MSGene: Estimating Lifetime Risk using multistate models in the UK Biobank EHR

Sarah M. Urbut, MD PhD, Shaan Khurshid, MD,MPH, So Mi Jemma Cho, PhD, Akl C. Fahed, MD MPH, Giovanni Parmigiani, PhD, Ludovic Trinquart, PhD, Alexander Sasha Gusev, PhD, and Pradeep Natarajan, MD, MMSc

Massachusetts General Hospital, Boston, MA

Broad Institute, Cambridge, MA

Introduction

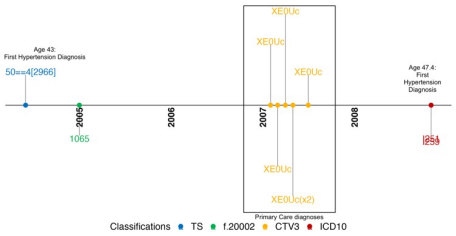
- Younger individuals at high lifetime risk are substantially underrecognized by existing risk stratification algorithms which tend to focus on short-term risk and don't reflect underlying primordial risk
- Lifetime risk algorithms are sorely missing, and don't reflect genetics
- Ability to estimate risk in younger individuals to guide earlier implementation of preventive strategies to those who will benefit the most

Goals

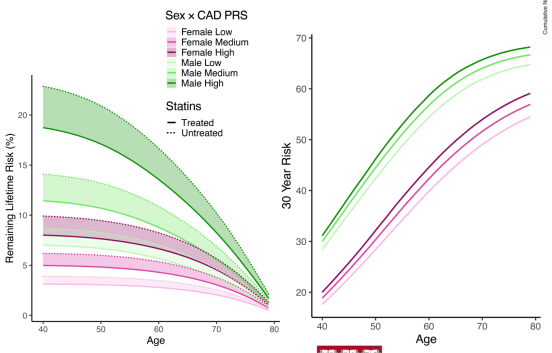
- Estimating CVD-free years gained following preventive treatment strategies
- Ability to assess the effect of individual treatment strategies on CVD free years gained

Data

- 481,686 Individuals enrolled in UKB Primary Care linking outpatient, prescription, hospitalization and operational codes from 1940 to present
- Median FU: 44y [30-58]
- Median first observation age
  - 24.5 [18-37.4]



Results: Lifetime Risk Estimates over the Age Span



HARVARD MEDICAL SCHOOL

Methods: Multistate Model

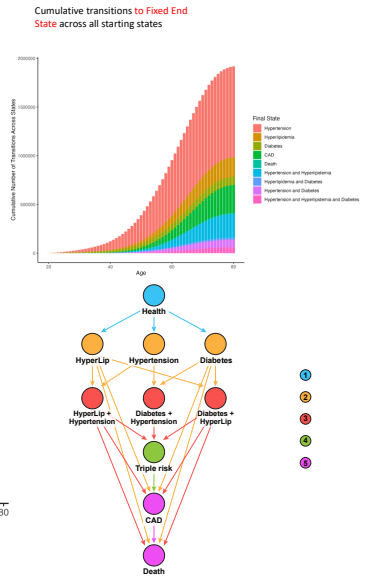
- Generalizability: Good internal and external validation in international cohorts
- Natural adjustment for competing risk of *any* alternative conditions
- Account for accumulated exposures through time in state

And lifetime risk for an individual of progressing to state k from state j where L is the maximum age of life and a is the currently observed age.

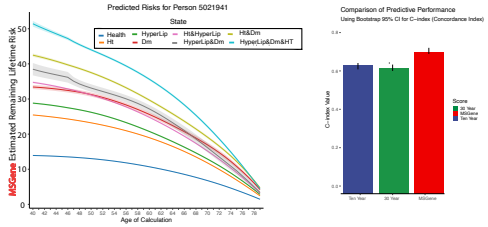
$$1 - \prod_{a=Age}^{L} (1 - \pi_{a,j,k}) \qquad (4)$$

- 10 starting states and 10 unique ending states
- Fit an age-specific generalized linear model with binomial error term for the odds of progressing from every starting state **i** to ending state **j** conditional on chosen covariates
- Covariates are time-fixed (primordial) covariates: sex, genetics, smoking status
- Obviates the need for time-correlated and medication-confounded biometric measurements
- Avoid parametric assumptions about proportional hazards, or linear interactions with time
- Allow for dynamically updated projections with flexible time estimates
- Easy to use R package with fast compute times (no memoryless Markov Assumptions)

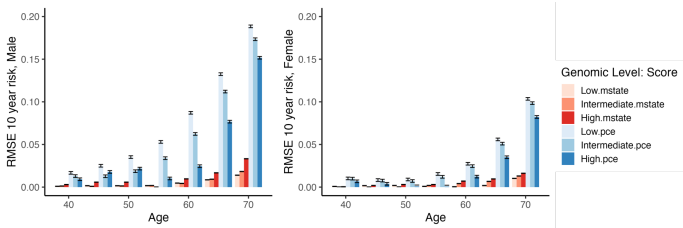
Results: Integrative Software



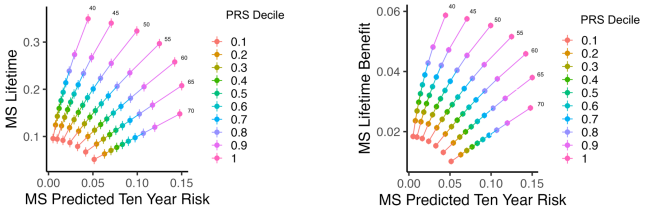
Results: Interactive and accurate flexible risk assessments



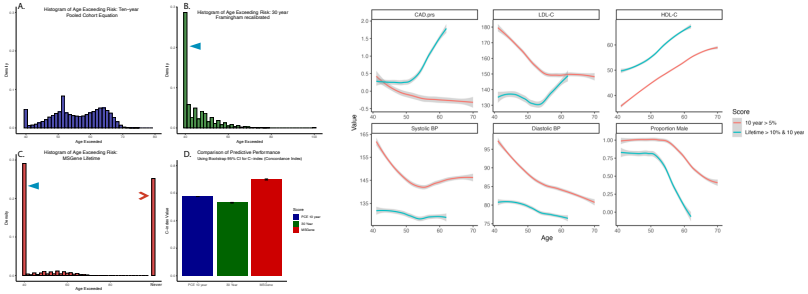
- Extract smoothed per year, covariate combination annual risk of transition to Coronary Artery Disease
- Allow for updated state-specific computation: i.e., how does my trajectory change if I become hypertensive, Diabetic,. etc,



Results: Intuitive Lifetime Risk assessment and Absolute Risk Reduction



Conclusions: Improved Identification of those at high lifetime risk dynamic stratification over time



- Uniquely identify subgroups of individuals at high lifetime risk unrecognized by existing equations
- Rapid recalibration for new populations and an intuitive **R package**