diabetesresearchandclinicalpractice 209 (2024) 111574

Contents lists available at ScienceDirect

Diabetes Research and Clinical Practice

journal homepage: www.journals.elsevier.com/diabetes-research-and-clinical-practice

A literature review of quality assessment and applicability to HTA of risk prediction models of coronary heart disease in patients with diabetes

Li Jiu [a], Junfeng Wang [a], Francisco Javier Somolinos-Simón [b], Jose Tapia-Galisteo [b,c], Gema García-Sáez [b,c], Mariaelena Hernando [b,c], Xinyu Li [a,d], Rick A. Vreman [a,e], Aukje K. Mantel-Teeuwisse [a], Wim G. Goettsch [a,e,*]

[a] Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, Netherlands
[b] Bioengineering and Telemedicine Group, Centro de Tecnología Biomédica, ETSI de Telecomunicación, Universidad Politécnica de Madrid, Parque Científico y Tecnológico de la UPM, Crta. M40, Km. 38, 28223 Pozuelo de Alarcón, Madrid, Spain
[c] CIBER-BBN: Networking Research Centre for Bioengineering, Biomaterials and Nanomedicine, Parque Científico y Tecnológico de la UPM, Crta. M40, Km. 38, 28223 Pozuelo de Alarcón, Madrid, Spain
[d] University of Groningen, Faculty of Science and Engineering, Groningen Research Institute of Pharmacy, Broerstraat 5, 9712 CP Groningen, the Netherlands
[e] National Health Care Institute (ZIN), Diemen, Willem Dudokhof 1, 1112 ZA Diemen, Netherlands

ARTICLE INFO

Keywords:
Risk prediction model
Coronary heart disease
Diabetes
Health technology assessment
Systematic review

ABSTRACT

This literature review had two objectives: to identify models for predicting the risk of coronary heart diseases in patients with diabetes (DM); and to assess model quality in terms of risk of bias (RoB) and applicability for the purpose of health technology assessment (HTA). We undertook a targeted review of journal articles published in English, Dutch, Chinese, or Spanish in 5 databases from 1st January 2016 to 18th December 2022, and searched three systematic reviews for the models published after 2012. We used PROBAST (Prediction model Risk Of Bias Assessment Tool) to assess RoB, and used findings from Betts et al. 2019, which summarized recommendations and criticisms of HTA agencies on cardiovascular risk prediction models, to assess model applicability for the purpose of HTA. As a result, 71 % and 67 % models reporting C-index showed good discrimination abilities (C-index >= 0.7). Of the 26 model studies and 30 models identified, only one model study showed low RoB in all domains, and no model was fully applicable for HTA. Since the major cause of high RoB is inappropriate use of analysis method, we advise clinicians to carefully examine the model performance declared by model developers, and to trust a model if all PROBAST domains except analysis show low RoB and at least one validation study conducted in the same setting (e.g. country) is available. Moreover, since general model applicability is not informative for HTA, novel adapted tools may need to be developed.

1. Introduction

Coronary heart disease (CHD) is a heart disease featured by narrowing or blockage of coronary arteries [1,2]. CHD is one of the major complications of diabetes mellitus (DM), and it is diagnosed in more than one-fifth of patients with type-2 diabetes across all socioeconomic statuses [3]. Previous reviews show that CHD is the leading cause of diabetes mortality, and it doubles the economic burden of patients with diabetes [4,5].

To reduce CHD morbidity, mortality, and relevant costs in DM patients, early recognition of high CHD risks and appropriate selection of prevention strategies are highly needed [6,7]. To achieve this, clinicians need to take into account as many risk factors as possible, including the traditional ones (e.g. age and gender) and those gaining popularity (e.g. C-reactive protein and coronary artery calcium) [8]. Due to the existence of a number of risk factors, risk prediction models have been

Abbreviations: DM, Diabetes mellitus; CHD, Coronary heart disease; MI, Myocardial infarction; ACS, Acute coronary syndrome; HTA, Health technology assessment; ML, Machine learning; RoB, Risk of bias; PROBAST, Prediction model Risk Of Bias Assessment Tool; CHARMS, Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology; TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis; SQ, Signaling question.

* Corresponding author at: Universiteitsweg 99, 3584 CG Utrecht, Netherlands.
E-mail address: W.G.Goettsch@uu.nl (W.G. Goettsch).

https://doi.org/10.1016/j.diabres.2024.111574
Received 15 September 2023; Received in revised form 17 January 2024; Accepted 6 February 2024
Available online 10 February 2024

widely used by clinicians to take into account all relevant risk factors when estimating probabilities of the occurrence of CHD in DM patients [9,10].In addition, risk prediction models, especially those developed for estimating outcomes occurring within a specified time frame (i.e. prognostic model) [11], can be applied for the purpose of health technology assessment (HTA). More specifically, by functioning as a part of a microsimulation cost-effectiveness model, risk prediction models allow estimation of resource use of a healthcare intervention and estimation of heterogeneous costs and outcomes among different subpopulations of interest [12–15]. Such models can be utilized to inform clinicians and patients on the cost burden of an intervention within a time frame (e.g. 10 years), or to inform HTA agencies on whether costs of a healthcare intervention should be reimbursed [16,17].

As prognostic risk prediction models for DM patients started to emerge, the variety in techniques used for developing the models increased as well. The most frequently used technique is statistical modelling, which can be further categorized into Cox regression analysis, Logistic regression analysis, Weibull regression analysis, Gompertz regression analysis, etc. [18,19]. Another modelling technique that emerges is machine learning (ML), including but not limited to neural networks, random forest, decision-tree, support vector machine, etc. [19]. ML models are gaining popularity in the field of diabetes due to their capability to capture the complex relationships among a vast number of predictors from different data sources [20,21].

However, the ever-increasing number of models and variety of modelling techniques have placed a heavy burden on model evaluation and raised concerns about model quality in terms of risk of bias (RoB) and applicability. A high RoB, which is pervasive among risk prediction models [21], could increase the likelihood that models yield inaccurate prediction, and decrease the confidence of users (e.g. clinicians, HTA researchers and agencies, patients, etc.) in model performance [22]. In addition, model users are at risk of selecting and applying a suboptimal model for their own purposes, as they often miss information on therapeutic, geographic, or temporal settings in which the models can be applied [23].

To assess quality of CHD risk prediction models for DM patients, Van Dieren et al. conducted a systematic review to summarize the structure and predictive performance of existing models for predicting type 2 diabetes published before May 2012 [24]. More recently, Galbete et al. updated the review conducted by Van Dieren et al. by searching for the models published before July 2021 [25]. They summarized model performance and assessed RoB and generic model applicability. However, these two reviews adopted similar search strategies using a single data source (either Medline or PubMed) and did not systematically search for risk prediction models developed with ML techniques. Also, they did not provide an assessment of model applicability for the purpose of HTA, which requires special considerations, such as appropriateness of subgroup populations [26].

Hence, the aim of our study was to identify the prognostic risk prediction models developed recently with statistical or machine learning techniques, and to assess their RoB and applicability for HTA. This research was performed as part of the HTx project [27]. The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825162.

2. Material and methods

2.1. Protocol

A research protocol of this study was registered in the PROSPERO database (ID CRD42021273240), then rigorously followed. To conduct the systematic review, we followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [28] and the guidance developed by Debray et al. 2017 [29].

2.2. Data source & search strategy

We searched PubMed, Embase, Scopus, Web of Science, and IEEE Xplore for journal articles and conference papers predicting CHD risks in DM patients, in two rounds (published from 1st January 2016 to 31st May 2021; published from 1st June 2021 to 18th December 2022). We used a search strategy (***Appendix 1***) with three concepts (i.e. risk prediction, CHD, and diabetes), which was developed from published strategies to retrieve relevant publications for CHD [30,31] or prediction models [32,33]. The search strategy was developed by two reviewers (LJ & JW), then edited by an experienced librarian in document retrieval from Utrecht University. We also checked citations in all identified relevant studies. In addition to the database search, we searched three recently published systematic reviews (i.e. Galbete et al. 2022 [25], Faizal et al. 2021 [34], and Lenselink et al. 2022 [35]) which identified models predicting the risk of cardiovascular diseases, in DM patients or general population.

2.3. Eligibility criteria

A study was eligible if (1) it described the development of a prediction model; (2) the target population was patients with diabetes; (3) the outcome of prediction was CHD or a CHD component (i.e. myocardial infarction, acute coronary syndrome, and/or angina); (4) the study was published in English, Dutch, Spanish, or Chinese, based on the review team's language proficiencies; (5) the study was published after 2012;(6) the full-text was available. Exclusion criteria included non-human studies or studies only describing model validation. Studies using heart or cardiovascular disease as a combined outcome only were also excluded because the risk of CHD could not be predicted.

2.4. Study selection & data collection

For study selection, one reviewer (LJ) screened titles and abstracts of all identified studies, while another (GGS) independently screened a random set of 20 %. Then two reviewers (LJ and FJSS or JTG) independently scanned full texts of studies that might be eligible. Any disagreement between reviewers was solved through consensus. For each model identified, one reviewer (LJ) extracted model characteristics (e.g. target population, outcome, and number of predictors, etc.), based on a data collection form developed from the CHARMS (CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) [36].

2.5. Quality assessment

For assessing RoB, several appraisal tools can be used, such as the PROBAST (Prediction model Risk Of Bias Assessment Tool) [22], TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) [37], and STROBE (Strengthening the Reporting of Observational studies in Epidemiology) [38]. We determined to use the PROBAST, because it was carefully and specifically designed for evaluating RoB of prediction models. It consists of four RoB domains (i.e. participants, predictors, outcomes, and analysis), and each domain-specific RoB is graded as low, high, or unclear.

For assessing model applicability for the HTA purpose, we did not use PROBAST, because it mainly addressed applicability concerns regarding medical settings, i.e., whether population, predictors, or outcomes of a study differed from those specified in a systematic review question [22]. Additionally, we did not find any specifically designed tool but only a review conducted by Betts et al. [26], which summarized reasons why HTA agencies recommended or criticized models for predicting cardiovascular diseases. Betts et al. mentioned three aspects of concern regarding applicability for the purpose of HTA, including geographic and therapeutic generalizability, whether the model was up-to-date, and appropriateness of model covariates. According to Betts

et al., seven signaling questions (SQs) were formulated by two reviewers (LJ and JW), and then edited by five reviewers (GGS, FJSS, JTG, XL, LJ) after a pilot quality assessment of one-third of the eligible models. The questions and their rationales are attached in ***Appendix 2***.

Quality assessment was conducted independently by two reviewers (LJ and FJSS, JTG, or XL), and any discrepancy was solved through discussion with at least three reviewers. Before the formal RoB assessment, two training sessions with three example modelling studies (e.g. Covid-19) were conducted by six reviewers (JW, LJ, FJSS, JTG, GGS).

2.6. Data analysis

For data analysis, we narratively synthesized characteristics of the eligible models by presenting all variables as numbers and percentages, and summarized the model discrimination (i.e. C-index) reported by the identified studies. We did not conduct a *meta*-analysis to summarize model performance, because a *meta*-analysis is only needed when the aim is to summarize the predictive performance of the model being validated across different settings and populations [29]. In the current review, we focused on the methodological quality of model development studies, and model validation studies were out of the scope. The results were presented separately for ML and statistical models in both tables and graphs.

3. Results

3.1. Study selection

Among a total of 12,784 records identified from the five databases, 1381 were eliminated as duplicates, leaving a total of 11,403 initial records (Fig. 1). After adding records from the three published reviews and reviewing titles and abstracts, we selected 183 records for full-text screening, then excluded 157 records with the reasons such as inappropriate population (n = 58). No new references were obtained through the reference lists of the remaining articles. Therefore, 26 studies, which described 30 models, were included for data extraction. Twenty-one, three, and two of the studies were identified from database search only, published reviews only, or both. Since RoB of the three studies [39–41], which were identified from the published reviews, had been previously assessed [24,42], we included 23 model studies for RoB assessment and all the 30 models for applicability assessment.

3.2. Model characteristics

The summary of study and model characteristics is presented in Table 1, and ***the reference list and model*** details can be found in ***Appendix 3 & 4***. The model discrimination, which was reported as C-index by the identified studies, was listed in Appendix 5.

Regarding model characteristics, models developed with the statistical approaches, such as Cox (17, 57 %) and Logistic (4, 13 %), accounted for the most, while the other five used various ML techniques, that is, Multi-task Learning [43], Random Forest [44], Neural Network [45], Recurrent Neural Network Gated Recurrent Unit [46], and K-nearest Neighbor models [47]. Regarding outcomes, about half of the models predicted CHD, while other predicted MI or acute coronary syndrome. Also, only 16 (53 %) models reported the duration of risk they could predict (e.g. 5-year CHD risk), and only four models provided the equations to predict the annual risk. In addition, 11 variables were included as final predictors in at least three models, and the most common predictors were age, history of disease, gender, and smoke. Regarding model discrimination (Appendix 5), 15 (71 %) and four (67 %) models reporting the C-index showed good discrimination (C-index >= 0.7).

3.3. RoB assessment - PROBAST

The quality assessment in terms of RoB is shown in Fig. 2. In ***Appendix 6 & 7***, the results were split if the model studies were developed with statistical or ML techniques. In summary, only one model, i.e., the model by Quan et al. 2019 [48], was graded as having low RoB in all the four PROBAST domains, and seven models were graded as having low RoB in all domains except the analysis.

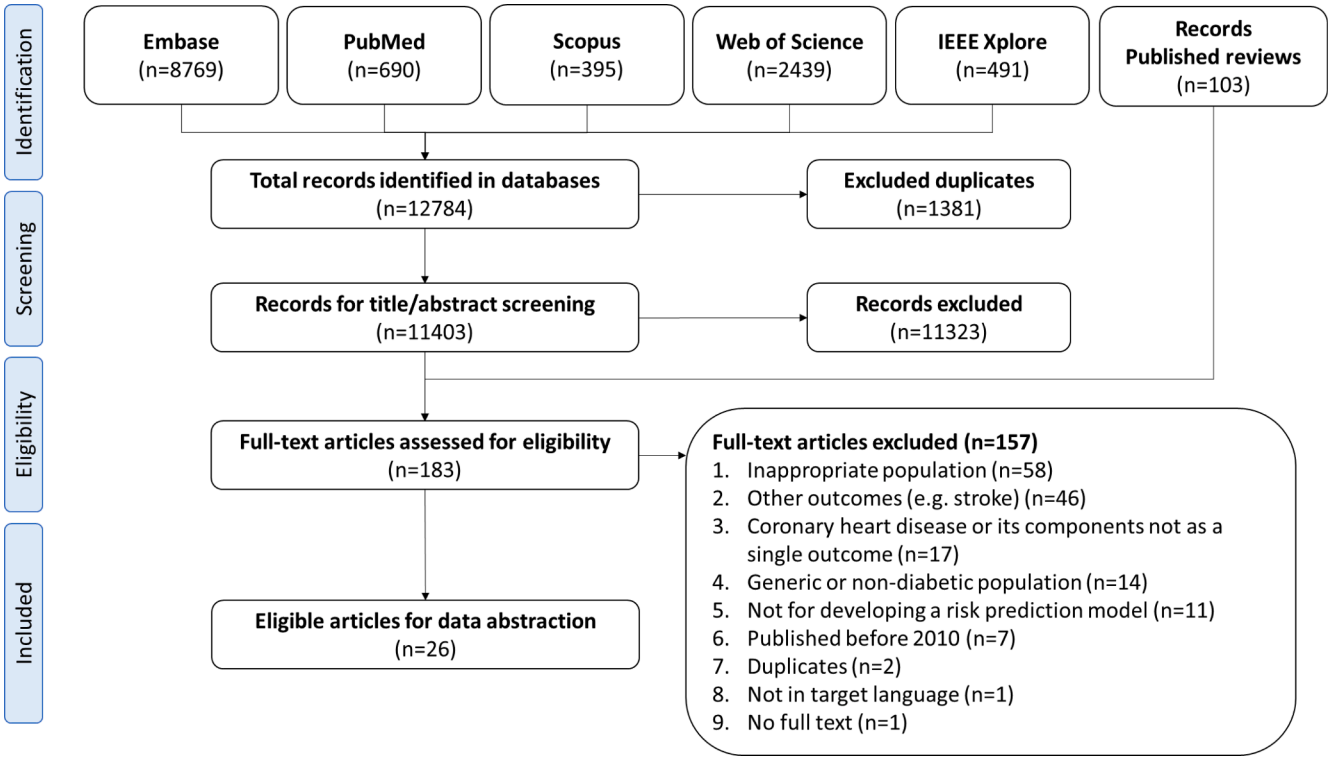

Fig. 1. Flowchart of included studies.

Table 1
Characteristics of eligible models.

	Number	Percentage (%)
Study characteristics (n = 26)		
Study design		
Observational study	21	81
Trial + Observational study	4	15
Trial	1	4
Disease		
Type-2 diabetes	22	85
Type-1 diabetes	1	4
All diabetes	3	12
Region		
Asia	12	46
North America	8	31
Europe	5	19
Intercontinental	1	4
Whether a reporting guideline was followed?		
Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD)	2	8
NR	24	92
Handling of continuous predictors		
Continuous	8	31
Continuous or categorized	10	38
Not reported	8	31
Non-linearity		
Restricted cubic splines	2	8
Logarithm transformation,linear plus splines, or quadratic functions	1	4
Continuous spline function	1	4
Not reported	22	85
Handling of missing data		
Complete case analysis	11	42
Multiple imputation	7	27
Not reported	8	31
Method for selection of predictors for inclusion in multivariable modelling		
All candidate variables	18	69
Based on literature	3	12
Based on literature & expert opinion	1	4
Based on previous models	1	4
Not reported	2	8
Method for selection of predictors during multivariable modelling		
Full model approach	7	27
Backward selection	5	19
Forward selection	4	15
Not reported	10	38
Method of internal validation		
Cross-validation	6	23
Bootstrapping	5	19
Sample split	7	27
Sample split & Cross-validation	1	4
Not reported	7	27
Model calibration		
Calibration plot & Hosmer-Lemeshow test	8	31
Calibration plot	6	23
Hosmer-Lemeshow test	1	4
Not reported	11	42
Model discrimination		
C-statistic	12	46
C-statistic & Area under the receiver operating characteristic curve	4	15
Area under the receiver operating characteristic curve	3	12
C-statistic & Bayesian information criteria	1	4
Not reported	6	23
Model presentation		
Hazard ratio	9	35
Full equation	9	35
Odds ratio	2	8
Not reported	6	23
Alternative model presentation		
Nomogram	4	15
Sum score	3	12
Web server	2	8
Nomogram & Web server	2	8
Not reported	15	58
Analysis done in reviewed study		
Development and internal validation	14	54
Development and external validation	8	31
Development only	4	15
Median sample size (development)	5521 (74—891095)	
Median sample size (external validation)	3803 (962—386425)	
Model characteristics (n = 30)		
Model type		
Machine learning[a]	5	17
Cox	17	57
Logistic	4	13
Other (e.g. Linear)	4	13
Outcome of interest[b]		
CHD	16	53
MI	11	37
ACS	3	10
Age of simulated individuals		
All	17	57
With a range (e.g. 40–64)	8	27
Not reported	5	17
Time cycle of prediction		
> 1 year	12	40
1 year at maximum	4	13
Not reported	14	47
Events per variable		
>=20	18	60
<20	6	20
Not reported	6	20
Final predictors included by at least 3 models		
Age	20	67
History of disease (e.g. cardiovascular disease)	17	57
Gender	14	47
Smoke	14	47
HbA1c	13	43
Cholesterol	13	43
Blood pressure	12	40
Diabetes duration	7	23
Use of drugs (e.g. Statins)	7	23
Ethnicity	6	20
Body mass index	5	17

[a] *"Machine learning" indicates the machine learning techniques used for developing the included prediction models, including Multi-task Learning (MTL), Random Forest (RF), Neural Network (NN), and Recurrent Neural Network Gated Recurrent Unit (RNN GRU).*

[b] *"CHD" indicates coronary heart disease, "MI", myocardial infarction, "ACS", acute coronary syndrome.*

In the *Participants* domain (SQ 1.1–1.2), three (60 %) ML model studies and 12 (67 %) statistical model studies were rated as having low RoB. Appropriate data sources, such as prospective cohort and randomized controlled trial data, were used (SQ 1.1) in four (80 %) ML model studies and 16 (89 %) statistical model studies. The included patients were considered representative of the target population (SQ 1.2) in four (80 %) ML model studies and 13 (72 %) statistical model studies. A total of four models included patients who were already known to have the CHD-related outcomes at the time of predictor measurement (e.g. patients with CHD history), and one model excluded sicker patients based on number of hospitalization [46]. Consequently, the predictive performance of these models could be overestimated or underestimated, respectively.

In the *Predictors* domain (SQ 2.1–2.3), four (80 %) ML model studies and 15 (83 %) statistical model studies were rated as low RoB. Predictors were defined and assessed in a similar way for all participants (SQ 2.1) in four (80 %) ML model studies and 17 (94 %) statistical model studies. Predictor assessments were made without knowledge of outcome data (SQ 2.2) in 21(91 %) studies. All predictors were considered available at the time the model is intended to be used (SQ 2.3) in all ML models, but only in 15 (83 %) statistical models. The remaining three statistical

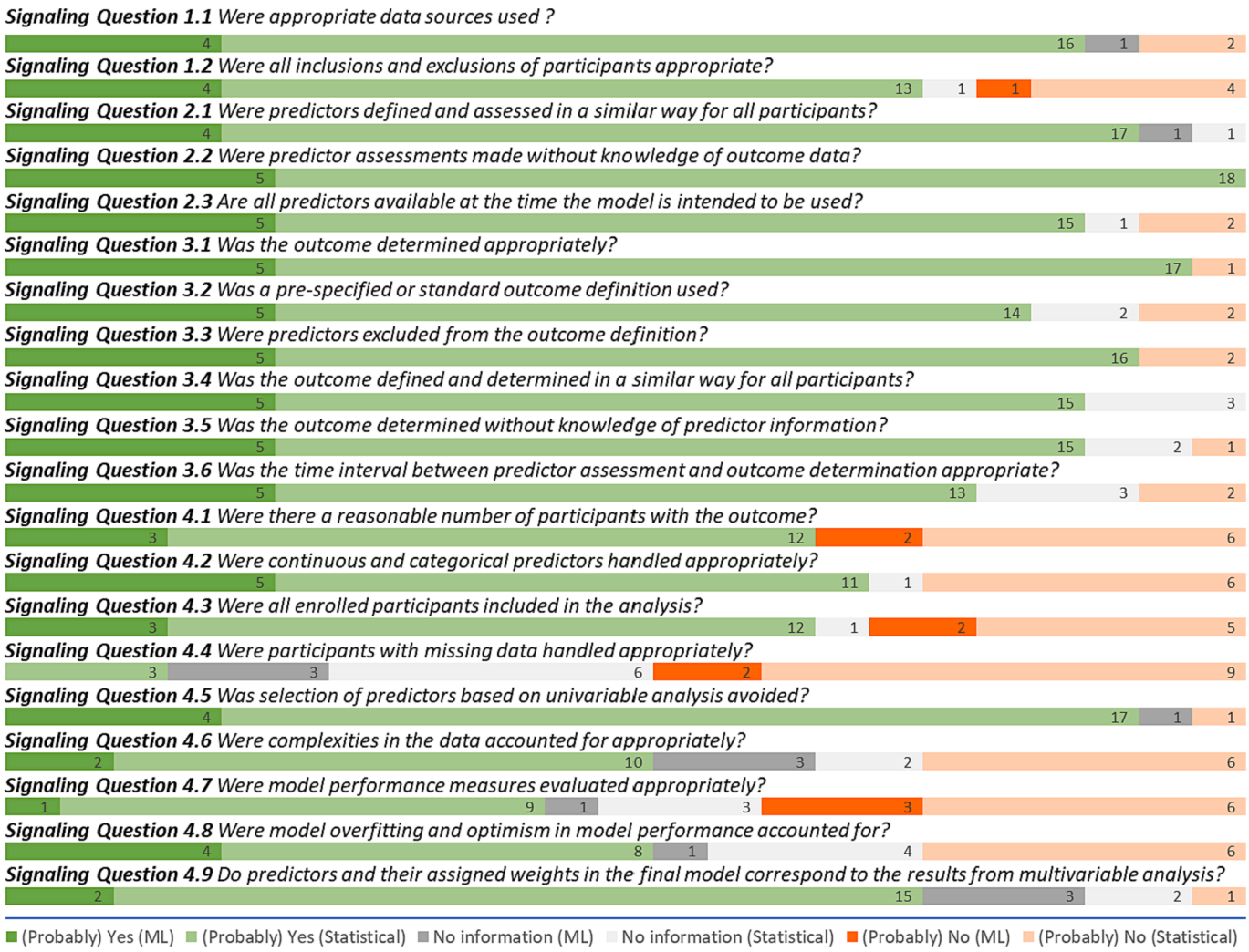

Fig. 2. PROBAST signaling questions for the 23 studies investigating models developed with statistical or machine learning techniques. *Note ML* indicates machine learning models; *Statistical* indicates statistical models. The signaling questions *1.1–1.2, 2.1–2.3, 3.1–3.6*, and *4.1–4.9* corresponds to the risk of bias domain of participants, predictors, outcome, and analysis, respectively.

models were considered high RoB, because two of them included predictors that were unlikely to be available in clinical practice (e.g. anthropometric measurement) [48,49], and one did not mention when the model would be used [50].

In the *Outcome* domain (SQ 3.1–3.6), all the ML model studies (100 %) were rated as low RoB, and were considered high-quality in Signaling questions from 3.1 to 3.6. Comparably, 10 (56 %) statistical model studies were rated as low RoB. Only one statistical model [41] did not use appropriate methods to determine the outcome, thus increasing the risk of misclassification (SQ 3.1). Similarly, only two statistical model studies [51,52] missed prespecified or standard definitions to determine the outcome (SQ 3.2). Likewise, predictors were mistakenly included in the outcome definition (SQ 3.3) in two statistical model studies [53,54]. Outcomes were defined and measured in a similar way (SQ 3.4) in 15 (83 %) statistical model studies, except three which provided no information [55–57]. According to SQ 3.5, prediction information was known only in one statistical model [53] when determining the outcome status. In SQ 3.6, the time interval between predictor assessment and outcome determination was considered too short in two statistical models [48,51].

In the *Analysis* domain (SQ 4.1–4.9), most concerns regarding RoB were identified. All the ML model studies and 16 (89 %) statistical model studies were rated as high RoB. SQ 4.7 showed that three (60 %) ML model studies and six (33 %) statistical model studies did not appropriately evaluate model performance, because they missed calibration evaluation [44,46,50,54,57–60], only used the Hosmer–Lemeshow test for calibration evaluation [61], or presented classification measures with predicted probability thresholds derived from the data set at hand [45]. According to SQ 4.4, two (40 %) ML model studies [43,47] and nine (50 %) statistical model studies handled missing data inappropriately by simply excluding them, while three (60 %) ML model studies [44–46] and six (33 %) statistical model studies suffered from no information. SQ 4.2 revealed that continuous and categorical predictors were handled appropriately in all ML model studies, but only in 11 (61 %) statistical models. Six (33 %) studies did not examine non-linearity for continuous variables [53,58,60,62] or categorized continuous variables [58,61]. Similarly, model overfitting and optimism were considered (SQ 4.8) in four (80 %) ML model studies, but only in eight (44 %) statistical model studies. Six (33 %) studies did not use internal validation techniques [50,58–62], or the validation did not include the whole model development procedures [48,51]. According to SQ 4.3, two (40 %) ML models [43,46] and five (28 %) statistical models [50,51,53,60,63] inappropriately excluded patients due to uninterpretable findings, outliers, or missing data. SQ 4.6 finally showed that none of the ML model studies but six statistical model studies inappropriately addressed censoring or competing risks. Three models simply used logistic regression for censoring [53,58,62] and three ignored competing risks [48,58,63]. Also, five studies provided no

information. In response to SQ 4.1, two ML model [44,47] and six (33 %) statistical models did not have a reasonable number of participants with the outcome (i.e. event per predictor parameter < 10).

The remaining questions contributed relatively less to the overall RoB. Only one model [61] selected predictors based on univariable analysis, and another [45] provided no information (SQ 4.5). Based on SQ 4.9, information was missing on whether predictors and their assigned weights in the final model correspond to the results from multivariable analysis in three (60 %) ML models [43,45,46] and two (11 %) statistical models [53,62]. Additionally, in only one statistical model [61], the predictors did not correspond to the results.

3.4. Applicability to HTA assessment

The assessment in terms of applicability for the purpose of HTA is shown in Fig. 3 and ***Appendix 8***.

In summary, the applicability of the models for HTA was quite limited, as none of the 30 models were fully applicable (i.e. "Yes" or "Probably Yes" in all the seven signaling questions). Only six models in three studies [39,63,64] had an "Yes" or "Probably Yes" in at least four signaling questions. The major barrier of model applicability was the lack of feasibility to calculate the annual risk of CHD or its component, either directly or indirectly (SQ f & SQ g), as only three models [39,64] could provide the option. The direct calculation indicates that, an equation or tool (e.g. an online user interface) is provided in the original study to calculate the annual risk of disease. The indirect calculation indicates that, though equations or a tool for predicting the annual risk are not provided, users could calculate the risk, using evidence provided in the original study (e.g. using hazard functions to calculate the accumulated risk). Another barrier of model applicability was inappropriate exclusion of major CHD risk factors as candidate predictors (SQ d),as only one model [50] considered all the factors as candidate predictors. The CHD risk factors or features refer to those defined by a recently published overview, Hajar 2017 [65], including blood pressure, high blood cholesterol levels, smoking, overweight or obesity, lack of physical activity, unhealthy diet and stress, age, gender, family history, and race. Additionally, external validation was performed within the same study in only one ML [43] and eight statistical models (SQ a). Although all the identified models were published from 2013 onwards, one ML [32] or 12 statistical models were considered relatively obsolete (SQ b), because all the follow-ups of their target populations ended before 2012. Additionally, about one-third of models were attached with examples on how these models could be used. For example, the model from the United Kingdom Prospective Diabetes Study (UKPDS) [39] and from Ye et al. 2022 [64] created an artificial patient with assumed value on its characteristics, and illustrated how the CHD-related risk was calculated using the model. Finally, according to SQ c, four (22 %) statistical models [50,55,56,62] included predictors that were not likely to be reported as trial outcomes, such as biomarkers.

4. Discussion

4.1. Findings

We conducted a literature review of prognostic models which predicted CHD risk in patients with diabetes to assess quality, in terms of risk of bias, and applicability for the purpose of HTA. We identified 25 statistical and five ML models, with overall relatively poor model quality. Only one study [48] showed low RoB in all domains of the PROBAST checklist, and none of the 30 models were fully applicable for HTA. The models with low RoB in at least three RoB domains and a user-friendly presentation (e.g. nomogram) included those by Quan et al. 2019, Choi et al. 2020, and Zhong et al. 2022 [47,48,61]. Additionally, the models with relatively wide HTA applicability included the UKPDS model, and the models by Hirai et al. 2019, and Ye et al. 2022 [39,63,64].

We discovered that most of the major contributors of high RoB were located in the analysis domain. Similar findings were also reported by Galbete et al., Haider et al., and Van der Heijden et al., who assessed RoB of 65, 14, and 16 models, respectively, for predicting the risk of cardiovascular disease or retinopathy in general or DM populations [22,42,66]. This finding is as expected, because the analysis domain, which addresses statistical considerations of model development, is the most complicated, with the most (n = 9) signaling questions [22]. We did not identify similar research that assessed the model applicability for the HTA purpose. Since use of inappropriate analysis methods may lead to biased estimates on model performance, we advise clinicians to carefully examine the model performance declared by model developers

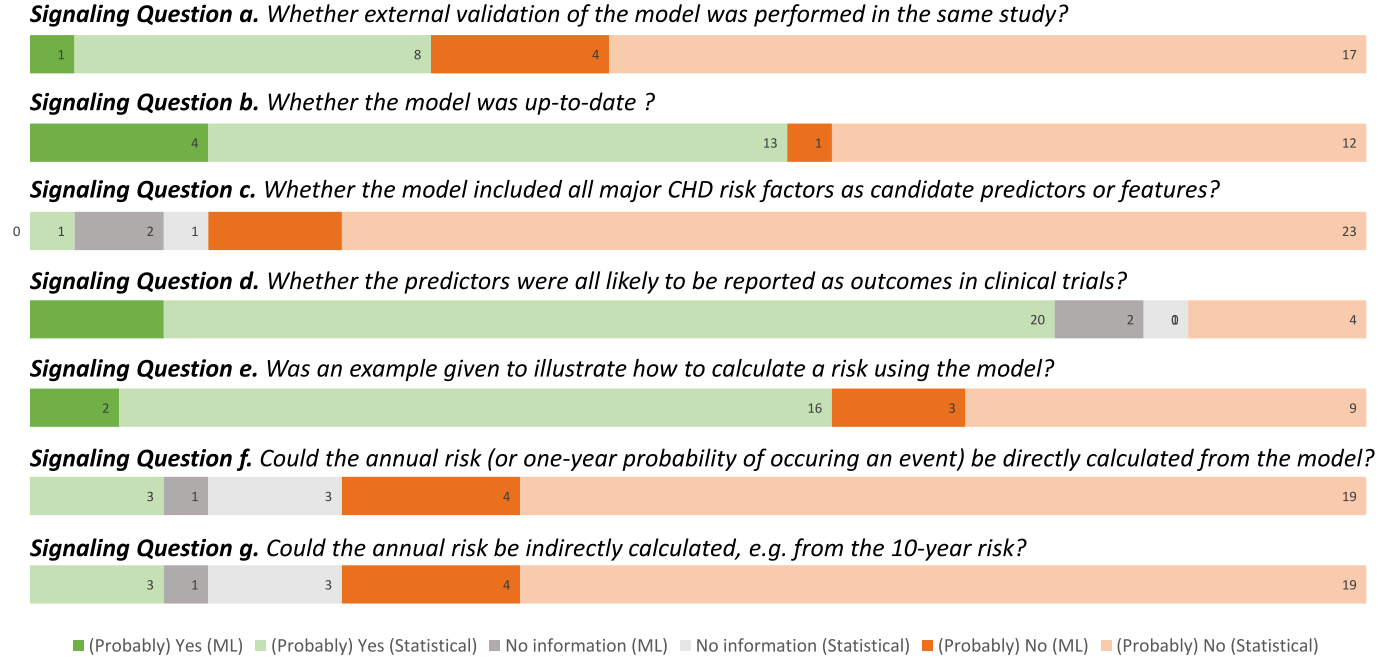

Fig. 3. Model applicability for the purpose of HTA. *Note ML* indicates machine learning models; *Statistical* indicates statistical models.

(e.g. good discrimination based on the C-index). Clinicians might trust the models developed by Quan et al. 2019, Choi et al. 2020, and Zhong et al. 2022 [47,48,61], if at least one validation study conducted in the same setting (e.g. country) is available.

The overall high RoB of the identified model studies implied that PROBAST might not be strictly followed by model developers. To go one step further, we could speculate that, model developers did not fully comply with some other published appraisal tools either, because, as mentioned in the method part, these tools also highlighted similar RoB concerns that were not adequately addressed by the identified models. For example, the CHARMS discouraged the complete-case analysis for addressing missing data, emphasized the importance of recoding model performance in terms of calibration and discrimination, and recommended the use of bootstrapping and cross-validation against overfitting [36]. Also, the STROBE guideline emphasized the importance of reporting missing data in modelling studies [38]. In contrast, almost half of the identified model studies reported no information regarding missing data.

One explanation for the lack of compliance might be the failure of disseminating the appraisal tools. The publications that described the successful external validation cannot prove the success of dissemination, as the potential model developers who do not understand an appraisal tool would never use it or describe their confusion in their own modelling study. Alternatively, the lack of compliance may be attributed by the lack of feasibility to apply the tools. For example, although all the above-mentioned tools discouraged the use of complete case analysis for addressing missing data, the complete case analysis might not lead to biases. In certain conditions, it achieved precision similar to or better than multiple imputation, and high statistical coverage [67]. If this was true, the existing tools might need to be adapted to approve the use of complete case analysis in some scenarios. Hence, further research may be needed to analyze whether the model developers have understood the existing tools, and how they have used them. We expect that future research could contribute to improved appraisal tools and the relevant dissemination strategies.

Additionally, it seems that most developers of risk prediction models did not fully understand the needs for the HTA purpose, so the potential of these models was not fully explored. For example, health economics models in the diabetes field, especially for those with a Markov structure, often need empirical data or risk prediction models for predicting the CHD risk, with an aim to accurately calculate the overall costs and effectiveness of a cohort. Compared to aggregate data, risk prediction models could enable the estimation of cost-effectiveness at individual level, and they are a good alternative to empirical patient data from real-world databases [68,69]. However, to apply risk prediction models to health economics modelling, the original mathematical equations should be provided and called repeatedly. Our results showed that, some studies only provided an online user interface without an equation, which could not satisfy the relevant needs [48,58,64]. Also, it is worth noting that, Cox models are a popular type of risk prediction models for the HTA purpose, as they could predict time to an event and an event risk within a time interval of any length. In particular, Cox models are suitable to discrete-event simulation models, an increasingly popular health economics model featured by great flexibility to handle time-to-event data [69]. However, most of our identified Cox models (n = 17) were not applicable, because they only provided a cumulative hazard function with fixed model coefficients, for estimating the CHD-related risk for 3,5,or 10 years. They did not provide the original hazard function, which enabled estimation of an instantaneous risk. These models could satisfy the needs for clinical decision-making, as information on a 5- or 10-year event risk, could help health-care providers or patients decide on which treatment to receive. However, these models could not be incorporated into a health economics model, unless assumptions are made on the instantaneous event risk (e.g. constant overtime), which would increase RoB. Therefore, we highly recommend developers of risk prediction models not to develop, but to improve their existing models, by reporting their mathematical equations more transparently, or by at least providing a cumulative function that could predict an annual event risk. Although the models with relatively wide HTA applicability (e.g. the UKPDS model [39]) have shortcomings in terms of RoB, developers may refer to these models for information on how to improve model applicability for the purpose of HTA.

Another finding regarding model applicability for HTA was that, although all the models included some CHD risk factors as model covariates, they were not in full agreement on which risk factors should be included. For example, while all the models included age, sex, and smoking as covariates, they differed in whether to include diet, physical activity, or mental health. The appropriate inclusion of CHD risk factors as model covariates has been considered by HTA agencies as evidence of wide model applicability. For example, the Dental and Pharmaceutical Benefits Agency in Sweden and the Dutch Healthcare Insurance Board in the Netherlands commented on the absence of any cholesterol measure as a covariate in a cardiovascular risk prediction model called REACH. [15] However, we identified no clear statement from HTA agencies, or even from clinical guidelines, on what risk factors should at least be included as model covariates. Indeed, many studies have investigated the issue by providing a list of major CHD risk factors [70–73], but their recommendations vary. Consequently, the lack of agreement on CHD risk factors to be included in a risk prediction model would confuse model developers, and ultimately reduce model applicability. Hence, we suggest developing a generic framework which summarizes clinical risk factors as model covariates in the diabetes field. The framework may not only address risk factors of CHD, but also those of other major DM complications.

We found that the concerns regarding model applicability for HTA cannot be simply addressed by the assessment of generic applicability. As mentioned by PROBAST, the generic applicability considers the extent to which the population, outcome, and definition and assessment of predictors match a review question. However, the generic applicability doesn't imply much regarding how to develop a model with wide applicability, as the PROBAST could not expect what review questions can be imposed by the HTA stakeholders. Consequently, the model users might only select and apply the least unsatisfactory model, while losing the opportunity of acquiring a perfect one. One solution for this applicability concern is to account for needs of HTA stakeholders in appraisal tools. This could be achieved by adapting existing appraisal tools or developing new tools. However, given the various needs of model users, innovating an one-size-fits-all appraisal tool which defines an one-size-fits-all risk prediction model may not be feasible. Therefore, to account for various needs, we recommend closer collaboration among model developers, tool developers, and HTA stakeholders, and suggest the involvement of all stakeholders in development and implementation of appraisal tools.

4.2. Comparison of ML and statistical models

Since the numbers of ML and statistical models we identified are small, we could not compare quality of the two. However, we, as reviewers, feel that it is harder to assess quality of ML models than statistical models. One obvious reason is that ML models include unique features that could not be highlighted by generic quality appraisal tools. For example, ML models might have built-in capabilities for handling missing data [51]. To address this concern, several tools specifically designed for ML models are being developed, such as the PROBAST-AI and STROBE-AI [74]. Another reason for the difficult quality assessment is that ML models normally adopt a black box approach that prevent users from interpreting the reasoning behind a models' prediction [75]. To address this concern, a research topic – Explainable AI – has emerged, and novel approaches for improved interpretability are being developed [76]. However, as model users often need to compare quality of models developed with various techniques, we suggest exploring methods to compare quality of statistical and ML models while taking

into account their particularities.

5. Limitations

Our study still has limitations. One limitation is that we might have missed models, as only one reviewer scanned all titles and abstracts, while another scanned a random set of 20 %. However, tracking references of included studies did not yield additional references. Our findings regarding overall model quality were supported by other studies and will not be disturbed by the potentially missing models. Another limitation is that our results regarding model applicability for the purpose of HTA is explorative, and the evaluation criteria were from a single review (i.e. Betts et al [15]) and authors' opinions. While our results cover key concerns of HTA stakeholders, some concerns may not be covered. Hence, extra efforts are needed if HTA stakeholders apply the models based on our results.

6. Conclusion

Both models based on machine learning and statistical techniques have been developed to predict the CHD risk in DM patients, but the quality, in terms of risk of bias and model applicability for the purpose of HTA, is relatively low. Model developers do not fully understand the needs from HTA stakeholders, and we recommend further research to explore the reasons. In addition, novel tools are needed, as the existing tools which only address generic model applicability could not satisfy the needs of HTA stakeholders. To achieve this, model developers, tool developers, and HTA stakeholders may need closer collaboration.

CRediT authorship contribution statement

Li Jiu: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Junfeng Wang:** Formal analysis, Investigation, Methodology, Supervision, Validation, Writing – review & editing. **Francisco Javier Somolinos-Simón:** . **Jose Tapia-Galisteo:** Formal analysis, Investigation, Validation, Writing – review & editing. **Gema García-Sáez:** Writing – review & editing. **Mariaelena Hernando:** Writing – review & editing. **Xinyu Li:** Formal analysis, Investigation, Validation, Writing – review & editing. **Rick A. Vreman:** Writing – review & editing. **Aukje K. Mantel-Teeuwisse:** Writing – review & editing, Supervision. **Wim G. Goettsch:** Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Not applicable.

Funding: This research was performed as part of the HTx project. The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825162. This dissemination reflects only the authors' view and the Commission is not responsible for any use that may be made of the information it contains.

Data statement

The data set used and analyzed for this review will be available from the.

corresponding author upon a reasonable request.

Author contributions.

- LJ designed the search strategy, scanned all hits, conducted full-text review of potential eligible studies, collected data, participated in quality assessment of risk prediction models, analyzed and interpreted data, and wrote the manuscript;
- JW proposed the research idea, designed the search strategy, organized training sessions of quality assessment, and edited the manuscript;
- FJSS conducted full-text review, data collection and quality assessment of risk prediction models;
- JTG conducted full-text review, data collection and quality assessment of risk prediction models;
- GGS scanned 20 % of hits, and participated in quality assessment of risk prediction models;
- MH participated in quality assessment of risk prediction models;
- XL participated in quality assessment of risk prediction models;
- RV edited the manuscript;
- AM edited the manuscript;
- WG edited the manuscript;
- All authors read and approved the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.diabres.2024.111574.

References

[1] **National Cancer Institute**: Coronary heart disease. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/coronary-heart-disease. (Accessed 2022 Jun 11).
[2] **Centers for Disease Control and Prevention**: Coronary Artery Disease (CAD). https://www.cdc.gov/heartdisease/coronary_ad.htm. (Accessed 2022 Jun 11).
[3] Einarson TR, Acs A, Ludwig C, Panton UH. Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007–2017. Cardiovasc Diabetol 2018 Dec;17(1):1–9. https://doi.org/10.1186/s12933-018-0728-6.
[4] Khalil CA, Roussel R, Mohammedi K, Danchin N, Marre M. Cause-specific mortality in diabetes: recent changes in trend mortality. Eur J Prev Cardiol 2012 Jun 1;19(3):374–81. https://doi.org/10.1177/1741826711409324.
[5] Einarson TR, Acs A, Ludwig C, Panton UH. Economic burden of cardiovascular disease in type 2 diabetes: a systematic review. Value Health 2018 Jul 1;21(7):881–90. https://doi.org/10.1016/j.jval.2017.12.019.
[6] Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2019 Sep 10;74(10):e177–232. https://doi.org/10.1161/CIR.0000000000000678.
[7] Marshall T. Coronary heart disease prevention: insights from modelling incremental cost effectiveness. BMJ 2003 Nov 27;327(7426):1264. https://doi.org/10.1136/bmj.327.7426.1264.
[8] Degrell P, Sorbets E, Feldman LJ, Steg PG, Ducrocq G. Screening for coronary artery disease in asymptomatic individuals: Why and how? Arch Cardiovasc Dis 2015 Dec 1;108(12):675–82. https://doi.org/10.1016/j.acvd.2015.10.001.
[9] Van Der Heijden AA, Ortegon MM, Niessen LW, Nijpels G, Dekker JM. Prediction of coronary heart disease risk in a general, pre-diabetic, and diabetic population during 10 years of follow-up: accuracy of the Framingham, SCORE, and UKPDS risk functions: The Hoorn Study. Diabetes Care 2009;32(11):2094–8. https://doi.org/10.2337/dc09-0745.
[10] Bhopal R, Fischbacher C, Vartiainen E, Unwin N, White M, Alberti G. Predicted and observed cardiovascular disease in South Asians: application of FINRISK, Framingham and SCORE models to Newcastle Heart Project data. J Public Health (Oxf) 2005;27(1):93–100. https://doi.org/10.1093/pubmed/fdh202.
[11] Hendriksen JM, Geersing GJ, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. J Thromb Haemost 2013 Jun;11:129–41. https://doi.org/10.1111/jth.12262.
[12] Stevanovic J, Postma MJ, Pechlivanoglou P. A systematic review on the application of cardiovascular risk prediction models in pharmacoeconomics, with a focus on primary prevention. *Eur*. J Prev Cardiol 2012;19(2_suppl):42–53. https://doi.org/10.1177/2047487312448994.
[13] Palmer AJ, Roze S, Valentine WJ. The CORE Diabetes Model: projecting long-term clinical outcomes, costs and cost-effectiveness of interventions in diabetes mellitus (types 1 and 2) to support clinical and reimbursement decision-making. Curr Med Res Opin 2004 Jan 1;20(sup1):S5–. https://doi.org/10.1185/030079904X1980.
[14] Eddy DM, Schlessinger L. Archimedes: a trial-validated model of diabetes. Diabetes Care 2003;26(11):3093–101. https://doi.org/10.2337/diacare.26.11.3093.

[15] Mueller E, Maxion-Bergemann S, Gultyaev D, et al. Development and validation of the Economic Assessment of Glycemic Control and Long-Term Effects of diabetes (EAGLE) model. Diabetes Technol Ther 2006;8(2):219–36. https://doi.org/10.1089/dia.2006.8.219.

[16] Boodoo C, Zhang Q, Ross HJ, Alba AC, Laporte A, Seto E. Evaluation of a heart failure telemonitoring program through a microsimulation model: cost-utility analysis. J Med Internet Res 2020 Oct 6;22(10):e18917.

[17] Conrads-Frank A, Schnell-Inderst P, Neusser S, et al. Decision-analytic modeling for early health technology assessment of medical devices–a scoping review. Ger Med Sci 2022;20. https://doi.org/10.3205/000313.

[18] Xu Q, Wang L, Sansgiry SS. A systematic literature review of predicting diabetic retinopathy, nephropathy and neuropathy in patients with type 1 diabetes using machine learning. J Med Artif Intell 2020;3(6).

[19] Kwon O, Na W, Kim YH. Machine learning: a new opportunity for risk prediction. Korean Circ J 2020 Jan 1;50(1):85–7. https://doi.org/10.4070/kcj.2019.0314.

[20] Collins GS, Moons KG. Reporting of artificial intelligence prediction models. Lancet 2019 Apr 20;393(10181):1577–9. https://doi.org/10.1016/S0140-6736(19)30037-6.

[21] Venema E, Wessler BS, Paulus JK. Large-scale validation of the prediction model risk of bias assessment Tool (PROBAST) using a short form: high risk of bias models show poorer discrimination. J Clin Epidemiol 2021;138:32–9. https://doi.org/10.1016/j.jclinepi.2021.06.017.

[22] Moons KG, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019 Jan 1;170(1):W1–. https://doi.org/10.7326/M18-1377.

[23] Austin PC, van Klaveren D, Vergouwe Y, Nieboer D, Lee DS, Steyerberg EW. Geographic and temporal validity of prediction models: different approaches were useful to examine model performance. J Clin Epidemiol 2016 Nov;1(79):76–85. https://doi.org/10.1016/j.jclinepi.2016.05.007.

[24] 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(5):360–9. https://doi.org/10.1136/heartjnl-2011-300734.

[25] Galbete A, Tamayo I, Librero J, Enguita-Germán M, Cambra K, Ibáñez-Beroiz B. Cardiovascular risk in patients with type 2 diabetes: A systematic review of prediction models. Diabetes Res Clin Pract 2021;109089. https://doi.org/10.1016/j.diabres.2021.109089.

[26] Betts MB, Milev S, Hoog M, et al. Comparison of recommendations and use of cardiovascular risk equations by health technology assessment agencies and clinical guidelines. Value Health 2019 Feb 1;22(2):210–9. https://doi.org/10.1016/j.jval.2018.08.003.

[27] **HTx**: About HTx project. https://www.htx-h2020.eu/about-htx-project. (Accessed 2022 Oct 25).

[28] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. https://doi.org/10.1136/bmj.n71.

[29] Debray TP, Damen JA, Snell KI, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ 2017 Jan;5:356. https://doi.org/10.1136/bmj.i6460.

[30] De Waure C, Lauret GJ, Ricciardi W, et al. Lifestyle interventions in patients with coronary heart disease: a systematic review. Am J Prev Med 2013 Aug 1;45(2):207–16. https://doi.org/10.1016/j.amepre.2013.03.020.

[31] Wolters FJ, Segufa RA, Darweesh SK, et al. Coronary heart disease, heart failure, and the risk of dementia: a systematic review and meta-analysis. Alzheimers Dement 2018 Nov 1;14(11):1493–504. https://doi.org/10.1016/j.jalz.2018.01.007.

[32] Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. J Am Med Inform Assoc 2001 Jul 1;8(4):391–7. https://doi.org/10.1136/jamia.2001.0080391.

[33] Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS One 2012 Feb 29;7(2):e32844.

[34] Faizal AS, Thevarajah TM, Khor SM, Chang SW. A review of risk prediction models in cardiovascular disease: conventional approach vs. artificial intelligent approach. Comput Methods Programs Biomed 2021 Aug;1(207):106190. https://doi.org/10.1016/j.cmpb.2021.106190.

[35] Lenselink C, Ties D, Pleijhuis R, van der Harst P. Validation and comparison of 28 risk prediction models for coronary artery disease. Eur J Prev Cardiol 2022 Mar 1;29(4):666–74. https://doi.org/10.1093/eurjpc/zwab095.

[36] 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(10):e1001744.

[37] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Br J Surg 2015;102(3):148–58. https://doi.org/10.1136/bmj.g7594.

[38] Vandenbroucke JP, Von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PloS Med 2007;4(10):e297.

[39] Hayes AJ, Leal J, Gray AM, et al. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. Diabetologia 2013 Sep;56:1925–33. https://doi.org/10.1007/s00125-013-2940-y.

[40] Nishimura K, Okamura T, Watanabe M, et al. Predicting coronary heart disease using risk factor categories for a Japanese urban population, and comparison with the randomized risk score: the suita study. J Atheroscler Thromb 2014 Aug 26;21(8):784–98. https://doi.org/10.5551/jat.19356.

[41] Piniés JA, González-Carril F, Arteagoitia JM, et al. Development of a prediction model for fatal and non-fatal coronary heart disease and cardiovascular disease in patients with newly diagnosed type 2 diabetes mellitus: the Basque Country Prospective Complications and Mortality Study risk engine (BASCORE). Diabetologia 2014 Nov;57:2324–33. https://doi.org/10.1007/s00125-014-3370-1.

[42] Haider S, Sadiq SN, Moore D, Price MJ, Nirantharakumar K. Prognostic prediction models for diabetic retinopathy progression: a systematic review. Eye 2019 May;33(5):702–13. https://doi.org/10.1038/s41433-018-0322-x.

[43] Kim E, Caraballo PJ, Castro MR, Pieczkiewicz DS, Simon GJ. Towards more accessible precision medicine: building a more transferable machine learning model to support prognostic decisions for micro-and macrovascular complications of type 2 diabetes mellitus. J Med Syst 2019 Jul;43(7):1–2. https://doi.org/10.1007/s10916-019-1321-6.

[44] Fan R, Zhang N, Yang L, Ke J, Zhao D, Cui Q. AI-based prediction for the risk of coronary heart disease among patients with type 2 diabetes mellitus. Sci Rep 2020;Sep 2;10(1):1–8. https://doi.org/10.1038/s41598-020-71321-2.

[45] Longato E, Fadini GP, Sparacino G, Gubian L, Di Camillo B. Prediction of cardiovascular complications in diabetes from pharmacy administrative claims. In2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON) 2020 Jun 16 (pp. 315-320). IEEE.

[46] Ljubic B, Hai AA, Stanojevic M, et al. Predicting complications of diabetes mellitus using advanced machine learning algorithms. J Am Med Inform Assoc 2020 Sep 1;27(9):1343–51. https://doi.org/10.1093/jamia/ocaa120.

[47] Zhong Z, Sun S, Weng J, et al. Machine learning algorithms identifying the risk of new-onset ACS in patients with type 2 diabetes mellitus: A retrospective cohort study. *Front.* Public Health 2022:10. https://doi.org/10.3389/fpubh.2022.947204.

[48] Quan J, Pang D, Li TK, et al. Risk prediction scores for mortality, cerebrovascular, and heart disease among Chinese people with type 2 diabetes. J Clin Endocrinol Metab 2019 Dec;104(12):5823–30. https://doi.org/10.1210/jc.2019-00731. Doi: 10.1002/clc.22991.

[49] Rådholm K, Chalmers J, Ohkuma T, et al. Use of the waist-to-height ratio to predict cardiovascular risk in patients with diabetes: R esults from the ADVANCE-ON study. Diabetes Obes Metab 2018 Aug;20(8):1903–10. https://doi.org/10.1111/dom.13311.

[50] Segar MW, Patel KV, Vaduganathan M, et al. Development and validation of optimal phenomapping methods to estimate long-term atherosclerotic cardiovascular disease risk in patients with type 2 diabetes. Diabetologia 2021 Jul;64(7):1583–94. https://doi.org/10.1007/s00125-021-05426-2.

[51] Shi R, Wu B, Niu Z, Sun H, Hu F. Nomogram based on risk factors for type 2 diabetes mellitus patients with coronary heart disease. Diabetes Metab Syndr Obes 2020;13:5025. https://doi.org/10.2147/DMSO.S273880.

[52] El Sanadi CE, Pantalone KM, Ji X, Kattan MW. Development and Internal Validation of A Prediction Tool To Assist Clinicians Selecting Second-Line Therapy Following Metformin Monotherapy For Type 2 Diabetes. Endocr Pract 2021 Apr 1;27(4):334–41. https://doi.org/10.1016/j.eprac.2020.10.015.

[53] Basu S, Sussman JB, Berkowitz SA, Hayward RA, Yudkin JS. Development and validation of Risk Equations for Complications Of type 2 Diabetes (RECODe) using individual participant data from 5 randomized trials. Lancet Diabetes Endocrinol 2017 Oct 1;5(10):788–98. https://doi.org/10.1016/S2213-8587(17)30221-8.

[54] Hu WS, Lin CL. Use of the progression of adapted Diabetes Complications Severity Index to predict acute coronary syndrome, ischemic stroke, and mortality in Asian patients with type 2 diabetes mellitus: A nationwide cohort investigation. Clin Cardiol 2018 Aug;41(8):1038–43. https://doi.org/10.1002/clc.22991.

[55] Xiao S, Dong Y, Huang B, Jiang X. Predictive nomogram for coronary heart disease in patients with type 2 diabetes mellitus. Front Cardiovasc Med 2022:9. https://doi.org/10.3389/fcvm.2022.1052547.

[56] Koteliukh M. Predictive model for recurrent myocardial infarction in patients with type 2 diabetes mellitus. Med Sci 2022.

[57] Kazemian P, Wexler DJ, Fields NF, Parker RA, Zheng A, Walensky RP. Development and validation of PREDICT-DM: a new microsimulation model to project and evaluate complications and treatments of type 2 diabetes mellitus. Diabetes Technol Ther 2019 Jun 1;21(6):344–55. https://doi.org/10.1089/dia.2018.0393.

[58] Lee SH, Han K, Kim HS, Cho JH, Yoon KH, Kim MK. Predicting the development of myocardial infarction in middle-aged adults with type 2 diabetes: a risk model generated from a nationwide population-based cohort study in Korea. Endocrinol Metab (Seoul) 2020 Sep;35(3):636. https://doi.org/10.3803/EnM.2020.704.

[59] Tam CH, Lim CK, Luk AO, et al. Development of genome-wide polygenic risk scores for lipid traits and clinical applications for dyslipidemia, subclinical atherosclerosis, and diabetes cardiovascular complications among East Asians. Genome Med 2021 Dec;13(1):1–8. https://doi.org/10.1186/s13073-021-00831-z.

[60] Lithovius R, Antikainen AA, Mutter S, et al. Genetic Risk Score Enhances Coronary Artery Disease Risk Prediction in Individuals With Type 1 Diabetes. Diabetes Care 2022 Mar;45(3):734–41. https://doi.org/10.2337/dc21-0974.

[61] Choi Y, Yang Y, Hwang BH, et al. Practical cardiovascular risk calculator for asymptomatic patients with type 2 diabetes mellitus: PRECISE-DM risk score. Clin Cardiol 2020 Sep;43(9):1040–7. https://doi.org/10.1002/clc.23405.

[62] Ferreira JP, Sharma A, Mehta C, et al. Multi-proteomic approach to predict specific cardiovascular events in patients with diabetes and myocardial infarction: findings from the EXAMINE trial. Clin Res Cardiol 2021 Jul;110(7):1006–19. https://doi.org/10.1007/s00392-020-01729-3.

[63] Hirai H, Asahi K, Yamaguchi S, et al. New risk prediction model of coronary heart disease in participants with and without diabetes: assessments of the Framingham

risk and Suita scores in 3-year longitudinal database in a Japanese population. Sci Rep 2019 Feb 26;9(1):1–6. https://doi.org/10.1038/s41598-019-39049-w.
[64] Ye W, Ding X, Putnam N, et al. Development of clinical prediction models for renal and cardiovascular outcomes and mortality in patients with type 2 diabetes and chronic kidney disease using time-varying predictors. J Diabetes Complications 2022 May 1;36(5):108180. https://doi.org/10.1016/j.jdiacomp.2022.108180.
[65] Hajar R. Risk factors for coronary artery disease: historical perspectives. Heart Views 2017;18(3):109. https://doi.org/10.4103/HEARTVIEWS.HEARTVIEWS_106_17.
[66] Van der Heijden AA, Nijpels G, Badloe F, et al. Prediction models for development of retinopathy in people with type 2 diabetes: systematic review and external validation in a Dutch primary care setting. Diabetologia 2020;63(6):1110–9. https://doi.org/10.1007/s00125-020-05134-3.
[67] Mukaka M, White SA, Terlouw DJ, Mwapasa V, Kalilani-Phiri L, Faragher EB. Is using multiple imputation better than complete case analysis for estimating a prevalence (risk) difference in randomized controlled trials when binary outcome observations are missing? Trials 2016;17(1):1–12. https://doi.org/10.1186/s13063-016-1473-3.
[68] Li X, Li F, Wang J, van Giessen A, Feenstra TL. Prediction of complications in health economic models of type 2 diabetes: a review of methods used. Acta Diabetol 2023 Jul;60(7):861–79.
[69] Caro JJ, Möller J, Karnon J, Stahl J, Ishak J. Discrete event simulation for health technology assessment. CRC Press; 2015 Oct 16.
[70] Roeters van Lennep JE, Westerveld HT, Erkelens DW. **van der Wall EE.** Risk factors for coronary heart disease: implications of gender. Cardiovasc Res 2002;53(3):538–49. https://doi.org/10.1016/s0008-6363(01)00388-1.
[71] Albus C. Psychological and social factors in coronary heart disease. Ann Med 2010;42(7):487–94. https://doi.org/10.3109/07853890.2010.515605.
[72] Kannel WB. Coronary heart disease risk factors in the elderly. Am J Geriatr Cardiol 2002;11(2):101–7. https://doi.org/10.1111/j.1076-7460.2002.00995.x.
[73] Hopkins PN, Williams RR. A survey of 246 suggested coronary risk factors. Atherosclerosis 1981;40(1):1–52. https://doi.org/10.1016/0021-9150(81)90122-2.
[74] Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021;11:e048008.
[75] Wanner J, Herm LV, Janiesch C. How much is the black box? The value of explainability in machine learning models. *ECIS 2020 Research-in-Progress Papers* 2020; 85.
[76] Xu F, Uszkoreit H, Du Y, Fan W, Zhao D, Zhu J. Explainable AI: A brief survey on history, research areas, approaches and challenges. InCCF international conference on natural language processing and Chinese computing 2019:563–74.