BMJ Open

Can risk be predicted? An umbrella systematic review of current risk prediction models for cardiovascular diseases, diabetes and hypertension

Francesca Lucaroni,[1] Domenico Cicciarella Modica,[2] Mattia Macino,[2] Leonardo Palombi [ID],[2] Alessio Abbondanzieri,[2] Giulia Agosti,[2] Giorgia Biondi,[2] Laura Morciano,[1] Antonio Vinci [ID] [2]

¹Biomedicine and Prevention, University of Rome Tor Vergata, Roma, Lazio, Italy
²Biomedicine and Prevention, Università degli Studi di Roma Tor Vergata Facoltà di Medicina e Chirurgia, Roma, Lazio, Italy

Correspondence to
Professor Leonardo Palombi;
palombi@uniroma2.it

ABSTRACT

Objective To provide an overview of the currently available risk prediction models (RPMs) for cardiovascular diseases (CVDs), diabetes and hypertension, and to compare their effectiveness in proper recognition of patients at risk of developing these diseases.

Design Umbrella systematic review.

Data sources PubMed, Scopus, Cochrane Library.

Eligibility criteria Systematic reviews or meta-analysis examining and comparing performances of RPMs for CVDs, hypertension or diabetes in healthy adult (18–65 years old) population, published in English language.

Data extraction and synthesis Data were extracted according to the following parameters: number of studies included, intervention (RPMs applied/assessed), comparison, performance, validation and outcomes. A narrative synthesis was performed. Data were reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.

Study selection 3612 studies were identified. After title/abstract screening and removal of duplicate articles, 37 studies met the eligibility criteria. After reading the full text, 13 were deemed relevant for inclusion. Three further papers from the reference lists of these articles were then added.

Study appraisal The methodological quality of the included studies was assessed using the AMSTAR tool.

Risk of bias in individual studies Risk of Bias evaluation was carried out using the ROBIS tool.

Results Sixteen studies met the inclusion criteria: six focused on diabetes, two on hypertension and eight on CVDs. Globally, prediction models for diabetes and hypertension showed no significant difference in effectiveness. Conversely, some promising differences among prediction tools were highlighted for CVDs. The Ankle-Brachial Index, in association with the Framingham tool, and QRISK scores provided some evidence of a certain superiority compared with Framingham alone.

Limitations Due to the significant heterogeneity of the studies, it was not possible to perform a meta-analysis. The electronic search was limited to studies in English and to three major international databases (MEDLINE/PubMed, Scopus and Cochrane Library), with additional works derived from the reference list of other studies; grey literature with unpublished documents was not included

Strengths and limitations of this study

► This is the most comprehensive umbrella systematic review on risk prediction models for cardiovascular diseases, hypertension and diabetes to date.
► Available studies, although apparently of medium-to-high quality, were based on primary studies of debatable quality, several of which lack discrimination and calibration assessments.
► Grey literature was not searched.
► Heterogeneity was too high for meta-analysis; results are, therefore, reported narratively.

in the search. Furthermore, no assessment of potential adverse effects of RPMs was carried out.

Conclusions Consistent evidence is available only for CVD prediction: the Framingham score, alone or in combination with the Ankle-Brachial Index, and the QRISK score can be confirmed as the gold standard. Further efforts should not be concentrated on creating new scores, but rather on performing external validation of the existing ones, in particular on high-risk groups. Benefits could be further improved by supplementing existing models with information on lifestyle, personal habits, family and employment history, social network relationships, income and education.

PROSPERO registration number CRD42018088012.

INTRODUCTION

Cardiovascular diseases (CVDs), hypertension and diabetes represent a major health concern, throughout the world and across income levels, as a silent epidemic responsible for millions of deaths every year.

CVDs, excluding hypertension, are the leading cause of death worldwide; they have a global prevalence rate of 6.6% and account for 17.6 million deaths per year.[1]

Hypertension alone is the leading preventable cause of premature death worldwide,[2] and causes 7.5 million deaths each year.[3]

In 2010, almost one of every three adults (31.1%) had hypertension, although there was a significant gap in prevalence rates between high-income and low-income countries: 28.5% (27.3%–29.7%) and 31.5% (30.2%–32.9%), respectively, worldwide.[2]

Diabetes (types 1 and 2 combined) has a global prevalence rate of 5.4% and is responsible for 1.4 million deaths every year.[3]

In addition to being a significant cause of mortality worldwide, CVDs, hypertension and diabetes are also a leading cause of disability. Together they account for almost 40% of disability-adjusted life years (DALYs): CVDs alone are responsible for 32.3% of total DALYs, hypertension for 3.7% and diabetes for 2.4%.[4]

The epidemiology of these three diseases helps explain the substantial economic impact they have on national health services: in the USA, CVD-related direct costs amount to approximately US$444 billion per year,[5] whereas the costs of diabetes are estimated as US$327 billion per year; hypertension has an annual estimated cost of US$51 billion,[6] most of which (nearly US$48 billion) represents direct medical expenses.[7]

In recent decades, hypertension and diabetes have shown an increasing trend in both prevalence and mortality rates. It has been estimated that the prevalence of diabetes will continue to grow: one in five to one in three adults will be affected by 2050. The same is true for mortality rate trends: in the last 15 years, diabetes-related deaths increased by 1%, and these data are expected to climb dramatically over the coming decades.[8]

The prevalence of hypertension and the associated mortality rates have increased significantly in both men and women,[9] and by 2030, the prevalence is projected to be 44%.[10]

Conversely, prevalence and mortality rate trends for CVDs have significantly decreased,[11] including in countries that had seen considerable increases until the beginning of the 21st century.[12]

CVDs, hypertension and diabetes are strongly related to each other: diabetes is associated with an increased risk of CVDs, which is exaggerated by concomitant hypertension. These conditions also share the same pathogenic pathways, at both macroscopic and molecular level: oxidative stress, inflammation and fibrosis, cause microvascular and macrovascular complications in diabetes, and also lead to vascular remodelling and dysfunctions in hypertension.[13]

Because of the high prevalence and mortality rates associated with CVDs, hypertension and diabetes, and their related direct and indirect costs, early identification of individuals at high risk for these diseases is crucial; it results in terms of obtaining significant savings in both global health outcomes and economic expenditures.

A number of prediction models focused on these three non-communicable diseases (NCDs) are available, but there is no consensus as to the gold-standard tools best used in practice.

Table 1 Search strategy

Database	Search string(s)	Filters
Cochrane Library	risk (prediction OR assessment) (model OR score)	/
	diabetes risk prediction	
	cardiovascular risk prediction	
	hypertension risk prediction	
PubMed	(risk(Title/abstract)OR "risk assessment"(MeSH Terms)) AND predict*(Title/abstract) AND (model(Title/abstract)OR score(Title/abstract))	Systematic reviews
		Meta-analysis
		Abstract
		Humans
		English
Scopus	(risk AND prediction AND model AND (systematic AND review OR meta-analysis))	Review
		English
		Medic

AMSTAR, A MeaSurement Tool to Assess systematic Reviews tool.

The aim of this study is to provide an overview of the currently available risk prediction models (RPMs) for CVDs, diabetes and hypertension and to compare their effectiveness at properly recognising vulnerable people, at risk of developing these NCDs.

METHODS

This umbrella systematic review was performed following a protocol designed a priori, and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.[14]

The PubMed, Scopus and Cochrane databases were searched electronically on, 24 September 2019, using combinations of the relevant medical subject heading terms, key words and word variants for 'risk prediction scores' and 'CVD', 'diabetes' and 'hypertension', as shown in table 1. The search and selection criteria were restricted to the following: systematic reviews, with or without meta-analysis, as the type of study; general population aged 18–65 with no major illness; comparison of at least two RPMs; English language. For the PubMed database, only two further filters were added: only articles on humans and with abstracts available were included. No restrictions were applied in terms of publication date in any of the databases.

Reference lists of relevant articles and reviews were hand-searched for additional reports.

Two different authors independently screened the article titles and abstracts in each database: MM and AA for Scopus, AV and DCM for PubMed/Medline, AV and AA for Cochrane Library.
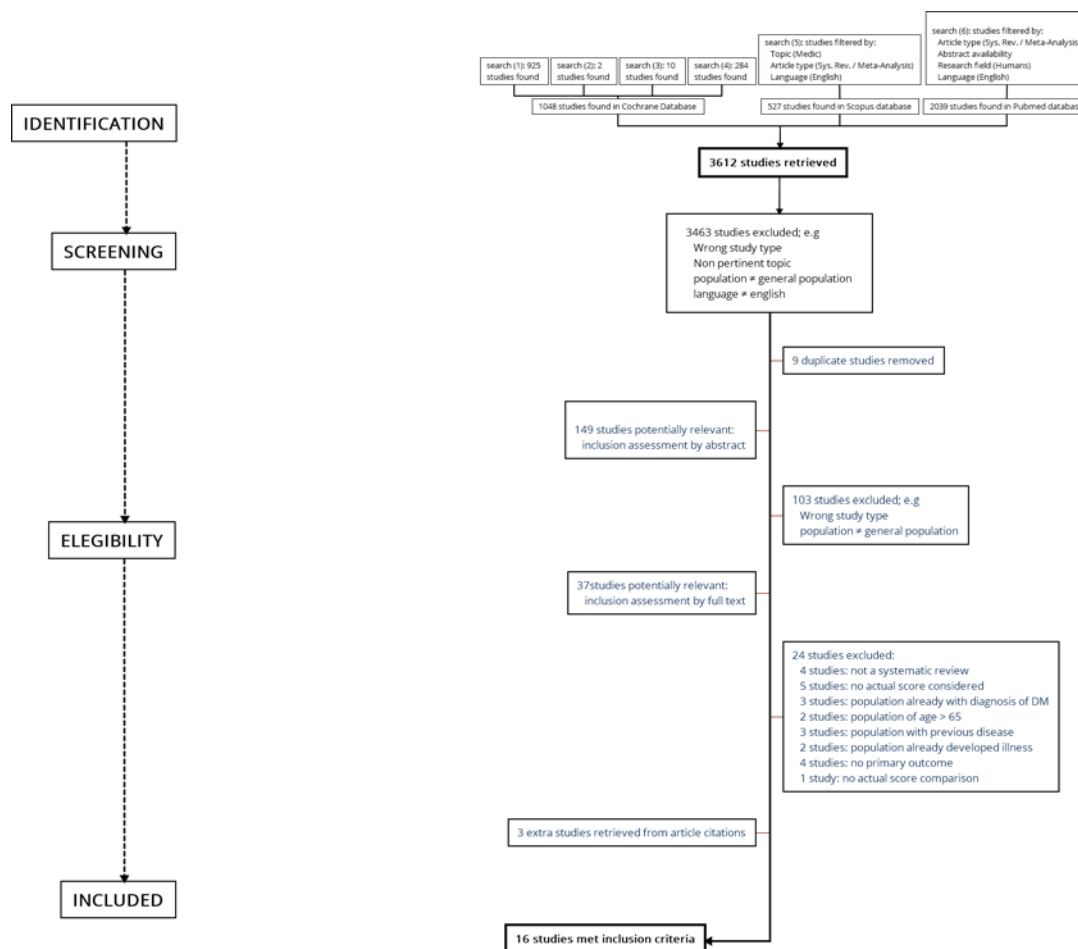
Figure 1 Study workflow.

Disagreements were discussed by the authors and resolved by consensus or by recourse to a third author (FL). Studies were then labelled for inclusion or exclusion.

Every article meeting the eligibility criteria—systematic reviews/meta-analyses on RPMs on CVDs, diabetes and hypertension, evaluated by comparison with other RPMs, in adults with no relevant illness—was considered for subsequent qualitative synthesis; duplicate records were removed, as—were articles that included the exclusion criteria: any study carried out with the sole purpose of developing a new RPM, or validating one, or to propose a diagnostic/prognostic tool, without any comparisons to other prediction models.

Studies were henceforth labelled for inclusion or exclusion.

The selection process described above is summarised by the flow diagram shown in figure 1.

Data extraction

Four authors (AA, DCM, MM and AV) extracted the data, including the following variables: number of the included studies; intervention (RPMs); comparison; performance: area under the receiver-operating characteristic curve (AUC), C-statistic, D-statistic); validation: internal, external, both or not provided; outcomes: incidence, prevalence, mortality.

Assessment of study quality

Quality assessment of individual studies was performed by applying the AMSTAR tool.[15] According to their score, articles were classified into three groups: low (AMSTAR score <4), medium (AMSTAR score ≥4 and ≤7) and high quality (AMSTAR score ≥8).

Four authors (AA, DCM, MM and AV) independently assigned score. Disagreements were resolved by consensus or by discussion with a fifth author (FL). No reviews were excluded ex-post for quality reasons.

A qualitative (narrative) synthesis of the selected reviews was then performed; the guidelines for umbrella review from the Joanna Briggs Institute were applied.[16]

Risk of bias in individual studies

Risk of bias in individual studies was independently evaluated by four authors (AA, DCM, MM and AV), using the ROBIS tool.[17] Any disagreements were resolved by consensus or by discussion with a fifth author (FL).

Patient and public involvement

As the study design was a systematic review, neither patients nor the public were involved.

RESULTS

A total of 3612 studies were identified through a search of the electronic databases.
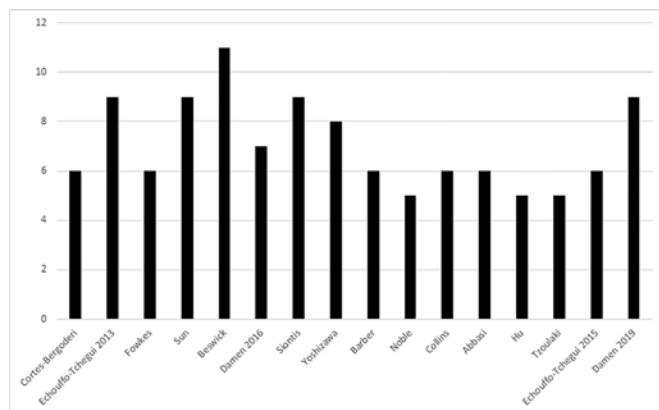
Figure 2 Quality assessment scores.

After title screening, 3463 studies were excluded because they did not meet the eligibility criteria and nine duplicate articles were removed. Of the 149 studies passing the first evaluation stage, 102 studies were excluded because the topic was not pertinent, sample characteristics were inadequate (ie, some articles included non-healthy populations), the article type did not meet the inclusion criteria (ie, some studies were not systematic reviews), or there was no comparison between models.

After reading the full text of the 37 remaining studies, 13 were deemed relevant for inclusion. Twenty-four studies were excluded because of a lack of comparison model, related but non-pertinent topic or wrong article type.

Three additional studies were included after consulting reference lists of relevant articles and reviews overall, 16 studies met the eligibility criteria and were included in qualitative synthesis.

Quality assessment of the studies

According to the AMSTAR tool, all of the studies were of medium-to-high quality, with a mean score of 7.14, (range 5–11). Specifically, 8 of 15 (53.4%) were of medium quality (AMSTAR ≥4 and ≤7), and 7 of 15 (46.6%) were of high quality (AMSTAR >8).

The results of quality assessment have been summarised in graph form (figure 2).

Risk of bias within studies

The results of the risk-of-bias evaluation are shown in table 2 and figure 3.

Globally, only 9 articles of 16 (56.25%) had a low risk of bias, according to the ROBIS tool.

Two articles had a high risk of bias due to the eligibility criteria: search limitation on English language. Eight articles had a high risk of bias due to the identification and selection process: the most common source of bias was the search limitation to a single database. Nine articles had a high risk of bias due to data collection and study appraisal, in particular because of the lack of formal appraisal tools. Finally, one article had a high risk of bias due to the synthesis and identification process, mainly due to significant heterogeneity of primary sources.

Table 2	Risk of bias				
	Phase 2				Phase 3
Review	**1. study eligibility criteria**	**2. identification and selection of the studies**	**3. data collection and study appraisal**	**4. synthesis and findings**	**Risk of bias in the systematic reviews included**
Abbasi et al[18]	☺	☹	☹	☺	☹
Barber et al[19]	☺	☹	☹	☺	☹
Beswick et al[29]	☺	☺	☺	☺	☺
Collins et al[35]	☺	☹	☹	☺	☺
Cortes-Bergoderi et al[24]	☺	☹	☹	☺	☹
Damen et al[28]	☹	☹	☺	☺	☺
Damen et al[31]	☺	☺	☺	☺	☺
Echouffo-Tcheugui et al[33]	☺	☺	☹	☺	☺
Echouffo-Tcheugui et al[30]	☺	?	☺	☺	?
Fowkes et al[26]	☺	☺	☹	?	?
Hu et al[21]	?	?	?	☹	?
Noble et al[22]	☺	☺	☹	☺	☺
Siontis et al[27]	☺	☹	☺	☺	☺
Sun et al[32]	☹	☹	☹	☺	☹
Tzoulaki et al[25]	☺	☹	☺	☺	☺
Yoshizawa et al[23]	☺	☺	☹	☺	☺

☺=low risk; ☹=high risk; **?**=unclear risk.
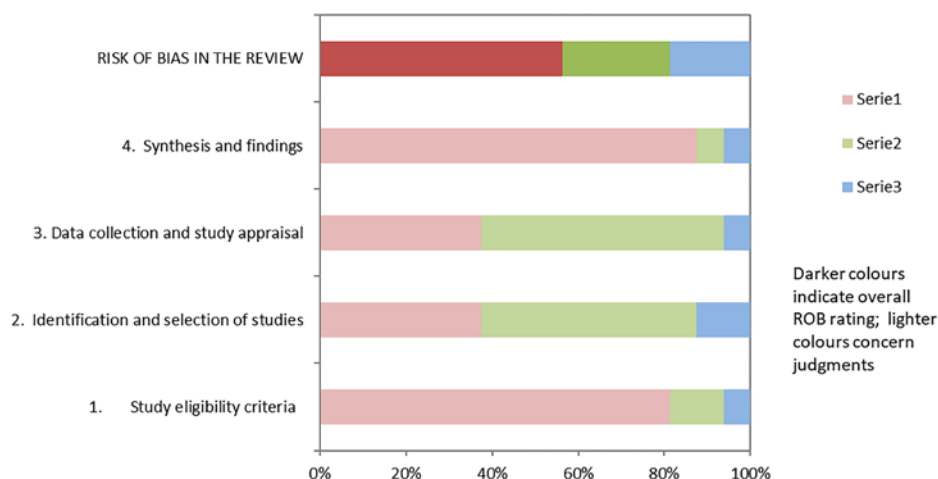ROBIS, Risk Of Bias In Studies tool; SRs, Systematic Revews.

Figure 3 Graphical representation of risk of bias according to the ROBIS tool.

Given the overall heterogeneity of the retrieved studies, it was impossible to conduct a meta-analysis.

Risk of bias across systematic reviews

We identified all of the ongoing SRs that met our inclusion criteria by searching the PROSPERO database, in order to assess publication bias.

Eight ongoing systematic reviews were found, including the present study, and none have been published to date.

Results of individual studies

Overall, 16 studies met the inclusion criteria, 8 of them concerning diabetes, 6 CVDs and 2 hypertension.

The results for all the included studies are summarised qualitatively in table 3.

Studies on diabetes

Abbasi et al[18] (AMSTAR 6/11) focused on 16 prospective cohort studies, in order to validate 25 risk-predictive models for type 2 diabetes mellitus(T2DM), by means of an external validation cohort. The sample was included 38379 people aged 20–70 with no diabetes at the baseline. Incidence of type 2 diabetes was evaluated as outcome. All included studies reported a C-statistic, ranging from 0.74 to 0.84 for risk at 7.5 years, indicating a good discriminatory capability. The risk models had an estimate of calibration, the Hosmer-Lemeshow test, generally indicating good calibration.

Barber et al[19] (AMSTAR 6/11) assessed the applicability of 18 risk assessment tools in individuals with prediabetes. Their systematic review included 12 studies, with sample sizes ranging from 1351 to 7092. Incidence of prediabetes, defined according to American Diabetes Association criteria, was considered as the primary outcome.

Validation (either internal or external) of the risk scores was achieved by evaluating both discrimination and calibration. The internal C-statistic ranged from 0.66 to 0.75. Calibration was described by Hosmer-Lemeshow

goodness-of-fit test value, reported by only two studies and with discordant results.

Collins et al[20] (AMSTAR 6/11) evaluated RPMs for type 2 diabetes, including 39 studies comparing 47 different risk tools. The studies had a median sample size of 2562 people, with an IQR from 1426 to 4965. No quantitative information was available on discrimination or calibration.

Hu et al[21] (AMSTAR 5/11) evaluated the effectiveness of risk-predictive models for type 2 diabetes in the Asian population. Their systematic review included 43 studies examining 12 risk-predictive models, derived from population samples ranging from 2677 to 73961.

Discrimination was evaluated by the AUC: this showed a high variability of results (AUC 0.66 to 0.91).

Noble et al[22] (AMSTAR 5/11) conducted a systematic review assessing 94 risk models for type 2 diabetes.

They evaluated 43 perspective cohort studies (sample size ranging from 399 to 2.54 million people) and incidence of diabetes was the primary outcome. Some of the risk models had been externally validated on a different population. The C-statistic index showed high variability, fluctuating between not acceptable (0.60) and good quality (0.91) scores. The same results applied for calibration indicators.

Yoshizawa et al[23] (AMSTAR 8/11) focused on evaluating the predictive ability of a non-blood-based RPM for incidence of T2DM. The 18 eligible studies included an overall number of 184011 participants aged 42.4–68.4 years. Discrimination, evaluated by the AUC, was adequate to good (0.72–0.81).

Studies on CVDs

Cortes-Bergoderi et al[24] (AMSTAR 6/11) assessed the validity of RPMs in Latin America and in US people of Hispanic descent. Their review included five cohort studies, comparing the Framingham score with three risk models for CVD and one for Chagas disease, and investigating incidence and mortality as outcomes. Risk score calibration measured by C-statistic index was good,

Table 3 Summary of results

Authors	Year of publication	No of studies	Intervention	Comparison	Formance (AUC, C	Validation	Outcomes: incidence, prevalence, mortality	Quality assessment score
Abbasi et al[18]	2012	16	25 type 2 diabetes risk prediction models.	Prospective cohort study, with a case cohort study random subcohort.	C-statistics: 0.74–0.92	External validation cohort	Type 2 diabetes morbidity (incidence)	6
Barber et al[19]	2014	12	18 type 2 diabetes risk prediction models.	Seven risk tools were validated using an external dataset. 14 of the tools were validated internally using resampling techniques such as bootstrapping and cross-validation. 4 of the tools also split their dataset into a 'training' set, which was used to create the tool. One risk tool was validated using a partially independent dataset; all cases were used in both the 'training' and 'test' cohorts, however, the non-cases were split into these two cohorts.	AUROC: 0.69	Internal and external validation cohort.	Pre-diabetes morbidity (incidence).	6
Beswick et al[29]	2008	110	110 risk scoring methods with potential for use in primary prevention (among them, 70 methods specifically designed for use in primary prevention).	internal?	AUROC min 0.57, max 0.88	Internal and convergent validation	CHD, MI/sudden ischaemic death CHD death, CVD, stroke, CHD, CVD death MI or CHD death CHD death CHD Death all causes Stroke (ABI) MI Heart attack CVD MI or stroke (ABI) Sudden death Hard CHD All causes of death CHD (CAD) fatal MI non-fatal MI sudden coronary death chronic CHD IHD fatal/ non-fatal MI CHD death or non fatal MI hard CHD hard CVD hard cerebrovascular macrovascular disease fatal MI/ sudden death MI (fatal or non-fatal).	11
Collins et al[35]	2011	39	Use of 47 different risk predictors.	10 studies randomly split the cohort into development and validation cohorts. 8 of these studies split the original cohort equally into development and validation cohorts. 21 studies conducted and published an external validation of their risk prediction models within the same article, and 8 of these studies used two or more data sets in an attempt to demonstrate the external validity of the risk prediction models.	NA	Internal and external validation cohort	Diabetes morbidity (incidence and prevalence).	6

Table 3 Continued

Authors	Year of publication	No of studies	Intervention	Comparison	Formance (AUC, C	Validation	Outcomes: incidence, prevalence, mortality	Quality assessment score
Cortes-Bergoderi et al[24]	2012	5	Framingham Novel RPM based on local cohort WHO/ISH cardiovascular risk score system Chagas RPM.	Framingham not for Chagas.	C-stat MAX 0.81 (95% CI, 0.72 to 0.90); MIN 0.69	N/A (oChagas only)	General mortality, CVD mortality and morbidity, (incidence), and CHD mortality and morbidity (incidence).	6
Damen et al[28]	2016	212	363 different models.	Internal?	For 143 (39%) models, discrimination was reported as a Cstatistic or area under the receiver operating characteristic curve; D statistic 5 (1%) other (sensitivity, specificity, etc) 24 (7%).	80 of the 363 developed models (22%) were internally validated; 132 (36%) models were externally validated	Fatal or non-fatal CHD Fatal or non-fatal: CHD, CVD, myocardial infarction and stroke fatal: CHD, CVD and non-CHD Fatal CVD Fatal or non-fatal stroke fatal or non-fatal: CHD, myocardial infarction and stroke; claudication; coronary artery bypass grafting; percutaneous transluminal coronary angioplasty; transient ischaemic attack fatal CHD.	7
Damen et al[31]	2019	38	Framingham Adult Treatment Panel III, the Framingham Wilson model and the pooled cohort equations.	Internal	OE from 0.58 to 0.79; C-STAT from 0,58 to 0,82.	External	Fatal or non-fatal MI, Fatal or non-fatal stroke, fatal, hard or non-fatal CHD, angina pectoris, TIA, ASCVD.	9
Echouffo-Tcheugui et al[30]	2015	13	22 RISK SCORES: FHS risk score Health ABC Score Kaiser Permanente FHS Risk score 1 FHS Risk score 2 FHS Risk score 3 CHS score 1 CHS risk score 2 MDCS score 1 MDCS score 2 MDCS score 3 MDCS score 4 MDCS score 5 MDCS score 6 MDCS score 7 ABC score 2 Proactive trial score CHS? FHS ARIC NAVIGATOR study British Regional Heart Study.	Internal	AUC from 0.71 to 0.87	HL $\chi 2$ from 2.98 to 9.45	Heart failure incidence.	6
Echouffo-Tcheugui et al[33]	2013	11	Johns Hopkins Framingham score Women's Health Study (WHS) Inclusive model WHS simplified model with lipids WHS- simplified model Whitehall II risk score Whitehall II Repeat measures risk score Whitehall II Average blood pressure measure risk score Whitehall II usual measure risk scores ARIC/CHS Score Iran BP risk score Taiwan BP clinical risk model Taiwan BP biochemical risk model Korean risk model Swedish risk model.	Internal?	MAX 0.803, MIN 0.707	External? C-statistic in validation studies (0.71 to 0.81)	Hypertension morbidity (incidence).	9
Fowkes et al[26]	2008	20	Baseline ABI measurements.	Framingham risk score	N/A		General mortality, CVD mortality and CHD morbidity (incidence).	6
Hu et al[21]	2016	12	12 type 2 diabetes risk prediction models.	To be seen	AUROC: 0.66–0.91	Internal and external validation cohort	Diabetes morbidity (incidence).	8
Noble et al[22]	2011	43	94 type 2 diabetes risk prediction models.		AUROC: 0.74–0.85	Internal and external validation cohort	Diabetes morbidity (incidence).	5
Siontis et al[27]	2012	20	Framingham ASSIGN score SCORE score PROCAM score Reynolds risk score QRISK1 QRISK2.	Internal?	AUC MIN 0.55 MAX 0.83	N/A	CVD mortality, CHD incidence, cerebrovascular disease incidence, CABG or PTCA.	9

Continued

Table 3 Continued

Authors	Year of publication	No of studies	Intervention	Comparison	Formance (AUC, C	Validation	Outcomes: incidence, prevalence, mortality	Quality assessment score
Sun et al[32]	2017	26	Anthropometric indices risk prediction ARIC/CHC risk score biomarker-based risk-prediction model China risk prediction model 1 China risk prediction model 2 China risk prediction score conditional model Demographic indices risk prediction model one for men demographic indices risk prediction model two for men demographic indices risk prediction model three for women demographic indices risk prediction model 1 for women demographic indices risk prediction model 2 for women demographic indices risk prediction model 3 for women Framingham risk score genetic risk prediction model 1 genetic risk prediction model 2 genetic risk prediction model 3 IDH risk prediction model InterASIA risk prediction ISH risk prediction model Japanese risk prediction model Japanese risk score sheet Johns Hopkins KoGES risk score Korean genetic risk score marginal model northeastern Han Chinese genetic risk score Prediction for men Prediction for women rural India risk score SHIP risk model Swedish genetic risk model Swedish nongenetic risk model Swedish risk model 2 Taiwan BP clinical risk model the 'usual' blood pressure risk score the average blood pressure risk score the PFAA index TLGS risk multivariable models TLGS risk prediction for IDH TLGS risk prediction for ISH TLGS risk score Whitehall II repeat measures risk score Whitehall II risk score WHS inclusive risk prediction WHS Simplified Model WHS Simplified Model with Lipids.	Internal?	AUC=0.767, 95%CI(0.742, 0.792)	Variable	Hypertension morbidity (incidence)	9
Tzoulaki et al[25]	2009	79	Application of FRS plus a candidate additional predictor.	Framingham risk score	AUC 0.77	n.p.	General mortality, CVD mortality and morbidity (incidence) and CHD mortality and morbidity (incidence).	5
Yoshizawa et al[23]	2016	18	Non-blood-based risk model for type 2 diabetes	None of the included studies met all criteria on the QUADAS list. Among five studies that had both derivation and validation cohorts, three studies conducted the calibration in both cohorts; two studies did it only in derivation cohort.	AUC: 0.71–0.79	Internal and external validation cohort.	Diabetes morbidity (incidence)	8

ABC, Age, Biomarker, Clinical History score; ABI, Ankle-Brachial Index; ARIC, Atherosclerosis Risk in Communities study; ASCVD, Arteriosclerotic Cardiovascular Disease; ASSIGN, Assessing cardiovascular risk using SIGN (Scottish Intercollegiate Guidelines Network) guidelines score; AUC, AUROC, area under the curve/area under the receiver-operating characteristic curve; BP, blood pressure; CABG, coronary artery bypass grafting; CHD, coronary heart disease; CHS, Cardiovascular Health Study; C-STAT, C-statistic; CVD, cardiovascular disease; D-STAT, D-statistic; FHS, Framingham Hearth Study; FRS, Framingham Risk Score; ISH, Isolated systolic Hypertension; MI, Miocardial Infarction; N/A, Not Available; PROCAM, Cardiovascular Risk PROCAM Score; PTCA, Percutaneous transluminal coronary angioplasty; RPM, Risk Prediction Model; TLGS, Tehran Lipid and Glucose Study.

(0.69–0.80). While the Authors openly admit that the Framingham score needs to be recalibrated for Latin American populations, they also recognise that evidence regarding CVD risk models is 'modest at best'. Indeed, all of the included studies showed a ratio of predicted/observed that was not significant.

Tzoulaki et al[25] (AMSTAR: 5/11) focused on 79 studies on Framingham-based improving models, derived from populations of less than 1000 to over 10 000 subjects from the USA and the UK. Incidence and mortality for coronary heart disease (CHD) were measured as outcome.

The discrimination ability of the examined scores, evaluated by AUC, varied from not acceptable to good: the FRS alone model showed an area under the curve between 0.50 and 0.83, whereas FRS with additional predictors ranged from .57 to 0.84.

Fowkes et al[26] (AMSTAR 6/11) evaluated the Ankle-Brachial Index (ABI) as predictor of cardiovascular events and mortality, compared with the Framingham Risk Score. They included 20 prospective studies involving general populations from the EU and the USA, with sample sizes that ranged from 554 to 14 109.

The combination of ABI and FRS risk prediction scores had a higher discriminating power compared with FRS alone (0.655 vs 0.646 among men, and 0.658 vs 0.605 among women).

Incidence of CVDs was assessed, as primary outcome, using adjusted HR estimates.

The study results showed that ABI measurement can be used in addition to FRS to improve its predictive power, and the Authors suggested that a combined tool could be useful.

Siontis et al[27] (AMSTAR 9/11) performed a comparison of eight RPMs for CVD. Their review included 20 prospective and retrospective studies, with sample sizes ranging from 403 to 1 072 800. The main outcomes considered were CVD mortality and CVD-related incidence. The probability for prediction of outcome varied significantly among the studies, from poor (0.55) to good (0.85).

Damen et al[28] (AMSTAR: 7/11) conducted a systematic review examining 212 studies that described the development of 363 different prediction models for CVD and CHD. Sample size was extremely variable, ranging between 51 and 1 189 845 people, mainly from Europe, Canada and the USA.

Measures of predictive performance were reported in 53% only of the studies, with discriminatory ability from 0.61 to 1.00.

In addition, an external validation test was performed on 136 articles and most often concerned four models: Framingham, SCORE, QRISK and Adult Treatment Panel (ATP III).

The median discriminative ability was always acceptable (0.70–0.79), except for the ATPIII score (C-statistic index: 0.66). Calibration was estimated as observed: expected ratio, ranged from 0.59 of Framingham-Wilson to 0.94 of QRISK.

Beswick et al[29] (AMSTAR 11/11) included 30 articles that evaluated several risk prediction methods for CHD and CVD: 16 studies using convergent validation of Framingham-Anderson-based methods and 21 comparisons used different risk scoring methods. The enrolled samples involved 4540 to over 205 000 people, aged 5–70 years, from USA, Australia, Europe and India. Incidence and mortality for CHD and CVD were estimated as primary outcomes.

Only the most recent updates to the Sheffield tables and the Joint British charts showed acceptable sensitivity and specificity compared with the Framingham-Anderson model.

In addition, Beswick et al performed a second systematic review of external validation of Framingham-based risk scoring methods, based on 62 longitudinal or cross-sectional studies conducted on 112 different populations.

The results indicated extreme variability in discriminatory ability, with areas under the curve ranging from not acceptable (0.58) to good (0.85), the results were better in women than in men or in people with more recent baseline examinations.

Concerning calibration, the predicted:observed ratios ranged from an underprediction of 0.43 to an overprediction of 2.87. Generally speaking, underprediction was greater in people at higher risk, such as subjects with a family history of premature CVD, and lower in people at lower risk.

Echouffo-Tcheugui et al[30] (AMSTAR: 6/11) focused on 13 studies that evaluated 28 heart failure RPMs. Studies were based on a US and European cohort of 725 to 3 59 947 subjects, over 18 years of age. Assessed scores had acceptable-to-good discriminatory ability, with C-statistics ranging between 0.71 and 0.87.

Calibration, when reported, was generally acceptable. Only two models were externally validated and showed modest-to-acceptable discrimination, with C-statistics from 0.61 to 0.79.

Damen et al[31] (AMSTAR:9/11) included 38 studies and compared the performance of the Framingham ATP III, the Framingham Wilson model and the pooled cohort equations (PCE) for fatal or non-fatal CHD (Framingham Wilson and ATP III) and hard atherosclerotic CVD (PCE). Results for men and women were compared separately. The authors performed meta-analyses of the included studies, calibration was assessed through the observed versus expected (OE) ratio and discriminative power, through the C-statistic, for 10-year risk predictions. The OE ratio results were very heterogeneous, ranging from 0.58 to 0.79. C-statistic values were highly variable as well (from 0.58 to 0.82). Most of the studies showed overprediction of the expected events, especially in high-risk groups. According to the authors, RPMs for CVDs and CHDs showed a similar performance.

Studies on hypertension

Sun et al[32] (AMSTAR 9/11) included 26 cohort studies on hypertension that assessed 48 risk models and included both traditional risk factors—body mass index, age, smoking, blood pressure level and parental history of hypertension—with biochemical parameters and genetic

factors. Evaluated articles were based on samples drawn mainly from populations in the USA, Eastern Asia and the UK and included a population aged over 20, with a sample size that ranged from 443 to 17.471. All the studies included reported a C-statistic index ranging from 0.74 to 0.79, indicating a good discriminatory capability. Furthermore, calibration estimates of most studies by Hosmer-Lemeshow test showed no significant results.

Echouffo-Tcheugui et al 2013[33] (AMSTAR 9/11) assessed 11 prospective cohort studies that evaluated 15 different risk models for hypertension in population samples from 1135 to 11 407 subjects from US and Eastern Asian populations. Incidence of hypertension was considered as primary outcome. The C-statistic ranged from 0.80 to 0.70, indicating good performance and discrimination. Ten models also estimated calibration, using the Hosmer-Lemeshow test, and generally reported good calibration.

DISCUSSION AND CONCLUSIONS
Summary of evidence
Developing a good predictive score to enable early identification of diabetes, hypertension and CVDs are the major public health concern across countries of all income levels because of the extremely high rates of incidence and prevalence of these diseases, their upward trend worldwide and their massive consumption of social, health and economic resources.

In spite of the amount of evidence on the issue, the average quality of the existing primary studies is poor: they lack external validation, model calibration and standardised study design, and suffer from optimism bias. In fact, a number of searches showed that older and more limited RPMs performed better than newer, more complex models.[27]

The majority of studies, in particular those predicting diabetes, reported comparisons that were often achieved with scores that were very similar in prediction model tools, including those differing by a very small number of items (sometimes only one) and those focused on the same population. It is, therefore, not surprising that no RPMs on diabetes and hypertension have seemed to excel: no significant difference was found in the majority of studies on these two diseases.

Conversely, some promising differences among prediction tools were highlighted for CVDs and CHDs. The new RPMs investigated generally used Framingham scores as the main comparison tool.

According to Fowkes et al,[26] the ABI, in association with the Framingham tool, improved performance results, although only slightly. In addition, QRISK scores provided some evidence of superiority compared with Framingham, in particular in the areas of calibration and discrimination performance.[28] However, it should be pointed out that Framingham-based methods underestimated risk in diabetics, socioeconomically deprived populations, and in patients with a strong family history of premature CVD.[29] Because of the limitations described

in the available studies, and because no predictive model was clearly identified as superior, it seems legitimate to question whether investing in new risk models is still a good practice, or if it would be a better approach to focus our efforts on external validation of existing tools.

Strengths and limitations
To the best of our knowledge, this is the most comprehensive umbrella systematic review on RPMs for NCDs, such as diabetes and CVDs, with high incidence, prevalence and mortality worldwide. In fact, no other umbrella systematic reviews are available on this issue. The only umbrella systematic review found[34] was focused solely on hypertension. For this review, topics were selected according to the following criteria:
1. Epidemiological relevance in terms of incidence and prevalence.
2. The significant link between Diabetes, CVD and hypertension in terms of pathogenic pathway and clinical presentation.

This is also the reason why cancer was not considered among the inclusion criteria.

Many studies had been conducted on NCDs, in particular during the last decade, and the authors have, therefore, chosen to use an umbrella methodology for systematic reviews and meta-analyses. Unfortunately, available studies, although reported to be of medium-to-high quality according to AMSTAR score (mean 8.07 out of 11, ranging from 5 to 11), were based on primary studies of debatable quality, with a large proportion of them lacking discrimination and calibration assessments.

Due to the significant heterogeneity of study designs, the risk models involved and the outcomes reported, it was not possible to conduct a meta-analysis. The results were, therefore, reported narratively.

Our electronic search was limited to studies in English and to three major international databases (MEDLINE/PubMed, Scopus and Cochrane Library), with additional works derived from the reference list of studies, and did not include grey literature with unpublished documents. However, a publication bias was estimated by the quantification of ongoing and non-completed systematic reviews in the PROSPERO database.

No assessment of potential adverse effects of RPMs has been carried out, with a potential risk of bias.

General interpretation of results
The increasing global growth in prevalence of chronic diseases, as a direct consequence of epidemiological transition, has led to broader use of predictive tools as a major aid for health workers. Indeed, these instruments can be very important and should be regularly implemented in medical settings to support the activity of general practitioners and public health authorities involved in monitoring and evaluation of patients. Specific benefits of RPMs could emerge in prevention and health promotion for specific populations—such as workers and students—and social settings.

It must be pointed out that the studies evaluated in this systematic review, although of medium-to-high quality, are not primary studies, and therefore, could be affected by significant bias. Furthermore, there is no evidence in the scientific literature for the evaluation of the effectiveness of RPMs on long-term patient outcomes.[35] Therefore, the results from this study should be carefully applied by health workers, in order to minimise the risk of over or undertreatment.[36]

Scientific literature in the past 30 years has produced an abundance of evidence on other powerful health determinants, such as social relationships networks, stress, unemployment, education and income,[37–40] however, none of these variables have been included in all the available predictive tools. Moreover, very few instruments considered lifestyle variables like smoking, alcohol, physical activity and drug use or addiction. A strictly biological perspective should be considered as a serious limitation in terms of forecasting and predicting the development of CVDs, CHDs, diabetes and hypertension. A new generation of predictive tools, conceptually developed around biological and non-biological determinants, could consistently ameliorate the assessment of risk and the detection of risk stratification groups.

Conclusions and future perspectives

The wide range of available studies that have tested RPMs for CVDs, CHDs, hypertension and diabetes compare almost overlapping tools (which often differ by only a single entry), does not really increase our knowledge of the issue; rather it merely increases uncertainty.

More precise evidence is available only for CVD prediction: the Framingham score, alone or in combination with the ABI, and QRISK score can be confirmed as the gold standard.

Further efforts should not be concentrated on creating new scores, but rather on performing external validation of the existing ones. Promising future possibilities could then involve testing risk scores on wider samples and on certain target populations, such as workers, with specific exposure risks and for which no robust scientific evidence is currently available. These individuals could definitely benefit from early detection of chronic disease, since the conditions are often worsened by occupational exposure and result in disability and absence from work. Benefits could be further improved by supplementing existing models with information on lifestyle, personal habits and family history,[23] social network relationships, income, education and employment history.

ORCID iDs
Leonardo Palombi http://orcid.org/0000-0003-1806-5278
Antonio Vinci http://orcid.org/0000-0002-4915-7733

REFERENCES

1. Health statistics and information systems. Disease burden and mortality estimates. Cause-specific mortality, 2000–2016. Available: http://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html [Accessed 17 Jan 2019].
2. Mills KT, Bundy JD, Kelly TN, *et al*. Global disparities of hypertension prevalence and control: a systematic analysis of population-based studies from 90 countries. *Circulation* 2016;134:441–50.
3. Global Health Observatory (GHO) data. Raised blood pressure. Situation and trends. Available: http://www.who.int/gho/ncd/risk_factors/blood_pressure_prevalence_text/en/ [Accessed 18 Jun 2018].
4. WHO - Health statistics and information systems. Available: https://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html [Accessed 17 Jan 2019].
5. Shaw LJ, Goyal A, Mehta C, *et al*. 10-Year Resource Utilization and Costs for Cardiovascular Care. *J Am Coll Cardiol* 2018;71:1078–89.
6. American Diabetes Association. Economic costs of diabetes in the U.S. in 2017. *Diabetes Care* 2018;41:917–28.
7. High Blood Pressure. Blood pressure: make control your goal Infographic. Available: https://www.cdc.gov/bloodpressure/infographic.htm [Accessed 18 Jun 2018].
8. Mehta S, Ghosh S, Sander S, *et al*. Differences in all-cause health care utilization and costs in a type 2 diabetes mellitus population with and without a history of cardiovascular disease. *J Manag Care Spec Pharm* 2018;24:280–90.
9. Kung H-C, Xu J. Hypertension-related mortality in the United States, 2000-2013. *NCHS Data Brief* 2015:1–8.
10. Gupta R, Gupta VP, Prakash H, *et al*. 25-Year trends in hypertension prevalence, awareness, treatment, and control in an Indian urban population: Jaipur heart Watch. *Indian Heart J* 2018;70:802–7.
11. Bhatnagar P, Wickramasinghe K, Wilkins E, *et al*. Trends in the epidemiology of cardiovascular disease in the UK. *Heart* 2016;102:1945–52.
12. European Heart Network. European cardiovascular disease statistics 2017. Available: http://www.ehnheart.org/CvD-statistics/CvD-statistics-2017.html [Accessed 18 Jun 2018].
13. Petrie JR, Guzik TJ, Touyz RM. Diabetes, hypertension, and cardiovascular disease: clinical insights and vascular mechanisms. *Can J Cardiol* 2018;34:575–84.
14. Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
15. Sharif MO, Janjua-Sharif FN, Ali H, *et al*. Systematic reviews explained: AMSTAR-how to tell the good from the bad and the ugly. *Oral Health Dent Manag* 2013;12:9–16.
16. The Joanna Briggs Institute Reviewers' Manual 2014 edition / supplement methodology for JBI umbrella reviews. Available: http://joannabriggs.org/assets/docs/sumari/ReviewersManual-Methodology-JBI_Umbrella%20Reviews-2014.pdf [Accessed 17 Jan 2019].
17. Whiting P, Savović J, Higgins JPT, *et al*. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;69:225–34.
18. 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.
19. Barber SR, Davies MJ, Khunti K, *et al*. Risk assessment tools for detecting those with pre-diabetes: a systematic review. *Diabetes Res Clin Pract* 2014;105:1–13.
20. Collins GS, Mallett S, Omar O, *et al*. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med* 2011;9:103.
21. Hu PL, Koh YLE, Tan NC. The utility of diabetes risk score items as predictors of incident type 2 diabetes in Asian populations: an evidence-based review. *Diabetes Res Clin Pract* 2016;122:179–89.

22 Noble D, Mathur R, Dent T, *et al*. Risk models and scores for type 2 diabetes: systematic review. *BMJ* 2011;343:d7163.

23 Yoshizawa S, Kodama S, Fujihara K, *et al*. Utility of nonblood-based risk assessment for predicting type 2 diabetes mellitus: a meta-analysis. *Prev Med* 2016;91:180–7.

24 Cortes-Bergoderi M, Thomas RJ, Albuquerque FN, *et al*. Validity of cardiovascular risk prediction models in Latin America and among Hispanics in the United States of America: a systematic review. *Rev Panam Salud Publica* 2012;32:131–9.

25 Tzoulaki I, Liberopoulos G, Ioannidis JPA. Assessment of claims of improved prediction beyond the Framingham risk score. *JAMA* 2009;302:2345–52.

26 Fowkes FGR, Murray GD, Butcher I, *et al*. Ankle brachial index combined with Framingham risk score to predict cardiovascular events and mortality: a meta-analysis. *JAMA* 2008;300:197–208.

27 Siontis GCM, Tzoulaki I, Siontis KC, *et al*. Comparisons of established risk prediction models for cardiovascular disease: systematic review. *BMJ* 2012;344:e3318.

28 Damen JAAG, Hooft L, Schuit E, *et al*. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016;353.

29 Beswick AD, Brindle P, Fahey T, *et al*. *A systematic review of risk scoring methods and clinical decision aids used in the primary prevention of coronary heart disease (supplement) [Internet]*. London: Royal College of General Practitioners, UK: National Institute for Health and Clinical Excellence: Guidance, 2008.

30 Echouffo-Tcheugui JB, Greene SJ, Papadimitriou L, *et al*. Population risk prediction models for incident heart failure: a systematic review. *Circ Heart Fail* 2015;8:438–47.

31 Damen JA, Pajouheshnia R, Heus P, *et al*. Performance of the Framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis. *BMC Med* 2019;17:109.

32 Sun D, Liu J, Xiao L, *et al*. Recent development of risk-prediction models for incident hypertension: an updated systematic review. *PLoS One* 2017;12:e0187240. eCollection 2017.

33 Echouffo-Tcheugui JB, Batty GD, Kivimäki M, *et al*. Risk models to predict hypertension: a systematic review. *PLoS One* 2013;8:e67370.

34 Pocock SJ, McCormack V, Gueyffier F, *et al*. A score for predicting risk of death from cardiovascular disease in adults with raised blood pressure, based on individual patient data from randomised controlled trials. *BMJ* 2001;323:75–81.

35 Collins DRJ, Tompson AC, Onakpoya IJ, *et al*. Global cardiovascular risk assessment in the primary prevention of cardiovascular disease in adults: systematic review of systematic reviews. *BMJ Open* 2017;7:e013650.

36 Studziński K, Tomasik T, Krzysztoń J, *et al*. 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;19:11.

37 Forouzanfar MH, Afshin A, Alexander LT, *et al*. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the global burden of disease study 2015. *The Lancet* 2016;388:1659–724.

38 Kontis V, Mathers CD, Bonita R, *et al*. Regional contributions of six preventable risk factors to achieving the 25 × 25 non-communicable disease mortality reduction target: a modelling study. *Lancet Glob Health* 2015;3:e746–57.

39 Loef M, Walach H. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. *Prev Med* 2012;55:163–70.

40 Holt-Lunstad J, Smith TB, Layton JB, *et al*. Social relationships and mortality risk: a meta-analytic review. *PLoS Med* 2010;7:e1000316.