Supplemental Online Content

Helgason H, Eiriksdottir T, Ulfarsson MO, et al. Evaluation of large-scale proteomics for prediction of cardiovascular events. *JAMA*. doi:10.1001/jama.2023.13258

eAppendix 1. Methods

eAppendix 2. Results

eAppendix 3. References

eFigure 1. Age at Plasma Collection and Follow-Up in the Icelandic Population

eFigure 2. Cross-Validation for Selecting Penalization Strength for Fitting Protein ASCVD Risk Prediction Models

eFigure 3. Protein Risk Score Weights and Relationship With Polygenic Risk Scores and Age in the Primary Event Test Set

eFigure 4. Protein Risk Scores With Respect to Age and Sex in Test Set for Primary Event Population

eFigure 5. Polygenic Risk Scores for CAD and Stroke With Respect to Age and Sex in Primary Event Population

eFigure 6. AUCs and ROC Curves with Different Handling of Censoring and Competing Risk in the Primary Event Population

eFigure 7. Calibration in the Primary Event Population

eFigure 8. Reclassification in the Primary Event Population

eFigure 9. Predicted 10-Year Risk by Different Models in Three Separate Groups in the Primary Event Population

eFigure 10. Cumulative Rate of Cardiovascular Events in Primary Event Population

eFigure 11. AUCs and ROC Curves With Different Handling of Censoring in the Secondary Event Population

eFigure 12. Calibration in the Secondary Event Population

eFigure 13. Polygenic Risk Scores for CAD and Stroke and Protein Risk Score in the Secondary Event Population

eFigure 14. Cumulative Rate of Cardiovascular Events in Secondary Event Population

eFigure 15. Cumulative Event Rate of Cardiovascular Events Stratified by Protein Risk Scores in the Primary Event Population

eFigure 16. Prediction of Age and Sex Using Proteomics Data in the Primary Event Population

eFigure 17. Comparison of Models to Calculate Protein Risk Scores

eFigure 18. Cumulative Event Rate of Cardiovascular Events for Several Clinical Risk Scores in the Secondary Primary Event Population

eAppendix 4. List of eTables

This supplemental material has been provided by the authors to give readers additional information about their work.

eAppendix 1. Methods

Populations for Derivation of Polygenic Risk Score for CAD

Icelandic study: CAD case status (*N*=43,969) was assigned based on ICD code discharge diagnoses (ICD10: I20.0, I21.x, I22.X, I24.x, I25.x. or ICD9: 410.*, 411.*, 412.*, and 414.*) from Landspitali – The National University Hospital of Iceland (LUH), or based on the same ICD codes for CAD listed as the cause, or contributing cause of death, in the Icelandic death registry. The controls included 307,272 individuals recruited through different genetic studies at deCODE genetics and their relatives. The study was approved by The Icelandic Data Protection Authority and the National Bioethics Committee of Iceland (approvals no. VSNb201508003/03.01, VSNb2015010033/03.12, and VSNb2014100020/03.12 with amendments). All participating subjects donating biological samples signed informed consents. The personal identities of the participants and biological samples were encrypted with a third-party system approved and monitored by the Icelandic Data Protection Authority.

UK Biobank: CAD case status (*N*=55,282) was assigned based on ICD codes indicative of CAD (ICD10: I20.0, I21.x, I22.X, I24.x, I25.x. or ICD9: 410.*, 411.*, 412.*, and 414.*), obtained from primary or secondary diagnoses codes a participant has had recorded following every admission to hospital. The UK Biobank resource was accessed under Application Number '24711'.

The UK Biobank study is a large prospective cohort study of approximately 500,000 individuals (age range 40-69) from across the UK (recruited in England, Wales, and Scotland). Extensive phenotype and genotype data have been collected for participants, including ICD diagnosis codes. The UK Biobank data were obtained under application number 24898. Phenotype and genotype data were collected following an informed consent obtained from all participants. The North West Research Ethics Committee reviewed and approved UK Biobank's scientific protocol and operational procedures (REC Reference Number: 06/MRE08/65).

Danish study: This study is a part of the 'Genetics of cardiovascular disease' – a genome-wide association study on repository samples from Copenhagen Hospital Biobank (CHB) approved by The National Ethical Committee (1708829). CHB is part of the Danish National Biobank and has been approved by the Danish Data Protection Agency (general approval number 2012-58-0004, and local number: RH-2007-30-4129/ I-suite 00678). Patients included in CHB were informed about their right to refuse the use of their samples for research. CHB includes >330,000 adult patients (age above 18 years). The current study involves a targeted selection of patients with cardiovascular disease identified through health registries (the Danish National Patient Registry (NPR)). CAD case status (N=38,581) was assigned based on ICD codes indicative of CAD (ICD10: I20.0, I21.x, I22.X, I24.x, I25.x. or ICD9: 410.*, 411.*, 412.*, and 414.*). The control group included blood donors from "The Danish Blood Donor Study" (CVK-1700407) (*N* > 91,000)[1].

CARDIoGRAMplusC4D: Data on coronary artery disease (CAD) involving 60,801 cases and 123,504 controls have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG. The data includes a meta-analysis of GWAS studies of mainly European, South Asian, and East Asian, descent, reported in Nat Genet 2015 47:1121-1130.

FinnGen study: The clinical endpoint used in the analysis was Ischaemic heart disease, wide definition (I9_IHD; based on the following diagnostic codes ICD10: I20*, I21.*, I22.*, I23.*, I24.*, I25.* and ICD9 and

ICD8: 410.*, 411.*, 412.*, 413.*, 414.*) (https://www.finngen.fi/en/researchers/clinical-endpoints). The study included 25,366 cases and 176,899 controls. FinnGen is a large public-private partnership aiming to collect and analyze genome and health data from 500,000 Finnish biobank participants. The summary statistics for available phenotypes were imported in May 2021 from a source available to consortium partners (version 5; http://r5.finngen.fi).

The FinnGen study is an ongoing cohort study that integrates the genetic data generated from biobank samples and health-related data from social and health care registers. Detailed information such as participating biobanks/cohorts, genotyping, and data analysis is available on their website (https://www.finngen.fi/en/).

Populations for Derivation of Polygenic Risk Score for Stroke

We obtained genetic association results for ischemic stroke from the following datasets: deCODE (Icelandic study), UKBiobank, Finngen (freeze 5) and the Megastroke consortium (European participants excluding deCODE) and conducted a genome wide association meta-analysis including 60,092 cases and 1,096,793 controls to generate a PRS weights for the secondary event population (FOURIER). For the score used for the primary event population, the Icelandic study was excluded, resulting in a meta-analysis including 52,827 cases and 765,607 controls.

Icelandic study: Icelandic ischemic stroke cases (7,265), were identified from a registry of individuals diagnosed with ischemic stroke or TIA at LUH in Reykjavik. Control comprised 331,186 individuals recruited through different genetic projects at deCODE. Individuals with confirmed stroke (identified by cross-matching with hospital lists) were excluded as controls. The study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. All participants gave informed consent.

UK Biobank: In the UKBiobank study, the ischemic stroke cases (7,211) were selected using the ICD diagnosis I63 (ICD-10) primarily from hospital data codes, surgical records, and General Practice clinical records. The number of controls used were 423,727.

UK Biobank data were obtained under application number 24898.

FinnGen study: The clinical endpoint used in the analysis was Ischemic stroke excluding hemorrhages (I9_STR_EXH, n=10,551) defined by the ICD-9 (4330, 4331, 4339, 4340, 4341, 4349, 436) and ICD-10 codes (I63, I64) (see: https://www.finngen.fi/en/researchers/clinical-endpoints).

Megastroke consortium: Ischemic stroke cases of European ancestry from the Megastroke consortium[2] were included in the analysis (35,065 cases). Cases from deCODE were excluded from the Megastroke data set. The phenotype definition of ischemic stroke was based on clinical and imaging criteria.

Icelandic Proteomics Study

The study includes 39,155 Icelanders with measurements of protein levels in plasma. The plasma was sampled at deCODE from the years 2000 through 2006 (*N* = 23,474) and from the years 2010 through 2019 (*N* = 15,681). The older samples were mostly sampled through the Icelandic cancer project (ICP)[3] (*N* = 20,610) and most of the newer samples through the deCODE health study (dHS)[4] (*N* = 8,818) from the years 2016 through 2019. The rest of the samples came from various smaller projects (**eFigure 1**).

All participants who donated samples gave informed consent, and the National Bioethics Committee of Iceland approved the study, which was conducted in agreement with conditions issued by the Data Protection Authority of Iceland (VSN_14-015, VSN_15-130, and VSN_15-214). Personal identities of the participants' data and biological samples were encrypted with a third-party system (Identity Protection System), approved and monitored by the Data Protection Authority.

Protein measurements: The SomaScan v4 platform was used to measure the levels of around 5,000 proteins in each of the Icelandic participants as has been previously described[5]. At plasma collection, the participants were non-fasting. All plasma samples were analyzed in the time period from January to June 2019, with one thawing cycle for the majority of the samples during preparation. Changes in protein levels by time of plasma collection was estimated with linear regression for each aptamer individually. Only 106 out of the 5,284 SOMAscan aptamers showed a significant trend by time of sampling and the effects were modest; the maximum absolute trend was observed for GPD with a decrease of 0.0045 units per year.

In the data analysis, we excluded aptamers that did not target human proteins and those that were reported as deprecated or unreliable by SomaLogic, resulting in a total of 4,963 protein aptamers targeting 4,718 different proteins or combinations of proteins, (i.e., some aptamers targeted multiple proteins and some proteins were targeted by multiple aptamers). A total of 4,728 unique proteins, based on UniProt IDs, were targeted. All protein levels were log-transformed.

Baseline Features: All features were collected as close to the time of plasma collection as possible. Age was included as a decimal number calculated from the date of birth and time of plasma collection. BMI values were available within one year of plasma collection for 31,135 participants. For 6,094 participants where all BMI measurements were more than one year from the time of plasma collection, we used the median of all available BMI measurements. For 1,926 participants there were no available BMI measurements. BMI for these participants was imputed with the mean of the subset used for that particular analysis. Information on the diagnosis of type 2 diabetes was based on the first date of any of the following: Oral diabetes medication prescription from The Icelandic Prescription Medicine Register up to 2014; diagnostic codes ICD9 250 or ICD10 E11 from LUH, Register of Primary Health Care Contacts, or Register of Contacts with Medical Specialists in Private Practice; HbA1C levels at least 6.5% (LUH laboratory); or self-report. Individuals with type 1 diabetes and women with gestational diabetes were excluded. Where information about smoking was not available, it was estimated based on prediction from proteomics data. Information on hypertension treatment and statin use were obtained from the Icelandic Prescription Medicines Register. Since the register only goes back to 2003, we assumed that those with plasma sampled before the middle of 2003 who were receiving hypertension-lowering drugs during the first half of 2003 were already being treated at the time of plasma collection (*N* = 5,812). We used the proteomics data for those with plasma sampled after the middle of 2003 to estimate statin use for the older samples where statin use was uncertain (*N* = 1,879), our estimator had an AUC of 0.93 in a holdout test set. Information on missing data and imputation is provided in **eTable 1.**

Endpoints: The primary endpoint was the first hard atherosclerotic cardiovascular disease (ASCVD) event, defined as the first myocardial infarction (MI), coronary heart disease death, or fatal or non-fatal stroke. MI was assigned based on ICD diagnosis codes (ICD10: I21.x, I22.X, I23.x, I25.2. or ICD9: 410.*, 412) from the Icelandic health care system (LUH, or the Icelandic Cause of Death Register). Icelanders with stroke, admitted to LUH over the period 1983-2020, were identified from the hospital registry using the following ICD codes for ischemic stroke: ICD9: 433.0, 433.1, 433.2, 433.3, 433.8, 433.9, 434.0, 434.1, 434.9 and ICD10: I63.0, I63.1, I63.2, I63.3, I63.4, I63.5, I63.8; and intracerebral hemorrhagic stroke (ICD9: 431 and ICD10: I61.0, I61.1,

I61.2, I61.3, I61.4, I61.5, I61.6, I61.8, I61.9. Individuals with stroke probably caused by alcohol or drug dependence syndrome, trauma, or brain tumor were excluded. Coronary heart disease death was assigned based on ICD primary cause of death codes (ICD10: I20.x, I21.x, I22.x, I23.x, I24.x, I25.x, ICD9: 410.*, 411.*, 412.*, 413.*, 414.*) from the Icelandic Cause of Death Register. Dates of death from other causes, resulting in the end of the follow-up period, were obtained from the Icelandic Cause of Death Register.

Primary event population data set: The ASCVD protein risk score was developed using a subset of 13,540 subjects in the Icelandic proteomics data set. The subset consisted of participants aged 40 to 75 years who had not had previous MI (participants with codes ICD10: I25.2 and ICD9: 412 were assumed to have had MI at the time of plasma collection and were therefore always excluded) or stroke or undergone a percutaneous coronary intervention or coronary artery bypass graft. Participants, where there was uncertainty about whether events happened before or after plasma collection, were excluded. For all participants, follow-up information was available until the end of 2018 which was chosen as the end of the study. Because of limited follow-up information for the more recent measurements, only participants with plasma sampled from the years 2000 through 2006 were included (**eFigure 1**, **eTable 1**).

This data set of 13,540 subjects was split into training and test sets (training: about 70% of total, *N* = 9,522; test: about 30% of the total, *N* = 4,018). The derivation set was used for all model selection, hyperparameter tuning, and training of the final model while the test set was reserved for testing the prediction performance. Each feature was standardized with its mean and standard deviation in the derivation set before training and applying the predictor.

Secondary Event Population - FOURIER Trial

As part of the data collection for FOURIER, medical history of prior MI, stroke, peripheral arterial disease due to atherosclerosis (PAD), CAD, cerebrovascular disease, type 1 and type 2 diabetes, coronary artery bypass graft, congestive heart failure, and hypertension were collected. Common clinical chemistry tests were also performed on samples extracted from the subjects in the FOURIER trial. As for the Icelandic study, the SomaScan v4 platform was used to measure the levels of around 5,000 proteins in a subset of the trial participants. The data used in the joint analysis of protein risk scores, medical history, and laboratory tests corresponded to baseline information (i.e., from day one of the trial). At plasma collection, the participants in the FOURIER trial had to be fasting for at least 9 hours. All plasma samples were analyzed in the time period from August 2020 to January 2021, with one thawing cycle for all sample preparations.

The MACE endpoint was defined as the composite of cardiovascular death, myocardial infarction, or stroke. All potential endpoint events were adjudicated by a central clinical events committee led by the TIMI Study Group, whose members were unaware of the treatment assignment and lipid levels of the participants. Details of the definitions of the endpoints and further information about the trial protocol are provided in previous publications[6].

The study focused on individuals of European ancestry. In the analysis that included PRS data, genetic ancestry assignments by supervised ADMIXTURE[7] were used to classify whether an individual was of European ancestry. ADMIXTURE was run in supervised mode with five contemporary reference populations (EUR: Utah residents with Northern and Western European ancestry (CEU); EAS: Han Chinese in Beijing, China (CHB); SAS: Indian Telugu from the UK (ITU); AMR: Peruvian from Lima, Peru (PEL); AFR: Yoruba in Ibadan, Nigeria (YRI)). Individuals assigned less than 0.93 of the CEU component were deemed ancestry outliers and excluded from the analysis. Since genotypic data were not available for all the individuals with

proteomics data, self-reported ethnicity (white race, including both Hispanics and non-Hispanics) was used for the analysis that did not involve PRS data.

Development of Protein Risk Score for ASCVD

The ASCVD protein risk scores were developed in the derivation set using a Cox proportional hazards model[8] with a lasso penalty[9] by solving

$$\min_{\beta}(-\log L\,(\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1),$$

where $\boldsymbol{\beta}$ is a vector of variable coefficients, the parameter $\lambda > 0$ controls penalization strength (large λ correspond to greater penalization), and $\log L\,(\boldsymbol{\beta})$ is the partial likelihood function of the Cox model. The parameter λ was chosen to minimize mean partial likelihood deviance in the hold-out sets of ten-fold cross-validation in the derivation set. The glmnet[10] package in R was used to perform the fit.

Two versions of the protein scores were constructed. The first version was fitted using age, sex, and the 4,963 protein measurements (ProtRS). To ensure that the coefficients for age and sex, i.e., β_{age} and β_{sex}, would be non-zero, they were exempted from the penalization. This was done to prevent the model from trying to capture the effects of age and sex on the endpoint using the protein variables. For the chosen penalization strength, the model had 70 other non-zero coefficients β_k. The three models where at least one of the proteins NT-proBNP and MMP-12 was removed from the set of proteins during training, were fitted in the same manner as was done for the main protein score ProtRS.

Another version was fitted using only the 4,963 protein measurements ($\text{ProtRS}_{\text{unadj}}$). The resulting model included 199 non-zero coefficients β_k, making it considerably larger than the model trained with age and sex given. It is worth noting that small changes in penalization strength can affect model size considerably without strongly affecting the fit (**eFigure 2**).

After fitting the model, protein risk scores were calculated for everyone, according to the following formula;

$$\text{ProtRS} = \sum_{k=1}^{p} w_k x_k,$$

i.e., the linear part of the Cox model whose exponential represents the hazard ratio. Where $w_k = \beta_k, k = 1, \ldots, p$, are the non-zero coefficients for protein levels, x_k; i.e., we do not use age and sex to calculate the protein risk score (the coefficients and proteins for our main ProtRS are provided in **eTable 3**).

Our motivation for using the lasso is that it provides a well-established framework for dimensionality reduction that selects linear models systematically and objectively by fulfilling mathematical optimality criteria. This choice of method as a starting point in our analysis allows us to infer whether the underlying phenotype can be explained well by a subset of proteins. This could lead the way in a search for even sparser models with the hope of building better risk predictors and gaining insight into the biological processes at play.

Construction of Polygenic Risk Score for CAD and Stroke

The polygenic risk score for CAD (PRS_{CAD}) was generated using methods previously described[11]. The PRS was calculated using genotypes for about 630,000 autosomal markers that are included on the Illumina SNP microarrays to avoid uncertainty due to imputation quality. Linkage disequilibrium (LD) between markers was estimated using 14,938 phased Icelandic samples and this LD information was used to calculate adjusted effect estimates using LDpred[11,12]. The effect estimates used for PRS_{CAD} were based on a combination of five study cohorts for a primary event of CAD (*N* affected: 223,999; *N* controls: 1,123,388), ensuring that individuals that had been used in more than one study were only used once. Several PRSs assuming different fractions of causal variants were calculated (using a tuning parameter in LDpred) and the best one was selected based on a prediction of the corresponding disease in the Icelandic data that had been used in calculating the effect estimates for the two PRSs. The weights for the most predictive scores were then used to calculate PRS_{CAD}. Icelandic samples were not used in the construction of the PRS used in the primary event population in Iceland but were used in the construction of the PRS used for the secondary event population (FOURIER).

Development of Clinical Risk Factor Scores in Secondary Event Population

In the secondary event population, we considered two published secondary prevention risk scores, the updated Secondary Manifestations of ARTerial disease (SMART2) score[13] and the TIMI Risk Score for Secondary Prevention (TRS2P)[14]. The only variable in the SMART2 score we did not have information on was medical history for abdominal aortic aneurysms. For the SMART2 score, we considered both the score with the weights of the individual risk variables provided in the original publication[13] and a refitted score ($SMART2_{refit}$) based on estimating weights using a Cox proportional hazards model using individuals from the secondary event population that were separate from the test set (individuals in the placebo arm of the FOURIER trial that did not have proteomics data, *N* = 5,403, *N* events = 413).

The TIMI Risk Score for Secondary Prevention (TRS2P)[14] is a sum of indicator variables for congestive heart failure, hypertension, age over 75 years, diabetes mellitus, prior stroke, prior coronary artery bypass graft, PAD, renal dysfunction (estimated glomerular filtration rate < 60 ml x min^{-1} x 1.73 m^{-2}), and current smoking. Since not all individuals in our dataset had previous MI we also included previous MI in the sum as has been done before in a similar situation[15].

For calculating clinical risk scores, missing variable values were imputed using the mean values for continuous variables and zero for binary variables. At most 4% and 6% were missing for continuous and binary variables, respectively, except that 13% had no coronary artery bypass graft status data.

In addition to considering published scores, we derived an optimized clinical risk scores using a Cox model with lasso penalization for model selection in the secondary event populations. The derivation involved selection from 66 variables (listed in **eTable 15**) that included clinical information, lab measurements, age, sex, the number of prior MIs and strokes, and other variables. The optimized risk score model included 20 variables after being trained on the same set as was used in the refitting of variable weights for the SMART2 model.

Weights for models based on clinical risk factors in the secondary event population are provided in **eTable 2**.

Bootstrapped Lasso

To estimate the robustness of proteins selected by the lasso-penalized Cox method, we performed bootstrapping. Using resampling with replacement we sampled 1,000 different sets of 9,522 participants from the derivation set in the primary event population. In each set, we trained lasso-penalized Cox models for 20 different values of model size penalization parameter λ using all the protein measurements, age, and sex, where age and sex were excluded from penalization (i.e., forced into the model). For the λ closest to the λ used to train ProtRS, we counted how often each protein was included in the 1,000 different models.

eAppendix 2. Results

Reclassification in Secondary Event Population

The reclassification metrics for the addition of the ProtRS to a model with $SMART2_{refit}$ showed a significant increase for 2-year risk (IDI: 0.004, 95% CI: 0.002 to 0.007; category-free NRI: 0.182, 95% CI: 0.072 to 0.288) (**Table 3** and **eTable 8**). Adding both the ProtRS and PRSs to $SMART2_{refit}$ significantly improved classification accuracy for 2-year risk as measured by category-free NRI (0.339, 95% CI: 0.203 to 0.476) and the IDI index (0.007, 95% CI: 0.003 to 0.010) (**eTables 8** and **10**).

Analysis Clinical Risk Factor Scores in Secondary Event Population

Here we present additional analysis involving clinical risk factors in the secondary event population, focusing on the clinical scores described in the section **Development of Clinical Risk Factor Scores in Secondary Event Population** above (results using $SMART2_{refit}$ are in the main text).

Out of the 64 clinical variables considered, 50 had a significant correlation with the ProtRS and 16 were significantly associated with the MACE endpoint, after adjusting for age and sex (**eTable 15** and **16)**. When adjusting for ProtRS, only one of the clinical variables was significant (time since first CVD event). To investigate to what extent the risk predicted by ProtRS could be explained when considering all the clinical variables, we employed machine learning on data from the secondary event population to derive an optimized clinical risk score. In joint model with this risk score, the ProtRS remained significant with an adjusted HR of 1.36 per SD (95% CI: 1.22 to 1.50; *P* = 3.7e-9). The ProtRS also remained significant when considering two other clinical risk scores, the original SMART2 score and TRS2P score (**eTables 2** and **17** and **eFigure 18**).

Proteomics and Data Redundancy

Care has to be taken when interpreting the presence of specific proteins in the model. High correlations between levels of proteins (multicollinearity) and variable measurement accuracy of proteins can lead to models containing different sets of proteins that are equivalent in terms of prediction performance. This phenomenon is evident when randomly resampling from the derivation set and constructing a new ProtRS model for each resampling, where we observed great variability in the proteins selected. For example, out of the full set of 4,963 proteins, 3,077 were included at least once in a model and 424 different proteins were included in at least 10% of the models when resampling 1,000 times. The two proteins that stood out in terms of frequency of occurrence in the resampling models were NT-proBNP, which was the only protein that was included in all of the models, and MMP-12, which was included in more than 90% of the models. Five additional proteins, encoded by the genes *LILRA6*, *CA10*, *WFDC2*, *EPOR*, and *PHGDH* were included in more than 75% of the models (**eFigure 3**).

NT-proBNP and MMP-12 are among the proteins with the most significant association in univariate analysis with the ASCVD endpoint in the primary event population (**eTable 3**). To investigate how important these two proteins are for proteomics-based risk prediction, we trained ProtRSs with NT-proBNP or MMP-12 removed from the set of 4,963 plasma proteins (age and sex were included in the model). This increased the number of proteins included from 70 to 137 when we excluded NT-proBNP, 108 when we excluded MMP-

12, and 181 when we excluded both. In the primary event population, the Harrell's C-index decreased from 0.79 to 0.78 when we excluded both proteins and from 0.64 to 0.61 in the secondary event population, mostly driven by NT-proBNP (**eTable 12**).

Excluding Information on Age and Sex During Protein Score Training

To determine how information about sex and age affects the structure and performance of ProtRSs, we also considered a different approach for creating the protein score. In this second approach, a linear protein risk predictor ($ProtRS_{unadj}$) was trained without providing information about the subjects' age and sex. In the absence of age and sex information during training, the machine learning algorithm selected proteins and weights into the model that capture the effect of these covariates concerning ASCVD risk; this is perhaps not surprising given that levels of proteins in plasma can predict age[16,17] and sex very well (see **Predicting Sex with Proteomics Data** and **Predicting Age at Plasma Extraction with Proteomics Data**).

When we calculated the two ProtRSs in the primary event population test set (*N* = 4,018), we observed that in a linear model, age and sex explained 55% of the variance of $ProtRS_{unadj}$ while they explained considerably less of the variance of ProtRS, or 22%. The score $ProtRS_{unadj}$ had highly significant association with age and sex, consistent with ASCVD incidence increasing with age and males having a higher risk of ASCVD than females in the general population. On the other hand, ProtRS only had a suggestive association with sex and a significantly smaller age effect than observed for $ProtRS_{unadj}$ (**eFigure 4**).

There is considerable difference in the number of proteins in the models, where $ProtRS_{unadj}$ is comprised of 199 proteins, whereas ProtRS is considerably sparser with 70. The two ProtRSs were highly correlated (ρ = 0.87, 95% CI: 0.86, 0.88) and had 61 proteins in common. As for ProtRS, the two proteins with the largest standardized weights for $ProtRS_{unadj}$ were NT-proBNP and MMP-12.

A comparison of event rates for ASCVD events for the two protein scores in linear risk predictors including age and sex as covariates did not show better performance for ProtRS than for $ProtRS_{unadj}$. We also observed that adding age and sex as covariates to a risk model with only $ProtRS_{unadj}$ did not improve the risk prediction. This indicates that including age and sex information during the training step might not necessarily result in a better proteomics-based risk predictor, although the model consists of fewer proteins (**eFigure 15** and **eTable 14**).

Predicting Sex with Proteomics Data

To predict sex using proteomics measurements, we used a logistic regression model with lasso penalty, trained with all 4,963 protein measurements as candidate features. The penalization strength was selected to minimize binomial deviance in 10-fold cross-validation. The resulting model had an ROC AUC of 0.999 in the test set; out of the 4,018 participants in the test set only 9 were wrongly predicted, or 0.22%, when using a threshold of 0.5 in a binary classifier based on the model (**eFigure 16**). The model included 244 protein measurements and the 5 proteins with the highest absolute weight in the model corresponded to the genes *KLK3*, *LEP*, *PZP*, *SLITRK4*, and *IL3RA*.

Predicting Age at Plasma Extraction with Proteomics Data

To predict age at plasma extraction, we trained a linear regression model with lasso penalty using all 4,963 protein measurements as candidate features. The penalization strength was chosen to minimize the mean squared error in 10-fold cross-validation. In the test set, 90.0% of the variance in age at plasma extraction

could be explained by the model (**eFigure 16**). The model included 1,305 proteins and the 5 proteins with the highest weights in the model corresponded to the genes *PTN*, *ADAMTSL1*, *CHRDL1*, *SCARF2*, and *CD93*.

Protein Risk Scores and Stability Selection

We also tried constructing protein risk scores using smaller, carefully selected sets of proteins. The proteins were chosen using stability selection, where we ordered the proteins by how often they appeared in models for the ten highest λs in our lasso bootstrapping experiment (see **Bootstrapped Lasso** above). We trained three different Cox models using the proteins selected in at least 50% of models (8 proteins), the 10 proteins selected most frequently, and the 20 proteins selected most frequently. The models were trained with l2-penalization, where penalization strength was selected with cross-validation. We also tried training without any penalization but did not find that to improve the models.

These experiments with smaller models composed of 8, 10, or 20 proteins showed that protein risk scores utilizing only a few proteins can achieve good prediction performance. In the primary event population, all three protein scores added significantly to prediction with only the clinical risk factors where the smallest model had the smallest increase in prediction performance with the Harrell's C-index increasing by 0.012 (95% CI: 0.002 to 0.023). In the secondary event population, only the 10-protein model resulted in significant discrimination improvement over $SMART2_{refit}$ where the C-index increased by 0.016 (95% CI: 0.002 to 0.031) (**eTable 13**).

Alternative Methods for Developing Protein Risk Scores

In light of their good performance in previous studies[15], we also tried non-linear methods to create a ProtRS. As a non-linear approach for creating ProtRSs, we trained a Cox proportional hazards model with gradient boosted trees using XGBoost[18]. We used random search with cross-validation to adjust the following tree hyperparameters: max tree depth, learning rate, data instance subsampling, variable subsampling, minimum split loss, and minimum child weight. To further prevent overfitting, we used early stopping to determine the number of boosting iterations. The non-linear and linear models were compared using five-fold cross-validation on the derivation set and we found the performance of both models to be similar (**eFigure 17**).

Analysis of Protein Risk Scores Trained on Secondary Event Population

For comparison to protein risk scores trained on primary event populations, we developed protein risk scores using the placebo arm of the secondary event population (*N* = 6,307, *N* events = 432). As in the case of the primary event population, we developed two scores using Cox models with a lasso penalization: one with age and sex adjustment ($ProtRS_{adj,FOURIER}$) and the other without ($ProtRS_{unadj,FOURIER}$). For developing the scores we used 70% of the data for training the protein weights and reserved the remaining 30% for testing. The number of proteins in the scores was 72 for $ProtRS_{unadj,FOURIER}$ and 85 for $ProtRS_{adj,FOURIER}$. These two scores were highly correlated in the test set ($\rho > 0.96$) and had similar hazard ratios in Cox models with age and sex as covariates and using the MACE endpoint ($ProtRS_{unadj,FOURIER}$ HR = 1.54, *P* = 8.5e-7; $ProtRS_{adj,FOURIER}$ HR = 1.59, *P* = 2.4e-7). Those hazard ratios were not significantly different from the hazard ratios for the scores trained on the Icelandic primary event population ($ProtRS_{unadj}$ HR = 1.61, *P* = 1.5e-6; ProtRS HR = 1.48, *P* = 1.4e-5) (**eTable 14**). In the subset of the secondary event population that was used for testing (*N* = 1,863),

$ProtRS_{adj,FOURIER}$ had a Harrell's C-index (C-index = 0.620, 95% CI: 0.573 to 0.673) that was similar to the one for the score derived in the primary event population (C-index = 0.611, 95% CI: 0.556 to 0.666). This indicates that the difference in the C-indices observed for the the primary and secondary event populations for the score trained in the primary event population is probably not due FOURIER being an external validation cohort but rather due to differences between primary and secondary event populations.

Protein Risk Scores When Sex and Age Information is Optional

When generating ProtRS, adjustment was done for age and sex by forcing these variables into the Cox model while using lasso-penalization to select proteins along with their weights into the score (see **Development of Protein Risk Score for ASCVD**, where the regression coefficients for age and sex are excluded from the penalization). To investigate the added informational content of age and sex for cardiovascular risk prediction when proteomics data is available, we trained a model where the algorithm was given the option to include the age and sex variables in the model as opposed to forcing them in; this was done by adding a lasso-penalty to their model coefficients as well. When this model was trained on a secondary event population training set (see **Analysis of Protein Risk Scores Trained on Secondary Event Population**), this resulted in a linear predictor consisting of 70 proteins and the sex variable but the age variable was not selected (i.e., the lasso algorithm set its coefficient to zero); for downstream analysis, we used these 70 proteins along with their weights to form the score $ProtRS_{adjsex,FOURIER}$. In the secondary event population test set, the score $ProtRS_{adjsex,FOURIER}$ was highly correlated with the other two protein scores, $ProtRS_{unadj,FOURIER}$ and $ProtRS_{adj,FOURIER}$, that were both also trained on the secondary event population ($\rho > 0.96$) and had a similar hazard ratio to them (HR for $ProtRS_{adjsex,FOURIER}$ = 1.54, 95%CI: 1.30, 1.83; *P* = 6.3e-7).

eAppendix 3. References

1. Hansen TF, Banasik K, Erikstrup C, et al. DBDS Genomic Cohort, a prospective and comprehensive resource for integrative and temporal analysis of genetic, environmental and lifestyle factors affecting health of blood donors. *BMJ Open*. 2019;9(6):e028401. doi:10.1136/bmjopen-2018-028401

2. Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50(4):524-537. doi:10.1038/s41588-018-0058-3

3. Rafnar T, Thorlacius S, Steingrimsson E, et al. The Icelandic Cancer Project – a population-wide approach to studying cancer. *Nature Reviews Cancer*. 2004;4(6):488-492. doi:10.1038/nrc1371

4. Ivarsdottir EV, Benonisdottir S, Thorleifsson G, et al. Sequence variation at ANAPC1 accounts for 24% of the variability in corneal endothelial cell density. *Nature Communications*. 2019;10(1):1284. doi:10.1038/s41467-019-09304-9

5. Ferkingstad E, Sulem P, Atlason BA, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet*. 2021;53(12):1712-1721. doi:10.1038/s41588-021-00978-w

6. Sabatine MS, Giugliano RP, Keech AC, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *New England Journal of Medicine*. 2017;376(18):1713-1722. doi:10.1056/NEJMoa1615664

7. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*. 2011;12(1):246. doi:10.1186/1471-2105-12-246

8. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1972;34(2):187-202. doi:10.1111/j.2517-6161.1972.tb00899.x

9. Tibshirani R. The Lasso Method for Variable Selection in the Cox Model. *Statistics in Medicine*. 1997;16(4):385-395. doi:10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3

10. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*. 2010;33:1-22. doi:10.18637/jss.v033.i01

11. Kong A, Frigge ML, Thorleifsson G, et al. Selection against variants in the genome associated with educational attainment. *Proc Natl Acad Sci U S A*. 2017;114(5):E727-E732. doi:10.1073/pnas.1612113114

12. Vilhjálmsson BJ, Yang J, Finucane HK, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *The American Journal of Human Genetics*. 2015;97(4):576-592. doi:10.1016/j.ajhg.2015.09.001

13. Hageman SHJ, McKay AJ, Ueda P, et al. Estimation of recurrent atherosclerotic cardiovascular event risk in patients with established cardiovascular disease: the updated SMART2 algorithm. *Eur Heart J*. 2022;43(18):1715-1727. doi:10.1093/eurheartj/ehac056

14. Bohula EA, Bonaca MP, Braunwald E, et al. Atherothrombotic Risk Stratification and the Efficacy and Safety of Vorapaxar in Patients With Stable Ischemic Heart Disease and Previous Myocardial Infarction. *Circulation*. 2016;134(4):304-313. doi:10.1161/CIRCULATIONAHA.115.019861

15. Rousset A, Dellamonica D, Menuet R, et al. Can machine learning bring cardiovascular risk assessment to the next level? A methodological study using FOURIER trial data. *European Heart Journal - Digital Health*. Published online November 15, 2021:ztab093. doi:10.1093/ehjdh/ztab093

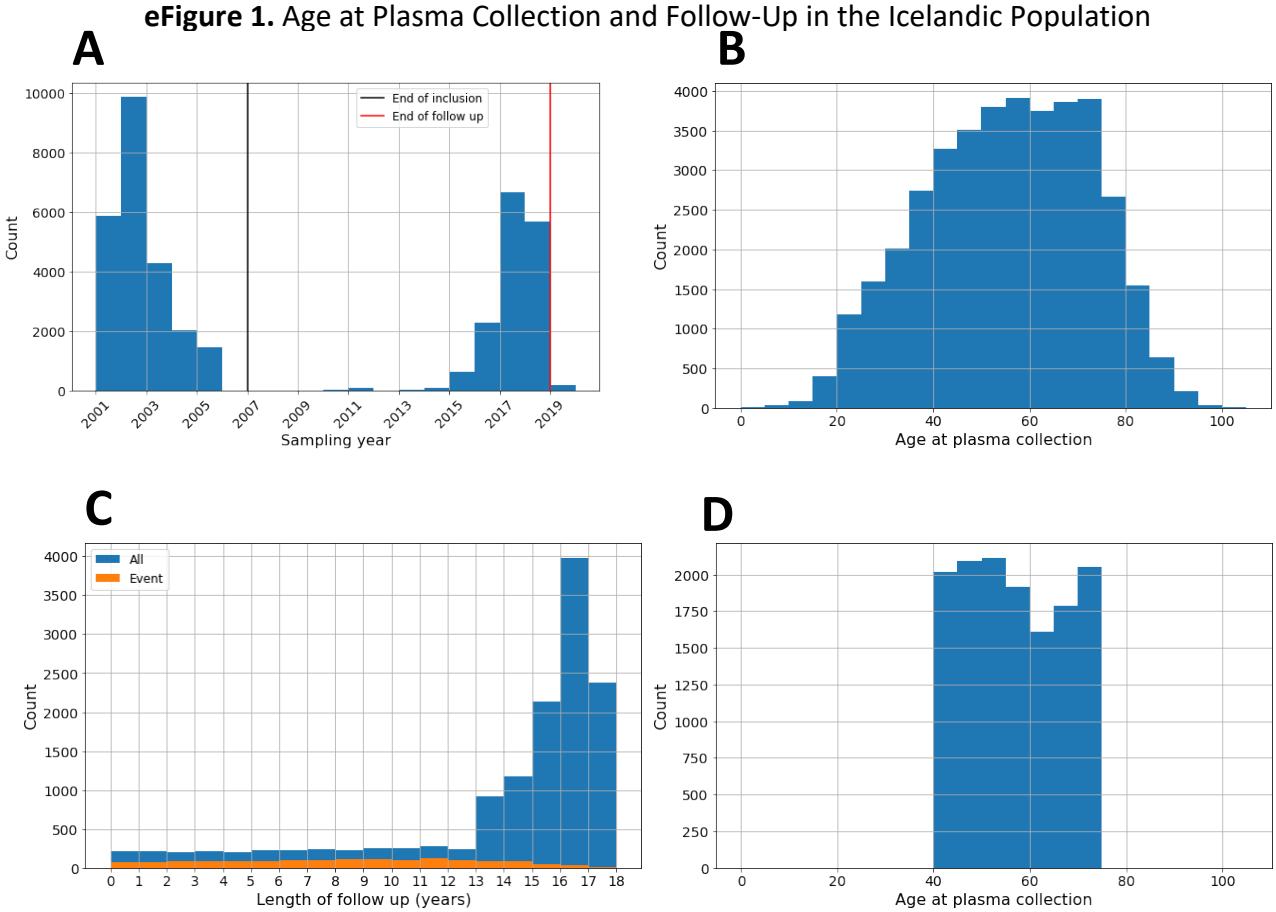
16. Tanaka T, Biancotto A, Moaddel R, et al. Plasma proteomic signature of age in healthy humans. *Aging Cell*. 2018;17(5):e12799. doi:10.1111/acel.12799

17. Lehallier B, Gate D, Schaum N, et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nature Medicine*. 2019;25(12):1843-1850. doi:10.1038/s41591-019-0673-2

18. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. Association for Computing Machinery; 2016:785-794. doi:10.1145/2939672.2939785

19. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine*. 2013;32(30):5381-5397. doi:10.1002/sim.5958

eFigure 1. Age at Plasma Collection and Follow-Up in the Icelandic Population



Panel A: Shown is a histogram showing the time of plasma collection of all 39,155 participants in the Icelandic study. The plasma was sampled at deCODE from the year 2000 throughout 2006 (*N* = 23,474) and from the year 2010 throughout 2019 (*N* = 15,681) (see **eMethods** for further details).

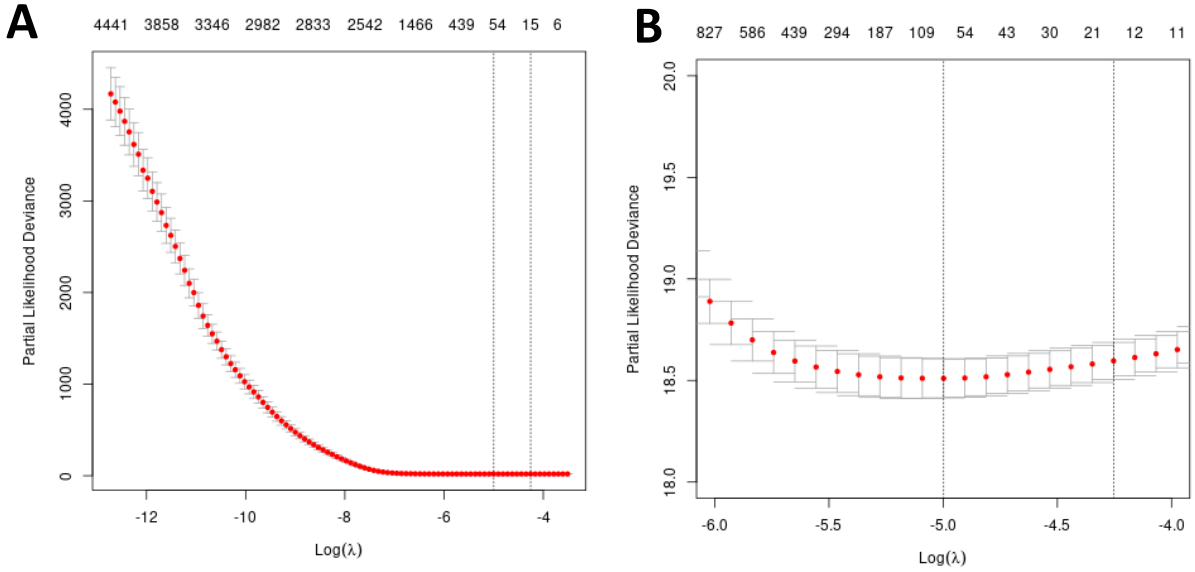
The black vertical line at the end of year 2006 corresponds to the end of inclusion in the primary event population train and test sets. Follow-up information was available for all participants until the end of 2018 (red vertical line in figure); this year was chosen as the end of the study.

Panel B: Shown is a histogram of the age at sample collection of all 39,155 participants in the Icelandic study. The age count is shown in five-year bins from 0 – 110. The mean age is 55.4 (SD: 17.1).

Panel C: Shown is a histogram of the follow-up time of the 13,588 participants included in the primary event population (blue bars). The mean follow-up is 14.0 (SD: 4.4) years. Most participants have follow-up until the end of the follow-up period at the end of 2018. The end of follow-up is 12 years after end of inclusion (see **Panel A**), therefore most participants have more than 12 years of follow up. Follow-up can also end with an event occurring (*N* = 1,522). The orange bars in the figure show separately the length of follow-up for the group that experience an event during the study. Events occur at a similar rate through most of the follow-up. During every year of follow-up, a median of 90.5 events occurs. Event rates drop for the longest follow-up since only a part of the participants have the full-length follow-up. The mean length of follow-up until event is 8.4 (SD: 4.5) years. End of follow-up because of death from other causes than cardiovascular (*N* = 2,389) is similarly distributed as follow-up until an event occurs, i.e., evenly where follow-up is available for everyone and drops for the longest follow-up times. A median of 137 deaths from other causes occurred during every year of follow-up and the mean follow-up until death from other causes is 8.8 (SD: 4.9) years.

Panel D: Shown is a histogram of the age at sample collection of the 13,588 included in the primary event population. The age count is shown in five-year bins from 0 – 110 but due to the inclusion criteria for the population all participants are between 40 – 75 years. The mean age is 57.2 years (SD: 10.2).

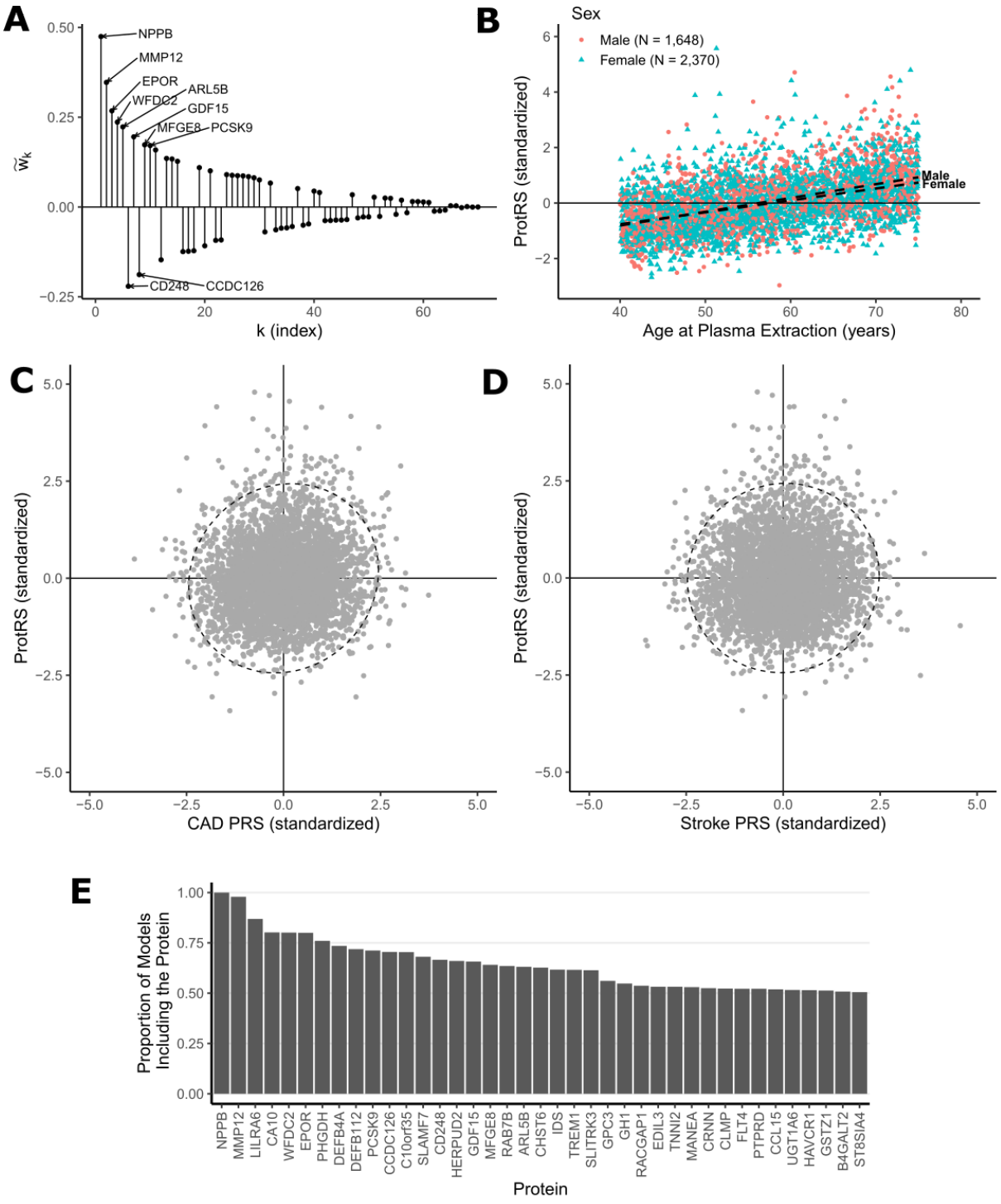
eFigure 2. Cross-Validation for Selecting Penalization Strength for Fitting Protein ASCVD Risk Prediction Models



The figure demonstrates how the penalization strength λ for fitting the ProtRS using all the protein measurements, age, and sex, where age and sex are excluded from the penalization, is selected based on the partial likelihood deviance (PLD) in ten-fold cross-validation. The red dots represent the mean PLD for each λ that was used and the bars represent one standard deviation in each direction for the cross-validations. The top x-axis shows how the number of nonzero coefficients changes with the penalization strength.

For each panel, the vertical dotted line to the left indicates the λ that gave the lowest mean PLD; this is the penalization strength λ_{min} that was used to develop the protein risk score. The vertical line to the right indicates the highest λ that gave a mean PLD within one standard deviation from the PLD corresponding to λ_{min}. **Panel A** shows the PLDs for all 100 tested λs while **Panel B** shows in close up the PLDs for the λs closest to λ_{min}.

eFigure 3. Protein Risk Score Weights and Relationship With Polygenic Risk Scores and Age in the Primary Event Test Set



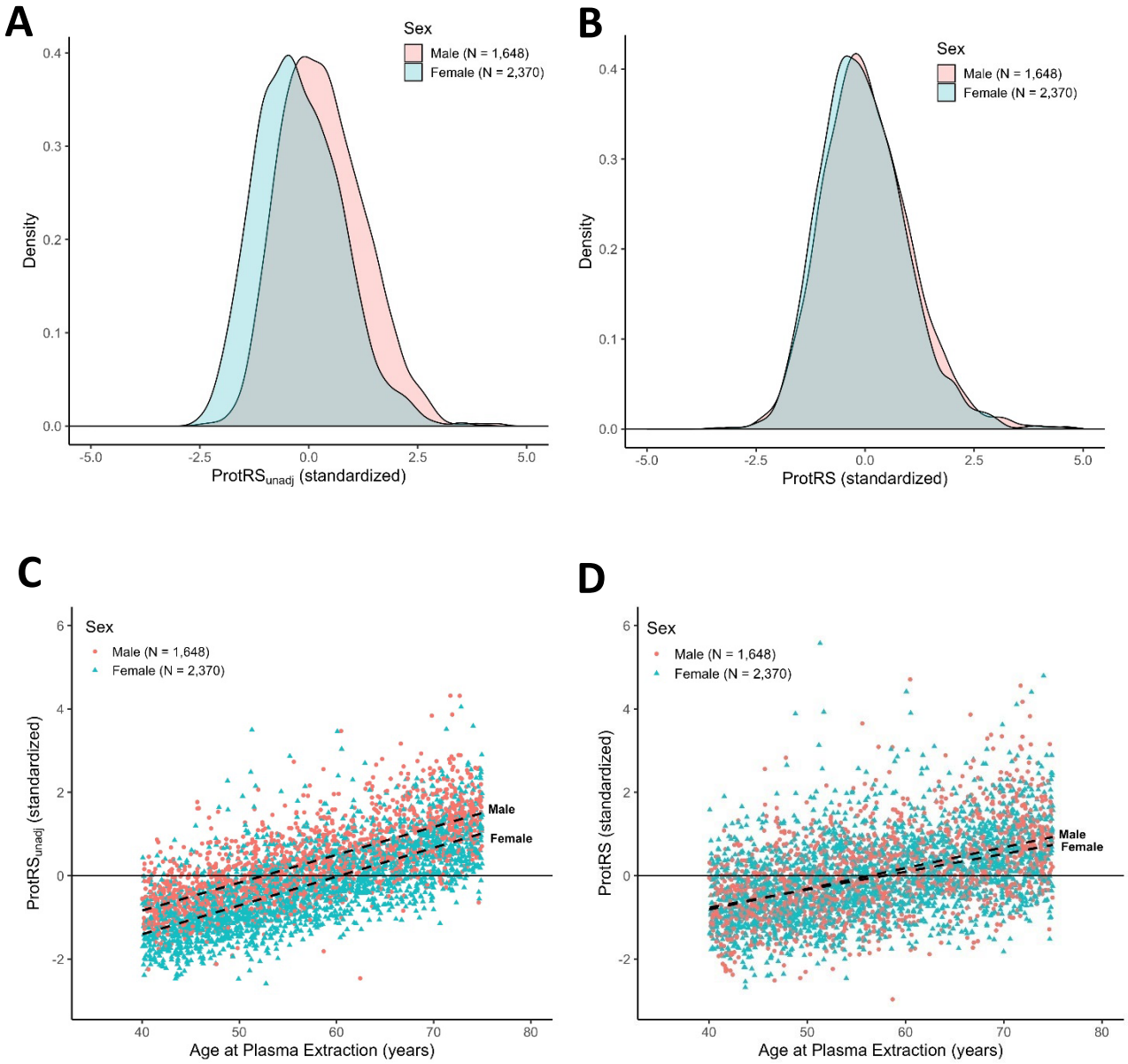
Panel A shows a plot of protein score weights in the formula for ProtRS. The plotted weights correspond to standardized protein measurements that have been normalized with the sum of squares over all protein weights in the score. The weights are ordered by their absolute value in the model and the ten proteins with the greatest absolute weight are indicated with a label with the corresponding gene name for the protein

Panel B shows the behavior of ProtRS with age and sex in the primary event population test set of 4,018 individuals (*N* female = 2,370 (blue); *N* male: 1,648 (red)). The broken lines show a positive trend with age for males and females that almost coincide.

Panels C and **D** are based on proteomics data in the primary event population test set. ProtRS and PRSs were standardized so that they had a mean zero and standard deviation of one within the test set**. Panel C** shows a scatterplot for ProtRS and PRS for CAD (Pearson correlation = 0.08, 95% CI: 0.05, 0.11) and **Panel D** shows a scatterplot for ProtRS and PRS for stroke (Pearson correlation = 0.02, 95% CI: -0.01 to 0.05).

Panel E: The panel shows the frequency of inclusion of the different proteins in multiple ASCVD risk models. The models are Lasso penalized Cox models using age, sex, and the 4,963 proteins as candidate features, where age and sex were excluded from the penalization, trained on 1,000 different resamplings of the derivation data. Out of the twenty tried penalization strengths, these figures show results for the λ closest to λ_{min} used to train ProtRS. Shown is the proportion of models that include the 39 proteins that are included in at least 50% of the models; the labels on the x-axis are the gene names corresponding to the proteins. N-terminal pro-BNP (gene name *NPPB*) is the only protein that is included in every model and MMP-12 (gene name *MMP12*) is the only other protein included in more than 90% of the models.

eFigure 4. Protein Risk Scores With Respect to Age and Sex in Test Set for Primary Event Population



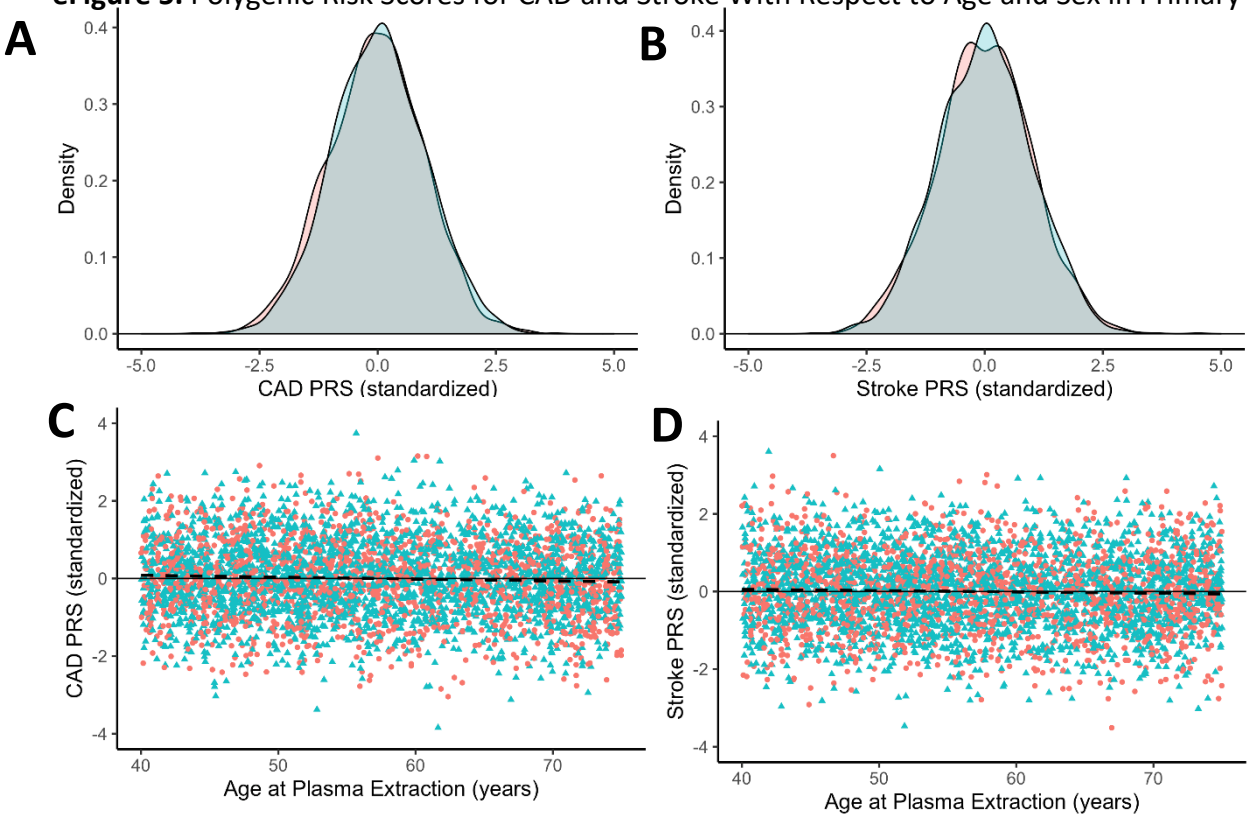
These figures are based on proteomics data in the test set of 4,018 individuals (*N* female = 2,370; *N* male: 1,648) and show plots of protein risk scores for cardiovascular events calculated with the formula developed in the derivation set using lasso-penalized Cox model, where ProtRS corresponds to the model that takes age and sex into account by including them as covariates and ProtRS$_{unadj}$ is only based on proteomics data. The protein risk scores were standardized so that they had mean zero and standard deviation of one within the test set.

Panel A corresponds to ProtRS$_{unadj}$ and shows protein score density estimates for females and males, where the density for females is shifted to the left by β = 0.53 SD (95% CI: 0.49, 0.57 SD, *P* = 2.8e-124). On **Panel B**, corresponding to ProtRS, we see that there is almost no difference in the protein score densities for females when information about sex and age was taken into account during construction of the score (β = 0.07 SD for males, 95% CI: 0.02, 0.13, *P* = 0.013).

Panel C corresponds to ProtRS$_{unadj}$ and shows the behavior of the protein score with age. The broken lines show positive trend with age for males ($\beta_{age,male}$ = 0.067 SD per year, 95% CI: 0.064, 0.070 SD per year) and females ($\beta_{age,female}$ = 0.070 SD per year, 95% CI: 0.066, 0.072 SD per year); there was not a significant difference in the age trends between the genders (*P* interaction = 0.27). On **Panel D**, corresponding to ProtRS, we see that the trend with age has attenuated when information about age and sex was taken into

account and the broken lines corresponding to the sex-specific fits almost coincide ($\beta_{age,males}$ = 0.050 SD per year, 95% CI: 0.046, 0.054 SD per year; $\beta_{age,females}$ = 0.043 SD per year, 95% CI: 0.039, 0.047 SD per year; *P* interaction = 0.013).

eFigure 5. Polygenic Risk Scores for CAD and Stroke With Respect to Age and Sex in Primary Event Population



The plots on **Panels A** and **B** correspond to the test set of 4,018 individuals in the primary event population (*N* female = 2,370; *N* male: 1,648). The PRSs were standardized so that they had a mean of zero and standard deviation of one within the set.

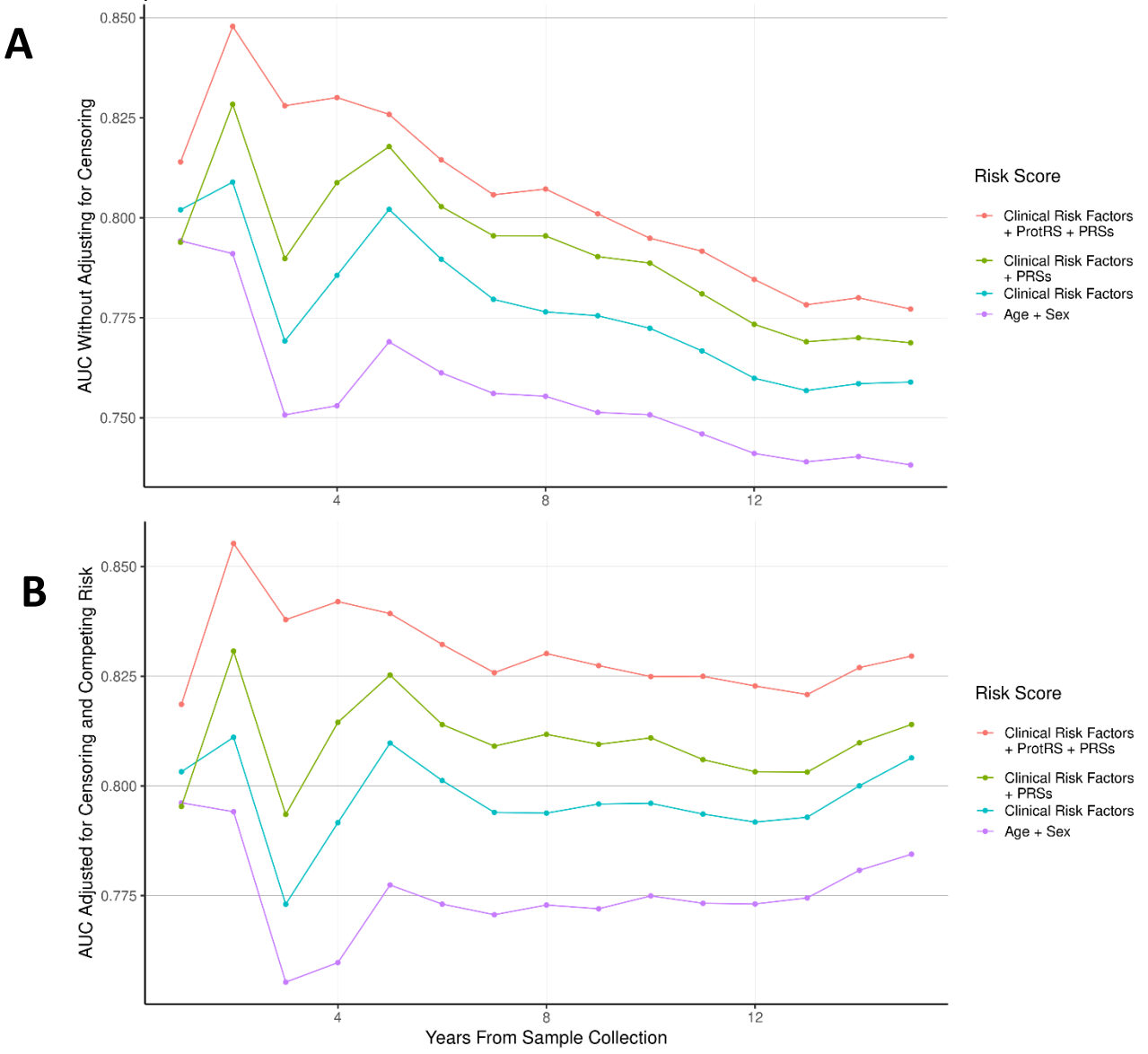
Panel A shows density estimates for PRS_{CAD} in the test set. There was not a statistically significant shift in density between females and males (β_{Sex} = 0.06 SD for females, 95% CI: -0.01, 0.12 SD; *P* = 0.073).

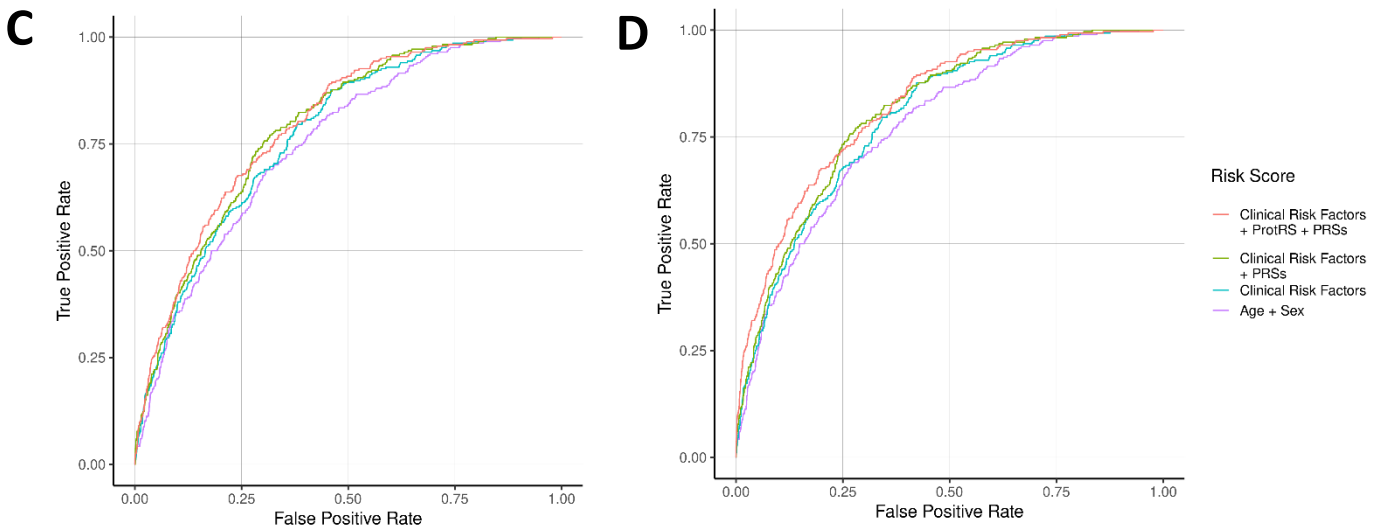
Panel B shows density estimates for PRS_{Stroke} in the test set. There was not a statistically significant shift in density between females and males (β_{Sex} = 0.01 SD for females, 95% CI: -0.06, 0.07 SD; *P* = 0.86).

Panel C shows the behavior of PRS_{CAD} with age at plasma extraction in the test set. The broken line showed slight negative trend with age (β_{age} = -0.0048 SD per year, 95% CI: -0.0079, -0.0018 SD per year; *P* = 0.0019); no difference in trends was observed when stratifying on sex (*P* interaction = 0.091).

Panel D shows the behavior of PRS_{Stroke} with age at plasma extraction in the test set. The broken line shows a marginal negative trend with age (β_{age} = -0.0034 SD per year, 95% CI: -0.0065, -0.0004 SD per year; *P* = 0.028); no difference in trends was observed when stratifying on sex (*P* interaction = 0.86).

eFigure 6. AUCs and ROC Curves with Different Handling of Censoring and Competing Risk in the Primary Event Population



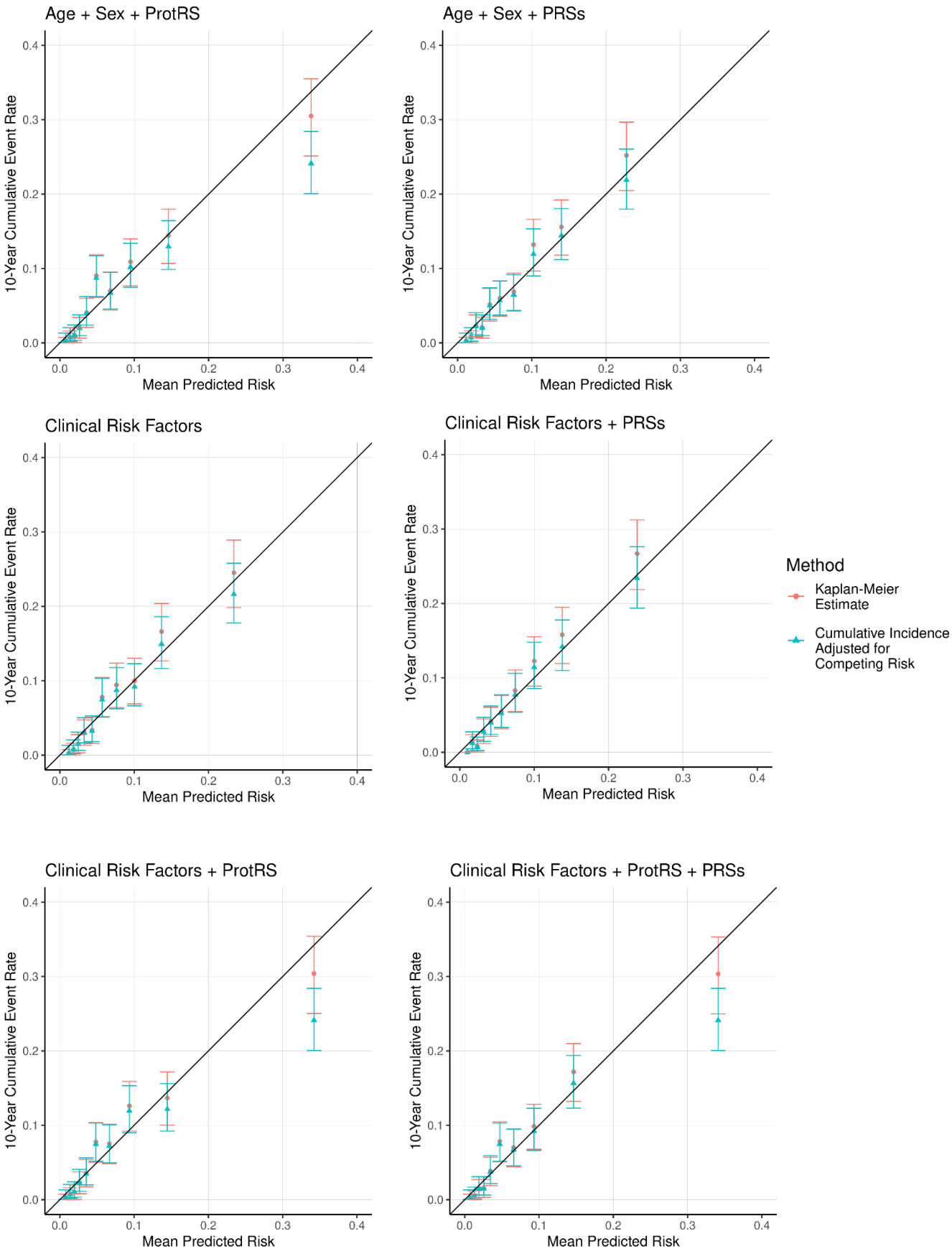


Shown are the discrimination performance of the ProtRS, PRSs and clinical risk factors in the primary event population test set (N = 4,018, N events = 465). AUCs and ROC curves were both calculated with and without adjustment for censoring and competing risk. In the adjustment, competing events (i.e., death from non-ASCVD causes) were not included as controls when calculating the specificity and inverse probability of censoring weighting was used. In the curves without adjustment, all non-events before each timepoint were taken as controls. The censoring and competing risk adjusted AUCs and ROC curves were calculated using the timeROC package in R[19].**Panels A** and **B** show the AUC for the events within 1-15 years. In **Panel A** the simple binary AUC is shown with no adjustment for censoring while in **Panel B** we adjust for censoring and competing risk.

Panels C and **D** show the ROC curves for 10-year risk prediction without (**Panel C**) and with (**Panel D**) adjustment for censoring and competing risk.

eFigure 7. Calibration in the Primary Event Population

Shown are predicted 10-year risk and observed 10-year event rate in the primary event population test



dataset (*N* = 4,018, *N* events = 465). The event rates were estimated in deciles of predicted risk with Kaplan

Meier method (red) and using cumulative incidence that accounts for competing risk from death from other causes than coronary heart disease (blue). The estimates are show for different combinations of risk scores where the baseline survival is estimated in the same dataset as we are estimating calibration in, i.e., the primary event population test dataset. All the observed models seem to be mostly well calibrated except for the top 10% at greatest risk whose risk seems to be consistently overestimated by all the models, especially when we account for competing risk of death from other causes.

eFigure 8. Reclassification in the Primary Event Population

A

		Clinical risk factors	Clinical risk factors + ProtRS		
			< 20.0%	>= 20.0%	Total number of participants (%)
Free of ASCVD events	ASCVD events	< 20.0%	183	40	223 (78.5)
		>= 20.0%	11	50	61 (21.5)
		Total number of participants (%)	194 (68.3)	90 (31.7)	284 (100.0)
	Dead from non-ASCVD causes	< 20.0%	215	93	308 (86.0)
		>= 20.0%	6	44	50 (14.0)
		Total number of participants (%)	221 (61.7)	137 (38.3)	358 (100.0)
	Alive and free of ASCVD events	< 20.0%	3163	91	3254 (96.4)
		>= 20.0%	61	61	122 (3.6)
		Total number of participants (%)	3224 (95.5)	152 (4.5)	3376 (100.0)

B

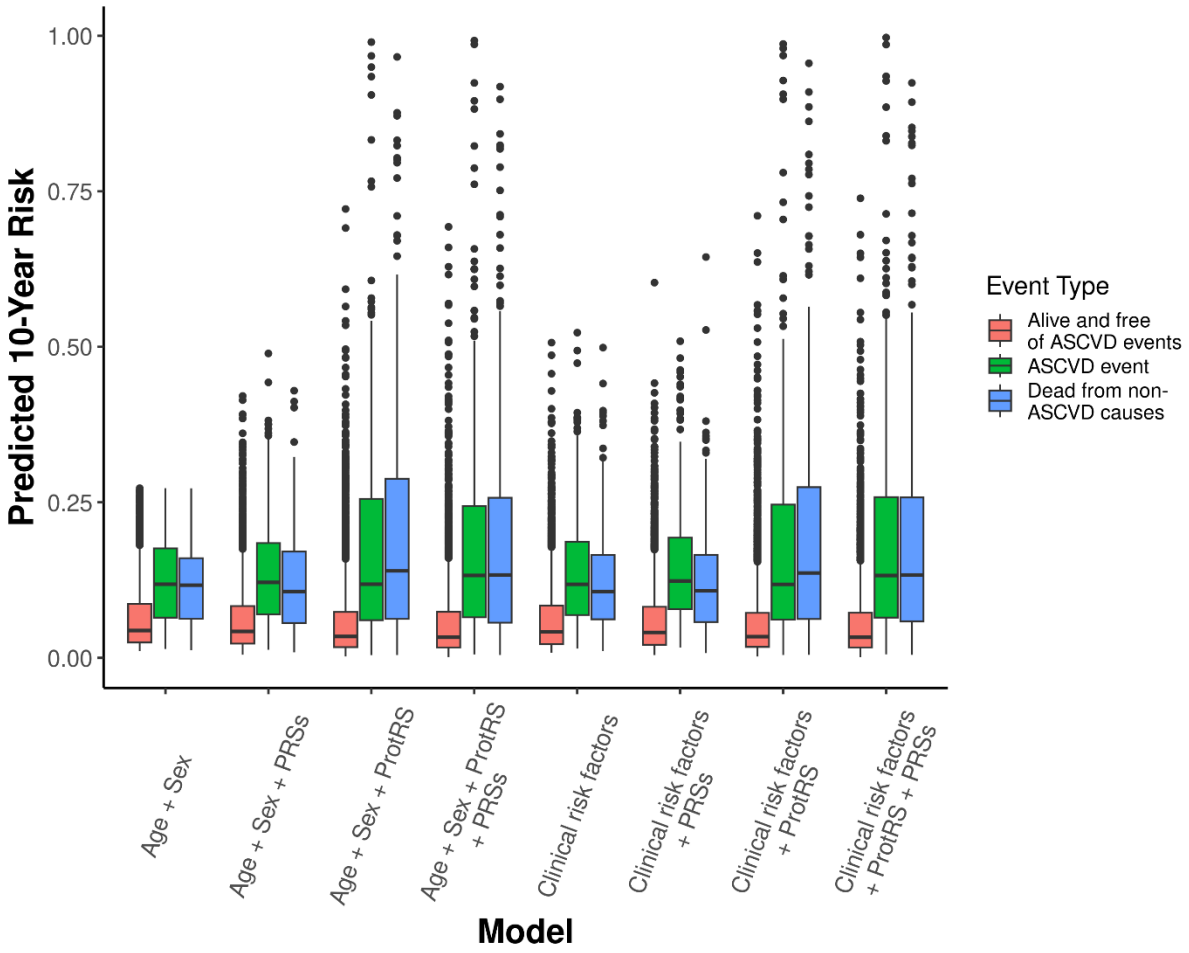
		Clinical risk factors	Clinical risk factors + ProtRS + PRSs		
			< 20.0%	>= 20.0%	Total number of participants (%)
Free of ASCVD events	ASCVD events	< 20.0%	179	44	223 (78.5)
		>= 20.0%	12	49	61 (21.5)
		Total number of participants (%)	191 (67.3)	93 (32.7)	284 (100.0)
	Dead from non-ASCVD causes	< 20.0%	219	89	308 (86.0)
		>= 20.0%	13	37	50 (14.0)
		Total number of participants (%)	232 (64.8)	126 (35.2)	358 (100.0)
	Alive and free of ASCVD events	< 20.0%	3155	99	3254 (96.4)
		>= 20.0%	67	55	122 (3.6)
		Total number of participants (%)	3222 (95.4)	154 (4.6)	3376 (100.0)

The figure shows reclassification results when ProtRS and PRSs are added on top of clinical risk factors (age, sex, statin use, hypertension treatment, type 2 diabetes, BMI, and smoking status at the time of plasma collection) in the primary event population test set (*N* = 4,018, *N* events = 465, *N* events within 10 years = 284). The results are shown for predicted 10-year risk using a 20.0% risk threshold to identify people at very high risk. Three groups are considered: (i) those with ASCVD event within 10 years, (ii) those who die from other causes than ASCVD within 10 years, and (iii) those who survive 10 years without ASCVD event.

Panel A shows a reclassification table for the addition of ProtRS on top of the clinical risk factors. The total categorical net reclassification improvement (NRI) is 0.071 (0.022 to 0.121) and 0.093 (0.047 to 0.143) when excluding those who die from non-ASCVD causes. NRI for ASCVD events is 0102 (0.054 to 0.152), -0.031 (-0.039 to -0.023) for ASCVD non-events (groups (ii) and (iii)), and -0.009 (-0.016 to -0.002) for those who survive 10 years without ASCVD event (group (iii)).

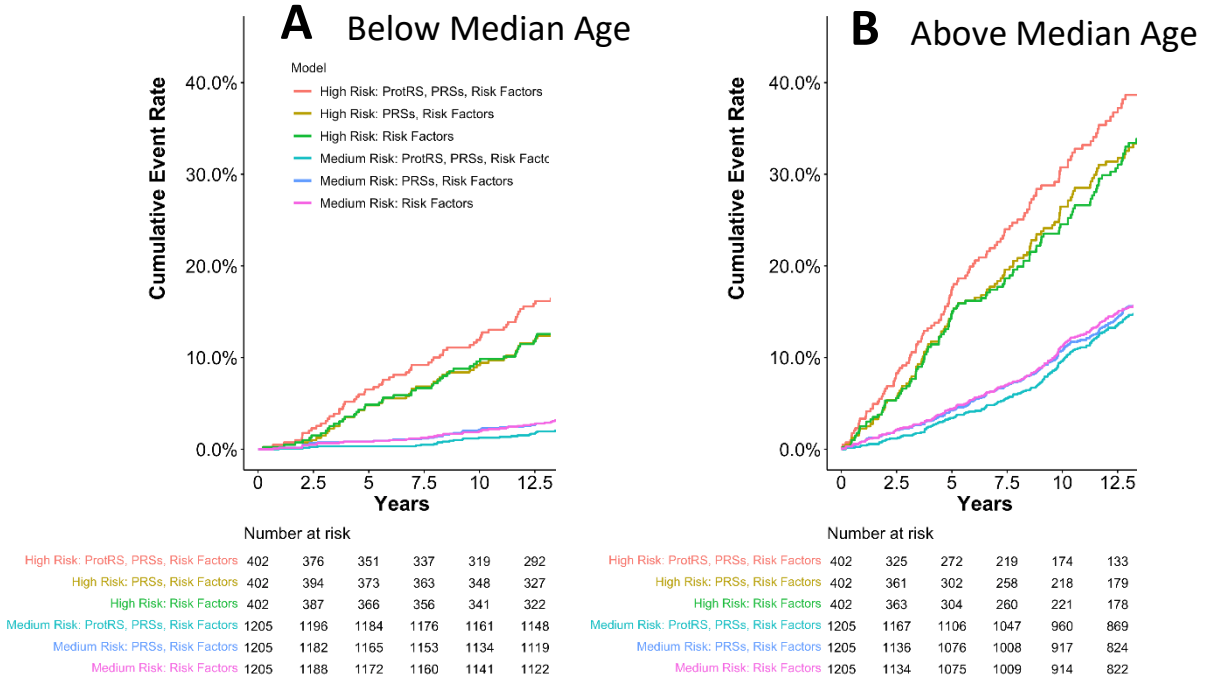
Panel B shows a reclassification table when both ProtRS and PRSs are added on top of the clinical risk factors. The total NRI is 0.084 (0.034 to 0.134) and 0.103 (0.056 to 0.153) when excluding those who die from non-ASCVD causes. NRI for ASCVD events is 0.113 (0.063 to 0.162), -0.029 (-0.037 to -0.020) for ASCVD non-events (groups (ii) and (iii)), and -0.009 (-0.017 to -0.002) for those who survive 10 years without ASCVD event (group (iii)).

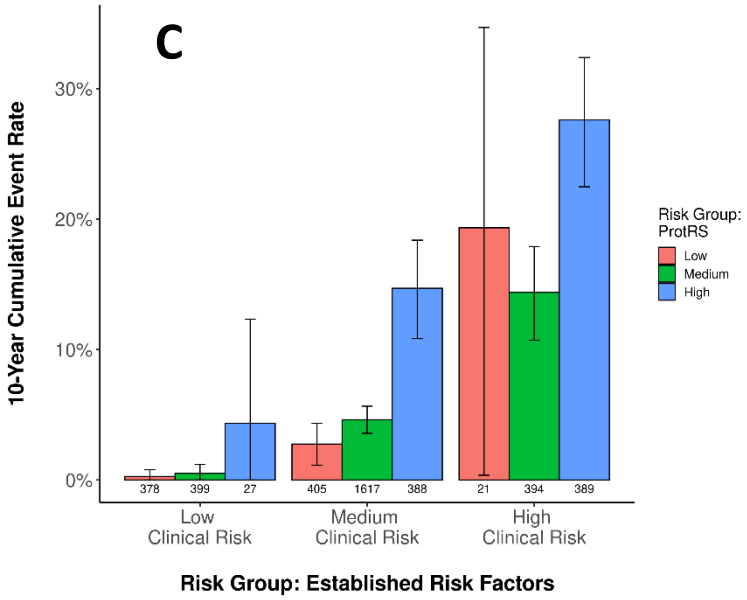
eFigure 9. Predicted 10-Year Risk by Different Models in Three Separate Groups in the Primary Event Population

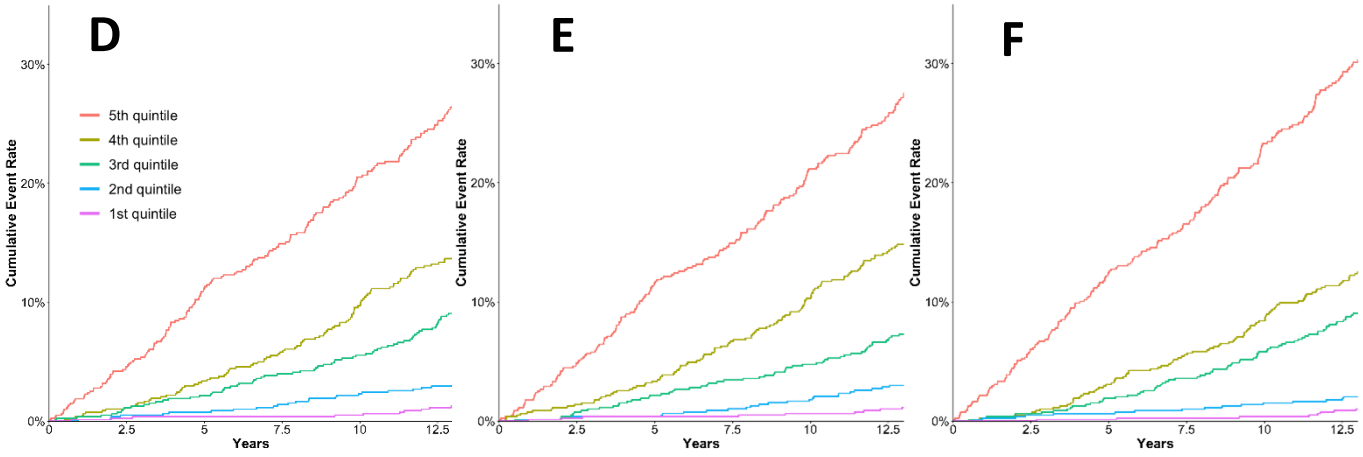


Shown are boxplots demonstrating predicted 10-year risk by different models for three separate groups in the primary event population. The groups considered are: (i) those with ASCVD event within 10 years (*N* = 284, green), (ii) those who die from other causes than ASCVD within 10 years (*N* = 358, blue), and (iii) those who survive 10 years without ASCVD event (*N* = 3,376, red).
The boxes range from the first quartile to the third quartile and the whiskers extend to the smallest and largest value or the 1.5 times the interquartile range from the box if the smallest or largest values are beyond that. Datapoints beyond the inter-quartile range are represented by dots. The black horizontal bar inside the box represents the median.

eFigure 10. Cumulative Rate of Cardiovascular Events in Primary Event Population







The plots show results for cardiovascular endpoint in the test set for the primary event population (*N* = 4,018, *N* events = 465) using risk models with coefficients fitted in the training set. Risk groups correspond to the 1st quintile (low risk), 2nd-4th quintiles (medium risk), and the 5th quintile (high risk). Clinical/established risk factors considered are sex, age, BMI, smoking, type 2 diabetes, hypertension treatment, and statin treatment. PRSs corresponds to polygenic risk scores for CAD and stroke. Cumulative event rates are based on Kaplan-Meier estimates.

Cumulative event rates for cardiovascular end point when restricting to individuals below or above the median age at plasma sampling in the test set (median age: 56.6 years; *N* events for age < 56.6 = 108; *N* events for age ≥ 56.6 = 357) are shown in **Panel A** (below median age) and **Panel B** (above median age).

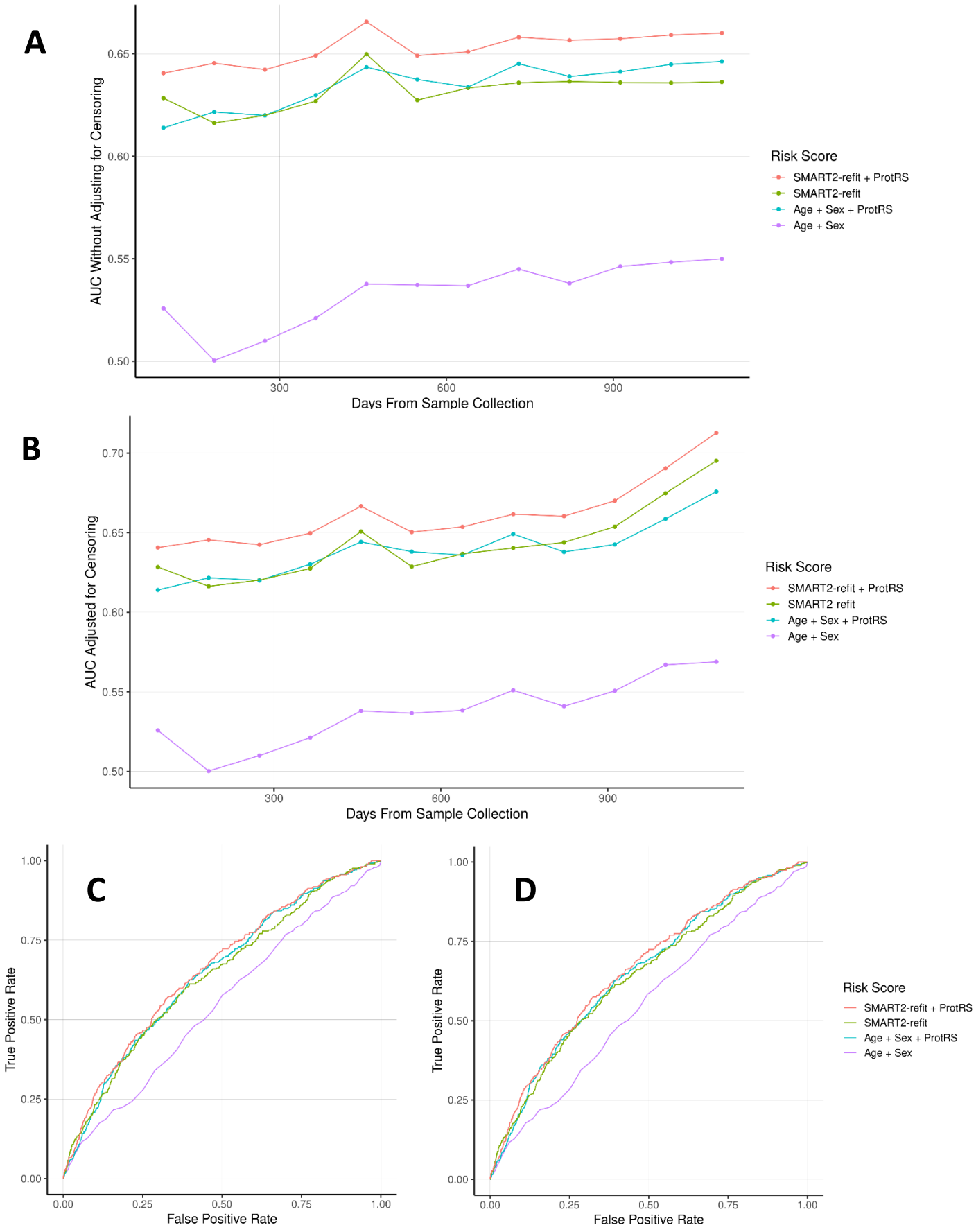
The curves are based on Kaplan-Meier estimates for different risk groups corresponding to the linear predictor values resulting from three different Cox models: (i) clinical risk factors; (ii) PRSs and clinical risk factors, and (iii) ProtRS, PRSs, and clinical risk factors. Shown are curves for High and Medium risk groups.

When considering both age groups and the top quintile of predicted risk, the model with clinical risk factors and the PRSs had a 10-year cumulative event rate of 21.2% (95% CI: 18.1% to 24.1%) and when adding the ProtRS to the cumulative event rate was 23.3% (95% CI: 20.0% to 26.4%). Among those that were above median age, the 10-year cumulative event rate for the top quintile for the full model with clinical risk factors, PRSs, and ProtRS was 30.8% (95% CI: 25.4% to 35.7%), that was considerably higher than when considering the same risk group for those below median age, or 11.9% (95% CI: 8.6% to 15.1%). Further counts and risk estimates for years 5 and 10 are shown in **eTable 9**.

Panel C: Shown are the 10-year cumulative event rates in the intersection of risk groups based on ProtRS and risk groups based on clinical risk factors. The numbers below the bars correspond to the number of individuals in the intersection. The addition of the ProtRS to clinical risk groups defined by medium and high risk estimated by clinical risk factors, resulted in a significant gradient in 10-year event incidence (*P* log-rank < 3.3e-6 when comparing medium and high ProtRS within the medium and high clinical risk groups). Patients with both high clinical risk and a high ProtRS, about 9.7% of the test population, had a 10-year cumulative CVD event rate of 27.6% (95% CI: 22.4% to 32.4%) compared to 0.3% (95% CI: 0.0% to 0.8%) for those with low clinical risk and a low ProtRS (about 9.4% of the population)

Panels D, **E** and **F** show the cumulative event rates for all risk quintiles of the whole test set for the clinical risk factor model (**Panel D**), clinical risk factors, $PRS_{CAD,}$ and PRS_{Stroke} model (**Panel E**), and clinical risk factors, PRS_{CAD}, PRS_{Stroke}, and ProtRS model (**Panel F**).

eFigure 11. AUCs and ROC Curves With Different Handling of Censoring in the Secondary Event Population



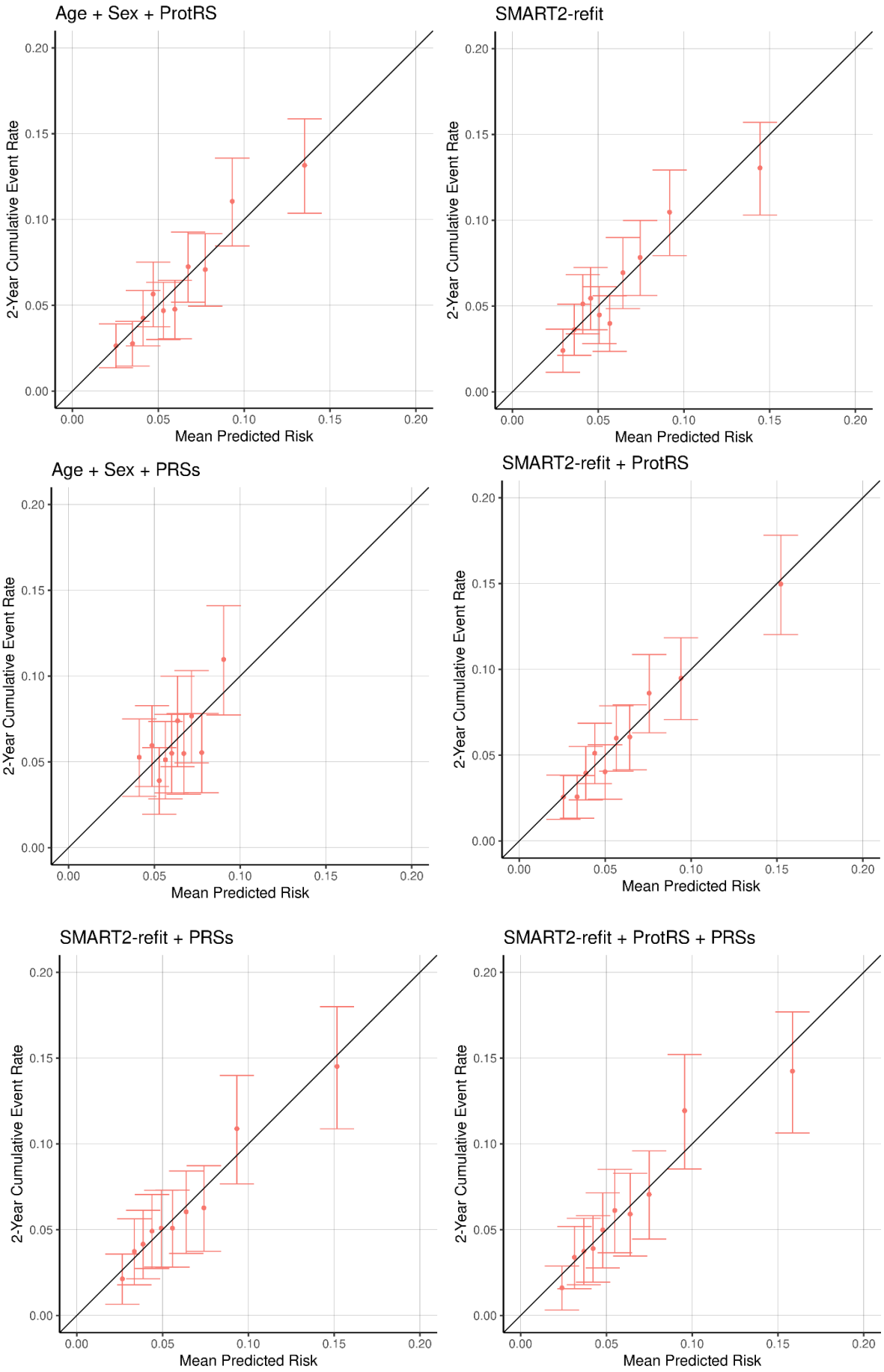
Shown are the discrimination performance of the ProtRS, PRSs and clinical risk factors in the secondary event population test set (N = 6,307, N events = 432). AUCs and ROC curves were both calculated with no

consideration of censoring, considering all non-events before each timepoint as controls, as well as with adjustment for censoring. When adjustments were used, censoring was adjusted for with inverse probability of censoring weighting. The censoring adjusted AUCs and ROC curves were calculated using the timeROC package in R[19]

Panels A and **B** show the AUC for the events within 3-36 months. In **Panel A** the simple binary AUC is shown with no adjustment for censoring while in **Panel B** we adjust for censoring.

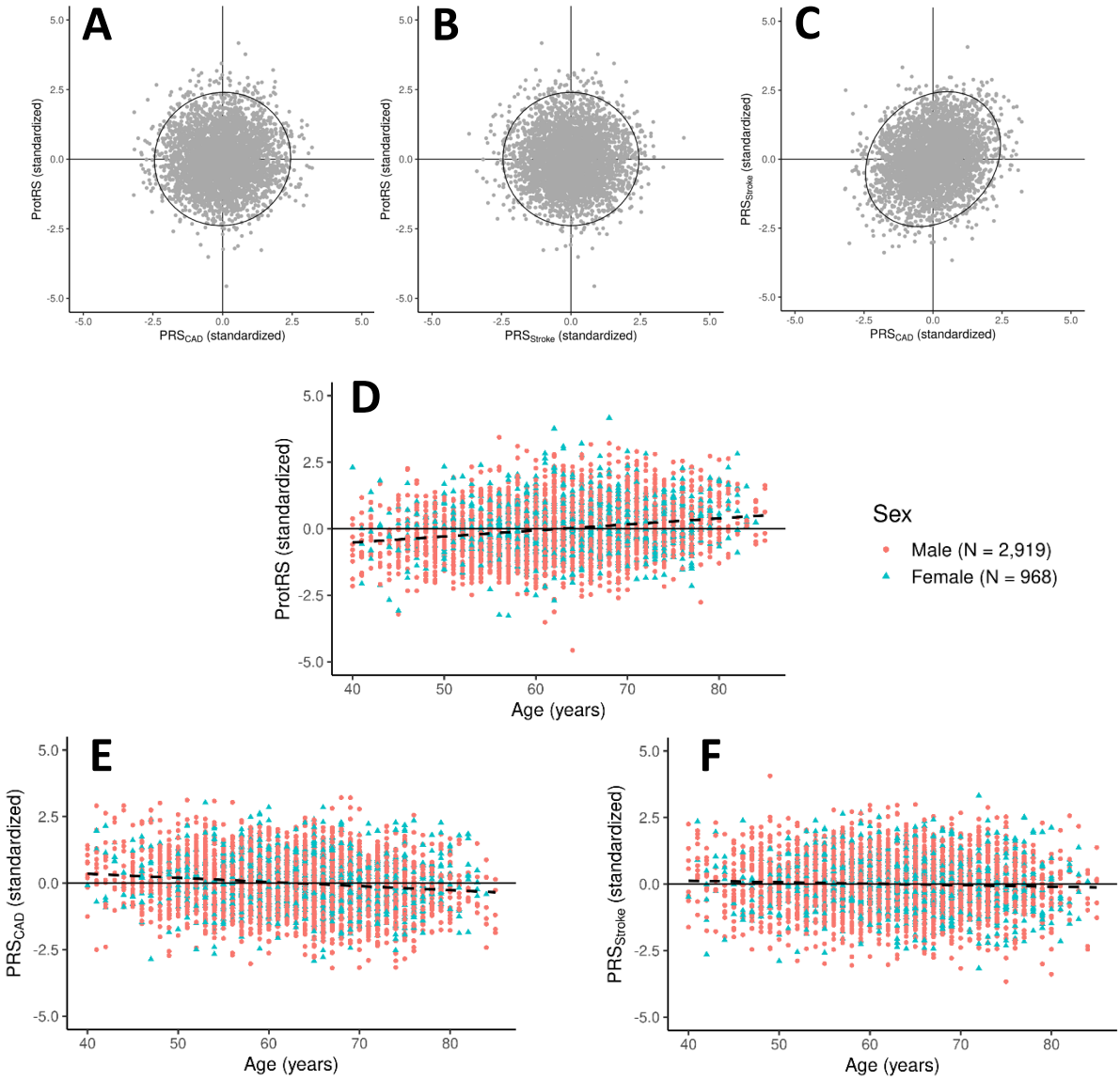
Panels C and **D** show the ROC curves for 10-year risk prediction without (**Panel C**) and with (**Panel D**) adjustment for censoring.

eFigure 12. Calibration in the Secondary Event Population



Shown are predicted 2-year risk and observed 2-year event rate in the secondary event population proteomics dataset (*N* = 6,307, *N* events = 432) or proteomics and genomics intersection dataset (*N* = 3,887, *N* events = 271) when the PRSs are included. The event rates were estimated in deciles of predicted risk with Kaplan-Meier method. The estimates are shown for different combinations of risk scores where the baseline survival is estimated in the same dataset as we are estimating calibration in. All the observed models seem to be mostly well calibrated.

eFigure 13. Polygenic Risk Scores for CAD and Stroke and Protein Risk Score in the Secondary Event Population



Shown are scatterplots for protein risk score (ProtRS), and polygenic risk scores for CAD and Stroke in **Panels A**, **B**, and **C** (one plot for each pair). The data in the scatterplots are based on individuals in the secondary event population that were of European descent based on ADMIXTURE analysis using genetic information and had both proteomics and genotypic data available. The total number of individuals was 3,887.

Panels D, **E**, and **F** show the behavior of the protein risk score, and polygenic risk scores for CAD and Stroke with age and sex.

Panel D shows how ProtRS increases with the age of the individuals in the secondary event population that have genotype measurements, where the age effect was 0.022 SD per year (95% CI: 0.019, 0.026; *P* = 2.1e-35) and sex effect was 0.12 SD for females (95% CI: 0.05, 0.19; *P* = 8.1e-4); variance explained by age and sex was 4.3%.

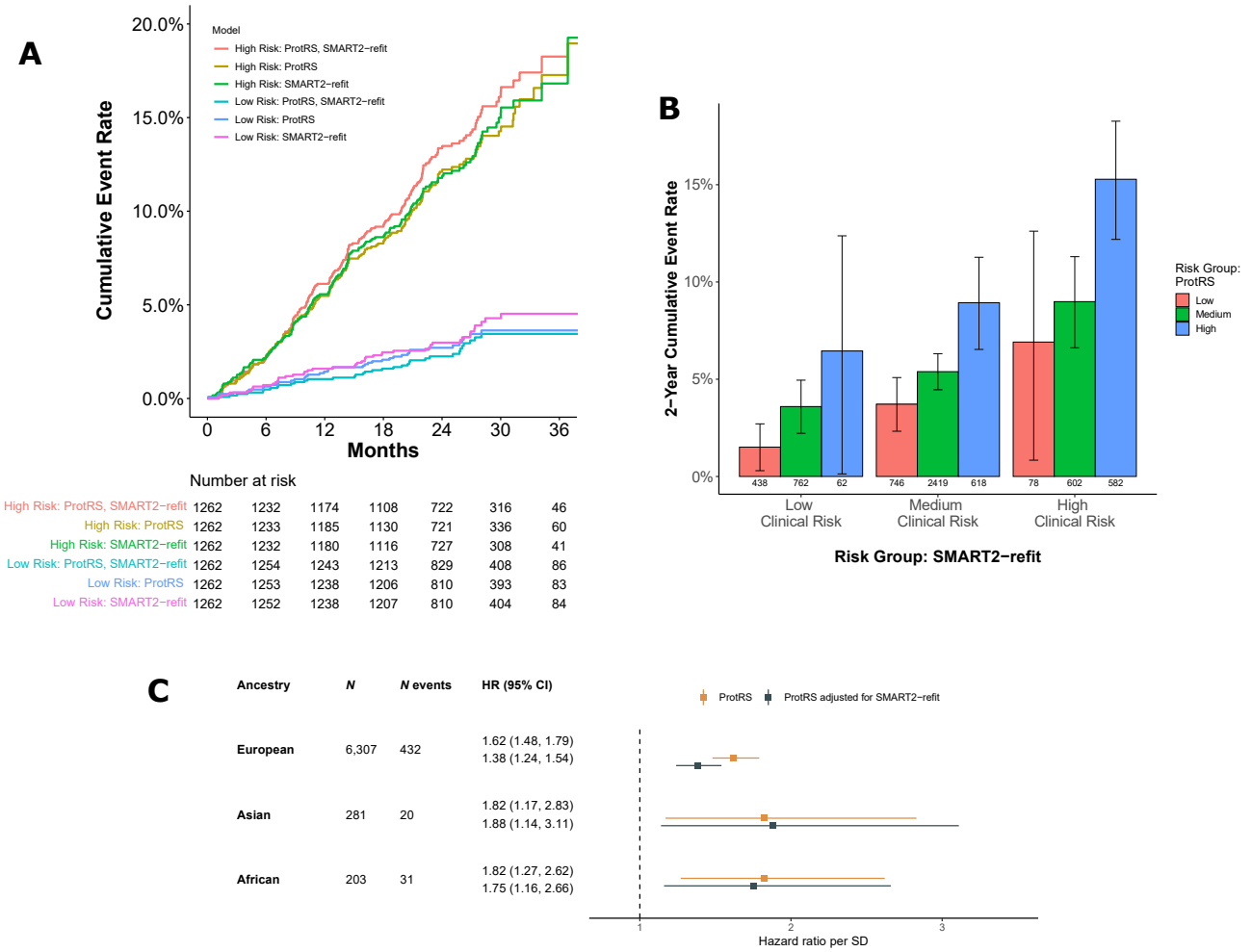
Panel E shows how PRS for CAD decreases with the age of the individuals in the secondary event population that have proteomics measurements, where the age effect was -0.016 SD per year (95% CI: -0.019, -0.012; *P* = 2.7e-18) and sex effect was 0.13 SD for females (95% CI: 0.06, 0.20; *P* = 4.9e-4); variance explained by age

and sex was 2.1%. This is consistent with what was observed when considering all individuals that had PRS available (i.e., not limiting to those that also had proteomics data available).

Panel F shows how PRS for stroke decreases slightly with the age of the individuals in the secondary event population that have proteomics measurements, where the age effect was -0.006 SD per year (95% CI: -0.009, -0.002; *P* = 0.0019); no sex effect was observed (0.04 SD for females (95% CI: -0.03, 0.12; *P* = 0.23); variance explained by age and sex was 0.3%.

The broken lines in the plots on **Panels D**, **E**, and **F** correspond to linear fit through the data, with adjustment for sex. The red color corresponds to males and the blue color to females.

eFigure 14. Cumulative Rate of Cardiovascular Events in Secondary Event Population



Risk groups correspond to the 1st quintile (Low risk), 2nd-4th quintiles (Medium risk), and the 5th quintile (High risk). For the $SMART2_{refit}$ score, coefficients of the SMART2 score were refitted using individuals of European ancestry in the secondary event population that did not have proteomics measurements. Cumulative event rates are based on Kaplan-Meier estimates.

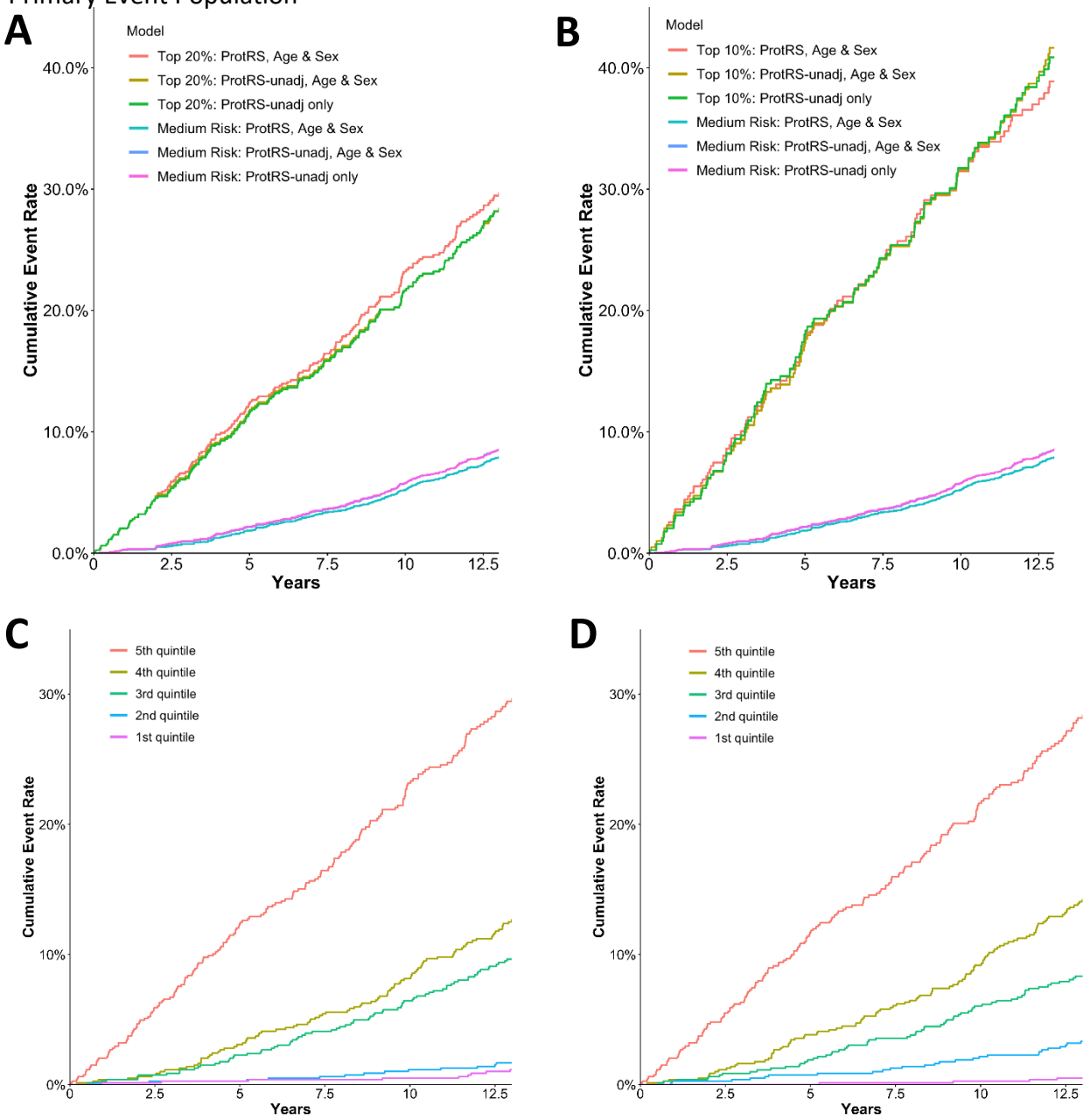
Panel A: Cumulative event rates for the secondary event population (*N* = 6,307, *N* events = 432) and the MACE endpoint. The curves are based on Kaplan-Meier estimates for different risk groups corresponding to the linear predictor values resulting from three different Cox models that all included age and sex as covariates. Shown are curves for the high and low risk groups. The high risk group for the ProtRS, age, and sex model had a 2-year cumulative event rate of 12.1% (95%CI: 10.2% to 14.0%). The high risk group for the model with only the $SMART2_{refit}$ score had a similar 2-year event rate, or 11.8% (95%CI: 9.9% to 13.6%), and the model including both the ProtRS and $SMART2_{refit}$ had a 2-year event rate of 13.4% (95% CI: 11.4%, 15.3%). For the low risk group, the 2-year cumulative event rates were 3.0% (95%CI: 2.0%, 3.9%), 2.7% (95%CI: 1.8%, 3.6%), and 2.2% (95%CI: 1.4%, 3.1%) for the ProtRS model, $SMART2_{refit}$ model, and the joint model, respectively.

Panel B: Shown are the 2-year cumulative event rates in the intersection of risk groups based on ProtRS and risk groups based on the $SMART2_{refit}$ score. The numbers below the bars correspond to the number of individuals in the intersection. The results are based on individuals in the secondary event population that had proteomics data (*N* = 6,307, *N* events = 432). The addition of the ProtRS to $SMART2_{refit}$ risk groups

resulted in a significant gradient in event incidence (*P* log-rank < 0.003 when comparing medium and high ProtRS within the medium and high $SMART2_{refit}$ risk group). Within the low $SMART2_{refit}$ risk group, the 2-year cumulative CVD event rate increased from 1.5% (95% CI: 0.3% to 2.7%) for those with low ProtRS to 6.5% (95% CI: 0.1% to 12.4%) for those with high. Patients with both high $SMART2_{refit}$ risk and a high ProtRS had a 2-year cumulative CVD event rate of 15.3% (95% CI: 12.2% to 18.3%).

Panel C: Comparison of hazard ratios for ProtRS in different ancestries and the MACE endpoint in the secondary event population. The analysis was based on self-reported ancestries. The Cox models included age and sex as covariates. No heterogeneity was observed when comparing the hazard ratios for the three ancestries (P_{het} = 0.74) and combined

eFigure 15. Cumulative Event Rate of Cardiovascular Events Stratified by Protein Risk Scores in the Primary Event Population

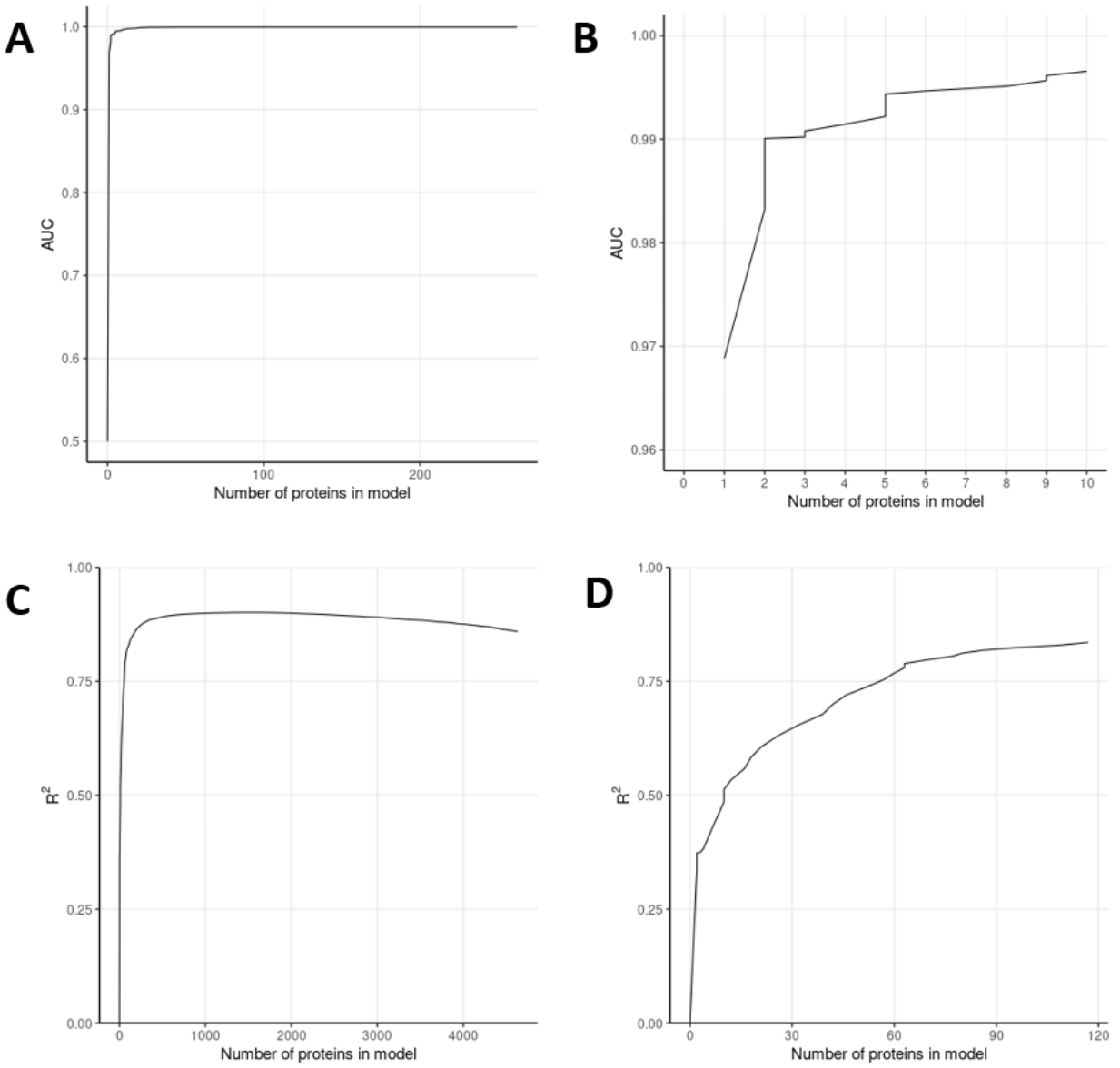


The plots show cumulative event rates for cardiovascular end point in the test set for the primary event population in Iceland (*N* = 4,018, *N* events = 465). The curves are based on Kaplan-Meier estimates for different risk groups corresponding to the linear predictor values resulting from four different Cox models: (i) $\text{ProtRS}_{\text{unadj}}$ only (ii) $\text{ProtRS}_{\text{unadj}}$, age , and sex, and (iii) ProtRS, age, and sex.

Shown are curves for the 20% at highest risk (**Panel A**), the 10% at highest risk (**Panel B**), and medium risk (20% to 80% percentile) as a reference.

Panels C and **D** show all five quintiles for the ProtRS, age, and sex model (**Panel C**) and the $\text{ProtRS}_{\text{unadj}}$, age , and sex model (**Panel D**).

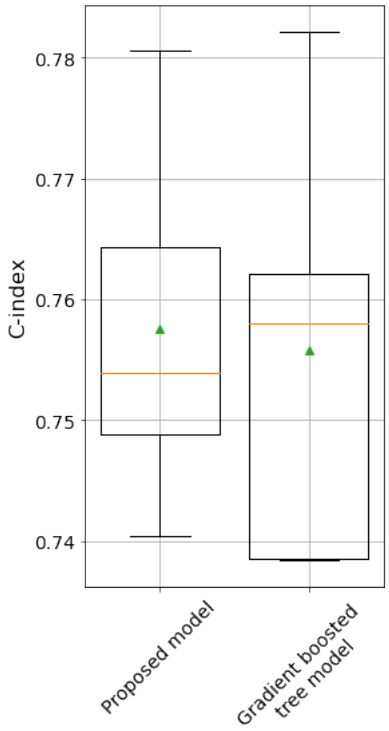
eFigure 16. Prediction of Age and Sex Using Proteomics Data in the Primary Event Population



Panels A and **B**: Shown are plots of AUC as a function of model size when predicting sex using proteomics data. To predict sex using proteomics measurements, we used a logistic regression model with lasso penalty, trained in the derivation set with all 4,963 protein measurements as candidate features. **Panel B** shows that with only one protein one can reach an AUC of almost 0.97 and only two proteins are necessary to reach an AUC of 0.99. The best one protein model consisted of PSA (prostate-specific antigen) and the best two protein model consisted of PSA and PZP (pregnancy zone protein).

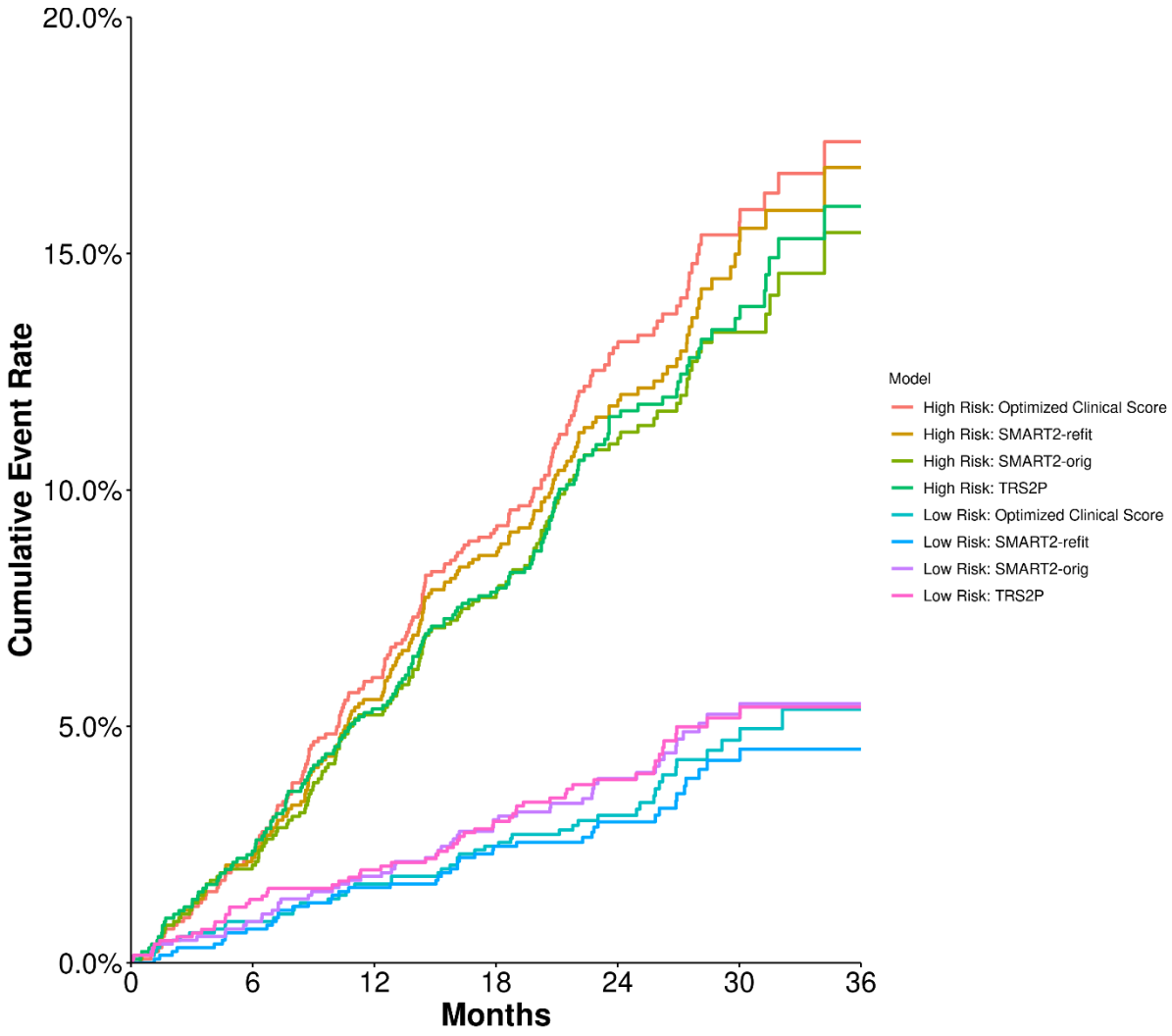
Panels C and **D**: Shown are plots of proportion of the variation (R^2) in age that is predictable from proteomics data in Iceland as a function of model size. To predict age at plasma extraction, we trained a linear regression model with lasso penalty in the derivation set using all 4,963 protein measurements as candidate features. **Panel D** shows that with less than 60 proteins one can explain 75% of the variance in age.

eFigure 17. Comparison of Models to Calculate Protein Risk Scores



The figure shows the Harrell's C-indices for each fold in a five-fold cross-validation performed in the derivation set (*N* = 9,522). The green triangles represent the mean C-index over the five folds and the yellow lines are the median. These results demonstrate no obvious benefit from using the non-linear gradient boosted tree cox model over the simpler penalized linear cox model.

eFigure 18. Cumulative Event Rate of Cardiovascular Events for Several Clinical Risk Scores in the Secondary Primary Event Population



The plot shows comparisons of cumulative event rates for the MACE endpoint in the secondary event population (*N* = 6,307, *N* events = 432) for several different clinical risk models. The curves are based on Kaplan-Meier estimates for risk groups corresponding to the 1st quintile (Low risk) the 5th quintile (High risk) for the linear predictors based on the models. 'Optimized Clinical Score' corresponds to a model that was derived within the secondary prevention population itself using a set of clinical variables that were available. The other scores are based on two published secondary prevention risk scores, the SMART2 score and the TIMI Risk Score for Secondary Prevention (TRS2P) (**eMethods**). Two scores based on the SMART2 score were considered: SMART2-orig uses the weights provided in the original publication for the score and SMART2-refit uses weights estimated using a Cox proportional hazards model using individuals from the secondary event population that were separate from the test set (the placebo arm of the FOURIER trial that excluded individuals with proteomics data, *N* = 5,403, *N* events = 413).

eAppendix 4. List of eTables

See separate Excel document in Supplement 2.

Table Contents:

eTable 1. Inclusion Criteria and Missing Data and Imputation in the Icelandic Proteomics Population

eTable 2. Weights in ASCVD Risk Scores Based on Clinical Risk Factors in the Secondary Event Population

eTable 3. Probes and Weights for Protein Risk Score Trained Including Age and Sex Covariates

eTable 4. Harrell's C-indices in Primary and Secondary Event Populations

eTable 5. Cox Proportional Hazard Joint Survival Analysis for Clinical Risk Factors, PRSs and ProtRS in the Primary Event Population Test Set (N = 4,018, N events = 465)

eTable 6. Association of Clinical Risk Factors With ProtRS and the ASCVD Event in the Primary Event Population Test Set (N = 4018, N events = 465)

eTable 7. Prediction Performance in the Primary Event Population for Events within 2-, 5-, or 10-Years

eTable 8. Category-free Net Reclassification Index in Primary and Secondary Event Populations

eTable 9. Risk Estimates for Models with Protein Risk Score, Clinical Risk Factors, and Polygenic Risk Scores for CAD and Stroke in the Test Set of the Primary Event Population

eTable 10. C-indices and Reclassification in the Secondary Event Population With Protein and Genotype Data Available

eTable 11. Cox Survival Analysis of the ProtRS in Secondary Event Populations of European, Asian, and African Ancestries

eTable 12. Cox Survival Analysis of Protein Risk Scores Trained With and Without Including NT-proBNP and MMP-12 in Primary and Secondary Event Populations

eTable 13. Association and Prediction Performance of Protein Risk Scores Trained Using Small Sets of Proteins Selected With Stability Selection

eTable 14. Protein Risk Scores Trained in Secondary Event Population or Primary Event Population Tested in Secondary Event Population Test Subset (N = 1863, N events = 127)

eTable 15. Correlations of Lab Measurements and Medical History at Baseline with ProtRS in Secondary Event Population

eTable 16. Association of Lab Measurements and Medical History Parameters, and Protein Risk Score with MACE Endpoint in the Secondary Event Population

eTable 17. Cox Survival Analysis of Protein Risk Scores, PRSs, and Clinical Risk Factor Scores in the Secondary Event Population