Prediction models for cardiovascular disease risk in the general population: systematic review

Johanna A A G Damen,[1, 2] Lotty Hooft, [1, 2] Ewoud Schuit,[1, 2, 3] Thomas P A Debray,[1, 2] Gary S Collins,[4] Ioanna Tzoulaki,[5] Camille M Lassale,[5] George C M Siontis,[6] Virginia Chiocchia,[4 ,7] Corran Roberts,[4] Michael Maia Schlüssel,[4] Stephen Gerry,[4] James A Black,[8] Pauline Heus,[1, 2] Yvonne T van der Schouw,[1] Linda M Peelen,[1] Karel G M Moons[1, 2]

[1]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands

[2]Cochrane Netherlands, University Medical Center Utrecht, PO Box 85500, Str 6.131, 3508 GA Utrecht, Netherlands

[3]Stanford Prevention Research Center, Stanford University, Stanford, CA, USA

[4]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

[5]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

[6]Department of Cardiology, Bern University Hospital, 3010 Bern, Switzerland

[7]Surgical Intervention Trials Unit, University of Oxford, Oxford, UK

[8]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK

Correspondence to: J A A G Damen j.a.a.damen@umcutrecht.nl

ABSTRACT

OBJECTIVE
To provide an overview of prediction models for risk of cardiovascular disease (CVD) in the general population.

DESIGN
Systematic review.

DATA SOURCES
Medline and Embase until June 2013.

ELIGIBILITY CRITERIA FOR STUDY SELECTION
Studies describing the development or external validation of a multivariable model for predicting CVD risk in the general population.

RESULTS
9965 references were screened, of which 212 articles were included in the review, describing the development of 363 prediction models and 473 external validations. Most models were developed in Europe (n=167, 46%), predicted risk of fatal or non-fatal coronary heart disease (n=118, 33%) over a 10 year period (n=209, 58%). The most common predictors were smoking (n=325, 90%) and age (n=321, 88%), and most models were sex specific (n=250, 69%). Substantial heterogeneity in predictor and outcome definitions was observed between models, and important clinical and methodological information were often missing. The prediction horizon was not specified for 49 models (13%), and for 92 (25%) crucial information was missing to enable the model to be used for individual risk prediction. Only 132 developed models (36%) were externally validated and only 70 (19%) by independent investigators. Model performance was heterogeneous and measures such as discrimination and calibration were reported for only 65% and 58% of the external validations, respectively.

CONCLUSIONS
There is an excess of models predicting incident CVD in the general population. The usefulness of most of the models remains unclear owing to methodological shortcomings, incomplete presentation, and lack of external validation and model impact studies. Rather than developing yet another similar CVD risk prediction model, in this era of large datasets, future research should focus on externally validating and comparing head-to-head promising CVD risk models that already exist, on tailoring or even combining these models to local settings, and investigating whether these models can be extended by addition of new predictors.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Several well known prediction models estimate the risk of developing cardiovascular disease (CVD) in the general population

Such models include the Framingham risk score, SCORE, and QRISK

No comprehensive overview has described all competitive models in this domain, how these models have been developed, how many were externally validated, and their predictive performance

WHAT THIS STUDY ADDS

Although there is an over-abundance of CVD risk prediction models for the general population, few have been externally validated, making them currently of unknown value for practitioners, policy makers, and guideline developers

Most developed models are inadequately reported to allow external validation or implementation in clinical practice

Rather than developing new models, researchers should make better use of available evidence by validating, making head-to-head comparisons, and tailoring the promising existing models

Introduction

Cardiovascular disease (CVD) is a leading cause of morbidity and mortality worldwide,[1] accounting for approximately one third of all deaths.[2] Prevention of CVD requires timely identification of people at increased risk to target effective dietary, lifestyle, or drug interventions. Over the past two decades, numerous prediction models have been developed, which mathematically combine multiple predictors to estimate the risk of developing CVD—for example, the Framingham,[3-5] SCORE,[6] and QRISK[7-9] models. Some of these prediction models are included in clinical guidelines for therapeutic management[10 11] and are increasingly advocated by health policymakers. In the United Kingdom, electronic health patient record systems now have QRISK2 embedded to calculate 10 year CVD risk.

Several reviews have shown that there is an abundance of prediction models for a wide range of CVD outcomes.[12-14] However, the most comprehensive review[12] includes models published more than 10 years ago (search carried out in 2003). More recent reviews have shown that the number of published prediction models has increased dramatically since then; furthermore, these reviews have not systematically described the outcomes that the models intended to predict, the most common predictors, the predictive performance of all these models, and which developed prediction models have been externally validated.[13 14]

We carried out a systematic review of multivariable prediction models developed to predict the risk of developing CVD in the general population, to describe the characteristics of the models' development, included predictors, CVD outcomes predicted, presentation, and whether they have undergone external validation.

Methods

We conducted our systematic review following the recently published guidance from the Cochrane Prognosis Methods Group, using the CHARMS checklist, for reviews of prediction model studies.[15]

Literature search

We performed a literature search in Medline and Embase on 1 June 2013 using search terms to identify primary articles reporting on the development and/or validation of models predicting incident CVD, published from 2004 onwards (see supplementary table 1). Articles published before 2004 were identified from a previously published comprehensive systematic review,[12] and a cross reference check was performed for all reviews on CVD prediction models identified by our search. For external validation studies where the development study was not identified by our search, we manually retrieved and included in the review the original article describing the development of the model.

Eligibility criteria

We included all primary articles that reported on one or more multivariable (that is, including at least two predictors[16]) prediction models, tools, or scores, that have been proposed for individual risk estimation of any future CVD outcome in the general population. We differentiated between articles reporting on the development[17-19] or external validation[19-21] of one or more prediction models (box 1). Studies reporting on the incremental value or model extension—that is, evaluating the incremental value of one or more new predictors to existing models,[26] were excluded. We classified articles as development studies if they reported the development of a model in their objectives or conclusions, or if it was clear from other information in the article that

they developed a prediction model for individual risk estimation (eg, if they presented a simplified risk chart). Included articles had to report original research (eg, reviews and letters were excluded), study humans, and be written in English. Articles were included if they reported models for predicting any fatal or non-fatal arterial CVD event. We excluded articles describing models for predicting the risk of venous disease; validation articles with a cross sectional study design that, for example, compared predicted risks of two different models at one time point without any association with actual CVD outcomes; and articles describing models developed from or validated exclusively in specific diseased (patient) populations, such as patients with diabetes, with HIV, with atrial fibrillation, or undergoing any surgery. Furthermore, we excluded methodological articles and articles for which no full text was available through a license at our institutes. Impact studies identified by our search were excluded from this review but were described in a different review.[27] External validation articles were excluded if the corresponding development article was not available.

A single article can describe the development and/or validation of several prediction models, and the distinction between models is not always clear. We defined reported models as separate models whenever a combination of two or more predictors with unique predictor-outcome association estimates were presented. For example, if a model was fitted after stratification for men and women yielding different predictor-outcome associations (that is, predictor weights), we scored it as two separate models. Additionally, two presented models yielding the same predictor-outcome associations but with a different baseline hazard or risk estimate, were considered separately.

Screening process

Initially pairs of two reviewers (JAB, TPAD, CML, LMP, ES, GCMS) independently screened retrieved articles for eligibility on title and subsequently on abstract. Disagreements were resolved by iterative screening rounds. After consensus, full text articles were retrieved and one reviewer (JAB, GSC, VC, JAAGD, SG, TPAD, PH, LH, CML, CR, ES, GCMS, MMS, IT) screened the full text articles and extracted data. In case of doubt, a second (JAAGD or GSC) or third (ES or KGMM) reviewer was involved.

Data extraction and critical appraisal

We categorised the eligible articles into two groups: development articles, and external validation (with or without model recalibration) articles.

The list of extracted items was based on the recently issued Cochrane guidance for data extraction and critical appraisal for systematic reviews of prediction models (the CHARMS checklist[15]) supplemented by items obtained from methodological guidance papers and previous systematic reviews in the specialty.[15 28 29-31] The full list of extracted items is available on request. Items extracted from articles describing model development included study design (eg, cohort, case-control), study population, geographical location, outcome, prediction

---

Box 1: Definitions of technical terms

Internal validation—testing a model's predictive accuracy by reusing (parts of) the dataset on which the model was developed. The aim of internal validation is to assess the overfit and correct for the resulting "optimism" in the performance of the model. Examples are cross validation and bootstrapping[22]

External validation—testing a model's predictive accuracy in a population other than the development population[23]

Prediction horizon—time frame for which the model is intended to predict the outcome[15]

Discrimination—ability of the model to distinguish between people who do and do not develop the outcome of interest[24]

Calibration—agreement between predicted and observed numbers of events[22]

Updating—adjusting a previously developed model to a new setting or study population, to improve model fit in that population. Several forms of updating exist, including intercept recalibration, slope recalibration, and refitting all coefficients of a model.[25] It is also possible to combine and update existing models
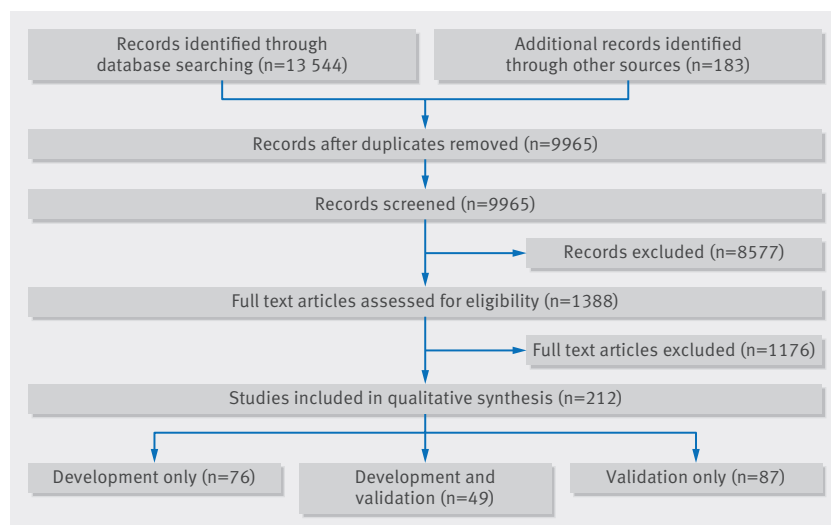
Fig 1 | Flow diagram of selected articles

Descriptive analyses
Results were summarised using descriptive statistics. We did not perform a quantitative synthesis of the models, as this was beyond the scope of our review, and formal methods for meta-analysis of prediction models are not yet fully developed.

Patient involvement
No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Results
The search strategy identified 9965 unique articles, of which 8577 were excluded based on title and abstract. In total, 1388 full texts were screened, of which 212 articles met the eligibility criteria and were included in this review (fig 1). In total, 125 articles concerned the development of one or more CVD risk prediction models and 136 articles described the external validation of one or more of these models (see supplementary table 2). Frequently, articles described combinations of development or external validation (fig 1), therefore the total number does not sum up to 212. The number of development and external validation studies increased over time (fig 2).

Studies describing the development of CVD prediction models
Study designs and study populations
Overall, 125 articles described the development of 363 different models. Most of the prediction models (n=250, 69%) were developed using data from a longitudinal cohort study (see supplementary figure 1A); most originated from Europe (n=168, 46%) or the United States and Canada (n=132, 36%, see supplementary figure 1B). No models were developed using data from Africa. Several cohorts were used multiple times for model development—for example, the Framingham cohort, yielding 69 models in 23 papers.

horizon, modelling method (eg, Cox proportional hazards model, logistic model), method of internal validation (eg, bootstrapping, cross validation), number of study participants and CVD events, number and type of predictors, model presentation (eg, full regression equation, risk chart), and predictive performance measures (eg, calibration, discrimination). For articles describing external validation of a prediction model we extracted the type of external validation (eg, temporal, geographical[21 32]), whether or not the validation was performed by the same investigators who developed the model, study population, geographical location, number of participants and events, and the model's performance before and (if conducted) after model recalibration. If an article described multiple models, we carried out separate data extraction for each model.

To accomplish consistent data extraction, a standardised data extraction form was piloted and modified several times. All reviewers were extensively trained on how to use the form. A second reviewer (JAAGD) checked extracted items classed as "not reported" or "unclear," or unexpected findings. We did not explicitly perform a formal risk of bias assessment as no such tool is currently available for studies of prediction models.
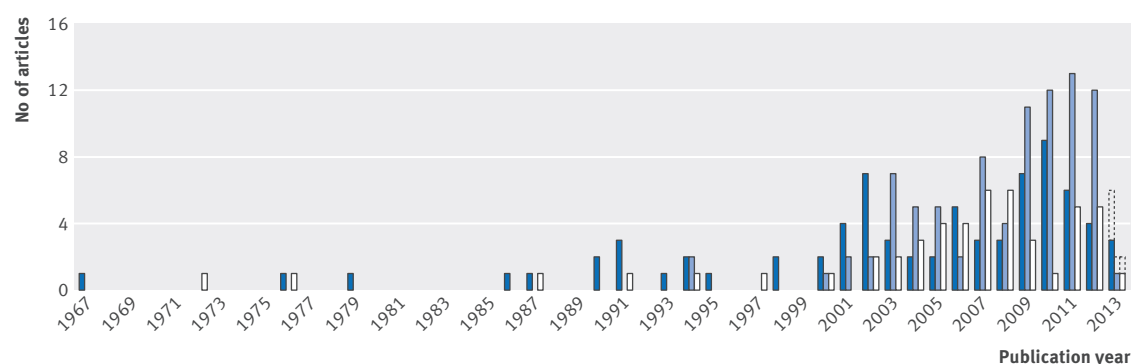

Fig 2 | Numbers of articles in which only one or more models were developed (dark blue), only one or more models were externally validated (light blue), or one or more models were developed and externally validated (white), ordered by publication year (up to June 2013). Predictions of the total numbers in 2013 are displayed with dotted lines

Study populations (that is, case mix) differed noticeably between studies, mainly for age, sex, and other patient characteristics. Most models were developed for people with ages ranging from 30 to 74 years (n=206, 57%), although 69 different age ranges were reported (see supplementary figure 1C). The majority of models was sex specific (men n=142, 39%; women n=108, 30%), and for most models (n=230, 63%), investigators explicitly stated they excluded study participants with existing CVD (including coronary heart disease, stroke, other heart diseases, or combinations of those), or with other diseases such as cancer (n=21, 6%) or diabetes (n=43, 12%).

CVD outcomes

We observed large variation in predicted outcomes. Although the majority of prediction models focused on (fatal or non-fatal) coronary heart disease or CVD (n=118, 33% and n=95, 26%), 19 other outcomes were identified, such as (fatal or non-fatal) stroke, myocardial infarction, and atrial fibrillation (see supplementary table 3). On top of this, the definitions of these outcomes showed considerable heterogeneity, with, for example, more than 40 different definitions for fatal or non-fatal coronary heart disease (see supplementary table 4). International classification of disease codes were specified for 82 out of 363 models (23%).

Predictors

The median number of predictors included in the developed models was 7 (range 2-80). In total, more than 100 different predictors were included (fig 3). Sex was included in 88 (24%) models; however, 250 (69%) models were explicitly developed only for men or only for women. Most of the models (n=239, 66%) included a set of similar predictors, consisting of age, smoking, blood pressure, and blood cholesterol measurements. Other prevalently selected predictors were diabetes (n=187, 52%) and body mass index (n=107, 29%). Treatment

modalities were included in a few prediction models; 56 models (15%) included use of antihypertensive treatment and no models included use of lipid lowering drugs.

Sample size

The number of participants used to develop the prediction models ranged from 51 to 1 189 845 (median 3969), and the number of events ranged between 28 and 55 667 (median 241). The number of participants and the number of events were not reported for 24 (7%) and 74 (20%) models, respectively. The number of events for each variable included in the final prediction model could be calculated for 252 (69%) models and ranged from 1 to 4205. For 25 out of these 252 (10%) models, this number of events for each variable was less than 10.[33 34]

Modelling method and prediction horizon

We found that most prediction models were developed using Cox proportional hazards regression (n=160, 44%), accelerated failure time analysis (n=77, 21%), or logistic regression (n=71, 20%). For 36 models (10%) the method used for statistical modelling was not clear (see supplementary table 5). The prediction horizon ranged between 2 and 45 years, with the majority of studies predicting CVD outcomes for a five year or 10 year horizon (n=47, 13% and n=209, 58%, respectively). For 49 models (13%), the prediction horizon was not specified (see supplementary table 6).

Model presentation

For 167 models (46%) the complete regression formula, including all regression coefficients and intercept or baseline hazard, were reported. Of the other 196 models, 104 (53%) were presented as online calculator, risk chart, sum score, or nomogram to allow individual risk estimation. For the remaining models (n=92, 25%) insufficient information was presented to allow calculation of individual risks.
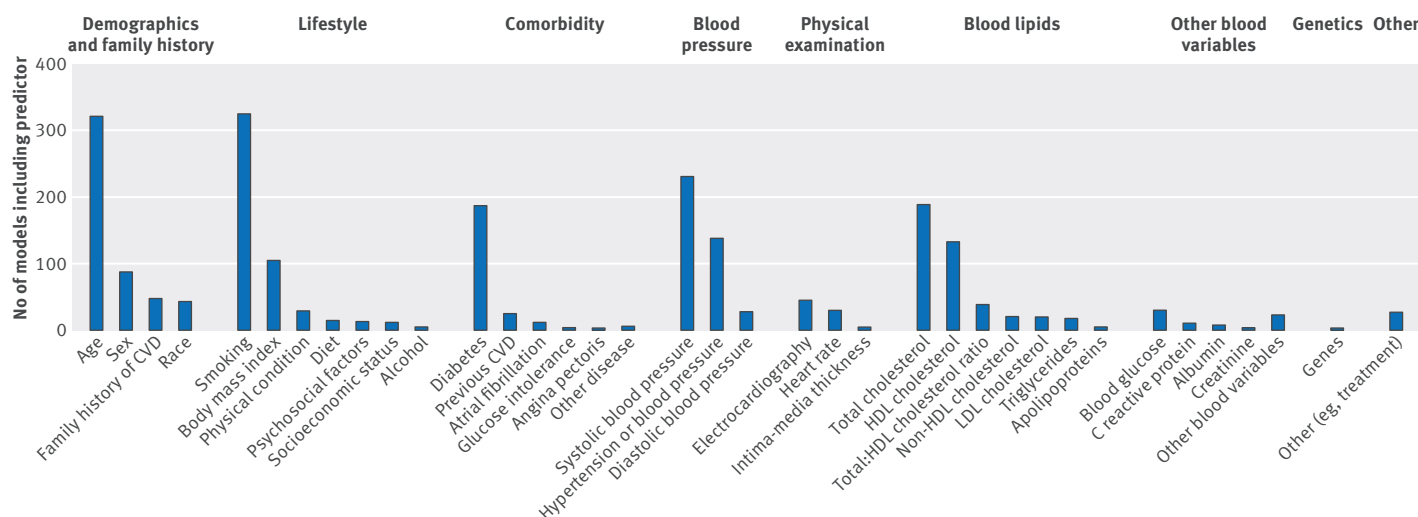

Fig 3 | Main categories of predictors included in developed models. CVD=cardiovascular disease; HDL=high density lipoprotein; LDL=low density lipoprotein

Predictive performance

At least one measure of predictive performance was reported for 191 of the 363 (53%) models (table 1). For 143 (39%) models, discrimination was reported as a C statistic or area under the receiver operating characteristic curve (range 0.61 to 1.00). Calibration was reported for 116 (32%) models, for which a variety of methods was used, such as a Hosmer-Lemeshow test (n=60, 17%), calibration plot (n=31, 9%) or observed:-expected ratio (n=12, 3%). For 99 (27%) models, both discrimination and calibration were reported. Table 2 shows that reporting of discriminative performance measures seems to have increased over time, whereas reporting of calibration seems to remain limited.

Internal validation

In total, 80 of the 363 developed models (22%) were internally validated, most often using a random split of the dataset (n=27), bootstrapping (n=23), or cross validation (n=22).

Studies describing external validation of a prediction model

In 136 articles, 473 external validations were performed. However, the majority of the 363 developed models (n=231, 64%) has never been externally validated. Out of the 132 (36%) models that were externally validated, 35 (27%) were validated once, and 38 (29%) (originally developed and described in seven articles) were validated more than 10 times. The most commonly validated models were Framingham (Wilson 1998, n=89),[5] Framingham (Anderson 1991, n=73),[3] SCORE (Conroy 2003, n=63),[6] Framingham (D'Agostino 2008, n=44),[36] Framingham (ATP III 2002, n=31),[37] Framingham (Anderson 1991, n=30),[4] and QRISK (Hippisley-Cox 2007, n=12)[8] (table 3).

Out of the 132 externally validated models, 45 (34%) were solely externally validated in the same paper in which their development was described, 17 (13%) were externally validated in a different paper but with authors overlapping between the development and validation paper, and 70 (53%) were validated by independent researchers. Sample sizes of the validation studies ranged from very small (eg, 90 participants or one event) to very large (eg, 1 066 127 participants or 51 340 events). Most external validations were performed in a different geographical area from the development study—for example, the Framingham (Anderson 1991)[3] model (developed on data from the United States) was often validated outside North America, namely in Europe (71% of its validations), Australia (16%), or Asia (4%) (table 4). There was considerable heterogeneity in eligibility criteria for patients between validation and development studies. For example, for the seven aforementioned models, 13% of the validation studies were performed in the same age range for which the model was originally developed. For Framingham (Anderson 1991)[3] only few (n=12, 16%) validations were performed in people outside these age ranges, whereas for Framingham (Wilson 1998)[5] and SCORE (Conroy 2003)[6] this happened more often (n=34, 38% and n=33, 52%, respectively; see supplementary figure 2).

In external validation studies, the C statistic was reported for 303 (64%) models. For 277 models (58%) a calibration measure was reported by using a calibration plot (n=122, 26%), an observed:expected ratio (n=124, 26%), the Hosmer-Lemeshow test (n=68, 14%), a calibration table (that is, a table with predicted and observed events; n=62, 13%), or a combination of those (table 1). Both discrimination and calibration were reported for 185 (39%) external validations. The discriminative ability and calibration of the three most often validated models (Framingham (Wilson 1998),[5] Framingham (Anderson 1991),[3] and SCORE (Conroy 2003[6])) varied between validation studies, with C statistics between 0.57 and 0.92, 0.53 and 0.99, and 0.62

Table 1 | Performance measures reported for developed models. Values are numbers (percentages) unless stated otherwise

Performance measures	Development	Validation
Discrimination measures:		
C statistic/AUC	143 (39)	303 (64)
D statistic	5 (1)	45 (9)
Other*	24 (7)	8 (2)
Any	163 (45)	306 (65)
Calibration measures:		
Plot	31 (9)	122 (26)
Table	34 (9)	62 (13)
Slope	3 (1)	7 (1)
Intercept	2 (1)	7 (1)
Hosmer Lemeshow test	60 (17)	68 (14)
Observed:expected ratio	12 (3)	124 (26)
Other†	7 (2)	20 (4)
Any	116 (32)	277 (58)
Overall performance measures:		
R^2	13 (4)	49 (10)
Brier score	15 (4)	45 (9)
Other‡	10 (3)	1 (<0.5)
Any	35 (10)	68 (14)
Any performance measure	191 (53)	398 (84)
Total	363	474

AUC=area under receiver operating characteristic curve.
Numbers add up to over 363 since papers may have reported more than one predictive performance measure.
*For example, sensitivity, specificity.
†For example, Grønnesby-Borgan χ^2 test.
‡For example, Akaike information criterion, bayesian information criterion.

Table 2 | Reporting of performance measures for models across years of publication. Values are numbers (percentages) unless stated otherwise

Performance measures	Publication year			
	1967-2001	2002-05	2006-08	2009-13
Development:				
Discrimination	12 (14)	46 (55)	41 (44)	64 (64)
Calibration	13 (15)	41 (49)	25 (27)	37 (37)
Overall performance*	0 (0)	2 (2)	12 (13)	21 (21)
Any performance	25 (29)	48 (58)	42 (45)	76 (76)
Total	87	83	93	100
Validation:				
Discrimination	12 (32)	41 (44)	71 (68)	182 (77)
Calibration	29 (76)	45 (48)	64 (61)	139 (59)
Overall performance	0 (0)	0 (0)	22 (21)	46 (19)
Any performance	31 (82)	56 (60)	98 (93)	213 (90)
Total	38	93	105	237

*Performance measures giving overall indication of goodness of fit of a model, such as R^2 and brier score.[35]

Table 3 | List of the models that were validated at least three times, and their predicted outcomes (sorted by number of validations)

Reference (No of developed models)	Predicted outcomes	No of validations
Framingham Wilson 1998[5] (n=2*)	Fatal or non-fatal CHD	89
Framingham Anderson 1991[3] (n=12)	Fatal or non-fatal: CHD, CVD, myocardial infarction, and stroke	73
SCORE Conroy 2003[6] (n=12)	Fatal: CHD, CVD, and non-CHD	63
Framingham D'Agostino 2008[36] (n=4)	Fatal CVD	44
Framingham ATP III 2002[37] (n=2)	Fatal or non-fatal CHD	31
Framingham Anderson 1991[4] (n=4)	Fatal or non-fatal CHD	30
QRISK Hippisley-Cox 2007[8] (n=2)	Fatal CVD	12
PROCAM Assman 2002[38] (n=1)	Fatal or non-fatal CHD	8
Framingham Wolf 1991[39] (n=2)	Fatal or non-fatal stroke	8
Chambless 2003[40] (n=4)	Fatal or non-fatal CHD	7
Friedland 2009[41] (n=7)	Fatal or non-fatal: CHD, myocardial infarction, and stroke; claudication; coronary artery bypass grafting; percutaneous transluminal coronary angioplasty; transient ischaemic attack	6
QRISK Hippisley-Cox 2010[7] (n=2)	Fatal CVD	6
Keys 1972[42] (n=4)	Fatal or non-fatal CHD	6
Leaverton 1987[43] (n=4)	Fatal CHD	6
Asia Pacific cohort studies 2007[44] (n=4)	Fatal CVD	4
Woodward 2007[45] (n=2)	Fatal CVD	4
Levy 1990[46] (n=4)	Fatal or non-fatal CHD	4
Chien 2012[47] (n=3)	Fatal or non-fatal CHD	3
Framingham unspecified†	—	32

CHD=coronary heart disease; CVD=cardiovascular disease.
*Number of models developed in this article.
†Authors stated they externally validated the Framingham model without referencing the specific model.

and 0.91, respectively, and observed:expected ratios between 0.37 and 1.92, 0.18 and 2.60, and 0.28 and 1.50, respectively (table 4).

Models that were external validated differed in many respects from the non-validated models (see supplementary table 7). Ninety three per cent of validated models were developed using longitudinal cohort data versus 81% of non-validated models, 34% versus 15% were internally validated, and 83% versus 70% were presented in a way that allowed the calculation of individual risk. The median publication year for validated models was 2002 (or 2003 after excluding the earliest Framingham models) versus 2006 for models that were not validated. In addition, validated models were developed in studies with a median of 364 events versus 181 for non-validated models. More than half (75 out of 132, 57%) of the models developed in the United States or Canada were validated, compared with 24% (40 out of 168) of models developed from Europe and 16% (7 out of 43) from Asia; excluding the Framingham prediction models did not influence these percentages. None of the models developed in Asia was validated by independent researchers, whereas 41 out of 132 (31%) models from the United States and 26 out of 168 (15%) from Europe were validated by independent researchers.

Discussion

This review shows that there is an abundance of cardiovascular risk prediction models for the general population. Previous reviews also indicated this but were conducted more than a decade ago,[12] excluded models that were not internally or externally validated,[13] or excluded articles that solely described external validation.[14]

Clearly, the array of studies describing the development of new risk prediction models for cardiovascular disease (CVD) in the general population is overwhelming, whereas there is a paucity of external validation studies for most of these developed models. Notwithstanding a few notable exceptions, including the Framingham and SCORE models, most of the models (n=231, 64%) have not been externally validated, only 70 (19%) have been validated by independent investigators, and only 38 (10%)—from only seven articles—were validated more than 10 times.

Healthcare professionals and policymakers are already in great doubt about which CVD prediction model to use or advocate in their specific setting or population. Instead of spending large amounts of research funding on the development of new models, in this era of large datasets, studies need to be aimed at validating the existing models and preferably using head-to-head comparisons of their relative predictive performance, be aimed at tailoring these models to local settings or populations, and focus on improving the predictive performance of existing models by the addition of new predictors.[48]

We found much variability in geographical location of both model development and model validation, but the majority of models were developed and validated in European and Northern American populations. Although the World Health Organization states that more than three quarters of all CVD deaths occur in low income and middle income countries,[49] a prediction model for people from Africa or South America has only recently been developed.[50] Several prediction models have been developed using data from Asia (eg,[44 51 52]) but none has yet been externally validated by independent researchers. Models tailored to these countries are important, as it is known that predictor-outcome associations vary among ethnic groups.[53]

With respect to outcome definitions, most models aimed to predict the risk of fatal or non-fatal coronary heart disease or the combined outcome of CVD. But we identified over 70 different definitions for these two outcomes. In addition, most outcomes were not fully defined and ICD codes were presented for only a few of the predicted outcomes. Without direct head-to-head comparison studies, these differences make it difficult to compare and choose between the existing prediction models based on our review, let alone to decide on which model to choose or advocate in a particular setting. Different definitions of CVD outcome lead to different estimated predictor effects, thus to different predicted probabilities and model performances, and consequently indicate different treatment strategies based on these prediction models. A more uniform definition and reporting of the predicted outcomes, preferably by explicit reporting of the ICD-9 or ICD-10 codes for each outcome, would help the comparison of developed risk models, and their recommendation for and translation into clinical practice. Providing clear

Table 4 | Description of study populations and design characteristics used to validate seven most often (>10 times, see table 3) validated models. Values are numbers (percentages) unless stated otherwise

Characteristics	Framingham Wilson 1998[5] (n=89)†	Anderson 1991[3] (n=73)	SCORE: Conroy 2003[6] (n=63)	Framingham D'Agostino 2008[36] (n=44)	ATP III 2002[37] (n=31)	Anderson 1991[4] (n=30)	QRISK: Hippisley-Cox 2007[8] (n=12)
Location:							
Asia	9 (10)	3 (4)	2 (3)	8 (18)	2 (6)	2 (7)	0 (0)
Australia	0 (0)	12 (16)	4 (6)	2 (5)	1 (3)	2 (7)	0 (0)
Europe	34 (38)	52 (71)	47 (75)	20 (45)	6 (19)	18 (60)	12 (100)
North America	46 (52)	6 (8)	10 (16)	14 (32)	22 (71)	8 (27)	0 (0)
Age:							
Same age range as development study*	2 (3)	21 (29)	4 (6)	5 (11)	0 (0)	0 (0)	12 (100)
Young people (<50 years)	3 (3)	6 (8)	4 (6)	3 (7)	3 (10)	1 (3)	0 (0)
Older people (>60 years)	5 (6)	7 (10)	4 (6)	3 (7)	10 (32)	0 (0)	0 (0)
Other	79 (89)	39 (53)	51 (81)	33 (25)	18 (58)	29 (97)	0 (0)
Sex:							
Men	38 (43)	30 (41)	23 (37)	11 (25)	10 (32)	16 (53)	6 (50)
Women	29 (33)	25 (34)	23 (37)	11 (25)	10 (32)	13 (43)	6 (50)
Men and women	22 (25)	18 (25)	17 (27)	22 (50)	11 (35)	1 (3)	0 (0)
Median (range) No of participants	2716 (100-163627), n=87	2423 (262-797373), n=71	8025 (262-44649), n=63	2661 (272-542987), n=44	3029 (534-36517), n=31	3573 (331-542783), n=30	536,400 (301,622-797373), n=12
Median (range) No of events	146 (8-24659), n=65	128 (1-42408), n=59	224 (16-1722), n=54	164 (15-26202), n=35	415 (35-2343), n=29	188 (4-26202), n=28	29057 (18027-42408), n=6
Median (range) C statistic	0.71 (0.57-0.92), n=61	0.75 (0.53-0.99), n=46	0.75 (0.62-0.91), n=28	0.77 (0.58-0.84), n=28	0.66 (0.60-0.84), n=21	0.75 (0.63-0.78), n=6	0.79 (0.76-0.81), n=12
Median (range) observed:expected	0.59 (0.37-1.92), n=14	0.68 (0.18-2.60), n=42	0.68 (0.28-1.50), n=26	0.80 (0.62-0.96), n=3	0.47 (0.47-0.47), n=1	0.71 (0.32-3.92), n=14	0.94 (0.87-1.00), n=4

*30-74 (Framingham Wilson 1998[5], Anderson 1991,[3,4] D'Agostino 2008,[36] ATP III 2002[37]), 40-65 (SCORE Conroy 2003[6]), 35-74 (QRISK Hippisley-Cox 2007[8]).
†Number of times model was externally validated.
‡Number of models for which this information was reported.

outcome definitions enhances not only the reporting of the development studies but also the conduct of external validation of developed models and, most importantly, the clinical implementation of the models by others.[30]

Most models (66%) were based on a common set of predictors, consisting of age, smoking, blood pressure, and cholesterol levels. Additional to this set, a large number (>100) of predictors have been included in models only once or twice. Interestingly, all these extended models have rarely been externally validated. This suggests that there is more emphasis placed on repeating the process of identifying predictors and developing new models rather than validating, tailoring, and improving existing CVD risk prediction models.

Strengths and limitations of this study

The major strengths of this review include the comprehensive search, careful selection of studies, and extensive data extraction on key characteristics of CVD risk prediction models, including the predictors, outcomes, and studied populations. However, this review also has some limitations. Firstly, we performed our search almost three years ago, and since then more than 4000 articles have been published that matched our search strategy. Therefore, some newly developed prediction models, such as the Pooled Cohort Equations[10] and GLOBORISK,[50] are not included in this overview. However, considering the large number of included models, including these articles is unlikely to change our main conclusions and recommendations. Moreover, it is this large number of newly identified articles in only two years, that actually underlines our main conclusions and reaffirms the necessity for changes regarding CVD risk prediction and a shift in focus from model development to model validation, head-to-head comparison, model improvement, and assessment of modelling impact. Secondly, we excluded articles not written in English (n=65) and for which no full text was available (n=124). This may have led to some underestimation of the number of models and external validations in the search period, and it might have affected the geographical representation. Thirdly, for external validations of a model published in an article in which several models were developed, it was often not stated exactly which of these models was validated. We therefore assumed all developed models in such articles as validated, which could even have resulted in an overestimation of the number of validated models.

Comparison with other studies

As with previous reviews in other specialties,[29 54 55] we found that important clinical and methodological information needed for validation and use of a developed model by others, was often missing. Incomplete reporting is highlighted as an important source of research waste, especially because it prevents future studies from summarising or properly building on

previous work, and guiding clinical management.[56] We have already dealt with the poor reporting of predicted outcome definitions and measurement. Although we observed an improvement in the reporting of discriminative performance measures over time, for 10% of the developed models, the modelling method was not described, for 13% the time horizon (eg, 10 years) for which the model was predicting was not described, and for 25% information for calculating individual CVD risks (eg, full regression equation, nomogram, or risk chart) was insufficient, making it impossible to validate these models or apply them in clinical practice. For external validation of a model, the full regression equation is needed, which was presented for only 46% of the developed models. To improve the reporting of prediction model studies, the TRIPOD statement was recently published (www.tripod-statement.org).[30 57]

Since the publication of the review by Beswick et al[12] in 2008, in which they searched the literature until 2003, several major things have changed. The number of developed prediction models has more than tripled, from 110 to 363, revealing problems such as the overwhelming number of prediction models, predictor definitions, outcome definitions, prediction horizons, and study populations, and showing how poorly researchers make use of available evidence or existing models in the discipline. Although Beswick et al stated that "New prediction models should have multiple external validations in diverse populations with differing age ranges, ethnicity, sex and cardiovascular risk",[12] we still found a great lack of validation studies for most developed CVD risk prediction models.

Presumably there are various reasons why researchers continue to develop a new CVD risk prediction model from scratch, such as the perceived lack of prediction models for their specific population (eg, ethnic minority groups) or specific outcomes (eg, ischaemic stroke), newly identified predictors, published articles reporting on bad performance of existing models in another setting, availability of data with higher quality (eg, greater sample size, prospectively collected data), funding priorities, or merely self-serving to generate another publication. Nevertheless, our review clearly indicates that many of these studies are still similar in design and execution, as corresponding models often include the same (or similar) predictors, target the same (or similar) patient populations, and predict the same (or similar) outcomes. Therefore, researchers are often—perhaps without knowing—repeating the same process and mostly introduce implicit knowledge when developing a prediction model from scratch. Given that there is a huge amount of literature on prediction of CVD outcomes for the general population, we think it is time to capitalise on prediction modelling research from scratch in this specialty. Over the past few decades, statistical methods for building prediction models using established knowledge have substantially improved, and these can be achieved by refining, updating, extending, and even combining the most promising existing models for prediction of CVD in the general population.

Recommendations and policy implications

Ideally, systematic reviews also guide evidence informed health decision making, in this case leading to recommendations on which models to advocate or even use in different settings or countries. Given the lack of external validation studies (notably by independent investigators) of the majority of CVD risk prediction models, the even bigger lack of head-to-head comparisons of these models (even of the well known CVD risk prediction models such as Framingham, SCORE, and QRISK), the poor reporting of most developed models, and the large variability in studied populations, predicted outcomes, time horizons, included predictors, and reported performance measures, we believe it is still impossible to recommend which specific model or models should be used in which setting or location. Guided by this review, we will continue to focus on quantitatively summarising the predictive performance of the identified CVD risk prediction models that were externally validated across various different locations, and ideally of models that were validated head-to-head and compared in the same dataset. Such meta-analysis of CVD risk prediction models should attempt to identify boundaries of the external validity and thus eventual applicability of these frequently validated models.

This leads to a number of new recommendations in the discipline of CVD risk prediction research and practice. Firstly, this area would benefit from the formulation of guidance with clear definitions of the relevant outcomes (eg, similar to the CROWN initiative in obstetrics[58]), predictors, and prediction horizons. Secondly, the validity, and thus potential impact, of cardiovascular risk prediction models could substantially be improved by making better use of existing evidence, rather than starting from scratch to develop yet another model.[59] Thirdly, the suitable and promising models for a particular targeted population, outcome, and prediction horizon, should be identified, and subsequently be validated (and if necessary tailored to the situation at hand), allowing for head-to-head comparisons such as previously done for prediction models for type 2 diabetes[60] and patients requiring cardiac surgery.[61] Fourthly, more work is needed to evaluate the presence of heterogeneity in performance of different models across countries, allowing for tailoring of prediction models to different subpopulations. This can be achieved by combining the individual participant data (IPD) from multiple sources, including the increasingly available large registry datasets, and performing the so called IPD meta-analysis.[62 63] Analysis of such combined or large datasets has the advantage not only of increased total sample size, but also of better tackling case mix effects, setting specific issues (eg, inclusion of setting specific predictors), and better tailoring of existing models to different settings and consequently improving the robustness and thus generalisability of prediction models across subgroups and countries. Recently,

prediction modelling methods for analysis of large, combined datasets have been proposed.[59-68] If, after these efforts, generalisability of a developed and validated prediction model is still not good enough (eg, because of too much differences between populations, treatment standards, or data quality), more advanced methods for redevelopment of models can be used. Promising techniques are dynamic prediction modelling,[69 70] modelling strategies that take into account treatment-covariate interactions,[71] or other techniques such as machine learning.[72 73] Finally, models with adequate generalisability—as inferred from external validation studies—should be evaluated for potential impact on doctors' decision making or patient outcomes, before being incorporated in guidelines.[16 74] A recently published systematic review showed that the provision of risk information increases prescribing of antihypertensive drugs and lipid lowering drugs, but to our knowledge there are yet no studies investigating the effect of the use of prediction models and risk information provision on actual incidences of CVD events.[27]

Conclusions

The current literature is overwhelmed with models for predicting the risk of cardiovascular outcomes in the general population. Most, however, have not been externally validated or directly compared on their relative predictive performance, making them currently of yet unknown value for practitioners, policy makers, and guideline developers. Moreover, most developed prediction models are insufficiently reported to allow external validation by others, let alone to become implemented in clinical guidelines or being used in practice. We believe it is time to stop developing yet another similar CVD risk prediction model for the general population. Rather than developing such new CVD risk prediction models, in this era of large and combined datasets, we should focus on externally validating and comparing head-to-head the promising existing CVD risk models, on tailoring these models to local settings, to investigate whether they may be extended with new predictors, and finally to quantify the clinical impact of the most promising models.

1 Eckel RH, Jakicic JM, Ard JD, et al. American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 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(Suppl 2):S76-99. doi:10.1161/01.cir.0000437740.48606.d1.

2 Alwan A. Global status report on noncommunicable diseases 2010: World Health Organization, 2011. http://apps.who.int/iris/bitstream/10665/44579/1/9789240686458_eng.pdf.

3 Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular disease risk profiles. Am Heart J 1991;121:293-8. doi:10.1016/0002-8703(91)90861-B.

4 Anderson KM, Wilson PW, Odell PM, Kannel WB. An updated coronary risk profile. A statement for health professionals. Circulation 1991;83:356-62. doi:10.1161/01.CIR.83.1.356.

5 Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97:1837-47. doi:10.1161/01.CIR.97.18.1837.

6 Conroy RM, Pyörälä K, Fitzgerald AP, et al. SCORE project group. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 2003;24:987-1003. doi:10.1016/S0195-668X(03)00114-3.

7 Hippisley-Cox J, Coupland C, Robson J, Brindle P. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. BMJ 2010;341:c6624. doi:10.1136/bmj.c6624.

8 Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007;335:136. doi:10.1136/bmj.39261.471806.55.

9 Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ 2008;336:1475-82. doi:10.1136/bmj.39609.449676.25.

10 Goff DC Jr, , Lloyd-Jones DM, Bennett G, et al. American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 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(Suppl 2):S49-73. doi:10.1161/01.cir.0000437741.48606.98.

11 Lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. NICE guidelines [CG181], National Institute for Health and Clinical Excellence 2014. https://www.nice.org.uk/guidance/cg181.

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).Royal College of General Practitioners, 2008, http://www.ncbi.nlm.nih.gov/books/NBK55818/.

13 Matheny M, McPheeters ML, Glasser A, et al. Systematic Review of Cardiovascular Disease Risk Assessment Tools. Rockville MD, 2011. http://www.ncbi.nlm.nih.gov/books/NBK56166/.

14 Wessler BS, Lai Yh L, Kramer W, et al. Clinical Prediction Models for Cardiovascular Disease: Tufts Predictive Analytics and Comparative Effectiveness Clinical Prediction Model Database. Circ Cardiovasc Qual Outcomes 2015;8:368-75. doi:10.1161/CIRCOUTCOMES.115.001693.

15 Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014;11:e1001744. doi:10.1371/journal.pmed.1001744.

16 Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;338:b606. doi:10.1136/bmj.b606.

17 Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012;98:683-90. doi:10.1136/heartjnl-2011-301246.

18 Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ* 2009;338:b604. doi:10.1136/bmj.b604.

19 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35:1925-31. doi:10.1093/eurheartj/ehu207.

20 Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;19:453-73. doi:10.1002/(SICI)1097-0258(20000229)19:4<453::AID-SIM350>3.0.CO;2-5.

21 Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605. doi:10.1136/bmj.b605.

22 Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013;13:33. doi:10.1186/1471-2288-13-33.

23 Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;338:b375. doi:10.1136/bmj.b375.

24 Harrell FE Jr, , Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

25 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;23:2567-86. doi:10.1002/sim.1844.

26 Steyerberg EW, Pencina MJ, Lingsma HF, Kattan MW, Vickers AJ, Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. *Eur J Clin Invest* 2012;42:216-28. doi:10.1111/j.1365-2362.2011.02562.x.

27 Usher-Smith JA, Silarova B, Schuit E, Moons KG, Griffin SJ. Impact of provision of cardiovascular disease risk estimates to healthcare professionals and patients: a systematic review. *BMJ Open* 2015;5:e008717. doi:10.1136/bmjopen-2015-008717.

28 van Dieren S, Beulens JW, Kengne AP, et al. Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review. *Heart* 2012;98:360-9. doi:10.1136/heartjnl-2011-300734.

29 Bouwmeester W, Zuithoff NP, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med* 2012;9:1-12. doi:10.1371/journal.pmed.1001221.

30 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;162:55-63. doi:10.7326/M14-0697.

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

32 Steyerberg EW, Moons KG, van der Windt DA, et al. PROGRESS Group. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10:e1001381. doi:10.1371/journal.pmed.1001381.

33 Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995;48:1503-10. doi:10.1016/0895-4356(95)00048-8.

34 Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373-9. doi:10.1016/S0895-4356(96)00236-3.

35 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128-38. doi:10.1097/EDE.0b013e3181c30fb2.

36 D'Agostino RB Sr, , Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;117:743-53. doi:10.1161/CIRCULATIONAHA.107.699579.

37 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;106:3143-421.http://circ.ahajournals.org/content/106/25/3143.citation.

38 Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Münster (PROCAM) study. *Circulation* 2002;105:310-5. doi:10.1161/hc0302.102575.

39 Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke: a risk profile from the Framingham Study. *Stroke* 1991;22:312-8. doi:10.1161/01.STR.22.3.312.

40 Chambless LE, Folsom AR, Sharrett AR, et al. Coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC) study. *J Clin Epidemiol* 2003;56:880-90. doi:10.1016/S0895-4356(03)00055-6.

41 Friedland DR, Cederberg C, Tarima S. Audiometric pattern as a predictor of cardiovascular status: development of a model for assessment of risk. *Laryngoscope* 2009;119:473-86. doi:10.1002/lary.20130.

42 Keys A, Aravanis C, Blackburn H, et al. Probability of middle-aged men developing coronary heart disease in five years. *Circulation* 1972;45:815-28. doi:10.1161/01.CIR.45.4.815.

43 Leaverton PE, Sorlie PD, Kleinman JC, et al. Representativeness of the Framingham risk model for coronary heart disease mortality: a comparison with a national cohort study. *J Chronic Dis* 1987;40:775-84. doi:10.1016/0021-9681(87)90129-9.

44 Barzi F, Patel A, Gu D, et al. Asia Pacific Cohort Studies Collaboration. Cardiovascular risk prediction tools for populations in Asia. *J Epidemiol Community Health* 2007;61:115-21. doi:10.1136/jech.2005.044842.

45 Woodward M, Brindle P, Tunstall-Pedoe H. SIGN group on risk estimation. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart* 2007;93:172-6. doi:10.1136/hrt.2006.108167.

46 Levy D, Wilson PW, Anderson KM, Castelli WP. Stratifying the patient at risk from coronary disease: new insights from the Framingham Heart Study. *Am Heart J* 1990;119:712-7, discussion 717. doi:10.1016/S0002-8703(05)80050-X.

47 Chien KL, Hsu HC, Su TC, et al. 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;157:263-8. doi:10.1016/j.ijcard.2012.03.017.

48 Collins GS, Moons KG. Comparing risk prediction models. *BMJ* 2012;344:e3186. doi:10.1136/bmj.e3186.

49 WHO. 2015. Cardiovascular diseases (CVDs) Fact sheet N°317. http://www.who.int/mediacentre/factsheets/fs317/en/. Last access: 09-12-2015

50 Hajifathalian K, Ueda P, Lu Y, 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;3:339-55. doi:10.1016/S2213-8587(15)00081-9.

51 Liu J, Hong Y, D'Agostino RB Sr, et al. Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-Provincial Cohort Study. *JAMA* 2004;291:2591-9. doi:10.1001/jama.291.21.2591.

52 Wu Y, Liu X, Li X, et al. USA-PRC Collaborative Study of Cardiovascular and Cardiopulmonary Epidemiology Research Group China Multicenter Collaborative Study of Cardiovascular Epidemiology Research Group. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. *Circulation* 2006;114:2217-25. doi:10.1161/CIRCULATIONAHA.105.607499.

53 Gijsberts CM, Groenewegen KA, Hoefer IE, et al. Race/Ethnic Differences in the Associations of the Framingham Risk Factors with Carotid IMT and Cardiovascular Events. *PLoS One* 2015;10:e0132321. doi:10.1371/journal.pone.0132321.

54 Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med* 2011;9:103. doi:10.1186/1741-7015-9-103.

55 Collins GS, Omar O, Shanyinde M, Yu LM. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. *J Clin Epidemiol* 2013;66:268-77. doi:10.1016/j.jclinepi.2012.06.020.

56 Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014;383:267-76. doi:10.1016/S0140-6736(13)62228-X.

57 Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1-73. doi:10.7326/M14-0698.

58 Khan KS, Romero R. Chief Editors of Journals participating in CROWN Initiative. The CROWN initiative: journal editors invite researchers to develop core outcomes in women's health. *Am J Obstet Gynecol* 2014;211:575-6. doi:10.1016/j.ajog.2014.09.015.

59 Debray TP, Koffijberg H, Nieboer D, Vergouwe Y, Steyerberg EW, Moons KG. Meta-analysis and aggregation of multiple published prediction models. *Stat Med* 2014;33:2341-62. doi:10.1002/sim.6080.

60   Abbasi A, Peelen LM, Corpeleijn E, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *BMJ* 2012;345:e5900. doi:10.1136/bmj.e5900.

61   Ettema RG, Peelen LM, Schuurmans MJ, Nierich AP, Kalkman CJ, Moons KG. Prediction models for prolonged intensive care unit stay after cardiac surgery: systematic review and validation study. *Circulation* 2010;122:682-9, 7, 689. doi:10.1161/CIRCULATIONAHA.109.926808.

62   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;68:279-89. doi:10.1016/j.jclinepi.2014.06.018.

63   Debray TP, Riley RD, Rovers MM, Reitsma JB, Moons KG. Cochrane IPD Meta-analysis Methods group. Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. *PLoS Med* 2015;12:e1001886. doi:10.1371/journal.pmed.1001886.

64   Debray TP, Koffijberg H, Vergouwe Y, Moons KG, Steyerberg EW. Aggregating published prediction models with individual participant data: a comparison of different approaches. *Stat Med* 2012;31:2697-712. doi:10.1002/sim.5412.

65   Debray TP, Moons KG, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Stat Med* 2013;32:3158-80. doi:10.1002/sim.5732.

66   Snell KI, Hua H, Debray TP, et al. 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. doi:10.1016/j.jclinepi.2015.05.009.

67   Royston P, Parmar MK, Sylvester R. Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. *Stat Med* 2004;23:907-26. doi:10.1002/sim.1691.

68   Sauerbrei W, Royston P. A new strategy for meta-analysis of continuous covariates in observational studies. *Stat Med* 2011;30:3341-60. doi:10.1002/sim.4333.

69   Nicolaie MA, van Houwelingen JC, de Witte TM, Putter H. Dynamic prediction by landmarking in competing risks. *Stat Med* 2013;32:2031-47. doi:10.1002/sim.5665.

70   Teramukai S, Okuda Y, Miyazaki S, Kawamori R, Shirayama M, Teramoto T. Dynamic prediction model and risk assessment chart for cardiovascular disease based on on-treatment blood pressure and baseline risk factors. *Hypertens Res* 2016;39:113-8. doi:10.1038/hr.2015.120.

71   van Klaveren D, Vergouwe Y, Farooq V, Serruys PW, Steyerberg EW. Estimates of absolute treatment benefit for individual patients required careful modeling of statistical interactions. *J Clin Epidemiol* 2015;68:1366-74. doi:10.1016/j.jclinepi.2015.02.012.

72   Wolfson J, Bandyopadhyay S, Elidrisi M, et al. A Naive Bayes machine learning approach to risk prediction using censored, time-to-event data. *Stat Med* 2015;34:2941-57. doi:10.1002/sim.6526.

73   Guo Y, Wei Z, Keating BJ, Hakonarson H. Genetic Consortium for Anorexia Nervosa Wellcome Trust Case Control Consortium 3 Price Foundation Collaborative Group. Machine learning derived risk prediction of anorexia nervosa. *BMC Med Genomics* 2016;9:4. doi:10.1186/s12920-016-0165-x.

74   Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;98:691-8. doi:10.1136/heartjnl-2011-301247.

Web appendix: supplementary information