#AHA24

Aladyn(oulli) individual: Dynamic individual comorbidity modeling for genomic discovery and clinical prediction

Sarah Urbut, MD, PhD | Massachusetts General Hospital

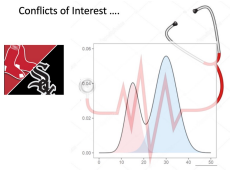
American Heart Association. Mass General Brigham

100 YEARS Bold Hearts

DISCLOSURES

I have nothing to disclose, except …

Conflicts of Interest ….

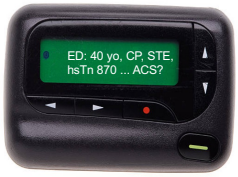


Mass General Brigham

01 THE STORY

#AHA24

THE CALL



- Short term focus
- No dynamic trajectory
- Missing Genetics ….

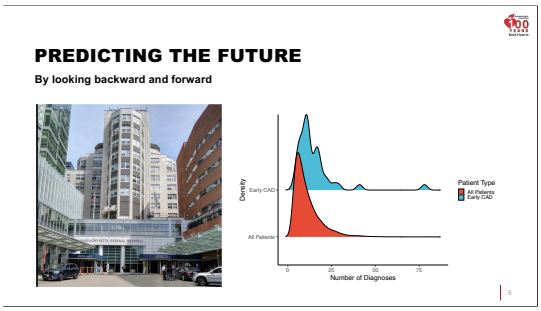
Mass General Brigham

An early shift on the consult pager as a first-year cardiology fellow and I was startled by this call:

40 year old gentleman, classic inferior ST elevations, crushing substernal chest pain, and markedly elevated biomarkers.

Perhaps not a diagnostic dilemma, but as a budding cardiologist and trained statistician, I was struck by the irony:

Sound clinical reasoning would alarm urgent coronary angio, but the day prior to his presentation he wouldn't have even qualified for primary prevention due to our reliance on short term risk. To address this concern, some have suggested a longer interval focus, however this still fails to capture the dynamic trajectory of an individuals' changing risk profile, time varying effects, or the primordial power of genomics.

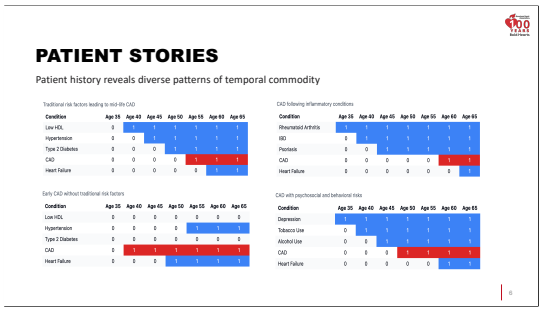


In order to to both understand and predict, we need to look globally and, paradoxically, consider both the past and the future.

Our patients can be seen by multiple other floors in our hospital before and after their diagnoses, giving us insight into their future trajectory across many diseases.

Not unlike Chicago voting, coming early can also mean coming often; patients with early disease are enriched for total all-cause lifetime diagnoses.

We would be shortsighted if we ignored underlying genomics in combination with this longiduijnal information when seeking to understand trajectories.'

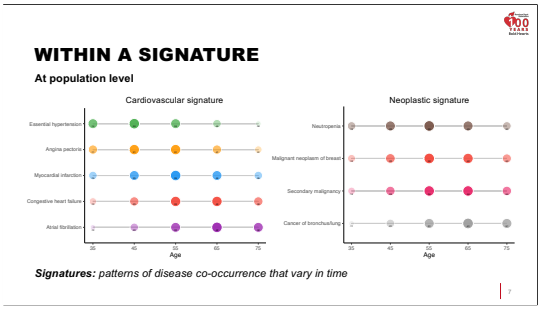


When looking at patient timelines, no one characteristic pattern of disease co-occurrence emerges, and the temporal dimension adds a degree of complexity.

For example, here I display four patient timelines, all for patients with CAD in red but whose disease coexisted at different times and with different comorbid 'partners':

upper left following a more traditional metabolic cascade, lower left without any clear precipitating RF,

upper right, following an inflammatory cascade, and finally following environmental insults.

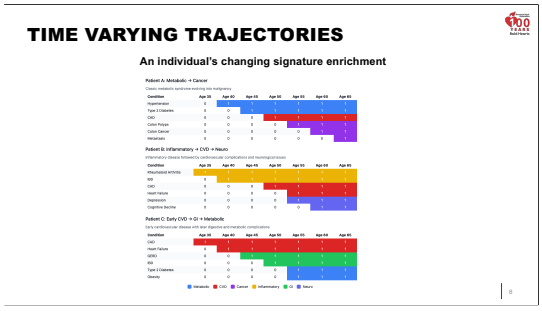


If we're seeking to find summary 'patterns' of disease co-occurrence in the overall data, we need to ask not only what but when.

We can define Signatures as: Patterns of disease co-occurrence that vary in time.

A signature asks with which comorbidities and WHEN a given diagnosis tends to occur, rather than just its presence or absence.

For example, in the toy plots featured here, if we observed a cardiovascular and neoplastic signature, we might note that disease prevalence generally tends to increase with time, but that the peak age of onset varies for each condition within the signature. This aligns well with the heterogeneity in onset and prevalence we observe in real population incidence data.

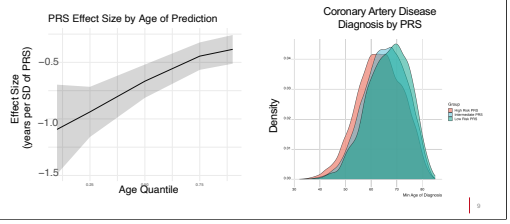


Similarly, the signature or profile an individual predominantly exhibits varies dynamically over time: characterizing a patient as 'vascular or "neoplastic' depends on when the question is asked, rather than being neatly summarized at a single snapshot.

Here, we show three unique sequences of progression through distinct disease signatures as indicated by the colors which connote disease processe.

GENETICS MATTERS

Genetics impacts early disease and all disease, early



So, where does genomics come in? An individual's relative enrichment in any one of these signatures evolves as it is influenced by genetics, environment, and previous diagnostic history.

Genetics tells much of this story.

If we examine the effect size of Afib PRS, for example, on age of onset as stratified by age of prediction at left, we see that the effects are more negatively predictive when considered among younger individuals.

Furhtermore, early CAD is enriched in high polygenic risk, but it tells only part of the story: as an isolated biomarker, PRS leaves much to be desired.

So how can population-level tendencies, underlying genetic predisposition, and new diagnostic data inform our ability to transition between signatures over time?

IDENTITY CRISIS

Joint consideration: discovery and prediction

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

While the statistician inside might jump to any number of clustering algorithms, the careful clinician begs to consider the data more closely in a model-based approach.

The power of an unsupervised analysis can also be its peril, and a black box, strictly machine-learning-based analyses would miss the generative story.

Enter our dear friend and personal hero, Thomas Bayes

BAYES THEOREM

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

Continuously updated posteriors

Individual data likelihood (EHR, clinical data)

Individual predilection to a signature

Population level signatures

The crux of Bayes theorem is to combine new specific data with the prior belief based on universal patterns.

In this case, our 'new' data is the streaming set of updated diagnoses acquired from the EHR for a given patient

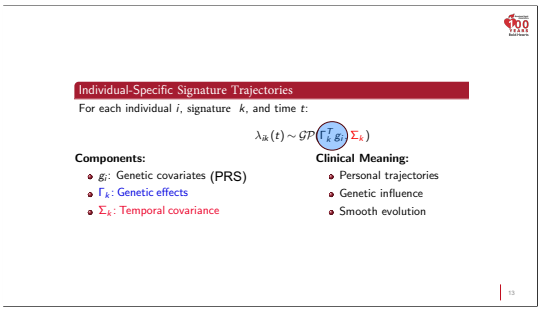
The prior here combines population-level patterns of disease occurrence over time with our naive estimation of an individual's underlying signature predilection centered around germline genetic variation.

02 RESULTS: OUR MODEL

ALADYNOULLI

#AHA24

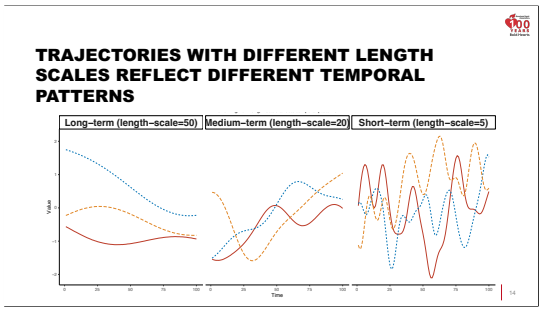
To accomplish all these goals, we adopted a model-based approach in building our new method, Aladynoulli. It's beauty is the ability to describe the generative story in math that we can then approach computationally. The Greeks help us follow biology to inference, and so it's worth taking a moment to understand the intricacy that we have created



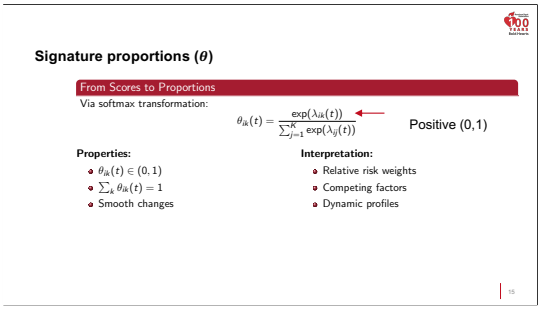
We allow for a set of latent or underlying signatures, each representing a pattern of disease cooccurrence over time. Individuals are indexed by i, signatures by k, and time by t.

We can think of Lambda, or a signature's relative importance to a given individual, as a draw from a a stochastic process centered around the underlying genetic influence, her conveyed by polygenic risk scores in blue. We learn the importance of each genetic score through the set of effects encoded in gamma.

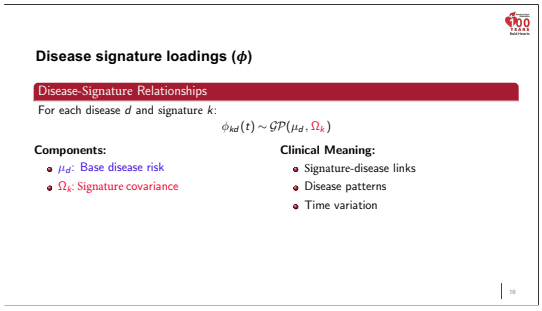
Where does time come in?



To help us accomplish this goal, We take advantage of Gaussian processes: these are simply distributions that help us with smoothing by using a special covariance matrix to capture time.
This covariance matrix is called a kernel, and it reflects how quickly or slowly individual predilection towards a signature can change. The size of these changes is learned through the length scales.



We then transform lambda to a value between 0 and 1, which allows us to interpret the relative weight of each signature for an individual at a given time point



Similarly, a given disease d also arises from a Gaussian process, now centered around the time-specific disease prevalence in the population, here mu_d.

Our special kernel again captures the variation of this signature-disease loading over time.

At each time step, these loadings can be transformed to probability and are used to generate the absence or presence of disease as a bernoulli process, without any competitive restrictions.

Disease probabilities (π)

Individual Disease Risk

Probability for individual i, disease d, at time t:

$$\pi_{id}(t) = \sum_{k=1}^{K} \theta_{ik}(t) \cdot \text{sigmoid}(\phi_{kd}(t))$$

Components:
- Personal risk profile
- Topic contributions
- Temporal dynamics

Clinical Use:
- Risk prediction
- Trajectory planning
- Intervention timing

17

These relative individual-weights and disease loadings combine across signatures to generate an individual probability, pi_idt: This reflects the individual, disease, and time-specific probability of occurrence, given that an individual is still at risk. These are naturally discrete-time hazards.

BAYESIAN MAGIC SHARING OF INFORMATION

Pt Genetics. λ_{ik}

$$\textit{Remaining Risk} = 1 - \prod_{t=Start}^{T} (1 - \pi_{idt})$$

Disease ϕ_{kd}

Time K_k

18

Not only do we understand disease patterns, but you might recognize that these discrete hazards can be used in a survival context to calculate the conditional or marginal probability among any set of conditions over any horizon. The risk of disease occurrence in any horizon is simply 1 – the probability of survival in the same interval.

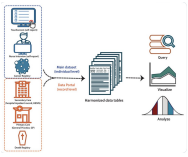
And herein lies the Bayesian magic: we incorporate sharing across individual genomics, diseases, and time points to both discover and predict.

REAL DATA

From EHR to *N* of 1

N: 460,000 individuals in UK Biobank (UKB)
D: 348 conditions, > 2% lifetime prevalence
T: 51 time points (ages 30–80 years)
G: 36 unique externally validated PRS

Validation:
N: ~340,000 individuals in AoU



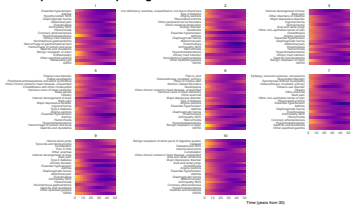
UKB EHR: credit MW Yeung et al. 2022

19

After developing this approach, we applied to nearly ½ M Individuals in UKB, 350 diseases over 50 years, using 36 Polygenic risk score to anchor our loadings and validated in the AllofUs database.

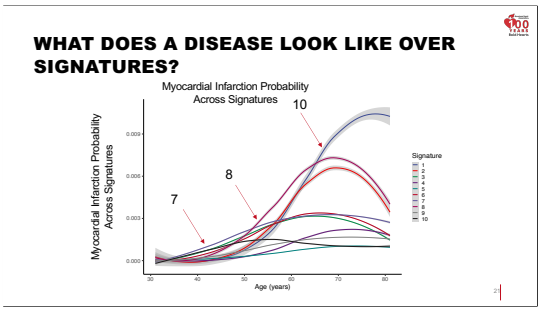
WHAT DOES A SIGNATURE LOOK LIKE?

Top 20 diseases per signature over time



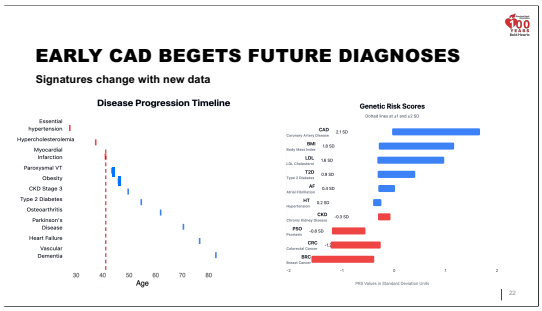
20

From this model, we retrieve a beautiful but complicated set of heatmaps summarizing the loadings of all diseases over 50 time points on each signature. These are the novel patterns of latent sharing across disease and time, the phi that we introduced.

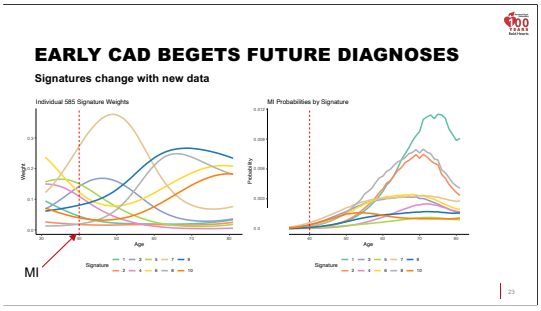


like a good Bayesian, we must translate back to the clinic. We take an example of one disease, myocardial infarction, and examine its loading across signatures over time.

We notice that those with very early MI feature heavily on signature 7, mid onset 8, and later onset signature 10, just the like the patterns we observed in clinic. An individual with MI's weighting on each of these signatures would depend on the age of onset and the diseases with which his particular diagnosis coexists.

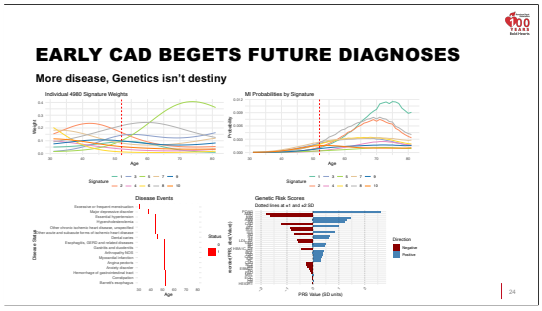


Returning to our early case study, let's take an individaul patient example. Here we encounter a patient with early MI (41) who in fact had a high CAD PRS, 2 SD above population.

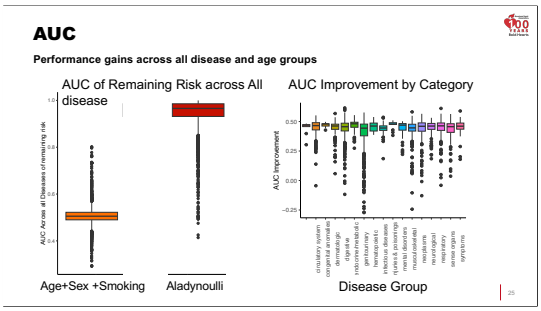


our model appropriately assigns the highest individual weight to the signature with which he has a high prior genetic probability and to which the data lends support through the co-occurrence of signature-time matching diseases.
As we saw in the previous slide, for MI this is signature 7, exactly as we see for this patient.
As he accumulates new diagnoses, his signature weights appropriately change, reflecting the ability of the model to adapt to incorporate new information.

** solidify that



In this multimorbid pt with an MI at 54, he is appropriately weighted on the signature with highest probability of MI, but you'll notce his signatured distribution changes as he accumulates diverse new diagnoses, conveying genetics is NOT only destiny.
A model that would choose only average summary of signature weights would miss this individuals changing comorbidity weighting and the multiple diseases processes he may undergo.
=



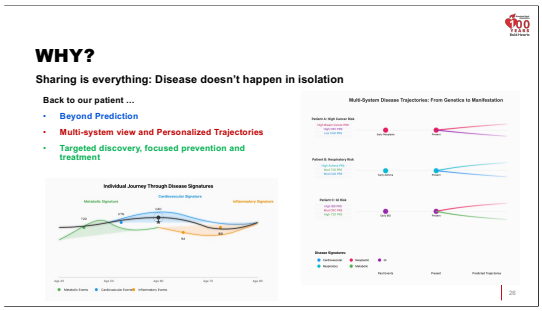
In both case histories, we can follow the trajectories to understand the biology at any given point. But how can we also measure prediction?

At each age, we can use Aladynoulli to estimate the remaining lifetime risk for any disease.

The plot on the left is the area under the receiver operator curve estimating the performance on remaining lifetime risk over all 350 diseases calculated every 5 years for at-risk patients between 30-80.
Red is Aladynoulli when compared to age and sex-based predictions over the same horizon in orange, The plot on the right separates these gains by pre-determined ICD disease-grouping.

The message is clear: borrowing information dynamically across all ages, conditions and patients drastically improves performance for individual disease as well.

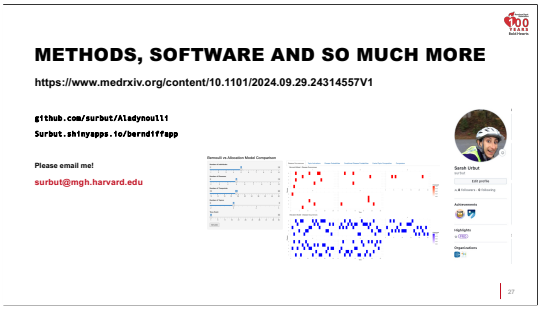
25



Your mamma was right – sharing is caring. Why consider only one condition when your patients and their history bring a wealth of data.

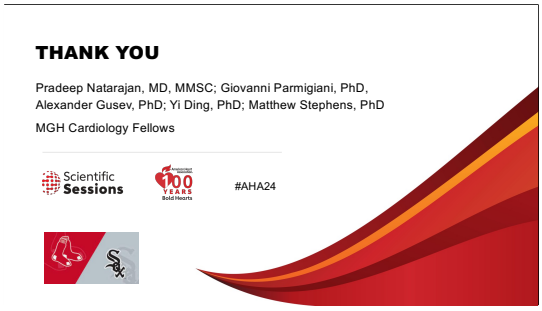
While this story might begin in the germline, more information is gradually uncovered as we ride the elevator with our patientsaha. This goes far beyond GWAS into insight. The beauty of our approach is that in improving prediction, we can learn the underlying relationships between genetics, unique disease processes, and time to ideally stratify what might seem like the same condition into biologically heterogeneous and potentially actionable groups. Understanding heterogeneity is the crux of precision medicine, and we must look globally across all patients, histories, and conditions to help the n of 1 before us.

26



WE've made all of our models and software publicly available, and I have created an application to compare the differences here between our model and an allocation based approach.

27



28