Sex-Specific Survival Bias and Interaction Modeling in Coronary Artery Disease Risk Prediction

Ida Surakka, PhD; Brooke N. Wolford, PhD; Scott C. Ritchie, PhD; Whitney E. Hornsby, PhD; Nadia R. Sutton, MD; Maiken Elvenstad Gabrielsen, PhD; Anne Heidi Skogholt, PhD; Laurent Thomas, PhD; Michael Inouye, PhD; Kristian Hveem, MD, PhD*; Cristen J. Willer, PhD*

BACKGROUND: The 10-year Atherosclerotic Cardiovascular Disease risk score is the standard approach to predict risk of incident cardiovascular events, and recently, addition of coronary artery disease (CAD) polygenic scores has been evaluated. Although age and sex strongly predict the risk of CAD, their interaction with genetic risk prediction has not been systematically examined. This study performed an extensive evaluation of age and sex effects in genetic CAD risk prediction.

METHODS: The population-based Norwegian HUNT2 (Trøndelag Health Study 2) cohort of 51,036 individuals was used as the primary dataset. Findings were replicated in the UK Biobank (372,410 individuals). Models for 10-year CAD risk were fitted using Cox proportional hazards, and Harrell concordance index, sensitivity, and specificity were compared.

RESULTS: Inclusion of age and sex interactions of CAD polygenic score to the prediction models increased the C-index and sensitivity by accounting for nonadditive effects of CAD polygenic score and likely countering the observed survival bias in the baseline. The sensitivity for females was lower than males in all models including genetic information. We identified a total of 82.6% of incident CAD cases by using a 2-step approach: (1) Atherosclerotic Cardiovascular Disease risk score (74.1%) and (2) the CAD polygenic score interaction model for those in low clinical risk (additional 8.5%).

CONCLUSIONS: These findings highlight the importance and complexity of genetic risk in predicting CAD. There is a need for modeling age- and sex-interaction terms with polygenic scores to optimize detection of individuals at high risk, those who warrant preventive interventions. Sex-specific studies are needed to understand and estimate CAD risk with genetic information.

Key Words: atherosclerosis • cardiovascular disease • coronary artery disease • risk • sex
There is significant additive value of integrating genome-wide genetic data to enhance risk prediction using polygenic scores (PGS).7–10 Additionally, individuals with a PGS in the highest 8% of score distribution have a risk of CAD comparable to having monogenic familial hypercholesterolemia (3-fold increased risk).10 To date, investigators have shown that adding PGS11–17 to standard risk prediction algorithms enhances the power of the model to predict CAD, consistent with the estimated contribution of genetic factors responsible for 40% to 50% of CAD risk.18

The most predictive components of CAD prediction models are age and sex.2 The interplay of these 2 factors with the other traditional risk factors has been evaluated extensively in epidemiologic studies.2–6 However, only a limited amount of genetic risk prediction studies have evaluated the CAD risk separately for males and females.19–21 Of these 3 studies, 2 used PGSs based on significant CAD single nucleotide polymorphisms only, whereas, Huang et al19 additionally evaluated the performance of polygenic score with >1 million variants. All these studies concluded that genetic prediction works better for males than for females. However, these studies did not model longitudinal time to event which is required to evaluate the usefulness of the score in the clinical setting. To systematically evaluate the age and sex effects in the genetic CAD risk prediction, we used a longitudinal population-based data set of 51 036 samples from Norway and performed Cox proportional hazards models to explore how age and sex impact CAD genetic risk prediction (Table S1 and Supplemental Methods). The objective was to identify whether CAD genetic risk scores’ performance in the prediction of incident CAD depends on a patient’s age and sex.

RESULTS

CAD Polygenic Score and Correlations With Age and Sex

We initially tested whether the genome-wide polygenic score for CAD estimated using metaGRS weights12 (hereafter called the coronary artery disease polygenic score [PGSCAD]) was associated with sex or age in HUNT2 before examining how to optimally account for age and sex in CAD risk-stratification model. In principle, one would expect PGSs for various diseases to show equivalent distributions among males and females, provided age, sex, and ancestry are corrected in the underlying summary statistics and sex chromosomes are excluded from the evaluation. However, the relationship between the PGS, sex, and age could be impacted by ascertainment bias, undetected population stratification, or survival bias by genotype.

Significant associations were observed between enrollment age and PGSCAD and between sex and PGSCAD when assessing baseline data only (Table S2A and S2B). These results suggest that there is nonrandom selection in the cohort related to PGSCAD possibly ascertainment bias or survival effects. Based on the results from these 2 models, males at baseline had an average of 0.06 SD-units lower PGSCAD than females, and the PGSCAD was 0.003 SD-units lower per year of age. The association of age with PGSCAD was significant for both males and females (Table S2C and S2D), although the effect for males was marginally higher (PGSCAD 0.0032 SD-units lower per year for males, 0.0025 for females; Figure 1A). However, when adding the interaction term to the linear regression model, age×sex term was not significant (P=0.212). The sex–PGSCAD association was partly age-dependent as the effect of sex on PGSCAD becomes non-significant (P=0.404) when the interaction term is added (Table S2E). The age and sex associations were confirmed by testing the models in the UK Biobank. In UK Biobank, the interaction term age×sex on the PGSCAD was statistically significant (P=1.0×10^-6; Table S3A through S3C, Figure S1A). The trends were reduced when including the prevalent cases (together with the baseline statin users in the UK Biobank) in the baseline analysis for both datasets (N=1455 in HUNT2, N=84 292 in UK Biobank; Figure 1B, Table S4A and S4B, Figure S1B). The slight gradual decrease in mean PGSCAD by age, particularly in men, could be due to lower survival of older males with high PGSCAD, who are probably absent from the cohort at baseline (and some of whom were excluded from analyses due to prevalent or earlier-onset CAD).
CAD POLYGENIC SCORE AND AGE- AND SEX-INTERACTION MODELS

The PGS\text{CAD} performance was tested while explicitly modeling age, sex, and a comprehensive set of interaction terms to counter the survival bias observed at baseline. We used an additive PGS\text{CAD} model (model C1, Table S5) as the comparison model to test the effect of added interaction effects to the model performance (Table S6). To fully capture any potential interactions, a model including all interaction terms of age, sex, and PGS\text{CAD} was examined. Specifically, we aimed to test for the possible age-dependent (age×PGS\text{CAD} term) and sex-dependent (sex×PGS\text{CAD} term) behavior of the PGS\text{CAD}, and age-effects of the PGS\text{CAD} predictive performance that may differ between males and females (age×sex×PGS\text{CAD} term; model C2).

The sensitivity (78.4\%) increased in the full interaction model (model C2, Table S7) compared with the model with additive genetic effects only (model C1; sensitivity

---

**Figure 1.** Raw coronary artery disease polygenic score (PGS\text{CAD}) by age and sex in the cohort baseline.
Illustration of the selection bias in the cohort baseline using Lowess curves for PGS\text{CAD} by age for males and females separately. The plot has been zoomed in relative to the y axis to better show the trends. A, The trends in the analysis dataset used in the Cox models (prevalent cases excluded) and (B) when prevalent cases are included.
77.0%, Table 1), whereas, the C-index did not show a significant increase (model C2 C-index 0.839 [0.833–0.845], model C1 C-index 0.838 [0.832–0.844]). Similarly in the UK Biobank, the C-index did not change. The sensitivity increased after adding the interaction terms while specificity decreased. The overall sensitivity and specificity values between the 2 cohorts are different likely due to the well-known bias towards healthier individuals in the UK Biobank dataset (consistent with later analyses where the 7.5% risk threshold classifies a smaller proportion of individuals into the high-risk group). However, the proportional gain in the sensitivity was consistent between the 2 cohorts (1.8% increase in HUNT2 and 1.2% in UK Biobank). Figure 2 illustrates the effect of the age×sex×PGS\textsubscript{CAD} term in the HUNT2 dataset. The hazard ratios (HRs) for the PGS\textsubscript{CAD} on CAD 10-year risk with a model fit separately for males and females were not significantly different. However, we observed significant differences in model performance between the age groups when stratifying the dataset into 3 age bins, demonstrating an age interaction. When further stratifying both males and females separately into age bins (approximating the age×sex×PGS\textsubscript{CAD} interaction term in model C2), we observed small differences in the HRs between males and females in the same age bins, which, however, were not statistically significant due to the lower number of samples in the subgroups. Simultaneously, we added an age×sex-interaction term to ensure the model was valid by including all lower-level effects. The positive beta of this term indicates that irrespective of the PGS\textsubscript{CAD} age increases the CAD risk more substantially for females compared to males. This observation may be due to the effect of menopause on increased CAD risk in females.\(^{22}\)

### CAD POLYGENIC SCORE WITH ASCVD RISK SCORE

**Modeling Clinical and Genetic Risk Together**

We expect that genetic risk will most likely be used in conjunction with or in addition to already existing risk estimates. With this in mind, we modeled the ASCVD risk score with the PGS\textsubscript{CAD}. Our model with additive effects only (model C3, Table S8) had a higher C-index (0.842 [0.836–0.848]) and slightly lower sensitivity than model C1 (76.8%), which suggests that including the ASCVD risk score (ie, clinical score) on top of the PGS\textsubscript{CAD} does not increase the number of identified cases but rather affects the specificity, which increases from 76.8% (in model C1) to 77.4% (in model C3 which includes ASCVD risk), and is observable as an increase in the C-index. This finding could be caused by reduced transferability of Pooled Cohort Equation into Norwegian population. However, to evaluate the impact of the genetic interaction terms in a model with the clinical risk included, we tested the improvement in the model metrics by including the ASCVD risk score in the full PGS\textsubscript{CAD} interaction model (model C4, Table 2). This model had the highest C-index (0.845 [0.839–0.851]) and sensitivity (79.6%) of the combined prediction models. Moreover, when comparing the model with PGS\textsubscript{CAD} and ASCVD predictors but without the interaction terms (model C3) to the same model with genetic interaction terms (model C4) the sensitivity increased from 76.8% to 79.7% while specificity decreased from 77.4% to 76.0%. Almost all of the interaction terms showed significant P value in this model (Table 2), suggesting that the interactions between the genetic risk and age and sex should be taken into account when modeling the clinical and genetic risk together.

### Two-Step Approach

We also tested a scenario where the PGS\textsubscript{CAD} could be added as an independent risk estimation tool to identify additional cases that were not already identified by their ASCVD risk score. This 2-step case identification procedure is based on 2 sequential and independent risk estimates. After identifying high-risk individuals by the ASCVD risk score, we applied the PGS\textsubscript{CAD} risk model to the remaining individuals, including interaction terms (effects coming from the model C2 for the full population with full population variable distributions). This staged
approach, where ASCVD is first applied and PGSCAD with genetic interaction terms is then applied, newly classified 3235 individuals as high-risk (8.3% of the remaining dataset or 27.2% of the total dataset) totaling to 82.6% of the cases identified with the 2-step approach. Among those newly classified, we observed 253 additional future cases during the 10-year follow-up (32.9% of the cases missed by the ASCVD score; Figure 3). If we used model C1 (the model without interaction terms) in the second step instead of the model C2 with the interactions, we would identify 81.5% of the total cases instead of 82.6%, highlighting the importance of interaction modeling also when using the sequential approach. The 253 additional incident cases identified using model C2 had a mean ASCVD risk of 4.73% ranging from 1.09% to 7.49%, suggesting the PGS CAD provides information orthogonal to the ASCVD. We identified the same number of cases when applying the PGSCAD model first (model C2) and then the ASCVD risk score (individuals that have either high PGSCAD model C2 risk or high ASCVD risk). However, using the ASCVD first and then applying the genetic model may be more cost-efficient as the number of samples needed to be genotyped is lower (only those with low ASCVD risk), and follows the current standard clinical practice for the first stage.

**SEX-SPECIFIC MODELS AND SEX-SPECIFICITY OF MODEL METRICS**

The currently applied clinical risk scores are typically applied to males and females separately instead of using sex-interaction models. To test the applicability of our PGSCAD interaction models in the similar manner, we applied the models separately for males and females.

---

**Figure 2. Age dependence of the sex-effect.**

This figure shows the hazard ratios with 95% CIs for coronary artery disease polygenic score (PGSCAD) in models fitted in 11 different subsets. Subsets were separated by sex (males, females), by age (<45-year-old, between 45 and 70, >70-year-old) and finally stratified by both. All models have been adjusted for within-bin age and age2-effects and additionally the 3 age-bin models for sex.

| Variable/term (units) | Effect  | SE    | HR    | P value |
|-----------------------|---------|-------|-------|---------|
| Age², y²              | $-2.8\times10^{-3}$ | $1.5\times10^{-4}$ | 0.997 | $9.4\times10^{-83}$ |
| Sex (female=0, male=1)| 3.24    | 0.255 | 25.47 | $7.1\times10^{-37}$ |
| Age, y                | 0.417   | 0.018 | 1.517 | $3.25\times10^{-13}$ |
| PGSCAD (SD-unit)      | 1.21    | 0.182 | 3.369 | $2.3\times10^{-11}$ |
| ASCVD (risk/100)      | 3.01    | 0.199 | 20.33 | $1.2\times10^{-51}$ |
| Sex×age (year for males compared with females) | $-0.041$ | $3.8\times10^{-3}$ | 0.960 | $7.3\times10^{-28}$ |
| Sex×PGSCAD (SD-units for males compared to females) | $-0.436$ | 0.221 | 0.647 | 0.048 |
| Age×PGSCAD (SD-units per year) | $-0.013$ | 2.6×10⁻³ | 0.987 | $6.9\times10^{-7}$ |
| Sex×age×PGSCAD (SD-units per year for males compared with females) | $6.2\times10^{-3}$ | $3.2\times10^{-3}$ | 1.006 | 0.052 |

ASCVD indicates atherosclerotic cardiovascular disease risk score; HR, hazard ratio; HUNT2, Trøndelag Health Study; PGSCAD, coronary artery disease polygenic score; and SE, SE of the effect.
we tested the performance of models allowing for age dependence of the PGS\textsubscript{CAD} separately in males and females. First, we evaluated the PGS\textsubscript{CAD} model without interactions (model S1, Table S9A and S9B). The C-indexes observed were 0.850 (0.840–0.860) for females and 0.816 (0.808–0.824) for males, and the magnitude of the HR for the PGS\textsubscript{CAD} was similar for both sexes (HR females, 1.41 [1.33–1.49], HR males, 1.43 [1.37–1.50]).

The inclusion of the PGS\textsubscript{CAD}×age interaction into the models (Tables S10A and S10B) did not notably change the C-indexes, although the interaction term was significant for both sexes (P in females=1.85×10\textsuperscript{-6}, in males=8.91×10\textsuperscript{-4}). However, the sensitivity increased for both sexes in HUNT2 (Table 3A-B). In UK Biobank, the sensitivity only increased for males (Supplementary Tables 11A-B). Both the C-index and sensitivity increased for both males and females when adding the ASCVD risk score to the model (model S3, Tables S12A and S12B). Lastly, we performed the 2-step process described earlier for males and females separately by including (1) the conventional ASCVD risk and (2) PGS\textsubscript{CAD} with age-interaction term. Using the 2-step approach, we correctly reclassified an additional 194 and 59 future cases for males and females, respectively (38.3% and 22.4% of the cases missed by the ASCVD risk assessment). We observed increased sensitivity by the 2-step approach also in the UK Biobank. The corresponding numbers without the interaction terms in the second step were 183 future cases for males and 51 for females (36.1% and 19.4% of the cases missed by the ASCVD risk score).

**SEX-SPECIFIC MODEL METRICS AND THE EFFECT OF THE RISK THRESHOLD**

We saw lower sensitivity and higher specificity for females compared with males in all sex-stratified models that included genetic information. These 2 metrics are dependent on the risk threshold. Therefore, we tested how changing the threshold would affect the risk classification. The sensitivity and specificity for males and females for varied risk thresholds are presented in Figures S2 through S5. For all of the sex-stratified models, the percentages of individuals in the high-risk group were higher for males than for females at any given risk threshold. This finding was expected given that females have a lower overall prevalence of CAD. The proportion of individuals in the high-risk group based on ASCVD risk was close between the 2 sexes (Figure S5). This is most likely due to the underestimation of the ASCVD risk seen for males when applying the ASCVD risk calculation to our test dataset (Figure S6). Figure S7 shows the risk calibration for the model S3 as comparison. However, lower sensitivity was observed for females for models that include the PGS\textsubscript{CAD}. To achieve the same sensitivity observed for males at the 7.5% risk threshold (81.4%), we would need to lower the risk threshold in females to 5.0% (Figures S2, S3, and S4). In all 3 models (S1–S3), the specificity in females with the 5.0% risk threshold was better than the specificity in males with the 7.5% risk threshold.

**DISCUSSION**

This study evaluated statistical approaches in 2 population-based datasets to fine-tune the prediction of
individuals at risk for 10-year CAD events by accounting for different rates and age distributions of CAD in males and females. We found that the C-index and sensitivity of the 10-year prediction of CAD improved by including sex and age interactions when modeling PGS\textsubscript{CAD} compared to a model without interaction effects in both datasets. Inclusion of the interaction terms accounts for the nonadditive effects of age and sex that have been shown to exist for many of the complex traits.\textsuperscript{23} Additionally, including the interaction terms most likely corrects for the baseline survival bias observed. The implications of these results highlight the importance of modeling age and sex interactions in predicting CAD events with genetic information.

In our baseline correlation checks, we observed significant associations between the genetic score and both age and sex and replicated these findings in the UK Biobank. The observed associations suggest nonrandom selection related to genetics in the study cohorts, and we contend that the age association is derived from the survival bias of individuals with lower genetic risk of CAD. The sex association is most likely derived from the earlier onset of CAD in males, which enhances the survival bias of those with lower genetic risk in males. We expect these biases to be present in all cross-sectional studies where the age ranges over the expected age-of-onset of the studied disease. Moreover, the same biases are most likely also present in populations where risk estimates are applied to identify high-risk individuals.

We evaluated the potential incorporation of genetic information into identifying at-risk individuals by modeling clinical and genetic risk at the same time or by applying 2 risk estimates (clinical and genetic) in a sequential manner. Both of these approaches showed increased number of cases identified when the age- and sex-dependent behavior of the genetic risk was taken into account. This suggests that the genetic risk should be evaluated in a similar manner as the other clinically relevant risk factors; separately for males and females and individuals of different ages. Additionally, with genome-wide genotyping being translated into clinical settings, CAD risk prediction may be enhanced by the sequential 2-step approach we evaluate here (1) first apply the existing clinical score (ie, Pooled Cohort Equation/ASCVD risk score) and (2) from those identified with a low ASCVD risk, apply a second model incorporating age, sex, and genetic information with age and sex interactions to identify additional high-risk individuals (Figure 4). Using our 2-step approach with a set risk threshold of 7.5%, we identified a total of 82.6% of incident CAD diagnoses (74.1% by ASCVD risk estimation and an additional 8.5% by the PGS\textsubscript{CAD} interaction model). The newly identified future cases in the second step suggest that incorporating genetic information including age- and sex-interaction modeling captures cases that do not yet show clinical signs of atherosclerosis or hypertension (which are the biggest clinical contributors to the ASCVD risk after age and sex). The implications of these results could be 2-fold (1) clinicians maintain the

| Females risk model | Incident cases with risk >7.5% | Incident cases with risk <7.5% | Noncases with risk >7.5% | Noncases with risk <7.5% | Specificity/selectivity | Sensitivity/recall | C-index [95% CI] |
|-------------------|-------------------------------|-------------------------------|--------------------------|--------------------------|------------------------|-------------------|-----------------|
| ASCVD             | 865                           | 263                           | 5171                     | 21 256                   | 80.4%                  | 76.7%*            | NA†             |
| Additive PGS\textsubscript{CAD} model (model S1) | 807 | 321 | 4658 | 21 769 | 82.4% | 71.5% | 0.850 [0.840–0.860] |
| PGS\textsubscript{CAD} with age-interaction (model S2) | 823 | 305 | 4792 | 21 635 | 81.9% | 73.0% | 0.851 [0.843–0.859] |
| PGS\textsubscript{CAD} with age-interaction and ASCVD (model S3) | 833 | 295 | 4707 | 21 720 | 82.2% | 73.8% | 0.859 [0.851–0.867] |

| 2-step: ASCVD then PGS\textsubscript{CAD} with age-interaction‡ | 924 | 204 | 6032 | 20 395 | 77.2% | 81.9% | NA† |

| Males risk model | Incident cases with risk >7.5% | Incident cases with risk <7.5% | Noncases with risk >7.5% | Noncases with risk <7.5% | Specificity/selectivity | Sensitivity/recall | C-index [95% CI] |
|------------------|-------------------------------|-------------------------------|--------------------------|--------------------------|------------------------|-------------------|-----------------|
| ASCVD            | 1339                          | 507                           | 4808                     | 16 827                   | 77.8%                  | 72.5%*            | NA†             |
| Additive PGS\textsubscript{CAD} model (model S1) | 1504 | 342 | 6727 | 14 908 | 68.9% | 81.5% | 0.816 [0.808–0.824] |
| PGS\textsubscript{CAD} with age-interaction (model S2) | 1510 | 336 | 6745 | 14 890 | 68.8% | 81.5% | 0.816 [0.808–0.824] |
| PGS\textsubscript{CAD} with age-interaction and ASCVD (model S3) | 1537 | 309 | 6828 | 14 807 | 68.4% | 83.3%* | 0.822 [0.814–0.830] |

| 2-step: ASCVD then PGS\textsubscript{CAD} with age-interaction‡ | 1533 | 313 | 6928 | 14 707 | 68.0% | 83.0% | NA† |

ASCVD indicates atherosclerotic cardiovascular disease risk score; CI, confidence interval; C-index, Concordance Index; HUNT2, Trøndelag Health Study; and PGS\textsubscript{CAD} coronary artery disease polygenic score.

* Highest observed sensitivity for the given gender of the single step models.
†C-index not available as the risk score is not based on a fitted model.
‡This is not a model but a sequential identification of high-risk individuals using first ASCVD risk score and then genetic interaction model to estimate ASCVD risk.
ability to identify high-risk individuals using the ASCVD risk tool, and (2) clinicians with access to genetic information on patients are able to more accurately discern which additional individuals may benefit from timely prevention strategies (Figure 5). Implementation of this approach will require a large study with diverse populations to test risk factors including genetic information to ascertain population-level effects that can be applied to a single patient in clinical practice.

For both cohorts, the sensitivity for females was consistently lower than for males. In the HUNT2 dataset, we found that similar sensitivity to predict female cases could be achieved by lowering the risk threshold for preventive therapies from 7.5% to 5.0%. Additionally, this would not result in a higher proportion of females recommended for treatment relative to males. We suggest that the risk threshold used in the genetic screening should be independently evaluated in males and females before applying genetic information in an equal manner in the clinical setting. For example, in our dataset, if we changed the risk threshold from 7.5% to 5.0% in females when applying the 2-step sequential approach, we would increase the identification of cases from 81.9% to 86.2% without increasing the proportional amount of females suggested for treatment (25.3%) relative to males recommended for treatment (36.0%). However, as has been previously shown,24 the composition of CAD risk differs between males and females also in the non-genetic risk factors suggesting that the specificity of female CAD prediction could be improved by including additional CAD risk factors (ie, C-reactive protein and hormone therapies). Nonetheless, the inclusion of genetic information into clinical practice and the effect of the inclusion to the risk thresholds needs to be comprehensively validated in further clinical studies, ideally with sex-segregated groups.

Our study has important limitations. First, the datasets used in this study, HUNT2 and UK Biobank, are sampled from different populations than the datasets in which the ASCVD score was originally created (different ancestry, country of residence, younger, and healthier). Moreover, the ASCVD score was developed to evaluate the risk of developing CAD or stroke. In our study, we used the ASCVD score to predict CAD events or death during the 10-year follow-up time. This approach may have caused the miscalibration observed in the HUNT2 study, which limits our ability to perform unbiased one-to-one comparisons between the performance of these scoring methods. However, the trends and conclusions reported herein do not rely on the exact ASCVD risk, but rather, compare the change in the metrics when modeling genetics with and without age- and sex-interaction terms. Second, the participants in this study are of European ancestry, and therefore, the results may not be generalizable to populations with other ancestries.25 Additional studies are needed to determine the importance of interaction effects in the genetic prediction of other traits and diverse populations with different rates of clinical risk factors such as hypertension and high low-density lipoprotein cholesterol. Third, due
to the lack of family history information available in the hospital registry linked HUNT2 dataset, we tested the performance of the interaction models against only one clinical score, albeit the one recommended by the American Heart Association. Lastly, our models were based on only a single PGS, although the performance of several different genome-wide PGSs (ie, those derived from statistical methods, such as metaGRS, LDpred, or PRS-CS) have shown to be nearly equivalent in CAD prediction.

CONCLUSIONS

All populations screened for CAD risk are subject to survival bias that shows a depletion of high PGS individuals. We suggest using age and sex interactions with the PGS to account for the nonadditive effects in disease prediction to further increase the number of future cases identified. To predict future CAD events, the best performing models in this study utilized both clinical and genetic information including interactions—whether applied as a single model or in a sequential 2-step process. Moreover, CAD prediction studies with genetic information should focus on the sex-specific behavior of the predictors and prediction models to account for sex-specific genetic effects and differences in the incidence of CAD events between males and females.

ARTICLE INFORMATION

Received July 22, 2021; accepted September 29, 2022.

Affiliations

Division of Cardiovascular Medicine, Department of Internal Medicine (I.S., W.E.H., N.R.S., C.J.W.), Department of Computational Medicine & Bioinformatics (B.N.W., C.J.W.), Department of Human Genetics (C.J.W.), University of Michigan, Ann Arbor. Department of Biostatistics & Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor (B.N.W.). Cambridge Baker Systems Genomics Initiative, Department of Public Health & Primary Care (S.C.R., M.I.), British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health & Primary Care (S.C.R., M.I.), British Heart Foundation Centre of Research Excellence (S.C.R., M.I.), University of Cambridge, Cambridge Baker Systems Genomics Initiative, Baker Heart & Diabetes Institute, Melbourne, Victoria, Australia (S.C.R., M.I.). K.G. Jebsen Center for Genetic Epidemiology, Department of Clinical & Molecular Medicine (L.T.) and BioCore - Bioinformatics Core Facility (L.T.), Norwegian University of Science & Technology, Trondheim, Norway. Health Data Research UK Cambridge, Wellcome Genome Campus & University of Cambridge, United Kingdom (M.I.). Department of Clinical Pathology, University of Melbourne, Parkville, Victoria, Australia (M.I.). The Alan Turing Institute, London, United Kingdom (M.I.). HUNT Research Centre, Department of Public Health & Nursing, Norwegian University of Science and Technology, Levanger, Norway (K.H., C.J.W.).
Acknowledgments

The authors thank the HUNT (Trendelag Health Study) and UK Biobank partici- pants for their contributions to research. The authors thank Kuan-Han Wu for his important graphical contributions. Drs Surakka, Hveem and Willer designed the study. Drs Surakka, Ritche, and Wolford analyzed the data. Drs Hedi Skogohnt, Elvendad Gabrielsen, and Thomas contributed to the phenotype harmonization. Drs Sutt and Hornsby provided clinical expertise. Drs Surakka, Hornsby, Ritche, Inouye, Hveem, and Willer wrote the article. All the authors read and revised the article.

Sources of Funding

Dr Willer is supported by the National Institutes of Health (R01-HL127564, R35-HL135824, and R01-HL142023). Dr Surakka is supported by a Precision Health Scholars Award from the University of Michigan Medical School. Dr Ritche is funded by a British Heart Foundation (BHF) Programme Grant (RG/18/13/33943). Dr Sutton is supported by the National Institutes of Health (U1KCA064426-01A1). Dr Inouye is supported by the Munz Chair of Cardiovascular Prediction and Prevention and the National Institute for Health and Care Research (NIHR) Cambridge Biomedical Research Centre (BRC-1215-20014) [The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care]. HUNT-MI study, which comprises the genetic investigations of the HUNT2 (Trendelag Health Study), is a collaboration between investigators from the HUNT study and University of Michigan Medical School and the University of Michigan School of Public Health. The K.G. Jebsen Center for Genetic Epidemiology is financed by Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), and Central Norway Regional Health Authority. This research has been conducted using the UK Biobank Resource under Application Number 7439. This work was supported by core funding from the: BHF (RG/13/13/30194; RG/18/13/33943), Cambridge BHF Centre of Research Excellence (RE/13/6/30180), and NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. This research was also supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), BHF, and Wellcome.

Disclosures

Dr Willer’s spouse works for Regeneron Pharmaceuticals. Dr Sutton serves on governmental Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), BHF, and Wellcome.

Supplemental Material

Supplemental Methods

Tables S1–S14

Figures S1–S9

Text S1–S3

References 27–35

Figures

Supplemental Methods

Supplemental Material

Supplemental Information

References

1. Goff DC Jr, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, et al. 2014 AHA/ACC guideline for the assessment of cardiovascular risk: a report of the American college of cardiology/American heart association task force on practice guidelines. J Am Coll Cardiol 2014;63:2935–2959. doi: 10.1016/j.jacc.2013.11.005

2. Gibbons R, Greenland P, Lackland DT, Levy D, O'Donnell CJ, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American college of cardiology/American heart association task force on practice guidelines. J Am Coll Cardiol 2014;63:2935–2959. doi: 10.1016/j.jacc.2013.11.005

3. Piepoli MF, Hoes AW, Alsvgall S, Albus C, Brotons C, Catapano AL, Couney MT, Corrêa U, Cosyns B, Deaton C, et al. 2016 European guidelines on cardiovascular disease prevention in clinical practice: the sixth joint task force of the European society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of 10 societies and by invited experts) developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J 2016;37:2315–2381. doi: 10.1093/eurheartj/ehw106

4. Hippi-Sipilä J, Huvonen K, Kabir A, Jadon F, Johnson I, Kajantie J, et al. 2018. 10.1093/eurheartj/ehw106

5. Anderson KM, Wilson PW, Odell PM, Kannel WB. An updated coronary risk profile, a statement for health professionals. Circulation. 1991;83:356–362. doi: 10.1161/01.cir.83.2.356

6. Selmer R, Lindman AS, Tverdal A, Pedersen NJ, Njølstad I, Veierød MB. [Model for estimation of cardiovascular risk in Norway]. Tidsskr Nor Laegeforen. 2008;128:286–290.

7. Damen JA, Pauwels RA, Reuter Y, Hoeks S, Reitsma JB, Schooten RJPM, Hoof S, DeBray TPA, Performance of the Framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis. BMC Med. 2017;15:102. doi: 10.1186/s12916-017-1340-7

8. Wekesah FM, Muat MA, Boaingh D, Grobbee DE, Aski G, Kyobutungi CK, Kipstein-Grobush K. Comparative performance of pooled cohort equations and Framingham risk classification in cardiovascular disease risk classification in a slum setting in Nairobi Kenya. Int J Cardiol Heart Vasc. 2020;28:100521. doi: 10.1016/j.ijcha.2020.100521

9. Siontis GC, Toulouki I, Siontis KC, Ioannidis JP. Comparisons of established risk prediction models for cardiovascular disease: systematic review. BMJ 2012;344:e3318. doi: 10.1136/bmj.e3318

10. Sun LR, Pennells L, Kapトーe S, Nelson CP, Ritche SC, Abraham G, Arnold M, Bell S, Bolton T, Burgess S, et al. Polygenic risk scores in cardiovascular risk prediction: a cohort study and modelling analyses. PLoS Med 2021;18:e1003498. doi: 10.1371/journal.pmed.1003498

11. Khera AV, Chaffin M, Aragam KG, Haas ME, Rigamorti S, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equiva- lent to monogenic mutations. Nat Genet 2018;50:1219–1224. doi: 10.1038/s41588-018-0183-z

12. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, Li FY, Kapトーe S, Brozynska M, Wang T, et al. Genomic risk prediction of coronary artery disease in 480 000 adults: implications for primary prevention. J Am Coll Cardiol 2018;72:1883–1893. doi: 10.1016/j.jacc.2018.07.070

13. Elliott J, Bodinier B, Bond TA, Chaeade-Hyam M, Evangelou E, Moons KGM, Dehghan A, Muller DC, Elliott P, Toulouki I. Predictive accuracy of a polygenic risk score-enhanced prediction model vs a clinical risk score for coronary artery disease. JAMA 2020;323:636–646. doi: 10.1001/jama.2019.22241

14. Marks N, Kosiela JT, Ripatti P, Kiskinen TS, Havelinlin AS, Lindbom JV, Ahola-Olly A, Kurki M, Karjalainen J, Patta F, et al. Polygenetic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. Nat Med 2020;26:549–557. doi: 10.1038/s41591-020-0800-0

15. Mega JL, Stitziel NO, Smith DG, Chasman DI, Caufield M, Devlin JJ, Nordio F, Hyde C, Cannon CP, Sacks F, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. Lancet 2015;385:2264–2271. doi: 10.1016/S0140-6736(14)63730-X

16. Tada H, Melander O, Louie JZ, Catanesi JZ, Rowland CM, Devlin JJ, Kathiresan S, Shiffman D. Risk prediction by genetic risk scores for coro- nary heart disease is independent of self-reported family history. Eur Heart J 2017;38:561–567. doi: 10.1093/eurheartj/ehw492

17. Abraham G, Havelinlin AS, Bhalaia OG, Byars SG, De Livera AM, Yetukuri L, Tikkanen E, Perola M, Schunkert H, SJrandijs EJ, et al. Coronary disease: a systematic review and meta-analysis. Circ Genom Prev Med 2021;14:e003147. doi: 10.1161/CIRCGEN.120.003147

18. Peclihavins S, Lehmann N, Hoffmann P, Nöthem MM, Jöckel KH, Ertel R, Moebus S, Risk prediction for coronary heart disease by a genetic risk score - results from the Heinz Nixdorf recall study. BMC Med Genet 2020;21:178. doi: 10.1186/s12881-020-01113-y

19. Hajek C, Guo X, Yao J, Jai H, Johnson W, Frazier-Wood AC, Post WS, Pittay SB, Taylor KD, Rotter JI. Coronary artery disease genetic risk score predicts cardiovascular risk in men, not women. Circ Genom Prev Med 2018;11:e002234. doi: 10.1161/CIRCGEN.118.002234

20. El Khoudry SR, Aggarwal B, Beckie TM, Hods TD, Johnson AE, Langer RD, Limacher MC, Manson JE, Stefanick ML, Allison MA, et al. Menopause
transition and cardiovascular disease risk: implications for timing of early prevention: a scientific statement from the American heart association. Circulation. 2020;142:e506–e532. doi: 10.1161/CIR.0000000000000912

23. Jiang X, Holmes C, McVean G. The impact of age on genetic risk for common diseases. PLoS Genet. 2021;17:e1009723. doi: 10.1371/journal.pgen.1009723

24. Baart SJ, Dam V, Scheres LJJ, Damen JAAG, Spijker R, Schult E, Debray TPA, Fauser BCJM, Boersma E, Moons KGM, et al. Cardiovascular risk prediction models for women in the general population: a systematic review. PLoS One. 2019;14:e0210329. doi: 10.1371/journal.pone.0210329

25. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51:584–591. doi: 10.1038/s41588-019-0379-x

26. Wolford BN, Surakka I, Ritchie SC, Lambert S, Nielsen JB, Skogholt A, Gabrielsen ME, Brumpton B, Jonasson C, Hveem K, et al. Comprehensive benchmarking of integrated polygenic and conventional risk factor models for cardiovascular traits in the Nord-Trøndelag Health Study. (Abstract/1128). Paper presented at: 70th Annual Meeting of the American Society of Human Genetics, Oct 28th, Virtual 2020. Accessed Apr 13, 2021, URL: https://www.abstractsonline.com/pp8/#!/9070/presentation/1340

27. Krokkstad S, Langhammer A, Hveem K, et al. Cohort Profile: the HUNT Study, Norway. Int J Epidemiol. 2013;42:968–977. doi: 10.1093/ije/dys095

28. Zhou W, Fritsche LG, Das S, et al. Improving power of association tests using multiple sets of imputed genotypes from distributed reference panels. Genet Epidemiol. 2017;41:744–755. doi: 10.1002/gepi.22067

29. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47:121–130. doi: 10.1038/ng.3396

30. 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. doi: 10.1371/journal.pmed.1001779

31. 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;72:1883–1893. doi: 10.1016/j.jacc.2018.02.079

32. Eckel RH, Jakicic JM, Ard JD et al. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American college of cardiology/American heart association task force on practice guidelines. J Am Coll Cardiol. 2014;63:2960–2984. doi: 10.1016/j.jacc.2013.11.003

33. Khera R, Pandey A, Ayers CR, et al. Performance of the pooled cohort equations to estimate atherosclerotic cardiovascular disease risk by body mass index. JAMA Netw Open. 2020;3:e2023242. doi: 10.1001/jamanetworkopen.2020.23242

34. Albarqouni L, Doust JA, Maglione D, Barr EL, Shaw JE, Glasziou PP. External validation and comparison of four cardiovascular risk prediction models with data from the Australian diabetes, obesity and lifestyle study. Med J Australia. 2019;210:161–167. doi: 10.5694/mja2.12061

35. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American college of cardiology/American heart association task force on clinical practice guidelines. J Am Coll Cardiol. 2019;73:e285–e350. doi: 10.1016/j.jacc.2018.11.003