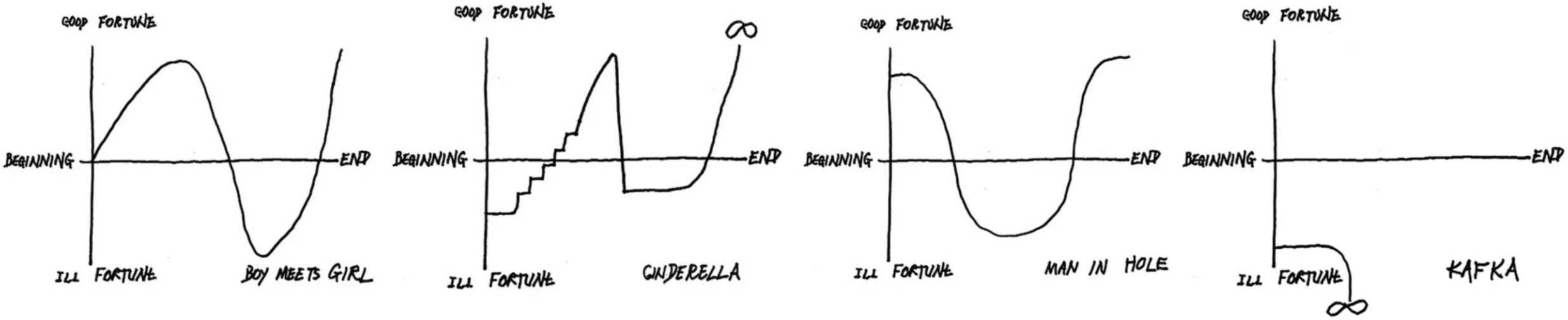
Conflicts of Interest ….



BROAD INSTITUTE

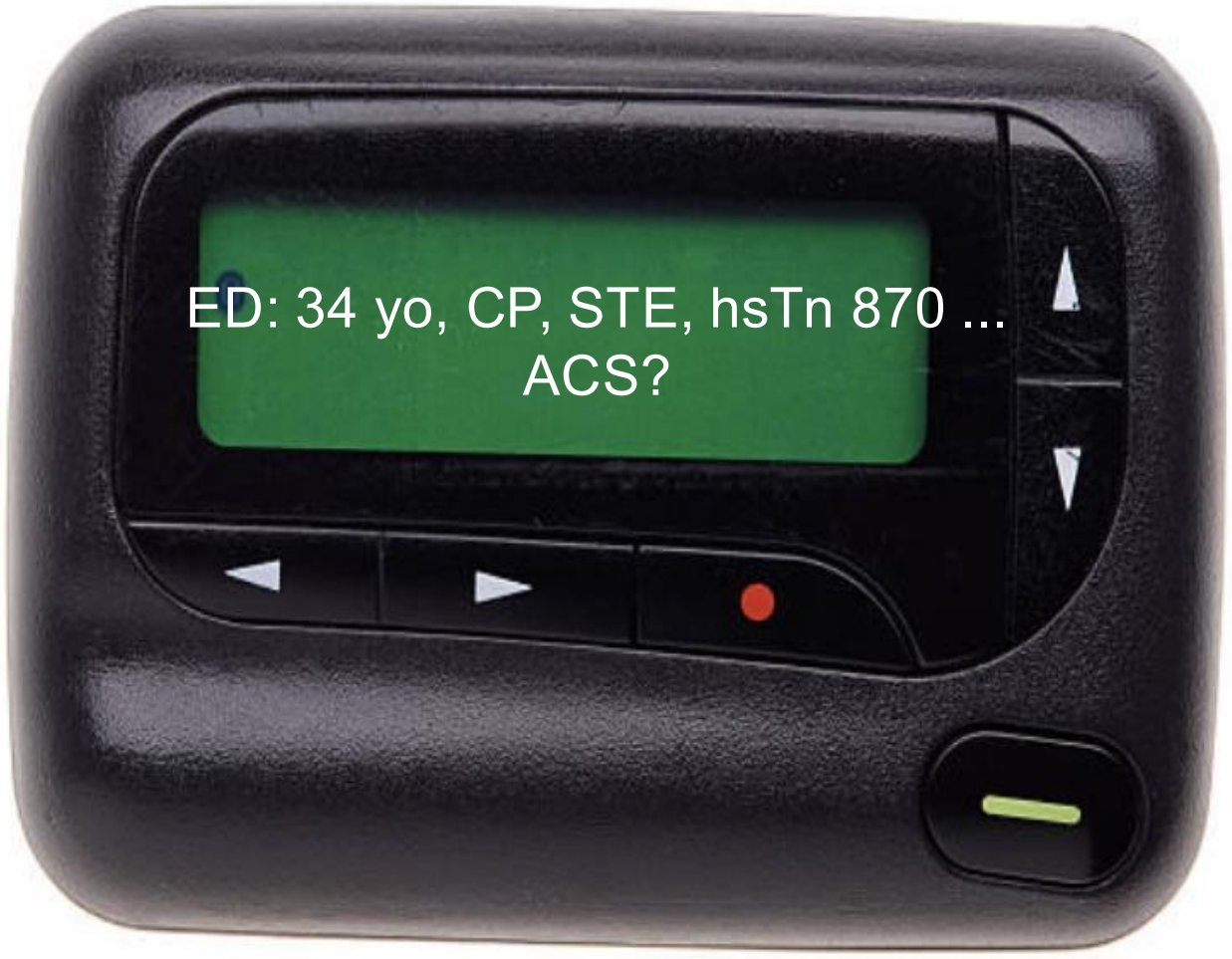
MASSACHUSETTS GENERAL HOSPITAL
CORRIGAN MINEHAN HEART CENTER

Shape of a story



Vonnegut, 1955

Everything is moving



JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY
© 2018 BY THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION
PUBLISHED BY ELSEVIER

VOL. 72, NO. 16, 2018

EDITORIAL COMMENT

Polygenic Risk Scoring for Coronary Heart Disease

The First Risk Factor*

Pradeep Natarajan, MD, MMSc

Absolute risk assessment for coronary heart disease (CHD) based on a composite of risk factors is the foundation of contemporary CHD prevention (1). Risk scores serve: 1) to identify individuals at greater risk of CHD over a given time frame; and 2) to establish candidacy for pharmacological preventive strategies. In this issue of the *Journal*, Inouye et al. (2) describe a framework of using polygenic risk scoring to complement clinical risk scoring to identify both high- and low-risk individuals.

SEE PAGE 1883

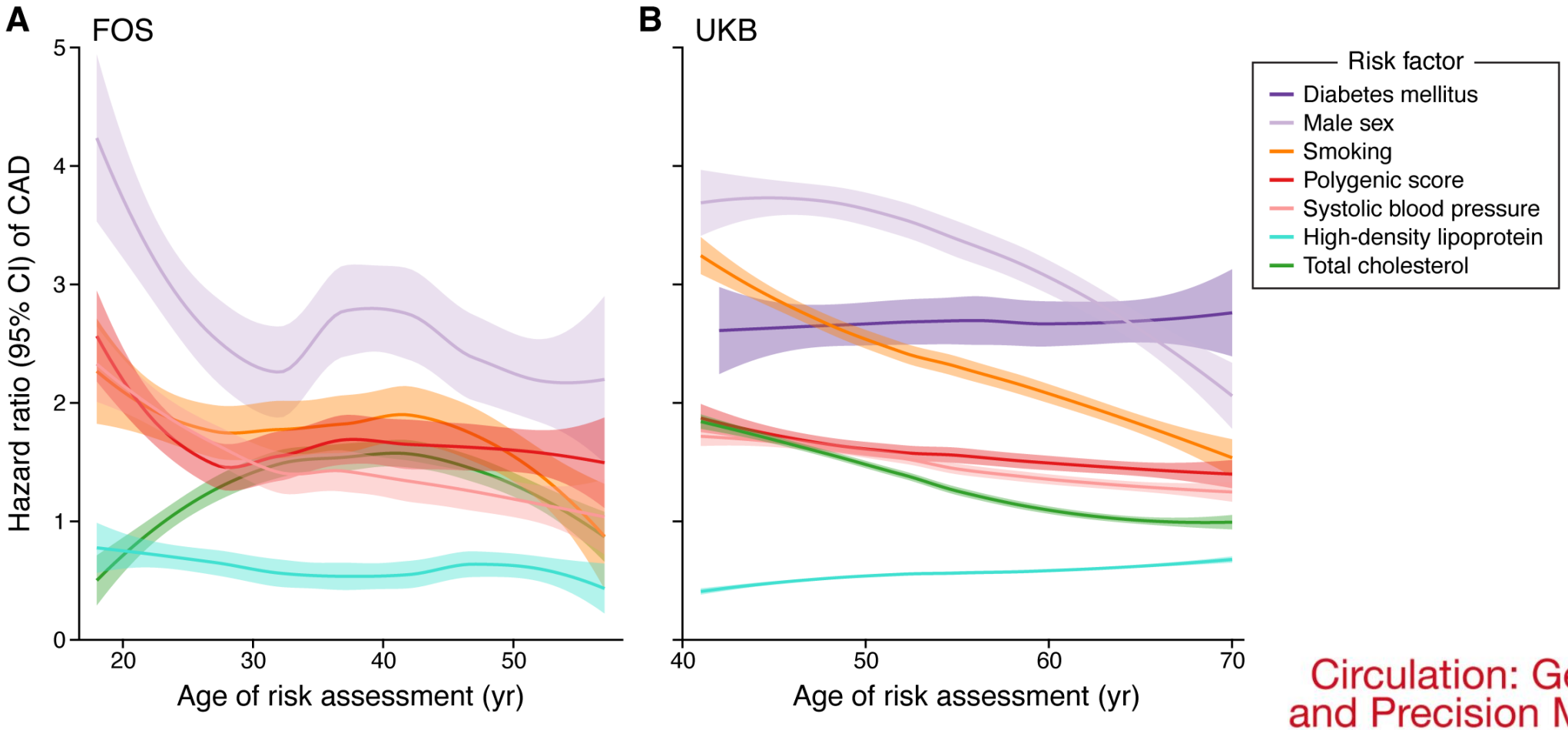
A HISTORICAL PERSPECTIVE OF

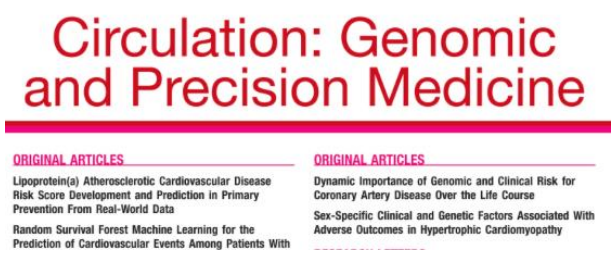
(LDL) cholesterol lowering among individuals with multiple CHD risk factors (5).

In the 1990s, the Framingham risk score, incorporating multiple risk categories to predict the onset of CHD within 10 years, was incorporated into the ATP-III (6). Using largely the same risk categories, the Pooled Cohort Equations incorporated additional cohorts and non-European Americans to develop a 10-year risk estimator for atherosclerotic cardiovascular disease. The Pooled Cohort Equations was adopted by the 2013 American College of Cardiology/American Heart Association joint cholesterol guidelines and is widely used in practice (1).

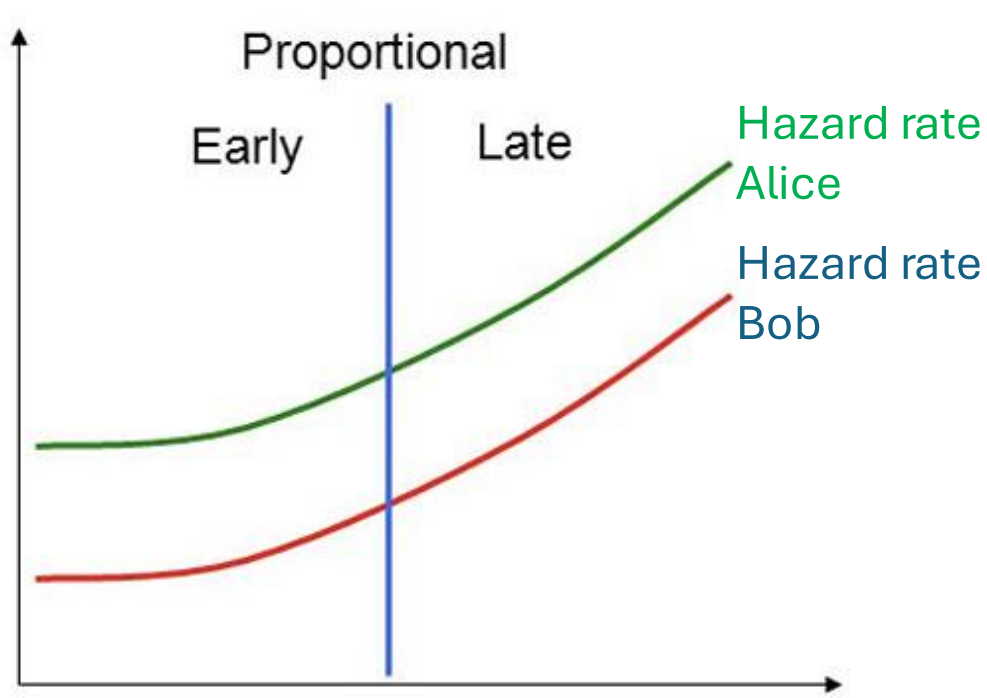
However, among younger individuals, the ability

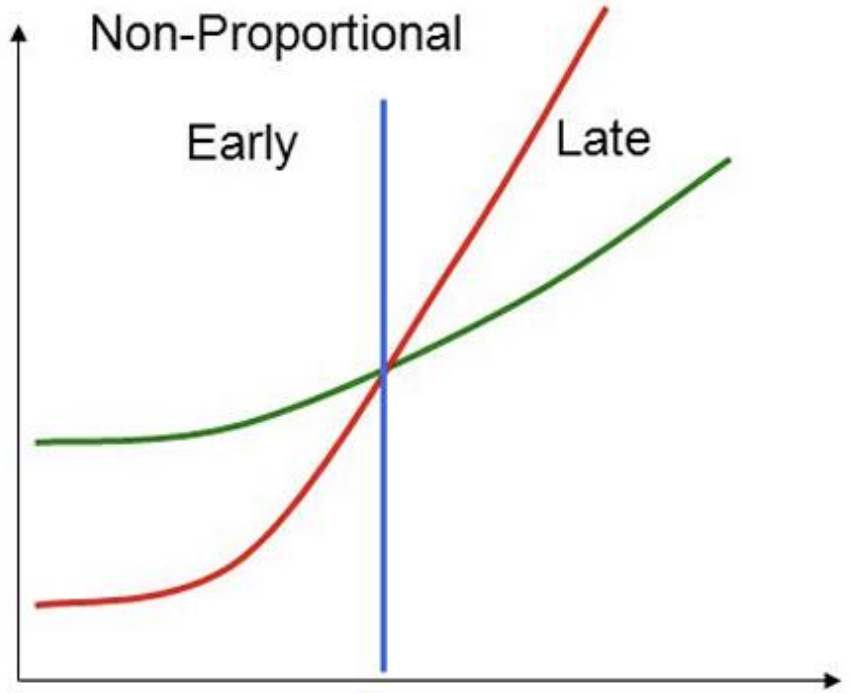
Dynamic hazard – different roles, same end





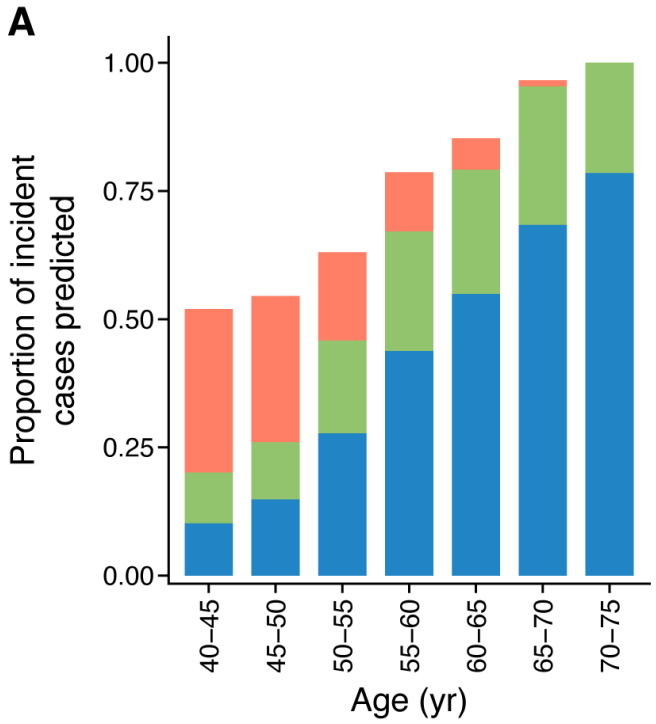
How does lifetime risk depend on … when the question is asked?



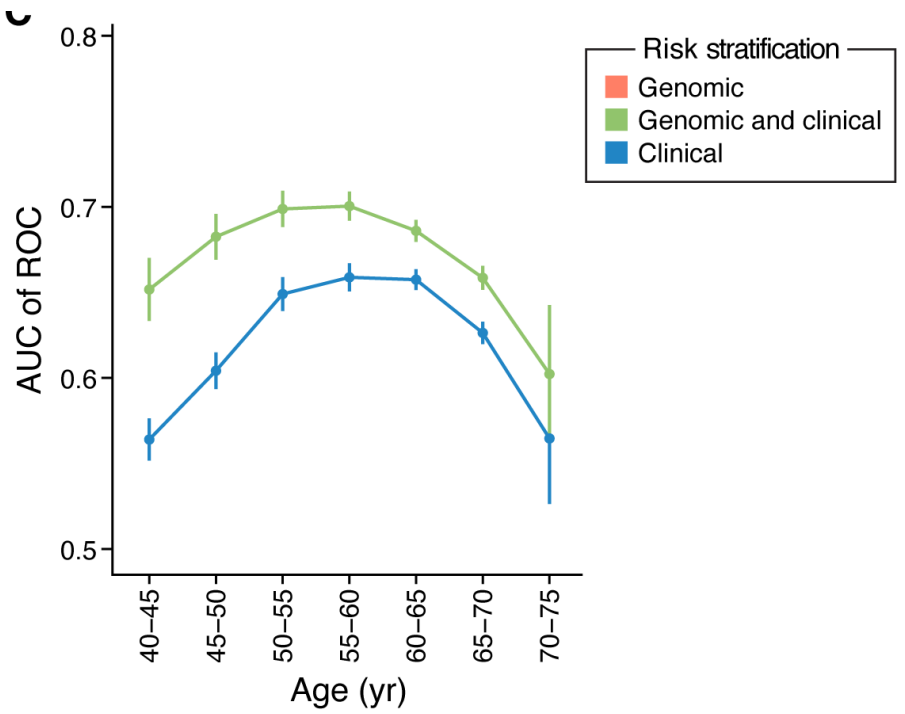


The Cox Proportional Hazards model assumes the hazards are ***proportional***:
the ***relative*** hazard ratio remains ***constant over time***

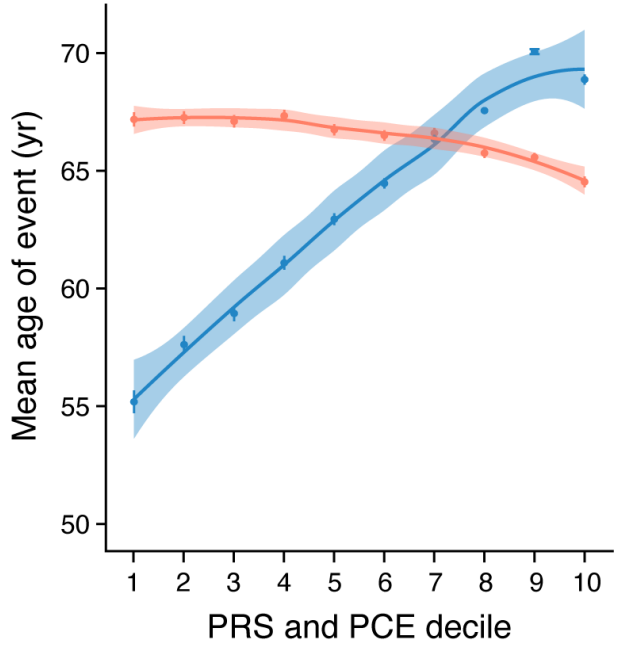
Predicts not only early disease, which we sometimes think of, but also all disease early



POWER

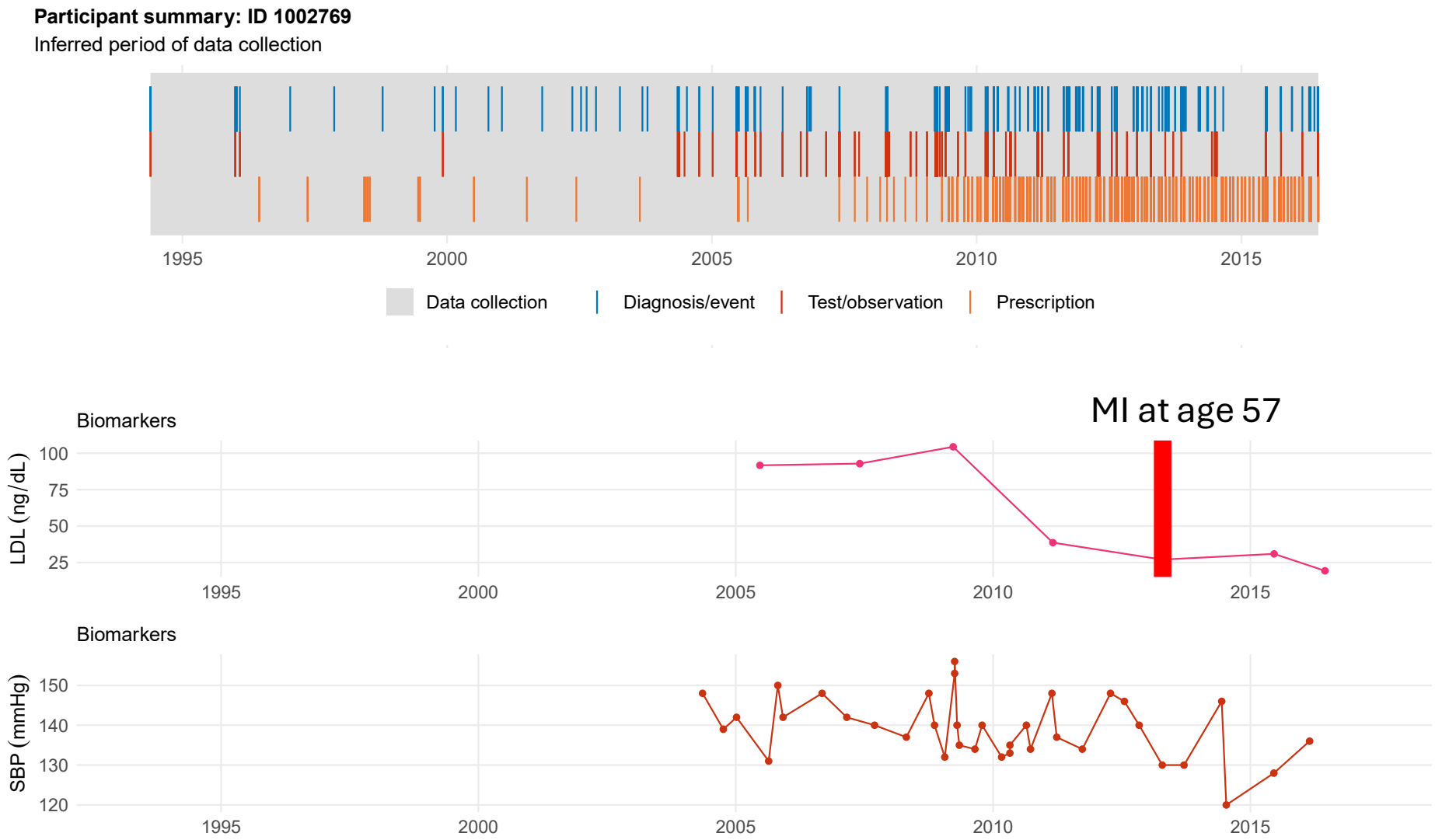


DISCRIMINATION

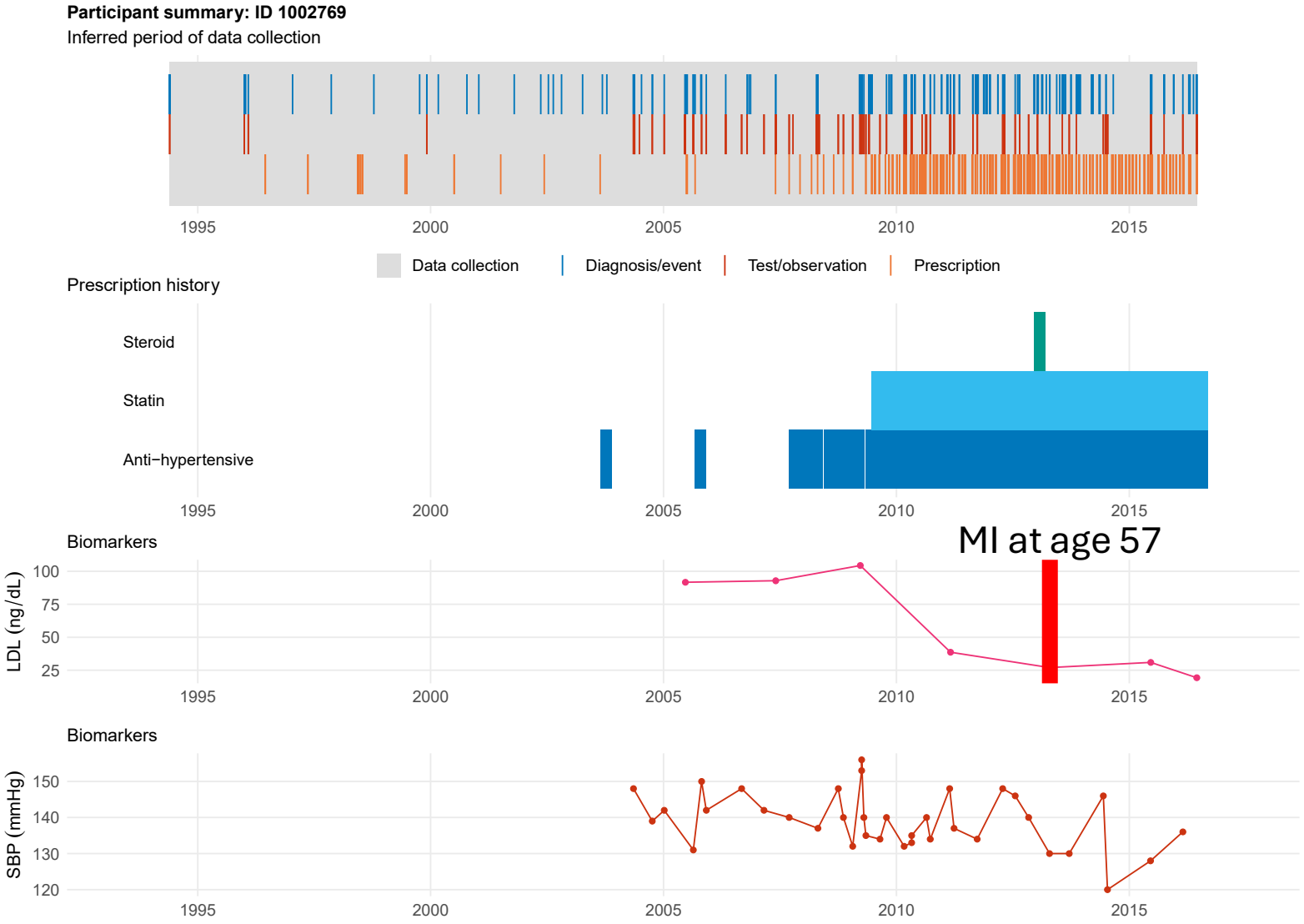


Urbut et al, Circulation GPM 2025

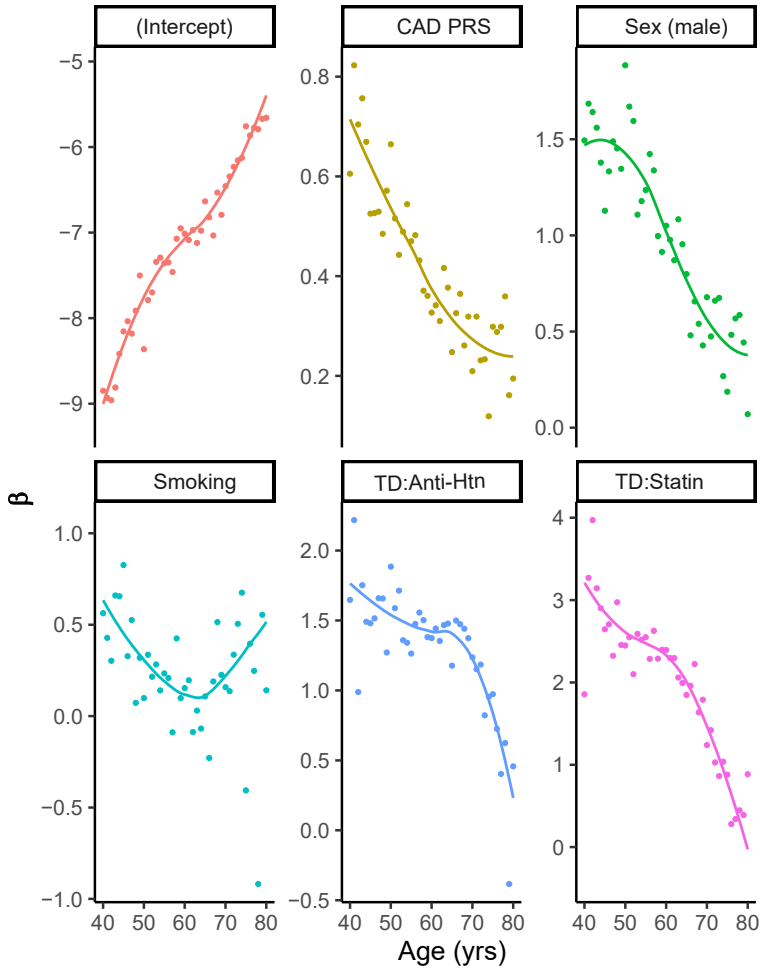
Using the EHR to be predictive, not responsive

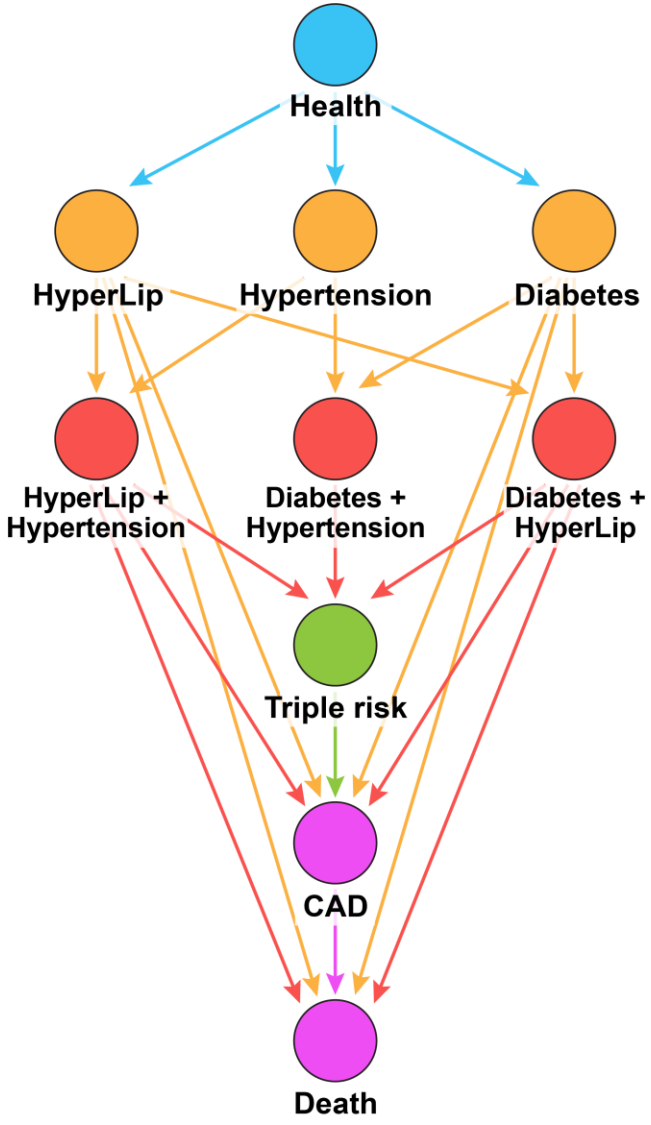


Using the EHR to be predictive, not responsive

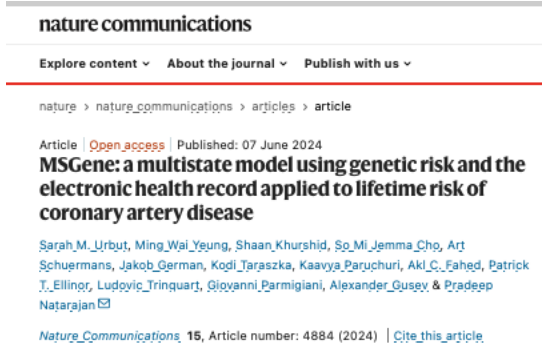


MSGene: multistate model using genetics for dynamic prediction





1. Health
2. Single risk factor
3. Double risk
4. Triple risk
5. Absorbing



nature communications

Explore content · About the journal · Publish with us

nature › nature communications › articles › article

Article | Open access | Published: 07 June 2024

MSGene: a multistate model using genetic risk and the electronic health record applied to lifetime risk of coronary artery disease

Sarah M. Urbut, Ming Wai Yeung, Shaan Khurshid, So Mi Jemma Cho, Art Schuermans, Jakob German, Kodi Taraszka, Kaavya Paruchuri, Akl C. Fahed, Patrick T. Ellinor, Ludovic Trinquart, Giovanni Parmigiani, Alexander Gusev & Pradeep Natarajan

Nature Communications 15, Article number: 4884 (2024) | Cite this article

Chapter 4: Everyone is a Bayesian

2^n = a numbers problem

Urbut et al, 2025 medRxiv

medRxiv

THE PREPRINT SERVER FOR HEALTH SCIENCES

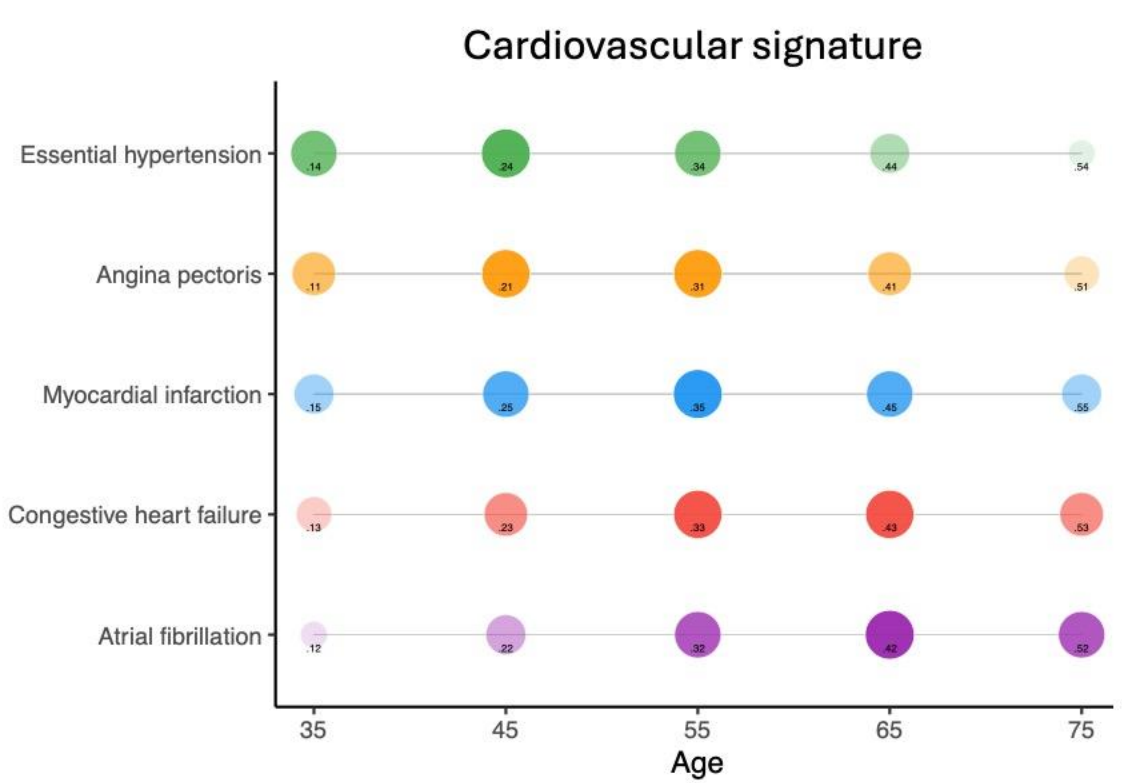
Follow this preprint

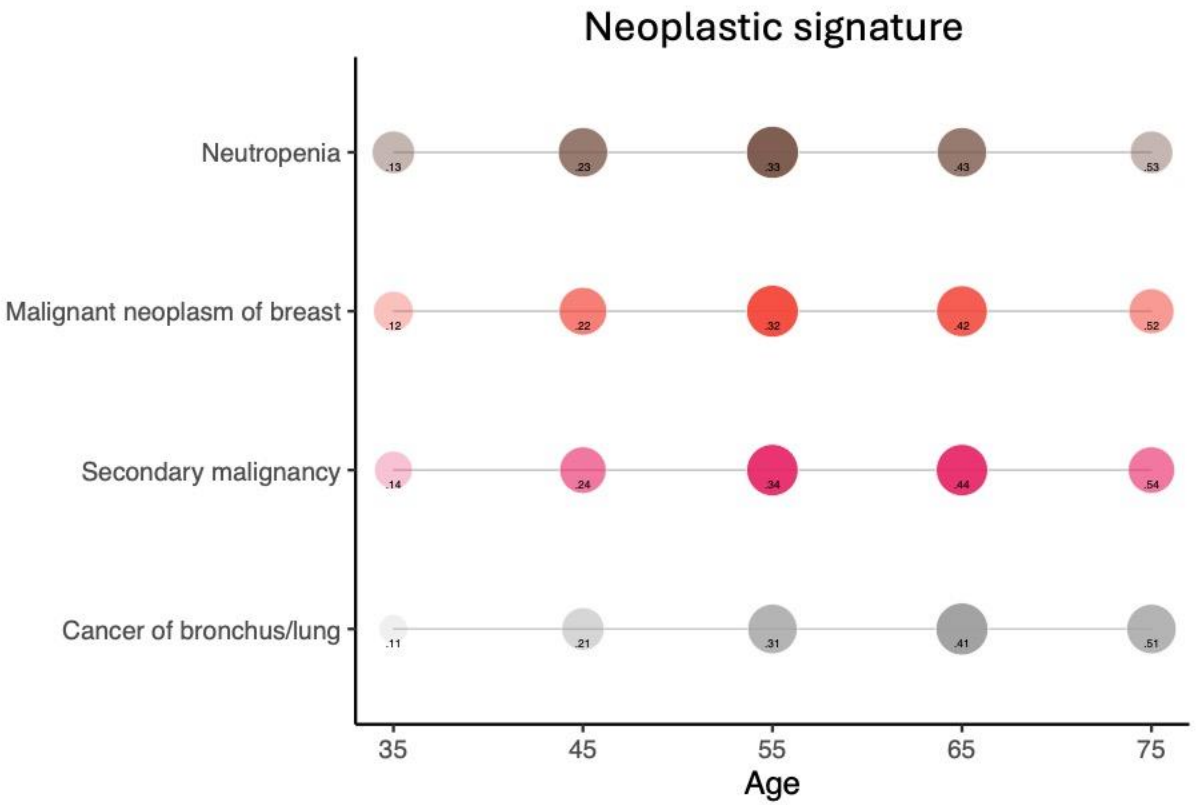
ALADYNOULLI: A Bayesian approach to disease progression modeling for genomic discovery and clinical prediction

Sarah M. Urbut, Yi Ding, Tetsushi Nakao, Xilin Jiang, Leslie Gaffney, Anika Misra, Whitney Hornsby, Jordan W. Smoller, Alexander Gusev, Pradeep Natarajan, Giovanni Parmigiani

doi: https://doi.org/10.1101/2024.09.29.24314557

Latent patterns of disease . . . within a signature





Signatures: patterns of disease co-occurrence that vary in time

Urbut et al, 2024

Time varying trajectories . . .

But when do you ask the question?

Patient A: Metabolic → Cancer

Classic metabolic syndrome evolving into malignancy

Condition	Age 35	Age 40	Age 45	Age 50	Age 55	Age 60	Age 65
Hypertension	0	1	1	1	1	1	1
Type 2 Diabetes	0	0	1	1	1	1	1
CAD	0	0	0	1	1	1	1
Colon Polyps	0	0	0	0	1	1	1
Colon Cancer	0	0	0	0	0	1	1
Metastasis	0	0	0	0	0	0	1

Patient B: Inflammatory → CVD → Neuro

Inflammatory disease followed by cardiovascular complications and neurological issues

Condition	Age 35	Age 40	Age 45	Age 50	Age 55	Age 60	Age 65
Rheumatoid Arthritis	1	1	1	1	1	1	1
IBD	0	1	1	1	1	1	1
CAD	0	0	0	1	1	1	1
Heart Failure	0	0	0	0	1	1	1
Depression	0	0	0	0	1	1	1
Cognitive Decline	0	0	0	0	0	1	1

Patient C: Early CVD → GI → Metabolic

Early cardiovascular disease with later digestive and metabolic complications

Condition	Age 35	Age 40	Age 45	Age 50	Age 55	Age 60	Age 65
CAD	1	1	1	1	1	1	1
Heart Failure	0	1	1	1	1	1	1
GERD	0	0	1	1	1	1	1
IBD	0	0	0	1	1	1	1
Type 2 Diabetes	0	0	0	0	1	1	1
Obesity	0	0	0	0	1	1	1

Metabolic CVD Cancer Inflammatory GI Neuro

- *ACC Statement on Inflammation*: Inflammation is therapeutic target in CVD
- CANTOS trial: IL-1β inhibition reduces CV events

Urbut et al, 2024

Everyone is Bayesian

Joint consideration: discovery and prediction

$$P(\Pi|Diagnoses) \propto P(Diagnoses|\Pi)\ p(\Pi)$$

Continuously updated posteriors

Individual data likelihood (EHR, clinical data)

Individual predilection to a signature

Population level signatures

Thomas Bayes, History of life insurance in its formative years
American Conservation Co:, 1936, Chicago

This is statistics: Aladynoulli

$$\phi \sim N(\mu_d + \psi_{kd}, K)$$

$$\lambda \sim N(\gamma_k \, g_i + r_i, K)$$

Hazard for individual *i* of disease *d* at time *t*:

$$\pi_{idt} = \sum_k f(\lambda_{idt}) f(\phi_{idt})$$

Individual Population

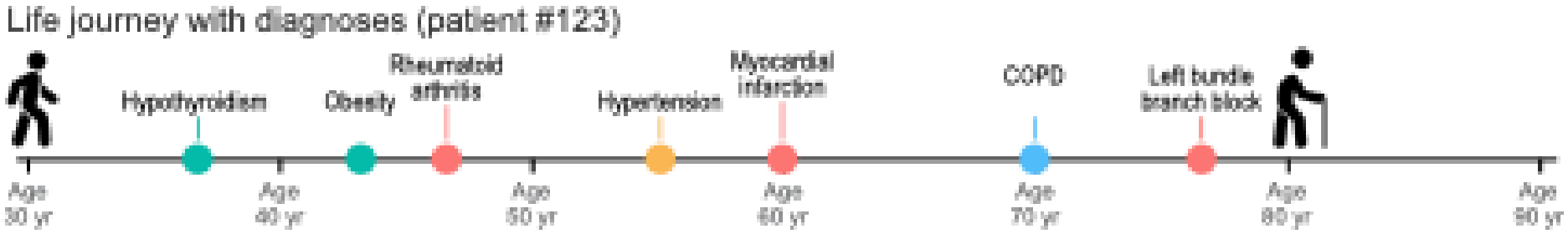
We are all Bayesians

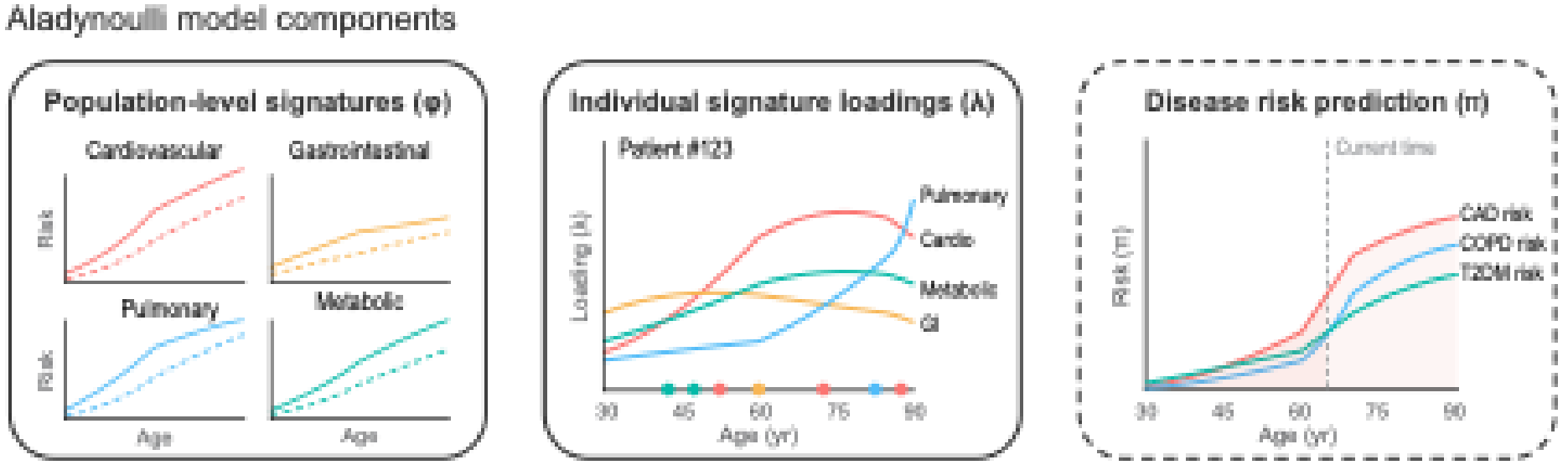
$$l_{id} = \sum_{t<E_{id}} log(1-\pi_{idt}) + Y_{idt}\pi_{idt} + (1-Y_{idE_{id}})(1-\pi_{idt})$$

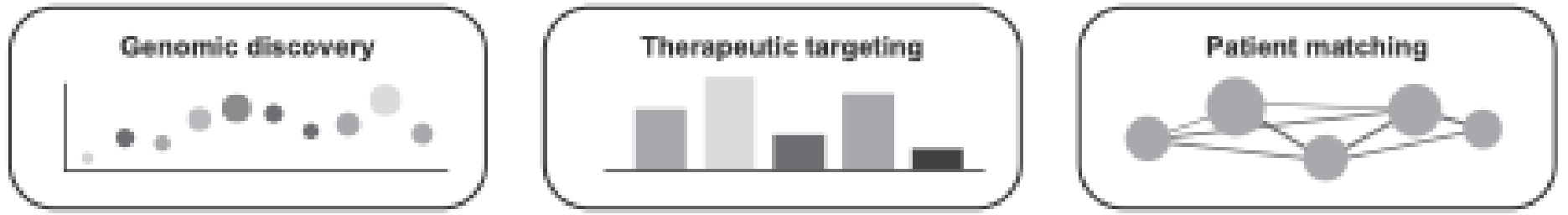
At risk Event Censored

$$\text{Post} \propto Likelihood \times Prior$$

$$P(Model|Diagnoses) = P(Diagnoses|Model) \cdot P(Model)$$


Life journey with diagnoses (patient #123)
Hypothyroidism
Obesity
Rheumatoid arthritis
Hypertension
Myocardial infarction
COPD
Left bundle branch block
Age 30 yr
Age 40 yr
Age 50 yr
Age 60 yr
Age 70 yr
Age 80 yr
Age 90 yr


Aladynoulli model components
Population-level signatures (φ)
Cardiovascular
Gastrointestinal
Pulmonary
Metabolic
Risk
Age
Individual signature loadings (λ)
Patient #123
Loading (λ)
Pulmonary
Cardio
Metabolic
GI
30
45
60
75
90
Age (yr)
Disease risk prediction (π)
Current time
Risk (π)
CAD risk
COPD risk
T2DM risk
Age (yr)


Applications
Genomic discovery
Therapeutic targeting
Patient matching

Signatures: Characteristic patterns of incidence and timing $f(\phi_{idt})$

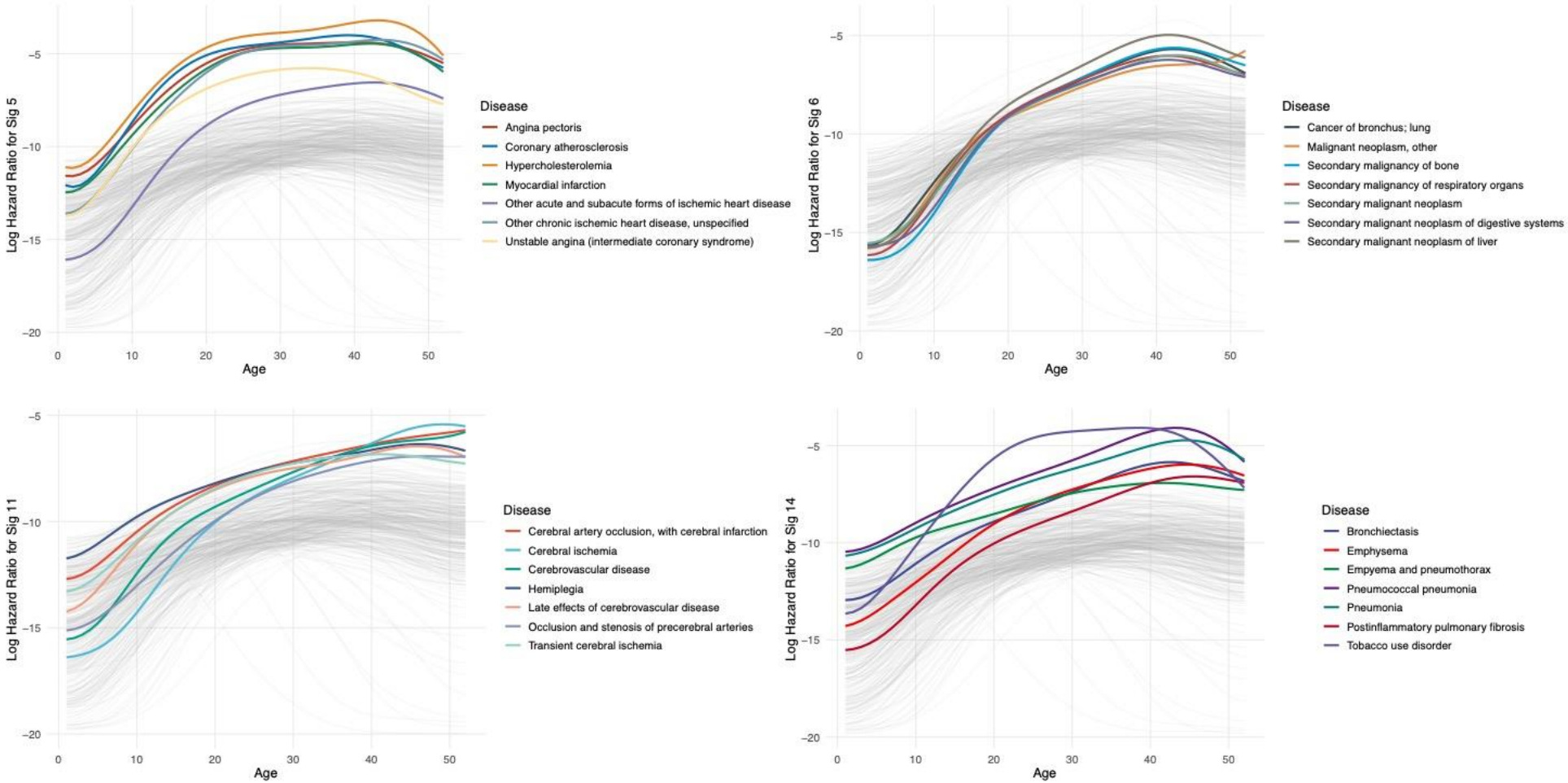
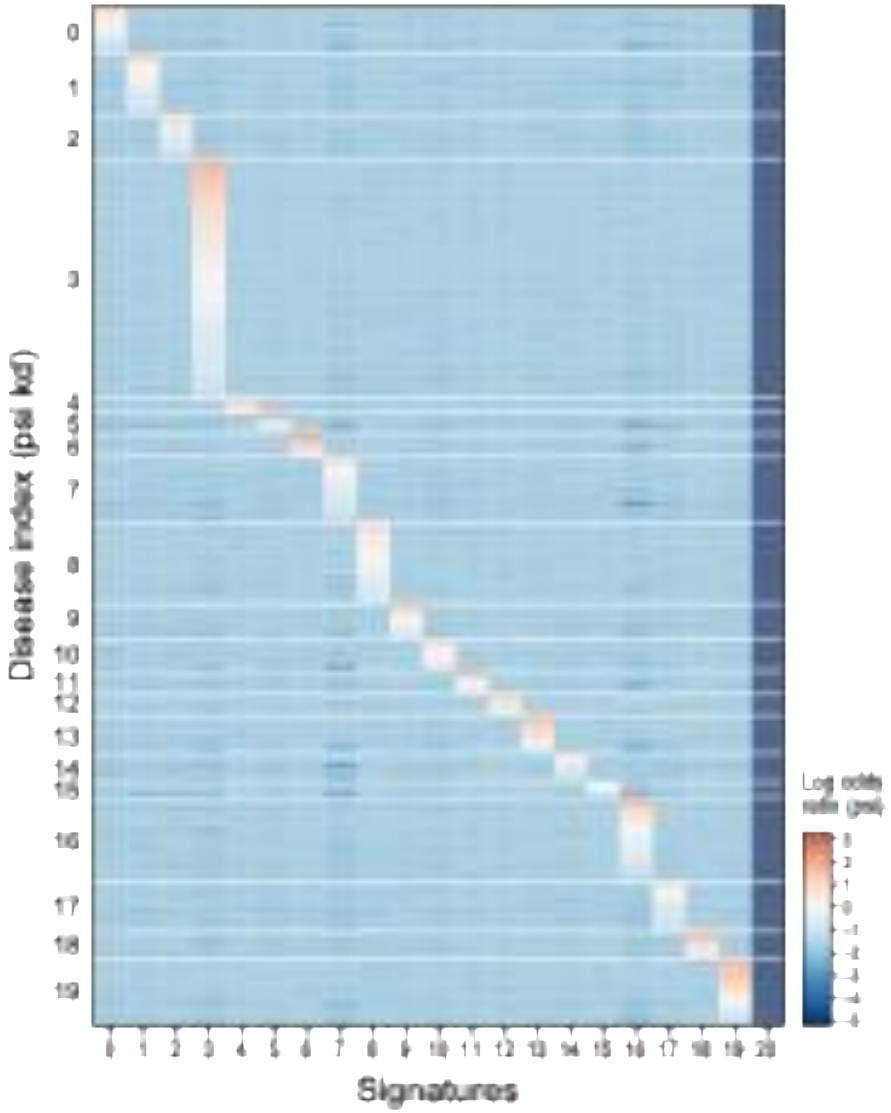
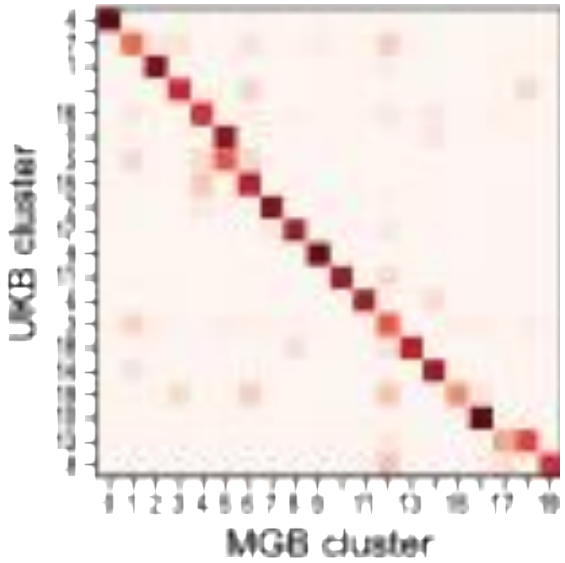


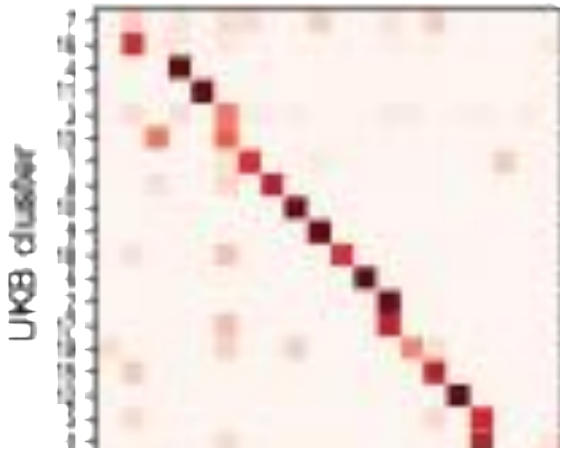
Figure: Temporal evolution of disease probabilities across signatures

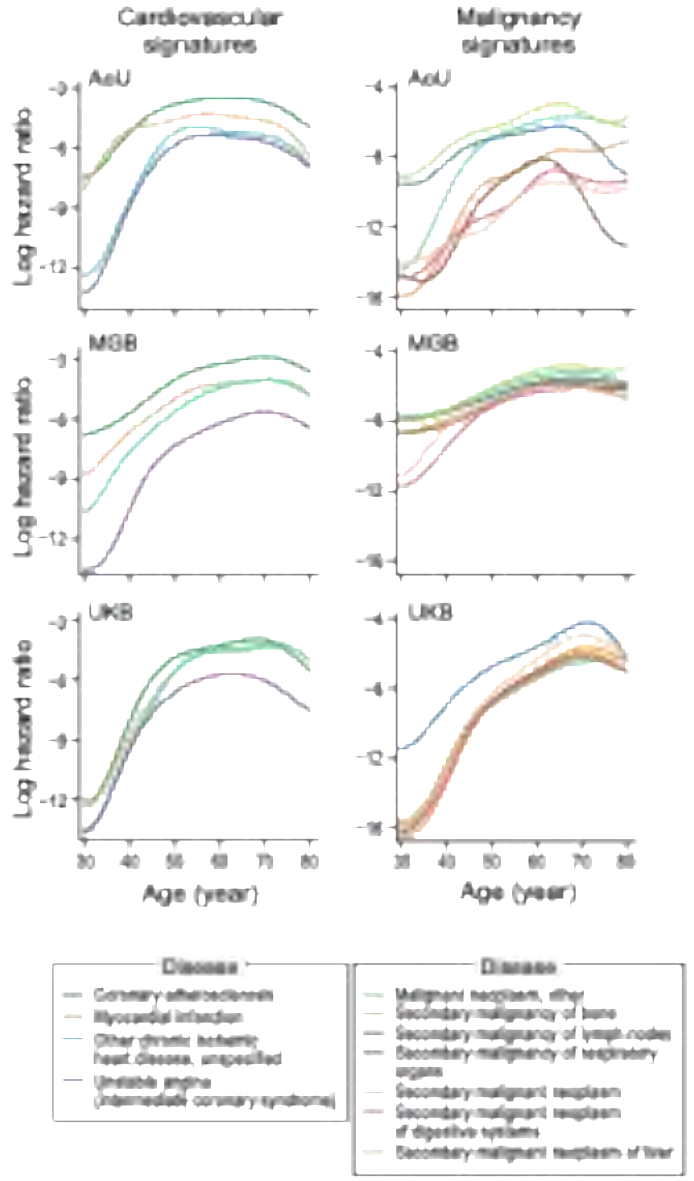
Varying degree of allegiance



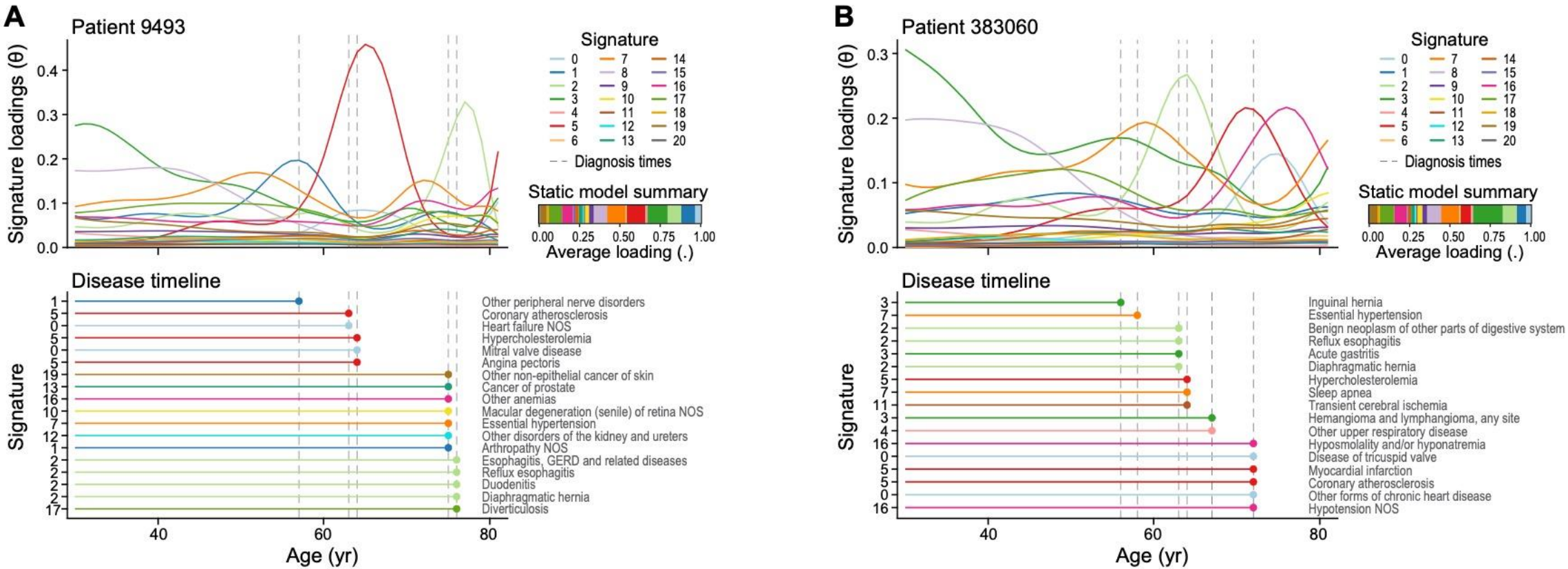
Consistency across biobanks



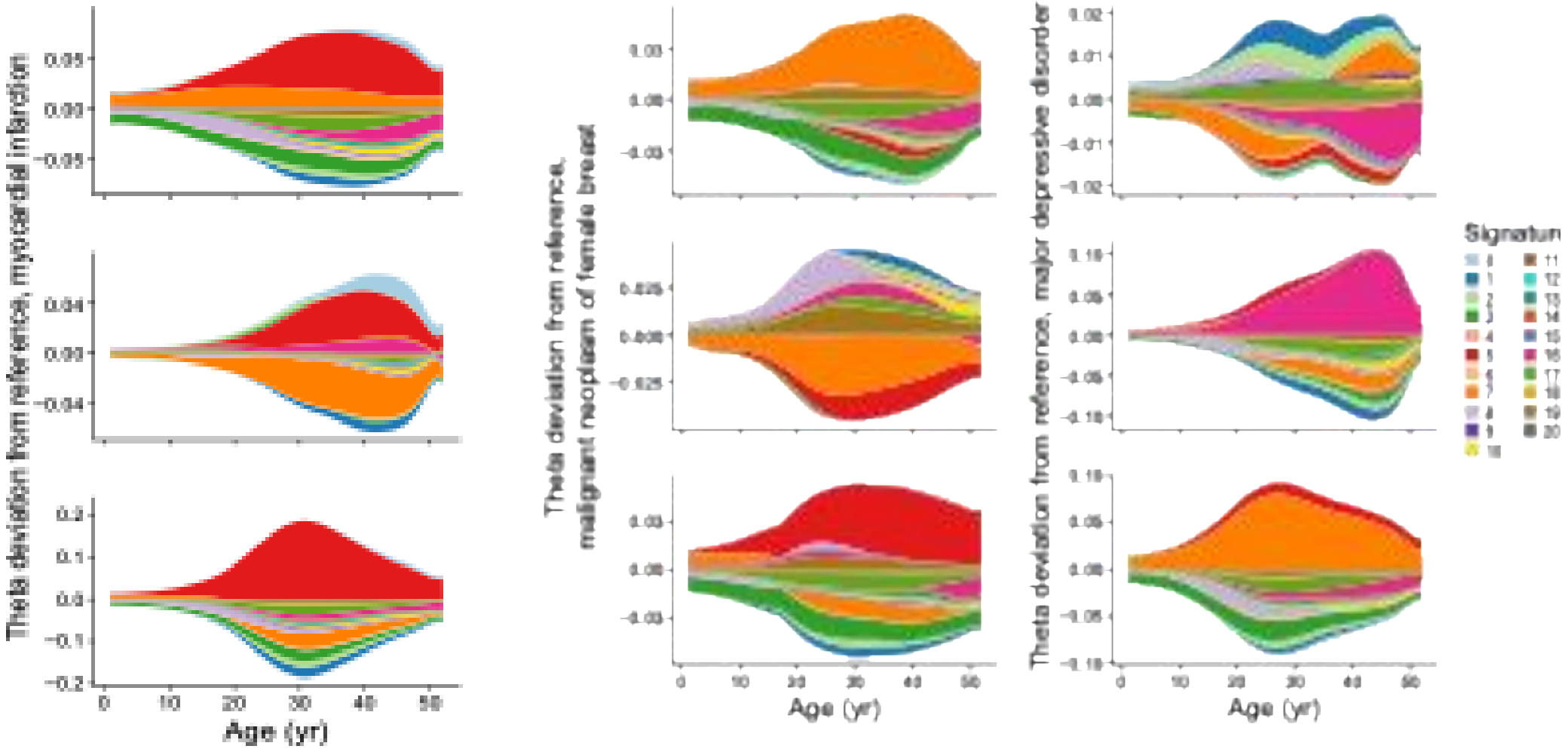




Walking the time-line… $f(\lambda_{idt})$

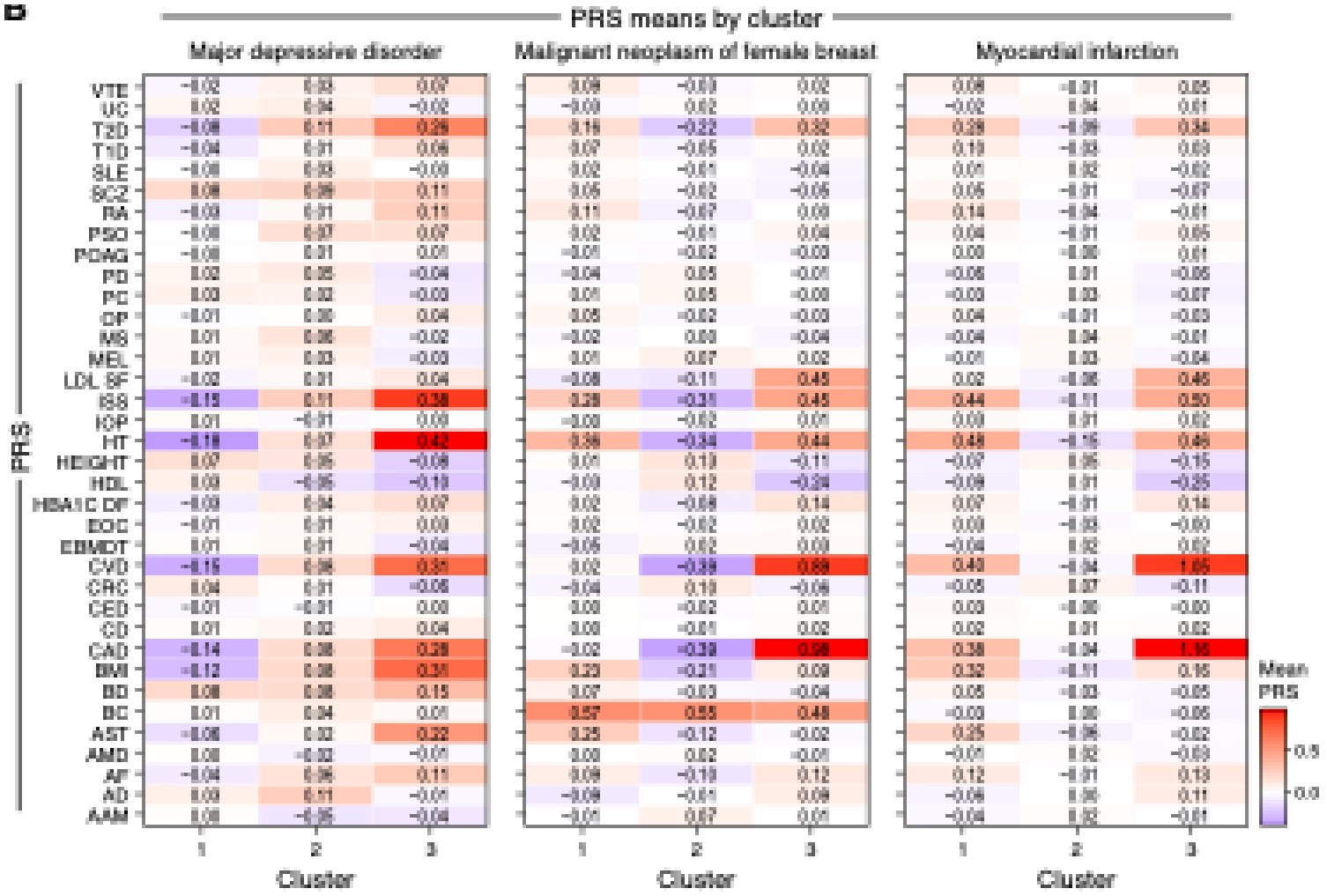


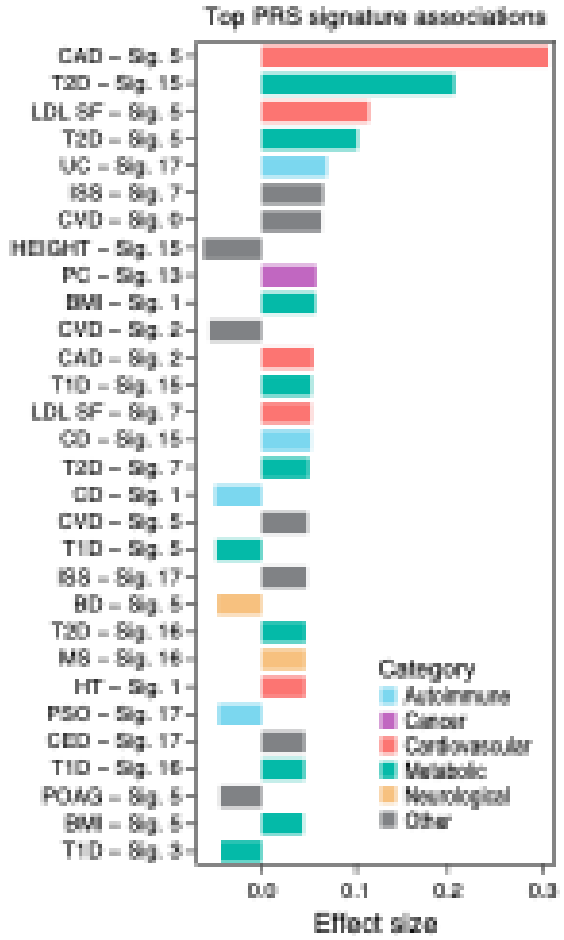
Heterogeneity within disease: Revealing biology



- *Ference et al., NEJM 2012*: Genetic variants that lower cholesterol even moderately over lifetime dramatically reduce CAD
- *Khera et al., Nat Genet 2018*: PRS identifies high-risk individuals missed by traditional screening

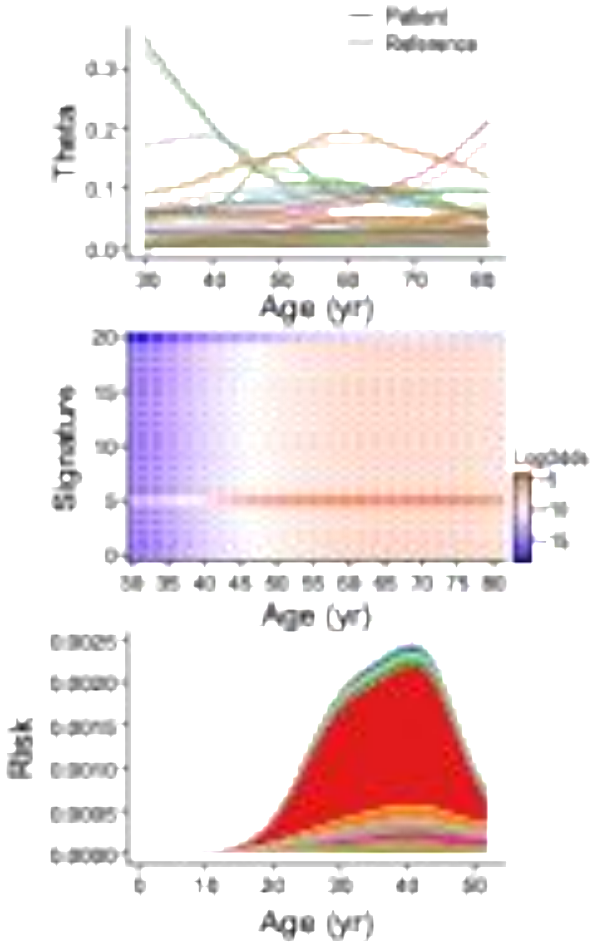
Heterogeneity in phenotype matches Genotype

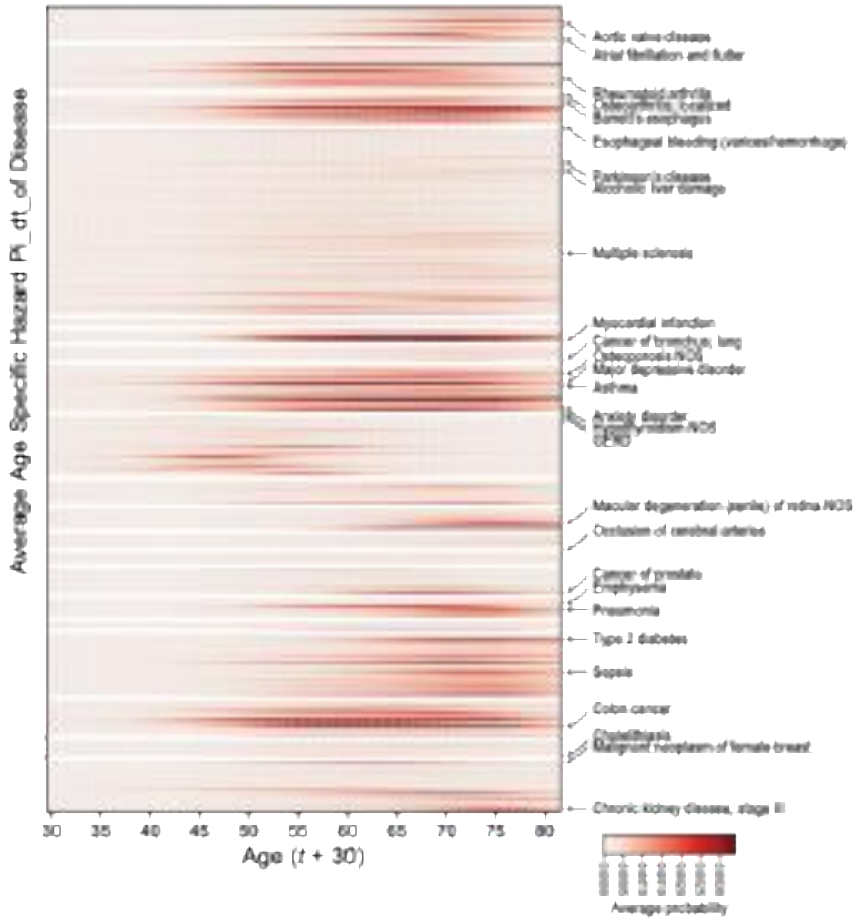




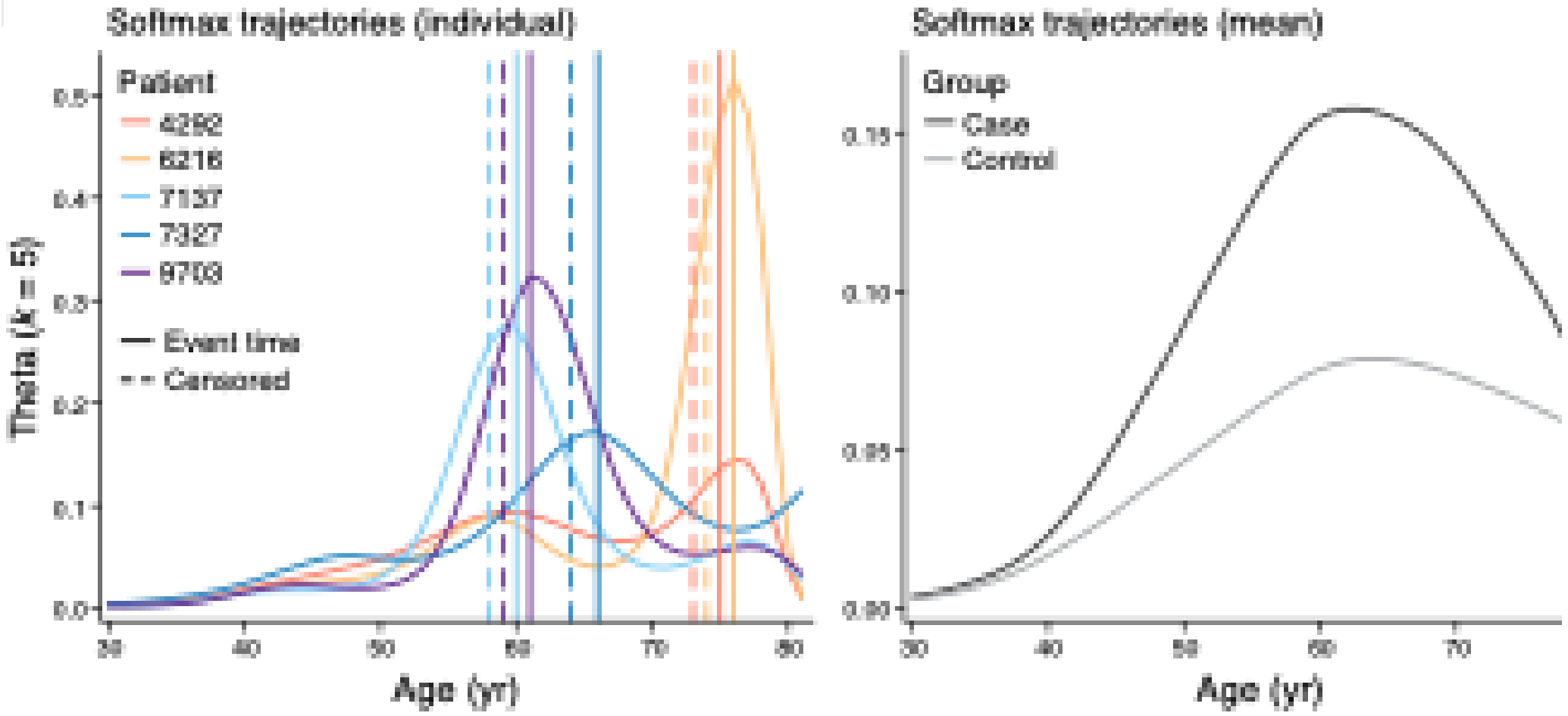
Making predictions is only part …

$$\pi_{idt} = \sum_{k} f(\lambda_{idt}) f(\phi_{idt})$$

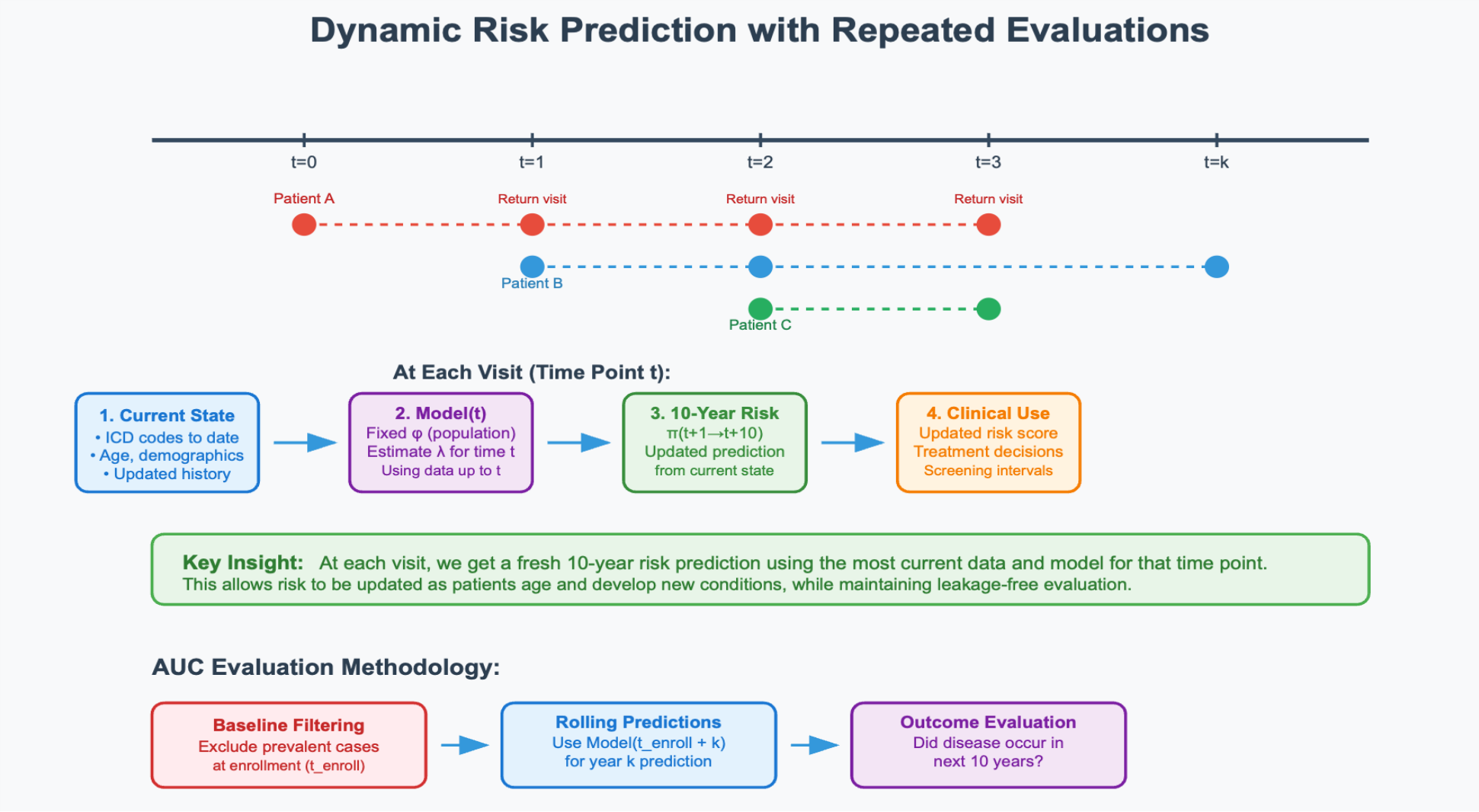


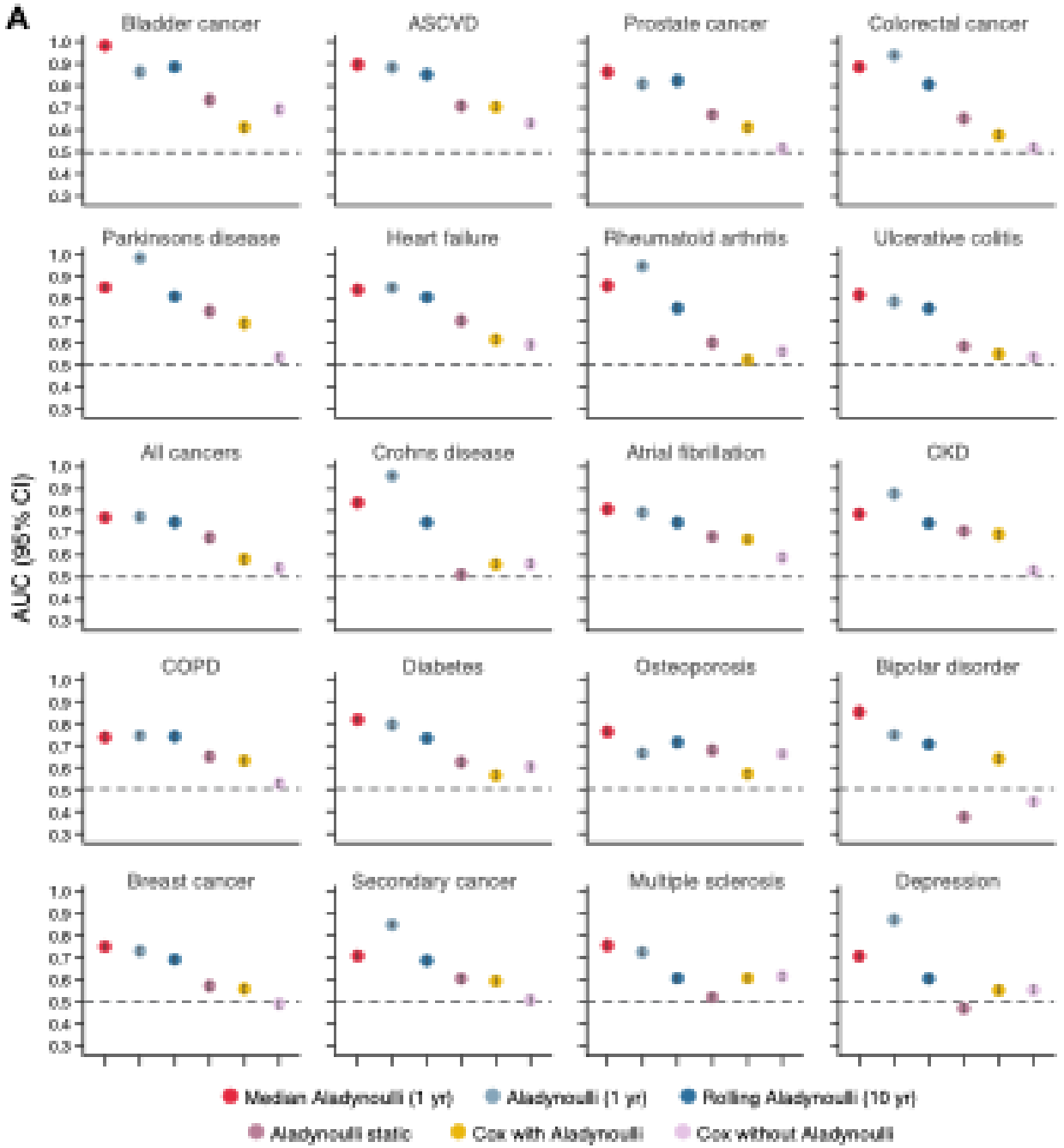


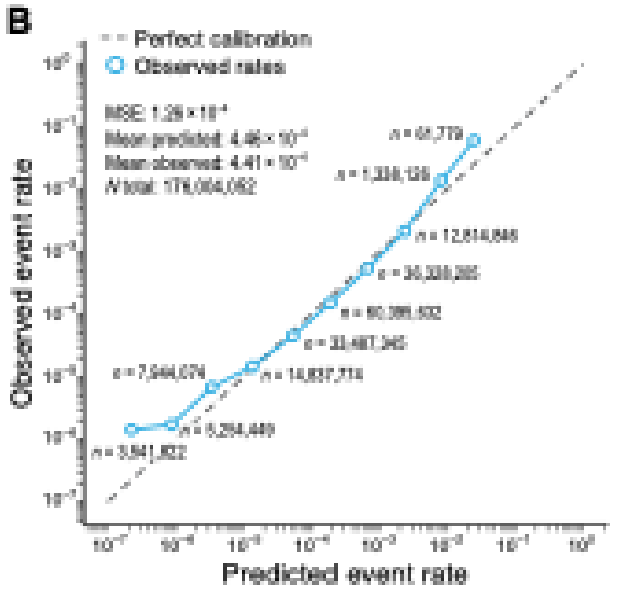
Aladynoulli: the genie works with a blindfold!

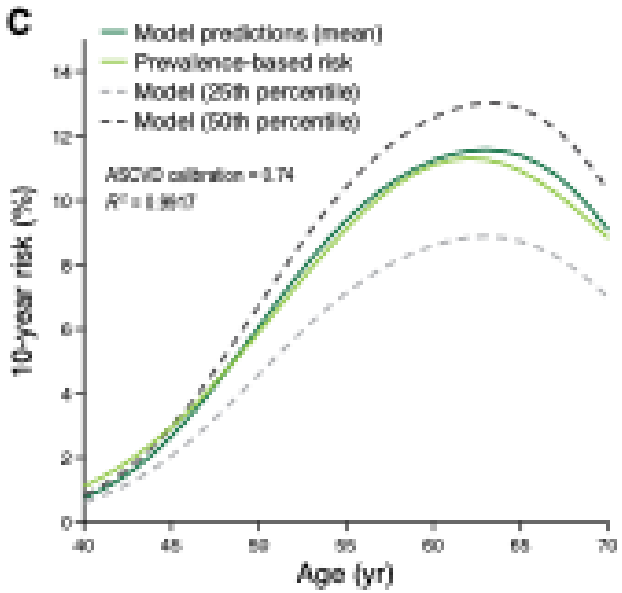


Performance assessment: Changing the paradigm



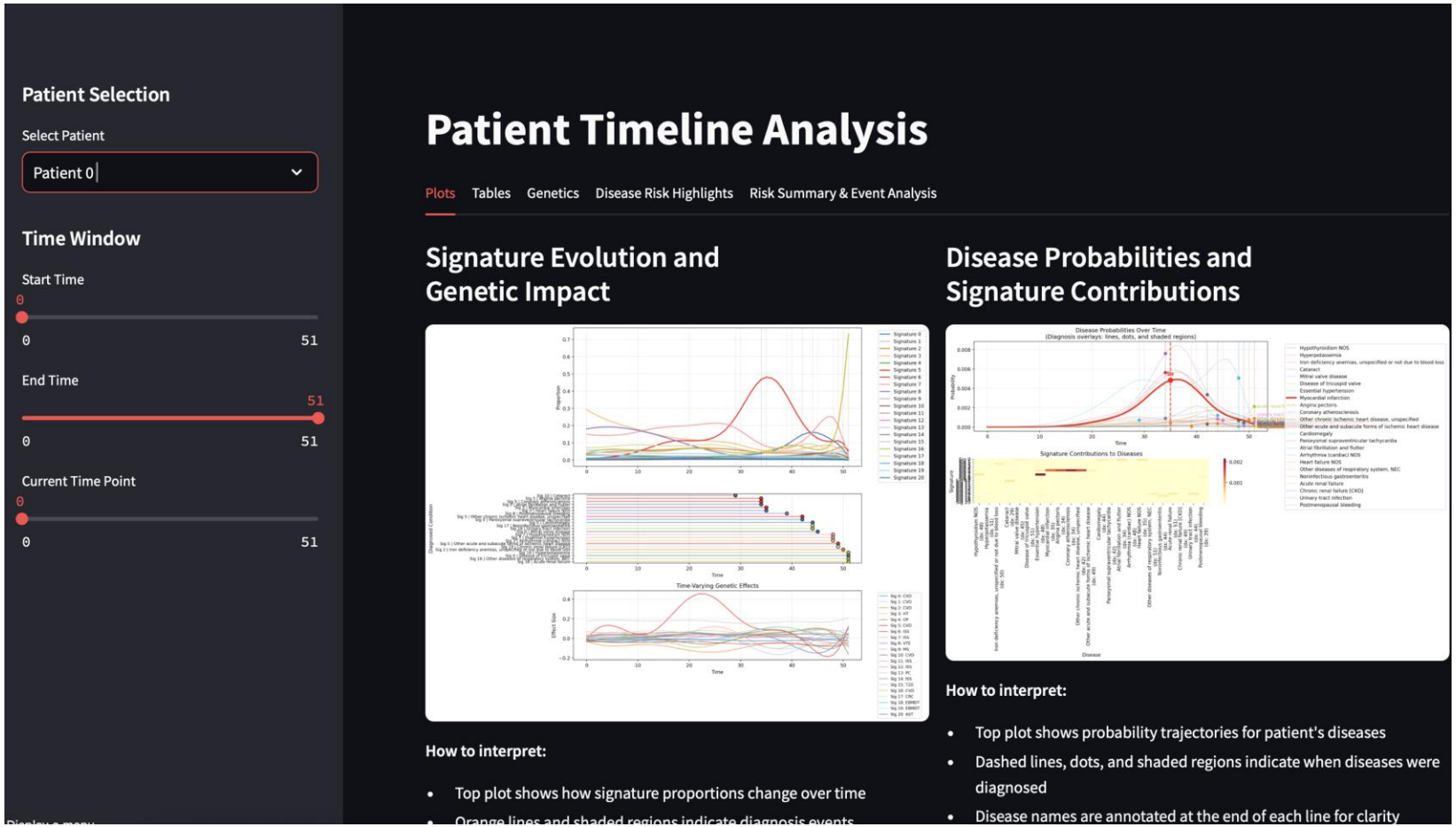






PREVENT: AUC 0.65
Aladynoulli median dynamic 0.901

What does Dynamic Look like?



http://aladynoulli.hms.Harvard.edu

Patient Selection

Select Patient

Patient 536

Time Window

Start Time

0 51

End Time

51

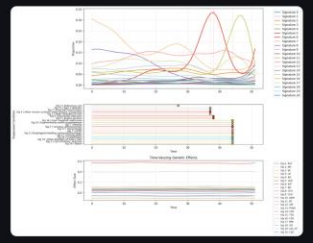
0 51

Deploy

Patient Timeline Analysis

Plots Tables Genetics Disease Risk Highlights Risk Summary & Event Analysis

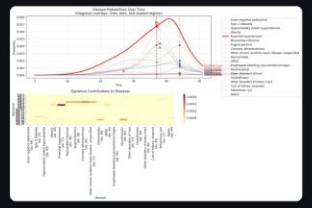
Signature Evolution and Genetic Impact



How to interpret:

- Top plot shows how signature proportions change over time

Disease Probabilities and Signature Contributions



How to interpret:

- Top plot shows probability trajectories for pat diseases
- Dashed lines, dots, and shaded regions indica diseases were diagnosed

Display a menu

ent Selection

Patient

ient 536

e Window

Time

51

ime

51

51

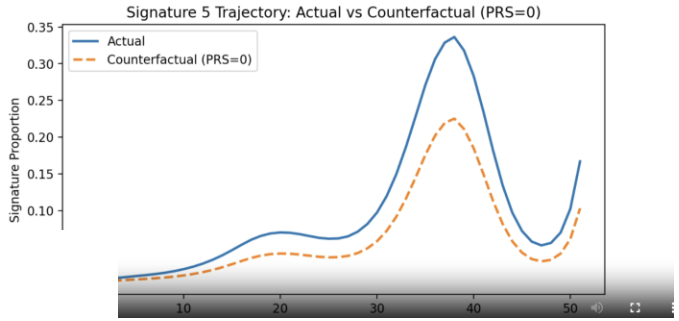
nt Time Point

Patient Index

536

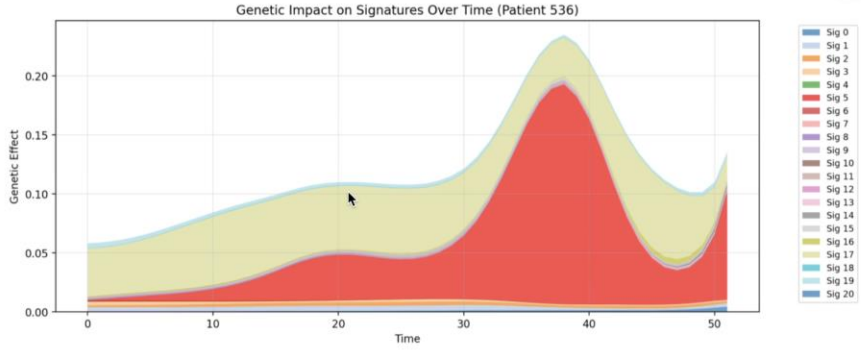
Signature Index

5



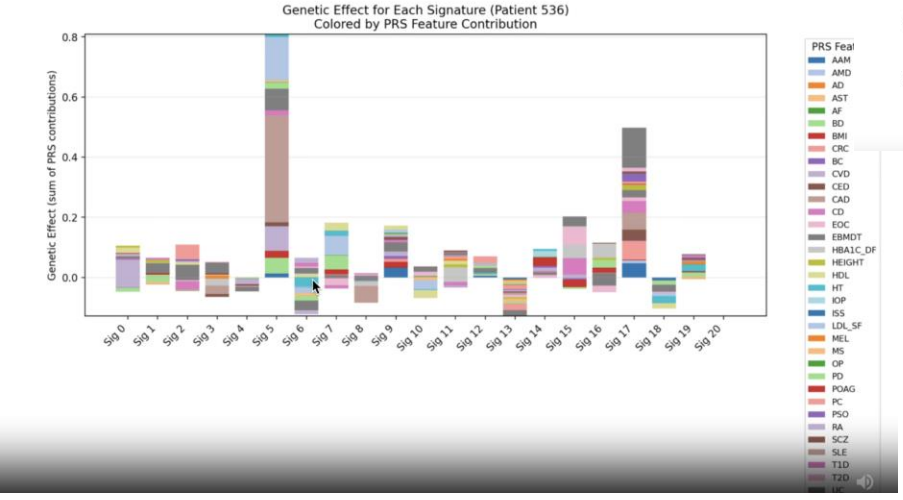
Genetic Impact on Signatures Over Time (Direct Weighting, Selected Patient)

This plot shows the direct genetic impact on each signature over time for the currently selected patient, using their PRS and the model's gamma weights.

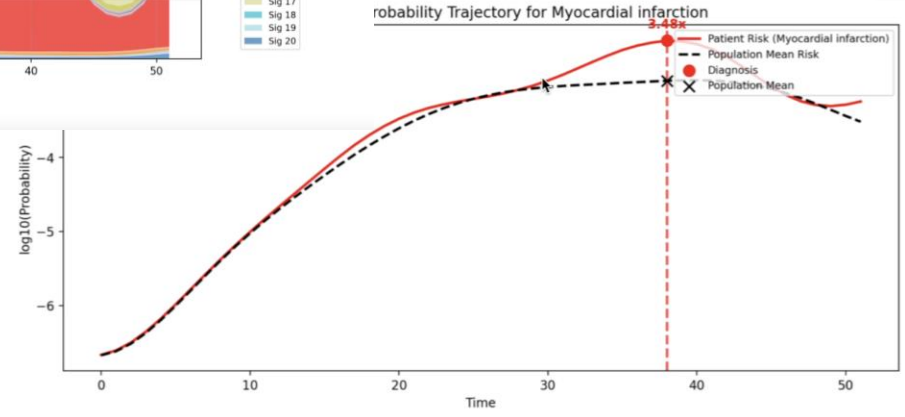


Patient, Stacked by PRS Feature)

This stacked barplot shows the overall genetic effect for each signature for the currently selected patie with each bar colored by the fractional contribution of each PRS feature (G*gamma).



[38]



Fold Enrichment of Disease Probability at Diagnosis

Discussion Points

- Tradeoffs of models where its highly precise, but difficult to communicate or models where it could potentially operate entirely behind the scenes and integrate into the EHR
  - Methods vs BlackBox

Figure 1: A fictional depiction of the "accuracy-interpretability trade-off," taken from the DARPA XAI (Explainable Artificial Intelligence) Broad Agency Announcement [18].

Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead

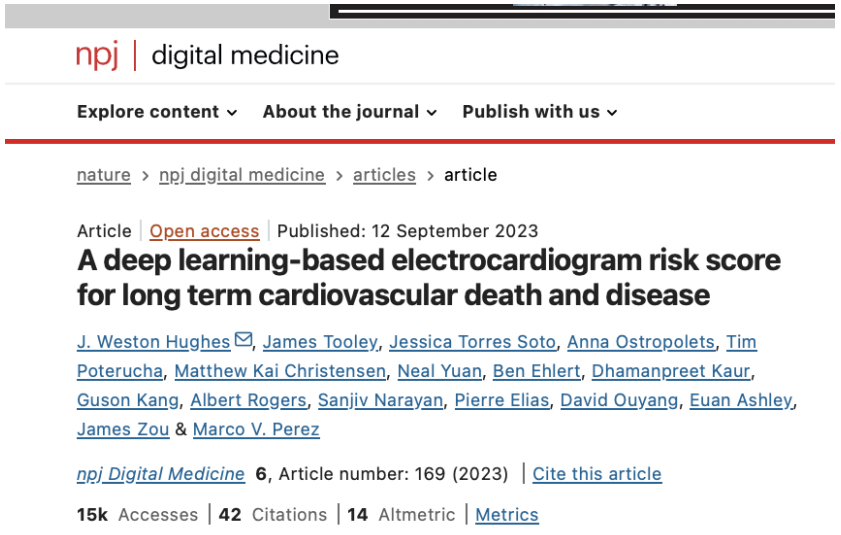
Cynthia Rudin
Duke University
cynthia@cs.duke.edu

Abstract

Implementing Machine Learning in Health Care — Addressing Ethical Challenges

Danton S. Char, M.D., Nigam H. Shah, M.B., B.S., Ph.D., and **David Magnus, Ph.D.**
Department of Anesthesiology, Division of Pediatric Cardiac Anesthesia (D.S.C.), the Center for Biomedical Ethics (D.S.C., D.M.), and the Center for Biomedical Informatics Research (N.S.), Stanford University School of Medicine, Stanford, CA

- What would be required to have a model that includes genomics, opportunistic imaging like CAC/CT-coronary, and AI image processing (like subtle signals on ECG and TTE)



npj | digital medicine

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › npj digital medicine › articles › article

Article | Open access | Published: 12 September 2023

A deep learning-based electrocardiogram risk score for long term cardiovascular death and disease

J. Weston Hughes ✉, James Tooley, Jessica Torres Soto, Anna Ostropolets, Tim Poterucha, Matthew Kai Christensen, Neal Yuan, Ben Ehlert, Dhamanpreet Kaur, Guson Kang, Albert Rogers, Sanjiv Narayan, Pierre Elias, David Ouyang, Euan Ashley, James Zou & Marco V. Perez

npj Digital Medicine **6**, Article number: 169 (2023) | Cite this article

15k Accesses | **42** Citations | **14** Altmetric | Metrics

Opinion

EDITORIAL

AI IN MEDICINE

AI in Medicine—*JAMA*'s Focus on Clinical Outcomes, Patient-Centered Care, Quality, and Equity

Rohan Khera, MD, MS; Atul J. Butte, MD, PhD; Michael Berkwits, MD, MSCE; Yulin Hswen, ScD, MPH; Annette Flanagin, RN, MA; Hannah Park; Gregory Curfman, MD; Kirsten Bibbins-Domingo, PhD, MD, MAS

Thank you!!!

Pradeep Natarajan, MD MMSc
Romit Bhattacharya,MD,
Giovanni Parmigiani, PhD
Sasha Gusev, PhD