Validation and comparison of 28 risk prediction models for coronary artery disease

Lenselink, Chris; Ties, Daan; Pleijhuis, Rick; van der Harst, Pim

Published in:
European journal of preventive cardiology

DOI:
10.1093/eurjpc/zwab095

Publication date:
2022

ESC European Society of Cardiology

European Journal of Preventive Cardiology (2022) **29**, 666–674
doi:10.1093/eurjpc/zwab095

FULL RESEARCH PAPER
Risk Prediction/Assessment & Stratification

Validation and comparison of 28 risk prediction models for coronary artery disease

Chris Lenselink [1], Daan Ties[1], Rick Pleijhuis[2] and Pim van der Harst [1]*

[1]Department of Cardiology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9700 RB Groningen, The Netherlands; and [2]Department of Internal Medicine, University Medical Center Groningen, Hanzeplein 1, 9700 RB, Groningen, The Netherlands

Received 15 March 2021; revised 29 April 2021; editorial decision 14 May 2021; accepted 18 May 2021; online publish-ahead-of-print 30 July 2021

Aims Risk prediction models (RPMs) for coronary artery disease (CAD), using variables to calculate CAD risk, are potentially valuable tools in prevention strategies. However, their use in the clinical practice is limited by a lack of poor model description, external validation, and head-to-head comparisons.

Methods and results CAD RPMs were identified through Tufts PACE CPM Registry and a systematic PubMed search. Every RPM was externally validated in the three cohorts (the UK Biobank, LifeLines, and PREVEND studies) for the primary endpoint myocardial infarction (MI) and secondary endpoint CAD, consisting of MI, percutaneous coronary intervention, and coronary artery bypass grafting. Model discrimination (C-index), calibration (intercept and regression slope), and accuracy (Brier score) were assessed and compared head-to-head between RPMs. Linear regression analysis was performed to evaluate predictive factors to estimate calibration ability of an RPM. Eleven articles containing 28 CAD RPMs were included. No single best-performing RPM could be identified across all cohorts and outcomes. Most RPMs yielded fair discrimination ability: mean C-index of RPMs was 0.706 ± 0.049, 0.778 ± 0.097, and 0.729 ± 0.074 ($P < 0.01$) for prediction of MI in UK Biobank, LifeLines, and PREVEND, respectively. Endpoint incidence in the original development cohorts was identified as a significant predictor for external validation performance.

Conclusion Performance of CAD RPMs was comparable upon validation in three large cohorts, based on which no specific RPM can be recommended for predicting CAD risk.

Keywords Myocardial ischaemia • Coronary artery disease • Preventive medicine • Risk assessment • Risk prediction models

Introduction

Coronary artery disease (CAD) is the leading cause of mortality globally, causing nearly 16% of all deaths worldwide.[1] The lifetime risk for developing CAD for 40-year-olds is 49% in men and 32% in women, as demonstrated by the Framingham Heart Study.[2] International guidelines on primary prevention of CAD recommend treating cardiovascular risk factors to reduce the risk of future CAD.[3] Adequate risk stratification is essential to guide initiation of primary and secondary preventive measures.[4] Risk prediction models (RPMs), estimating lifetime CAD risk from a set of predictive clinical variables, are useful tools to guide cardiovascular risk management by identifying individuals expected to benefit most from cardiovascular risk management.[4–6]

Over the past decades, the number of published CAD RPMs has increased exponentially. However, the majority of these RPMs lack external validation by both the developing and independent researchers. In addition, development and assessment of single RPMs in a single cohort makes it difficult to directly compare performance of different RPMs. The abundance of RPMs, the lack of external validation, and missing head-to-head comparison between RPMs obscure the view on which RPMs are most useful. Rather than keeping

* Corresponding author. Tel: +31503612355, Email: p.van.der.harst@umcg.nl

on developing new models, existing RPMs should be directly compared through external validation, in order to identify the best performing models suitable for clinical application.[5,7]

The aim of the present study is to identify and externally validate existing RPMs predicting CAD in the general population and to compare their performance in three large independent cohorts.

Methods

Study design

In this retrospective study, we externally validated and directly compared CAD RPMs identified through Tufts PACE CPM registry and a systematic literature search. For validation purposes, three independent cohorts from the general population were used: the UK Biobank, LifeLines, and The Prevention of Renal and Vascular Endstage Disease (PREVEND) cohort studies. Study design and procedures have been extensively described before.[8–10] The UK Biobank database contains demographic and medical history data of 502 536 participants aged 40–69 years from the UK, who were recruited from 2006 till 2010. Long-term follow-up was performed with the aim to investigate the association between environmental exposure or genetic predisposition and the development of disease, including cardiovascular diseases.[8] In the LifeLines biobank study, 167 729 inhabitants from the northern Netherlands were invited for visits and questionnaires between 2006 and 2013, with participants aged 6 months to 93 years. Follow-up visits are scheduled every 5 years. LifeLines aims to study environmental, phenotypic, and genotypic influences on the development of chronic diseases, including cardiovascular diseases, and healthy ageing.[10] The PREVEND community-based cohort study assessed 8507 inhabitants from the Dutch city of Groningen to study the association between presence of albuminuria and estimated glomerular filtration rate with the development of cardiovascular outcomes.[9] Follow-up in PREVEND was performed up to 10.5 years. The data underlying this article were provided by Lifelines, PREVEND and UK Biobank. Data will be shared on request to the corresponding author with permission of Lifelines, PREVEND and UK Biobank.

Data sources

First, Tufts PACE CPM registry was searched for RPMs using the general population as primary index condition and myocardial infarction (MI) or CAD as outcome.[11] Additionally, PubMed was systematically searched for studies describing RPMs not registered in the Tufts registry. Search terms are provided in the Supplementary material online.

RPM eligibility

Studies describing the development of a multivariable RPM for CAD in a prospective cohort study in the general population (>18 years) without history of major cardiac disease were eligible for this study. Studies were excluded if (i) the study was not published in the English language, (ii) existing RPMs were compared or reviewed, (iii) the outcome was not adequately specified, (iv) the outcome studied in the original paper included non-cardiac endpoints (e.g., stroke, peripheral vascular disease), (v) model variables were not available in at least two of the validation cohorts, or (vi) insufficient data were provided in the article to rebuild the model and model authors did not respond after two inquiries for lacking data. If articles described the development of multiple models, each individual model was evaluated for eligibility. Title-abstract screening and full-text screening were performed to select eligible studies for final analysis. For each eligible study, information regarding study design, model variables, sample size, in- and exclusion criteria, definition of endpoints, incidence of endpoints, follow-up time, exposure time, C-statistics, and performance of an external validation in the original study was extracted.

Endpoint definition

For external validation, we adhered to the definition of the variables as described in the original RPM studies as much as possible. Two endpoints were used: MI and CAD, with CAD defined as MI, percutaneous coronary intervention (PCI), or coronary artery bypass grafting (CABG). In the UK Biobank cohort, MI was defined as ICD10 codes I21, I22, I23, and I252, PCI as OPER codes K49, K50, and K75, and CABG as OPER codes K40 through K46. In the LifeLines and PREVEND cohorts self-reported outcomes were used.

Data analysis

RPM formulas were extracted from included studies and entered into the web-based platform Evidencio (Evidencio BV, version 2.19, Haaksbergen, The Netherlands) for validation purposes.[12] In brief, Evidencio offers free services to automatically convert RPM formulas into standardized user-friendly digital calculators, promoting clinical usage and integration in third party applications. In addition, it enables researchers to efficiently perform R-based external validations of RPMs.[13] After allocation of anonymized data to corresponding RPM variables, the platform generates multiple statistical and graphical outcomes, including model discrimination, calibration, and composite performance measures.

Discriminatory performance was described by C-indices with 95% confidence interval (CI). Calibration of RPMs was evaluated by regression analysis of estimated risk by RPMs and observed risk in validation cohorts. Intercept and regression slopes are presented. Regression slope above 1 indicates underestimation of risk, regression slope below 1 indicates overestimation. Overall model performance was described by the Brier score. Mean values of C-indices, Brier scores, regression slopes, and regression intercepts were compared with a one-way analysis of variance. Models using unspecified left ventricular hypertrophy criteria were tested with both the Cornell and Sokolow-Lyon criteria. Linear regression analyses were performed to investigate potential predictors (sample size, year of publication, and endpoint incidence in the original development cohort) of calibration as reflected by the regression slope. An alpha <0.05 was considered to indicate statistical significance. All analyses were run using Evidencio, STATA (StataCorp, version 16.0, College Station, USA) and GraphPad Prism (GraphPad Software Inc., version 8.4.2, San Diego, USA).

Results

Identification of RPMs

Eleven potentially eligible studies were identified in Tufts PACE CPM registry, whereas the PubMed search yielded 552 potentially eligible studies. After excluding duplicates and screening studies for eligibility, 11 studies describing 28 RPMs were included in the final analysis. Supplementary material online, *Figure S1* describes the inclusion of articles.

Table 1 Characteristics of original RPM studies

Author	Number of models (*n*)	Sample size (*n*)	Participants and age (years)	Definition of outcome, incidence of outcome	Follow-up time (years)	Exposure time (years)	C-statistic	External validation (yes/no), sample size of cohort (*n*)
Tunstall-Pedoe, 1991[14]	2	5203	Men, 40–59	CAD (not specified), 0.06	5	5	NA	Yes, 15 395
Anderson, 1991[15]	6	5573	Free of cancer, age 48.6 ± 11.7	CAD (MI and CAD death plus AP and coronary insufficiency); MI, CAD death, 0.15	12	10	NA	No
Bogle, 2018[16]	1	11 335	Race White or Black, age 55.4 ± 5.7	Sudden cardiac death (fatal MI or CAD), 0.01	10	10	0.82 (White), 0.75 (Black)	Yes, 5625
Onat, 2012[17]	1	2232	Turkish adults with metabolic syndrome, age 47.5 ± 11.3	MI, AP, 0.14	7.6 ± 2.5	10	M: 0.789 (0.754; 0.825) F: 0.806 (0.771, 0.840)	No
Chien, 2012[18]	3	3430	Inhabitants of Chinese community in Taiwan, age 54.8 ± 12.2	MI, PCI, CABG, CAD death, 0.05	15.9 (median), IQR 12.7–16.9	10	Clinical: 0.782 TC: 0.771 LDL: 0.731	Yes, 22 193
Wilson, 1998[19]	2	5345	Age 49.2 ± 11.8	AP, MI, coronary insufficiency, CAD death, 0.11	12	10	TC, M: 0.69 TC, F: 0.72 LDL, M: 0.68 LDL, F: 0.71	No
Nishimura, 2014[20]	2	5521	Inhabitants of Suita, Japan, age 55.3 ± 13.1	MI, sudden cardiac death, CAD followed by CABG or PCI, 0.04	11.8	10	TC: 0.835 LDL: 0.831	No
Jee, 2014[21]	1	268 315	Age 46.5 ± 9.7	MI, sudden cardiac death or other coronary deaths, <0.01	11.6	10	M: 0.764 F: 0.815	No
Koller, 2012[22]	2	9249	Age > 65	Nonfatal MI and coronary death, 0.20	16.5/14.9	10	M: 0.63 US F: 0.67 EU F: 0.68	Cross-validation
L'Italien, 2000[23]	1	6595	Age 45–64	Fatal CAD and nonfatal MI	NA	5	NR	No
Conroy, 2003[24]	1	205 178	12 European cohort studies	Coronary death and nonfatal MI, 0.03	2.7 million person years	10	0.71–0.84	No

AP, angina pectoris; CABG, coronary artery bypass grafting; CAD, coronary artery disease; F, female; LDL, low-density cholesterol; M, male; MI, myocardial infarction; NA, not available; TC, total cholesterol.

Table 2 Baseline and follow-up characteristics of the validation cohorts

Variable	UK Biobank (n = 481 628)[8]	LifeLines (n = 154 289)[9]	PREVEND (n = 8035)[10]
Age (years)	56.8 ± 8.1	44.9 ± 13.4	48.2 ± 12.4
Gender (m)	213 296 (44.3)	63 101 (40.9)	3895 (48.5)
Body mass index (kg/m^2)	27.4 ± 4.8	26.0 ± 4.3	26.0 ± 4.2
Smoking			
Never	266 120 (55.5)	75 412 (48.9)	2423 (30.2)
Former	162 436 (33.7)	47 748 (30.9)	2845 (35.4)
Current	50 335 (10.5)	31 129 (20.2)	2746 (34.2)
HDL (mmol/L)	1.46 ± 0.38	1.49 ± 0.40	1.33 ± 0.40
LDL (mmol/L)	3.59 ± 0.86	3.24 ± 0.92	3.49 ± 0.97
Total cholesterol (mmol/L)	5.47 ± 1.12	5.09 ± 1.01	5.65 ± 1.13
Cholesterol-lowering medication	72 578 (15.1)	6737 (4.4)	208 (2.6)
Diabetes mellitus	23 303 (4.8)	3475 (2.3)	113 (1.4)
Hypertension	132 886 (27.6)	30 738 (19.9)	2340 (29.1)
Systolic blood pressure (mmHg)	133 ± 18	125 ± 15	128 ± 20
Diastolic blood pressure (mmHg)	82 ± 9	74 ± 9	74 ± 10
ECG-LVH			
Cornell	892 (4.7)	2111 (1.4)	357 (4.4)
Sokolow-Lion	735 (3.9)	6403 (4.2)	610 (7.6)
Antihypertensive medication	96 109 (20.0)	13 110 (8.5)	921 (11.5)
Outcomes			
MI	5480 (1.1)	480 (0.3)	233 (2.9)
CAD	15 140 (3.1)	641 (0.4)	442 (5.5)

Continuous data are presented as mean ± standard deviation, categorical data as number (percentage).
CAD, coronary artery disease; ECG-LVH, electrocardiogram-confirmed left ventricular hypertrophy; HDL, high-density cholesterol; LDL, low-density cholesterol; m, male; MI, myocardial infarction.

Characteristics of original RPM studies

Basic characteristics of the included models are reported in *Table 1*. The sample size of the original cohorts in which RPMs were developed varied from 2232 to 268 315 participants. The follow-up time varied substantially between studies (5–16 years). Eight out of 11 (73%) studies reported the C-index as an outcome measure, and only three (27%) RPMs were externally validated in the original paper. Age, smoking, and gender were the most used variables (*Table 2*; Supplementary material online, *Figure S2*).

Model discrimination

The mean C-index of RPMs was 0.706 ± 0.049, 0.778 ± 0.097, and 0.729 ± 0.074 (P = 0.0021) for prediction of MI in UK Biobank, LifeLines, and PREVEND, respectively. The mean C-index was 0.698 ± 0.053, 0.786 ± 0.097, and 0.754 ± 0.076 (P = 0.0002) for prediction of CAD in UK Biobank, LifeLines, and PREVEND, respectively.

In UK Biobank, the C-index of the best performing RPM to predict MI was significantly higher compared to RPMs ranked 6th, 8th–10th, 15th, 19th, 20th, 22th till 28th (P < 0.05) (*Figure 1A*). For CAD, the C-index of the best performing RPM was significantly higher than the RPMs ranked 11th till 15th, as well as the models ranked 17th till 28th (P < 0.05) (*Figure 1D*).

In LifeLines, the C-index of the best performing RPM to predict MI was significantly higher compared to the C-index of the 20 worst performing RPMs (P < 0.05) (*Figure 1B*). For CAD, the C-index of the best performing RPM was significantly higher than the five worst performing RPMs (P < 0.05) (*Figure 1E*).

In PREVEND, the C-index (point estimate) of the best performing RPM to predict MI was significantly higher compared to the C-index of the seven worst performing RPMs (P < 0.05) (*Figure 1C*). For CAD, the C-index (point estimate) of the best performing RPM was significantly higher than all other RPMs (P < 0.05) (*Figure 1F*).

Model calibration

For MI, the mean regression slope was 0.25 ± 0.20, 0.19 ± 0.18, and 0.43 ± 0.30 (P = 0.003) in UK Biobank, LifeLines, and PREVEND, respectively (*Figure 3*). The mean intercept was 0.001 ± 0.003, 0.002 ± 0.014, and 0.010 ± 0.011 (P = 0.0039) for MI in UK Biobank, LifeLines, and PREVEND, respectively. For CAD, the mean regression slope was 0.56 ± 0.49, 0.14 ± 0.14, and 0.98 ± 1.04 (P = 0.0001) in UK Biobank, LifeLines, and PREVEND, respectively (*Figure 3*). The mean intercept was 0.005 ± 0.007, −0.001 ± 0.005, and 0.075 ± 0.329 (P = 0.2563) for CAD in UK Biobank, LifeLines, and PREVEND, respectively.

Overall performance

The mean Brier score was 0.018 ± 0.021, 0.029 ± 0.075, and 0.032 ± 0.014 (P = 0.450) for MI in UK Biobank, LifeLines, and PREVEND, respectively (*Figure 2A–C*). The mean Brier score was

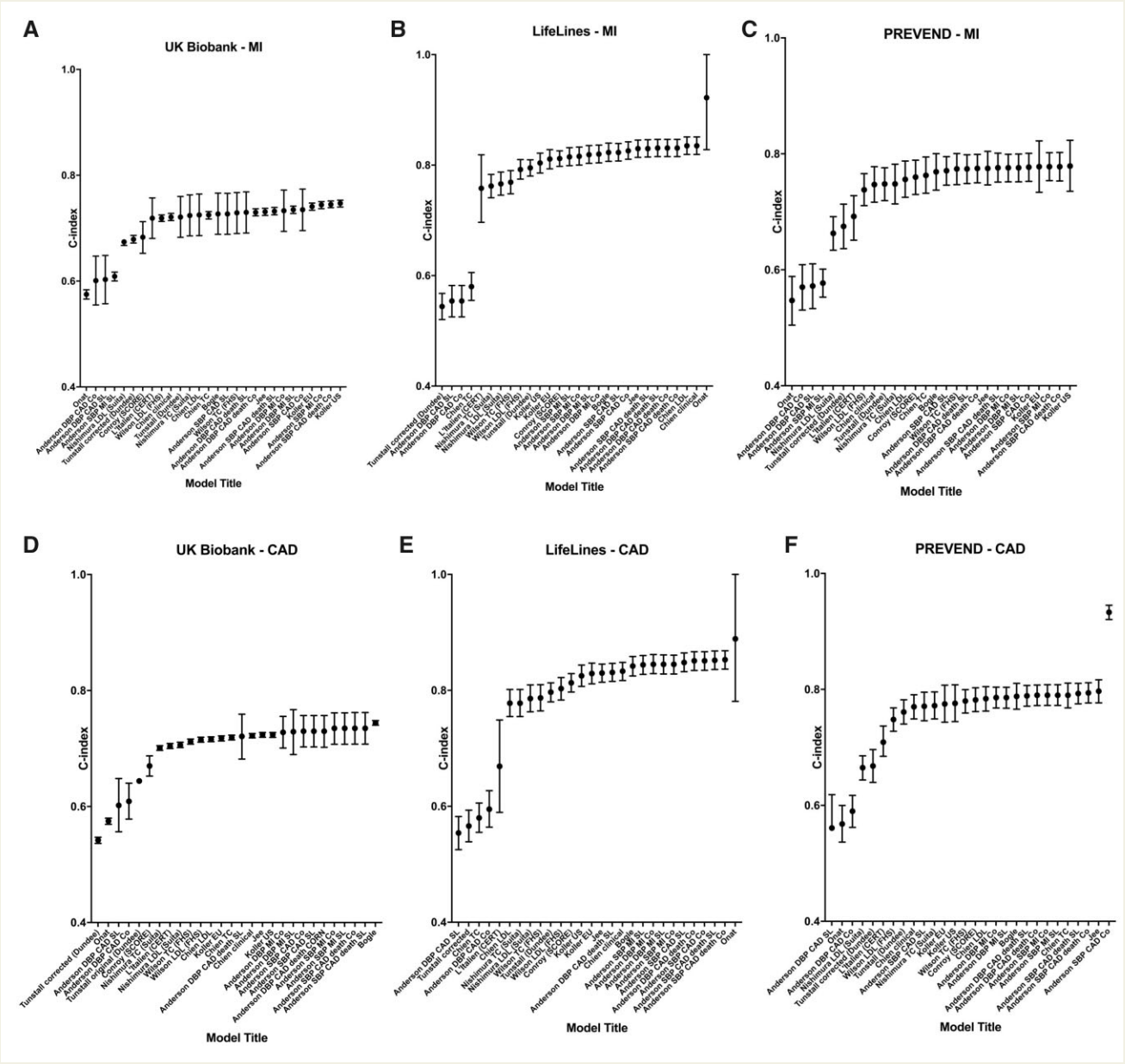

Figure 1 C-indices for every RPM in the three cohorts with outcomes MI and CAD. (*A*) results in UK Biobank for MI, (*B*) results in LifeLines for outcome MI, (*C*) results in PREVEND for outcome MI, (*D*) results in UK Biobank for CAD, (*E*) results in LifeLines for CAD, (*F*) results in PREVEND for CAD. CAD, coronary artery disease; CERT, Cardiovascular Event Reduction Tool; Co, Cornell; DBP, diastolic blood pressure; EU, European; FHS, Framingham heart Study; LDL, low-density cholesterol; MI, myocardial infarction; SBP, systolic blood pressure; SL, Sokolow-Lion; TC, total cholesterol; US, United States.

0.030 ± 0.021, 0.028 ± 0.075, and 0.045 ± 0.006 ($P = 0.302$) for CAD in UK Biobank, LifeLines, and PREVEND, respectively (*Figure 2D–F*).

Relationship between RPM study characteristics and model calibration

For MI, increased endpoint incidence in the development cohort was associated with decreased regression slopes in UK Biobank (β −2.11, 95% CI −3.23, −1.00, $P = 0.001$) and PREVEND (β −2.42, 95% CI −4.31, −0.54, $P = 0.014$), indicating increased overestimation of MI risk with increased endpoint incidence. For CAD, findings were similar as for MI (Supplementary material online, *Table S4*). Higher publication year of the original study and increased sample size of the discovery cohort was associated with decreased regression slopes for MI and CAD in UK Biobank and for CAD only in PREVEND (Supplementary material online, *Table S4*).

Discussion

The aim of the present study was to externally validate existing RPMs for prediction of CAD risk and compare their performance head-to-head, in order to further clarify potential differences in

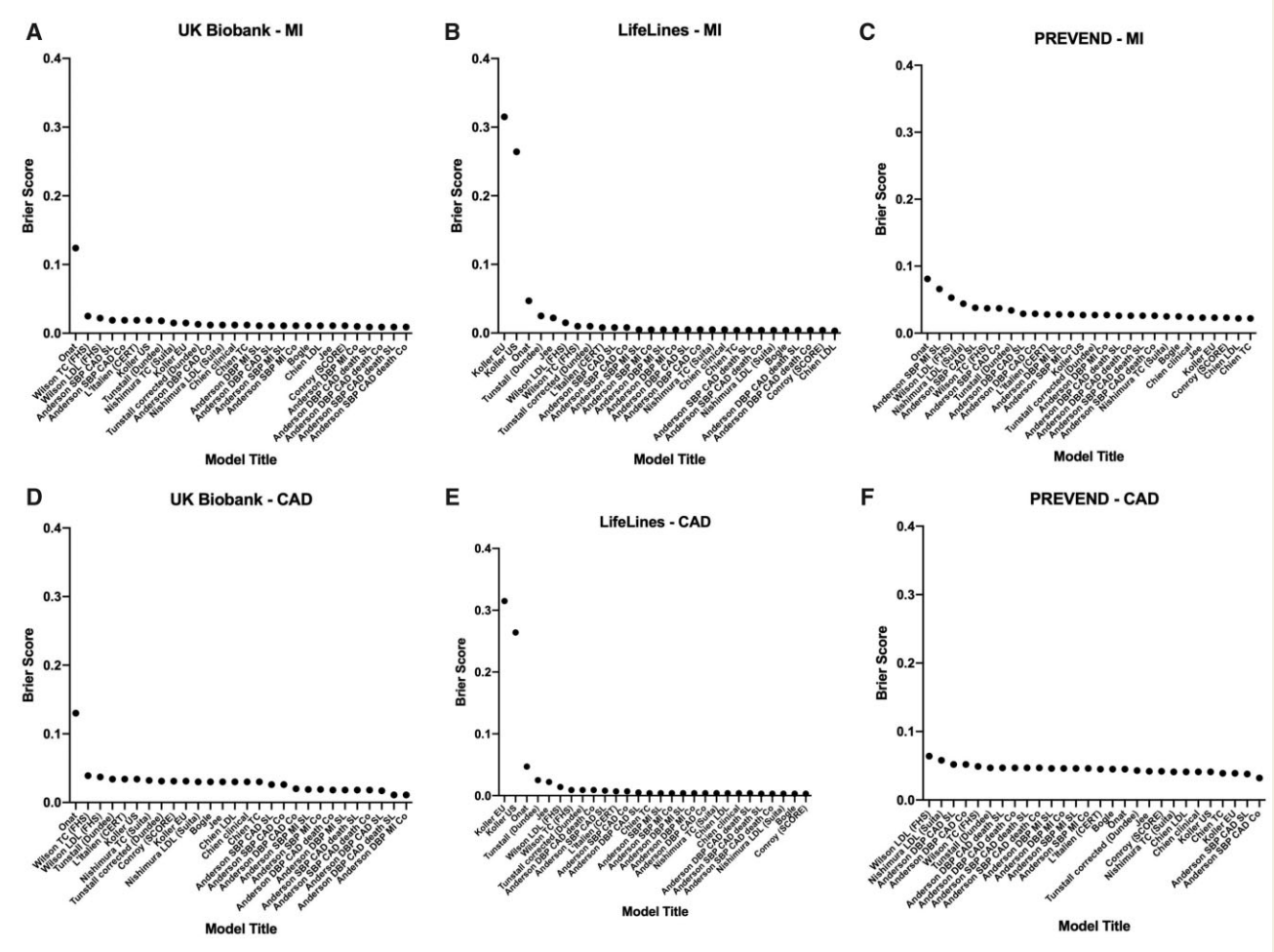

Figure 2 Brier scores for every RPM in the three cohorts with outcomes MI and CAD. (*A*) results in UK Biobank for MI, (*B*) results in LifeLines for outcome MI, (*C*) results in PREVEND for outcome MI, (*D*) results in UK Biobank for CAD, (*E*) results in LifeLines for CAD, (*F*) results in PREVEND for CAD. CAD, coronary artery disease; CERT, Cardiovascular Event Reduction Tool; Co, Cornell; DBP, diastolic blood pressure; EU, European; FHS, Framingham heart Study; LDL, low-density cholesterol; MI, myocardial infarction; SBP, systolic blood pressure; SL, Sokolow-Lion; TC, total cholesterol; US, United States.

predictive performance of RPMs in primary prevention of CAD. We performed external validation of RPMs in three large independent cohorts, effectively using a validation cohort of nearly 650 000 participants without baseline CAD. We identified 11 studies reporting on development of 28 RPMs predicting MI or incident CAD. The majority of RPMs showed fair discrimination and calibration for MI and CAD in all three cohorts. No single RPM performed consistently better across all cohorts compared to other RPMs.

An abundance of RPMs predicting CAD exist due to wide availability of clinical data from large-scale cohort studies, increased attention being paid to predictive algorithms, and the urge to publish scientific articles. Unfortunately, many studies reporting on development of RPMs lack adequate methodology, appropriate reporting and/or external validation by developing and independent authors.[7] In addition, to our knowledge, no studies comparing RPMs for identical outcomes within the same cohorts have been performed so far. The large number of reported RPMs, uncertainty regarding model robustness and generalizability in the absence of external models validations, and the lack of head-to-head comparisons between models impedes selection and implementation of the best performing RPMs in clinical practice.[5,25] To our knowledge, we are the first to independently investigate and directly compare model performance of numerous CAD RPMs though external validation in three large population-based cohorts.

We assessed several indicators of RPM performance.[26] Overall, C-indices found in the present study indicated moderate-to-good discrimination for most RPMs. However, evaluation of the C-index alone can easily overestimate RPM performance.[27] The C-index reflects the probability that a subject who experiences an event has a higher risk estimate compared to a subject who did not experience an event. In a general unselected population, chance of adequately predicting low risk is high due to the low incidence of the endpoint. This might thrive high C-indices, while this does not reflect accurate individual risk prediction by RPMs. Consequently, additional model

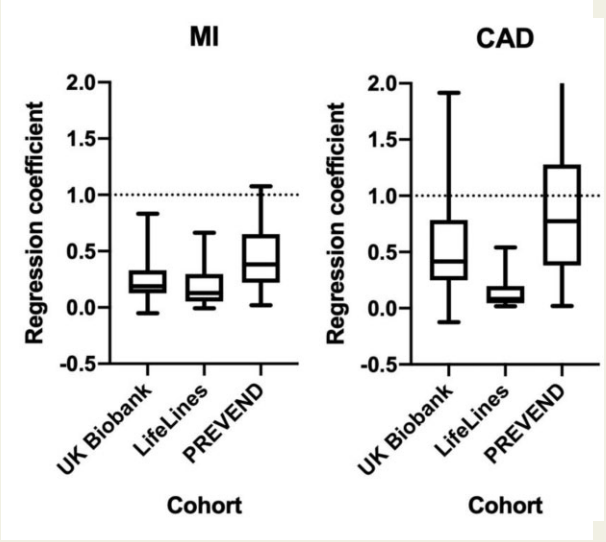

Figure 3 Regression slopes for every RPM in the three cohorts with outcomes MI and CAD. CAD, coronary artery disease; MI, myocardial infarction.

performance measures are highly recommended to ensure adequate allocation of individuals to further treatment or conservative management based on risk stratification, preventing under- or overtreatment.[28]

We additionally assessed the agreement between estimated and observed risk (i.e. calibration) to evaluate the ability of RPMs to correctly estimate probabilities of events. Overall, we observed that CAD risk was overestimated by RPMs, as indicated by mean regression slopes below 1 in all three cohorts. This is in line with previous literature, with a large randomized controlled clinical trial reporting overestimation of risk by the SCORE in comparison to coronary artery calcification as shown by cardiac computed tomography.[29]

In the present study, we also studied predictors of agreement between observed and estimated risk and found that increased endpoint incidence in the original RPM development cohort is associated with increased overestimation of risk by RPMs. Interestingly, one RPM showed good discriminatory performance in Lifelines, while performing significantly worse compared to other RPMs in UK Biobank and PREVEND.[17] This RPM was developed in a cohort with a relatively high incidence of the endpoint. Since the absolute number of events is low in the general unselected population, the better discriminatory performance of this RPM is driven by the relatively higher number of adequately labelled high-risk cases, at the cost of a relatively low number of inadequately labelled low-risk cases when endpoint incidence is very low, as is the case in Lifelines. A regression slope very close to zero was observed for this RPM in LifeLines, indicating substantial overestimation of risk, which is undesirable because of the risk of overdiagnosis. This underlines the importance of external validation of RPM performance in separate cohorts with differing incidences of predicted outcome to evaluate robustness of RPMs. When developing and applying RPMs, the target population and intended use should be taken into careful consideration.

We did not identify RPMs that consistently performed best across all three cohorts when compared to other RPMs. This suggest that development of new RPMs in the same fashion as before will not lead to new RPMs outperforming currently existing RPMs and that priority might be safely given to further implementation of commonly known and currently widely deployed RPMs for primary prevention, such as the FHS and SCORE. Final conclusions can be drawn after further evaluation of RPMs clinical utility by decision curve analysis.[30] Potential for substantial improvement in RPMs performance might lie in integrating artificial intelligence-algorithms to automatically validate and recalibrate RPMs.[7] Indeed, dynamic modelling involves continuous updating of models when new data becomes available, for example from coupling with electronic patient files. The new data leads to recalibration of the model, hereby decreasing the time between development of the model and its time of use. Rather than developing new RPMs, model updating allows for continuous and automated adaptation of existing RPMs to recent information.[7,31] In addition, recalibration of existing RPMs specific for the intended target population might improve model performance without a need to develop new RPMs. We did identify some RPMs that performed significantly worse across all three validation cohorts, suggesting that these models should not be used in clinical practice, depending on the population used.

The variables used in the models were similar to a large extent. Age, smoking, gender, and blood pressure/hypertension were used in the majority of RPMs, in line with current knowledge on traditional modifiable and non-modifiable risk factors for CAD.[25,32] Interestingly, established risk factors such as diabetes mellitus, low-density lipoprotein, and body mass index were used in a minority of RPMs. Yet, the prevalence of CAD in diabetics has been estimated at more than 20%.[33] This raises the question whether development of new or updating of existing RPMs should involve exploration of the value of diabetes mellitus as a variable. Relatively novel risk factors such as plasma concentrations of C-reactive protein, interleukins, microalbuminuria, and genetic markers were not frequently used.[34] Moreover, proven lifestyle factors such as grip strength and walking pace were not included in RPMs.[35,36]

Nearly all RPMs included gender as an independent variable with corresponding weighing to estimate risk. As such, we were unable to investigate gender-dependent differences in RPM performance, because gender is already accounted for in the RPMs and exclusion of this variable from the model would therefore seriously undermine validity of the original RPMs. In our literature search, we did not identify RPMs designed for males and females specifically. To further improve RPM performance, gender-specific selection and weighing of variables could be considered.

Many articles describe a basic RPM, as well as several submodels that have some differences in variables or methodology applied. We showed that submodels did not improve discrimination and calibration. This raises the question whether development of submodels is useful or only obscures vision on usefulness of separate RPMs.

Clinical implications

Advantages of RPMs include their low costs, easy use and non-invasiveness. RPMs are therefore important tools in primary prevention of MI and CAD. However, there is an abundance of RPMS available

and ambiguity on RPM performance in different populations exists.[37] Primary care doctors are reported to be insecure about which RPM to use and whether older scores should continue to be used.[38] For these reasons, implementation of systematic risk counselling in clinical practice is suboptimal. In addition, our data show that RPMs tend to overestimate the risk of future MI and CAD, especially in low-risk populations. Overestimation of risk might lead to unnecessary treatments with risk of adverse effects, which is undesirable in a primary preventive setting. Improved strategies, including dynamic modelling should be developed to enhance implementation of systematic RPM use or RPM-based population-screening.

Strengths and limitations

We used three independent population-based cohorts from two different European countries with large sample sizes. In addition, we used identical prespecified outcomes for comparisons between all RPMs, allowing for head-to-head comparison. Furthermore, the relatively large number of different RPMs evaluated allowed us to even study predictors of RPM performance. Most RPMs are developed to predict the occurrence of CAD in a certain time frame (exposure time), usually 10 years. We decided not to incorporate the exposure time in the validations, as this would have decreased power. This may have led to an over- or underestimation of the endpoint incidence in development and validation cohorts. However, this does not influence comparisons between RPMs, as this is consistently and equally present for all RPMs across all three cohorts. Data were missing in the validation cohorts for some variables of some RPMs, decreasing the number of participants used for validation of these RPMs. We decided not to impute values, as we chose to minimize bias and increase representativeness of results.

The current study evaluated RPMs for incident CAD in the general population. For future research, it may be worthwhile to also investigate RPMs predicting recurrences of CAD events, such as the EUROASPIRE risk calculator.[39]

Conclusions

We identified 11 articles containing 28 RPMs predicting CAD in the general population. External validation in the UK Biobank, LifeLines, and PREVEND cohorts yielded comparable moderate-to-good discrimination and calibration for most RPMs, indicating that no single best-performing RPM could be distinguished. RPM performance was found to depend on endpoint incidence in the original development and validation cohorts, underlining the importance of external validation in multiple independent cohorts to evaluate RPM robustness and generalizability, in addition to careful consideration of the target population and intended use when developing RPMs.

Supplementary material

Supplementary material is available at *European Journal of Preventive Cardiology* online.

Acknowledgements

The authors wish to acknowledge Evidencio, in particular Tom Hueting and Erik Verbeek, for their (technical) assistance with the validation analyses.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of interest: R.P. reports a minority stake in Evidencio B.V., an online platform offering free services regarding the creation, validation, and implementation of clinical prediction models. Evidencio was not involved in the development of any of the prediction models mentioned nor is expected to experience financial gain by the publication of this manuscript.

References

1. Sanchis-Gomar F, Perez-Quilis C, Leischik R, Lucia A. Epidemiology of coronary heart disease and acute coronary syndrome. *Ann Transl Med* 2016;**4**:256–256.
2. Lerner DJ, Kannel WB. Patterns of coronary heart disease morbidity and mortality in the sexes: a 26-year follow-up of the Framingham population. *Am Heart J* 1986;**111**:383–390.
3. Task Force Members, Montalescot G, Sechtem U, Achenbach S, Andreotti F, Arden C, Budaj A, Bugiardini R, Crea F, Cuisset T, Di Mario C, Ferreira JR, Gersh BJ, Gitt AK, Hulot JS, Marx N, Opie LH, Pfisterer M, Prescott E, Ruschitzka F, Sabaté M, Senior R, Taggart DP, van der Wall EE, Vrints CJ; ESC Committee for Practice Guidelines, Zamorano JL, Achenbach S, Baumgartner H, Bax JJ, Bueno H, Dean V, Deaton C, Erol C, Fagard R, Ferrari R, Hasdai D, Hoes AW, Kirchhof P, Knuuti J, Kolh P, Lancellotti P, Linhart A, Nihoyannopoulos P, Piepoli MF, Ponikowski P, Sirnes PA, Tamargo JL, Tendera M, Torbicki A, Wijns W, Windecker S; Document Reviewers, Knuuti J, Valgimigli M, Bueno H, ClaeysMJ, Donner-BanzhoffN, Erol C, Frank H, Funck-Brentano C, Gaemperli O, Gonzalez-Juanatey JR, Hamilos M, Hasdai D, Husted S, James SK, Kervinen K, Kolh P, Kristensen SD, Lancellotti P, Maggioni AP, Piepoli MF, Pries AR, Romeo F, Rydén L, Simoons ML, Sirnes PA, Steg PG, Timmis A, Wijns W, Windecker S, Yildirir A, Zamorano JL. 2013 ESC guidelines on the management of stable coronary artery disease: the Task Force on the management of stable coronary artery disease of the European Society of Cardiology. *Eur Heart J* 2013;**34**:2949–3003.
4. Ginghina C, Bejan I, Ceck CD. Modern risk stratification in coronary heart disease. *J Med Life* 2011;**4**:377–386.
5. Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, Tzoulaki I, Lassale CM, Siontis GCM, Chiocchia V, Roberts C, Schlüssel MM, Gerry S, Black JA, Heus P, van der Schouw YT, Peelen LM, Moons KGM. Prediction models for cardiovascular disease risk in the general population: systematic review. *Bmj* 2016;**353**:i2416.
6. Buccheri S, D'Arrigo P, Franchina G, Capodanno D. Risk stratification in patients with coronary artery disease: a practical walkthrough in the landscape of prognostic risk models. *Interv Cardiol* 2018;**13**:112–120.
7. Adibi A, Sadatsafavi M, Ioannidis JPA. Validation and utility testing of clinical prediction models: time to change the approach. *JAMA* 2020;**324**:235–236.
8. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. 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.
9. Smink PA, Lambers Heerspink HJ, Gansevoort RT, de Jong PE, Hillege HL, Bakker SJL, de Zeeuw D. Albuminuria, estimated GFR, traditional risk factors, and incident cardiovascular disease: the PREVEND (Prevention of Renal and Vascular Endstage Disease) study. *Am J Kidney Dis* 2012;**60**:804–811.
10. Scholtens S, Smidt N, Swertz MA, Bakker SJL, Dotinga A, Vonk JM, van Dijk F, van Zon SKR, Wijmenga C, Wolffenbuttel BHR, Stolk RP. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol* 2015;**44**:1172–1180.
11. CPM Registry. http://pacecpmregistry.org/registry/. (January 2019).
12. Evidencio. https://www.evidencio.com/home. (February 2020).

13. van Steenbeek CD, van Maaren MC, Siesling S, Witteveen A, Verbeek XAAM, Koffijberg H. Facilitating validation of prediction models: a comparison of manual and semi-automated validation using registry-based data of breast cancer patients in the Netherlands. *BMC Med Res Methodol* 2019;**19**:117.
14. Tunstall-Pedoe H. The Dundee coronary risk-disk for management of change in risk factors. *BMJ* 1991;**303**:744–747.
15. Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J* 1991;**121**:293–298.
16. Bogle BM, Ning H, Goldberger JJ, Mehrotra S, Lloyd-Jones DM. A simple community-based risk-prediction score for sudden cardiac death. *Am J Med* 2018;**131**:532–539.e5.
17. Onat A, Can G, Hergenç G, Uğur M, Yüksel H. Coronary disease risk prediction algorithm warranting incorporation of C-reactive protein in Turkish adults, manifesting sex difference. *Nutr Metab Cardiovasc Dis* 2012;**22**:643–650.
18. Chien K-L, Hsu H-C, Su T-C, Chang W-T, Chen P-C, Sung F-C, Lin H-J, Chen M-F, Lee Y-T. Constructing a point-based prediction model for the risk of coronary artery disease in a Chinese community: a report from a cohort study in Taiwan. *Int J Cardiol* 2012;**157**:263–268.
19. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;**97**:1837–1847.
20. Nishimura K, Okamura T, Watanabe M, Nakai M, Takegami M, Higashiyama A, Kokubo Y, Okayama A, Miyamoto Y. Predicting coronary heart disease using risk factor categories for a Japanese urban population, and comparison with the Framingham Risk Score: the Suita study. *J Atheroscler Thromb* 2014;**21**:784–798.
21. Jee SH, Jang Y, Oh DJ, Oh B-H, Lee SH, Park S-W, Seung K-B, Mok Y, Jung KJ, Kimm H, Yun YD, Baek SJ, Lee DC, Choi SH, Kim MJ, Sung J, Cho B, Kim ES, Yu B-Y, Lee T-Y, Kim JS, Lee Y-J, Oh J-K, Kim SH, Park J-K, Koh SB, Park SB, Lee SY, Yoo C-I, Kim MC, Kim H-K, Park J-S, Kim HC, Lee GJ, Woodward M. A coronary heart disease prediction model: the Korean Heart Study. *BMJ Open* 2014;**4**:e005025.
22. Koller MT, Leening MJG, Wolbers M, Steyerberg EW, Hunink MGM, Schoop R, Hofman A, Bucher HC, Psaty BM, Lloyd-Jones DM, Witteman JCM. Development and validation of a coronary risk prediction model for older U.S. and European persons in the Cardiovascular Health Study and the Rotterdam Study. *Ann Intern Med* 2012;**157**:389–397.
23. L'Italien G, Ford I, Norrie J, LaPuerta P, Ehreth J, Jackson J, Shepherd J. The cardiovascular event reduction tool (CERT)–a simplified cardiac risk prediction model developed from the West of Scotland Coronary Prevention Study (WOSCOPS). *Am J Cardiol* 2000;**85**:720–724.
24. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetière P, Jousilahti P, Keil U, Njølstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham IM. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003;**24**:987–1003.
25. Frohlich ED, Quinlan PJ. Coronary heart disease risk factors: public impact of initial and later-announced risks. *Ochsner J* 2014;**14**:532–537.
26. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;**21**:128–138.
27. Pencina MJ, D'Agostino RB Sr. Evaluating discrimination of risk prediction models: the C statistic. *JAMA* 2015;**314**:1063–1064.
28. Merkow RP, Hall BL, Cohen ME, Dimick JB, Wang E, Chow WB, Ko CY, Bilimoria KY. Relevance of the c-statistic when evaluating risk-adjustment models in surgery. *J Am Coll Surg* 2012;**214**:822–830.
29. van der Aalst CM, Denissen SJAM, Vonder M, Gratama JWC, Adriaansen HJ, Kuijpers D, Vliegenthart R, van Lennep JER, van der Harst P, Braam RL, van Dijkman PRM, van Bruggen R, Oudkerk M, de Koning HJ. Screening for cardiovascular disease risk using traditional risk factor assessment or coronary artery calcium scoring: the ROBINSCA trial. *Eur Heart J Cardiovasc Imaging* 2020;**21**:1216–1224.
30. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;**26**:565–574.
31. Jenkins DA, Martin GP, Sperrin M, Riley RD, Debray TPA, Collins GS, Peek N. Continual updating and monitoring of clinical prediction models: time for dynamic prediction systems? *Diagn Progn Res* 2021;**5**:1.
32. Hajar R. Risk factors for coronary artery disease: historical perspectives. *Heart Views* 2017;**18**:109–114.
33. Einarson TR, Acs A, Ludwig C, Panton UH. Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007–2017. *Cardiovasc Diabetol* 2018;**17**:83.
34. Mack M, Gopal A. Epidemiology, traditional and novel risk factors in coronary artery disease. *Heart Fail Clin* 2016;**12**:1–10.
35. Argyridou S, Zaccardi F, Davies MJ, Khunti K, Yates T. Walking pace improves all-cause and cardiovascular mortality risk prediction: a UK Biobank prognostic study. *Eur J Prev Cardiol* 2020;**27**:1036–1044.
36. Welsh CE, Celis-Morales CA, Ho FK, Brown R, Mackay DF, Lyall DM, Anderson JJ, Pell JP, Gill JMR, Sattar N, Welsh P, Gray SR. Grip strength and walking pace and cardiovascular disease risk prediction in 406,834 UK biobank participants. *Mayo Clin Proc* 2020;**95**:879–888.
37. Betts MB, Milev S, Hoog M, Jung H, Milenković D, Qian Y, Tai M-H, Kutikova L, Villa G, Edwards C. Comparison of recommendations and use of cardiovascular risk equations by health technology assessment agencies and clinical guidelines. *Value Health* 2019;**22**:210–219.
38. Liew SM, Blacklock C, Hislop J, Glasziou P, Mant D. Cardiovascular risk scores: qualitative study of how primary care practitioners understand and use them. *Br J Gen Pract* 2013;**63**:e401–e407.
39. De Bacquer D, Ueda P, Reiner Ž, *De Sutter J, De Smedt D, Lovic D, Gotcheva N, Fras Z, Pogosova N, Mirrakhimov E, Lehto S, Jernberg T, Kotseva K, Rydén L, Wood D, De Backer G, EUROASPIRE IV and V National Coordinators.*. Prediction of recurrent event in patients with coronary heart disease: the EUROASPIRE Risk Model: Results from a prospective study in 27 countries in the WHO European region—the EURObservational Research Programme (EORP) of the European Society of Cardiology (ESC). *Eur J Prev Cardiol* 2021;doi:10.1093/eurjpc/zwaa128.