

Construction and analysis of metaPRS on coronary heart disease from associated risks

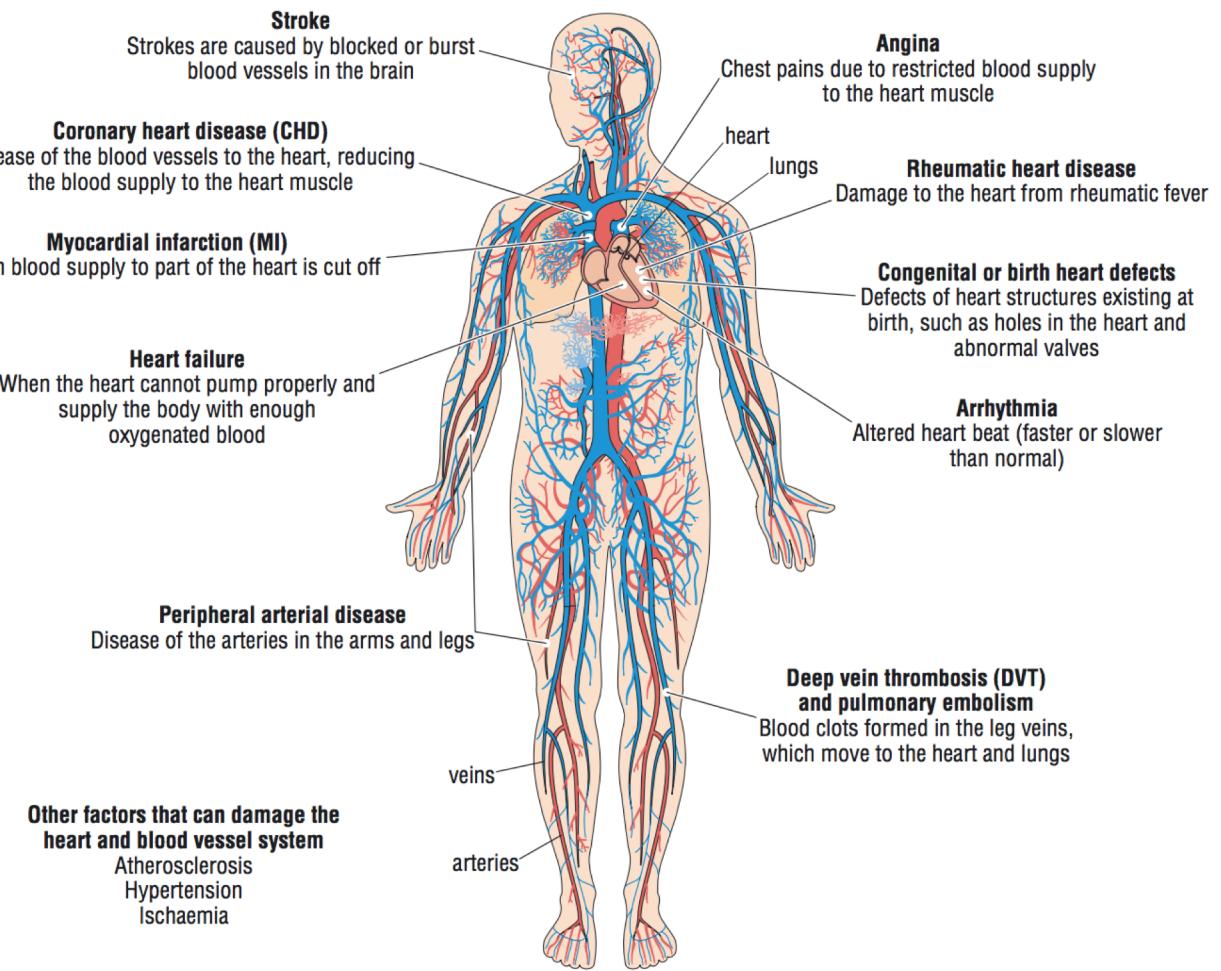
BOS-HEL 30.09.2020

Jake Lin, Pirinen Group, CoE@CDG

Aims

- Improve prediction
 - With pre-disease risk biomarkers
 - SNP contributions
- Cases – severe heart diseases
 - ICD10 codes
 - MI (heart attack)
 - By-pass surgery, cause of death

Umbrella of Cardiovascular diseases



Outline

- Workflow – selection/scoring
- Preliminary results

Recent history of CAD/CHD PRS

Table 4: Examples of polygenic risk models that have been developed for CAD/CHD in recent years.

Publication	Purpose	Brief overview of models examined			
		Models	No. of variants in PRS model	Outcome tested by models	Test population
Inouye <i>et al.</i> 2018 ³	Construct a PRS for CAD and estimate its potential as a screening tool for primary prevention	PRS adj. for sex, age, PC, genotyping array Clinical risk factors only (smoking, diabetes, family history, BMI, hypertension, high cholesterol) Combined model (PRS and clinical risk factors)	1,745,180 HR 1.71	Prediction of incident CAD	UK Biobank
Khera <i>et al.</i> 2018 ²¹	Develop and validate PRS for five common diseases (including CAD)	PRS and age, sex and PC	6,630,150 HR 1.72	Prediction of CAD	UK Biobank
Natarajan <i>et al.</i> 2017 ²³	Examining the impact of statin treatment at different levels of genetic risk	PRS adj. for age, sex, diabetes, smoking, LDL, HDL, BP, antihypertensive use and family history of MI or stroke	38-63	Incident nonfatal MI or death caused by CHD	WOSCOPS
Abraham <i>et al.</i> 2016 ¹²⁶	Construct and externally validate a CHD-PRS, examining lifetime CHD risk and comparison to traditional clinical risk factors	PRS PRS and FRS PRS plus ACC/AHA13 risk score	49,310 HR 1.74	Time to CHD event	Three FINRISK cohorts and two FHS cohorts
Khera <i>et al.</i> 2016 ¹¹⁶	Examining the relationship between genetic risk, CAD and healthy lifestyle	PRS	up to 50	Composite of CAD events (MI, coronary revascularization, and coronary cause death)	Tested in ARIC, WGHS, MDCS, BioImage
Tada <i>et al.</i> 2016 ¹¹⁸	Examining improvements in CHD risk prediction by inclusion of more SNPs and relationship with family history of CHD	PRS and age, BP, antihypertensive use, smoking, apolipoprotein A and B, diabetes	27 and 50	Time to first event of CHD	MDCS

LDL/TG CAD PRS (6M SNPs)

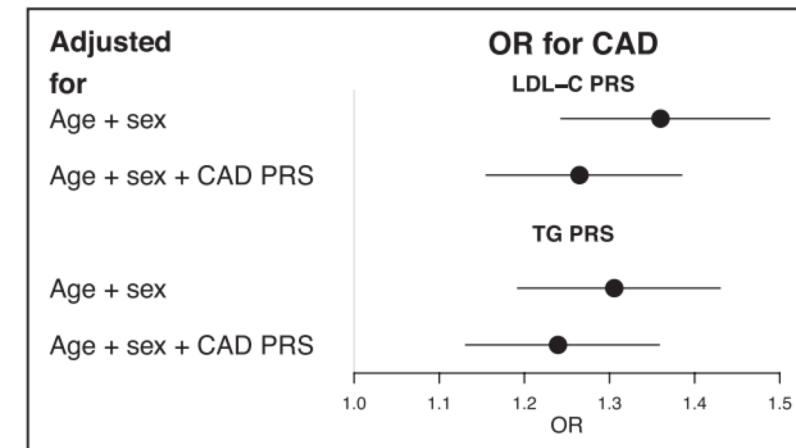
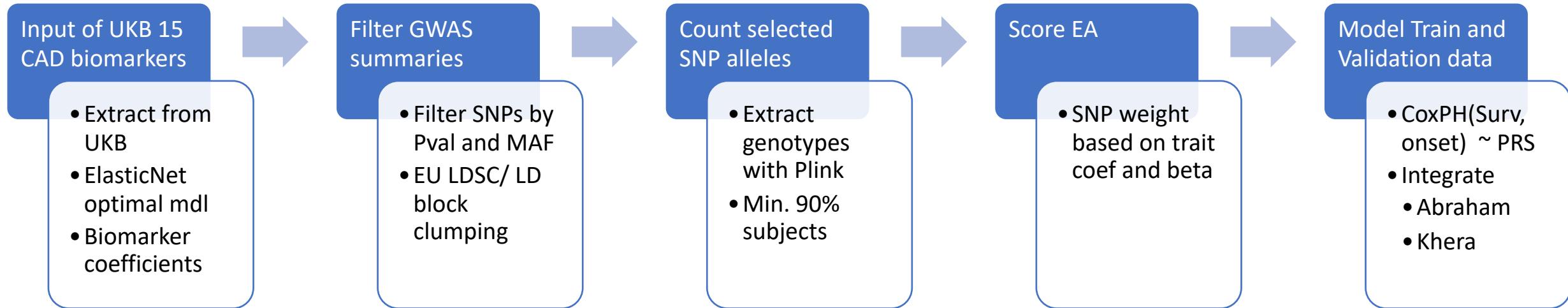


Figure 4. Odds ratios (ORs) for coronary artery disease (CAD)

Ripatti, Rämö, et al. Polygenic Hyperlipidemias and Coronary Artery Disease Risk FinnGen/FinnRisk (2020)

Workflow



Test PRS from Abraham et al. and Khera et al.

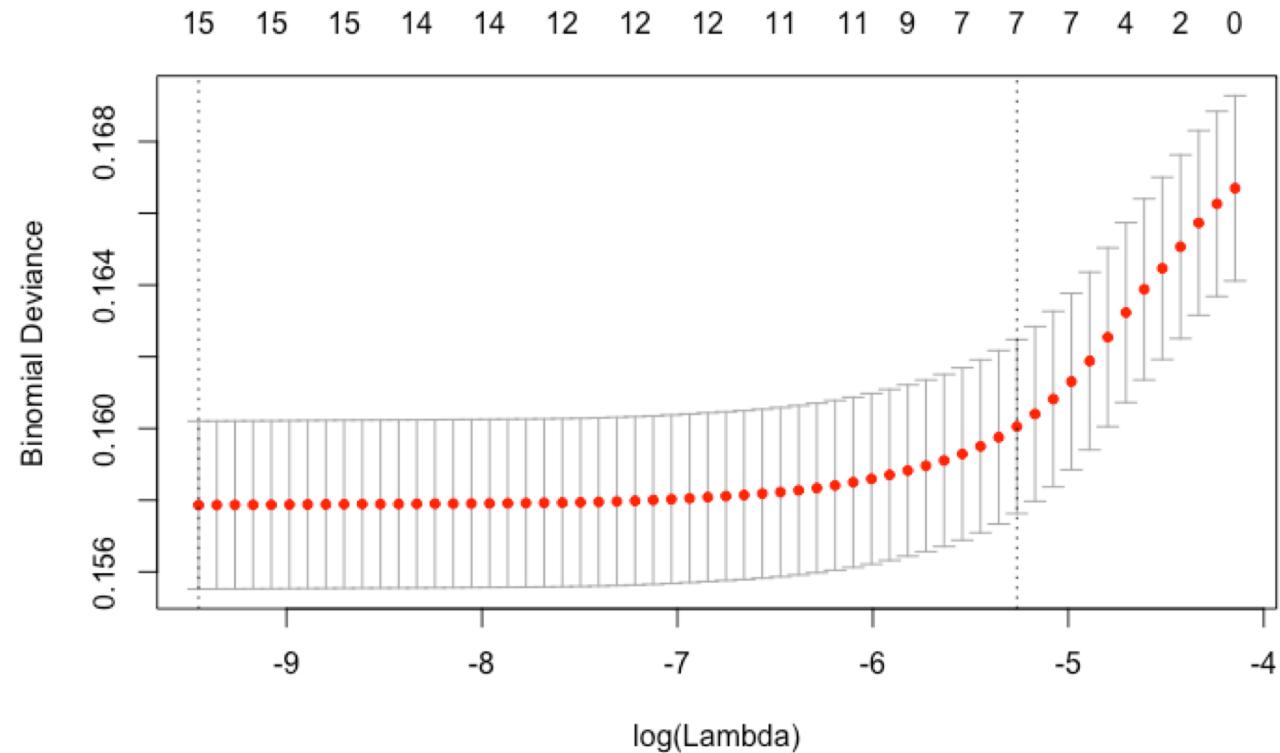
UKB 297K unrelated White British (5.1% CHD Case (22.5% Female))

Biomarker averages	Adj_statin control (female)	Adj statin Incident case (female)	+statin_adj
HDL	1.46 (1.60)	1.27 (1.44)	/1.053
LDL	3.78 (3.79)	3.99 (4.10)	/0.684
TG	1.78 (1.75)	2.19 (2.02)	/0.874
APOA	1.52 (1.61)	1.41 (1.54)	/1.065364
APOB	1.08 (1.07)	1.15 (1.17)	/0.721928
LipoA	34.31 (35.2)	38.4 (41.2)	/1.101954
Glucose	5.07 (5.09)	5.36 (5.38)	/1.028824
C-Reactive Prot	2.48 (2.61)	3.36 (4.16)	/1.2300281
BMI	27.4 (27.0)	28.6 (28.3)	
HBA1C	35.6 (35.4)	38.0 (38.5)	/1.0418022
T2D	4.2% (3.1%)	10.5% (8.9%)	
Systolic BP	141 (137.9)	151.4 (149.7)	+15 (BP meds)
Dialistolic BP	84.2 (82.3)	88.41 (85.9)	+10 (BP meds)
Cigs per day	1.05 (0.92)	2.33 (2.67)	NA

statin usage (16.2%)

Yu Fu; Sinnott-Armstrong et al. Genetics of 38 UKBB blood and urine biomarkers, (2019)

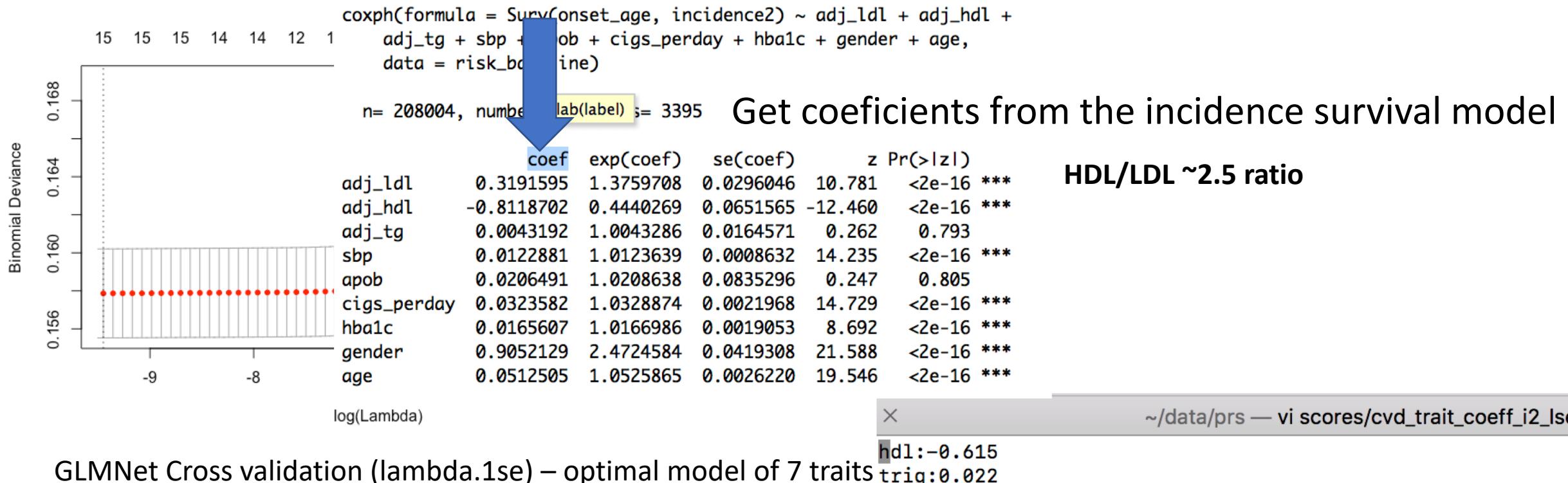
Shrink biomarkers Elastic.net and estimate coeffs



GLMNet Cross validation (lambda.1se) – optimal model of 7 traits

```
x ~/data/prs — vi scores/cvd_trait_01
hd1:-0.615
trig:0.022
ldl:0.136
sbp:0.013
hba1c:0.010
apob:0.079
cpd:0.0176
~
```

Shrunked biomarkers coefs from Cox survival



Redundancy check between selected biomarkers

- Correlation: (LDL, APOB) ~0.81, (HDL, Triglycerides) ~-0.43
- Variance inflation factor (VIF) seems ok* (< 3)

```
> car::vif(coxph_incident)
    adj_ldl      adj_hdl      adj_tg       sbp       apob cigs_perday
  2.538236    1.488951    1.384056    1.119472   2.328841   1.032309
    hba1c        age       gender
  1.059608    1.109322    1.174245
```

*Remove variable if VIF > 5

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2014.

An Introduction to Statistical Learning: With Applications in R. Springer Publishing Company, Incorporated.

Select SNPs from 7 GWAS summaries

- MAF>1%, clumped on LD blocks*/LDSC**
- ~17,000 SNPs
 - Annotate
 - intergenic, synonymous, nonsynonymous, start/stop sites,...
 - Map onto HLA, lipid, inflammation and nicotin curated pathways

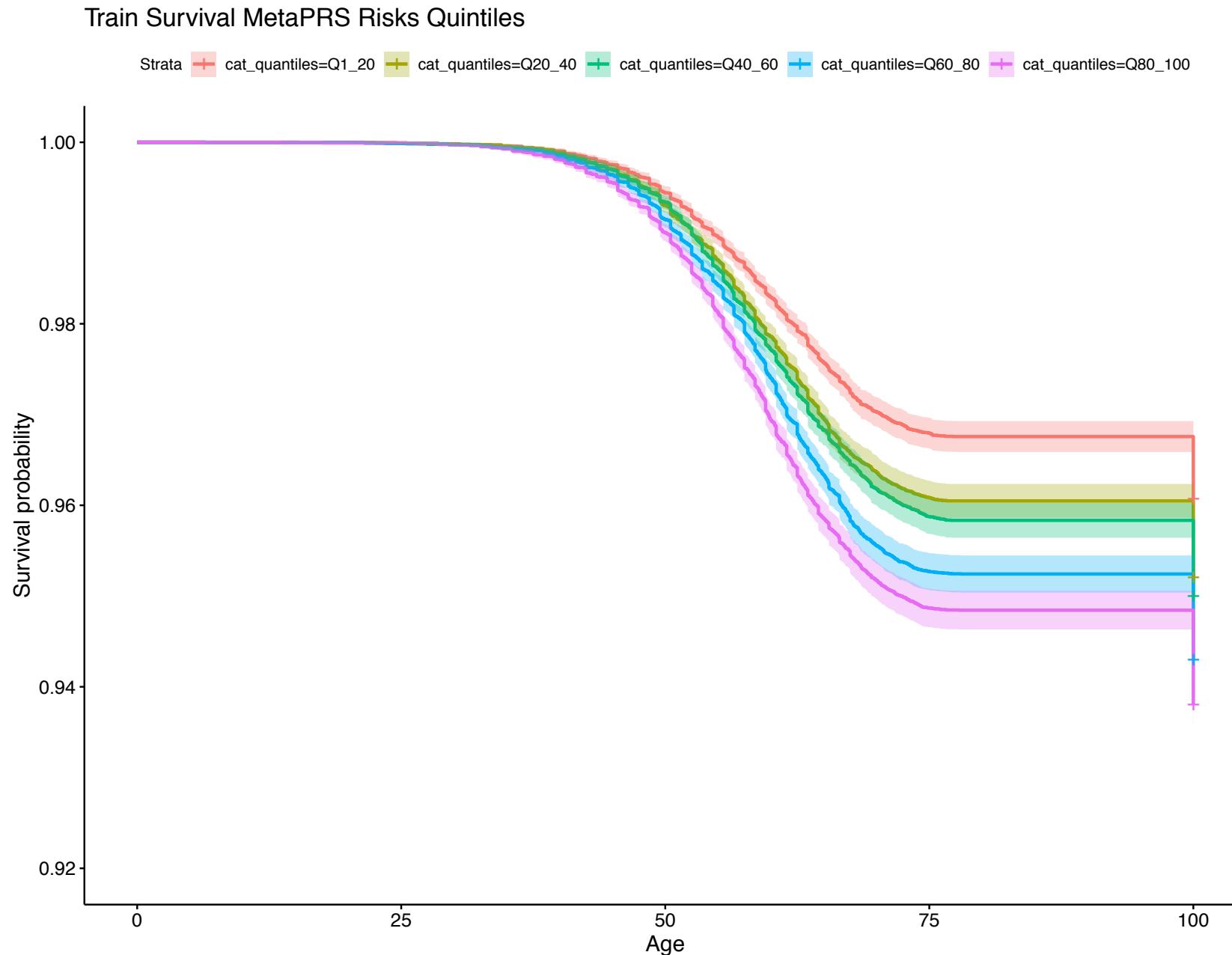
*<https://bitbucket.org/nygcresearch/ldetect/src/master/>

** Finucane et al, Alkes Group (https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2)

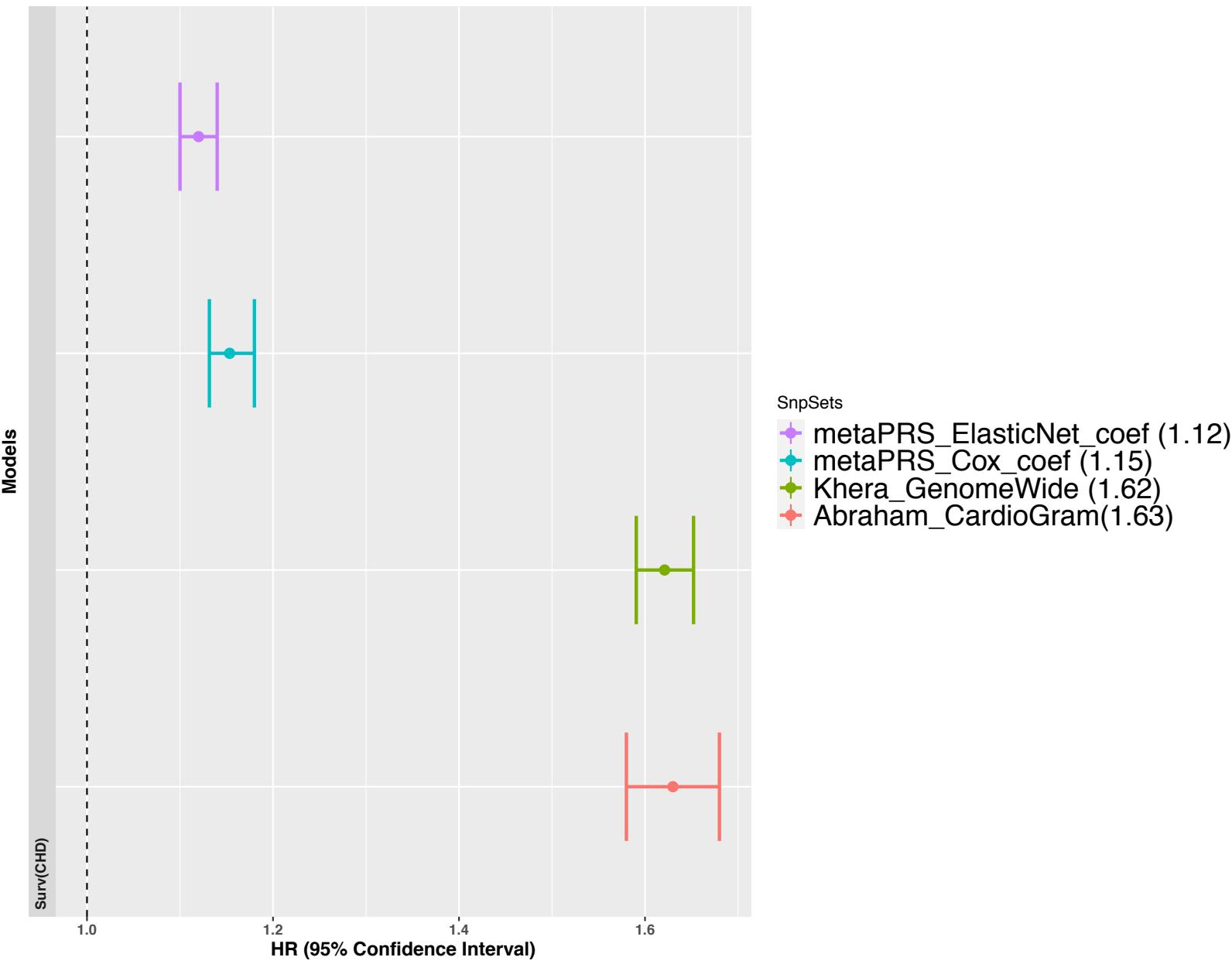
Meta Score SNPs with training data

- meta_prs scoring:
 - $\sum(\text{beta}) * \text{trait_coefficients}$
 - Remove less important snp:trait betas
 - Trim trait betas based on rankings [25,50,75,95]
 - Subject - $\sum n\text{Effect_allele} * \text{meta_score}$
- Cox survival coefficients performed better than ElasticNet coef

Model results - Higher the metaPRS score, probability of disease increases



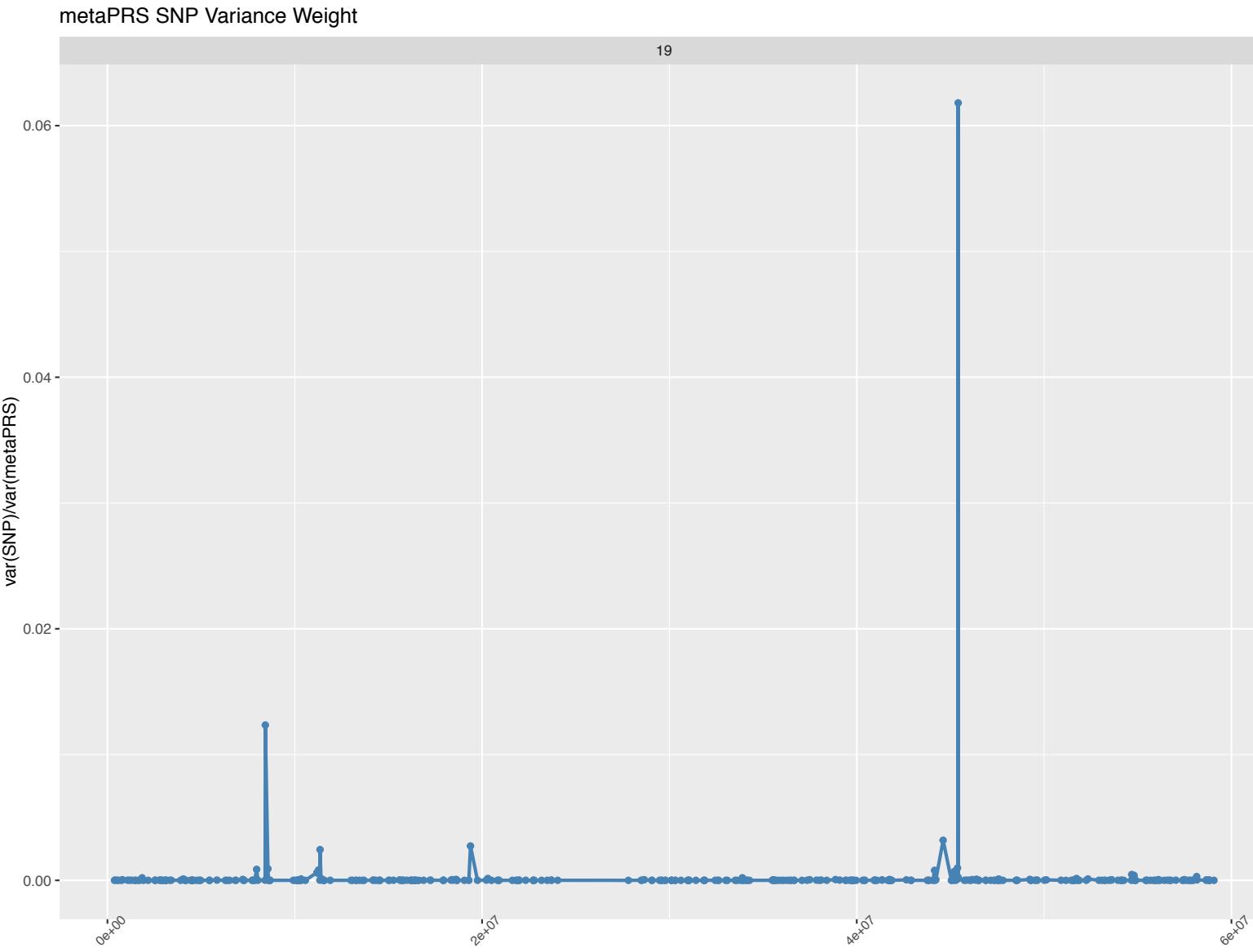
HR from survival(cadcase, onset) ~ prs[meta_elastic, meta_cox,
Khera,Abraham] + age + age2+ gender + pc1-10



SNP metaPRS contributions



Chr 19



Rsid rs429358 C:T missense

GWAS Significant for all 5 risks (beta sign): hdl+:ldl-:trig-:hba1c:apob

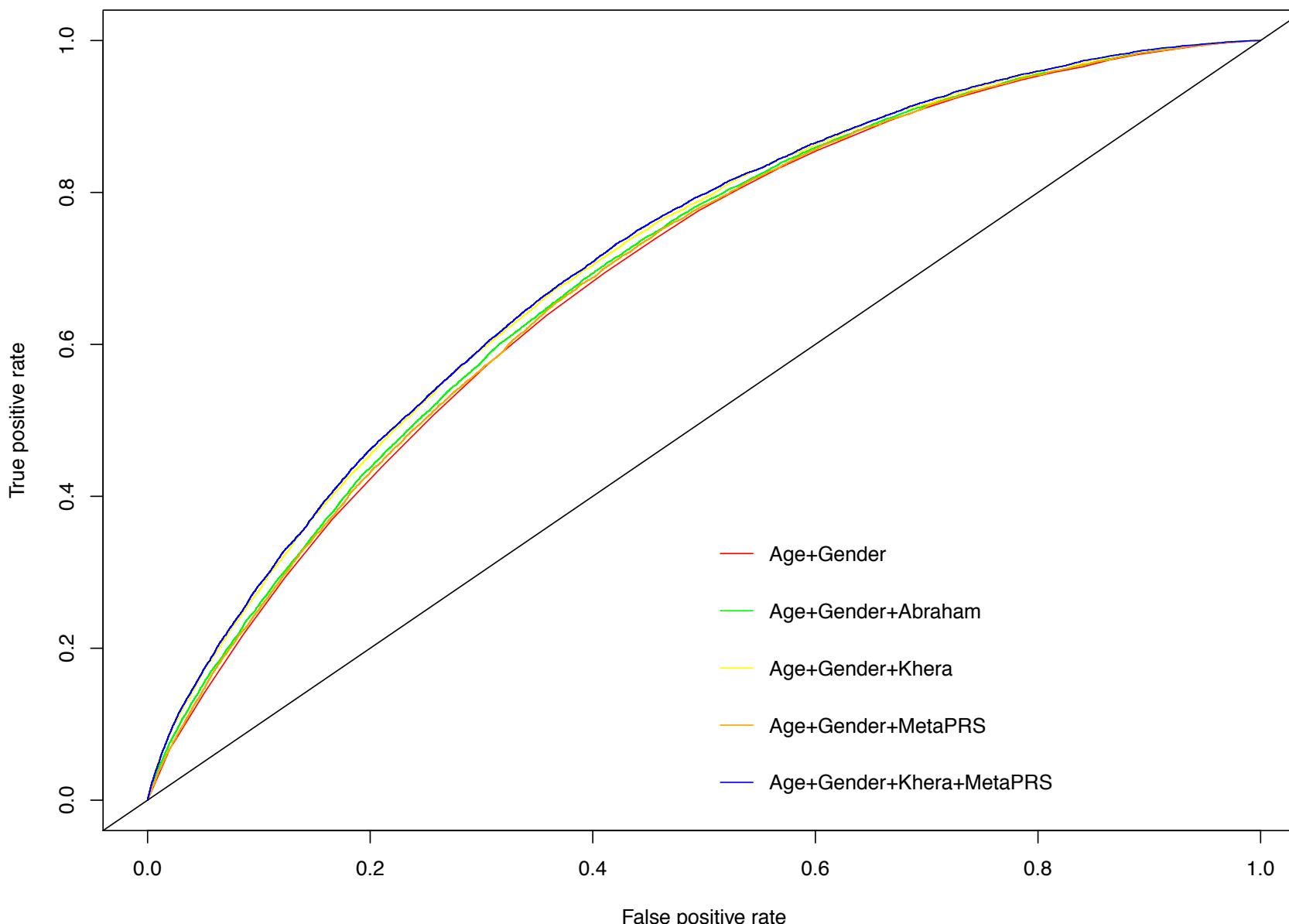
Abraham, Khera and Meta PRS integration summary

Surv (CAD, age_onset) ~normalize(PRS) + gender, age, age ² and pc1-10	HR	AUC	correlation	nSNPs
PRS_Abraham	1.64 ¹	0.73	na	49K
PRS_Khera	1.62 ¹	0.748	na	6M
PRS_meta PRS_meta_90% quantile	1.18 ^{1,2} 1.35	0.72	na	17K
PRS_Khera + PRS_Abraham	Khera 1.54 Abraham 1.28	0.749	0.23	
PRS_Khera + PRS_Meta	Khera 1.60 Meta 1.13	0.751	0.11	
PRS_Khera + PRS_Abraham + PRS_Meta	Khera 1.52 Abraham 1.26 Meta 1.12	0.753	(Abraham, PRS_Meta) 0.06	

Interesting that Khera+PRS_Meta has higher AUC than Khera+Abraham

1. P-val < 2e-16
2. Prelim. Sensitivity applied - 5 SNPs excluded

ROC

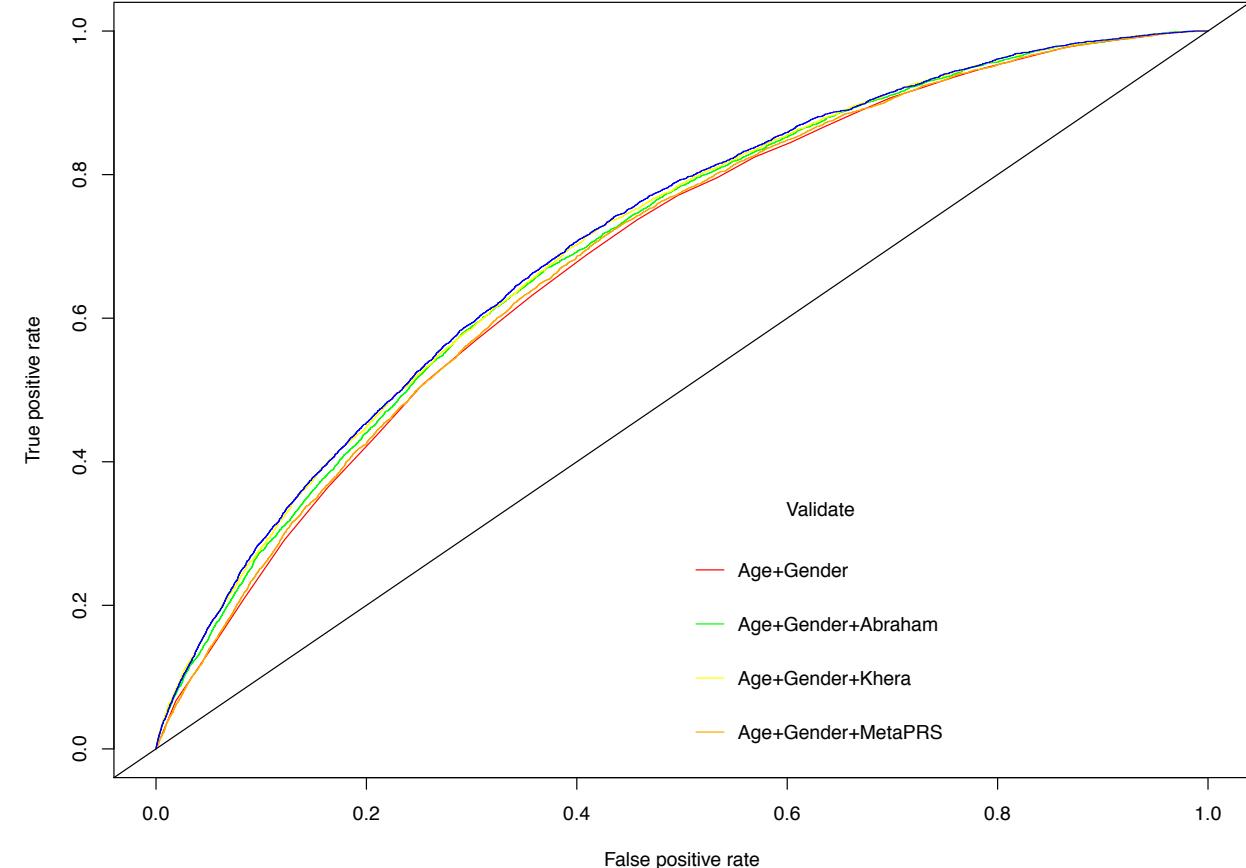
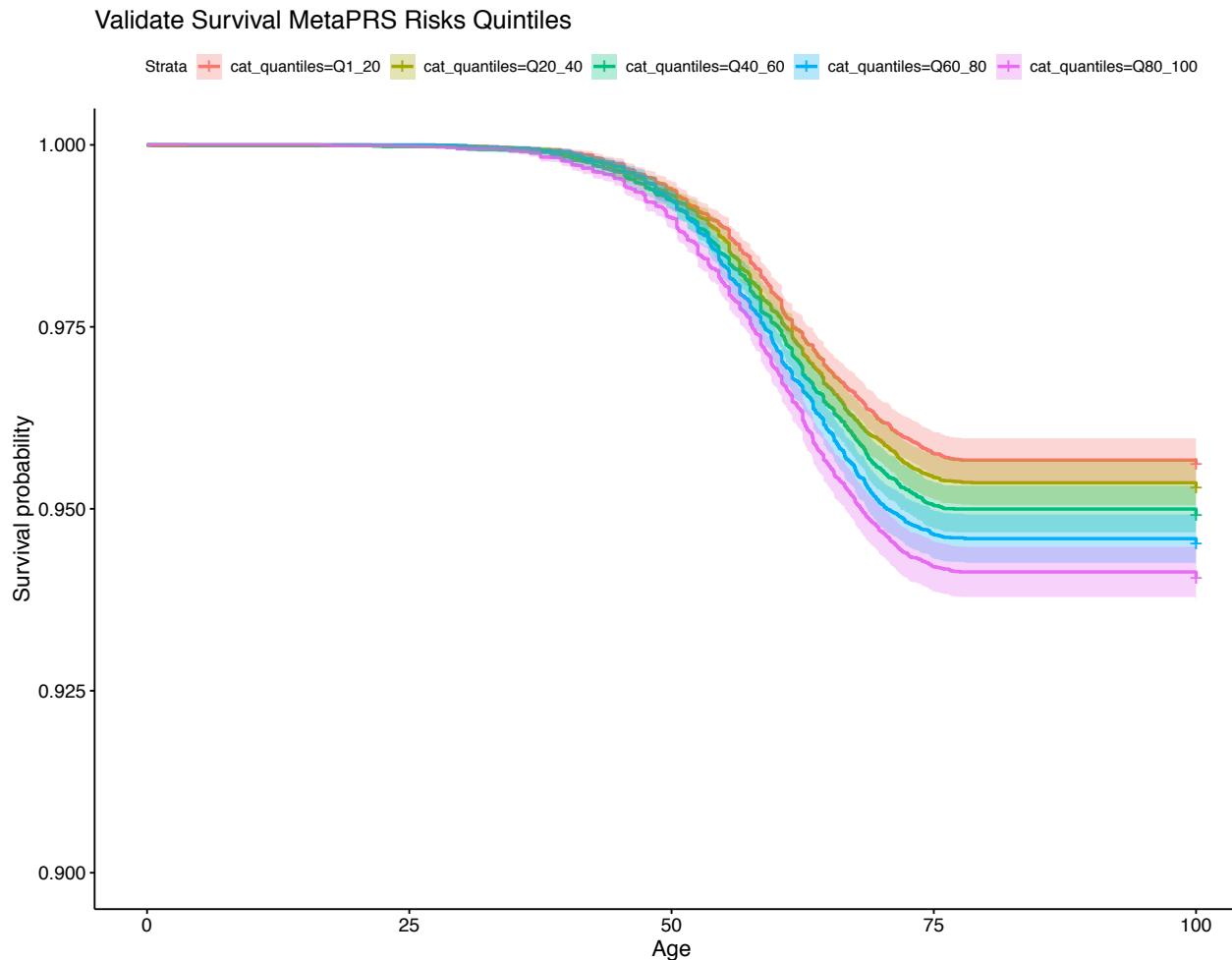


Validation data ~90K volunteers

- Independent of training data
 - Consistent proportion of cases (5.1%) (21.4% female)

Model~PRS_meta+gender +age+age2+pc1-10	HR (PVal)	Variance
Train	1.18 (< 2e-16)	0.51
Validate	1.12 (3.45e-15)	0.50

Validation: Survival and AUC consistent with training



Validation integration ~consistent with training

Surv (CAD, age_onset) ~normalize(PRS) + gender, age, age² and pc1-10	HR (PV)	AUC	Train AUC	corr
PRS_Abraham	1.43 (<e-16)	0.731	0.731	na
PRS_Khera	1.64 (<e-16)	0.745	0.748	na
PRS_meta	1.12 (3.45e-15)	0.711	0.715	na
PRS_Khera + PRS_Abraham	Khera 1.53 Abraham 1.23	0.751	0.749	0.36
PRS_Khera + PRS_Meta	Khera 1.63 Meta 1.10	0.746	0.751	0.04
PRS_Khera + PRS_Abraham + PRS_Meta	Khera 1.52 Abraham 1.23 Meta 1.10	0.752	0.753	(Abraham, PRS_Meta) 0.02

ToDos

- More tuning/sensitivity, benchmarking with this
 - SNPs with dubious combination of lipid beta signs
- Assess lipid/HLA pathways
 - Assess SNP2gene (Dey, Price, et al. 2020)
- Validate on UKBB (100K+) non-British whites
- FinnGen/FinnRisk cohorts

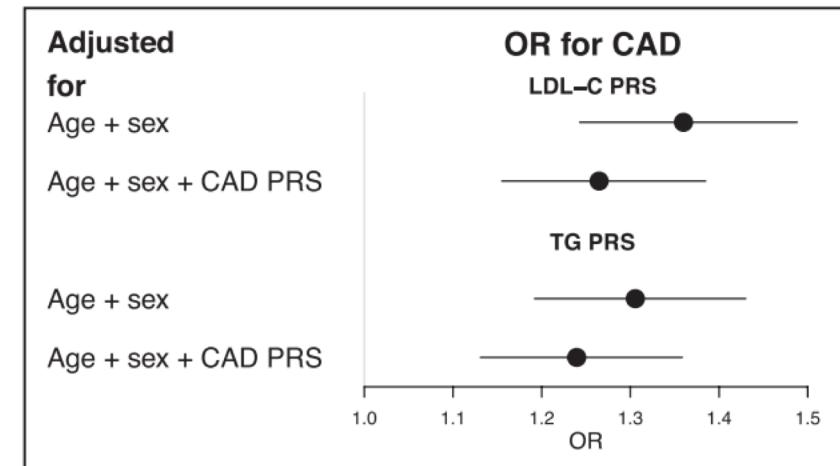


Figure 4. Odds ratios (ORs) for coronary artery disease (CAD)

Ripatti, Rämö, et al. Polygenic Hyperlipidemias and Coronary Artery Disease Risk FinnGen/FinnRisk (2020)

Acknowledgements

- Matti Pirinen
- Yu Fu
- Nina Mars
- Tuomo Kiiskinen
- Pietari Ripatti
- Samuli Ripatti
- UK Biobank
- TOPMed
- FinnGen
- Google Cloud Research
- Academy of Finland

