Damen et al. BMC Medicine (2019) 17:109
https://doi.org/10.1186/s12916-019-1340-7

BMC Medicine

RESEARCH ARTICLE Open Access

Performance of the Framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis

Johanna A. Damen[1,2*], Romin Pajouheshnia[2], Pauline Heus[1,2], Karel G. M. Moons[1,2], Johannes B. Reitsma[1,2], Rob J. P. M. Scholten[1,2], Lotty Hooft[1,2] and Thomas P. A. Debray[1,2]

Abstract

Background: The Framingham risk models and pooled cohort equations (PCE) are widely used and advocated in guidelines for predicting 10-year risk of developing coronary heart disease (CHD) and cardiovascular disease (CVD) in the general population. Over the past few decades, these models have been extensively validated within different populations, which provided mounting evidence that local tailoring is often necessary to obtain accurate predictions. The objective is to systematically review and summarize the predictive performance of three widely advocated cardiovascular risk prediction models (Framingham Wilson 1998, Framingham ATP III 2002 and PCE 2013) in men and women separately, to assess the generalizability of performance across different subgroups and geographical regions, and to determine sources of heterogeneity in the findings across studies.

Methods: A search was performed in October 2017 to identify studies investigating the predictive performance of the aforementioned models. Studies were included if they externally validated one or more of the original models in the general population for the same outcome as the original model. We assessed risk of bias for each validation and extracted data on population characteristics and model performance. Performance estimates (observed versus expected (OE) ratio and c-statistic) were summarized using a random effects models and sources of heterogeneity were explored with meta-regression.

Results: The search identified 1585 studies, of which 38 were included, describing a total of 112 external validations. Results indicate that, on average, all models overestimate the 10-year risk of CHD and CVD (pooled OE ratio ranged from 0.58 (95% CI 0.43–0.73; Wilson men) to 0.79 (95% CI 0.60–0.97; ATP III women)). Overestimation was most pronounced for high-risk individuals and European populations. Further, discriminative performance was better in women for all models. There was considerable heterogeneity in the c-statistic between studies, likely due to differences in population characteristics.

Conclusions: The Framingham Wilson, ATP III and PCE discriminate comparably well but all overestimate the risk of developing CVD, especially in higher risk populations. Because the extent of miscalibration substantially varied across settings, we highly recommend that researchers further explore reasons for overprediction and that the models be updated for specific populations.

Keywords: Cardiovascular disease, Prediction models, Prognosis, Systematic review, Meta-analysis

* Correspondence: J.A.A.Damen@umcutrecht.nl
[1]Cochrane Netherlands, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
[2]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, P.O. Box 85500, Str. 6.131, 3508, GA, Utrecht, The Netherlands

© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Background

Cardiovascular disease (CVD) is a major health burden, accounting for 17.5 million deaths worldwide in 2012 [1]. To effectively and efficiently implement preventive measures such as lifestyle advice and lipid-lowering drugs, early identification of high-risk individuals for targeted intervention using so-called CVD risk prediction models or risk scores is widely advocated [2]. Evidently, it is crucial that predictions of CVD risk provided by these models are sufficiently accurate. Inappropriate risk-based management may lead to overtreatment or undertreatment. Clinical guidelines from the National Cholesterol Education Program and the American College of Cardiology and American Heart Association (AHA) advise using the Framingham Adult Treatment Panel (ATP) III model [3] or the pooled cohort equations (PCE) to predict 10-year risk of CVD for all individuals 40 years or older [2]. Also, most clinical research focused on studying the Framingham Wilson model [4, 5].

All three models have been externally validated numerous times across different settings and populations, with most studies showing that their predicted risks are too high (i.e. poor calibration, see Table 1) [6–9], while other reports found adequate calibration for these same models [10, 11]. Previous reviews have summarized existing models for cardiovascular risk prediction without undertaking any formal comparison or quantitative synthesis [5, 12–14] or focussed solely on the performance of the PCE [15]. Systematic reviews, followed by a quantitative synthesis, have become a vital tool in the evaluation of a prediction model's performance across different settings and populations, and thus to better understand how the implementation of a developed model may affect clinical practice [16]. It may therefore help researchers, policy makers and clinicians to evaluate which models can be advocated in (new) guidelines for use in daily practice, and to what extent they should be updated prior to implementation.

In this review, we focus on the ATP III and PCE models as these are advocated in the clinical guidelines in the USA [2, 3]. Although Framingham Wilson is not mentioned in the clinical guidelines, it is relevant to review this prediction model, since many studies in the field of CVD risk prediction have externally validated this prediction model and have used it to assess the incremental value of new predictors or for comparison with newly developed prediction models [5]. We did not include other prediction models such as the SCORE [21] or QRISK [22] models in this review, as these were developed on European populations and we wanted to focus on truly competing models for the American population.

We, therefore, compared the predictive performance of the originally developed Framingham Wilson, Framingham ATP III and PCE models (see Additional file 1 for details on these prediction models and our review question). We conducted a systematic review, including critical appraisal, of all published studies that externally validated one or more of these three models, followed by a formal meta-analysis to summarize and compare the overall predictive performance of these models and the predictive performance across pre-defined subgroups. We explicitly did not intend to review all existing CVD risk prediction models but focused on these three most widely advocated and studied models in the USA.

Methods

We conducted our review based on the steps described in the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) [23] and in a guidance paper on the systematic review and meta-analysis of prediction models [16]. A review protocol is available in Additional file 2. This review is reported according to the MOOSE reporting checklist (Additional file 3).

Search and selection

We started with studies published before June 2013 that were already identified in two previously published systematic reviews [5, 12]. Studies published after June 2013 were identified according to the following strategy, developed by an information specialist working for Cochrane. First, a search was performed in MEDLINE and Embase (October 25, 2017, Additional file 4). In addition, a citation search in Scopus and Web of Science was performed to find all studies published between 2013 and 2017 that cited the studies in which the development of one of the original models was described (Additional file 4). All studies that were identified both by the search in MEDLINE and Embase, and the citation search were screened for eligibility, first on title and abstract by one reviewer and subsequently on full text by two independent reviewers. Disagreements were solved in group discussions. The reference lists of systematic reviews identified by our search were screened to identify additional studies.

Eligibility criteria

Studies were eligible for inclusion if they described the external validation of Framingham Wilson 1998 [4], Framingham ATP III 2002 [3] and/or PCE 2013 [10]. Studies were included if they externally validated these models for fatal or nonfatal coronary heart disease (CHD) in the case of Framingham Wilson and ATP III, and hard atherosclerotic CVD (here referred to as fatal or nonfatal CVD) in the case of PCE, separately for men and women, in a general (unselected) population setting. For Framingham Wilson and ATP III, we thus excluded external validation studies that used the combination of

CHD and stroke as an outcome, while we included studies with so called 'hard CHD' (myocardial infarction + fatal CHD). For PCE, we excluded studies that used CHD as outcome. Studies regarding specific patient populations (e.g. patients with diabetes) were excluded. Studies in which the model was updated or altered (e.g. recalibration or model revision [24, 25], see Table 1) before external validation were excluded if they did not provide any information on the original model's performance. Studies in which the models for men and women were combined in one validation (with one performance measure reported for men and women together instead of two separate performance measures) were excluded. Studies that assessed the incremental value of an additional predictor on top of the original model were also excluded, unless the authors explicitly reported on the external validity of the original model before adding the extra predictor. When a study population was used multiple times to validate the same model (i.e. multiple publications describing a certain study cohort), the external validation with eligibility criteria and predicted outcome that most closely resembled our review question was included, to avoid introducing bias because of duplicate data [26].

Data extraction and critical appraisal

For each included study, data were extracted on study design, population characteristics, participant enrolment, study dates, prediction horizon, predicted outcomes, predictors, sample size, model updating methods and model performance (Additional file 5). Risk of bias was assessed based on a combination of the CHARMS checklist [23] and a preliminary version of the Cochrane Prediction study Risk Of Bias Assessment Tool (PROBAST) [27, 28] (Additional file 5). Risk of bias was assessed for each validation, across five domains: participant selection (e.g. study design, in- and exclusions), predictors (e.g. differences in predictor definitions), outcome (e.g. same definition and assessment for every participant), sample size and participant flow (e.g. handling of missing data), analyses (e.g. handling of censoring). After several rounds of piloting and adjusting the data extraction form in a team of three reviewers, data were extracted by one of the three reviewers. Risk of bias was independently assessed by pairs of reviewers. Disagreements were solved after discussion or by a third reviewer.

Information was extracted on model discrimination and calibration, before and, if reported, after model updating, in terms of the reported concordance (c)-statistic and total observed versus expected (OE) ratio. If relevant information was missing (e.g. standard error of performance measure or population characteristics), we contacted the authors of the corresponding study. If no additional information could be obtained, we approximated missing information using formulas described by Debray et al. [16] (Additional file 6). If reported, calibration was also extracted for different risk categories. If the OE ratio was reported for shorter time intervals (e.g. 5 years), we extrapolated this to 10 years by assuming a Poisson distribution (Additional file 6).

Table 1 Terminology

	Definition
Case-mix/patient spectrum	Characteristics of the study population (e.g. age, gender distribution)
Prediction horizon	Time frame in which the model predicts the outcome (e.g. predicting 10-year risk of developing a CVD event).
External validation	Estimating the predictive performance of an existing prediction model in a dataset or study population other than the dataset from which the model was developed.
Predictive performance	Accuracy of the predictions made by a prediction model, often expressed in terms of discrimination or calibration.
Discrimination	Ability of the model to distinguish between people who did and did not develop the event of interest, often quantified by the c-statistic.
Concordance (c)-statistic	Statistic that quantifies the chance that for any two individuals of which one developed the outcome and the other did not, the former has a higher predicted probability according to the model than the latter. A c-statistic of 1 means perfect discriminative ability, whereas a model with a c-statistic of 0.5 is not better than flipping a coin [17].
Calibration	Agreement between observed event risks and event risks predicted by the model.
Observed versus expected (OE) ratio	The ratio of the total number of outcome events that occurred (e.g. in 10 years) and the total number of events predicted by the model. The OE ratio can be calculated for the entire study population (further referred to as 'total OE ratio'), or in categories of predicted risks.
Calibration slope	Measure that gives an indication of the strength of the predictor effects. The calibration slope ideally equals 1. A calibration slope < 1 indicates that predictions are too extreme (low-risk individuals have a predicted risk that is too low, and high-risk individuals are given a predicted risk that is too high). Conversely, a slope > 1 indicates that predictions are too moderate [18, 19].
Model updating/ recalibration	When externally validating a prediction model, adjusting the model to the dataset in which the model is validated, to improve the predictive performance of the model.
Updating the baseline hazard or risk	When externally validating a prediction model, adapting the original baseline hazard or intercept of the prediction model to the dataset in which the model is validated. This updating method corrects for differences in observed outcome incidence between the original development and external validation dataset.
Updating the common slope	When externally validating a prediction model, adapting the beta coefficients of the model using a single correction factor, to proportionally adjust for changes in predictor outcome associations [20].
Model revision	Taking the predictors of an existing previously developed model and fitting these in the external dataset by estimating the new predictor-outcome associations (e.g. regression coefficients).

Statistical analyses

We performed meta-analyses of the OE ratio and the c-statistic for 10-year risk predictions. Based on previous recommendations [16, 29], we pooled the log OE ratio and logit c-statistic using random-effects meta-analysis. Further, we stratified the meta-analysis by model and gender, resulting in six main groups: Wilson men, Wilson women, ATP III men, ATP III women, PCE men, PCE women. We calculated 95% confidence intervals (CI) and (approximate) 95% prediction intervals (PI) to quantify uncertainty and the presence of between-study heterogeneity. The CI indicates the precision of the summary performance estimate, and the PI provides boundaries on the likely performance in future model validation studies that are comparable to the studies included in the meta-analysis, and can thus be seen as an indication of model generalizability (Additional file 7) [30]. The observed and predicted probabilities were plotted in risk categories against each other and combined into a summary estimate of the calibration slope using mixed effects models (Additional file 7).

Since between-study heterogeneity in estimates of predictive performance is expected due to differences in the design and execution of validation studies [16], we investigated whether the c-statistic differed between validation studies with different eligibility criteria or actual case-mix. Furthermore, we performed univariable random effects meta-regression analyses to investigate the influence of case-mix differences (e.g. due to differences in eligibility criteria) on the OE ratio and c-statistic (Additional file 7). Several pre-specified sensitivity analyses were performed in which we studied the influence of risk of bias and alternative weighting methods in the meta-analysis on our findings (Additional file 7). All analyses were performed in R version 3.3.2 [31] using the packages metafor [32], mvmeta [33], metamisc [34] and lme4 [35].

Results

Identification and selection of studies

We first identified 100 potentially eligible studies from previously conducted systematic reviews. An additional search identified 1585 studies since June 2013 (Fig. 1). Of these 1685 studies, 304 studies were screened on full-text and data were extracted for 61 studies, describing 167 validations of the performance of one or more of the three models. Finally, 38 studies (112 validations) met our eligibility criteria [6–11, 36–67].

Description of included validations

In 112 validations (Additional file 10), the Framingham Wilson model was validated 38 times (men 23, women 15), Framingham ATP III 13 times (men 7, women 6) and PCE 61 times (men 30, women 31). One study performed a direct (head-to-head) comparison of all six prediction models [7] and one other study performed a direct comparison of the ATP III and PCE models [6]. Study participants were recruited between 1965 and 2008 and originated from North America (56), Europe (29), Asia (25) and Australia (2). All outcome definitions are described in Additional file 8. The median event rate across the included validations was 4.4% (IQR 2.8%–7.9%) and ranged between 0.5 and 29.4%. We excluded 18 and 9 external validations because the OE ratio and c-statistic, respectively, were not available, and subsequently excluded 20 and 26 external validations for the OE ratio and c-statistic, respectively, because cohorts were used multiple times to validate the same model (Additional file 9). This resulted in the inclusion of 74 validations in the analyses of the OE ratio and 77 validations in the analyses of the c-statistic (Fig. 1, Additional file 14).

Risk of bias

For participant selection, most validations scored low risk of bias ($n = 60$ (81%) and $n = 64$ (83%) for validations reporting OE ratio and c-statistic, respectively, Fig. 2). Risk of bias for predictors was often unclear ($n = 22$ (30%) and $n = 24$ (31%) for OE ratio and c-statistic) due to poor reporting of predictor definitions and measurement methods. Most validations scored low risk of bias on outcome ($n = 53$ (72%), $n = 59$ (77%)). More than three quarters of the validations scored high risk of bias for sample size and participant flow ($n = 59$ (80%) and $n = 60$ (78%)), often due to inadequate handling of missing data (i.e. simply ignoring). Low risk of bias was scored for the analysis in 51 (70%) and 50 (65%) validations, for OE ratio and c-statistic respectively. In total, 62 (84%) and 63 (82%) validations scored high risk of bias for at least one domain, and 4 (5%) and 6 (8%) validations scored low risk of bias for all five domains, for OE ratio and c-statistic, respectively.

Calibration

Figure 3 shows the calibration of the six main models, as depicted by their 10-year total OE ratio. For 24 out of 74 validations (32%), maximum follow-up was shorter than 10 years. For 20 out of these 24 (83%), information was available to extrapolate the OE ratio to 10 years. Most studies showed overprediction, indicating that 10-year risk predictions provided by the models were typically higher than observed in the validation datasets. For the Wilson model, the number of events predicted by the model was lower than the actual number of events in two studies (one in healthy siblings of patients with premature coronary artery disease [66], and one in community-dwelling individuals aged 70–79 [60]). For the PCE, underestimation of the number of events occurred in Chinese [51] and Korean [47] populations.

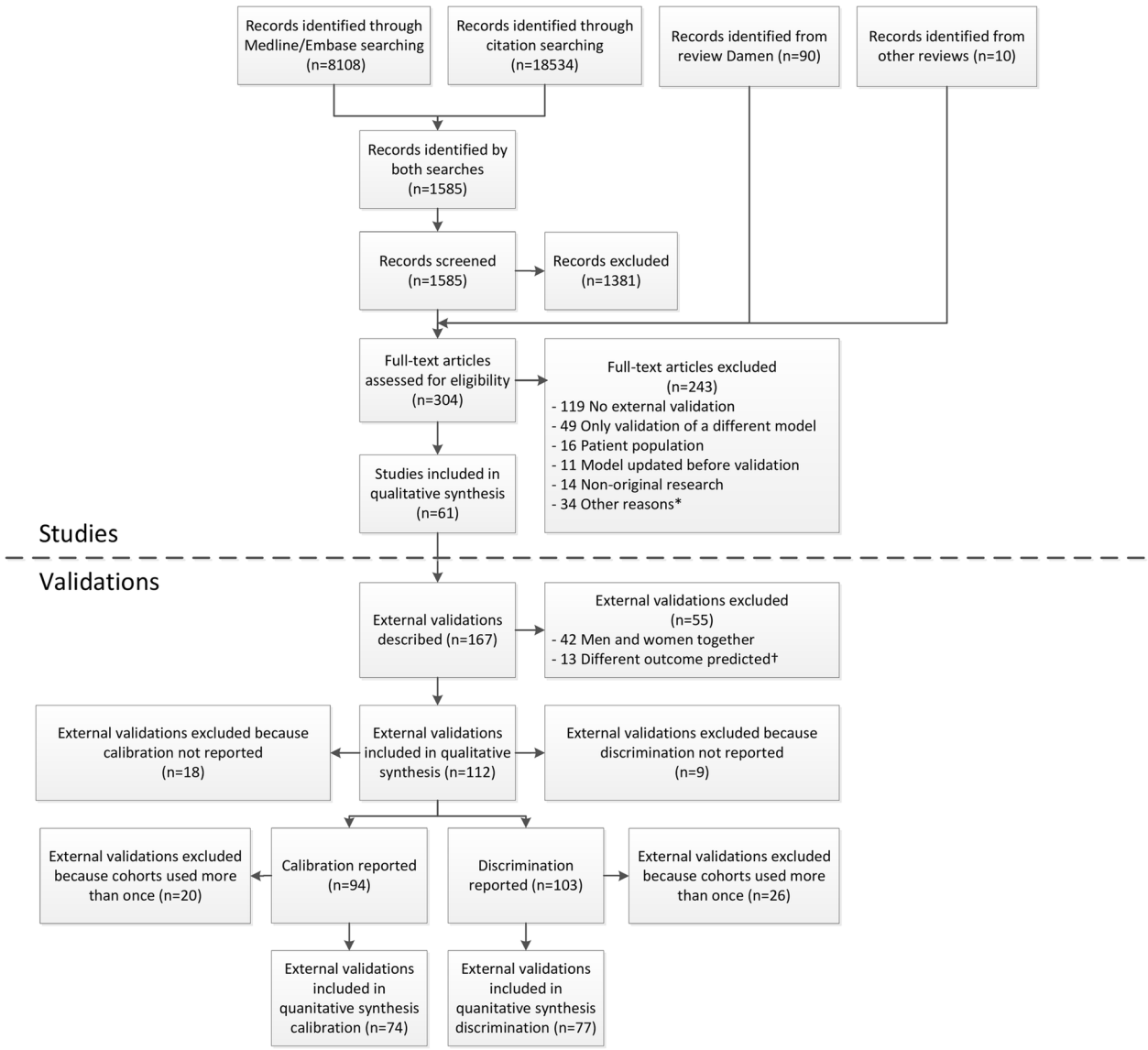

Fig. 1 Flow diagram of selected studies. Two searches were performed; one in MEDLINE and Embase and one in Scopus and Web of Science. Only studies identified by both searches were screened for eligibility, supplemented with records identified from previous systematic reviews. One study could describe more than one external validation (e.g. one for men and one for women); therefore, 61 studies described 167 external validations. Calibration was reported in 94 validations (41 directly reported, 19 provided by the authors on request, 34 estimated from calibration tables and calibration plots), and discrimination in 103 validations (91 c-statistics directly reported, 12 provided by the authors on request. Precision of c-statistic: 45 directly reported, 24 provided by the authors, 32 estimated from the sample size and 2 not reported). Some external validations were excluded because cohorts were used more than once to validate the same model (Additional file 9). *For example, no cardiovascular outcome and not written in English. †The Framingham Wilson and ATP III models were developed to predict the risk of fatal or nonfatal coronary heart disease, and the PCE model was developed to predict the risk of fatal or nonfatal cardiovascular disease. External validations that used a different outcome were excluded from the analyses (Additional file 8)

Meta-analysis revealed a considerable degree of between-study heterogeneity in OE ratios (Fig. 3), but with clear overprediction, as summary OE ratios ranged from 0.58 (Wilson men and ATP III men) to 0.79 (ATP III women). Additional analyses revealed that overprediction is more pronounced in high-risk patients for all models (Fig. 4). The results of the summary calibration slope suggest that miscalibration of the Framingham Wilson and ATP III models, and PCE men model was mostly related to heterogeneity in baseline risk (as the summary calibration slope is close to 1), while for PCE women we found a slope around 0.8, suggesting that this model was overfitted or does not transport well to new populations (Additional file 11).

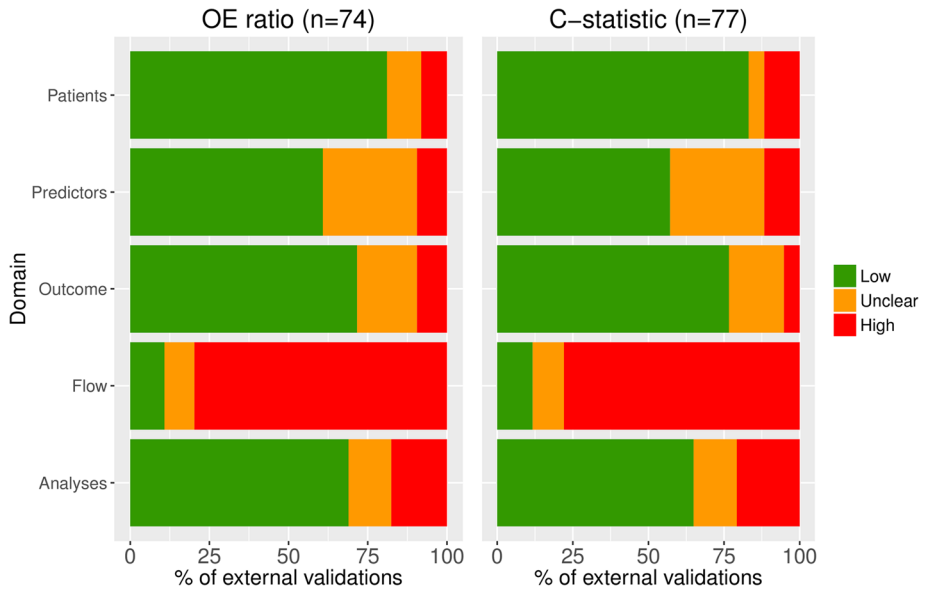

Fig. 2 Risk of bias assessment. Summary of risk of bias assessments for validations included in the meta-analyses of OE ratio (74 validations) and c-statistic (77 validations)

Discrimination

For all models, discriminative performance was slightly better for women than for men, although there was considerable variation between studies (Fig. 5). One head-to-head comparison of all three models showed worse discriminative performance for the Wilson model compared to the ATP III and PCE models [7].

Sensitivity analyses

Sensitivity analyses revealed no effect of study quality and different weighting strategies on the pooled performance of the models, both for calibration and discrimination (Additional file 12).

Factors that influence performance of the models

For women, the highest c-statistics were reported in studies with a large variety in case-mix. For men, such a trend was not visible (Fig. 6). The OE ratio for the Wilson model in the USA was closer to 1 compared to Europe, but the number of external validations per subgroup was very small (Additional file 13). Furthermore, the OE ratio appeared to decrease (further away from 1, i.e. more overprediction) with increasing mean total cholesterol. No evidence was found of an association between the OE ratio and other case-mix variables or the start date of participant recruitment. The c-statistic appeared to decrease with increasing mean age, mean systolic blood pressure and standard deviation of HDL cholesterol, and to increase with increasing standard deviation of age and total cholesterol (Additional file 13). No statistically significant associations were found between the c-statistic and other variables.

Performance after updating

For 40 validations, the model was subsequently updated, of which 24 reported the OE ratio and 15 reported the c-statistic after updating (Fig. 7). We found that substantial improvements in OE ratio were often obtained by simply re-estimating the baseline hazard or the common slope. More advanced methods of updating (e.g. entire revision of the model) did not offer much additional improvements. For the c-statistic, only advanced methods of updating resulted in limited improvement.

Discussion

Summary of findings

We systematically reviewed the performance of the Framingham Wilson, Framingham ATP III and PCE models for predicting 10-year risk of CHD or CVD for men and women separately in the general population. We found only small differences in pooled performance between the three models, but large differences in performance between validations of the same model across different populations. Although we mostly had to rely on indirect comparisons of the models, we found that performance of all three models was consistently better in women than in men for both discrimination and calibration. We found that all models overestimated the risk of CHD or CVD if they were not updated locally prior to implementation. This overestimation was more pronounced in European populations compared to the USA. Overprediction clearly declined when the validated models were adjusted (e.g. via updating the baseline hazard) to the validation setting at hand. This indicates that although the Framingham models and PCE have the

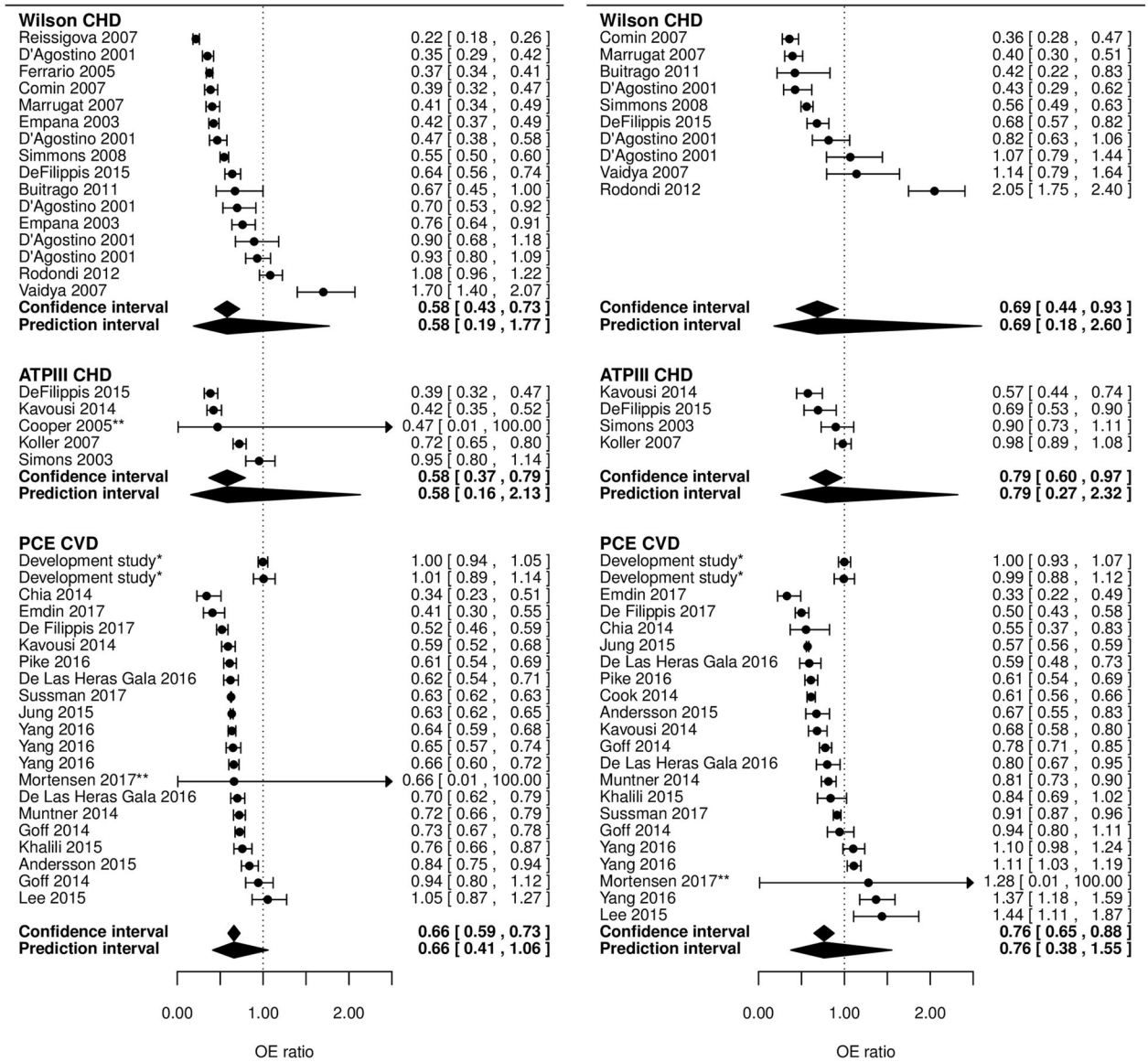

Fig. 3 Forest plots of the OE ratio in external validations. Ninety-five percent confidence intervals and 95% prediction intervals per model are indicated. The performance of the model in the development study is shown in the first rows (only reported for PCE). This estimate is not included in calculating the pooled estimate of performance. *Performance of the model in the development population after internal validation. The first row contains the performance of the model for Whites, the second for African Americans. **Standard error was not available. CHD: Coronary heart disease, CVD: cardiovascular disease

potential to be effective tools for patient management in clinical care, their use should be preceded by local adjustments of the baseline hazard. If aforementioned models are implemented as originally developed to guide treatment decisions, over- or undertreatment of individuals may occur and therefore introduce unnecessary costs or even harm. Unfortunately, current guidelines do not recommend local tailoring of Framingham and PCE and may therefore lead to suboptimal decision making. Although it was not possible to identify key sources of heterogeneity, we found that discriminative performance tends to increase as populations become more diverse, i.e. with a wider case-mix. This effect has previously been explained [69–71].

Comparison with previous literature

Our findings are in agreement with previous studies, which also found that the Framingham prediction models overestimate the risk of CHD in the general population [12, 13] and that (overall) calibration is better

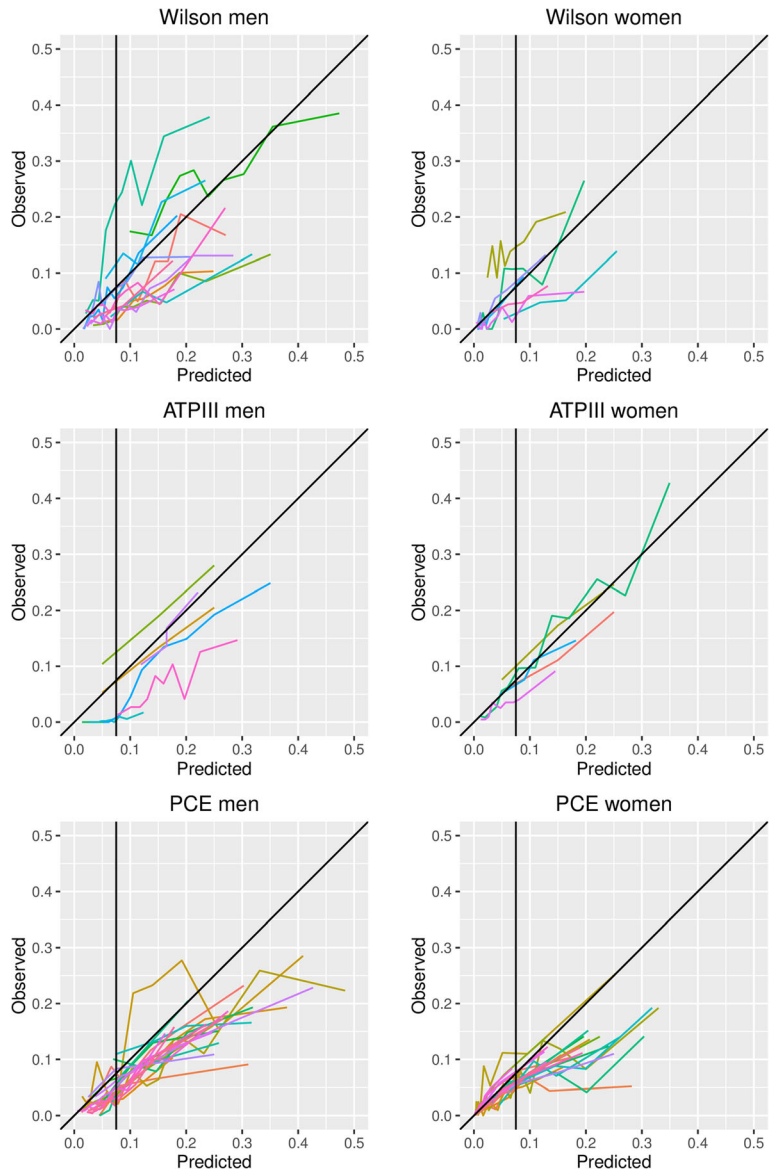

Fig. 4 Calibration plots of the Framingham Wilson, ATP III and PCE models. Each line represents one external validation. The diagonal line represents perfect agreement between observed and predicted risks. All points below that line indicate that more events were predicted than observed (overprediction) and points above the line indicate fewer events were predicted than observed (underprediction). The vertical black line represents a treatment threshold of 7.5% [68].

in the USA than in European populations [14]. Furthermore, the overestimation of risk was more pronounced in more recent populations than in earlier study populations [12, 13]. A recent study found that the PCE overestimate the risk of CVD [15], and in another study, the better discrimination seen in women was attributed to a stronger association between risk factors and CVD in women compared to men [72]. In addition to these previous reviews, we compared the predictive performance of the three prediction models that are currently in the medical guideline and that are most often evaluated in external validation studies, and found that these models have similar performance. Further, we did an extensive evaluation of factors possibly associated with heterogeneity in predictive performance and we found that predictive performance improved if these existing models were being updated.

Reasons for overprediction

There could be several reasons for the observed overprediction. These reasons have also been extensively discussed previously with regard to the PCE [15, 73]. First, differences in eligibility criteria (e.g. the exclusion of participants with previous CVD events) across validation

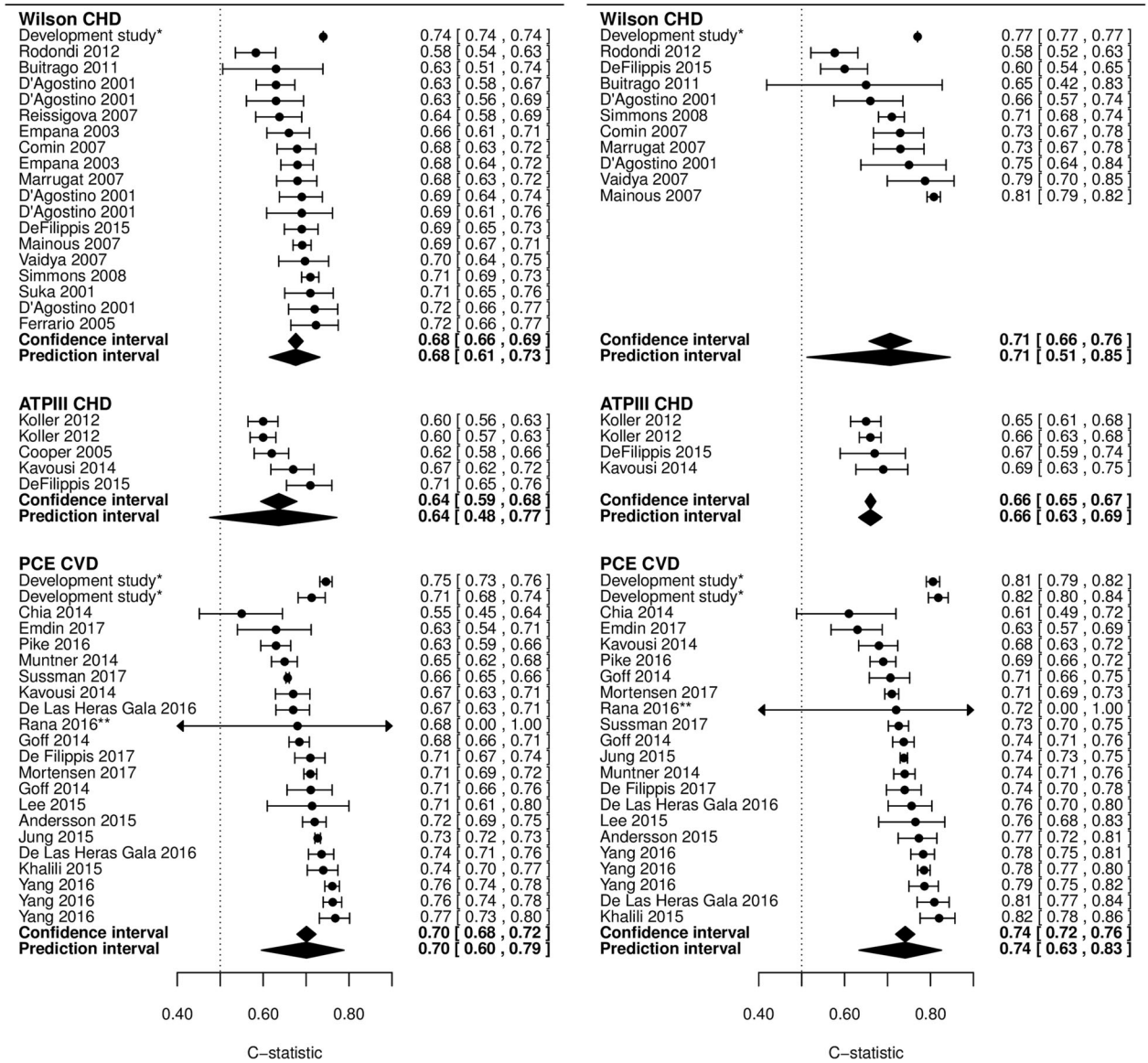

Fig. 5 Forest plots of c-statistic in external validations. Ninety-five percent confidence intervals and 95% prediction intervals per model are indicated. The performance of the model in the development study is shown in the first row(s) (not reported for the ATP III model) and is not included in the pooled estimate of performance. *Performance of the model in the development population (Wilson (no standard error reported)) and after 10 × 10 cross-validation (PCE). For the PCE, the first row contains the performance of the White model and the second the African American model. **Standard error was not available. CHD: coronary heart disease, CVD: cardiovascular disease

studies may have affected calibration. Second, the three prediction models have been (partly) developed using data from the 1970s, and since then, treatment of people at high risk for a CVD event has changed considerably, such as the introduction of statins in 1987 [74]. The increased use of effective treatments over time aimed at preventing CVD events will have lowered the observed number of events in more recent validation studies, resulting in overestimation of risk in these validation populations [39, 75, 76]. This would also explain why overprediction was most pronounced in high-risk individuals and why we found more overprediction in studies with increasing mean total cholesterol levels. We hypothesized that the degree of overprediction would increase over the years [12, 13, 77]; however, this could not be confirmed statistically. About one third of validations of the PCE excluded participants receiving treatment to lower CVD risk at baseline, but we found no difference in performance between validations that did or did not exclude these participants. However, as the use of risk-lowering medication during follow-up was

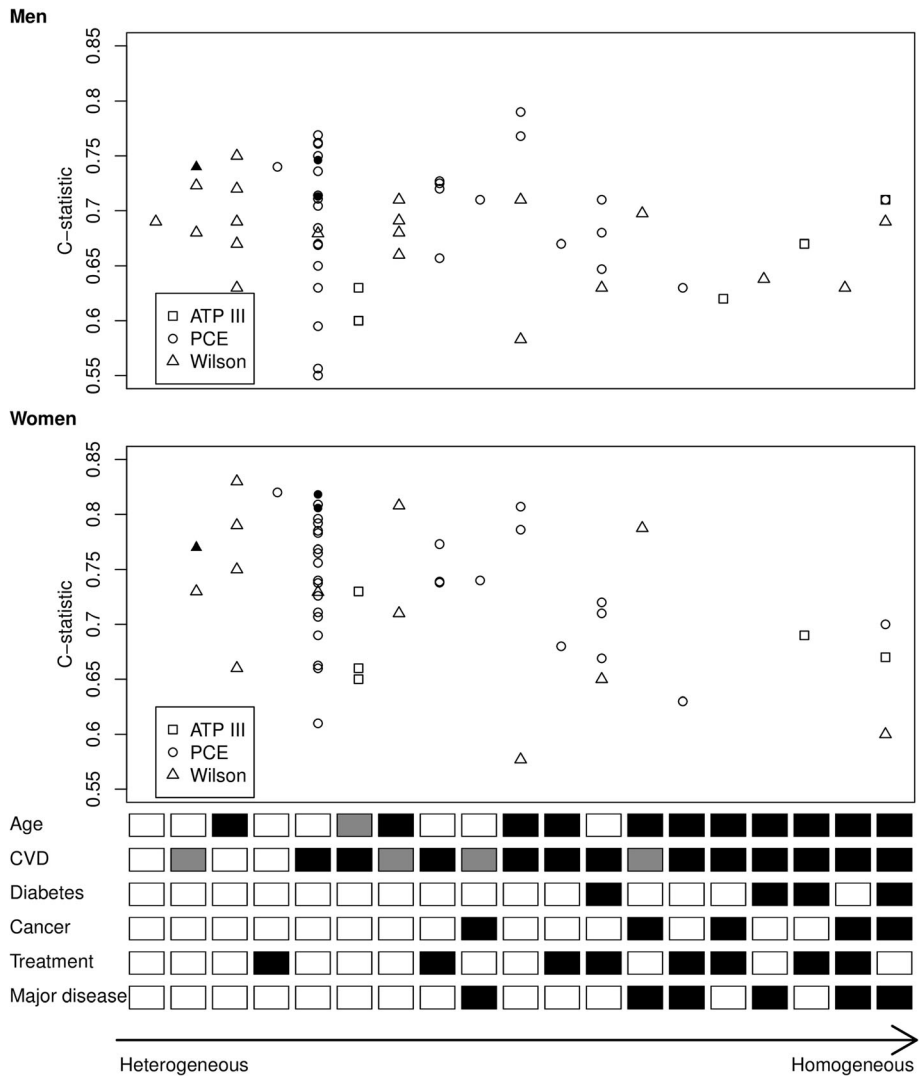

Fig. 6 C-statistic for different combinations of eligibility criteria. The open squares, circles and triangles represent validations of the ATP III, PCE and Wilson model, respectively. The black circles and triangles represent the performance of the PCE models for Whites and African-Americans, and Wilson models, in the development populations. Lower part: for age, white means a broad age range was included (difference between upper and lower age limit > 30 years), black means a narrow age range was included (difference between upper and lower age limit ≤ 30 years) and grey means age was not reported. For CVD, white means no exclusion of people with CHD or CVD, grey means people with previous CHD events were excluded from the study and black means people with previous CVD events were excluded from the study. For diabetes, cancer and major disease, white means that no restrictions were reported and black means that people with these conditions were excluded. For treatment, white means no restrictions and black means people who were receiving any treatment to lower their risk of CVD (e.g. anti-hypertensives) were excluded from the study

rarely reported in these studies, we cannot rule out an effect of incident treatment use on model performance [76]. Following the recently issued Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) guideline [78, 79] and the guidance on adjusting for treatment use in prediction modelling studies [75, 76, 80], we also strongly recommend investigators of future prediction model studies to record the use of treatment during follow-up. Third, in agreement with previously published reviews [12–15], we found more overestimation of risk in European populations compared to those of the USA whereas in some Asian populations an underestimation was seen. Both suggest that differences between these populations in, for example, unmeasured CVD risk factors and in the use of preventive CVD strategies (e.g. medical treatment or lifestyle programs) are responsible. Unfortunately, not enough information was available to study the role of ethnicity on the heterogeneity of the models' predictive accuracies. Finally, rather than overprediction by the

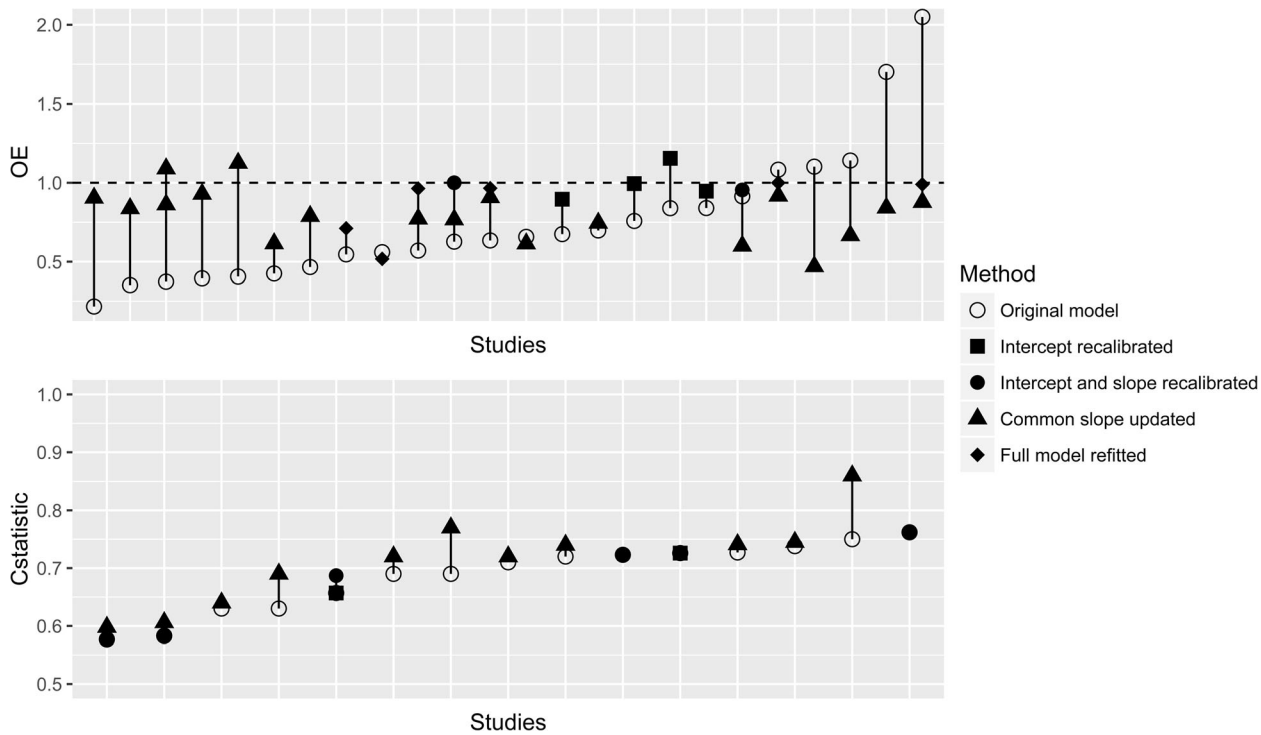

Fig. 7 Performance of models before and after update. The *x*-axis is sorted by performance before updating. The lines connect performance of models in the same cohort before and after updating

models, there could also be issues in the design of the external validation studies that give rise to a lower number of identified events. Underascertainment or misclassification of outcome events, unusually high rates of people receiving treatment, short follow-up duration, and inclusion of ethnicities not included in development of the models, have been mentioned as reasons for the overprediction we also observed [57, 81–84]. Also outcome definitions can be different between the development and validation studies. This seems particularly a problem for the Framingham Wilson model, as half of the validation studies did not include angina in their outcome definition. Researchers have however shown that overestimation can often not fully be explained by treatment use and misclassified outcome events [39, 85].

Implications for practice and research

According to the ACC-AHA guidelines [2], risk-lowering treatment is considered in people 40–75 years old, without diabetes, with LDL cholesterol levels between 70 and 189 mg/dl and with 10-year predicted risk of CVD ≥ 7.5%. After a discussion between clinician and patient about adverse effects and patient preferences, it is decided whether risk-lowering treatment is initiated. The overprediction observed in this review is problematic as this might change the population eligible for risk-lowering treatment. Unfortunately, this is true for all three CVD risk prediction models. As the meta-analysis indicates that overprediction does not consistently occur across different settings and populations, there is no simple solution to address this problem. From the studies that provided data on calibration in subgroups, we found that overestimation was more pronounced in high-risk individuals. When the (over)estimation of the absolute risk is already beyond the treatment probability threshold, it will not influence treatment decisions. However, overestimation of CVD risk might still influence the intensity (dose and frequency) of administered treatments and affect the patient's behaviour. Although excessive risk estimates could stimulate patients to adopt a more healthy lifestyle (similarly to patients with more risk factors [86]), it could also cause unnecessary anxiety for future cardiovascular events. For people at lower risk, this might, however, result in crossing the treatment probability boundary when, actually, they are at lower risk. If we assume that the observed risks reported in the validation studies are not influenced by factors that changed during follow-up such as treatment use, we can state that in 82% of individuals the average predicted and observed percentages were both below or both above the treatment boundary of 7.5% (i.e. concordant points); in 1% of individuals, the predicted risk was below 7.5% and the observed risk was above 7.5% which would, on average, lead to undertreatment in these groups of individuals; and in 17% of individuals, the average predicted risk was above

7.5% while the average observed risk was below 7.5% which would, on average, lead to overtreatment. Though this does not mean that 17% of individual will receive unnecessary treatment, it is important that doctors realize that some patients may receive unnecessary treatment, resulting in adverse effects, patient's burden and extra costs of treatment and monitoring.

The clinical implications of a certain c-statistic are even more difficult to predict. A low c-statistic (i.e. close to 0.5) indicates that the model discriminates no better than a coin toss, while a high c-statistic (close to 1.0) indicates that the model can perfectly separate people who develop an event from people who do not develop an event. If we have four pairs of individuals visiting the doctor, of which one individual will develop CVD in future (case) and one not (noncase), a c-statistic of 0.75 means that for 3 of these pairs, the case would indeed receive a higher predicted risk (which could still be too low or too high), while in one pair, the case would receive a lower predicted risk compared to the noncase.

In general, the performance of prediction models tends to vary substantially across different settings and populations, due to differences in case-mix and health care systems [87]. Hence, one external validation may not be sufficient to claim adequate performance and multiple validations are necessary to get insight in the generalizability of prediction models [71]. Primary external validation studies already showed the need for recalibration to improve a model's calibration and better tailor it to the setting at hand. In this systematic review, we now show that this also holds when all studies are pooled together in a meta-analysis. We found that none of the models offer reliable predictions unless (at least) their baseline risk or hazard (and, if applicable, population means of the predictors in the model) are recalibrated to the local setting. Studies that reported performance of the model before and after update showed that performance indeed improves after update; however, the necessary adjustments vary from setting to setting and thus need to be evaluated locally [8, 11, 36, 47, 48, 54]. As previously emphasized, more extensive revision methods are often not needed [24, 25, 88]. Hence, it appears that conventional predictors, such as age, smoking, diabetes, blood pressure and cholesterol, are still relevant indicators of 10-year CHD or CVD risk, and their associations with CVD events have largely remained stable. The need for updating CVD risk prediction models has already been discussed more than 15 years ago [11, 89], but still nothing has changed. We believe this should change now, especially since nowadays applying simple model updating is becoming increasingly possible, due to improvements in the storage of the information required to update a model. Clinical guidelines should advocate that the performance of the models is not appropriate for every patient population, and recalibration before using the models is necessary. Nice examples of the tailoring of CVD risk prediction models to specific populations are the Globorisk prediction model, which can easily be tailored to different countries using country-specific data on the population prevalence of outcomes and predictors [90], and the SCORE model, which has been tailored to many European countries using national mortality statistics [21, 91–93].

Furthermore, it is very important to gain insight into factors that cause the observed overestimation of events. This can be studied in empirical data, like the study by Cook et al. [38] where hypothetical situations are created in which a number of events are missed or prevented by risk-lowering drugs. Also simulation studies might give more insight into reasons for the overestimation of CVD risk.

These suggestions, however, offer no short-term solution for practitioners currently using the three reviewed prediction models. For now, we advise practitioners in the USA to continue using the PCE, though being aware of the potential miscalibration especially in high-risk individuals. The PCE are currently advised to be used by the ACC-AHA clinical guidelines, which is understandable because alternatives have not been studied in enough detail. However, the actual value of the PCE in clinical practice is limited by their calibration. We therefore advise to recalibrate the model in the near feature to specific populations. Fortunately, systematic reviews have shown that the prevalence of common CVD risk factors decreases (e.g. cholesterol levels drop) in populations where CVD risk prediction models and their corresponding treatment guidance are being used [94, 95]. Furthermore, statins have been proven effective with limited adverse events [96]. Finally, we advise practitioners to choose a model that predicts a clinically relevant outcome (for example, according to the AHA, CVD rather than only CHD, since stroke and CHD share pathophysiological mechanisms [10, 97]), consists of predictors available in their situation and is developed or updated in a setting that closely resembles their own setting.

Limitations

This study has several limitations. First, we focused on the three most validated and used prediction models in the USA, while in Europe many more prediction models are currently used for predicting cardiovascular risk, such as QRISK3 [22] and SCORE [21]. The differences between all these models are however limited, as most models include the same core set of predictors. Therefore, we believe our results can be generalized to other prediction models. Second, we had to rely on what is reported by the authors of primary validation studies and we unfortunately had to exclude relevant validations from our meta-analyses because of unreported information

which we could not obtain from the authors. Only 19 out of 61 authors were able to provide us with additional information, and we had to exclude 9 validations for the c-statistic and 18 for the OE ratio. Sometimes, incomplete reporting made it difficult to judge whether a study should be included in our analyses, for example when there was no explicit reference to the validated model and when it was not clear whether changes were made to the model before validation. In one specific study (D'Agostino et al. 2001 [14]), we decided to include the study because the validated model was similar in predictors and outcome definition. Although its eligibility could be debated, the exclusion of this study would not alter our conclusions regarding the performance of the Framingham Wilson model. Due to poor reporting, many validations scored unclear risk of bias, especially for the domain predictors and outcome. Many other validations scored high risk of bias which may hamper our conclusions. Unfortunately, evidence on the impact of these issues on model performance is still limited [98], which makes it difficult to argue in which direction the predictive performance of the models will change if all validations had low risk of bias. Third, the total OE ratio, while commonly reported, only provides an overall measure of calibration. To overcome this problem, we extracted information on the OE ratio in categories of predicted risk, which showed there was more overestimation of risk in the highest categories of predicted risk. Based on this information, we calculated the calibration slope, which suggested that miscalibration of the Framingham Wilson and ATP III models and PCE men model was mostly related to heterogeneity in baseline risk, while for PCE women the model is overfitted or does not transport well to new populations. In addition, more clinically relevant measures, such as net benefit, could not be considered in this meta-analysis due to the lack of reporting of these measures [5]. Fourth, because of the low number of external validation studies, we did not perform meta-regression analyses for the ATP III model. Unfortunately, the relatively small sample size makes it difficult to draw firm conclusions on the sources of observed heterogeneity. Fifth, the exclusion of non-English studies could have influenced the geographical representation. However, since only one full-text article was excluded for this reason, we believe the effect on our results is limited.

Conclusions

The Framingham Wilson, Framingham ATP III and PCE prediction models perform equally well in predicting the risk of CHD or CVD, but there is large variation between validations and only few direct comparisons have been performed so far. All three prediction models overestimate the risk of CHD or CVD if no local adjustments are made, which could lead to overtreatment in clinical practice. Therefore, we recommend that future studies should investigate reasons for overprediction and that guidelines offer advise how to make better use of existing models and subsequently tailor or recalibrate them to the setting at hand.

Additional files

Additional file 1: Review question and description of prediction models. Review question according to PICOTS format and a description of the three prediction models included in this systematic review and meta-analysis. (DOCX 26 kb)

Additional file 2: Review protocol. (DOCX 56 kb)

Additional file 3: MOOSE checklist. Reporting checklist for systematic reviews. (DOC 56 kb)

Additional file 4: Search strategy. Overview of search terms and databases searched. (DOCX 16 kb)

Additional file 5: Items for data extraction and risk of bias assessment. Overview and description of items for which data have been extracted and description of how risk of bias has been assessed. (DOCX 18 kb)

Additional file 6: Formulas used to estimate missing quantitative information. Overview of formulas used to estimate missing information on performance measures. (DOCX 23 kb)

Additional file 7: Statistical analyses. Details on the statistical analyses. (DOCX 24 kb)

Additional file 8: Description of outcomes. Overview showing which outcome definitions have been excluded and which definitions have been included. (DOCX 49 kb)

Additional file 9: Cohorts used multiple times to validate the same model. Overview of cohorts that were used more than once to validate the same model and which validation was subsequently chosen for which reason. (DOCX 310 kb)

Additional file 10: Characteristics of included validations. Table with an overview of the characteristics of included validations. (DOCX 315 kb)

Additional file 11: Summary calibration slope. Table with results from the pooled calibration slope. (DOCX 16 kb)

Additional file 12: Sensitivity analyses. Table with results from sensitivity analyses. (DOCX 18 kb)

Additional file 13: Meta-regression analyses. Figures with results from meta-regression analyses. (DOCX 141 kb)

Additional file 14: Dataset. Dataset used for analyses. (CSV 166 kb)

Abbreviations

AHA: American Heart Association; ATP: Adult Treatment Panel; CHD: Coronary heart disease; CI: Confidence interval; CVD: Cardiovascular disease; IQR: Interquartile range; OE ratio: Observed versus expected ratio; PCE: Pooled cohort equations; PI: Prediction interval

Acknowledgements

The authors would like to acknowledge René Spijker for performing the search in MEDLINE and Embase, and Gary Collins and Doug Altman for their valuable input in designing the study and interpreting the results. Furthermore, we acknowledge all authors of included studies, who provided additional information on their studies: Dr. Andersson, Dr. Asgari, Dr. Buitrago, Dr. Cook, Dr. De Las Heras Gala, Dr. DeFilippis, Dr. Ferrario, Dr. Hadaegh, Dr. Khalili, Dr. Koenig, Dr. Locatelli, Dr. Marrugat, Dr. Reissigová, Dr. Ridker, Dr. Rodondi, Dr. Simmons, Dr. Subirana, Dr. Sussman, Dr. Tan, Dr. Veronesi, Dr. Vila, and Dr. Zvárová.

Funding

Financial support was received from the Cochrane Methods Innovation Funds Round 2 (MTH001F) and Cochrane Trusted methods and Support for Cochrane

Reviews of Prognostic studies. KGMM received a grant from the Netherlands Organization for Scientific Research (ZONMW 918.10.615 and 91208004) and from the CREW NHS project, grant number 2013T083, co-funded by the Dutch Heart Society. TPAD was supported by the Netherlands Organization for Scientific Research (91617050).

Availability of data and materials
All data generated or analysed during this study are included in this published article and its supplementary information files.

Authors' contributions
JAAGD, KGMM, TPAD and LH conceived the study. All authors were involved in designing the study. JAAGD, RP and PH selected the articles. JAAGD and RP extracted the data. JAAGD analysed the data in close consultation with TPAD. All authors were involved in interpreting the data. JAAGD wrote the first draft of the manuscript which was revised by all authors. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 5 February 2019 Accepted: 7 May 2019
Published online: 13 June 2019

References
1. Cardiovascular diseases (CVDs) Fact sheet N°317. http://www.who.int/mediacentre/factsheets/fs317/en/ Accessed 24 Oct 2016.
2. Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, Goldberg AC, Gordon D, Levy D, Lloyd-Jones DM, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25 Suppl 2):S1–45.
3. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation. 2002;106(25):3143–421.
4. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation. 1998;97(18):1837–47.
5. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, Lassale CM, Siontis GC, Chiocchia V, Roberts C, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ. 2016;353:i2416.
6. Kavousi M, Leening MJ, Nanchen D, Greenland P, Graham IM, Steyerberg EW, Ikram MA, Stricker BH, Hofman A, Franco OH. Comparison of application of the ACC/AHA guidelines, Adult Treatment Panel III guidelines, and European Society of Cardiology guidelines for cardiovascular disease prevention in a European cohort. JAMA. 2014;311(14):1416–23.
7. DeFilippis AP, Young R, Carrubba CJ, McEvoy JW, Budoff MJ, Blumenthal RS, Kronmal RA, McClelland RL, Nasir K, Blaha MJ. An analysis of calibration and discrimination among multiple cardiovascular risk scores in a modern multiethnic cohort. Ann Intern Med. 2015;162(4):266–75.
8. Reissigova J, Zvarova J. The Framingham risk function underestimated absolute coronary heart disease risk in Czech men. Methods Inf Med. 2007;46(1):43–9.
9. Comin E, Solanas P, Cabezas C, Subirana I, Ramos R, Gene-Badia J, Cordon F, Grau M, Cabre-Vila JJ, Marrugat J. Estimating cardiovascular risk in Spain using different algorithms. Rev Esp Cardiol. 2007;60(7):693–702.
10. Goff DC Jr, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, Greenland P, Lackland DT, Levy D, O'Donnell CJ, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25 Suppl 2):S49–73.
11. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA. 2001;286(2):180–7.
12. Beswick AD, Brindle P, Fahey T, Ebrahim S. A systematic review of risk scoring methods and clinical decision aids used in the primary prevention of coronary heart disease (supplement). London: Royal College of General Practitioners; 2008.
13. Brindle P, Beswick A, Fahey T, Ebrahim S. Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review. Heart. 2006;92(12):1752–9.
14. Eichler K, Puhan MA, Steurer J, Bachmann LM. Prediction of first coronary events with the Framingham score: a systematic review. Am Heart J. 2007;153(5):722–31 731.e721–728.
15. Cook NR, Ridker PM. Calibration of the pooled cohort equations for atherosclerotic cardiovascular disease: an update. Ann Intern Med. 2016;165(11):786–94.
16. Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, Riley RD, Moons KG. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017;356:i6460.
17. Harrell F. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Switzerland: Springer; 2015.
18. Steyerberg E. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer-Verlag; 2008.
19. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925–31.
20. Su TL, Jaki T, Hickey GL, Buchan I, Sperrin M. A review of statistical updating methods for clinical prediction models. Stat Methods Med Res. 2018;27(1):185–197.
21. Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetiere P, Jousilahti P, Keil U, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J. 2003;24(11):987–1003.
22. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ. 2017;357:j2099.
23. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, Reitsma JB, Collins GS. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744.
24. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med. 2004;23(16):2567–86.
25. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol. 2008;61(1):76–86.
26. Tramer MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. BMJ. 1997;315(7109):635–40.
27. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51–8.
28. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med. 2019;170(1):W1–33.
29. Snell KI, Ensor J, Debray TP, Moons KG, Riley RD. Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the C-statistic and calibration measures? Stat Methods Med Res. 2018;27(11):3505–22.
30. Snell KI, Hua H, Debray TP, Ensor J, Look MP, Moons KG, Riley RD. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. J Clin Epidemiol 2016;69:40–50.
31. R Core Team. R: A language and environment for statistical computing. In: R Core Team. Vienna: R Foundation for Statistical Computing; 2016.

32. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36(3):1–48.
33. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. Stat Med. 2012;31(29):3821–39.
34. Debray TP. Metamisc: Diagnostic and Prognostic Meta-Analysis; 2017.
35. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48.
36. Andersson C, Enserro D, Larson MG, Xanthakis V, Vasan RS. Implications of the US cholesterol guidelines on eligibility for statin therapy in the community: comparison of observed and predicted risks in the Framingham Heart Study Offspring Cohort. J Am Heart Assoc. 2015;4(4):e001888.
37. Buitrago F, Calvo-Hueros JI, Canon-Barroso L, Pozuelos-Estrada G, Molina-Martinez L, Espigares-Arroyo M, Galan-Gonzalez JA, Lillo-Bravo FJ. Original and REGICOR Framingham functions in a nondiabetic population of a Spanish health care center: a validation study. Ann Fam Med. 2011;9(5):431–8.
38. Chia YC, Lim HM, Ching SM. Validation of the pooled cohort risk score in an Asian population - a retrospective cohort study. BMC Cardiovasc Disord. 2014;14:163.
39. Cook NR, Ridker PM. Further insight into the cardiovascular risk calculator: the roles of statins, revascularizations, and underascertainment in the Women's Health Study. JAMA Intern Med. 2014;174(12):1964–71.
40. Cooper JA, Miller GJ, Humphries SE. A comparison of the PROCAM and Framingham point-scoring systems for estimation of individual risk of coronary heart disease in the Second Northwick Park Heart Study. Atherosclerosis. 2005;181(1):93–100.
41. DeFilippis AP, Young R, McEvoy JW, Michos ED, Sandfort V, Kronmal RA, McClelland RL, Blaha MJ. Risk score overestimation: The impact of individual cardiovascular risk factors and preventive therapies on the performance of the American Heart Association-American College of Cardiology-Atherosclerotic Cardiovascular Disease risk score in a modern multi-ethnic cohort. Eur Heart J. 2017;38(8):598–608.
42. De Las Heras Gala T, Geisel MH, Peters A, Thorand B, Baumert J, Lehmann N, Jockel KH, Moebus S, Erbel R, Mahabadi AA, et al. Recalibration of the ACC/AHA risk score in two population-based German cohorts. PLoS One. 2016;11(10):e0164688.
43. Emdin CA, Khera AV, Natarajan P, Klarin D, Baber U, Mehran R, Rader DJ, Fuster V, Kathiresan S. Evaluation of the pooled cohort equations for prediction of cardiovascular risk in a contemporary prospective cohort. Am J Cardiol. 2017;119(6):881–5.
44. Empana JP, Ducimetiere P, Arveiler D, Ferrieres J, Evans A, Ruidavets JB, Haas B, Yarnell J, Bingham A, Amouyel P, et al. Are the Framingham and PROCAM coronary heart disease risk functions applicable to different European populations? The PRIME Study. Eur Heart J. 2003; 24(21):1903–11.
45. Ferrario M, Chiodini P, Chambless LE, Cesana G, Vanuzzo D, Panico S, Sega R, Pilotto L, Palmieri L, Giampaoli S, et al. Prediction of coronary events in a low incidence population. Assessing accuracy of the CUORE Cohort Study prediction equation. Int J Epidemiol. 2005;34(2):413–21.
46. Jee SH, Jang Y, Oh DJ, Oh BH, Lee SH, Park SW, Seung KB, Mok Y, Jung KJ, Kimm H, et al. A coronary heart disease prediction model: The Korean heart study. BMJ Open. 2014;4(5):e005025.
47. Jung KJ, Jang Y, Oh DJ, Oh BH, Lee SH, Park SW, Seung KB, Kim HK, Yun YD, Choi SH, et al. The ACC/AHA 2013 pooled cohort equations compared to a Korean Risk Prediction Model for atherosclerotic cardiovascular disease. Atherosclerosis. 2015;242(1):367–75.
48. Khalili D, Asgari S, Hadaegh F, Steyerberg EW, Rahimi K, Fahimfar N, Azizi F. A new approach to test validity and clinical usefulness of the 2013 ACC/AHA guideline on statin therapy: A population-based study. Int J Cardiol. 2015;184(1):587–94.
49. Koller MT, Steyerberg EW, Wolbers M, Stijnen T, Bucher HC, MGM H, JCM W. Validity of the Framingham point scores in the elderly: results from the Rotterdam study. Am Heart J. 2007;154(1):87–93.
50. Koller MT, Leening MJG, Wolbers M, Steyerberg EW, Hunink MGM, Schoop R, Hofman A, Bucher HC, Psaty BM, Lloyd-Jones DM, et al. 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(6):389–97.
51. Lee CH, Woo YC, Lam JKY, Fong CHY, Cheung BMY, Lam KSL, Tan KCB. Validation of the pooled cohort equations in a long-term cohort study of Hong Kong Chinese. J Clin Lipidol. 2015;9(5):640–6.
52. Lloyd-Jones DM, Wilson PWF, Larson MG, Beiser A, Leip EP, D'Agostino RB, Levy D. Framingham risk score and prediction of lifetime risk for coronary heart disease. Am J Cardiol. 2004;94(1):20–4.
53. Mainous AG 3rd, Koopman RJ, Diaz VA, Everett CJ, Wilson PWF, Tilley BC. A coronary heart disease risk score based on patient-reported information. Am J Cardiol. 2007;99(9):1236–41.
54. Marrugat J, Subirana I, Comin E, Cabezas C, Vila J, Elosua R, Nam BH, Ramos R, Sala J, Solanas P, et al. Validity of an adaptation of the Framingham cardiovascular risk function: the VERIFICA Study. J Epidemiol Community Health. 2007;61(1):40–7.
55. Mortensen MB, Afzal S, Nordestgaard BG, Falk E. Primary Prevention With Statins: ACC/AHA Risk-Based Approach Versus Trial-Based Approaches to Guide Statin Therapy. J Am Coll Cardiol. 2015;66(24):2699–709.
56. Mortensen MB, Nordestgaard BG, Afzal S, Falk E. ACC/AHA guidelines superior to ESC/EAS guidelines for primary prevention with statins in non-diabetic Europeans: the Copenhagen General Population Study. Eur Heart J. 2017;38(8):586–94.
57. Muntner P, Colantonio LD, Cushman M, Goff DC Jr, Howard G, Howard VJ, Kissela B, Levitan EB, Lloyd-Jones DM, Safford MM. Validation of the atherosclerotic cardiovascular disease pooled cohort risk equations. JAMA. 2014;311(14):1406–15.
58. Pike MM, Decker PA, Larson NB, St Sauver JL, Takahashi PY, Roger VL, Rocca WA, Miller VM, Olson JE, Pathak J, et al. Improvement in cardiovascular risk prediction with electronic health records. J Cardiovasc Transl Res. 2016;9(3):214–22.
59. Rana JS, Tabada GH, Solomon MD, Lo JC, Jaffe MG, Sung SH, Ballantyne CM, Go AS. Accuracy of the atherosclerotic cardiovascular risk equation in a large contemporary, multiethnic population. J Am Coll Cardiol. 2016;67(18):2118–30.
60. Rodondi N, Locatelli I, Aujesky D, Butler J, Vittinghoff E, Simonsick E, Satterfield S, Newman AB, Wilson PWF, Pletcher MJ, et al. Framingham risk score and alternatives for prediction of coronary heart disease in older adults. PLoS One. 2012;7(3):e34287.
61. Ryckman EM, Summers RM, Liu J, Munoz del Rio A, Pickhardt PJ. Visceral fat quantification in asymptomatic adults using abdominal CT: is it predictive of future cardiac events? Abdom Imaging. 2015;40(1):222–6.
62. Simmons RK, Sharp S, Boekholdt SM, Sargeant LA, Khaw K-T, Wareham NJ, Griffin SJ. Evaluation of the Framingham risk score in the European Prospective Investigation of Cancer-Norfolk cohort: does adding glycated hemoglobin improve the prediction of coronary heart disease events? Arch Intern Med. 2008;168(11):1209–16.
63. Simons LA, Simons J, Friedlander Y, McCallum J, Palaniappan L. Risk functions for prediction of cardiovascular disease in elderly Australians: the Dubbo Study. Med J Aust. 2003;178(3):113–6.
64. Suka M, Sugimori H, Yoshida K. Application of the updated Framingham risk score to Japanese men. Hypertens Res. 2001;24(6):685–9.
65. Sussman JB, Wiitala WL, Zawistowski M, Hofer TP, Bentley D, Hayward RA. The veterans affairs cardiac risk score: recalibrating the atherosclerotic cardiovascular disease score for applied use. Med Care. 2017;55(9):864–70.
66. Vaidya D, Yanek LR, Moy TF, Pearson TA, Becker LC, Becker DM. Incidence of coronary artery disease in siblings of patients with premature coronary artery disease: 10 years of follow-up. Am J Cardiol. 2007;100(9):1410–5.
67. Yang X, Li J, Hu D, Chen J, Li Y, Huang J, Liu X, Liu F, Cao J, Shen C, et al. Predicting the 10-year risks of atherosclerotic cardiovascular disease in chinese population: The China-PAR Project (Prediction for ASCVD Risk in China). Circulation. 2016;134(19):1430–40.
68. Eckel RH, Jakicic JM, Ard JD, de Jesus JM, Houston Miller N, Hubbard VS, Lee IM, Lichtenstein AH, Loria CM, Millen BE, et al. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25 Suppl 2):S76–99.
69. Knottnerus JA. Diagnostic prediction rules: principles, requirements and pitfalls. Prim Care. 1995;22(2):341–63.
70. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am J Epidemiol. 2010;172(8):971–80.
71. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol. 2015;68(3):279–89.
72. Paynter NP, Everett BM, Cook NR. Cardiovascular disease risk prediction in women: is there a role for novel biomarkers? Clin Chem. 2014;60(1):88–97.

73. Ridker PM, Cook NR. The Pooled Cohort Equations 3 Years On: Building a Stronger Foundation. Circulation. 2016;134(23):1789–91.
74. Tobert JA. Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors. Nat Rev Drug Discov. 2003;2(7):517–26.
75. Pajouheshnia R, Peelen LM, Moons KGM, Reitsma JB, Groenwold RHH. Accounting for treatment use when validating a prognostic model: a simulation study. BMC Med Res Methodol. 2017;17(1):103.
76. Pajouheshnia R, Damen JA, Groenwold RH, Moons KG, Peelen LM. Treatment use in prognostic model research: a systematic review of cardiovascular prognostic studies. Diagnostic Rognostic Res. 2017;1(1):15.
77. Rospleszcz S, Thorand B, de Las Heras Gala T, Meisinger C, Holle R, Koenig W, Mansmann U, Peters A. Temporal trends in cardiovascular risk factors and performance of the Framingham Risk Score and the pooled cohort equations. J Epidemiol Community Health. 2019;73(1):19–25.
78. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55–63.
79. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1–73.
80. Groenwold RH, Moons KG, Pajouheshnia R, Altman DG, Collins GS, Debray TP, Reitsma JB, Riley RD, Peelen LM. Explicit inclusion of treatment in prognostic modeling was recommended in observational and randomized settings. J Clin Epidemiol. 2016;78:90–100.
81. Muntner P, Safford MM, Cushman M, Howard G. Comment on the reports of over-estimation of ASCVD risk using the 2013 AHA/ACC risk equation. Circulation. 2014;129(2):266–7.
82. Krumholz HM. The new cholesterol and blood pressure guidelines: perspective on the path forward. JAMA. 2014;311(14):1403–5.
83. Goff DC Jr, D'Agostino RB Sr, Pencina M, Lloyd-Jones DM. Calibration and discrimination among multiple cardiovascular risk scores in a modern multiethnic cohort. Ann Intern Med. 2015;163(1):68.
84. Spence JD. Statins and ischemic stroke. JAMA. 2014;312(7):749–50.
85. Cook NR, Ridker PM. Response to Comment on the reports of over-estimation of ASCVD risk using the 2013 AHA/ACC risk equation. Circulation. 2014;129(2):268–9.
86. Thakkar J, Heeley EL, Chalmers J, Chow CK. Inaccurate risk perceptions contribute to treatment gaps in secondary prevention of cardiovascular disease. Intern Med J. 2016;46(3):339–46.
87. Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, Collins GS. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353:i3140.
88. Vergouwe Y, Nieboer D, Oostenbrink R, Debray TP, Murray GD, Kattan MW, Koffijberg H, Moons KG, Steyerberg EW. A closed testing procedure to select an appropriate method for updating prediction models. Stat Med. 2017;36(28):4529–39.
89. Grundy SM, D'Agostino RB Sr, Mosca L, Burke GL, Wilson PW, Rader DJ, Cleeman JI, Roccella EJ, Cutler JA, Friedman LM. Cardiovascular risk assessment based on US cohort studies: findings from a National Heart, Lung, and Blood institute workshop. Circulation. 2001;104(4):491–6.
90. Hajifathalian K, Ueda P, Lu Y, Woodward M, Ahmadvand A, Aguilar-Salinas CA, Azizi F, Cifkova R, Di Cesare M, Eriksen L, et al. A novel risk score to predict cardiovascular disease risk in national populations (Globorisk): a pooled analysis of prospective cohorts and health examination surveys. Lancet Diabetes Endocrinol. 2015;3(5):339–55.
91. van Dis I, Kromhout D, Geleijnse JM, Boer JM, Verschuren WM. Evaluation of cardiovascular risk predicted by different SCORE equations: the Netherlands as an example. Eur J Cardiovasc Prev Rehabil. 2010;17(2):244–9.
92. De Bacquer D, De Backer G. Predictive ability of the SCORE Belgium risk chart for cardiovascular mortality. Int J Cardiol. 2010;143(3):385–90.
93. Sans S, Fitzgerald AP, Royo D, Conroy R, Graham I. Calibrating the SCORE cardiovascular risk chart for use in Spain. Rev Esp Cardiol. 2007;60(5):476–85.
94. Usher-Smith JA, Silarova B, Schuit E, Gm Moons K, Griffin SJ. Impact of provision of cardiovascular disease risk estimates to healthcare professionals and patients: a systematic review. BMJ Open. 2015;5(10):e008717.
95. Studzinski K, Tomasik T, Krzyszton J, Jozwiak J, Windak A. Effect of using cardiovascular risk scoring in routine risk assessment in primary prevention of cardiovascular disease: an overview of systematic reviews. BMC Cardiovasc Disord. 2019;19(1):11.
96. Taylor F, Huffman MD, Macedo AF, Moore TH, Burke M, Smith GD, Ward K, Ebrahim S. Statins for the primary prevention of cardiovascular disease. Cochrane Database Syst Rev 2013(1):Cd004816.
97. Lackland DT, Elkind MS, D'Agostino R Sr, Dhamoon MS, Goff DC Jr, Higashida RT, McClure LA, Mitchell PH, Sacco RL, Sila CA, et al. Inclusion of stroke in cardiovascular risk prediction instruments: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2012;43(7):1998–2027.
98. Damen JAAG, Debray TPA, Pajouheshnia R, Reitsma JB, Scholten RJPM, Moons KGM, Hooft L. Empirical evidence of the impact of study characteristics on the performance of prediction models: a meta-epidemiological study. BMJ Open. 2019;9(4):e026160.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC