Neural network-based integration of polygenic and clinical information: development and validation of a prediction model for 10-year risk of major adverse cardiac events in the UK Biobank cohort

Jakob Steinfeldt, Thore Buergel*, Lukas Loock, Paul Kittner, Greg Ruyoga, Julius Upmeier zu Belzen, Simon Sasse, Henrik Strangalies, Lara Christmann, Noah Hollmann, Benedict Wolf, Brian Ference, John Deanfield†, Ulf Landmesser†, Roland Eils†*

Summary

Background In primary cardiovascular disease prevention, early identification of high-risk individuals is crucial. Genetic information allows for the stratification of genetic predispositions and lifetime risk of cardiovascular disease. However, towards clinical application, the added value over clinical predictors later in life is crucial. Currently, this genotype–phenotype relationship and implications for overall cardiovascular risk are unclear.

Methods In this study, we developed and validated a neural network-based risk model (NeuralCVD) integrating polygenic and clinical predictors in 395 713 cardiovascular disease-free participants from the UK Biobank cohort. The primary outcome was the first record of a major adverse cardiac event (MACE) within 10 years. We compared the NeuralCVD model with both established clinical scores (SCORE, ASCVD, and QRISK3 recalibrated to the UK Biobank cohort) and a linear Cox-Model, assessing risk discrimination, net reclassification, and calibration over 22 spatially distinct recruitment centres.

Findings The NeuralCVD score was well calibrated and improved on the best clinical baseline, QRISK3 (ΔConcordance index [C-index] 0·01, 95% CI 0·009–0·011; net reclassification improvement (NRI) 0·0488, 95% CI 0·0442–0·0534) and a Cox model (ΔC-index 0·003, 95% CI 0·002–0·004; NRI 0·0469, 95% CI 0·0429–0·0511) in risk discrimination and net reclassification. After adding polygenic scores we found further improvements on population level (ΔC-index 0·006, 95% CI 0·005–0·007; NRI 0·0116, 95% CI 0·0066–0·0159). Additionally, we identified an interaction of genetic information with the pre-existing clinical phenotype, not captured by conventional models. Additional high polygenic risk increased overall risk most in individuals with low to intermediate clinical risk, and age younger than 50 years.

Interpretation Our results demonstrated that the NeuralCVD score can estimate cardiovascular risk trajectories for primary prevention. NeuralCVD learns the transition of predictive information from genotype to phenotype and identifies individuals with high genetic predisposition before developing a severe clinical phenotype. This finding could improve the reprioritisation of otherwise low-risk individuals with a high genetic cardiovascular predisposition for preventive interventions.

Funding Charité–Universitätsmedizin Berlin, Einstein Foundation Berlin, and the Medical Informatics Initiative.

Lancet Digit Health 2022; 4: e84–94

*Contributed equally

†Contributed equally

Department of Cardiology, Campus Benjamin Franklin (J Steinfeldt MD, Prof U Landmesser MD), **Center for Digital Health** (T Buergel MSc, L Loock MSc, P Kittner BSc, G Ruyoga, J Upmeier zu Belzen MSc, S Sasse BSc, H Strangalies BSc, L Christmann MSc, N Hollmann BSc, B Wolf BSc, Prof R Eils PhD), **Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany; Centre for Naturally Randomized Trials, University of Cambridge, Cambridge, UK** (Prof B Ference MD); **Institute of Cardiovascular Science, University College London, London, UK** (Prof J Deanfield MD); **Health Data Science Unit, Heidelberg University Hospital and BioQuant, Heidelberg, Germany** (Prof R Eils)

Correspondence to:
Prof Roland Eils, Berlin Institute of Health, Charité–Universitätsmedizin Berlin, Digital Health Center, Berlin 10117, Germany
roland.eils@bih-charite.de

Introduction

Cardiovascular diseases, such as coronary heart disease, are consistently among the leading causes of death worldwide. A personalised risk assessment is fundamental to targeted prevention, intervention, and therapy. The early identification of high-risk individuals is crucial to reducing the disease burden on the population and increasing the effectiveness of interventions.[1]

Current prognostic models focus on prevalent classical cardiovascular risk factors, such as age, sex, blood pressure, cholesterol measurements, lifestyle factors such as smoking status, and medical history, which are analysed by linear models such as the semi-parametric Cox model.[2–4] Beyond these risk factors, many genetic variants have been associated with cardiovascular disease[5] and leveraged in polygenic scores (PGSs),[6] summarising genetic predisposition for cardiovascular disease at the time of birth. It has been shown that genetic risk captured in PGSs is associated with disease frequency for coronary heart disease and stroke.[7,8] The promise of PGSs to leverage genetic information in primary prevention for early disease detection has sparked the interest of regulatory authorities.[1,9]

However, the general applicability and benefit of PGSs for preventive cardiovascular medicine remain disputed.[10] One objection against a broad application of PGSs in

Research in context

Evidence before this study

In primary cardiovascular risk prediction, genetic predictors have already been added to traditional risk prediction models. Although neural networks have been applied previously on clinical variables, to date, no study has investigated their application for modelling the interaction of clinical and genetic predictors. We gathered evidence before this study using Google Scholar, searching all entries from the beginning of the database records until April 16, 2021, with no language restrictions. Relevant work identified by the Google Scholar search was considered the current state-of-the-art (thus reference material) for this study. To identify eligible studies, we used the keywords "cardiovascular disease", "polygenic scores", "survival analysis", and "neural networks".

Added value of this study

Neural network-based survival models represent the state-of-the art in time-to-event modelling. This study is, to our knowledge, the first to assess the applicability of these approaches for cardiovascular risk modelling in primary cardiovascular disease prevention. Furthermore, it is the first study to directly model the interaction of clinical risk and polygenic risk. We show that neural networks can model this genotype–phenotype relationship which could have direct consequence in the prioritisation of preventive interventions.

Implications of all the available evidence

Our proposed NeuralCVD score demonstrates that neural network-based survival models can learn expressive multimodal patient representations. Consequently, this paves the way for integrative models for cardiovascular primary prevention leveraging both clinical and polygenic information and for a clinical application of neural network-based risk models in general. Furthermore, this study motivates research in genetic variants and polygenic scores which maximise the residual information content over commonly assessed clinical predictors.

primary prevention is the low information content for most individuals in the population. In the long-tailed PGS distribution, only the individuals in the top percentiles show big changes in the associated disease frequencies. Five groups recently investigated the potential benefits of combining PGSs with conventional cardiovascular disease risk factors for cardiovascular risk prediction. Mosley and colleagues[11] found no additional benefit in adding a PGS against coronary heart disease to the features of the American Heart Association/Atherosclerotic Cardiovascular Disease (AHA/ASCVD) pooled cohort equation in two cohorts of US adults. Elliot and colleagues[12] found significant, yet modest improvements in discrimination (ie, the ability to differentiate individuals at low and high risk) and reclassification (ie, correct reclassification of predicted cases and non-cases based on the known ground truth compared with a baseline model) after adding their score against coronary artery disease to the features of the AHA/ASCVD pooled cohort equation and the QRISK3 score in the UK Biobank cohort. Sun and colleagues[13] found only incremental improvements in discrimination over the population, but notable reclassification after adding two PGSs against coronary heart disease stroke, respectively, to conventional predictors. With the additional information of genetic predisposition, the authors estimate prevention of additional 7% cardiovascular disease events compared with conventional scores based on the altered treatment recommendations.[13] Most recently, McKay and colleagues[14] developed a novel integrative risk tool combining a novel coronary artery disease PGS with the Pooled Cohort Equations (PCE) and QRISK3 score, respectively. The authors report a benefit in coronary artery disease prediction and estimate a net reclassification improvement of 0·137 at the 7·5% 10-year risk threshold for PCE, and 0·035 at the 10% 10-year threshold for QRISK3, and propose an effect of age and sex on reclassification.

Although PGSs bear an enormous potential for preventive medicine and risk modelling, their relationship to clinical phenotypes and known predictors remains elusive.[15] PGSs incorporate a wide range of variants over an individual's genome, distinguishing variants solely by effect size and dose, not by the mechanism of action. The incorporation of single nucleotide polymorphisms (SNPs) acting on known risk factors in PGSs has raised concerns about potential biases emerging from joint analysis with those same risk factors.[15] This concern calls for tools to model complex interactions to correct for these shortcomings in integrating PGSs and clinical predictors.

Neural networks represent state-of-the-art survival analysis.[16–19] If applied to real-world medical data, the model's increased complexity could facilitate the integration of polygenic information for primary prevention of cardiovascular disease by inherently accounting for the interaction of the polygenic information and the clinical parameters.

This study presents the development and validation of a novel neural network-based cardiovascular disease risk model, NeuralCVD, based on Deep Survival Machines,[19] for primary prevention based on a set of established cardiovascular disease risk factors. Comparing our model against existing risk scores and a Cox proportional hazards model[20] trained on the same data over the entire study population, we first demonstrated its discriminative capabilities. We subsequently assessed the integration of PGSs in risk modelling for primary cardiovascular

disease prevention by building on six well established PGSs against coronary artery disease and stroke.[7,8,21–24] After retraining our NeuralCVD risk score and the Cox model on clinical covariates and the PGSs, we demonstrate that our model can integrate the genetic information and learns the residual predictive contribution of the polygenic information over the manifested clinical phenotype.

Methods

Data source and outcome

We used data from the UK Biobank—a cohort of 273 383 women and 229 122 men aged between 37 years and 73 years at the time of their baseline assessment. The cohort is a sample of the UK's general population; participants were enrolled in 22 recruitment centres across the UK. Patients with pre-existing myocardial infarction, stroke, or lipid-lowering therapy were excluded from the analysis, but retained as auxiliary training data for our NeuralCVD score.

The outcome was 10-year cardiovascular disease risk defined by the earliest recorded event of fatal or non-fatal myocardial infarction (International Classification of Diseases [ICD]10 codes I21, I22, I23, I24, I25) or fatal or non-fatal transient ischaemic attack or ischaemic stroke (ICD10 codes G45, I63, I64) either in the primary care records, the hospital episode statistics, or death records. The study adhered to the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement for reporting.[25] The completed checklist can be found in the appendix (p 17).

Covariate selection

Predictors were selected to reflect traditional primary prevention risk models[2–4] (appendix p 16). Demographic information was extracted from primary care records and confirmed at the study's recruitment interview. Lifestyle information was extracted from the questionnaire at recruitment. Physical measurements and laboratory measures were taken at recruitment. Pre-existing medical conditions were extracted from the questionnaire or interview at recruitment, primary care records, and hospital episode statistics. Medications were extracted from the recruitment interview. PGSs (PGS000011,[21] PGS000018,[7] PGS000057,[22] PGS000058,[23] PGS000059[24]) for coronary artery disease and PGS000039[8] for stroke were selected from the PGS catalog[26] and calculated for all participants.

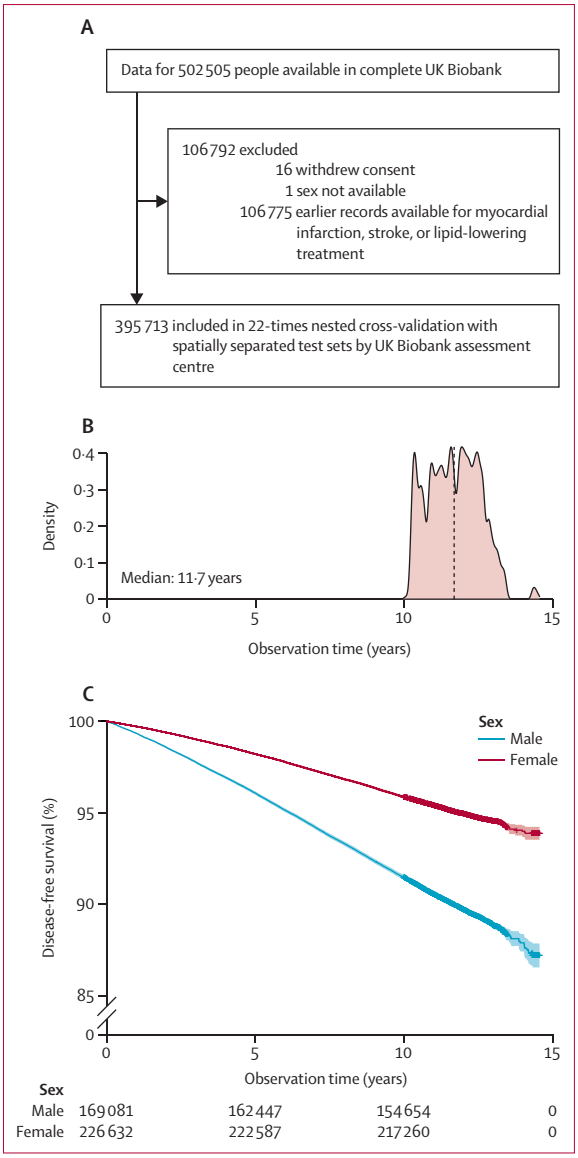


Figure 1: **Selection and characteristics of study population**
(A) Individuals in the UK Biobank population who withdrew consent, with missing information about their sex or with earlier records of incident myocardial infarction or stroke or lipid-lowering treatment at baseline were excluded. The remaining set was split into training, validation, and test sets in 22-fold nested cross-validation based on the assigned UK Biobank assessment centre. (B) Distribution of observation times for the derived study population. The median observation time was 11·7 years (IQR 11·0–12·3). (C) Kaplan-Meier estimates for the disease-free survival function stratified by sex. (D) Numbers at risk in 5-year intervals stratified by sex.

See Online for appendix

Dataset partitions and imputation

For model development and testing, we split the dataset into 22 spatially separated partitions based on the location of the assessment centre at recruitment. We analysed the data in 22-fold nested cross-validation, setting aside one of the spatially separated partitions as a test set, aggregating the remaining partitions and randomly selecting 10% of the aggregated data as the validation set. Within each of the 22 cross-validation loops, the individual test set (ie, the spatially disjunct partition) remained untouched throughout model development, while the validation set was used to validate the fitting progress and checkpoint selection. All 22 obtained models were then evaluated on their respective test sets. We assumed missing data occurred at random depending on the clinical variables and the

	Male (n=169 081)	Female (n=226 632)	Overall (N=395 713)
Age at recruitment	56 (48 to 62)	56 (49 to 62)	56 (49 to 62)
Ethnicity	..	..	..
Asian	3523 (2·1%)	3588 (1·6%)	7111 (1·8%)
Black	2782 (1·7%)	3871 (1·7%)	6653 (1·7%)
Chinese	491 (0·3%)	866 (0·4%)	1357 (0·3%)
Mixed	892 (0·5%)	1624 (0·7%)	2516 (0·6%)
White	158 761 (95%)	213 464 (96%)	372 225 (95%)
Missing	2632	3219	5851
Townsend deprivation index	−2·16 (−3·67 to 0·53)	−2·19 (−3·67 to 0·37)	−2·18 (−3·67 to 0·44)
Missing	230	276	506
Overall health rating	..	..	..
Excellent	30 892 (18%)	42 310 (19%)	73 202 (19%)
Good	98 290 (59%)	136 866 (61%)	235 156 (60%)
Fair	32 890 (20%)	39 057 (17%)	71 947 (18%)
Poor	5813 (3·5%)	6970 (3·1%)	12 783 (3·3%)
Missing	1196	1429	2625
Smoking status	..	..	..
Current	21 454 (13%)	20 012 (8·9%)	41 466 (11%)
Previous	58 717 (35%)	69 074 (31%)	127 791 (32%)
Never	87 924 (52%)	136 326 (60%)	224 250 (57%)
Missing	986	122	2206
Body-mass index, mg/kg^2	26·9 (24·7 to 29·5)	25·8 (23·2 to 29·3)	26·4 (23·8 to 29·4)
Missing	1161	118	2341
Weight, kg	84 (76 to 93)	68 (61 to 78)	75 (66 to 86)
Missing	1026	1105	2131
Standing height, cm	176 (172 to 181)	163 (158 to 167)	168 (162 to 175)
Missing	1017	959	1976
Systolic blood pressure, mm Hg	138 (128 to 150)	132 (120 to 146)	135 (124 to 148)
Missing	10 193	13 679	23 872

(Table 1 continues on next page)

cardiovascular events and performed multiple imputations using chained equations with random forests.[27] Continuous variables were standardised and mean centred; categorical variables were one-hot encoded. Imputation models were fitted on the training sets and applied to the respective validation and test set.

Model development and evaluation

We developed models on two distinct covariate sets, one including 29 cardiovascular risk factors in existing risk scores (table 1), the other with the addition of the computed values for the six PGSs. We constructed three models for each covariate set: a linear Cox model, a Cox model with interaction terms for age and each PGS, and our NeuralCVD score. The Cox model with the interaction terms allows to assess potential non-linear effects between age and the genetic information. For each assessment centre, and thus each cross-validation split, models were trained on the respective training set, and checkpoints selected on the respective validation set. For the final evaluation, predictions were then made for all participants in the test set. Harell's C-index was calculated with the lifelines package,[28] for both the aggregated test set and individual assessment centres. The net reclassification improvement was calculated with the nricens package.[29] 95% CIs were calculated based on 1000 bootstrapping runs and report the 2·5% and 97·5% quantile borders. For details on the implementation of NeuralCVD, the Cox models and the calibration, please refer to the appendix (pp 1–2).

Calculation of PGSs and the PGSMETA score

The PGSs were developed on multiple external cohorts and covered a diverse set of patients. Detailed information is available in the appendix (p 2). PGSs were calculated with the published weights from the PGS catalog,[26] the imputed genotype information from the UK Biobank, and the R package PRSice-2.[30] To analyse and visualise our model predictions by overall genetic risk, we sum the individual percentile ranks for each of the six PGS scores and calculate a new aggregated percentile rank over the sum to construct a polygenic meta score (PGSMETA). All models are trained by adding the six individual PGS scores, not the PGSMETA score.

Relative risk differences

To investigate the impact of the PGSs on individual predictions, we calculate relative risk differences between models trained with and without the polygenic information. By subtracting the clinical risk estimate from the model's prediction, which was trained on the clinical and polygenic information, we obtained an absolute risk difference. It is positive if the PGSs resulted in a higher risk estimate and negative if they lead to a lower risk estimate. Next, we normalised the absolute risk difference by dividing it by the clinical risk estimate to calculate the relative risk differences. Because absolute risk differences for individuals with clinical risk below 1% are close to zero and resulting relative risk differences in this group are thus prone to numerical instabilities in calibration, we did not calculate relative risk differences for these individuals. All patient data used throughout this study has been subject to patient consent as covered by the UK Biobank. All patient data used throughout this study was covered by the general patient consent of the UK Biobank, which applies to this study through the Material Transfer Agreement (MTA) of application 51157.Calibration was evaluated graphically by comparing predicted and observed risks.

Role of the funding source

The funders had no role in data collection, analysis, interpretation, writing, and the decision to submit.

Results

Participants were enrolled from March 13, 2006, to Oct 1, 2010. We extracted information on the demographics, clinical records, and outcomes of the complete

UK Biobank cohort.[31,32] 16 participants who withdrew their participation consent agreement; one participant without information on sex; and 106 775 participants with earlier records of myocardial infarction, stroke, or lipid-lowering treatment were excluded (figure 1). The remaining 395 713 participants had a median age of 56 years (IQR 49–62), with 95% being of White or British ethnicity, a median Townsend deprivation index of –2·18 (–3·67 to 0·44). 60% of the study population had good self-reported overall health (table 1; appendix p 10). The median follow-up time was 11·7 years (IQR 11·0–12·3). 28 083 (7·1%) participants had a major cardiovascular adverse event (MACE; defined as included fatal and non-fatal myocardial infarction, fatal and non-fatal transient ischaemic attack or stroke, and cardiovascular death; figure 1). Based on Deep Survival Machines,[19] we developed the NeuralCVD score (figure 2; appendix p 1) on a set of 29 cardiovascular risk factors used in well established scores, the ESC score,[2] the AHA/ASCVD score,[3] and the QRISK3 score[4] (table 1).

To determine whether neural networks improved risk discrimination over conventional approaches, we compared the NeuralCVD against established clinical baselines and a linear Cox[20] model trained on the same 29 cardiovascular risk factors. All scores were evaluated independently on all 22 assessment centres of the UK Biobank cohort with the Concordance Index and the categorical net-reclassification-improvement at the 10% threshold (following the NICE guidelines[33]) as metrics for the risk discrimination. We found that the NeuralCVD score outperformed SCORE with a difference in C-index of 0·037 (95% CI 0·034–0·039), ASCVD with 0·024 (0·023–0·026), and QRISK3 with 0·010 (0·009–0·011). At the 10% risk threshold the NRI over SCORE was 0·1043 (95% CI 0·0981–0·1103), resulting in an additional 4828 of 23 786 cases correctly identified as high-risk and 1106 cases incorrectly down classified. For non-cases, 14 572 of 371 889 were correctly down-classified, while 33 972 were incorrectly identified as high risk. The NRI of NeuralCVD over ASCVD was 0·0704 (0·0648–0·0765) and 0·0488 (0·0442–0·0534) over QRISK3 (figure 2; table 2). Absolute reclassification counts are provided in the appendix (pp 11, 15). Improvements were smaller over a linear Cox model fitted with the same set of covariates with a difference in C-index of 0·003 (0·002–0·004) and NRI of 0·0469 (0·0429–0·0511). The discrimination is stable over all 22 distinct assessment centres (appendix pp 4, 12). All models were well calibrated over the observed risk spectrum (figure 2; appendix p 4).

To assess the composite of clinical and genetic predictors, we rebuilt the model on an extended covariate set with six established PGS. We evaluated it against the occurrence of the first recorded MACE in the observation window.

To ensure the validity of the PGS in our cohort, we first confirmed the association of the applied PGS with the observed frequency of the MACE endpoint (appendix p 6). To test the potential of the NeuralCVD model to integrate polygenic information, we added six well established PGSs against coronary artery disease[7,21–24] and stroke[8] to the covariate set used in the previous analysis. The coefficients of the Cox model that included the PGS can be found in the appendix (p 14). Furthermore, to allow the Cox model to assess potential non-linear effects between age and the genetic information we additionally tested interaction terms between age and the PGSs. Additionally, we compared our models with the ASCVD-based model previously published by Sun and colleagues.[13]

Integrating PGSs in the NeuralCVD model improved risk discrimination over the clinical covariates alone with a difference in C-index of 0·006 (95% CI 0·005–0·007) and NRI of 0·0116 (95% CI 0·0066–0·0159; figure 2; table 3; appendix pp 11, 15). Although we observed improvements in discriminative performance for the Cox model after addition of the PGSs as well, the NeuralCVD model remained superior in C-index (COX plus PGS 0·002, 95% CI 0·002–0·003; COX plus PGS*age 0·002, 0·002–0·003) and NRI (COX plus PGS 0·0424, 95% CI 0·0383–0·0464; COX plus PGS*age 0·0359,

	Male (n=169 081)	Female (n=226 632)	Overall (N=395 713)
(Continued from previous page)			
Diastolic blood pressure, mm Hg	84 (78 to 91)	80 (74 to 87)	82 (75 to 89)
Missing	10 192	13 678	2387
Total cholesterol, mmol/L	5·72 (5·07 to 6·41)	5·93 (5·24 to 6·68)	5·84 (5·16 to 6·56)
Missing	1048	15 514	25 994
HDL cholesterol, mmol/L	1·26 (1·08 to 1·47)	1·57 (1·34 to 1·83)	1·43 (1·20 to 1·70)
Missing	22 622	34 937	57 559
LDL cholesterol, mmol/L	3·67 (3·17 to 4·20)	3·66 (3·13 to 4·25)	3·67 (3·15 to 4·23)
Missing	10 835	15 858	206 693
Triglycerides, mmol/L	1·68 (1·16 to 2·43)	1·30 (0·94 to 1·84)	1·44 (1·02 to 2·09)
Missing	10 659	15 634	26 293
Familial history of heart disease	57 025 (34%)	90 847 (40%)	147 872 (37%)
Antihypertensive treatment	1697 (1·0%)	1610 (0·7%)	3307 (0·8%)
Aspirin	1456 (0·9%)	935 (0·4%)	2391 (0·6%)
Atypical antipsychotics	2086 (1·2%)	3865 (1·7%)	5951 (1·5%)
Glucocorticoids	122 (<0·1%)	233 (0·1%)	355 (<0·1%)
Type 1 diabetes	795 (0·5%)	586 (0·3%)	1381 (0·3%)
Type 2 diabetes	3379 (2·0%)	2440 (1·1%)	5819 (1·5%)
Chronic kidney disease	6052 (3·6%)	8253 (3·6%)	14 305 (3·6%)
Atrial fibrillation	2687 (1·6%)	2303 (1·0%)	4990 (1·3%)
Migraine	7027 (4·2%)	20 879 (9·2%)	27 906 (7·1%)
Rheumatoid arthritis	6052 (3·6%)	16 916 (7·5%)	22 968 (5·8%)
Systemic lupus erythematosus	186 (0·1%)	725 (0·3%)	911 (0·2%)
Severe mental illness	14 303 (8·5%)	28 902 (13%)	43 205 (11%)
Erectile dysfunction	7731 (4·6%)	0 (0%)	7731 (2·0%)

Data are median (IQR) or n (%).

Table 1: **Study population**

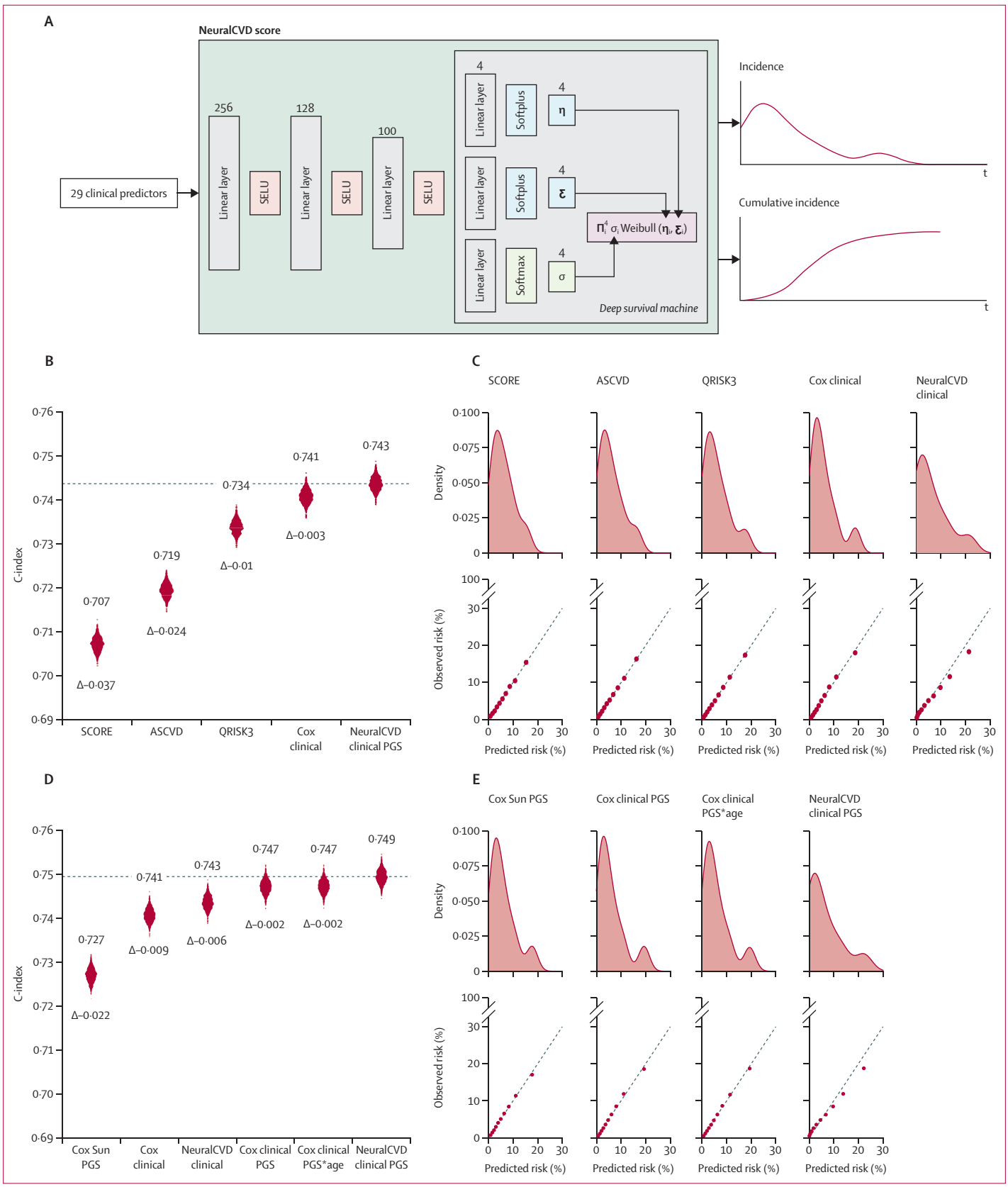

A
NeuralCVD score
29 clinical predictors
256
Linear layer
SELU
128
Linear layer
SELU
100
Linear layer
SELU
4
Linear layer
Softplus
4
η
Linear layer
Softplus
4
ξ
Linear layer
Softmax
4
σ
Π4 σi Weibull (ηi, ξi)
Deep survival machine
Incidence
t
Cumulative incidence
t
B
C-index
0·76
0·75
0·74
0·73
0·72
0·71
0·70
0·69
0·707
Δ–0·037
0·719
Δ–0·024
0·734
Δ–0·01
0·741
Δ–0·003
0·743
SCORE
ASCVD
QRISK3
Cox clinical
NeuralCVD clinical PGS
C
SCORE
ASCVD
QRISK3
Cox clinical
NeuralCVD clinical
Density
0·100
0·075
0·050
0·025
0
Observed risk (%)
100
30
20
10
0
0 10 20 30
Predicted risk (%)
D
C-index
0·76
0·75
0·74
0·73
0·72
0·71
0·70
0·69
0·727
Δ–0·022
0·741
Δ–0·009
0·743
Δ–0·006
0·747
Δ–0·002
0·747
Δ–0·002
0·749
Cox Sun PGS
Cox clinical
NeuralCVD clinical
Cox clinical PGS
Cox clinical PGS*age
NeuralCVD clinical PGS
E
Cox Sun PGS
Cox clinical PGS
Cox clinical PGS*age
NeuralCVD clinical PGS
Density
0·100
0·075
0·050
0·025
0
Observed risk (%)
100
30
20
10
0
0 10 20 30
Predicted risk (%)

Figure 2: **Comparison of the NeuralCVD score with established risk scores and after addition of PGSs**
(A) Our NeuralCVD score builds on the architecture of Deep Survival Machines,[19] learning a patient representation from the input features to parameterise a mixture of Weibull distributions to model the incidence function over a continuous time scale. (B) Our NeuralCVD score outperformed existing approaches in discrimination of major adverse cardiac event risk at 10 years measured by bootstrapped C-index. Over the entire population, this corresponded to an increment of 0·01 compared with the best-performing baseline model, the QRISK3 score (appendix p 4). (C–E) Calibration curves at 10 years. PGS=polygenic score. SELU=scaled exponential linear unit.

0·0321–0·0394). Compared with the model proposed by Sun and colleagues,[13] we see improvements in the C-index of 0·022 (95% CI 0·021–0·024) and in NRI of 0·0740 (95% CI 0·0687–0·0790). All models were well calibrated over the full spectrum of risk (figure 2), and the differences were consistent in the spatially separated assessment centres (appendix pp 4, 12).

To investigate the individual impact of the additional genetic information in both the Cox and the NeuralCVD models, we calculated relative risk differences with and without genetic information. The neural risk model predicted relative risk differences of up to 805% and –84% compared with up to 152% and –63% (249% and –72% with PGS*age interaction) for the Cox model in our study cohort (figure 3).

To examine these sizable risk differences in the NeuralCVD model, we investigated associations between the information added by the PGS (ie, the relative risk difference) and the observed clinical phenotype. Although we did not find pronounced associations of individual conventional risk factors with relative risk differences at 10 years (appendix p 7), we observed an association with the overall clinical risk (figure 3) and an association with age (figure 3). For high genetic risk individuals (top 5% PGSMETA), we saw pronounced risk differences in the predicted risk in younger individuals with low to intermediate clinical risk. With increasing clinical risk and age, the risk difference was diminished. This effect was predicted to be most pronounced for individuals with high genetic risk, decreasing with lower genetic predisposition (appendix p 8). These differences were non-existent in the linear Cox model without interaction terms, but observable in the Cox model with the PGS*age interaction terms.

Furthermore, the effect reflected the predicted cardiovascular risk trajectories stratified by clinical risk and age (figure 3). Young and low-risk individuals were predicted to have the highest relative risk increase from high genetic predisposition (RR[t_{10}] 2·64, 95% CI 2·52–2·76). Patients between 50 years and 60 years at intermediate clinical risk were predicted to have a lower impact (RR[t_{10}] 1·81, 1·78–1·85) and individuals older than 60 years at already high clinical risk see the smallest effect on their overall risk with the additional high genetic risk information 1·40 (1·37–1·42). To substantiate these findings, we calculated the number of events stratified by clinical risk and age at the end of the observation window for different genetic risk strata (appendix p 9). The relative risk for high genetic risk (top 5%) was 1·93 (95% CI 1·62–2·25) in the young and low risk subgroup, 1·65 (1·43–1·94) in the middle and intermediate risk supgroup, and 1·49 (1·34–1·63) in the older and high-risk subgroup at the end of the observation window.

These findings suggest that the additional predictive polygenic information depends on the clinical phenotype (ie, clinical risk) and that our NeuralCVD score can model this residual contribution. High genetic risk did

	SCORE *vs* NeuralCVD clinical	AHA/ASCVD *vs* NeuralCVD clinical	QRISK3 *vs* NeuralCVD clinical	Cox clinical *vs* NeuralCVD clinical
NRI	0·1043 (0·0981 to 0·1103)	0·0704 (0·0648 to 0·0765)	0·0488 (0·0442 to 0·0534)	0·0469 (0·0429 to 0·0511)
Cases (n=23 786)	15·65% (15·06 to 16·23)	11·41% (10·86 to 12·01)	9·62% (9·17 to 10·06)	10·75% (10·35 to 11·17)
Non-cases (n=371 889)	–5·22% (–5·33 to –5·11)	–4·38% (–4·47 to –4·27)	–4·74% (–4·83 to –4·65)	–6·06% (–6·14 to –5·98)

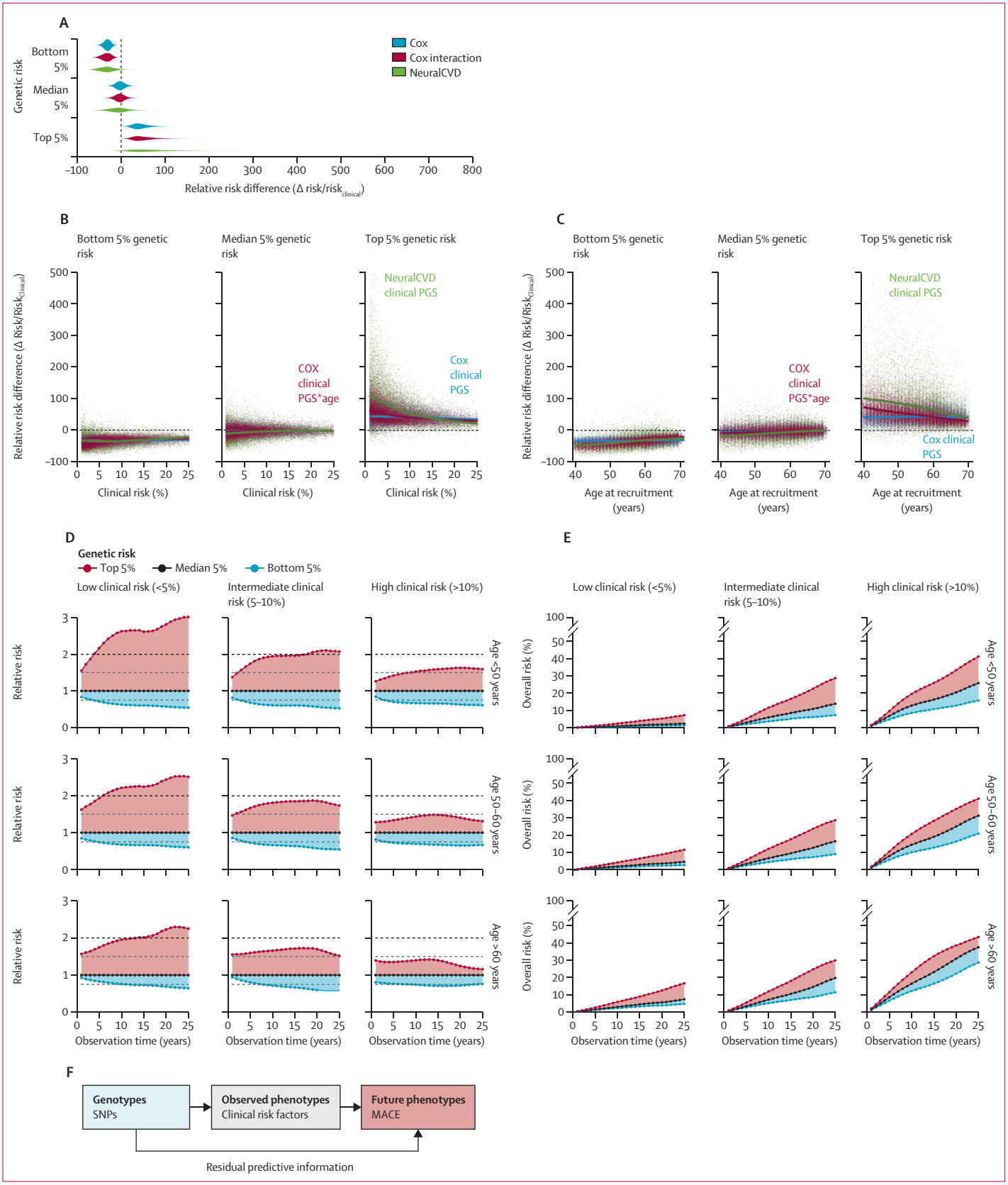
Data are NRI (95% CI) or % (95% CI). We assessed the categorical net reclassification improvement of our NeuralCVD score at the clinically relevant 10% risk threshold compared with the clinical baselines SCORE, ASCVD, QRISK3, and the linear Cox model. The NeuralCVD score substantially improves net reclassification and is particularly sensitive in detecting high-risk cases. NRI=net reclassification improvement. AHA/ASCVD=American Heart Association/ Atherosclerotic Cardiovascular Disease.

Table 2: **Categorical net reclassification improvement of NeuralCVD clinical at the 10% threshold**

	Cox Sun PGS *vs* NeuralCVD clinical plus PGS	Cox clinical PGS *vs* NeuralCVD clinical plus PGS	Cox clinical *vs* NeuralCVD clinical plus PGS	Cox clinical PGS*age *vs* NeuralCVD clinical plus PGS	NeuralCVD clinical *vs* NeuralCVD clinical plus PGS
NRI	0·0740 (0·0678 to 0·0795)	0·0424 (0·0383 to 0·0464)	0·0585 (0·0538 to 0·0625)	0·0359 (0·0321 to 0·0394)	0·0116 (0·0066 to 0·0159)
Cases (n=23 790)	12·92% (12·39 to 13·47)	10·34% (9·89 to 10·76)	11·87% (11·42 to 12·27)	9·00% (8·65 to 9·34)	1·12% (0·62 to 1·54)
Non-cases (n=371 909)	–5·52% (–5·64 to –5·42)	–6·10% (–6·16 to –6·00)	–6·01% (–6·11 to –5·92)	–5·41% (–5·48 to –5·34)	0·05% (–0·03 to 0·12)

Data are NRI (95% CI) or % (95% CI). Categorical net reclassification improvement of the NeuralCVD score with PGS at the 10% 10-year risk threshold compared with the American Heart Association/ Atherosclerotic Cardiovascular Disease-based model proposed by Sun and colleagues,[13] the linear Cox model with and without PGS addition, the non-linear Cox model with PGSs, and the NeuralCVD score without PGS addition. The NeuralCVD score with PGS improves net reclassification over all other scores. PGS=polygenic scores. NRI=net reclassification improvement.

Table 3: **Categorical net reclassification improvement of NeuralCVD clinical plus PGS at the 10% threshold**


A
Genetic risk
Bottom 5%
Median 5%
Top 5%
Cox
Cox interaction
NeuralCVD
−100 0 100 200 300 400 500 600 700 800
Relative risk difference (Δ risk/risk$_{clinical}$)
B
Bottom 5% genetic risk
Median 5% genetic risk
Top 5% genetic risk
Relative risk difference (Δ Risk/Risk$_{Clinical}$)
NeuralCVD clinical PGS
COX clinical PGS*age
Cox clinical PGS
0 5 10 15 20 25
Clinical risk (%)
C
Bottom 5% genetic risk
Median 5% genetic risk
Top 5% genetic risk
Relative risk difference (Δ Risk/Risk$_{Clinical}$)
NeuralCVD clinical PGS
COX clinical PGS*age
Cox clinical PGS
40 50 60 70
Age at recruitment (years)
D
Genetic risk
Top 5%
Median 5%
Bottom 5%
Low clinical risk (<5%)
Intermediate clinical risk (5–10%)
High clinical risk (>10%)
Relative risk
Age <50 years
Age 50–60 years
Age >60 years
0 5 10 15 20 25
Observation time (years)
E
Low clinical risk (<5%)
Intermediate clinical risk (5–10%)
High clinical risk (>10%)
Overall risk (%)
Age <50 years
Age 50–60 years
Age >60 years
0 5 10 15 20 25
Observation time (years)
F
Genotypes
SNPs
Observed phenotypes
Clinical risk factors
Future phenotypes
MACE
Residual predictive information

not significantly affect the overall risk in older individuals when their clinical risk was already high. In contrast, the risk in young individuals at low to intermediate clinical risk sharply increased.

Discussion

PGSs have been shown to inform on an individual's genetic predisposition for many common diseases. Their application in primary cardiovascular disease prevention suggests great potential for early identification of high-genetic-risk individuals and timely intervention before a clinical phenotype is developed. However, PGSs are approximations of the lifetime genetic risk and thus, to be applied clinically, it is imperative to understand the relationship between the information provided by PGSs, the observed clinical phenotype, and the overall risk later in life. Similarly, although neural networks represent the state-of-the-art performance in survival modelling to date, few medical studies exploit this potential.

In this study, we presented NeuralCVD, a novel neural-network-based cardiovascular disease risk model for primary cardiovascular disease prevention. On data from the UK Biobank cohort, we show that an application of NeuralCVD on phenotypic data improves discrimination and reclassification at the 10% risk threshold over currently available clinical scores and a Cox baseline model. These findings encourage the use of neural survival models in primary cardiovascular disease prevention, as this improvement in discrimination does not require any additional predictors. In agreement with previous studies,[12,13] we subsequently show that adding genetic information further improves discrimination and categorical reclassification at the 10% risk threshold resulting in more high-risk cases detected.

Established methods integrate clinical predictors and the polygenic information additively, irrespective of biological mechanisms of action and mediatory effects on the observed phenotype.[13,14] Although the effect of PGS addition on risk discrimination is small at the population level, it is greater at the individual level. Through further investigation of relative risk differences, we found that our NeuralCVD score accounted for the transition of predictive information from the genotype to the composite clinical phenotype by learning higher order interactions between clinical risk factors and PGS variables. Thereby, NeuralCVD captured an attenuating effect of observed phenotypes with increasing clinical risk on information gained by the PGS addition in the high genetic risk strata. We found a similar but weaker interaction with age, which could be modelled by interaction terms between age and the PGS in the Cox model. These findings imply that substantial parts of the genetic risk captured by PGS act through phenotypic manifestation, and age alone is not a sufficient approximation. It is the residual contribution of PGS information over the clinical risk factors (figure 3), which is relevant in an applied clinical setting.[15] This transition of the predictive information from the genotype to the clinical phenotype was first hypothesised by Jannsen and colleagues.[15] In their article, the central idea was that, although independent at birth, the effects of SNPs in the PGSs are mediated through clinical factors (eg, LDL cholesterol, blood pressure, and weight) and reduce the residual genetic risk contribution later in life. Analysing event rates in the UK Biobank, we can confirm this heterogeneity in residual genetic information (appendix pp 9, 13).

The implications of the findings are two-fold. First, PGSs allow for the identification of individuals who are still most susceptible to their genetic predisposition before developing a severe clinical phenotype. Second, when the predictive information has already transitioned from the genotype to the phenotype (ie, clinical risk), the future overall risk trajectory is just modestly informed by PGS.

Nevertheless, this study is subject to several limitations. First, as shown previously,[13] the UK Biobank study cohort is of generally lower risk for cardiovascular events than the general primary care population and recalibration with a relevant data source—eg, the UK Clinical Practice Research Datalink, should be performed before public application. Second, although the model was validated in spatially separated samples from the individual assessment centres, and we did not observe any signs of overfitting, the NeuralCVD model is yet to be evaluated in an entirely independent cohort. This is of particular importance for every model incorporating PGSs, as generalisation to ancestrally distinct populations is controversial.[34]

Figure 3: **Differences in relative and overall risk as modelled by NeuralCVD and the Cox models when stratified by age and clinical risk**

(A) Distributions of the RRD for three genetic strata (bottom, median, and top 5% PGSMETA). Higher genetic risk increases the RRD for all models. The distributions of RRDs for the NeuralCVD model are wide, with RRDs of up to 805% for the top 5% genetic stratum compared with the predicted risk based on the clinical factors. (B) RRDs within the two Cox models and the NeuralCVD score on PGS addition for the bottom, median, and top 5% of PGSMETA. Increasing genetic risk yields positive RRDs for both the Cox models and the NeuralCVD score. RRDs for the Cox model are constant over the spectrum of clinical risk. In contrast, the NeuralCVD learned the residual contribution of the polygenic risk over the clinical risk. In the high genetic risk group, RRDs were the highest for the low-to-intermediate clinical risk group and declined with clinical risk of more than 15%. (C) RRDs plotted against patient age at baseline. (D) 25-year risk ratios stratified by genetic risk (bottom, median, and top 5% PGSMETA), age, and clinical risk. Additional genetic information increased risk most in individuals with low-to-intermediate clinical risk, and age younger than 50 years. (E) 25-year overall risk stratified by genetic risk (bottom, median, and top 5% PGSMETA), age, and clinical risk. Risk ratios from (D) are reflected in the cardiovascular disease risk trajectories and in the proportion of polygenic risk in the overall risk. The difference in trajectories is most pronounced in individuals with low-to-intermediate clinical risk and age younger than 50 years. (F) Proposed mechanism for impact of polygenic information on overall risk, adapted from Janssens.[15] Parts of the SNPs included in PGS mediate through the manifestation of a clinical phenotype. As conventional risk factors contain this information, the information gained by PGS addition is the residual information. RRD=relative risk difference. PGSMETA=polygenic meta score as defined in methods. MACE=major adverse cardiac event. SNPs=single nucleotide polymorphisms.

Third, although discrimination, reclassification, and calibration are crucial criteria for evaluating predictive models and allow comparison over the established baselines, they are not quantifying an absolute clinical impact at the population level (eg, life-years saved by identifying the correct individuals for early intervention). This is relevant for primary prevention, because most individuals in a population are not expected to show strong risk modifications after adding PGSs to the predictors. Here prospective studies are required to show clinical utility. Additionally, clinical acceptance of genetic risk modification could be facilitated by further validation with phenotypic markers of subclinical disease.[35]

In summary, we introduced a clinically applicable neural-network-based risk model for primary cardiovascular disease prevention that outperformed conventional scores and learnt the residual genetic contribution to identify individuals at the highest risk of cardiovascular events. This opens up new opportunities for targeted primary cardiovascular disease prevention, integrating both clinical and genetic risk factors.

Contributors

RE, UL, and JD conceived, designed, and supervised the project. JS and TB implemented models, and did tests and data analysis. LL, PK, GR, HS, LC, and BW supported the analysis. JUzB, SS, and NH provided support with polygenic score calculation. BF provided methodological support and contributed to discussion of the results. JS, TB, RE, and UL wrote and prepared the Article. JS, TB, LL, PK, GR, HS, LC, BW, JUzB, SS, NH, RE, and UL had access to the raw data sets and verified the data. BF and JD were not covered by Charité Berlin's Material Transfer Agreement for the UK Biobank data application (number 51157) and therefore not permitted to access the raw data. JS, TB, JD, UL, and RE were responsible for the decision to submit this Article. All authors had access to the data presented, read, revised and approved the Article.

Declaration of interests

UL received grants from Bayer, Novartis, and Amgen; consulting fees from Bayer, Sanofi, Amgen, and Novartis; Daiichi Sankyo and honoraria from Novartis, Sanofi, Bayer, Amgen, and Daiichi Sankyo. JD received consulting fees from GENinCode UK; honoraria from Amgen, Boehringer Ingelheim, Merck, Pfizer, Aegerion, Novartis, Sanofi, Takeda, Novo Nordisk, and Bayer. He holds an Einstein Professorship, serves as fiduciary Senior Advisor and NHS Healthcheck Expert at Public Health England and chairs the Review of the National Health Check Programme at Public Health England. He is chief medical advisor to Our Future Health. All other authors declare no competing interests.

Data sharing

UK Biobank data are available to bona fide researchers on application at http://www.ukbiobank.ac.uk/using-the-resource/. Our code is available on https://github.com/thbuerg/NeuralCVD.

Acknowledgments

This research was conducted using data from UK Biobank, a major biomedical database, via application number 51157. This project was funded by the Charité–Universitätsmedizin Berlin and the Einstein Foundation Berlin. The study was supported by the Bundesministerium für Bildung und Forschung-funded Medical Informatics Initiative (HiGHmed, 01ZZ1802A, 01ZZ1802Z).

References

1 UK Department of Health and Social Care. Advancing our health: prevention in the 2020s. 2019. https://www.gov.uk/government/consultations/advancing-our-health-prevention-in-the-2020s/advancing-our-health-prevention-in-the-2020s-consultation-document (accessed Feb 5, 2021).

2 Conroy R. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003; **24:** 987–1003.

3 Goff D C, Lloyd-Jones D M, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk. *Circulation* 2014; **129:** S49–73.

4 Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017; **357:** j2099.10.1136/bmj.j2099.

5 Nikpay M, Goel A, Won HH, et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015; **47:** 1121–30.

6 Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018; **50:** 1219–24.

7 Inouye M, Abraham G, Nelson CP, et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J Am Coll Cardiol* 2018; **72:** 1883–93.

8 Abraham G, Malik R, Yonova-Doing E, et al. Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nat Commun* 2019; **10:** 5819.

9 Khoury MJ, Mensah GA. Is it time to integrate polygenic risk scores into clinical practice? Let's do the science first and follow the evidence wherever it takes us! 2019. https://blogs.cdc.gov/genomics/2019/06/03/is-it-time/ (accessed Feb 5, 2021).

10 Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* 2018; **19:** 581–90.

11 Mosley JD, Gupta DK, Tan J, et al. Predictive accuracy of a polygenic risk score compared with a clinical risk score for incident coronary heart disease. *JAMA* 2020; **323:** 627–35.

12 Elliott J, Bodinier B, Bond TA, et al. Predictive accuracy of a polygenic risk score-enhanced prediction model *vs* a clinical risk score for coronary artery disease. *JAMA* 2020; **323:** 636–45.

13 Sun L, Pennells LID, Kaptoge S, et al. Polygenic risk scores in cardiovascular risk prediction: a cohort study and modelling analyses. *PLoS Med* 2021; **18:** e1003498.

14 Riveros-Mckay F, Weale M E, Moore R, et al. An integrated polygenic tool substantially enhances coronary artery disease prediction. *Circ Genom Precis Med* 2021; **14:** e003304.

15 Janssens ACJW. Validity of polygenic risk scores: are we measuring what we think we are? *Hum Mol Genet 2019;* **28:** R143–50.

16 Luck M, Sylvain T, Lodi A, Bengio Y. Deep learning for patient-specific kidney graft survival analysis. *arXiv* 2017; published online May 29. https://arxiv.org/pdf/1705.10245 (preprint).

17 Gensheimer MF, Narasimhan B. A scalable discrete-time survival model for neural networks. *PeerJ* 2019; **7:** e6257.

18 Rietschel C, Yoon J, Van Der Schaar M. Feature selection for survival analysis with competing risks using deep learning. 2018; published online Nov 22. https://arxiv.org/abs/1811.09317 (preprint).

19 Nagpal C, Li XR, Dubrawski A. Deep survival machines: fully parametric survival regression and representation learning for censored data with competing risks. *IEEE J Biomed Health Inform* 2021; **25:** 3163–75.

20 Cox DR. Regression models and life-tables. *J R Stat Soc Series B Stat Methodol* 1972; **34:** 187–202.

21 Tada H, Melander O, Louie JZ, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur Heart J* 2016; **37:** 561–67.

22 Natarajan P, Young R, Stitziel NO, et al. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation* 2017; **135:** 2091–101.

23 Morieri ML, Gao H, Pigeyre M, et al. Genetic tools for coronary risk assessment in type 2 diabetes: a cohort study from the ACCORD clinical trial. *Diabetes Care* 2018; **41:** 2404–13.

24 Hajek C, Guo X, Yao J, et al. Coronary heart disease genetic risk score predicts cardiovascular disease risk in men, not women. *Circulation* 2018; **11:** e002324.

25 Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; **162:** W1–73.

26 Lambert SA, Gil L, Jupp S, et al. The Polygenic Score Catalog: an open database for reproducibility and systematic evaluation. *bioRxiv* 2020; published online May 23. https://doi.org/10.1101/2020.05.20.20108217 (preprint).

27 Wilson S. Miceforest. Github. https://github.com/AnotherSamWilson/miceforest (accessed Jan 4, 2021).

28 Davidson-Pilon C. lifelines: survival analysis in Python. *J Open Source Softw* 2019; **4:** 1317.

29 Inoue E. NRI for risk prediction models with time to event and binary response data. *CRAN* 2018; published online May 30. https://cran.r-project.org/web/packages/nricens/nricens.pdf (accessed Jan 16, 2021).

30 Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* 2020; **15:** 2759–72.

31 Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12:** e1001779.

32 Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562:** 203–09.

33 NICE. Overview, cardiovascular disease: risk assessment and reduction, including lipid modification. Guidance. NICE. 2016. https://www.nice.org.uk/guidance/cg181 (accessed Feb 3, 2021).

34 Schultz LM, Merikangas AK, Ruparel K, et al. Stability of polygenic scores across discovery genome-wide association studies. *bioRxiv* 2021; published online June 18. https://doi.org/10.1101/2021.09.10.459833 (preprint).

35 Figtree GA, Vernon ST, Nicholls SJ. Taking the next steps to implement polygenic risk scoring for improved risk stratification and primary prevention of coronary artery disease. *Eur J Prev Cardiol* 2020; published online Nov 4. DOI:10.1093/eurjpc/zwaa030.