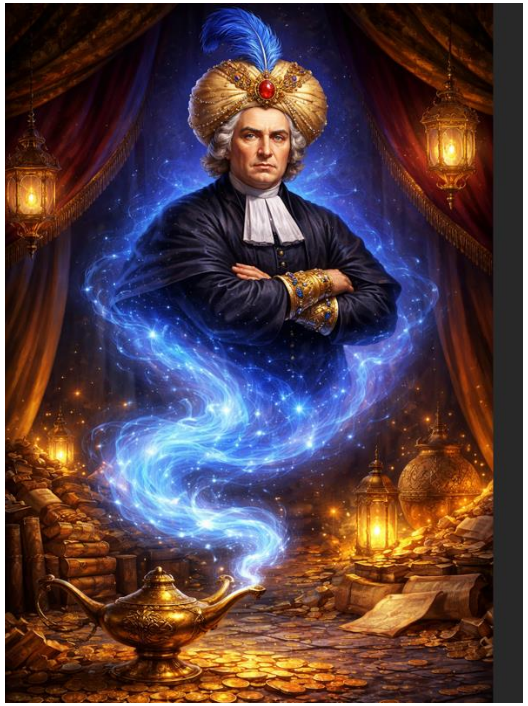
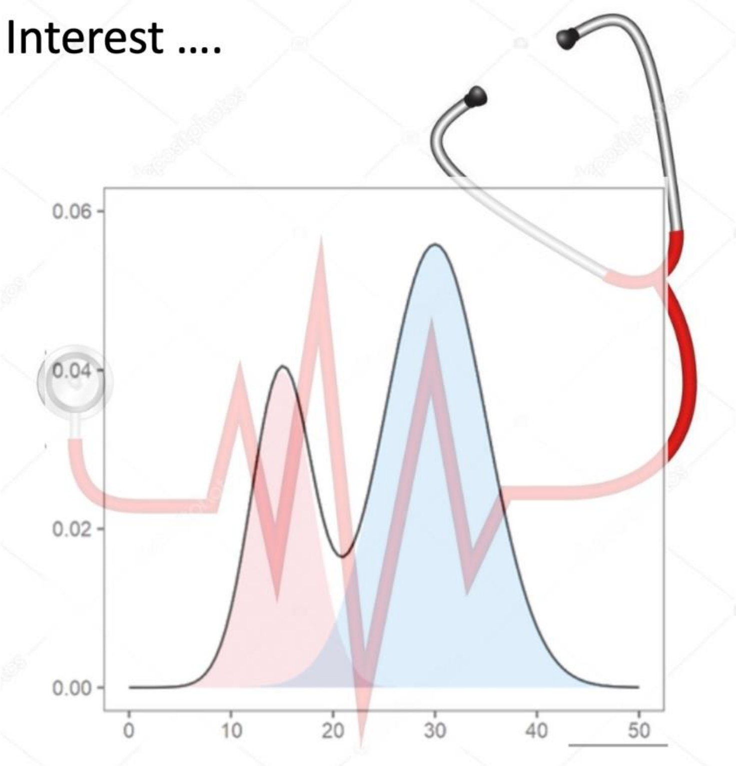
Aladynoulli: A bayesian voyage through the genome and the EHR.

Sarah Urbut, MD PhD
Natarajan Lab
5 February 2025



Disclosures
No disclosures to report

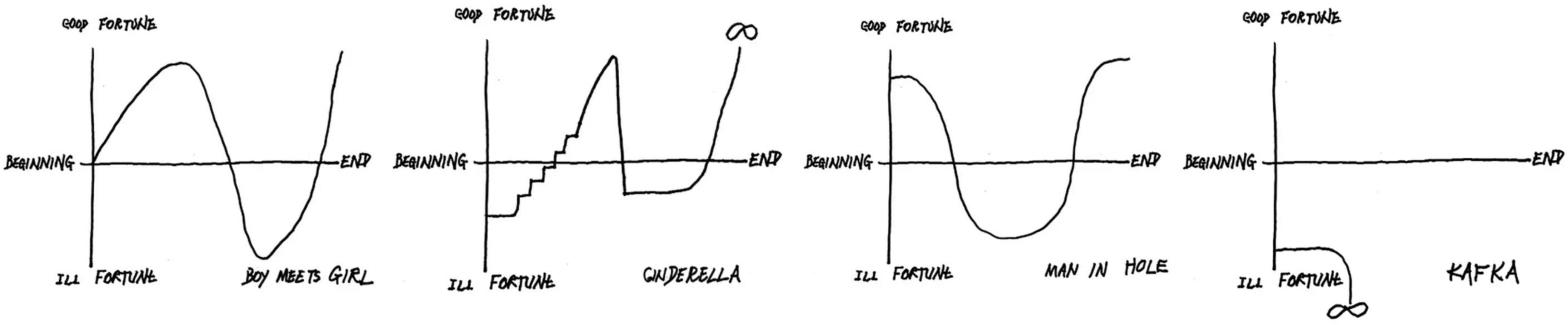
Conflicts of Interest ….



BROAD INSTITUTE

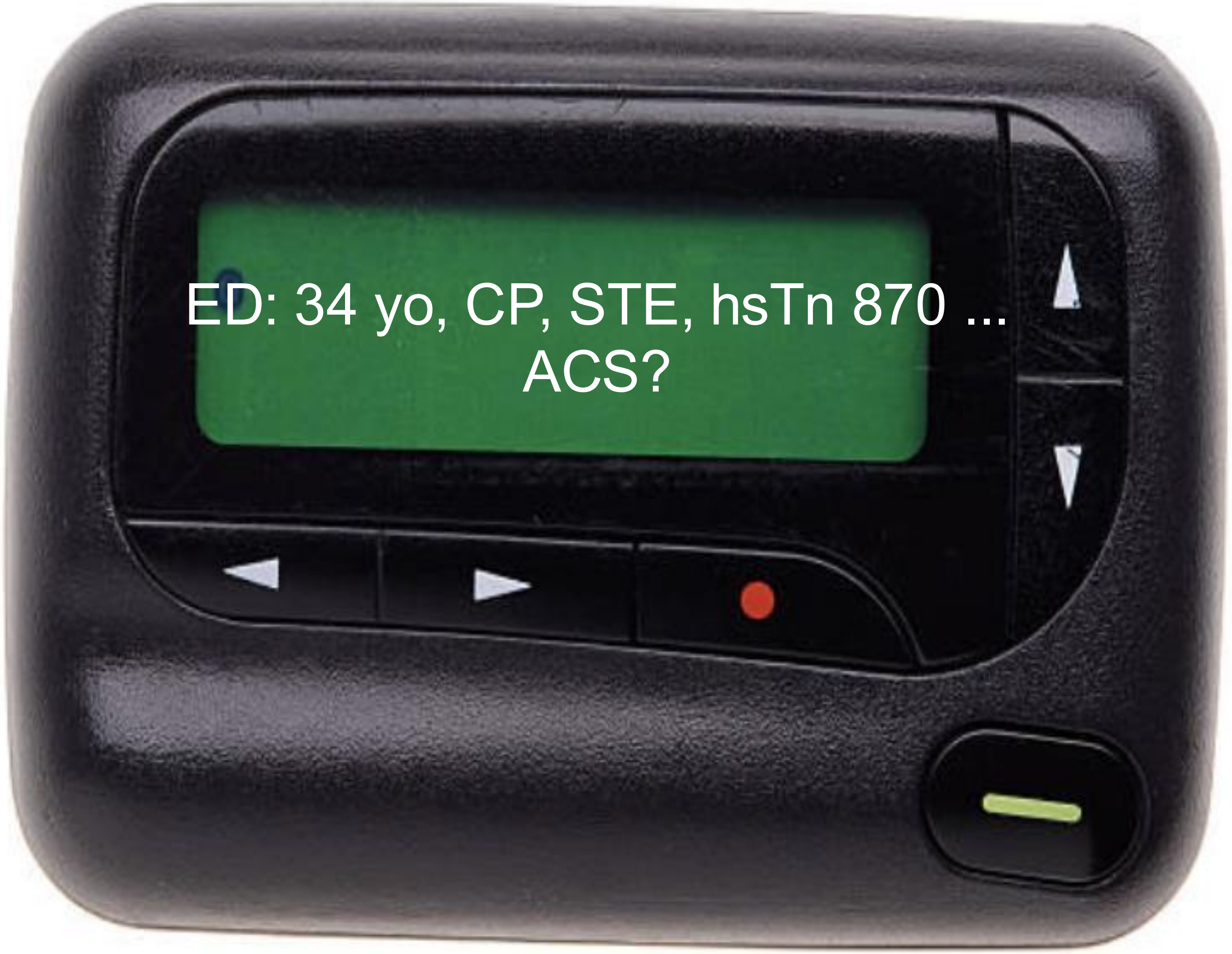
MASSACHUSETTS GENERAL HOSPITAL
CORRIGAN MINEHAN HEART CENTER

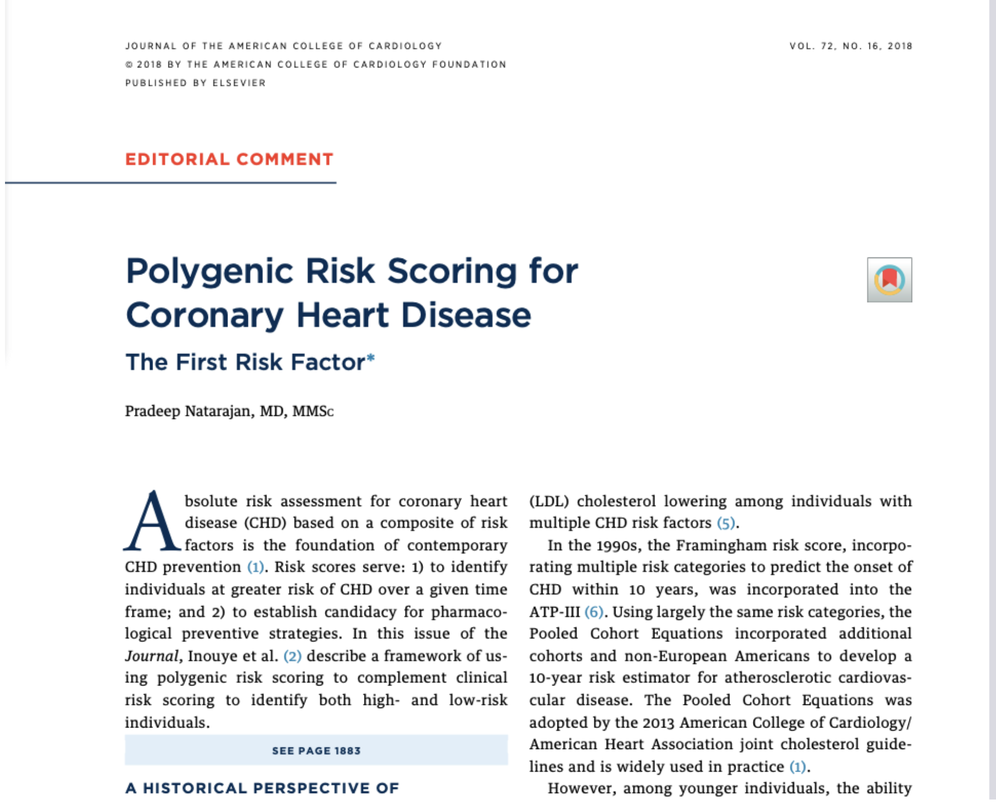
Shape of a story



Vonnegut, 1955

Everything is moving





JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY
VOL. 72, NO. 16, 2018

PUBLISHED BY ELSEVIER

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

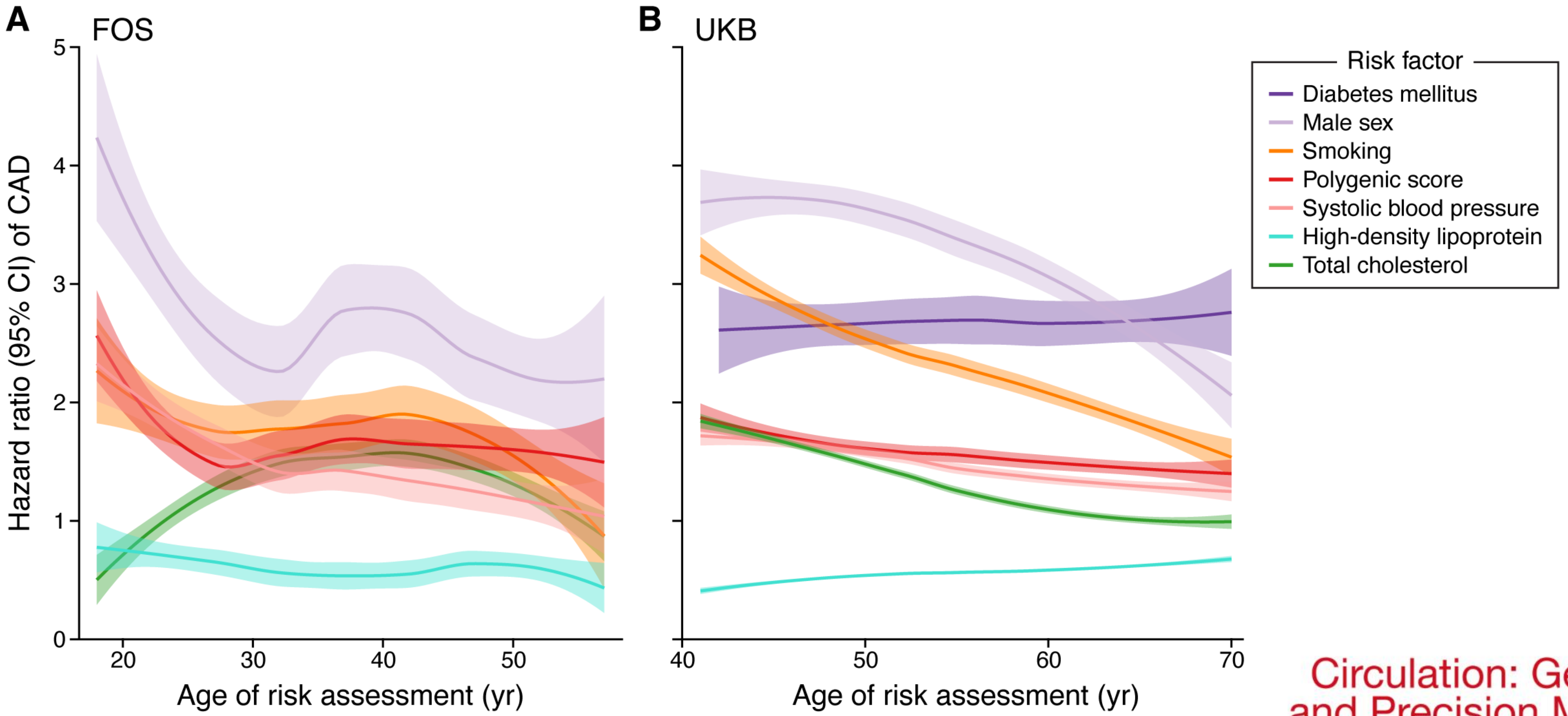
A HISTORICAL PERSPECTIVE OF

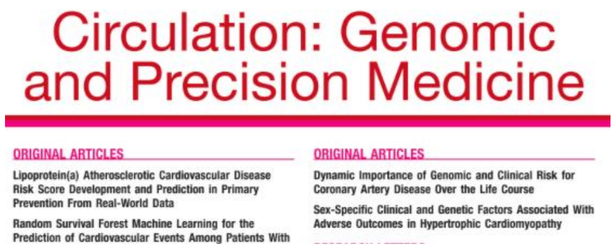
(LDL) cholesterol lowering among individuals with multiple CHD risk factors (5).

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




Circulation: Genomic and Precision Medicine

ORIGINAL ARTICLES

Lipoprotein(a) Atherosclerotic Cardiovascular Disease Risk Score Development and Prediction in Primary Prevention From Real-World Data

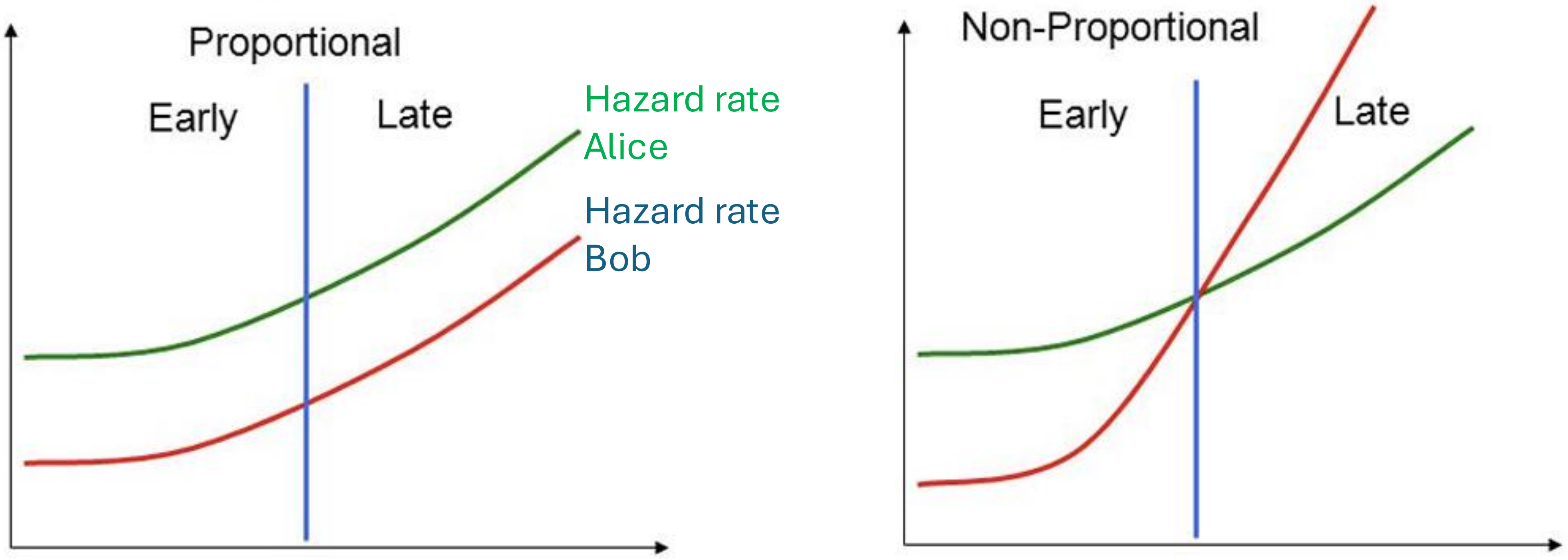
Random Survival Forest Machine Learning for the Prediction of Cardiovascular Events Among Patients With

ORIGINAL ARTICLES

Dynamic Importance of Genomic and Clinical Risk for Coronary Artery Disease Over the Life Course

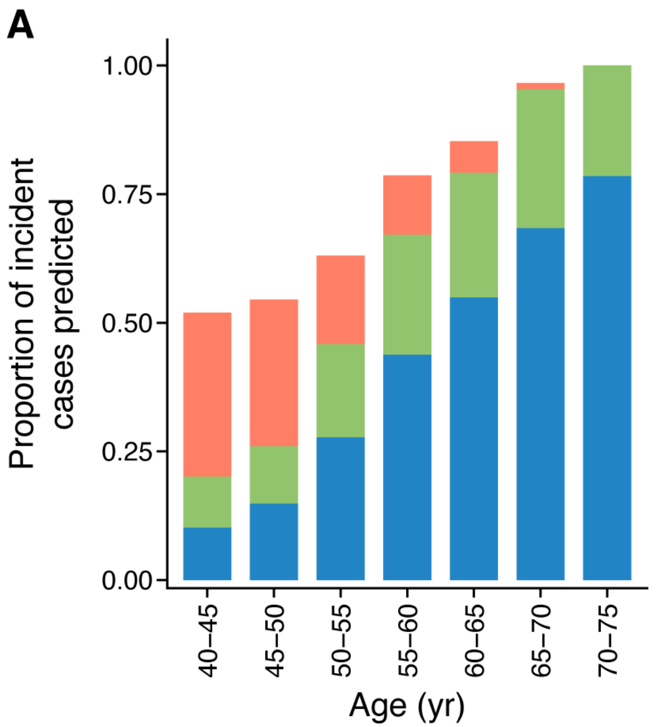
Sex-Specific Clinical and Genetic Factors Associated With Adverse Outcomes in Hypertrophic Cardiomyopathy

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

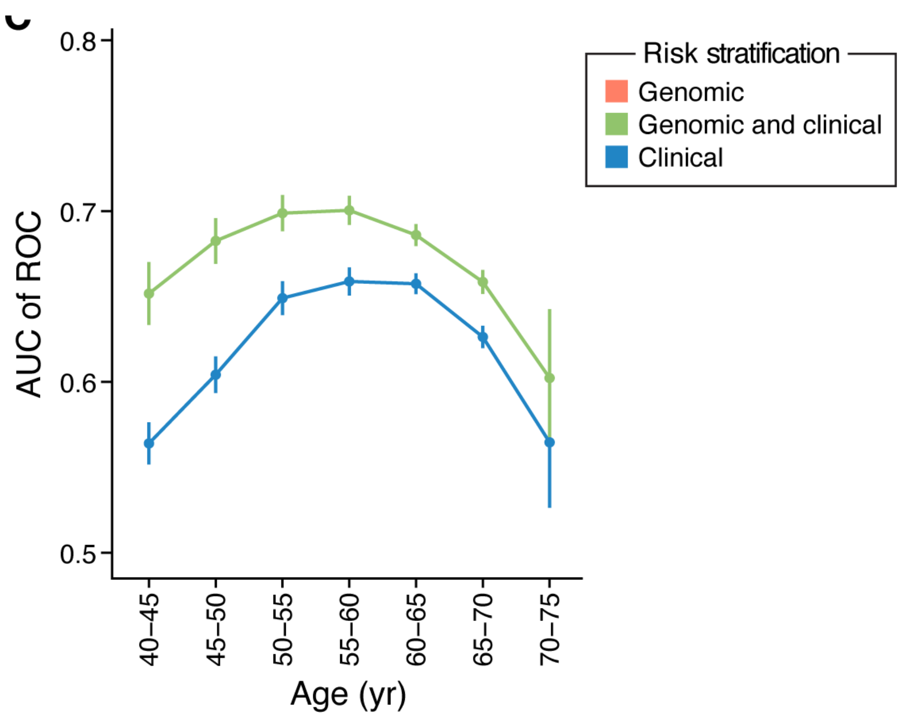


The Cox Proportional Hazards model assumes the hazards are ***proportional***: the ***relative*** hazard ratio remains ***constant over time***

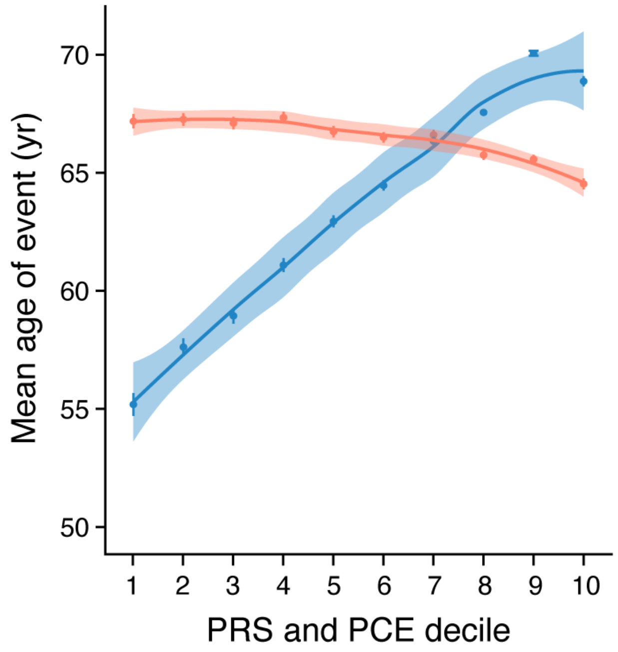
Predicts not only early disease, which we sometimes think of, but also all disease early



POWER

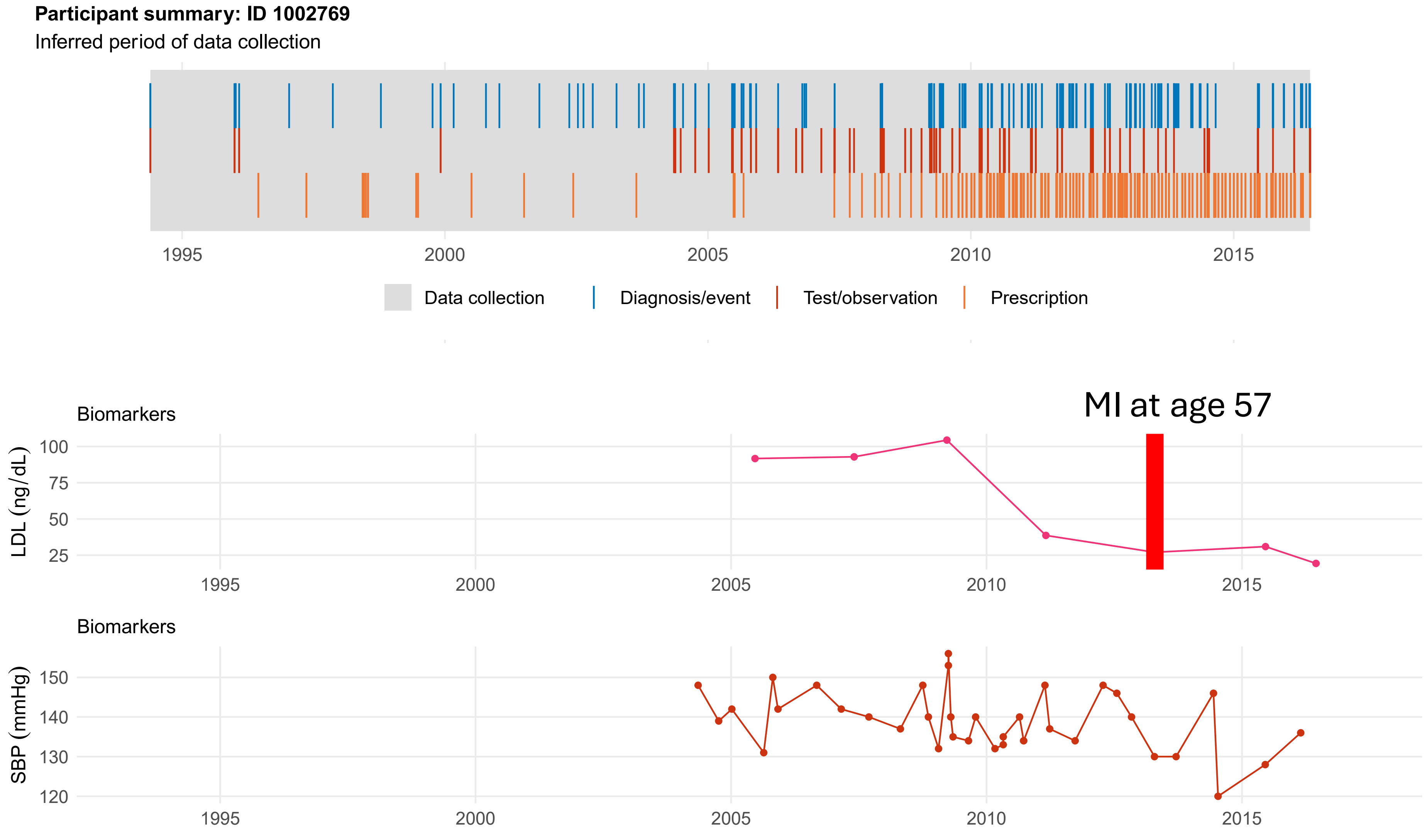


DISCRIMINATION

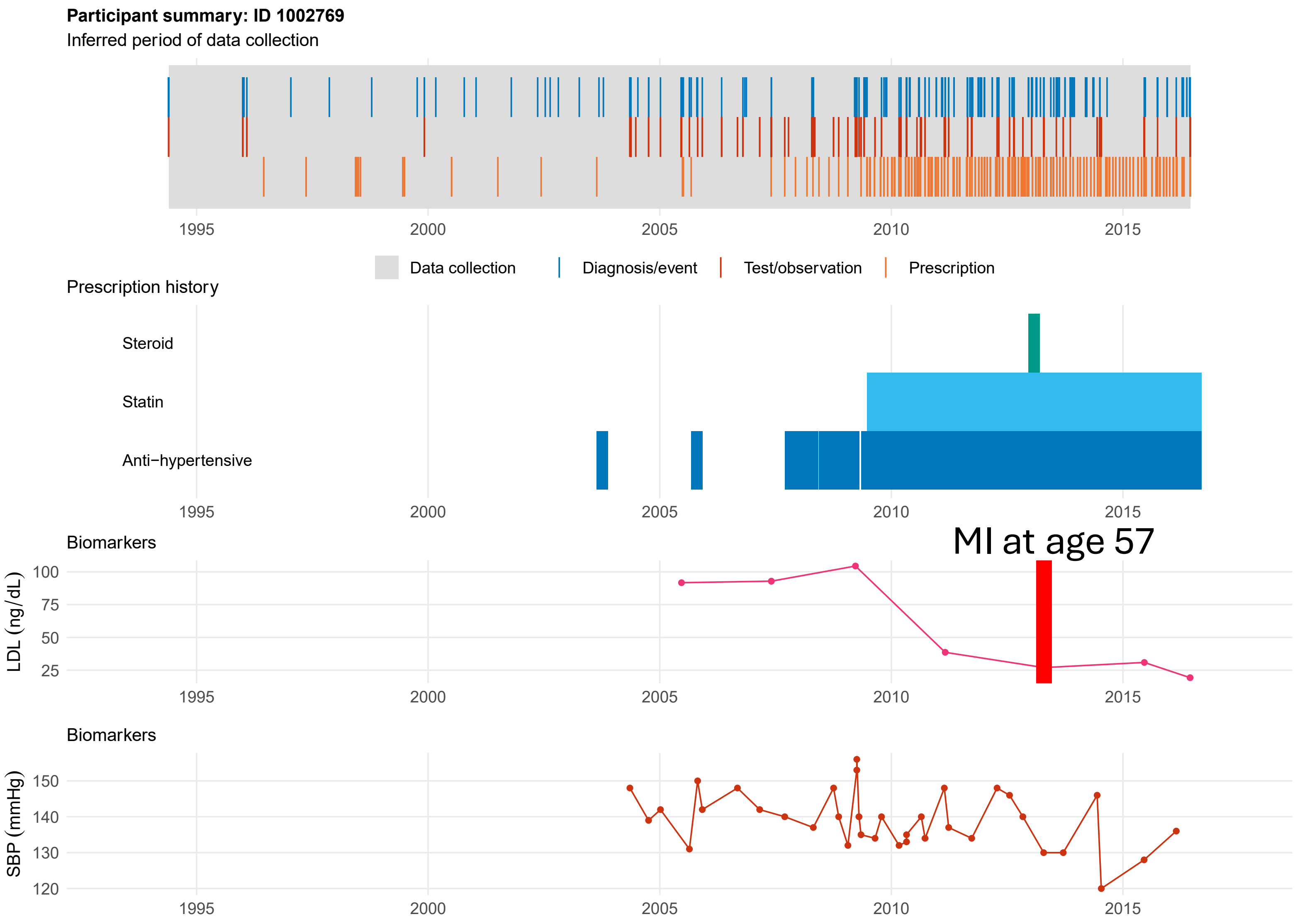


Urbut et al, Circulation GPM 2025

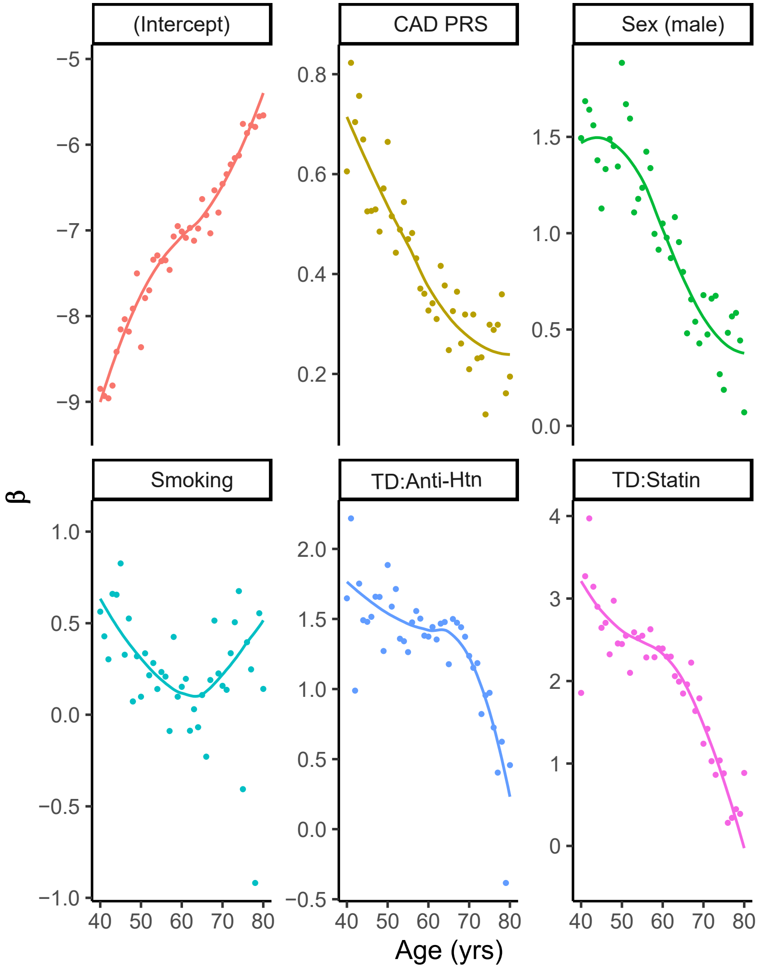
Using the EHR to be predictive, not responsive

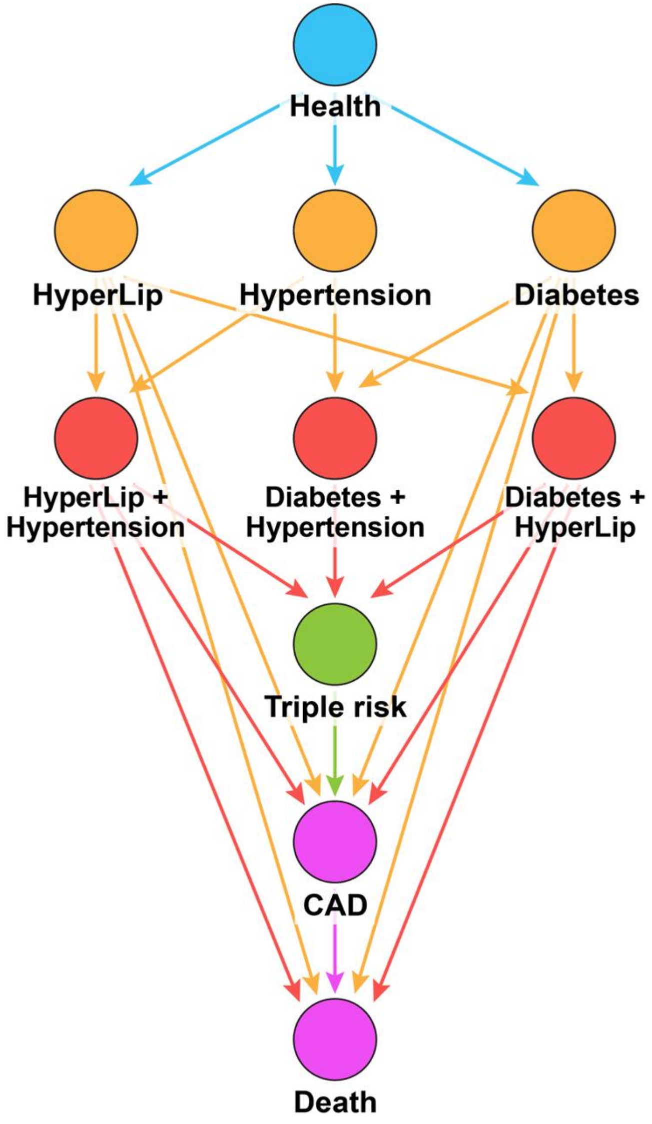


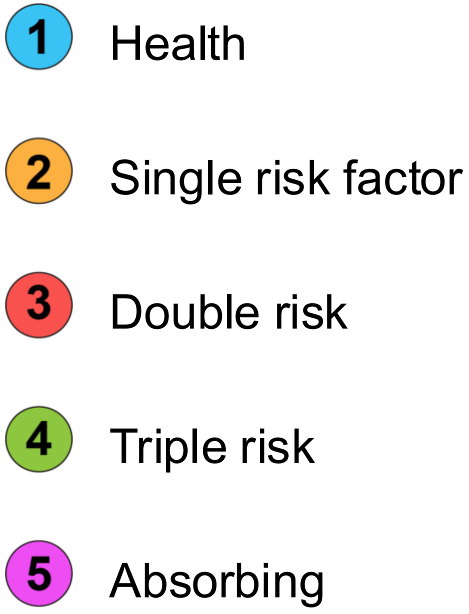
Using the EHR to be predictive, not responsive

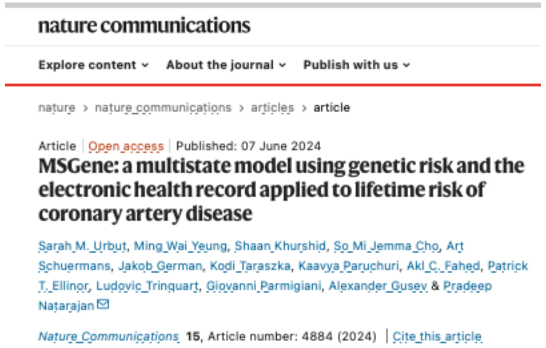


MSGene: multistate model using genetics for dynamic prediction









nature communications

Explore content About the journal Publish with us

nature › nature communications › articles › article

Article | Open access | Published: 07 June 2024

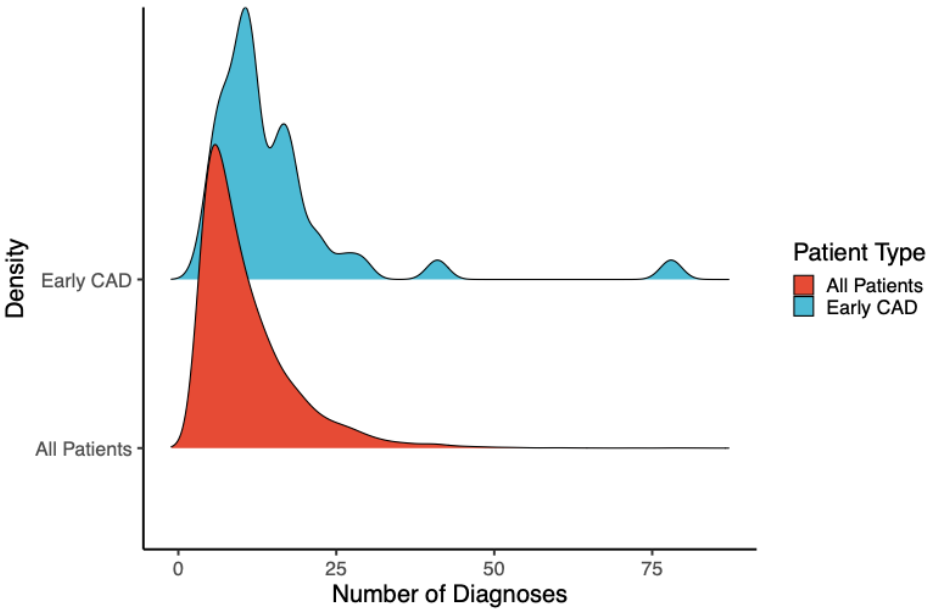
MSGene: a multistate model using genetic risk and the electronic health record applied to lifetime risk of coronary artery disease

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

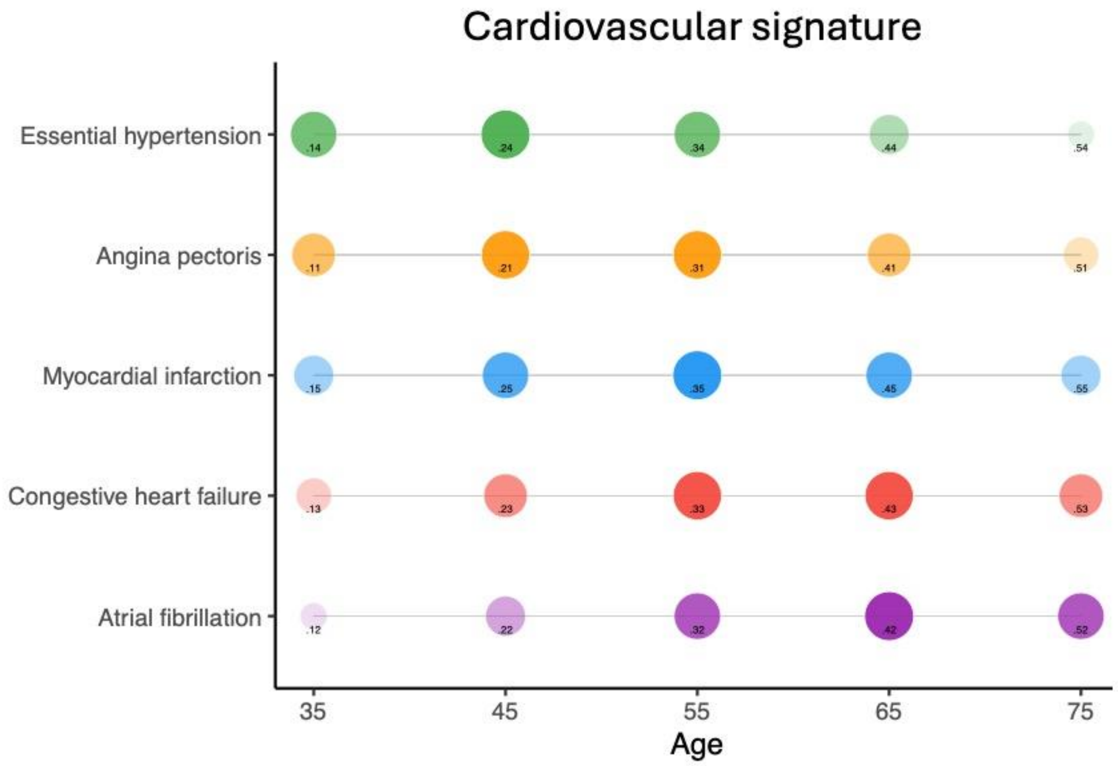
Aladyn Individual: Bayesian Hierarchical Dynamic Genetic Modeling of Comorbidity Progression

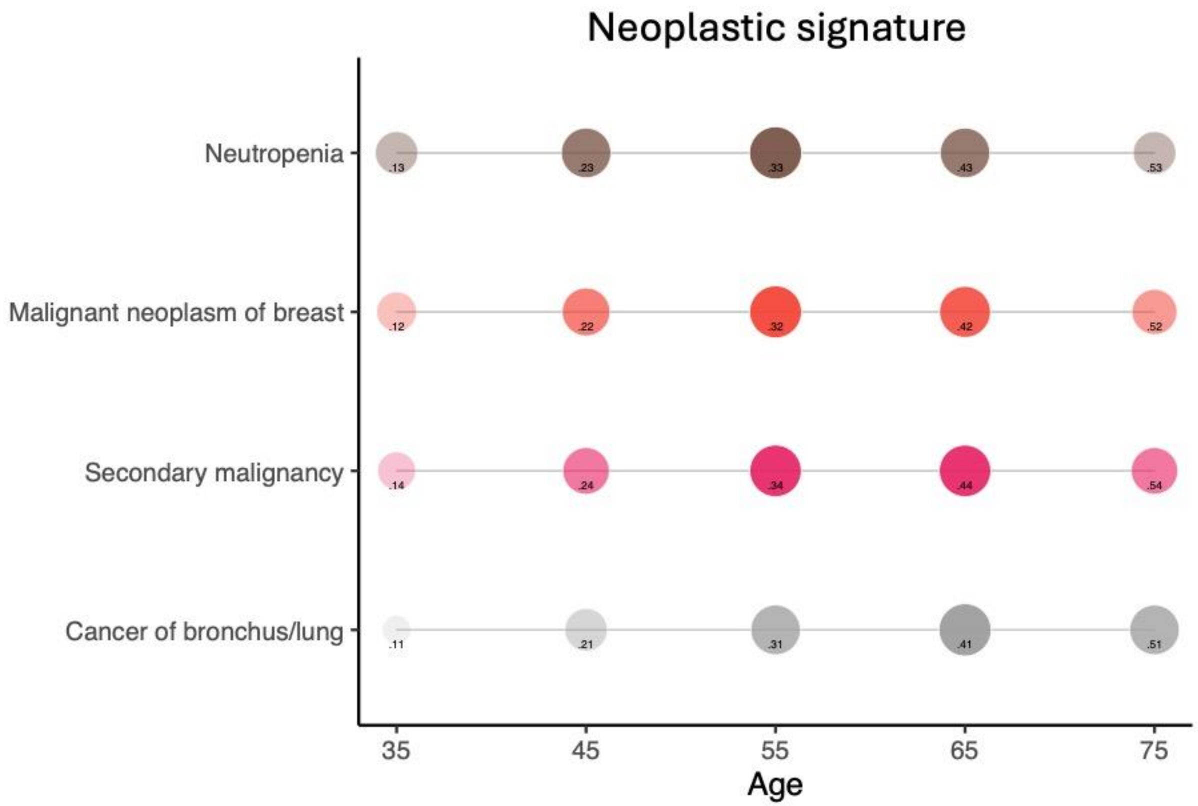
Sarah Urbut [1,2], Yi Ding [3], Xilin Jiang [2,4,5], Whitney Hornsby [1,2], Alexander Gusev [2,3], Pradeep Natarajan [1,2], Giovanni Parmigiani [3,5]

Affiliations + expand

PMID: 39568791 PMCID: PMC11577253 DOI: 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

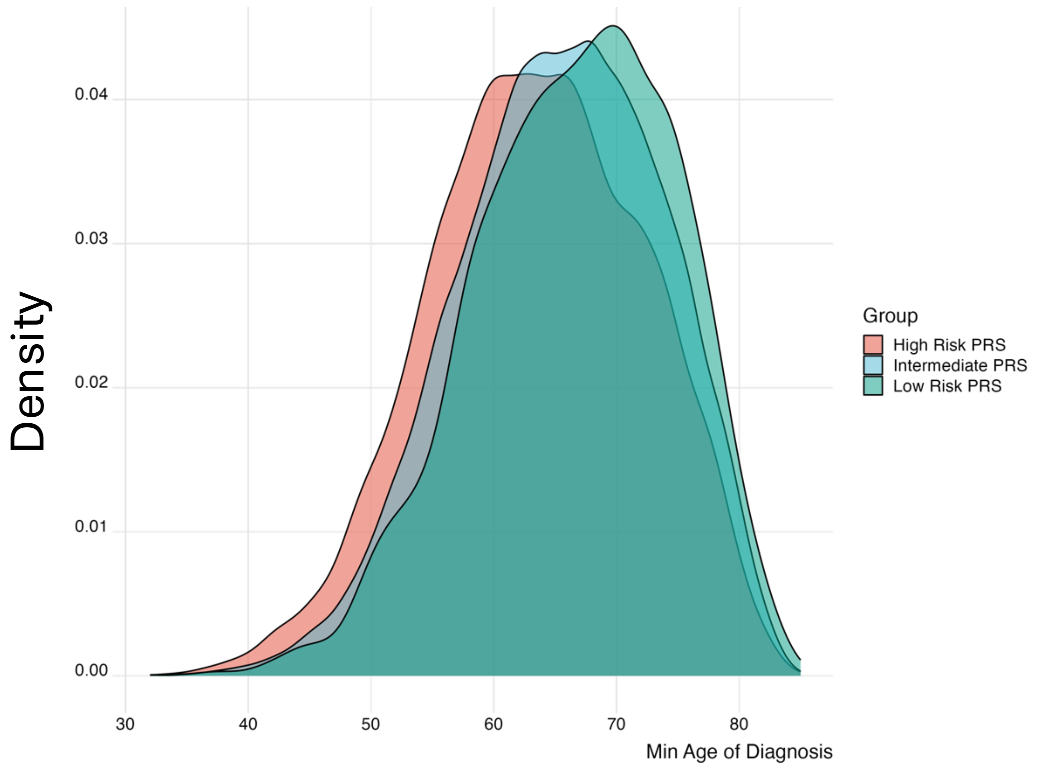
Condition	Age 35	Age 40	Age 45	Age 50	Age 55	Age 60	Age 65
CAD	1	1	1	1	1	1	1
Heart Failure	0	1	1	1	1	1	1
GERD	0	0	1	1	1	1	1
IBD	0	0	0	1	1	1	1
Type 2 Diabetes	0	0	0	0	1	1	1
Obesity	0	0	0	0	1	1	1

Metabolic CVD Cancer Inflammatory GI Neuro

GENETICS Matters

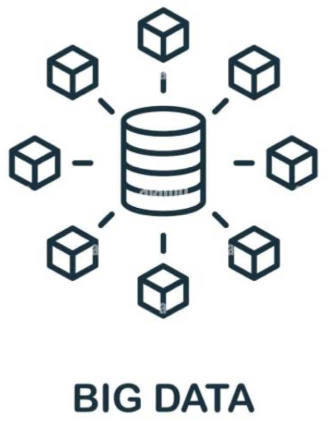
Genetics impacts early disease and all disease, early

Coronary Artery Disease
Diagnosis by PRS



Identity crisis

Joint consideration: discovery and prediction



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

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*:

$$\pi_{idt} = \sum_k \underbrace{f(\lambda_{idt})}_{\text{Individual}} \underbrace{f(\phi_{idt})}_{\text{Population}}$$

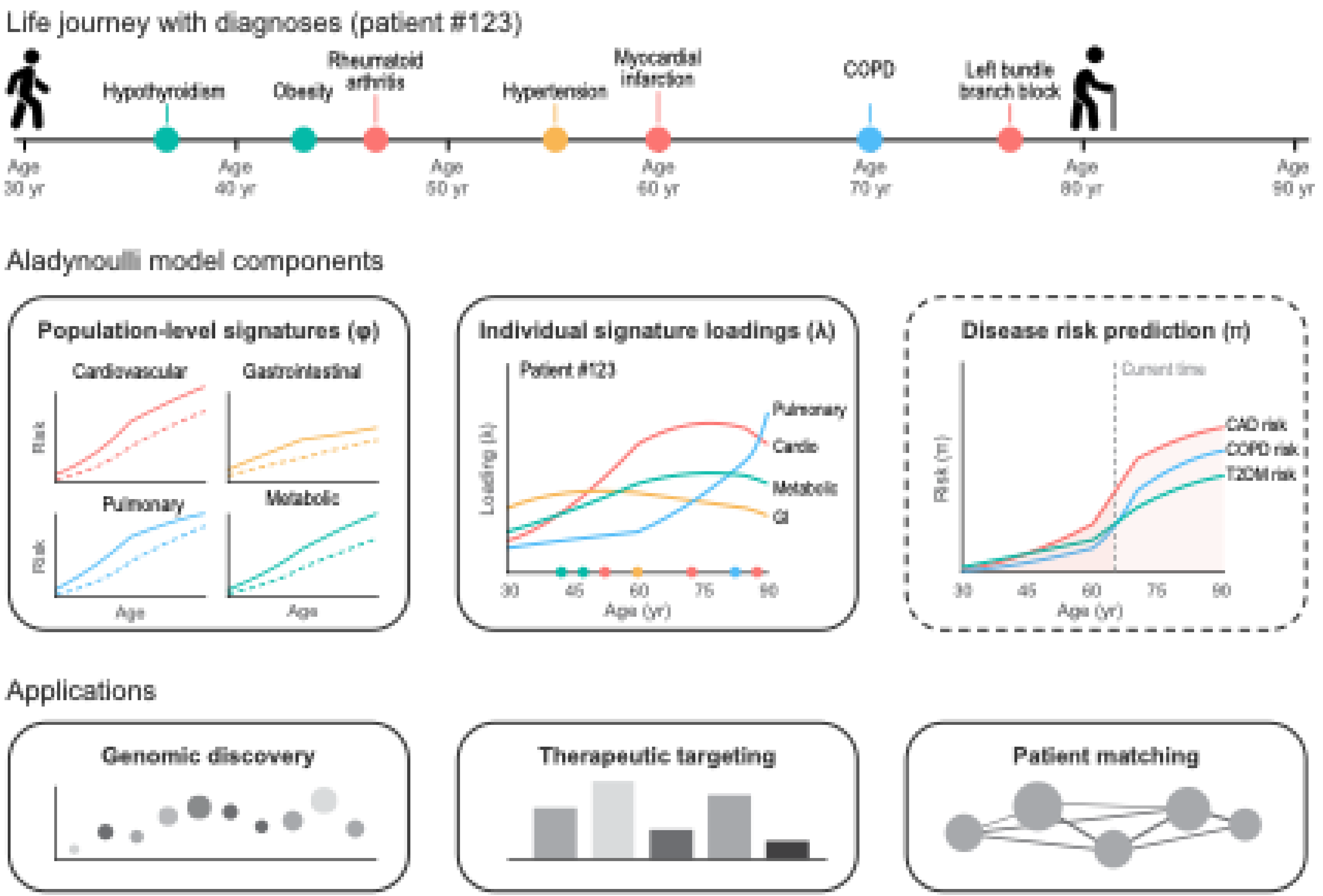
We are all Bayesians

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

At risk Event Censored

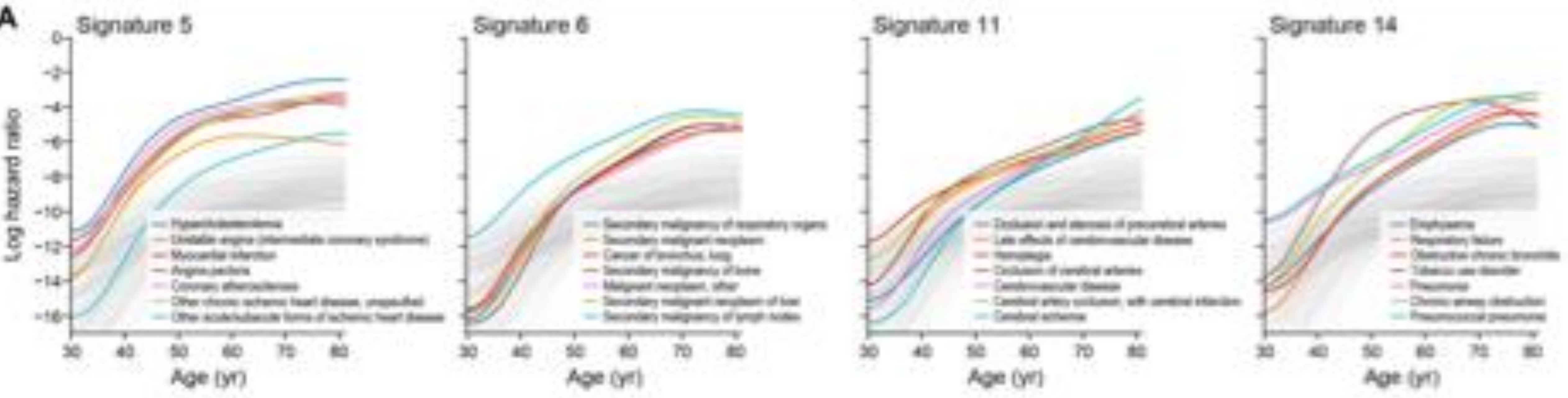
$$\text{Post} \propto Likelihood \times Prior$$

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

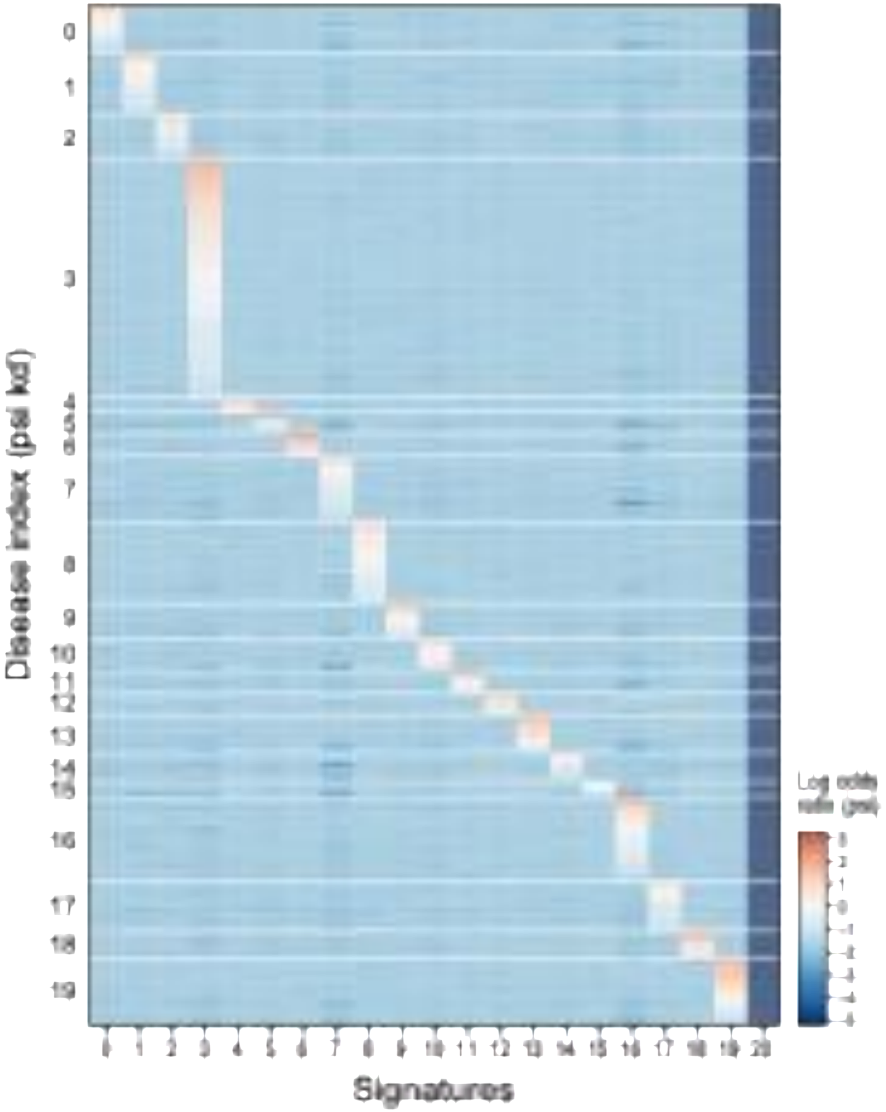


Life journey with diagnoses (patient #123)
Hypothyroidism
Obesity
Rheumatoid arthritis
Hypertension
Myocardial infarction
COPD
Left bundle branch block
Age 30 yr
Age 40 yr
Age 50 yr
Age 60 yr
Age 70 yr
Age 80 yr
Age 90 yr
Aladynoulli model components
Population-level signatures (φ)
Cardiovascular
Gastrointestinal
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
Applications
Genomic discovery
Therapeutic targeting
Patient matching

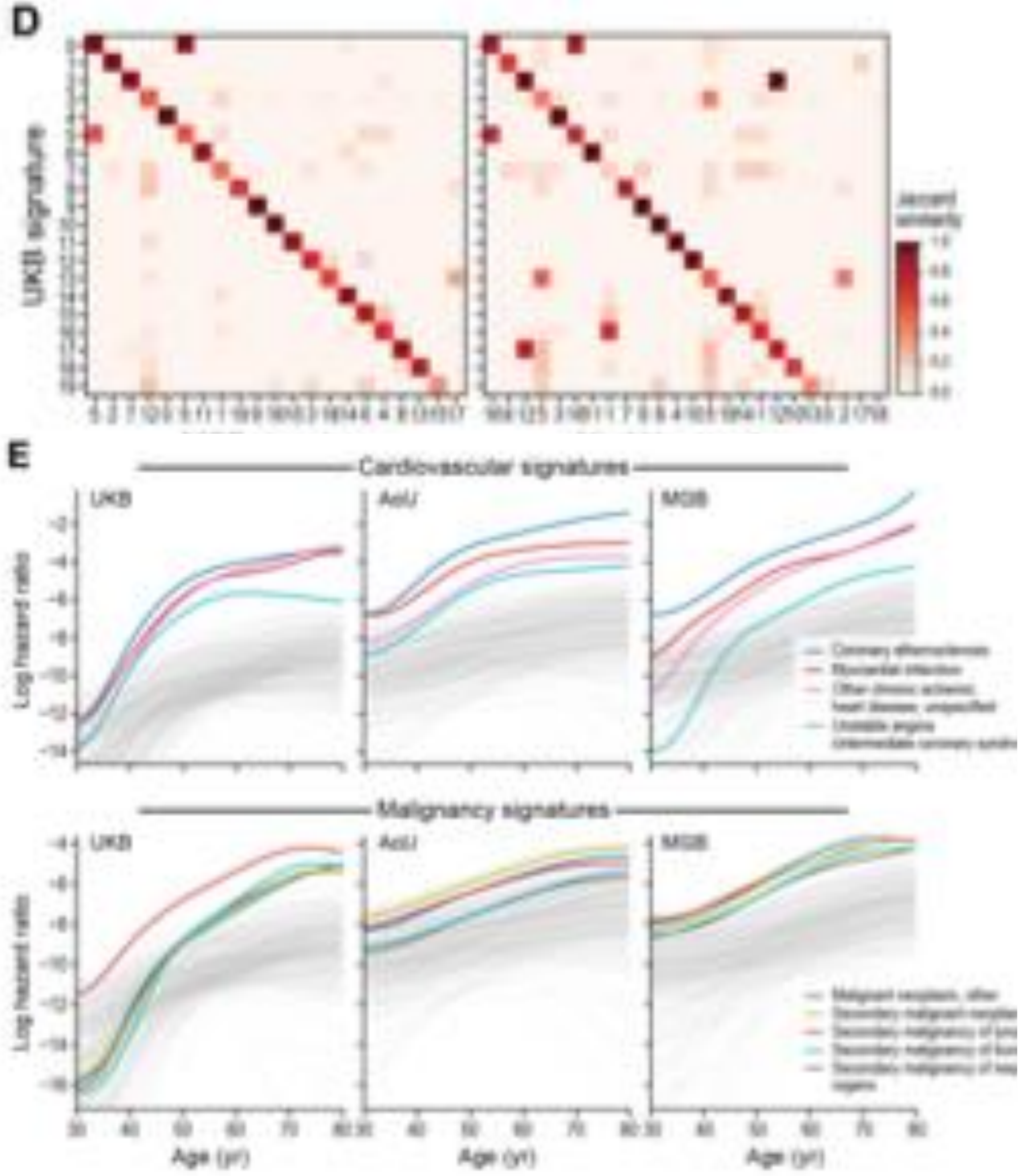
Signatures: Characteristic patterns of incidence and timing $f(\phi_{idt})$



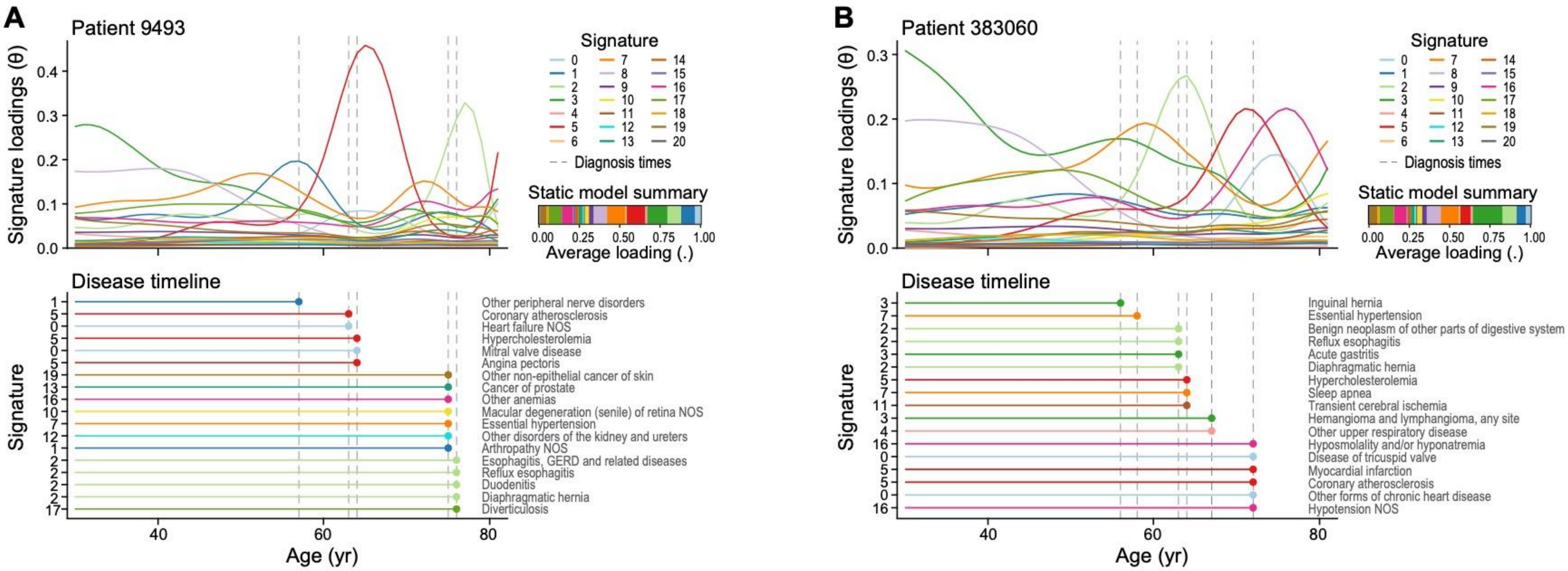
Varying degree of allegiance



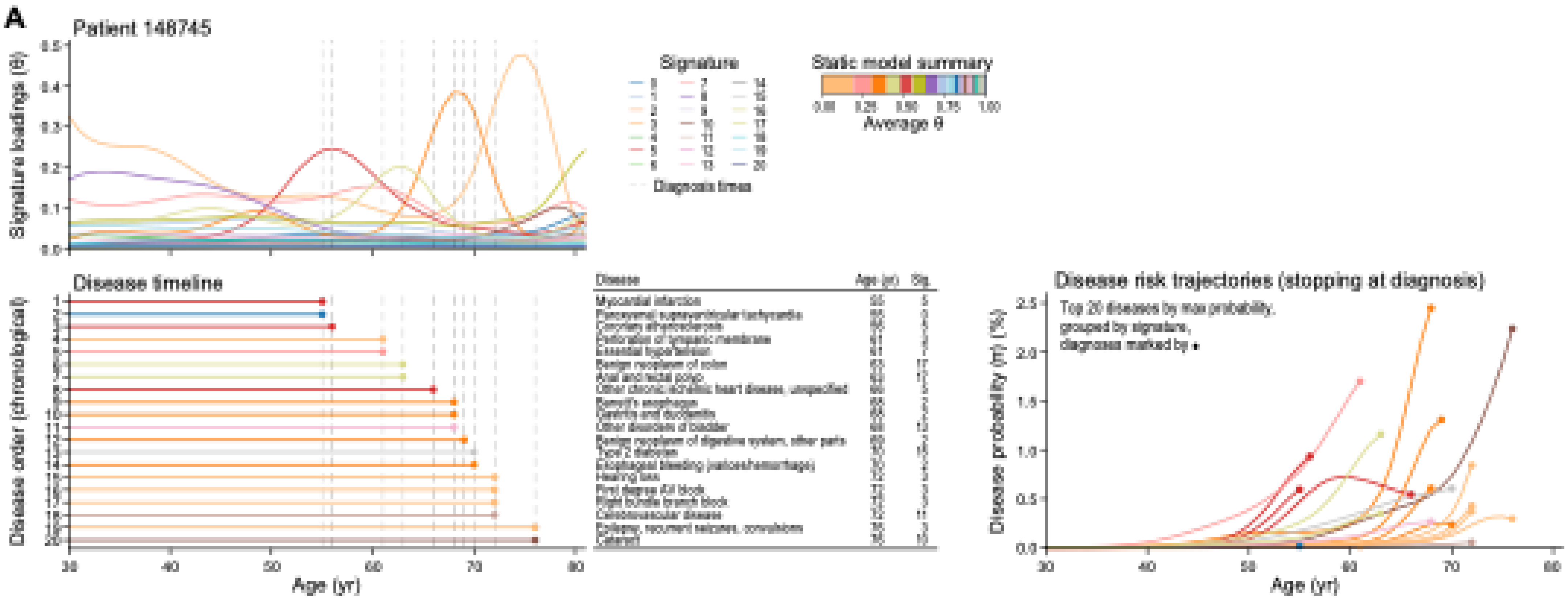
Consistency across biobanks



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

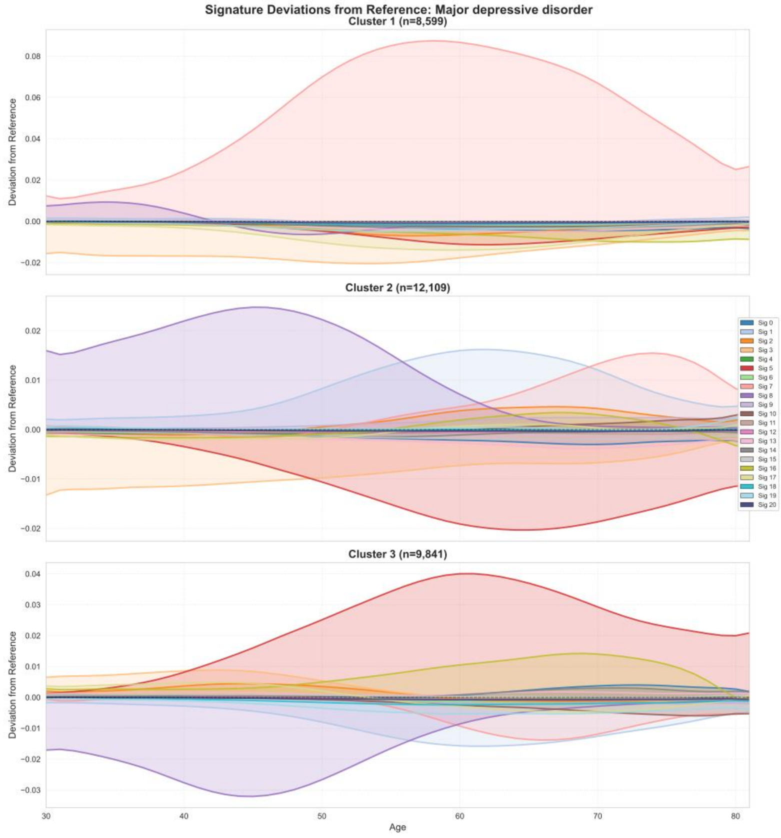


From the past to the future !

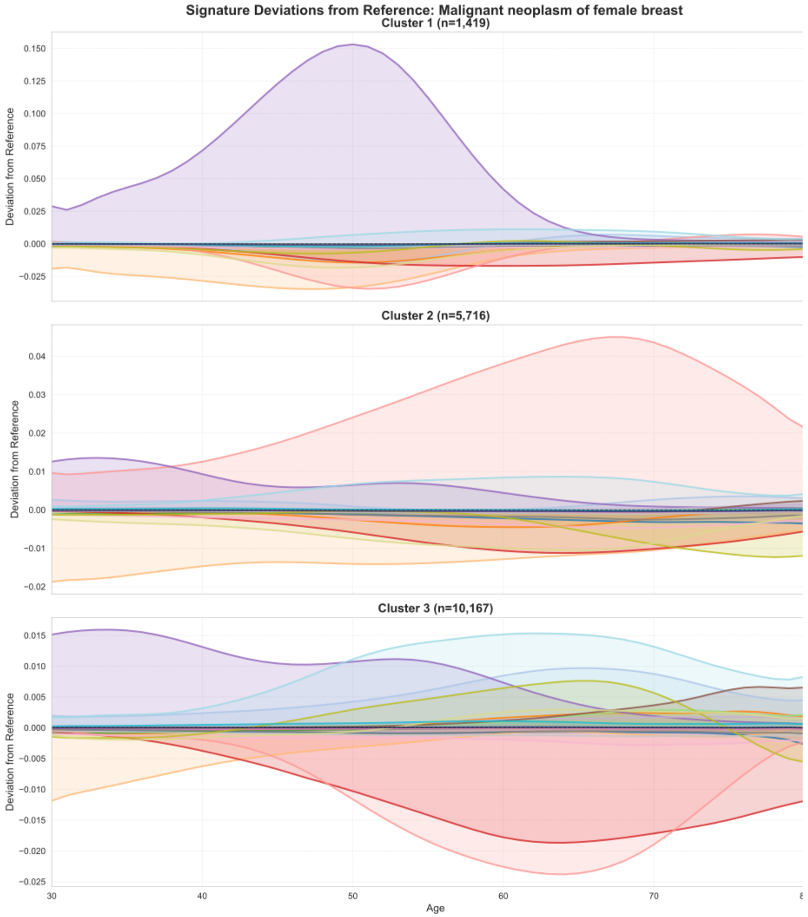


Heterogeneity within disease: Revealing biology

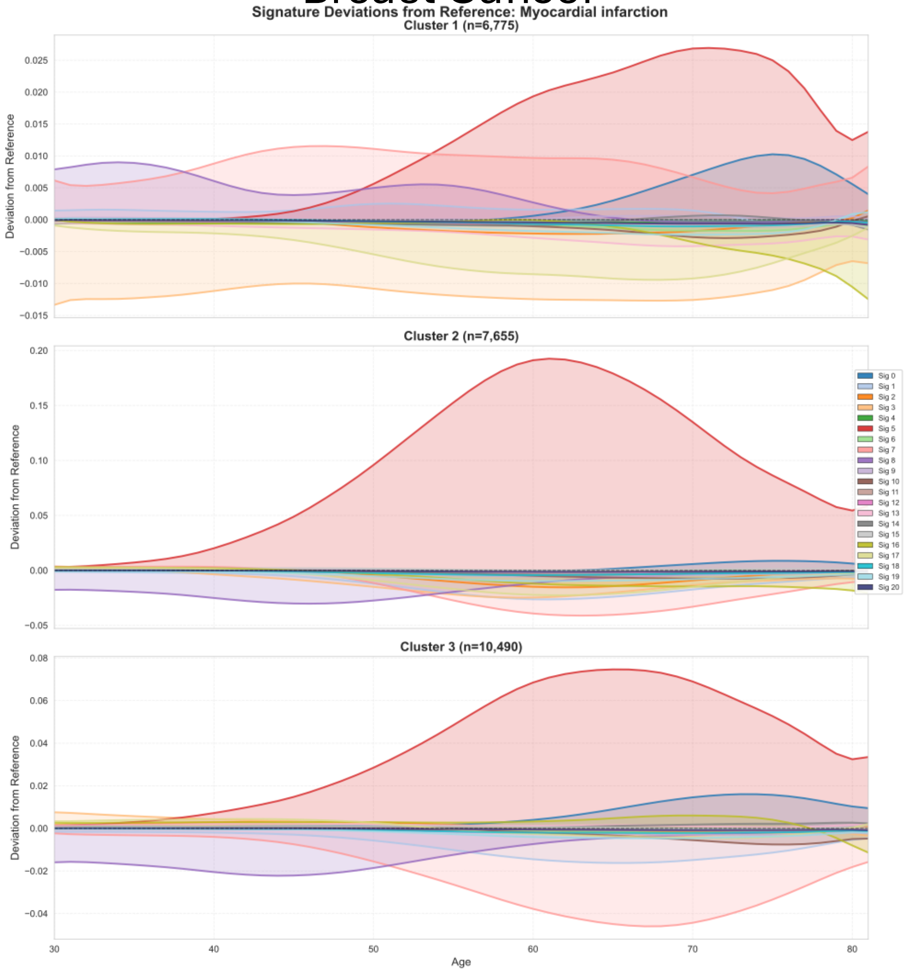
Major Depression



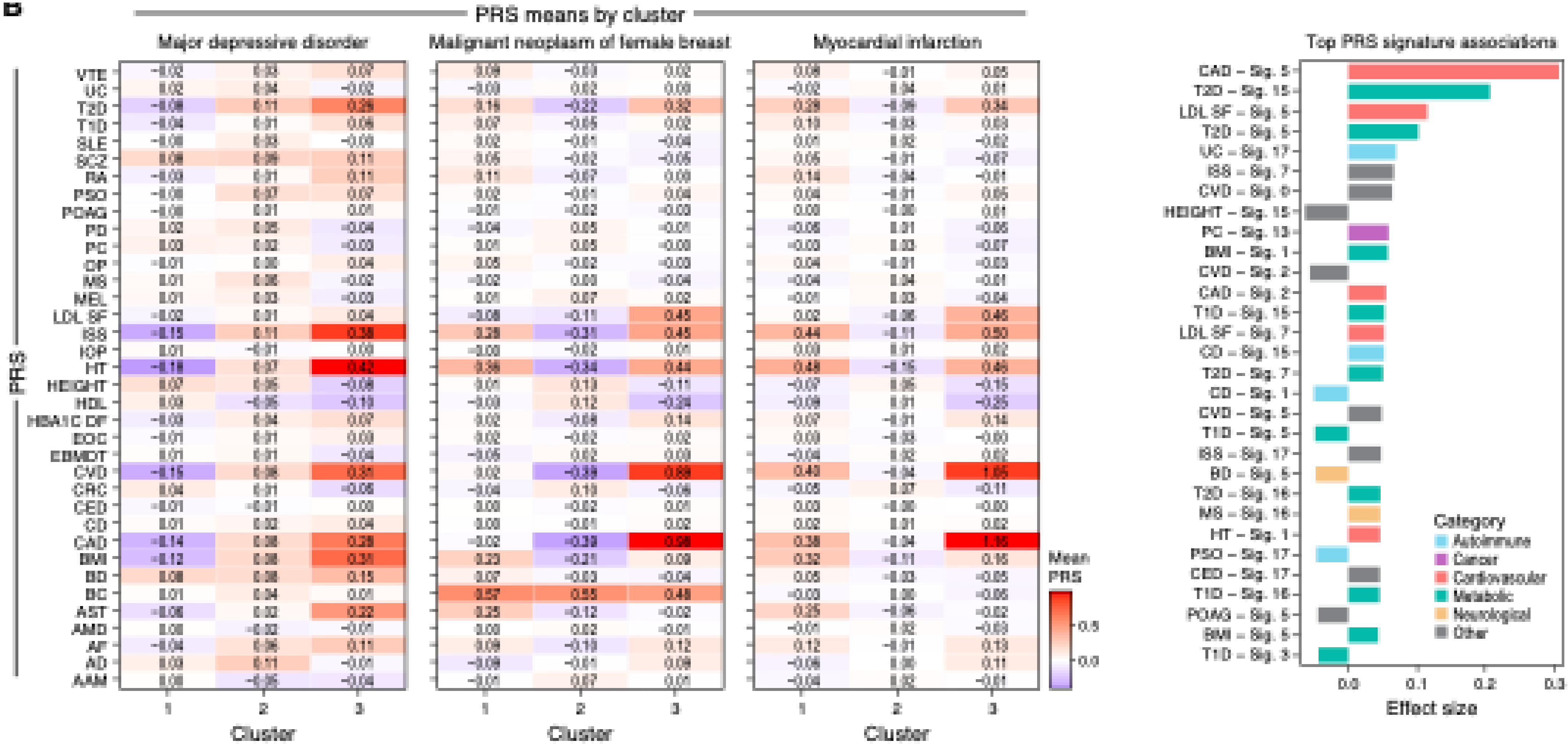
Myocardial Infarction



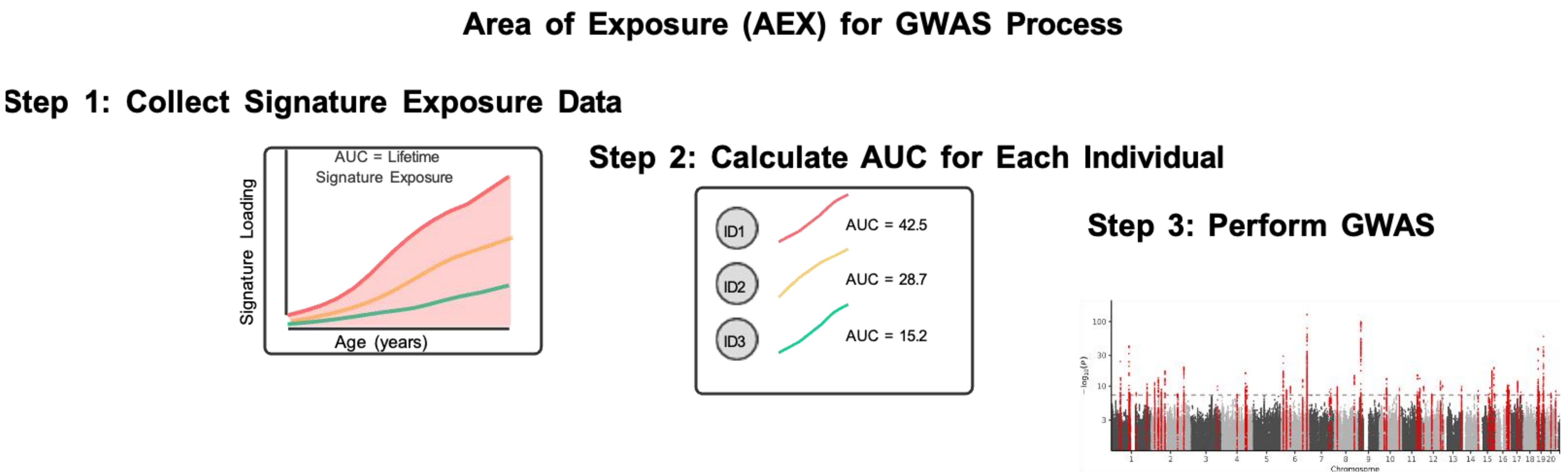
Breast Cancer



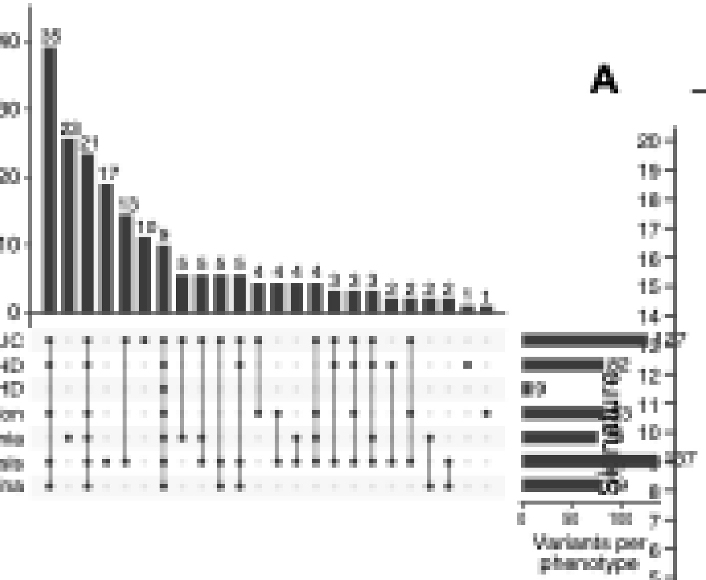
Heterogeneity in phenotype matches Genotype

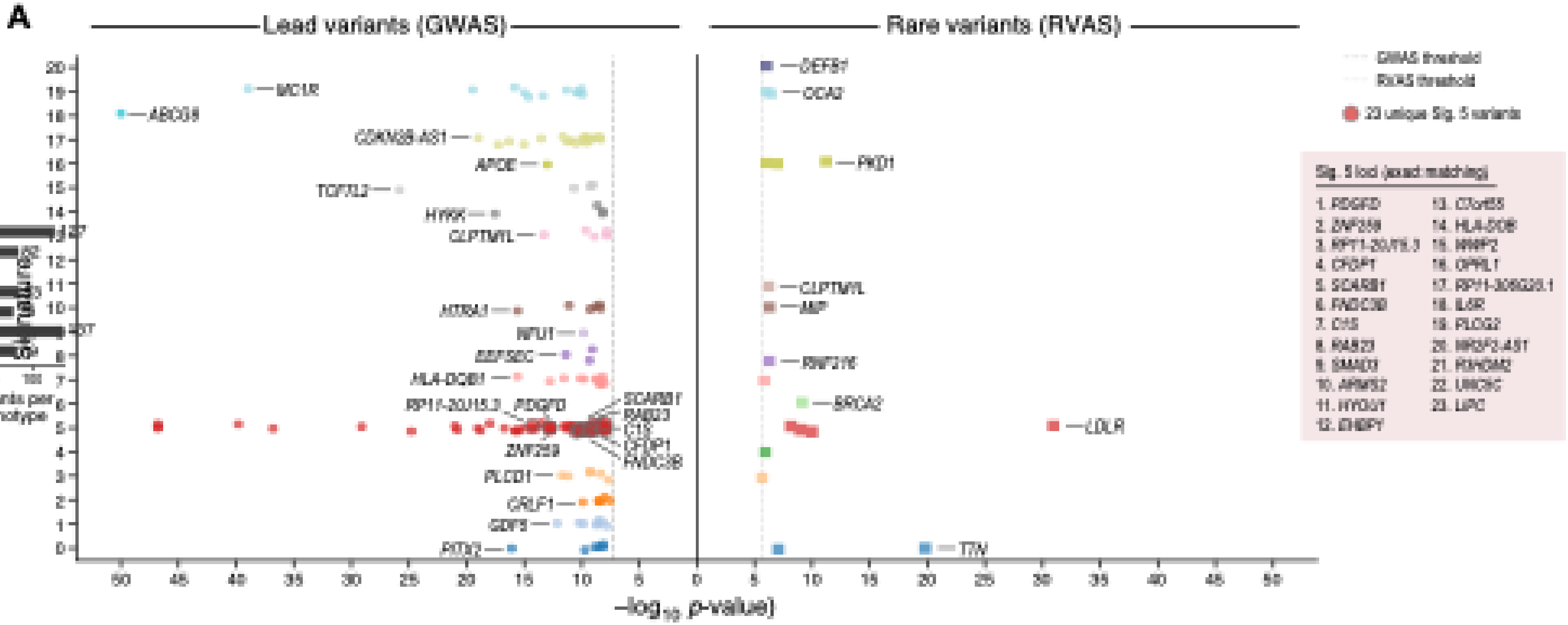


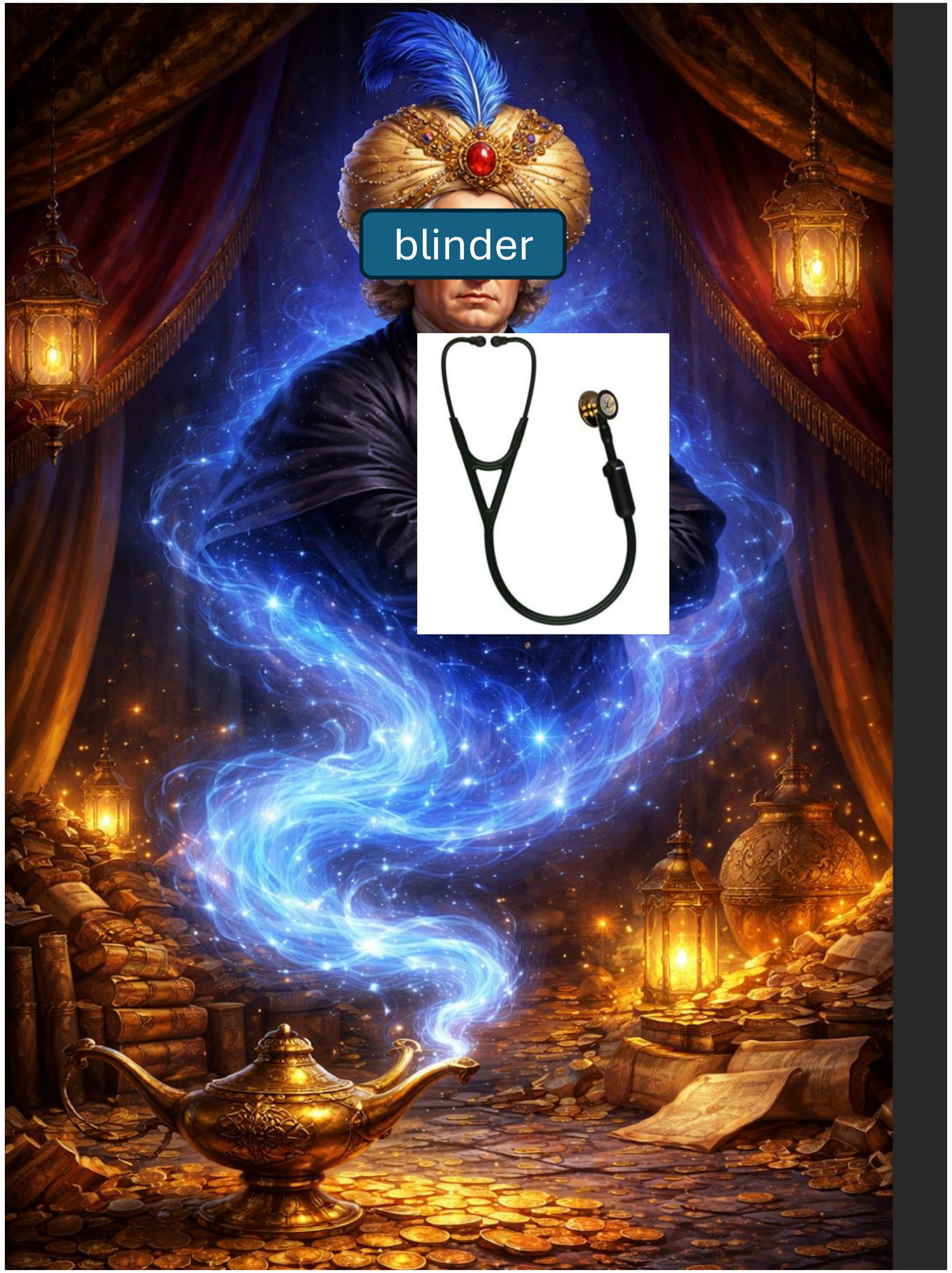
Cumulative Exposure GWAS



How well do we recapitulate known and Discover new?

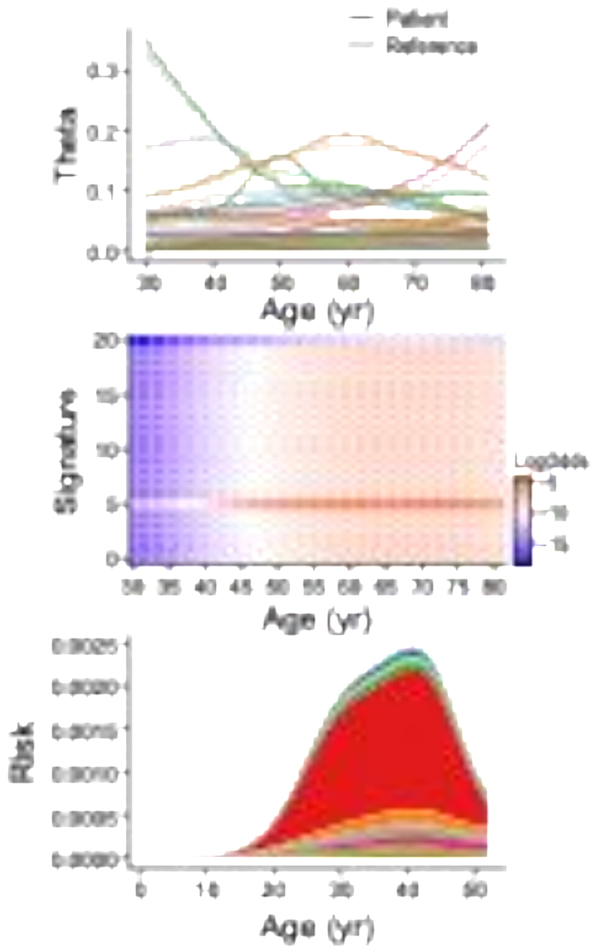


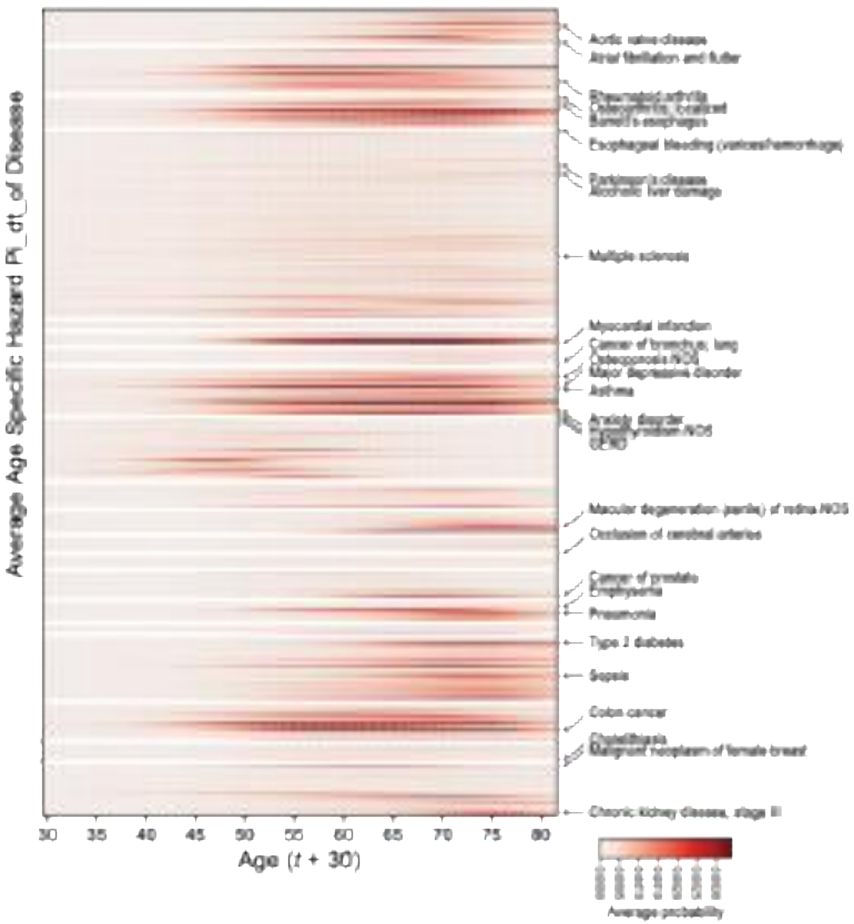



blinder

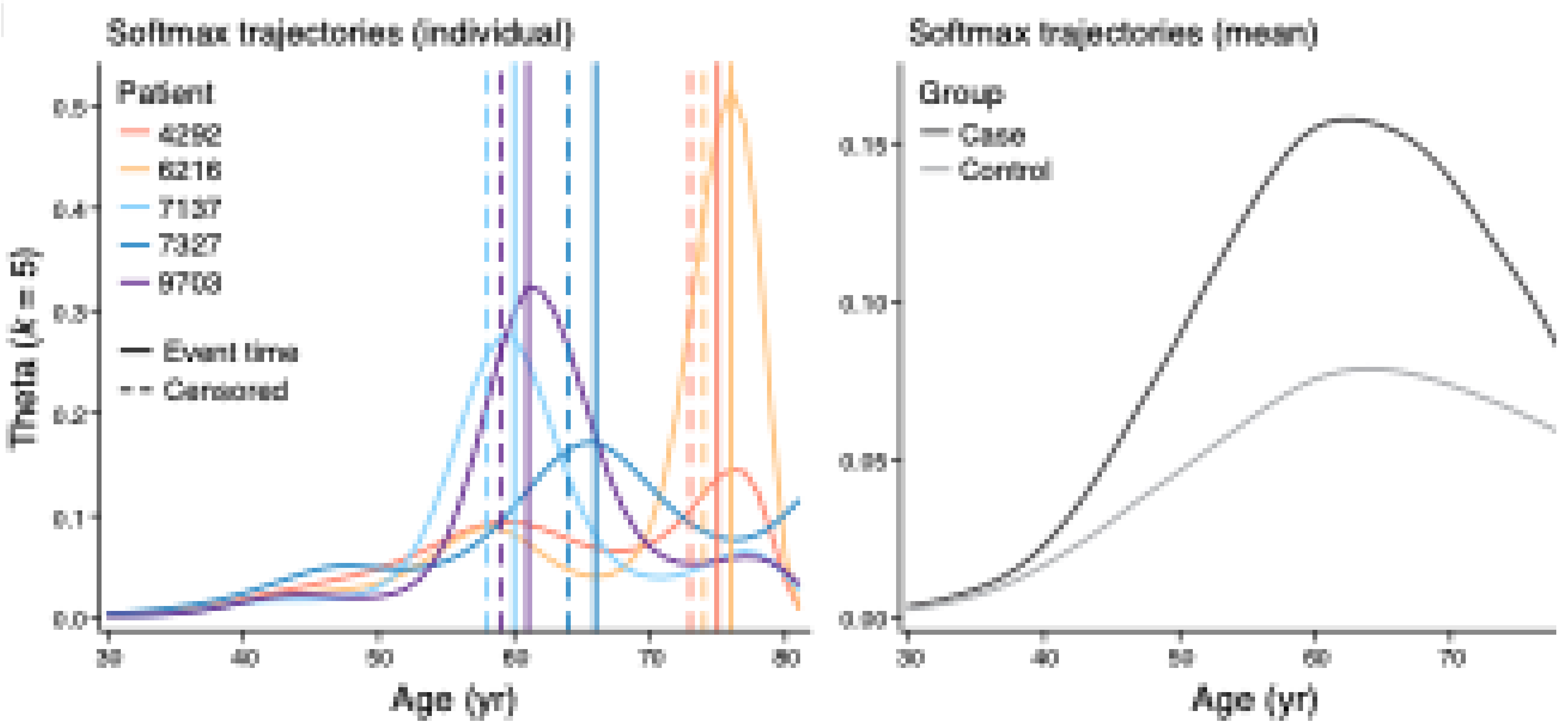
Making predictions is only part …

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

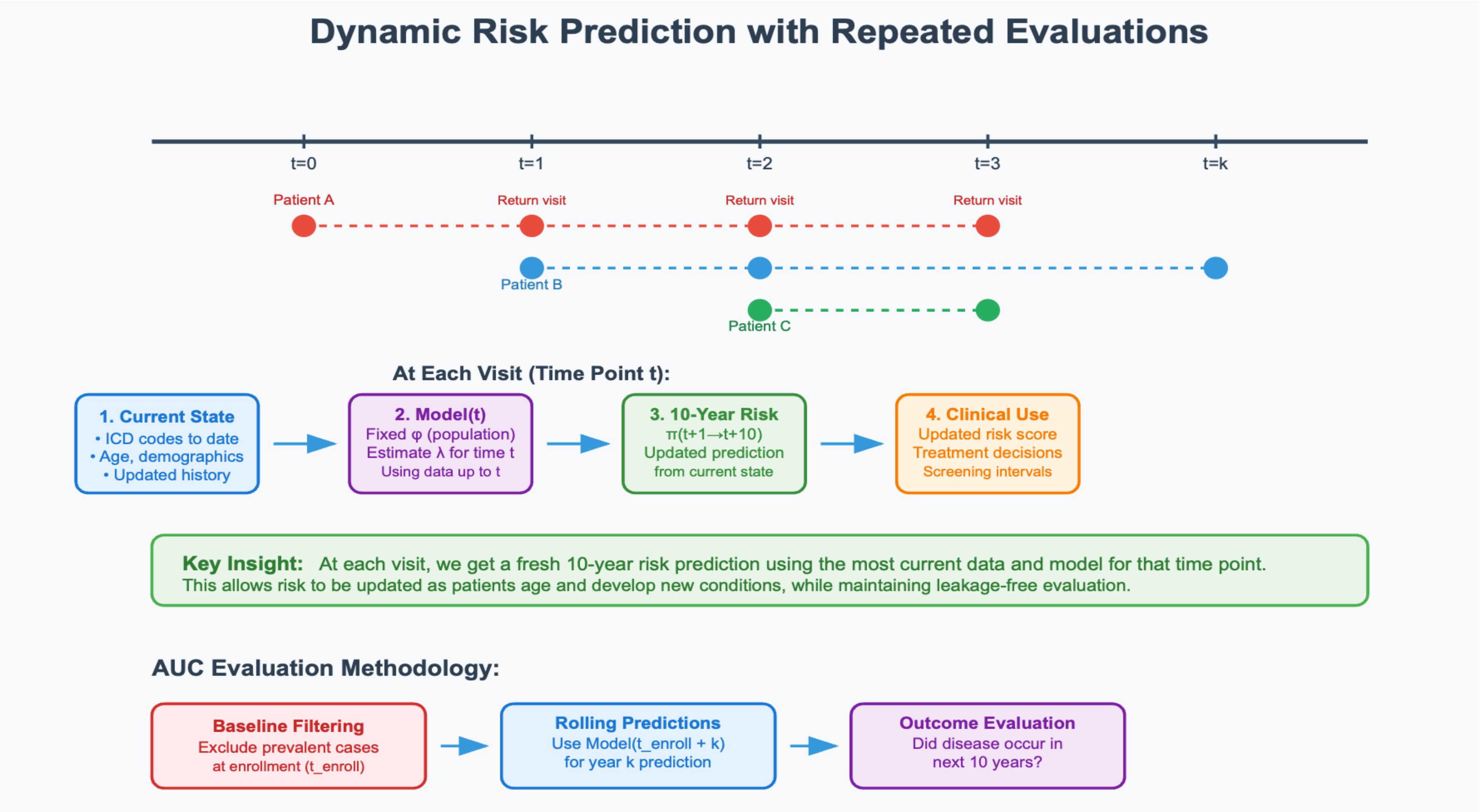


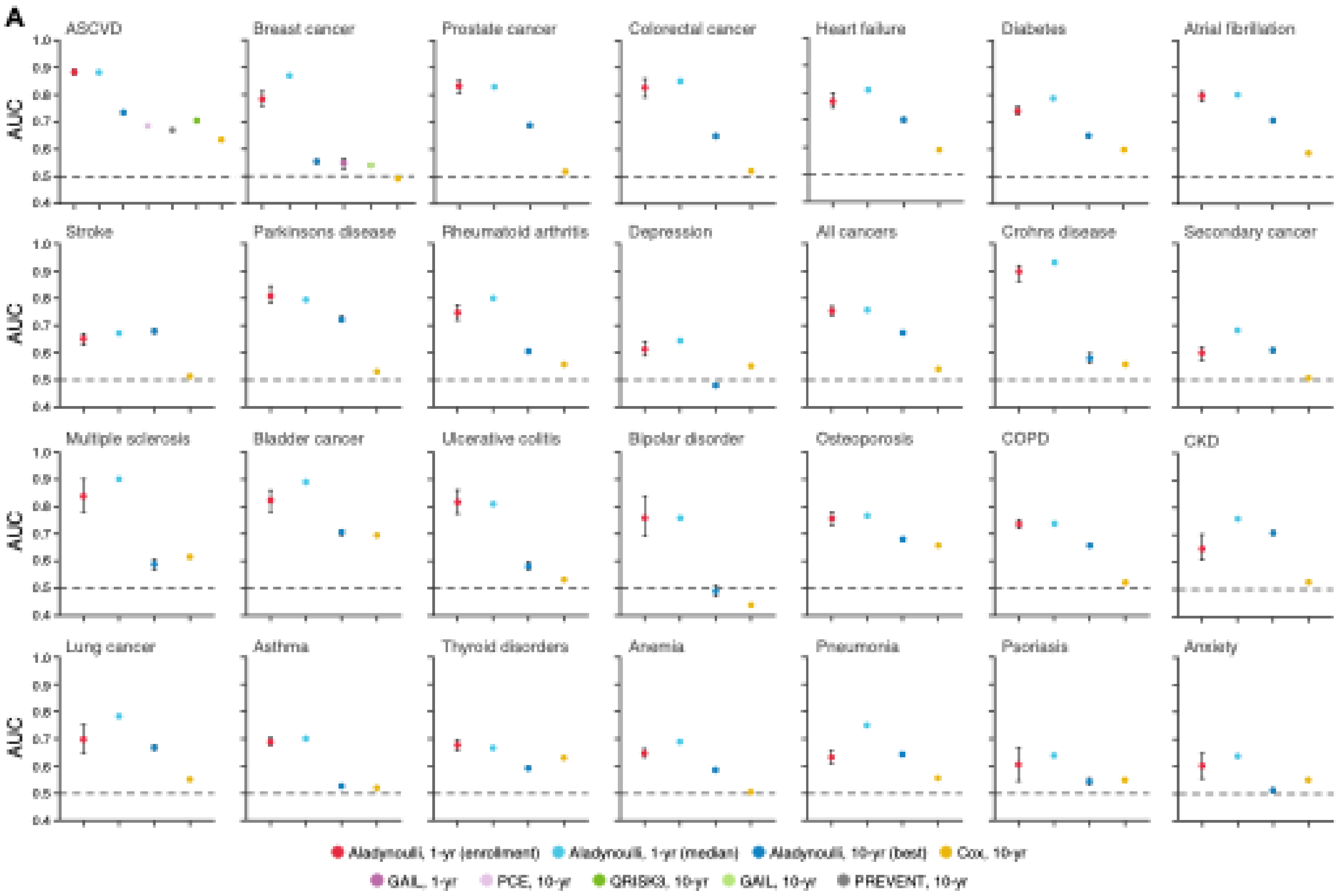


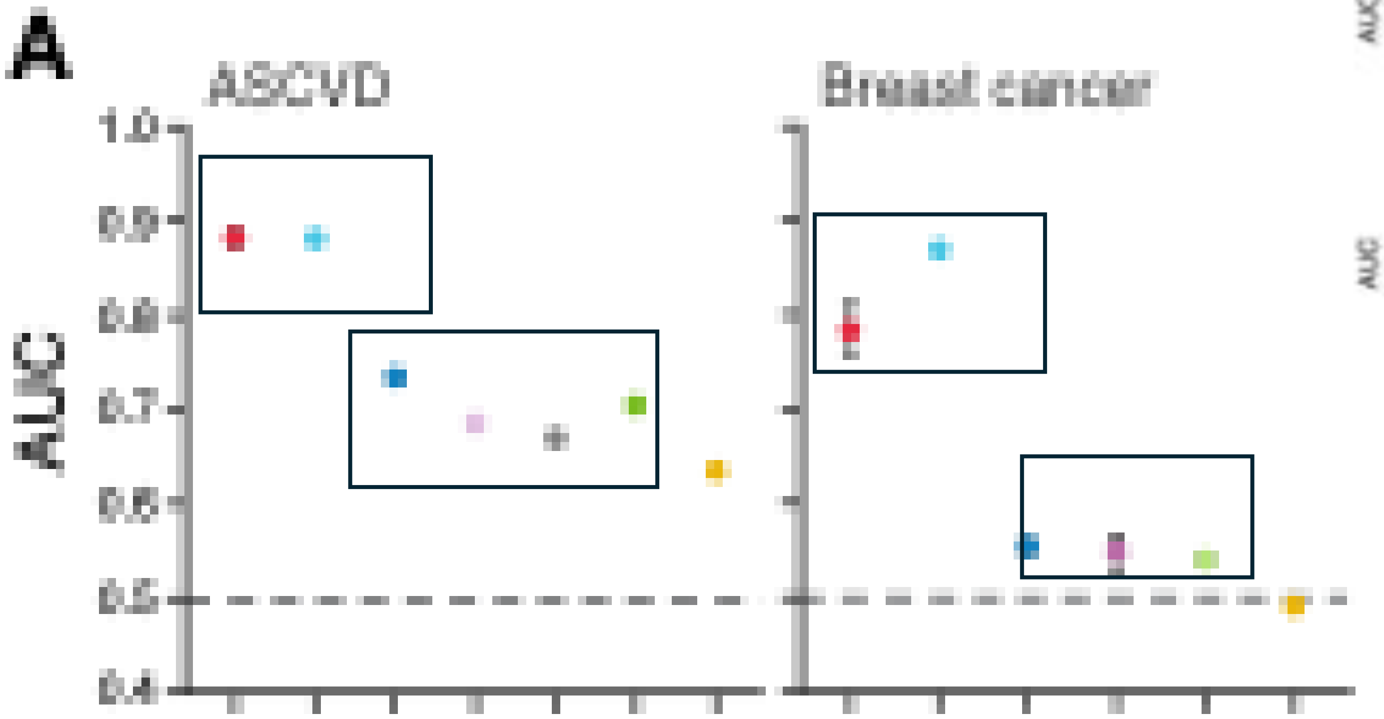
Aladynoulli: the genie works with a blindfold!



Performance assessment: Changing the paradigm




A
ASCVD
Breast cancer
Prostate cancer
Colorectal cancer
Heart failure
Diabetes
Atrial fibrillation
Stroke
Parkinsons disease
Rheumatoid arthritis
Depression
All cancers
Crohns disease
Secondary cancer
Multiple sclerosis
Bladder cancer
Ulcerative colitis
Bipolar disorder
Osteoporosis
COPD
CKD
Lung cancer
Asthma
Thyroid disorders
Anemia
Pneumonia
Psoriasis
Anxiety
AUC
1.0
0.9
0.8
0.7
0.6
0.5
0.4
Aladynoulli, 1-yr (enrollment)
Aladynoulli, 1-yr (median)
Aladynoulli, 10-yr (best)
Cox, 10-yr
GAIL, 1-yr
PCE, 10-yr
QRISK3, 10-yr
GAIL, 10-yr
PREVENT, 10-yr

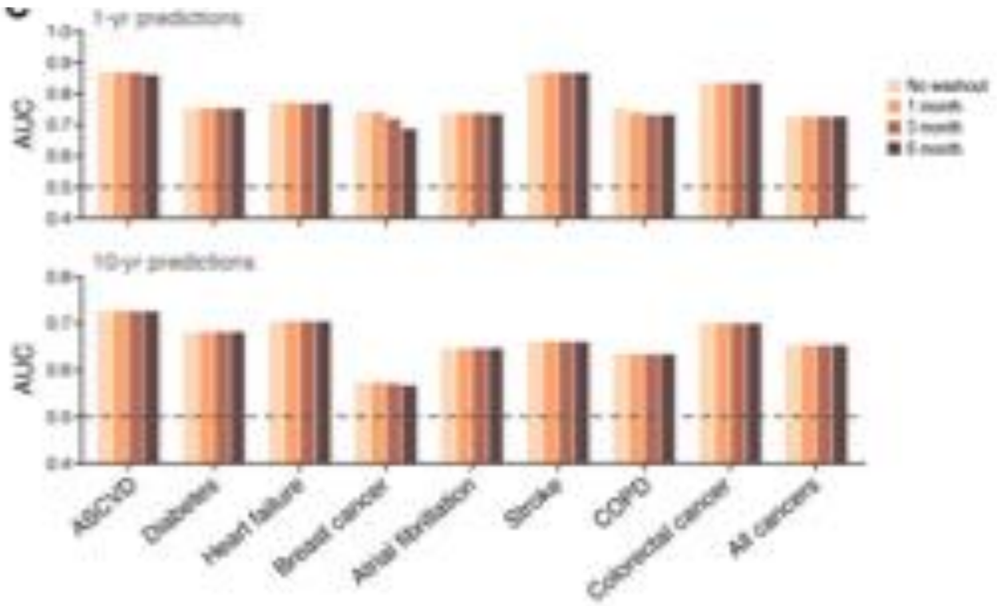


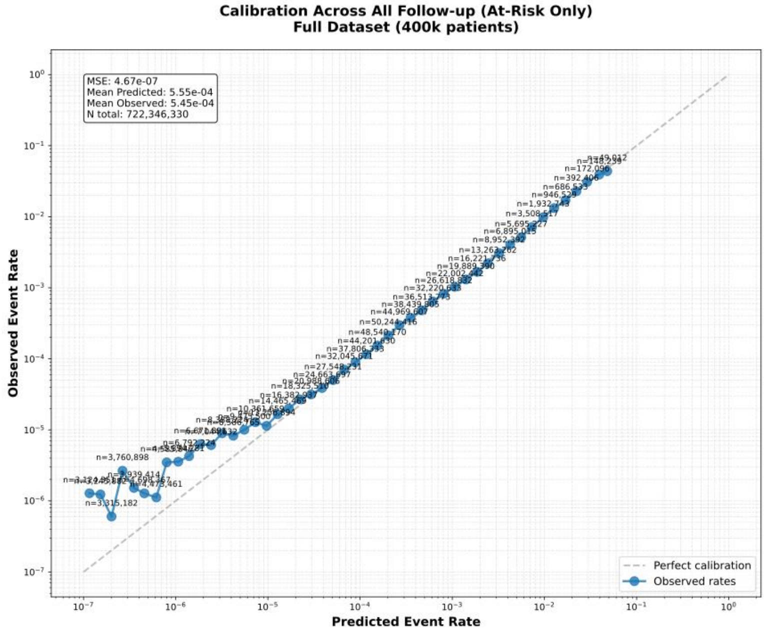
Noull Enrolment
Dynamic

Aladynoulli 10
PCE,
Prevent,
Qrisk,
Cox

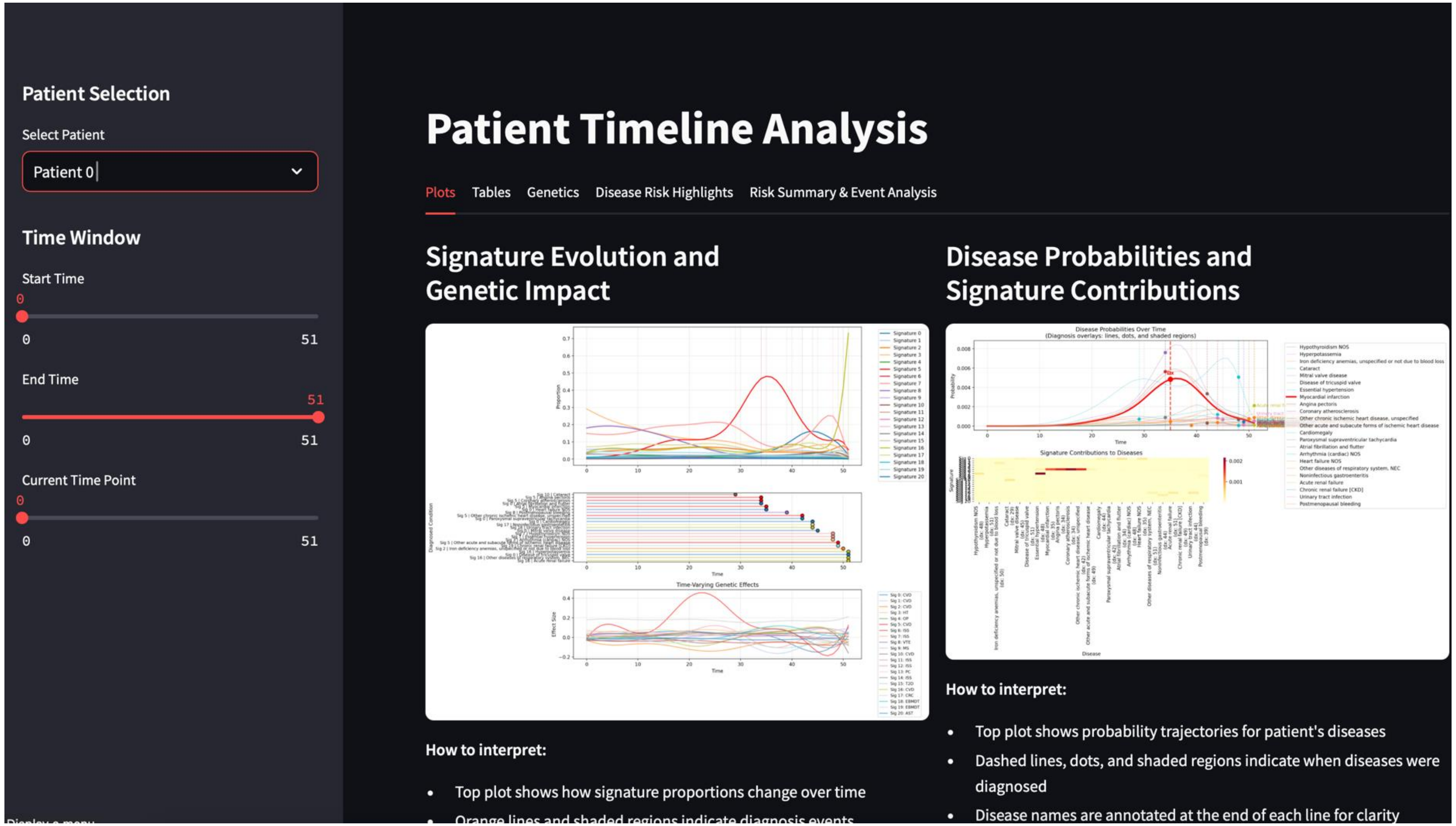
Noull Enrolment
Dynamic

Aladynoulli 10
Gail 1 yr
Gail 10 yr
Cox

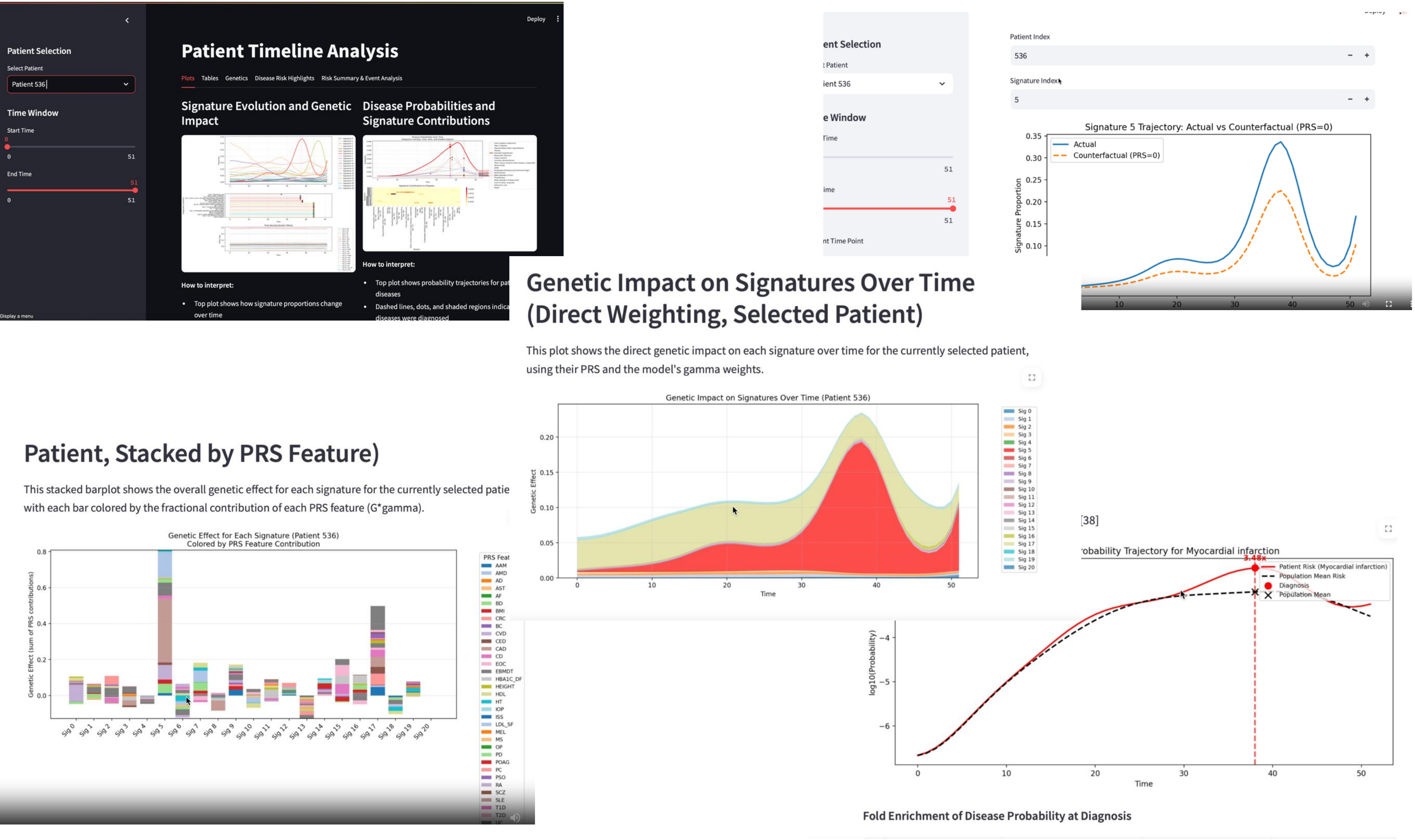




What does Dynamic Look like?



http://aladynoulli.hms.Harvard.edu


Deploy
Patient Selection
Select Patient
Patient 536
Time Window
Start Time
0
51
End Time
51
0
51
Display a menu
Patient Timeline Analysis
Plots
Tables
Genetics
Disease Risk Highlights
Risk Summary & Event Analysis
Signature Evolution and Genetic Impact
Disease Probabilities and Signature Contributions
How to interpret:
Top plot shows how signature proportions change over time
How to interpret:
Top plot shows probability trajectories for pat
diseases
Dashed lines, dots, and shaded regions indica
diseases were diagnosed
ent Selection
t Patient
ient 536
e Window
Time
51
ime
51
51
nt Time Point
Patient Index
536
Signature Index
5
Signature 5 Trajectory: Actual vs Counterfactual (PRS=0)
Actual
Counterfactual (PRS=0)
Signature Proportion
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.
Genetic Impact on Signatures Over Time (Patient 536)
Genetic Effect
Time
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).
Genetic Effect for Each Signature (Patient 536)
Colored by PRS Feature Contribution
Genetic Effect (sum of PRS contributions)
PRS Feat
AAM
AMD
AD
AST
AF
BD
BMI
CRC
BC
CVD
CED
CAD
CD
EOC
EBMDT
HBA1C_DF
HEIGHT
HDL
HT
IOP
ISS
LDL_SF
MEL
MS
OP
PD
POAG
PC
PSO
RA
SCZ
SLE
T1D
T2D
[38]
obability Trajectory for Myocardial infarction
3.48x
Patient Risk (Myocardial infarction)
Population Mean Risk
Diagnosis
Population Mean
log10(Probability)
Time
Fold Enrichment of Disease Probability at Diagnosis

Real life Genies!

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