Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

External validation and comparison of PREVENT and SCORE2 atherosclerotic cardiovascular risk scores in the MASHAD cohort study

Ali Tajik
Mashhad University of Medical Sciences
Hamidreza Rahimi
Mashhad University of Medical Sciences
Sara Saffar Soflaei
Mashhad University of Medical Sciences
Mohsen Moohebati
Mashhad University of Medical Sciences
Ashkan Bahrami
Isfahan University of Medical Sciences
Gordon A. Ferns
Brighton and Sussex Medical School
Habibollah Esmaily
Mashhad University of Medical Sciences
Majid Ghayour-Mobarhan
Mashhad University of Medical Sciences
Hedieh Alimi
Alimih@mums.ac.ir

Mashhad University of Medical Sciences

Research Article

Keywords: PREVENT risk score, SCORE2, Risk recalibration, External validation, Cohort

Posted Date: June 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10038994/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Contemporary atherosclerotic cardiovascular disease (ASCVD) risk prediction models require external validation before implementation in non-derivation populations. We externally validated and recalibrated PREVENT and SCORE2 in a large Middle Eastern cohort.

Methods

This prospective study included 9,069 adults from the Mashhad stroke and heart atherosclerotic disorder (MASHAD) cohort followed for 10 years. PREVENT was evaluated in the full eligible cohort and SCORE2 in 7,836 participants without diabetes. Sex-specific predictions were generated using published equations. Local recalibration preserved original coefficients using intercept updating for PREVENT and baseline-hazard updating for SCORE2. Performance was assessed using C-statistics, calibration slopes, observed-to-expected ratios, Brier scores, threshold-based metrics, categorical net reclassification improvement, Kaplan–Meier curves, and decision curve analysis.

Results

Within the 10-year analytic horizon, 678 ASCVD events occurred in the PREVENT sample and 470 in the SCORE2 sample. PREVENT showed modestly higher discrimination than SCORE2 in men (C-statistic: 0.714 [95% CI: 0.661–0.768] vs. 0.687 [0.622–0.752]) and women (0.731 [0.679–0.783] vs. 0.704 [0.639–0.769]). Both original models underestimated risk (overall O/E: PREVENT, 2.904; SCORE2, 4.002). Sex-specific recalibration corrected calibration-in-the-large for both models (O/E ≈ 1.00), improved Brier scores, sensitivity, net benefit, and categorical reclassification, while C-statistics remained unchanged.

Conclusions

PREVENT and SCORE2 preserved acceptable ASCVD risk-ranking ability in this Iranian cohort but required local recalibration for accurate absolute risk estimation. PREVENT showed modestly better discrimination and calibration slopes. Locally calibrated risk tools may improve ASCVD prevention in underrepresented regional populations.

Introduction

Cardiovascular disease (CVD) remains the leading cause of global mortality: in 2022, an estimated 19.8 million deaths, approximately 32% of all deaths worldwide, were attributable to CVD, and more than three-quarters occurred in low- and middle-income countries (1). In Iran, Global Burden of Disease 2021

estimates indicate that all-age CVD prevalence rose from 2.9 million in 1990 to 8.3 million in 2021, underscoring the need for population-specific prevention strategies (2).

Accurate estimation of absolute atherosclerotic cardiovascular disease (ASCVD) risk is central to preventive cardiology, lipid management, blood pressure control, and rational allocation of therapies. Current AHA/ACC and ESC guidelines recommend multivariable risk prediction to guide primary-prevention treatment intensity, particularly statin and antihypertensive therapy (3, 4). Earlier tools such as the Framingham Risk Score and pooled cohort equations established 10-year ASCVD risk estimation, whereas SCORE2 and PREVENT were developed to reflect contemporary epidemiology and improve risk stratification (5–7). However, prediction models are not automatically transportable: acceptable discrimination may coexist with poor calibration when baseline event rates, competing risks, treatment access, or risk-factor distributions differ from derivation cohorts (8, 9).

PREVENT was developed by the American Heart Association and incorporates cardiometabolic and kidney-related predictors, including body mass index, estimated glomerular filtration rate, diabetes, smoking, blood pressure, lipids, and medication use (5). SCORE2, developed by the ESC Cardiovascular Risk Collaboration, estimates 10-year fatal and non-fatal cardiovascular risk in adults without diabetes and is incorporated into European prevention guidance (3, 6, 7). Prior validations show that cardiovascular calculators frequently require population-specific recalibration, because miscalibration can alter treatment eligibility and preventive decision-making (8, 10–12). In the Tehran Lipid and Glucose Study, PREVENT showed acceptable-to-excellent discrimination but underestimated risk in men and improved after recalibration (13). In PURE-Colombia, several calculators had similar discrimination but substantial overprediction, again supporting local correction before implementation (14).

However, it remains unclear whether PREVENT and SCORE2 provide reliable ASCVD risk estimation in northeastern Iran, where cardiometabolic burden, smoking patterns, socioeconomic status, healthcare access, and preventive medication use may differ from Western cohorts and prior Iranian samples. Limited evidence exists regarding their head-to-head performance in a large Middle Eastern cohort with adjudicated long-term ASCVD outcomes. A critical gap also persists in determining whether recalibration improves calibration, risk stratification, reclassification, and decision-analytic utility.

Filling this gap is clinically important because uncalibrated calculators may under-identify high-risk individuals or expose lower-risk individuals to unnecessary treatment. Therefore, the aim of this study was to externally validate and compare the performance of PREVENT and SCORE2 for prediction of 10-year ASCVD risk in the MASHAD cohort, with secondary evaluation of recalibration, risk stratification, reclassification, and decision-analytic utility.

Methods

This external validation study was reported in accordance with the TRIPOD statement for prediction model validation; the completed checklist is provided in Supplementary Table 1 (15).

Study Population

Participants were selected from the Mashhad Stroke and Heart Atherosclerotic Disorder (MASHAD) cohort, an ongoing prospective study involving 9,704 residents of Mashhad, Iran, aged 30 to 65 years at baseline (16). In the current investigation, participants were monitored throughout a decade (2010–2020). From the initial cohort of 9,704 participants, exclusions were implemented sequentially: 201 with pre-existing cardiovascular disease (CVD) at baseline, 29 with total cholesterol levels below 130 or above 320 mg/dL, 8 with HDL cholesterol levels below 20 or above 100 mg/dL, 70 with systolic blood pressure below 90 or above 200 mmHg, 76 with body mass index below 18.5 or equal to or above 40 kg/m², 41 with estimated glomerular filtration rate below 15 or equal to or above 150 mL/min/1.73 m², 78 with incomplete baseline data, and 132 without follow-up for CVD outcomes. Subsequent to these exclusions, 9,069 individuals were retained and monitored until 2020, during which 912 ASCVD occurrences transpired (470 in women, 442 in males).

SCORE2 is designed for persons without diabetes; thus, participants having a prior history of diabetes at baseline (n = 1,233) were eliminated from the SCORE2 assessment, resulting in a final analytical cohort of 7,836 non-diabetic adults. Conversely, the PREVENT model was used over the complete follow-up sample (n = 9,069), encompassing both diabetes and non-diabetic participants. The Flow diagram of the study is shown is Supplementary figure 1.

Baseline Assessment

Demographic, anthropometric, and lifestyle data were gathered by two trained health workers and a nurse using structured questionnaires. These instruments captured information on demographic background, social and medical history, cardiovascular risk factors, and relevant family history of disease.

For biochemical measurements, venous blood was drawn following a 14-hour overnight fast. Samples were collected between 8:00 and 10:00 a.m. while participants were seated, using 20-mL vacuum tubes, and centrifuged within 30–45 minutes. Serum and plasma were separated into six aliquots and transported to the Bu Ali Research Institute in Mashhad. Additional serum aliquots were preserved at −80 °C for later analyses. Low-density lipoprotein cholesterol (LDL-C) was calculated from total cholesterol (TC), triglycerides (TG), and high-density lipoprotein (HDL) using the Friedewald formula, applied when TG levels were below 400 mg/dL.

Anthropometric measurements were performed by trained nurses. Participants wore light clothing and no shoes. Height was measured to the nearest 0.1 cm using a stadiometer, body weight to the nearest 0.1 kg with an electronic scale, and waist and hip circumferences using a flexible tape.

Follow-up Procedures

Participants were followed for a minimum of ten years; follow-up contacts occurred every three years to minimize attrition. At each wave, questionnaires were administered to record changes in health status

and lifestyle behaviors, and additional information was collected as required for specific sub studies. Cardiovascular outcomes and endpoints, including coronary artery diseases, were continuously monitored through community health records.

Definitions

Hypertension was defined as a systolic blood pressure of at least 140 mmHg, a diastolic blood pressure of at least 90 mmHg, or current use of antihypertensive medication. Diabetes was identified by a fasting plasma glucose (FPG) level exceeding 126 mg/dL or the use of glucose-lowering agents. Smoking status was categorized to distinguish current daily smokers from users of non-cigarette tobacco products.

Details regarding cardiovascular outcome ascertainment in the MASHAD cohort have been reported previously (16). In brief, participants were contacted annually by telephone to identify hospitalizations or deaths related to cardiovascular events. Initial information on reported events was collected by a trained nurse; when necessary, a physician-led home visit was conducted to review medical records and supporting documentation. All potential events were adjudicated by an independent outcomes committee comprising specialists in internal medicine, endocrinology, cardiology, epidemiology, and pathology, applying standardized criteria consistently throughout follow-up.

The primary outcome of this analysis was an ASCVD event, defined as a composite of fatal or non-fatal myocardial infarction (MI) and fatal or non-fatal stroke, consistent with the outcome definition used in the original PREVENT model. To ensure reproducibility, ASCVD events were operationalized using International Classification of Diseases, 10th Revision (ICD-10) codes derived from hospital discharge records and death certificates . MI was defined by codes I21 and I22, and stroke (ischemic or hemorrhagic) by codes I61, I62, and I63.

Risk scores

PREVENT is a sex-specific cardiovascular disease (CVD) risk prediction model designed to assess the absolute 10- and 30-year risks of atherosclerotic cardiovascular disease (ASCVD), heart failure (HF), and associated consequences. The foundational PREVENT model includes age, BMI, HDL-C, total cholesterol (TC), systolic blood pressure (SBP), eGFR, type 2 diabetes, current smoking status, antihypertensive therapy, and statin administration. The comprehensive model also incorporates HbA1c, the social deprivation index, and the urinary albumin-to-creatinine ratio. Risk estimates were initially obtained by cause-specific competing-risks modelling with sex-stratified Cox proportional hazards models on the age time scale, incorporating left truncation and right censoring. A logistic regression approximation of the competing-risks model was developed for dissemination and implementation, fitting the logit of the absolute 10-year risk with the identical set of predictors. This logistic formulation exhibited near-perfect concordance with the comprehensive survival-based model and was selected as the definitive operational version of the PREVENT equations. This work conducted external validation just utilising the published logistic PREVENT equations, without any refitting or alteration of model structure or coefficients. SCORE2 is a sex-specific CVD risk prediction model developed by the SCORE2 Working

Group and the European Society of Cardiology (ESC) Cardiovascular Risk Collaboration to estimate the absolute 10-year risk of fatal and non-fatal cardiovascular events (myocardial infarction, stroke, and cardiovascular death). The model incorporates age, sex, smoking status, systolic blood pressure, and non-HDL cholesterol (or total cholesterol and HDL-C) as core predictors, with diabetes included in extended versions. In the present study, external validation was performed exclusively using the published SCORE2 equations, without refitting or modification of model structure or coefficients. This approach for both models is consistent with their intended clinical use and with prior external validation studies.

Statistical analysis

Continuous variables were assessed for distributional normality and are presented as median and interquartile range (IQR), while categorical variables are reported as frequencies and percentages. Baseline characteristics were summarized for the overall study population and stratified by sex.

The performance of the PREVENT and SCORE2 risk prediction models for incident 10-year ASCVD was externally validated in the MASHAD cohort. PREVENT was evaluated in the full eligible analytic cohort, whereas SCORE2 was evaluated after excluding participants with diabetes mellitus at baseline, consistent with the intended target population of the original SCORE2 model. For PREVENT, the ASCVD Base 10-year equation was used. Sex-specific linear predictors were calculated for both PREVENT and SCORE2 using the original published equations and coefficients.

Model recalibration was performed to improve agreement between predicted and observed 10-year ASCVD risk in the study population while preserving the original model coefficients. For PREVENT, recalibration was performed using sex-specific intercept updating. Specifically, a logistic recalibration model including the original PREVENT linear predictor as an offset term was fitted separately by sex to estimate a local recalibration shift, which was added to the original linear predictor. For SCORE2, recalibration was performed separately by sex using Cox proportional hazards regression with the original SCORE2 linear predictor included as an offset term, allowing estimation of a population-specific 10-year baseline hazard while maintaining the original predictor coefficients.

Discrimination was evaluated separately by sex using C-statistics with corresponding 95% confidence intervals. Because recalibration preserved the original predictor coefficients and therefore did not alter individual risk ranking, C-statistics were expected to remain unchanged after recalibration.

Calibration was assessed using calibration slopes, observed-to-expected event ratios, Brier scores, and graphical calibration plots. Calibration slopes were estimated by modelling the observed time-to-event outcome against the original model linear predictor, with an ideal slope of 1.0 indicating appropriate spread of predicted risks. Prediction error was quantified using 10-year inverse probability of censoring weighted Brier scores with 95% confidence intervals. Observed-to-expected ratios were calculated overall and separately by sex by comparing Kaplan–Meier estimated 10-year observed ASCVD events with the total number of predicted events. An observed-to-expected ratio close to 1.0 indicated optimal

calibration, values greater than 1.0 indicated underestimation of risk, and values below 1.0 indicated overestimation.

Clinical performance was assessed at predefined risk thresholds by calculating the proportion of participants classified above each threshold, observed event rate among those classified as high risk, sensitivity, specificity, positive predictive value, negative predictive value, and net benefit. Net benefit was calculated as TP/N − FP/N × threshold/(1 − threshold), where TP denotes true positives, FP false positives, and N the number of participants in the relevant analysis group.

Risk stratification analyses were performed using clinically relevant categories. For PREVENT, participants were categorized as low risk (<5%), borderline risk (5%–<7.5%), intermediate risk (7.5%–<20%), and high risk (≥20%). For SCORE2, age-dependent thresholds were applied: for participants aged <50 years, low-to-moderate risk was defined as <2.5%, high risk as 2.5%–<7.5%, and very high risk as ≥7.5%; for participants aged 50-69 years, corresponding thresholds were <5%, 5%–<10%, and ≥10%. Kaplan–Meier survival curves were constructed across recalibrated risk categories and compared using the log-rank test.

Clinical utility was further evaluated using decision curve analysis. Decision curves were generated separately by sex for original and recalibrated PREVENT and SCORE2 predictions and compared with treat-all and treat-none strategies across clinically relevant threshold probabilities.

Categorical net reclassification improvement was calculated to assess whether local recalibration improved assignment of participants into clinically relevant risk categories compared with the original models. For PREVENT, reclassification was based on the 5%, 7.5%, and 20% risk thresholds. For SCORE2, age-specific risk categories were used. Reclassification results were interpreted alongside calibration, Brier scores, clinical threshold performance, and decision curve analysis because reclassification metrics may be sensitive to the selected thresholds.

All statistical analyses were conducted using R software version 4.5.1 (R Foundation for Statistical Computing, Vienna, Austria). Major R packages used in the analyses included survival, survminer, riskRegression, rmda, binom, ggplot2, dplyr, and related data-management packages. A two-sided p-value <0.05 was considered statistically significant.

Results

Baseline characteristics of the study population

A total of 9,069 participants were included in the present analysis, comprising 3,584 men (39.5%) and 5,485 women (60.5%). The median age of the study population was 47.0 years (IQR: 41.0–54.0 years). Overall, hypertension and diabetes mellitus were present in 30.5% and 13.6% of participants, respectively, while 21.6% were smokers. A positive family history of cardiovascular disease (CVD) was reported in 34.6% of the population.

Men had higher frequencies of smoking and employment, as well as higher median levels of uric acid, hemoglobin, and hematocrit. Women had higher median body mass index, total cholesterol, high-density lipoprotein cholesterol, anxiety score, and depression score. Additional baseline demographic, anthropometric, laboratory, and clinical characteristics stratified by sex are summarized in Table 1.

During the total available follow-up through 2020, 912 incident ASCVD events were observed in the eligible cohort. Within the 10-year analytic horizon, 678 observed ASCVD events occurred in the PREVENT analytic sample, including 333 events in men and 345 events in women. For SCORE2 analyses, participants with diabetes mellitus at baseline were excluded according to SCORE2 recommendations, leaving 7,836 non-diabetic participants with 470 observed 10-year ASCVD events, including 240 events in men and 230 events in women.

Discrimination, calibration slope, and prediction error

The discrimination, calibration slope, and prediction error of PREVENT and SCORE2 are summarized in Table 2. PREVENT demonstrated acceptable discrimination in both sexes. Among men, the C-statistic was 0.714 (95% CI: 0.661–0.768) for both the original and recalibrated PREVENT models. Among women, the corresponding C-statistic was 0.731 (95% CI: 0.679–0.783). Recalibration did not change the C-statistics because the original model coefficients were preserved and only the intercept or baseline-risk component was updated.

SCORE2 showed slightly lower discrimination than PREVENT. Among men, the C-statistic was 0.687 (95% CI: 0.622–0.752), and among women it was 0.704 (95% CI: 0.639–0.769), with identical values before and after recalibration.

Calibration slope analyses demonstrated imperfect calibration for both models, with slopes below 1.0. PREVENT showed calibration slopes of 0.861 (95% CI: 0.735–0.988) in men and 0.813 (95% CI: 0.710–0.917) in women. SCORE2 showed lower calibration slopes of 0.754 (95% CI: 0.603–0.905) in men and 0.700 (95% CI: 0.575–0.825) in women.

Brier score analyses showed improved prediction error after recalibration in both models. For PREVENT, the Brier score improved from 9.47% (95% CI: 8.49–10.45) to 8.28% (95% CI: 7.42–9.15) in men and from 6.76% (95% CI: 6.06–7.46) to 6.25% (95% CI: 5.59–6.90) in women. For SCORE2, the Brier score improved from 8.00% (95% CI: 7.02–8.99) to 7.17% (95% CI: 6.28–8.06) in men and from 5.35% (95% CI: 4.67–6.04) to 4.95% (95% CI: 4.32–5.59) in women.

Observed-to-expected event ratios and graphical calibration

Observed-to-expected event ratio analyses demonstrated substantial underestimation of 10-year ASCVD risk by both original models. In the overall PREVENT analytic sample, the original PREVENT model had an O/E ratio of 2.904, with sex-specific O/E ratios of 3.004 in men and 2.817 in women. After sex-specific local recalibration, the O/E ratio improved to approximately 1.000 overall and within each sex.

SCORE2 also substantially underestimated observed ASCVD risk. In the non-diabetic SCORE2 analytic sample, the original SCORE2 model had an overall O/E ratio of 4.002, with O/E ratios of 3.393 in men and 4.902 in women. After sex-specific local recalibration, the O/E ratio improved to approximately 1.000 overall and within each sex.

Graphical calibration plots (Figure 2) similarly demonstrated that the original PREVENT and SCORE2 models underestimated observed 10-year ASCVD risk, while recalibration improved calibration-in-the-large. However, calibration across the risk distribution remained imperfect, consistent with calibration slopes below 1.0.

Clinical performance at predefined risk thresholds

The clinical performance of PREVENT at predefined ASCVD risk thresholds is summarized in Supplementary Table 8. Recalibration substantially increased the proportion of participants classified above clinically relevant thresholds and improved sensitivity, although this occurred with reduced specificity.

Among men, at the 5% PREVENT threshold, recalibration increased the proportion classified above the threshold from 21.3% to 67.0%. Sensitivity increased from 46.2% (95% CI: 41.0–51.6) to 90.1% (95% CI: 86.4–92.9), while specificity decreased from 81.2% (95% CI: 79.8–82.5) to 35.4% (95% CI: 33.7–37.0). Net benefit increased from 0.0340 to 0.0529. At the 7.5% threshold, sensitivity increased from 24.9% to 77.2%, specificity decreased from 91.6% to 53.2%, and net benefit increased from 0.0170 to 0.0373.

Among women, at the 5% PREVENT threshold, recalibration increased the proportion classified above the threshold from 13.2% to 45.8%. Sensitivity increased from 36.2% (95% CI: 31.3–41.4) to 74.2% (95% CI: 69.3–78.5), while specificity decreased from 88.3% (95% CI: 87.4–89.2) to 56.1% (95% CI: 54.8–57.5). Net benefit increased from 0.0170 to 0.0250. At the 7.5% threshold, sensitivity increased from 19.7% to 65.5%, specificity decreased from 94.8% to 70.8%, and net benefit increased from 0.0084 to 0.0190. At the 20% PREVENT threshold, net benefit was small and confidence intervals included or approached zero, particularly among women.

The clinical performance of SCORE2 using age-specific thresholds is shown in Supplementary Table 9. Similar to PREVENT, recalibration increased the proportion classified into higher-risk categories and improved sensitivity, with corresponding reductions in specificity.

Among men aged less than 50 years, at the 2.5% SCORE2 threshold, recalibration increased the high-risk proportion from 7.0% to 70.9%. Sensitivity increased from 14.1% to 82.4%, specificity decreased from 93.3% to 29.7%, and net benefit increased from 0.0050 to 0.0216. Among women aged less than 50 years, sensitivity at the 2.5% threshold increased from 5.2% to 63.5%, specificity decreased from 99.5% to 65.9%, and net benefit increased from 0.0015 to 0.0119.

Among participants aged 50 years or older, recalibration produced larger shifts in risk classification. In men, at the 5% threshold, the proportion classified above the threshold increased from 28.7% to 96.1%,

sensitivity increased from 41.9% to 98.1%, and specificity decreased from 73.1% to 4.2%. Net benefit increased from 0.0365 to 0.0702. At the 10% threshold, sensitivity increased from 5.2% to 78.1%, specificity decreased from 96.8% to 40.5%, and net benefit increased from 0.0029 to 0.0330. In women aged 50 years or older, at the 5% threshold, sensitivity increased from 17.2% to 94.8%, specificity decreased from 94.3% to 19.8%, and net benefit increased from 0.0107 to 0.0356. At the 10% threshold, sensitivity increased from 1.5% to 58.2%, specificity decreased from 99.3% to 62.6%, and net benefit increased from 0.0005 to 0.0074.

Kaplan–Meier survival analyses

Kaplan–Meier survival analyses demonstrated progressive separation of ASCVD-free survival curves across recalibrated risk categories (Figure 1A-C). For the recalibrated PREVENT model, participants in higher-risk categories showed lower ASCVD-free survival over follow-up compared with those in lower-risk categories. Similar graded separation was observed for recalibrated SCORE2 categories in both younger and older age groups. Log-rank testing demonstrated significant differences in survival distributions across risk strata for all three Kaplan–Meier analyses.

Net reclassification improvement

Categorical net reclassification improvement after local recalibration is summarized in Table 3. For PREVENT, overall NRI was 0.296 (95% CI: 0.263–0.329). The event NRI was 0.813 (95% CI: 0.782–0.843), indicating that 81.3% of participants who experienced ASCVD events moved upward into higher risk categories after recalibration. The non-event NRI was −0.517 (95% CI: −0.527 to −0.507), reflecting upward movement among participants without events. Sex-specific total NRI values were 0.257 (95% CI: 0.220–0.292) in men and 0.291 (95% CI: 0.245–0.340) in women.

For SCORE2, overall NRI was 0.236 (95% CI: 0.201–0.269), with an event NRI of 0.851 (95% CI: 0.819–0.883) and a non-event NRI of −0.615 (95% CI: −0.626 to −0.604). Sex-specific total NRI was 0.098 (95% CI: 0.054–0.138) in men and 0.309 (95% CI: 0.255–0.360) in women. In age-stratified analyses, total NRI was 0.248 (95% CI: 0.185–0.316) among participants aged less than 50 years and 0.080 (95% CI: 0.048–0.110) among those aged 50 years or older. Among men aged 50 years or older, total NRI was 0.003 (95% CI: −0.041 to 0.045), indicating no meaningful net improvement in risk classification in this subgroup.

Decision curve analysis

Decision curve analysis demonstrated that recalibrated models generally provided greater net benefit than their original counterparts at lower and intermediate threshold probabilities (Figure 3). This pattern was observed for both PREVENT and SCORE2 in men and women. The improvement in net benefit was most evident across threshold ranges where preventive treatment decisions are clinically plausible. At higher threshold probabilities, the net benefit of all models approached zero, and the relative advantage of recalibrated models became less pronounced.

Discussion

The present study provides a large prospective external validation and direct comparison of PREVENT and SCORE2 for 10-year ASCVD risk prediction in the MASHAD cohort. The findings support the hypothesis that contemporary risk scores preserve clinically useful risk-ranking ability in this Middle Eastern population but require local recalibration to generate appropriate absolute risk estimates. PREVENT showed modestly stronger discrimination than SCORE2, with C-statistics higher by approximately 0.027 in both men and women. Recalibration improved calibration-in-the-large, prediction error, sensitivity at clinically relevant thresholds, categorical reclassification, and decision-analytic net benefit without changing discrimination. Thus, the principal limitation of the original models was not loss of prognostic signal, but misalignment between the original baseline-risk structure and local ASCVD incidence.

The slightly better performance of PREVENT is biologically and clinically plausible. Unlike SCORE2, PREVENT incorporates diabetes, estimated glomerular filtration rate, antihypertensive treatment, and statin use in addition to age, lipid measures, systolic blood pressure, and smoking. These predictors are highly relevant in MASHAD, where hypertension and diabetes were present in 30.5% and 13.6% of participants, respectively, while baseline statin use was only 1.3%. This cardiometabolic context may help explain why PREVENT showed modestly higher discrimination and better calibration slopes than SCORE2.

Calibration analyses demonstrated substantial underestimation of 10-year ASCVD risk by both original models. The original PREVENT model had an overall observed-to-expected ratio of 2.904, while the original SCORE2 model showed more pronounced underestimation, with an O/E ratio of 4.002. After sex-specific local recalibration, O/E ratios were approximately 1.00 overall and within each sex for both models. This indicates that local updating successfully corrected calibration-in-the-large while preserving the original model coefficients.

The clinical consequences of recalibration were substantial. For PREVENT, at the 5% threshold, recalibration increased sensitivity from 46.2% to 90.1% in men and from 36.2% to 74.2% in women. Net benefit also increased in both sexes, although specificity decreased as more participants were moved into actionable risk categories. SCORE2 showed a similar pattern. Among participants aged 50 years or older, recalibration increased sensitivity at the 5% threshold from 41.9% to 98.1% in men and from 17.2% to 94.8% in women. However, these improvements should be interpreted alongside the marked reductions in specificity, particularly among older men, where recalibration classified nearly all participants above the 5% threshold.

Comparison with prior validation studies reinforces the population-specific nature of miscalibration. In PURE-Colombia (14), several cardiovascular risk calculators overestimated risk, whereas in MASHAD both PREVENT and SCORE2 substantially underestimated 10-year ASCVD risk before recalibration. This contrast is informative: the direction and magnitude of miscalibration depend on local event incidence, risk-factor distribution, treatment uptake, socioeconomic structure, competing risks, and outcome

ascertainment. Therefore, correction factors from other settings should not be imported directly without empirical validation in the target population.

The Kaplan–Meier, reclassification, and decision-curve analyses extend the interpretation beyond C-statistics (17, 18). Recalibrated risk categories showed graded separation of ASCVD-free survival. Categorical NRI was positive for both models, indicating net improvement in risk-category assignment after recalibration, although upward movement among non-events partly offset the benefit of upward movement among events. Decision curve analysis further showed that recalibrated models generally provided greater net benefit than original models at lower and intermediate threshold probabilities. Together, these findings suggest that local recalibration improved the clinical usability of both prediction tools, particularly for identifying individuals who would otherwise be underclassified by the original equations.

The slightly better performance of PREVENT is biologically and clinically plausible. Unlike SCORE2, PREVENT incorporates diabetes, estimated glomerular filtration rate, antihypertensive treatment, and statin use in addition to age, lipid measures, systolic blood pressure, and smoking (5). These predictors are highly relevant in MASHAD, where hypertension and diabetes were present in 30.5% and 13.6% of participants, respectively, while baseline statin use was only 1.3%. This cardiometabolic context may help explain why PREVENT showed modestly higher discrimination than SCORE2, with C-statistics higher by approximately 0.027 in both men and women, as well as better calibration slopes. External validations similarly suggest that PREVENT maintains useful discrimination but still requires local calibration (19, 20).

Our results are consistent with, but not identical to, the Tehran Lipid and Glucose Study validation of PREVENT. In TLGS, PREVENT showed excellent discrimination in women and acceptable discrimination in men, with AUCs of 0.84 and 0.76, respectively, and recalibration increased mean predicted risk, particularly among men, from 4.2% to 7.6% (13). Recalibration also increased sensitivity at the 5% threshold from 62.4% to 79.9% in men and from 69.5% to 86.5% in women. In MASHAD, discrimination was more moderate, with C-statistics of 0.731 in women and 0.714 in men, approximately 0.109 and 0.046 lower than TLGS, respectively. However, recalibration produced substantial clinical gain: at the 5% PREVENT threshold, sensitivity increased from 46.2% to 90.1% in men and from 36.2% to 74.2% in women. This indicates that although PREVENT ranked individuals less strongly in MASHAD than in TLGS, local recalibration meaningfully improved identification of participants who subsequently developed ASCVD, supporting the presence of regional heterogeneity even within Iran.

SCORE2 performance should be interpreted according to its intended use. SCORE2 was developed from European cohorts for adults without diabetes and uses a parsimonious predictor set based on age, sex, smoking, systolic blood pressure, total cholesterol, and HDL cholesterol (6). We therefore restricted SCORE2 analyses to participants without diabetes mellitus at baseline, preserving methodological fidelity. Despite this restriction, SCORE2 showed slightly lower discrimination and weaker calibration slopes than PREVENT. The C-statistics for SCORE2 were 0.687 in men and 0.704 in women, compared

with 0.714 and 0.731 for PREVENT. Similarly, calibration slopes were 0.754 in men and 0.700 in women for SCORE2, compared with 0.861 and 0.813 for PREVENT. These findings suggest that although SCORE2 preserved clinically useful risk-ranking ability, its risk gradient was less well aligned with observed ASCVD outcomes in MASHAD than PREVENT. Observed-to-expected ratio analyses showed that the original SCORE2 model substantially underestimated absolute ASCVD risk, more markedly than PREVENT. In the overall SCORE2 analytic sample, the original O/E ratio was 4.002, compared with 2.904 for PREVENT. Underestimation was particularly pronounced among women, with an original SCORE2 O/E ratio of 4.902. However, after sex-specific local recalibration, SCORE2 achieved near-ideal calibration-in-the-large, with O/E ratios of approximately 1.00 overall and within both sexes. Therefore, the main limitation of SCORE2 in this cohort was not persistent failure after recalibration, but the poor transportability of its original European baseline-risk structure to the local Iranian population.

Our findings are consistent with the Tehran Lipid and Glucose Study external validation of SCORE2, which found good discrimination, particularly in women, but reported underestimation of CVD risk in both sexes and recommended recalibration and updating using national data (21). In TLGS, SCORE2 achieved a C-index of 0.80 in women and 0.70 in men, with O/E ratios of 1.50 and 2.05, respectively. Compared with TLGS, SCORE2 showed lower discrimination and stronger underestimation in MASHAD, especially among women. Together, these findings suggest that SCORE2 requires population-specific validation and recalibration before clinical implementation in Iran. Similar concerns have been reported in ethnically and socioeconomically diverse European populations, where SCORE2 calibration varied across population subgroups (22).

A central contribution of this study is demonstrating that recalibration improved clinically meaningful performance, not only statistical calibration. Discrimination and calibration address different questions: a model may rank individuals adequately while generating misleading absolute risks when transported to a population with different event incidence, competing risks, treatment access, or risk-factor distribution (23–25). For PREVENT, recalibration substantially increased the proportion of participants classified above clinically relevant thresholds and improved sensitivity. At the 5% threshold, the proportion classified above the threshold increased from 21.3% to 67.0% in men and from 13.2% to 45.8% in women. Sensitivity increased from 46.2% to 90.1% in men and from 36.2% to 74.2% in women, while net benefit increased from 0.0340 to 0.0529 in men and from 0.0170 to 0.0250 in women. This improvement occurred at the expected cost of lower specificity, as more individuals were moved into actionable risk categories. In primary prevention, this trade-off may be acceptable when the aim is to reduce missed high-risk individuals, provided treatment decisions remain individualized and are considered alongside patient preferences, comorbidity, and local treatment thresholds.

SCORE2 recalibration further highlights the need for age- and sex-specific validation. Among participants aged 50–65 years, recalibration markedly increased sensitivity at the 5% threshold, from 41.9% to 98.1% in men and from 17.2% to 94.8% in women. Net benefit increased in parallel, from 0.0365 to 0.0702 in men and from 0.0107 to 0.0356 in women. However, these gains were accompanied by substantial reductions in specificity, particularly among men, where specificity decreased from 73.1% to 4.2% at the

5% threshold. These findings suggest that the original SCORE2 model substantially under-identifies individuals at future ASCVD risk in MASHAD, but also that recalibrated SCORE2 thresholds require careful clinical interpretation because improved sensitivity may come with broad upward classification. Therefore, SCORE2 risk thresholds should not be transferred directly from European guidance without empirical validation and local calibration.

Comparison with prior validation studies reinforces the population-specific nature of miscalibration. In contemporary U.S. cohorts, pooled cohort equations often overestimated ASCVD risk and altered treatment eligibility (26–28). In PURE-Colombia, six calculators showed similar discrimination but substantial overprediction, reaching 61% for Globorisk-LAC and 59% for WHO in women, and 72% for Framingham and 71% for AHA/ACC PCE in men (14). In contrast, both original PREVENT and SCORE2 substantially underestimated 10-year ASCVD risk in MASHAD, and local recalibration shifted predicted risks upward, increasing high-risk classification and sensitivity. This apparent disagreement is informative rather than contradictory: the direction and magnitude of recalibration depend on local event rates, risk-factor distributions, socioeconomic structure, medication uptake, competing risks, healthcare access, and outcome ascertainment. Therefore, correction factors or recalibration parameters derived from one setting should not be imported into another population without direct empirical validation.

The Kaplan–Meier, reclassification, and decision-curve analyses extend the interpretation beyond C-statistics. Recalibrated risk categories showed graded separation of ASCVD-free survival, supporting their ability to stratify long-term event risk. Categorical reclassification analyses showed positive total NRI for both PREVENT and SCORE2, indicating net improvement in risk-category assignment after recalibration. However, this improvement should be interpreted carefully: recalibration shifted many future ASCVD cases upward into higher-risk categories, but it also shifted a substantial proportion of non-events upward, partly offsetting the event-based gain. Decision curve analysis further showed that recalibrated models generally provided greater net benefit than their original counterparts at lower and intermediate threshold probabilities. These findings are clinically relevant because risk calculators inform preventive decisions such as statin initiation, blood pressure intensification, lifestyle counseling, and follow-up intensity, rather than serving only as tools for risk ranking (29). Overall, local recalibration improved the clinical usability of both models by increasing detection of future ASCVD cases and improving net benefit, although this occurred with lower specificity and broader upward risk classification. Key strengths of this study include its prospective cohort design, large analytic sample, adjudicated ASCVD outcomes, conservative recalibration without refitting original predictor coefficients, and multidimensional validation using discrimination, calibration, observed-to-expected ratios, Brier scores, threshold-based clinical performance, Kaplan–Meier risk stratification, categorical reclassification, and decision curve analysis (23, 24).

Clinical Implications

PREVENT and SCORE2 may support ASCVD risk stratification in northeastern Iran, but their original risk estimates should be interpreted cautiously because both models substantially underestimated 10-year

ASCVD risk in MASHAD. Local recalibration improved absolute-risk agreement, sensitivity, risk-category assignment, and decision-analytic net benefit, although at the cost of lower specificity. Recalibrated PREVENT may be preferable when cardiometabolic, kidney-function, and medication-use data are available, whereas SCORE2 remains a simpler option for adults without diabetes if locally recalibrated. Before clinical or national implementation, these recalibrated equations should be independently validated in other Iranian and Middle Eastern populations and evaluated for their effects on treatment decisions, risk-factor control, and ASCVD outcomes.

Limitations

This study has several limitations. First, although MASHAD is a large prospective cohort, it represents a regional urban population from northeastern Iran; therefore, recalibrated estimates may not generalize directly to rural populations, other Iranian provinces, or other Middle Eastern countries. This limitation is particularly relevant for calibration, because baseline event rates, treatment access, and preventive medication patterns may vary geographically. Second, SCORE2 was restricted to participants without diabetes mellitus according to its intended use, strengthening methodological validity but limiting direct comparability with PREVENT. Third, we validated the published PREVENT ASCVD Base 10-year equation; additional optional variables available in some calculator implementations were not incorporated because they were not part of the published coefficient set used for this equation and/or were unavailable in MASHAD. Fourth, predictors were measured only at baseline, and changes in smoking, blood pressure, lipid levels, diabetes status, kidney function, or medication use during follow-up may have influenced observed risk. Fifth, although outcomes were adjudicated using standardized procedures and ICD-based definitions, some misclassification remains possible. Sixth, participants with prevalent CVD, missing data, unavailable follow-up, or predictor values outside model-valid ranges were excluded to avoid invalid extrapolation, which may reduce applicability to individuals at extreme risk. Finally, recalibration improved sensitivity and net benefit but reduced specificity, meaning that implementation may increase false-positive classification. This observational validation establishes predictive performance and estimated clinical utility, but not the effectiveness of treatment decisions based on recalibrated risk estimates.

Conclusion

In this large prospective validation study from the MASHAD cohort, PREVENT and SCORE2 showed acceptable discrimination for 10-year ASCVD risk prediction but substantially underestimated absolute risk before recalibration. Sex-specific local recalibration corrected calibration-in-the-large and improved prediction error, clinical sensitivity, risk classification, and decision-analytic net benefit without changing discrimination. PREVENT showed modestly better discrimination and calibration slopes than SCORE2, likely reflecting its broader cardiometabolic and kidney-related predictor set. These findings support the use of locally calibrated risk prediction tools for ASCVD prevention rather than direct adoption of imported risk estimates. Further validation in independent regional cohorts is needed before broader clinical implementation.

Declarations

Acknowledgements

The authors used AI-assisted tools solely to assist with language refinement, editing, code troubleshooting, and improving clarity of scientific writing. The AI system was not used to generate original scientific content, analyze data independently, interpret results, or determine study conclusions. All outputs were critically reviewed, verified, and revised by the authors, who assume full responsibility for the final manuscript.

Author Contributions

Ali Tajik: Conceptualization, Methodology, Formal analysis, Software, Validation, Visualization, Investigation, Data curation, Writing – Original Draft, Writing – Review & Editing.

Hamidreza Rahimi: Investigation, Data curation.

Sara Saffar Soflaei: Investigation, Data curation.

Mohsen Moohebati: Investigation, Data curation.

Ashkan Bahrami: Writing – Original Draft, Investigation.

Gordon A. Ferns: Investigation, Writing – Review & Editing.

Habibollah Esmaily: Project administration, Supervision, Resources, Investigation.

Majid Ghayour-Mobarhan: Conceptualization, Resources, Supervision, Project administration, Funding acquisition.

Hedieh Alimi: Supervision, Project administration, Writing – Review & Editing.

Ethical considerations

Prior to participation, all enrolled individuals provided written informed consent. In cases where participants were illiterate, the consent process was conducted in the presence of a literate first-degree family member, who signed the form on the participant's behalf. The study was approved by the Human Research Ethics Committee of Mashhad University of Medical Sciences (approval code: IR.MUMS.REC.1386.250) and adhered to the ethical principles outlined in the Declaration of Helsinki.

Clinical trial number

Not applicable.

Funding

This study is funded by Mashhad University of Medical Sciences (Grant No. 4050185).

Conflict of interest

None declared.

Data availability

The data used in this study will not be shared because of Mashhad University of Medical Sciences copyright policies. However, requests for de-identified data could be considered on a case-by-case basis upon receipt of a reasonable request.

References

1. Mensah GA, Fuster V, Murray CJ, Roth GA, Abate YH, Abbasian M, et al. Global burden of cardiovascular diseases and risks, 1990-2022. Journal of the American College of Cardiology. 2023;82(25):2350-473.
2. Heidari-Foroozan M, Farshbafnadi M, Golestani A, Younesian S, Jafary H, Rashidi M-M, et al. National and subnational burden of cardiovascular diseases in Iran from 1990 to 2021: results from global burden of diseases 2021 study. Global heart. 2025;20(1):43.
3. Visseren FL, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice: Developed by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies With the special contribution of the European Association of Preventive Cardiology (EAPC). European journal of preventive cardiology. 2022;29(1):5-115.
4. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Journal of the American College of cardiology. 2019;74(10):e177-e232.
5. Khan SS, Matsushita K, Sang Y, Ballew SH, Grams ME, Surapaneni A, et al. Development and validation of the American Heart Association's PREVENT equations. Circulation. 2024;149(6):430-49.
6. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J. 2021;42(25):2439-54.
7. D'Agostino Sr RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008;117(6):743-53.
8. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. bmj. 2016;353.

9. Austin PC, Harrell Jr FE, van Klaveren D. Graphical calibration curves and the integrated calibration index (ICI) for survival models. Statistics in Medicine. 2020;39(21):2714-42.
10. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Medical Decision Making. 2006;26(6):565-74.
11. Lee M-S, Onwuzurike J, Wu Y-L, Palmer-Toy DE, An J, Chen W. PREVENT and PCE Models for Estimating ASCVD Risk Stratified by Statin Exposure. JAMA Network Open. 2025;8(9):e2532164-e.
12. Temtem M, Mendonça MI, Santos M, Sá D, Sousa F, Freitas S, et al. Validation of the SCORE2 risk prediction algorithm in a Portuguese population: A new model to estimate 10-year cardiovascular disease incidence in Europe. Revista Portuguesa de Cardiologia. 2024;43(8):437-44.
13. Hasanpour A, Asgari S, Khalili D, Azizi F, Hadaegh F. External validation of the PREVENT risk score: performance and clinical utility in an Iranian population. Scientific Reports. 2026.
14. Lopez-Lopez JP, Garcia-Pena AA, Martinez-Bello D, Gonzalez AM, Perez-Mayorga M, Muñoz Velandia OM, et al. External validation and comparison of six cardiovascular risk prediction models in the Prospective Urban Rural Epidemiology (PURE)-Colombia study. European Journal of Preventive Cardiology. 2025;32(7):564-72.
15. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Journal of British Surgery. 2015;102(3):148-58.
16. Ghayour-Mobarhan M, Moohebati M, Esmaily H, Ebrahimi M, Parizadeh SMR, Heidari-Bakavoli AR, et al. Mashhad stroke and heart atherosclerotic disorder (MASHAD) study: design, baseline characteristics and 10-year cardiovascular risk estimation. International journal of public health. 2015;60(5):561-72.
17. Graham IM. Cardiovascular risk estimation: can a risk prediction model derived in one country be used in another? Eur J Prev Cardiol. 2025;32(7):573-4.
18. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. Epidemiology. 2014;25(1):114-21.
19. Scheuermann B, Brown A, Colburn T, Hakeem H, Chow CH, Ade C. External validation of the American Heart Association PREVENT cardiovascular disease risk equations. JAMA Network Open. 2024;7(10):e2438311.
20. Zhou H, Zhang Y, Zhou MM, Choi SK, Reynolds K, Harrison TN, et al. Evaluation and comparison of the PREVENT and Pooled Cohort Equations for 10-year atherosclerotic cardiovascular risk prediction. Journal of the American Heart Association. 2025;14(4):e039454.
21. Shahrestanaki E, Solaymani-Dodaran M, Mohammadzadeh M, Masoudkabir F, Azizi F, Hadaegh F, et al. Validation and clinical usefulness of the SCORE2 cardiovascular risk prediction algorithms in an Iranian population. Journal of Diabetes & Metabolic Disorders. 2025;24(2):280.
22. Kist JM, Vos RC, Mairuhu AT, Struijs JN, van Peet PG, Vos HM, et al. SCORE2 cardiovascular risk prediction models in an ethnic and socioeconomic diverse population in the Netherlands: an external validation study. EClinicalMedicine. 2023;57.

23. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC medicine. 2019;17(1):230.
24. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128-38.
25. DeFilippis AP, Young R, Carrubba CJ, McEvoy JW, Budoff MJ, Blumenthal RS, et al. An analysis of calibration and discrimination among multiple cardiovascular risk scores in a modern multiethnic cohort. Annals of internal medicine. 2015;162(4):266-75.
26. Yadlowsky S, Hayward RA, Sussman JB, McClelland RL, Min Y-I, Basu S. Clinical implications of revised pooled cohort equations for estimating atherosclerotic cardiovascular disease risk. Annals of internal medicine. 2018;169(1):20-9.
27. Cook NR, Ridker PM. Calibration of the pooled cohort equations for atherosclerotic cardiovascular disease: an update. Annals of internal medicine. 2016;165(11):786-94.
28. Rana JS, Tabada GH, Solomon MD, Lo JC, Jaffe MG, Sung SH, et al. Accuracy of the atherosclerotic cardiovascular risk equation in a large contemporary, multiethnic population. Journal of the American College of Cardiology. 2016;67(18):2118-30.
29. Pencina MJ, D'Agostino Sr RB, D'Agostino Jr RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in medicine. 2008;27(2):157-72.

Tables

Table 1. Baseline characteristics of the study population at baseline.

Variable	Total, N=9069 (100)	Gender	
		Men, N=3584 (39.5)	Women, N=5485 (60.5)
Age (Years), median [IQR]	47.0 (41.0-54.0)	48.0 (42.0-55.0)	47.0 (41.0-53.0)
Educational level, n (%)			
Low	4915 (54.2%)	1471 (41.0%)	3444 (62.8%)
Intermediate	3149 (34.7%)	1480 (41.3%)	1669 (30.4%)
High	1005 (11.1%)	633 (17.7%)	372 (6.8%)
Job status, n (%)			
Employed	3346 (36.9%)	2671 (74.5%)	675 (12.3%)
Unemployed	4862 (53.6%)	298 (8.3%)	4564 (83.2%)
Retired	861 (9.5%)	615 (17.2%)	246 (4.5%)
Smoker, n (%)	1960 (21.6%)	986 (27.5%)	974 (17.8%)
Height (m), median [IQR]	1.6 (1.54-1.67)	1.69 (1.64-1.73)	1.55 (1.51-1.59)
Weight (kg), median [IQR]	70.9 (62.9-79.6)	74.5 (66.1-82.6)	68.8 (61.2-77.0)
BMI (kg/m^2), median [IQR]	27.61 (24.66-30.76)	26.27 (23.47-28.85)	28.64 (25.55-31.73)
Waist circumference (cm), median [IQR]	95.0 (87.1-103.0)	94.0 (87.0-100.5)	96.0 (88.0-104.0)
Hip circumference (cm), median [IQR]	103.0 (98.0-109.0)	101.0 (96.0-106.0)	105.0 (99.0-111.0)
Systolic (mmHg), median [IQR]	120.0 (110.0-130.0)	120.0 (110.0-130.0)	120.0 (109.33-130.0)
Diastolic (mmHg), median [IQR]	80.0 (70.0-86.0)	80.0 (71.67-86.67)	80.0 (70.0-84.0)
Total cholesterol (mg/dl), median [IQR]	188.0 (165.0-214.0)	184.0 (161.0-209.0)	191.0 (167.0-218.0)
Total cholesterol (mmol/L), median [IQR]	4.86 (4.26-5.53)	4.76 (4.16-5.40)	4.94 (4.31-5.63)
Triglyceride (mg/dl), median [IQR]	119.0 (84.0-171.0)	124.0 (86.0-179.0)	116.0 (83.0-166.0)
Fasting plasma glucose (mg/dl), median [IQR]	82.0 (74.0-93.0)	82.0 (74.0-92.0)	82.0 (74.0-93.0)

Low Density Lipoprotein (mg/dl), median [IQR]	114.8 (94.1-137.6)	112.4 (91.8-135.4)	116.4 (95.6-139.5)
High Density Lipoprotein (mg/dl), median [IQR]	41.8 (36.0-48.4)	38.4 (33.5-44.7)	44.0 (38.0-50.4)
High Density Lipoprotein (mmol/L), median [IQR]	1.08 (0.93-1.25)	0.99 (0.86-1.15)	1.14 (0.98-1.30)
Uric acid (mg/dl), median [IQR]	4.5 (3.7-5.5)	5.2 (4.3-6.1)	4.1 (3.4-4.9)
Glomerular Filtraion Rate (mL/min), median [IQR]	83.62 (70.02-96.71)	87.53 (75.22-102.32)	79.85 (67.70-93.64)
White Blood Cell (*10^9/L), median [IQR]	5.9 (5.0-6.9)	5.9 (5.1-7.0)	5.8 (5.0-6.8)
Hemoglobin (g/dl), median [IQR]	13.7 (12.8-14.7)	14.9 (14.1-15.6)	13.1 (12.4-13.8)
Hematocrit (%), median [IQR]	41.1 (38.7-43.8)	44 (42.1-45.8)	39.5 (37.6-41.4)
Platelet (*109/L), median [IQR]	225.0 (190.0-265.0)	210.5 (179.0-245.0)	237.0 (200.0-277.0)
Highly sensitive C-Reactive Protein (mg/dl), median [IQR]	1.6 (0.99-3.42)	1.4 (0.89-2.79)	1.78 (1.05-3.8)
Diabetes mellitus, n (%)	1233 (13.6%)	453 (12.6%)	780 (14.2%)
Hypertension, n (%)	2762 (30.5%)	1017 (28.4%)	1745 (31.8%)
Statin use, n (%)	118 (1.3%)	40 (1.1%)	78 (1.4%)
Anti-hypertensive use, n (%)	313 (3.5%)	77 (2.1%)	236 (4.3%)
Positive family history of CVD, n (%)	3141 (34.6%)	1131 (31.6%)	2010 (36.6%)
Anxiety score, median [IQR]	8.0 (3.0-15.0)	6.0 (2.0-12.0)	9.0 (4.0-18.0)
Depression score, median [IQR]	10.0 (5.0-18.0)	8.0 (4.0-15.0)	12.0 (6.0-19.0)

Abbreviations: IQR; interquartile range, CVD; cardiovascular disease.

Note: The anxiety and Depression were assessed using Beck questionnaire.

Table 2. Discrimination, calibration slope, and prediction error of PREVENT and SCORE2 by sex.

Model	Sex	Version	n	Events	C-statistic	Brier score (%)	Calibration slope
PREVENT	Men	Original	3584	333	0.714 (0.661–0.768)	9.47 (8.49–10.45)	0.861 (0.735–0.988)
		Recalibrated	3584	333	0.714 (0.661–0.768)	8.28 (7.42–9.15)	0.861 (0.735–0.988)
	Women	Original	5485	345	0.731 (0.679–0.783)	6.76 (6.06–7.46)	0.813 (0.710–0.917)
		Recalibrated	5485	345	0.731 (0.679–0.783)	6.25 (5.59–6.90)	0.813 (0.710–0.917)
SCORE2	Men	Original	3131	240	0.687 (0.622–0.752)	8.00 (7.02–8.99)	0.754 (0.603–0.905)
		Recalibrated	3131	240	0.687 (0.622–0.752)	7.17 (6.28–8.06)	0.754 (0.603–0.905)
	Women	Original	4705	230	0.704 (0.639–0.769)	5.35 (4.67–6.04)	0.700 (0.575–0.825)
		Recalibrated	4705	230	0.704 (0.639–0.769)	4.95 (4.32–5.59)	0.700 (0.575–0.825)

[a]SCORE2 performance was calculated excluding participants with positive history of diabetes mellitus at baseline.

Table 3. Categorical net reclassification improvement (NRI) after local recalibration.

Model	Population	n	Events	Event NRI	Non-event NRI	Total NRI
PREVENT	Overall	9069	678	0.813 (0.782–0.843)	-0.517 (-0.527–-0.507)	0.296 (0.263–0.329)
	Men	3584	333	0.901 (0.868–0.932)	-0.644 (-0.661–-0.627)	0.257 (0.220–0.292)
	Women	5485	345	0.728 (0.683–0.773)	-0.436 (-0.451–-0.423)	0.291 (0.245–0.340)
SCORE2	Overall	7836	470	0.851 (0.819–0.883)	-0.615 (-0.626–-0.604)	0.236 (0.201–0.269)
	Men	3131	240	0.892 (0.851–0.929)	-0.793 (-0.809–-0.778)	0.098 (0.054–0.138)
	Women	4705	230	0.809 (0.758–0.857)	-0.500 (-0.514–-0.485)	0.309 (0.255–0.360)
	<50 years	4806	181	0.724 (0.660–0.792)	-0.475 (-0.490–-0.460)	0.248 (0.185–0.316)
	≥50 years	3030	289	0.931 (0.900–0.959)	-0.851 (-0.864–-0.838)	0.080 (0.048–0.110)
	Men, <50 years	1806	85	0.824 (0.741–0.904)	-0.703 (-0.726–-0.682)	0.120 (0.032–0.202)
	Men, ≥50 years	1325	155	0.929 (0.887–0.969)	-0.926 (-0.942–-0.911)	0.003 (-0.041–0.045)
	Women, <50 years	3000	96	0.635 (0.534–0.730)	-0.341 (-0.358–-0.324)	0.295 (0.191–0.392)
	Women, ≥50 years	1705	134	0.933 (0.886–0.972)	-0.795 (-0.815–-0.775)	0.138 (0.088–0.181)

Note: Categorical net reclassification improvement (NRI) was calculated by comparing original and recalibrated risk-category assignments. For PREVENT, risk categories were defined using 5%, 7.5%, and 20% 10-year ASCVD risk thresholds. For SCORE2, age-specific thresholds were applied: 2.5% and 7.5% for participants aged <50 years, and 5% and 10% for participants aged ≥50 years. Positive event NRI indicates upward movement among participants who experienced events, whereas negative non-event

NRI indicates upward movement among participants without events. Total NRI is the sum of the event and non-event components.

Figures

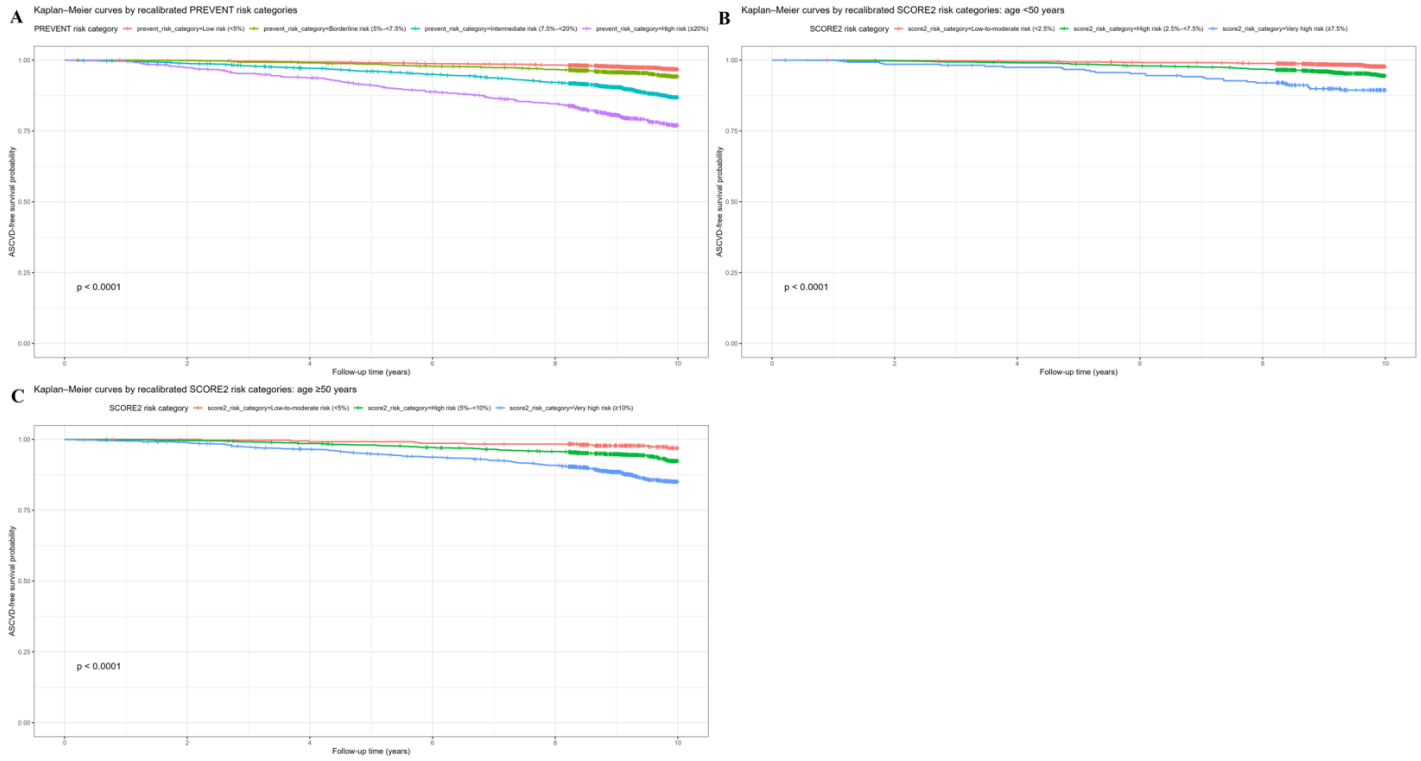


Figure 1

Kaplan–Meier survival curves according to recalibrated cardiovascular risk categories.
Kaplan–Meier curves demonstrating event-free survival across clinically defined risk categories derived from the recalibrated risk models. Panel (A) shows the recalibrated PREVENT model using predefined thresholds of <5%, 5%–<7.5%, 7.5%–<20%, and ≥20%. Panel (B) shows the recalibrated SCORE2 model in participants aged under 50 years using thresholds of <2.5%, 2.5%–<7.5%, and ≥7.5%. Panel (C) shows the recalibrated SCORE2 model in participants aged 50–69 years using thresholds of <5%, 5%–<10%, and ≥10%. Differences between risk strata were assessed using the log-rank test.

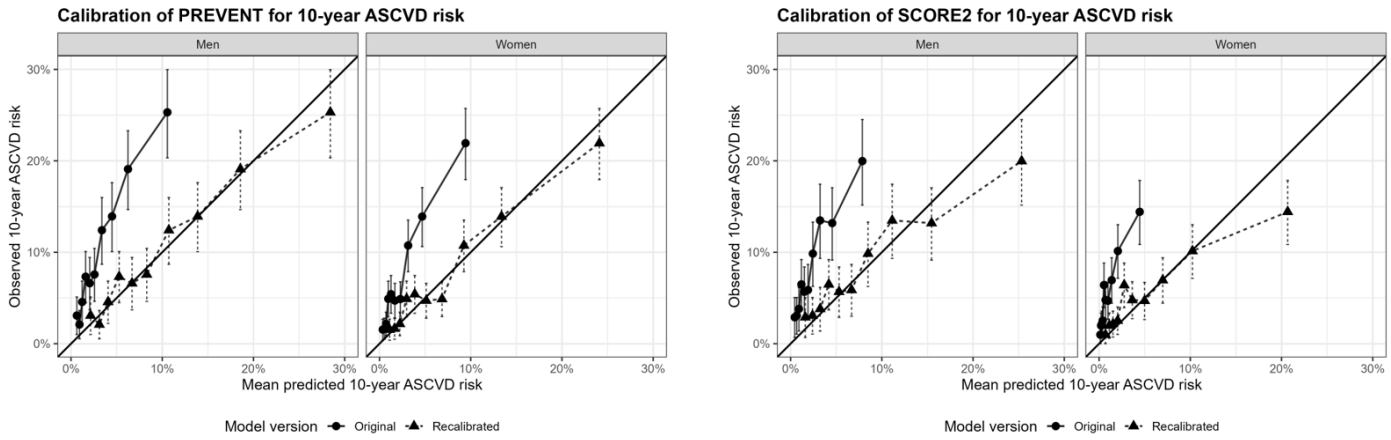


Figure 2

Calibration and recalibration plots of predicted and observed 10-year ASCVD risk. The plots show the relationship between mean predicted and Kaplan–Meier estimated observed 10-year ASCVD risk across groups of predicted risk, separately by sex and model version. Recalibration was performed using sex-specific local updating while preserving the original model coefficients.

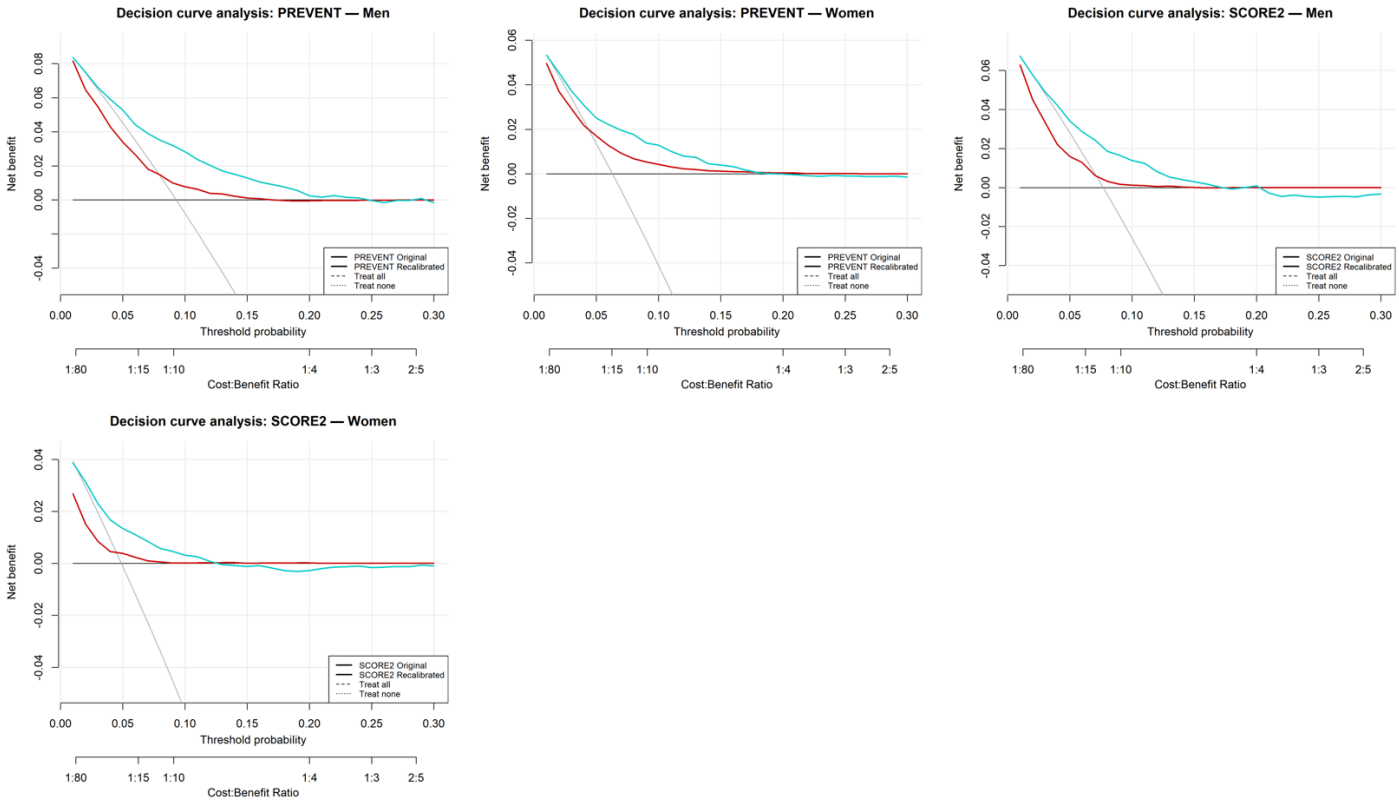


Figure 3

Decision curve analysis of original and recalibrated PREVENT and SCORE2 models by sex. Net benefit was plotted across threshold probabilities from 1% to 30%. Original and recalibrated predictions were

compared with treat-all and treat-none strategies. Recalibrated models generally showed higher net benefit at lower and intermediate threshold probabilities.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfile.docx
- Graphicalabstract.docx