HHS Public Access

Author manuscript

Comput Methods Programs Biomed Update. Author manuscript; available in PMC 2023 June 16.

Published in final edited form as:

Comput Methods Programs Biomed Update. 2023 ; 3: . doi:10.1016/j.cmpbup.2022.100087.

Evaluation of available risk scores to predict multiple cardiovascular complications for patients with type 2 diabetes mellitus using electronic health records

Joyce C Ho[a,*], **Lisa R Staimez**[b], **K M Venkat Narayan**[b], **Lucila Ohno-Machado**[c], **Roy L Simpson**[d], **Vicki Stover Hertzberg**[d]

[a] Department of Computer Science, Emory University, 400 Dowman Drive, Atlanta, GA 30322, United States

[b] Hubert Department of Global Health, Rollins School of Public Health, Emory University, United States

[c] Department of Biomedical Informatics, School of Medicine, University of California San Diego, United States

[d] Center for Data Science, Nell Hodgson Woodruff School of Nursing, Emory University, United States

Abstract

Aims: Various cardiovascular risk prediction models have been developed for patients with type 2 diabetes mellitus. Yet few models have been validated externally. We perform a comprehensive validation of existing risk models on a heterogeneous population of patients with type 2 diabetes using secondary analysis of electronic health record data.

Methods: Electronic health records of 47,988 patients with type 2 diabetes between 2013 and 2017 were used to validate 16 cardiovascular risk models, including 5 that had not been compared previously, to estimate the 1-year risk of various cardiovascular outcomes. Discrimination and calibration were assessed by the c-statistic and the Hosmer-Lemeshow goodness-of-fit statistic, respectively. Each model was also evaluated based on the missing measurement rate. Sub-analysis was performed to determine the impact of race on discrimination performance.

Results: There was limited discrimination (c-statistics ranged from 0.51 to 0.67) across the cardiovascular risk models. Discrimination generally improved when the model was tailored towards the individual outcome. After recalibration of the models, the Hosmer-Lemeshow statistic yielded p-values above 0.05. However, several of the models with the best discrimination relied on measurements that were often imputed (up to 39% missing).

* Corresponding author. joyce.c.ho@emory.edu (J.C. Ho).

Declaration of Competing Interest

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpbup.2022.100087.

Conclusion: No single prediction model achieved the best performance on a full range of cardiovascular endpoints. Moreover, several of the highest-scoring models relied on variables with high missingness frequencies such as HbA1c and cholesterol that necessitated data imputation and may not be as useful in practice. An open-source version of our developed Python package, cvdm, is available for comparisons using other data sources.

Keywords

Type 2 diabetes; Cardiovascular disease; Electronic health records; Risk models

1. Introduction

Despite concerted efforts by the American Diabetes Association and the American Heart Association to reduce its risk [1], cardiovascular disease (CVD) remains the most prevalent cause of morbidity and mortality in individuals with diabetes. For patients with diabetes, the identification of high CVD risk (CVDR) is crucial to early prevention, timely treatment, and effective management for reducing CVDR. Risk prediction models can identify high-risk patients with approaches ranging from simple equations with a small number of predictor variables to more complex approaches that capture interactions between different variables.

Various CVDR prediction models have been developed for patients with type 2 diabetes mellitus, or PT2DM [2]. However, the optimal algorithm for CVDR estimation and the need for diabetes-specific CVDR models are debatable. Each guideline recommends a different CVDR prediction model, despite limited evidence of external validation across these prediction models. Among the few studies to evaluate various CVDR scores for PT2DM on an external validation cohort [3–7], two studies [3,4] assessed 3–5 CVDR models on small cohorts (n ranging from 453 to 1174), yet small sample sizes can yield imprecise estimates of discrimination and calibration. Another study of 181,399 PT2DM (70.8% Caucasian or White) used 6 CVDR models [5] but the homogenous population can produce inaccurate risk estimates in multiethnic cohorts [8]. The last two studies investigated 3 and 22 CVDR models on electronic health record (EHR) cohorts but either had less missing measurements typically found in EHRs [6] or omitted several recently developed CVDR scores [7]. Supplemental Table 2 contains more details about these validation studies. Additionally, each risk model utilizes different outcome definitions which can hinder the validation of the various models. This coupled with the changing patterns of CVD presentation such as the higher prevalence of other CVD complications beyond coronary heart disease (CHD) and stroke [9] suggests the need for an in-depth study of various individual CVD endpoints. Therefore, a comprehensive evaluation of various risk prediction models on a large, multiethnic cohort across the full range of CVD outcomes is necessary to capture the contemporary presentation of CVD.

Another major limitation of the existing validation studies is the use of data from longitudinal cohorts, where the risk factors were measured using well-established protocols. In clinical practice, such factors are obtained from EHRs, in which data can be erroneous or incomplete. For example, in the Longitudinal Epidemiologic Assessment of Diabetes Risk (LEADR) study which captured EHR data from 1.4 M patients over six years, yet

17%, 65%, and 83% of the patients have no measurements associated with blood pressure, cholesterol, and HbA1c, respectively [10]. A recent study suggested that published CVDR models can be successfully applied to EHRs [11], yet whether a similar conclusion can be drawn for the CVDR scores developed for use among PT2DM is unclear.

Increasing availability of EHRs provides the opportunity to evaluate the performance of existing risk scores in a large, multiethnic, contemporary population of PT2DM. The individual-level nature of EHRs can also be used to understand performance differences that arise from 5 different CVD events: CHD, congestive heart failure (CHF), myocardial infarction (MI), stroke, and a composite endpoint, thereby capturing the full range of CVD presentation. We comprehensively evaluate the predictive performance of 13 diabetes-specific CVDR scores including 5 models, ARIC [12], DIAL [13], DMCx [14], HKDR [15–17], and UKPDS OM2 [18] that were not part of any previous validation study. Our study also contains the most diabetes-specific CVDR scores with 4 more than a previous study [7]. We are the largest heterogeneous study (n = 48,779) with more than 20% of the patients identifying as African American or Black and the first to assess performance based on ethnicity. Furthermore, we are the only study to assess the models on MI even though it is commonly found in existing CVD composite definitions. Finally, we developed an open-source Python package that implements all the scores in a cohesive software package.

2. Materials and methods

2.1. Study design and population

Our study was conducted using EHRs that did not contain identifiers, except for shifted dates, from the Emory Healthcare clinical data warehouse (CDW), a large healthcare delivery organization in the Southeast region of the United States. Secondary data analysis was approved by the Emory University Institutional Review Board. Adult patients seen in the clinic, outpatient, and inpatient settings aged 18 to 80 years old with type 2 diabetes (T2DM) (ICD-9 code of ‘250.XX’ or ICD-10 code of ‘E11.XX’) between January 1, 2013 and December 31, 2017 were included. For data security, certain records were omitted based on the date shifting logic. Patients who did not have at least one encounter after the initial T2DM diagnosis (earliest recorded T2DM billing code) were excluded from this study. Patients with a prior history of all outcomes of interest (CHD, CHF, MI, and Stroke) were excluded from the study. Thus, our study population was comprised of 48,779 patients.

2.2. Endpoint definitions

We used the combination of ICD-9 and ICD-10 clinical codes to identify cases of CHD (ICD-9 code of ‘414.XX’ or ICD-10 code of ‘I25. XX’), CHF (ICD-9 code of ‘428.XX’ or ICD-10 code of ‘I50.XX’), DCM (ICD-9 code of ‘425.XX’ or ICD-10 code of ‘I42.XX’), MI (ICD-9 code of ‘410.XX’, ‘411.1′, ‘411.8′ or ICD-10 code of ‘I21.XX’), and Stroke (ICD-9 code of ‘430.XX-434.XX’ or ICD-10 code of ‘I66.XX’). To ensure the patient likely experienced the event, the code needed to be classified as an admitting diagnosis or discharge diagnosis in the EHR. The CVD endpoint was defined as the presence of CHD, CHF, MI, or Stroke. The earliest recorded date in the EHR was used as the diagnosis date for each of the endpoints.

Due to limited follow-up in our cohort (full-scale EHR adoption occurred in 2012), we posed event prediction as a binary classification problem: whether an individual experienced a particular CVD endpoint in the next year. Although existing validation studies focused on predicting time-to-event, a previous study of cardiovascular risk also used event prediction [19]. For each outcome, the value 1 (i.e., positive) was assigned for any patient where there was either an encounter or death of the endpoint within the next year. The value 0 was assigned to subjects that either had an encounter or death at least one year in the future without CHD, CHF, MI, or stroke.

Each patient was limited to a single encounter, or index encounter, to avoid arbitrarily inflating or deflating discrimination ability. For the positive samples, the index encounter was the earliest recorded encounter within one year of the CVD endpoint. For the negative samples, the index encounter was a randomly sampled encounter occurring at least one year before the last recorded encounter.

2.3. Selected CVD risk models

A recent systematic reviews identified 19 CVDR prediction models specifically developed for use in PT2DM [7]. We benchmarked 13 of the 19 models: Advance [20], ARIC [12], DARTS [21], DCS [22], DIAL [13], DMCx [14], Fremantle [23], HKDR [15–17], NDR [24], QDiabetes [25], RECODe [26], UKPDS [27], and UKPDS OM2 [18]. The remaining 6 CVDR scores were not considered as some of the predictors were not readily available in our EHRs (e.g., exercise, years of formal education, alcohol consumption). We considered 3 general CVDR: ACC/AHA Pooled Cohort Equation (PCE) [28], the Framingham model [29], and the SCORE model [30] as comparisons due to their success in existing PT2DM validation studies. A short description of the model, the endpoints included in the outcome definition, and the method for dealing with missing data are summarized in Table 1. A comparison of the predictor variables used for each model can be found in the Supplemental Material.

2.4. Definition of predictors in cohort

All types of patient encounter data (i.e., clinics, outpatient, and inpatient visits) within one year prior to index encounter were considered for predictor value measurement. Data included billing codes, medication lists, laboratory results, and structured notes such as physiological measurements. Age, sex, and race were self-reported in the administrative records. For vital signs and laboratory values, the most recent measurement was used. Extreme values of vital signs were excluded (1st and 99th percentile measurements). For medication-related predictors (i.e., treatment of hypertension, insulin, etc.), the value was set to positive if there was any presence of the medication within the last year before the index encounter. A table of the predictors used by each CVDR model is provided in Supplemental Table 1.

2.5. Statistical analysis

We developed cvdm, an open-source Python 3.7 package, to implement the CVDR scores used here. Python is widely adopted by data analysts and researchers due to its ease of interactive data analysis and utility as a general-purpose programming language [31]. Code

is available as a GitHub repository (https://github.com/joyceho/cvdm) under a permissive MIT license. Thus the CVDR scoring systems can be validated on many datasets, promoting transparency of this work.

Pperformance of CVDR models was assessed both in terms of discrimination and calibration. The C-statistic was used to assess model discrimination. For model calibration, we used the Hosmer-Lemeshow χ^2 test, reliability-in-the-small, reliability-in-the-large, and visual inspection [32]. Ten random samples (70%-30% train-test) were drawn using the repeated random subsampling cross-validation method. We used CVDR scores from training data to calibrate an isotonic regression model [33], a calibration mechanism that only assumes the curve is monotonically increasing and is more general than the sigmoid method provided there is sufficient calibration data. Discrimination and calibration performance were calculated only on the test subjects for each random sample.

We imputed predictors that were missing for a subset of the patients using five different mechanisms: mean, median, k-nearest neighbors, multivariate imputation by chained equations (MICE), and stochastic regression imputation. The imputation methods were performed using the impute module in scikit-learn or the autoimpute package. Imputation parameters (e.g., mean values) were learned from the training data and applied to the test data. The practicality of the models was measured as the average percentage of missing measurements. Additional details regarding imputation and evaluation are available in the Supplemental Material.

3. Results

3.1. Descriptive statistics

Overall, 5923 out of 39,187 (15.1%) subjects were diagnosed with CHD; 4355 out of 43,607 (10.0%) with CHF; 2304 out of 45,930 (5.0%) had an MI; and 2390 out of 45,510 (5.3%) suffered a stroke; and 7840 out of 36,620 (21.4%) experienced a CVD endpoint in our study population (n = 48,779). Baseline characteristics and subgroups of interest are reported in Table 2. We reported descriptive statistics for the demographic and clinical data as means ± standard deviations, with percentages of missing values reported in parenthesis.

3.2. Discrimination

All CVDR scores had limited to moderate discrimination, the average C-statistic ranging between 0.53 and 0.67. Imputation method for missing measurements had a marginal impact on the resulting discrimination (shown in Supplemental Fig. 1). Mean imputation consistently performed the best across the 16 CVDR models and 6 endpoints. Discrimination results using mean imputation with the confidence intervals (CIs) for the C-statistic are summarized in Table 3. Although no single CVDR model offered the best performance across the five different endpoints, DMCx performed consistently well across all the endpoints. For CHD, DMCx had the best discrimination by best point estimate (0.67 with 95% CI of 0.66–0.67). Advance offered comparable performance within the 95% CI. DMCx also had the best discrimination for MI with point estimate (0.65 with 95% CI of 0.64–0.66) where HKDR, RECODE, and UKPDS OM2 yielded similar performance.

Similarly, the CHF variant of RECODe achieved the highest point estimate of discrimination for detecting CHF (0.67 with 95% CI of 0.67–0.68) and CVD (0.66 with 95% CI of 0.66–0.66). For CVD, Advance and DMCx offered comparable performance to RECODE. For predicting stroke, the CVD-variant of DCS and Fremantle had the highest point estimated C-statistic (0.63 with 95% CIof 0.62–0.63). Advance, DMCx, and HKDR also provided similar CVD discrimination ability to Advance and DMCx.

For the four individual CVD outcomes, Stroke was the most difficult endpoint to predict (average C-statistic of 0.60) followed by MI and CHF (average C-statistic of 0.61). Discriminating between patients who experienced CHD was the easiest of the outcomes with an average C-statistic of 0.63. The results also suggest that discrimination generally improves when the risk equation was tailored towards the individual outcome. For example, the top performers for CHF included CHF-variants of RECODe and HKDR. A similar pattern happened for MI with the MI-variants of UKPDS OM2 and RECODe and stroke with stroke-variants of UKPDS OM2, RECODe, and HKDR. However, a tailored risk equation for the event was not guaranteed to outperform other risk models, as evident with the CHF-variant of UKPDS OM2 and the MI-variant of DCS.

Comparison of the diabetes-specific CVDR equations and general-population CVDR equations (i.e., Framingham, PCE, and SCORE) did not yield a significant difference overall. For example, ARIC, DARTS, DIAL, and UKPDS were unable to distinguish high-risk subjects compared to Framingham, PCE, and SCORE. Only Advance, DMCx, RECODe, and HKDR were able to achieve better discrimination than Framingham, PCE, and SCORE across the 4 individual CVD endpoints.

3.3. Impact of missing values

Certain CVDR models relied on predictors that had a higher percentage of missing values. As shown in Table 2 Fig. 1, 68–73% of our study population did not have an HbA1c measurement in the year before the encounter date. Even the most available measurement, systolic blood pressure (SBP), was unobserved in 16–34% of our population. To assess the potential impact of missing measurements, the models were assessed on two subgroups of our study population: those with both diastolic (DBP) and SBP measurements, and those with SBP, body mass index (BMI), and smoking status available. For the CVD endpoint, this resulted in patient cohorts of 24,720 and 6753, respectively. There was no noticeable difference in discriminatory performance even with fewer missing measurement levels across the various models (results summarized in Supplemental Table 3).

Models with higher percentages required more data imputation and thus may be less practical when the discrimination performance is similar. Fig. 1 illustrates the summary statistics for the percentages of missing values across the 16 CVDR models for the CVD endpoint. Framingham and PCE required the least amount of data imputation, with 21.3% and 23.2% of the entries missing on average across the entire cohort. DIAL consistently had the highest missing percentage of the predictor values (39.4%), followed by NDR (38.3%), HKDR (34.8%), and DMCx (33.0%). Thus, three of the highest scoring models (DMCx, HKDR, and RECODe) relied on data that were often imputed. Moreover, the

diabetes-specific CVDR models also generally relied on measurements that were often not available.

3.4. Impact of ethnicity

Existing prediction models can yield inaccurate risk estimates on multiethnic cohorts [8]. As shown in Table 2, almost 44% of our population identified as African American or Black. We performed a sub-analysis to assess the impact of race on the results. We focused on two subgroups: Caucasian or White and African American or Black. Supplemental Table 4 summarizes the discrimination performance on the two subgroups across the 16 risk prediction models. Across the five different outcomes, the discrimination performance was generally better on the Caucasian or White population as opposed to the African American or Black. This was true even for models where race was a predictor value (i.e., ARIC, DCS, Fremantle, QDiabetes, RECODE, UKPDS, UKPDS OM2). The lone outlier was DIAL, which had better performance on African American or Black patients across the five outcomes. More surprisingly, all models except RECODE offered better performance for the African American or Black population on the MI endpoint.

3.5. Impact of prediction time

The existing prediction models are predominately designed for 5- or 10-year risk, and thus may have limited discrimination power due to the short time frame. We performed a sensitivity analysis to evaluate the impact of the time frame. Due to the limited follow-up of the cohort, we were only able to extend the time frame to 3 years. Supplemental Table 5 summarizes the discrimination performance across the 16 prediction models for 1-, 2-, and 3-year CVD prediction. There was no noticeable performance improvement between the 1-year endpoint and the 3-year endpoint.

3.6. Calibration

Prior to recalibration using isotonic regression, the calibration of all the scores was poor with p values below 0.05 (shown in Supplemental Fig. 2). After recalibration, Hosmer-Lemeshow χ^2 statistics improved and yielded p-values above 0.1 (summarized in Supplemental Table 3). For reliability-in-the-large and reliability-in-the-small, the value was almost 0 for many of the scores, suggesting slight differences between expected and observed risk. This is supported by the calibration plots shown in Fig. 2 as the predicted risk and observed risk were comparable across the various scores. However, the recalibrated models failed to generate predictions for higher-risk groups. We also assessed impact of calibration method by comparing isotonic regression and Platt's scaling. Isotonic regression provided noticeably better calibration results without impacting discrimination (shown in Supplemental Table 4).

4. Discussion

In this study, we conducted a comprehensive external validation of 16 CVDR models for PT2DM. We also performed the first validation of various diabetes-specific risk scores using only EHR data. Furthermore, we assessed predictive performance of the models for the full range of CVD outcomes to capture the contemporary presentation of CVD.

Across all CVDR scores, the ability to discriminate between patients who will and will not get any of the four individual CVD endpoints (i.e., CHD, CHF, MI, and Stroke) using EHRs in the next year was comparable, with the C-statistics < 0.68. After recalibration, there was a good agreement between predicted and observed risks for most models.

4.1. Findings from other studies

A previous study has suggested that diabetes-specific CVDR prediction models provide slightly better discriminative performance than general-population risk prediction models [34]. This discriminatory advantage is posited to arise from use of diabetes as a binary risk factor, ignoring the heterogeneity in diabetes. In contrast, diabetes-specific CVDR models include several diabetes-related predictors such as diabetes duration and HbA1c measurements. Our results suggest that model performance depends on the model itself as not all diabetes-specific models improved discriminative performance. A potential explanation is differences in outcome definitions, as the general population-derived scores included 3 or 4 of the five individual endpoints, whereas several of the diabetes-specific risk models only included 2 or 3 endpoints, as indicated in Table 1.

There have only been five previous validation studies that assessed performance of several CVDR scores on PT2DM. One validation study found similar discriminative ability with C-statistics of 0.54–0.69, with variability in performance arising from the cohort itself and not the risk model [3]. Two validation studies found higher discriminative performance on their respective cohorts with C-statistics ranging from 0.619 to 0.674 and no discriminatory advantage in using a diabetes-specific CVDR model [5,7]. However, the general population score was less well-calibrated as it overestimated the risk and may result in unnecessary interventions. A study that included an EHR cohort obtained C-statistics from 0.58 to 0.62 with estimates comparable between EHRs and population-based cohorts even though EHR cohorts tend to be sicker [6].

4.2. Strengths/limitations

There are several strengths to this study. First is inclusion of 13 published CVDR models designed for PT2DM. Previous validation studies only assessed the performance of 5–9 diabetes-specific risk scores [3–7]. Thus, our study serves as the most comprehensive and extensive external validation of existing CVDR scores. Moreover, our study evaluated the differences between diabetes-specific and general population-specific models for individual CVD complications. Two previous validation studies only compared against a single general population model [5] or only CHD [7] while another was a meta-analysis of existing published articles [34]. Finally, our study used EHR data from a diverse, heterogeneous population from a large healthcare system. By performing a secondary analysis of EHRs, our study cohort serves as one of the larger validation populations and offers a practical setting for assessing risk.

There are several limitations to our study that need to be addressed. The first limitation is the validation of prediction models for 5- or 10-year risk to predict short-term (1-year to 3-year) risk. Given the limited follow-up time for our study population, it was infeasible to predict 5-year or 10-year using EHR data. A follow-up study can be conducted after 7

years to better understand differences in short-term prediction. A second limitation is that not all predictors were commonly collected in our cohort. More than two-thirds of the cohort had missing measurements for one or more of the cholesterol values, smoking, eGFR, and HbA1c. We addressed this issue by exploring several common imputation methods using other risk predictors. In addition, we analyzed the amount of data imputation required to use each risk prediction model (see Fig. 1). However, data imputation itself could have resulted in some loss of discriminatory power. One remedy is to pool data across multiple healthcare sites to construct a sufficiently large and diverse cohort with less missing measurements for future work. A third issue is related to measurement quality. Unlike traditional cohort studies where the measurement of predictor values is performed using a standardized protocol, EHR data is filled with noisy measurements. Extreme measurement errors such as a height of 200 inches were excluded, but other measurement errors were not as readily detected. Recent development of data quality indicators [35] can potentially be incorporated into the assessment of measurement quality for future studies. There are also some concerns regarding the recording accuracy of the individual CVD events and patient history related to atrial fibrillation, CHF, hypertension. Outcomes and predictor variables were assessed using billing codes which can have low precision. Thus, the risk scores may not be an accurate assessment of the subject.

CVDR model predictive performance was slightly lower than the original C-statistic reported in the studies (0.68–0.85) and the C-statistics of the externally validated studies (0.59–0.86) [2]. This might be explained by the ethnicity differences in our study population, with 44% of the subjects identifying as African American or Black. Many CVDR scores had substantially less population diversity. Additional differences may also arise from using EHR-collected data which may be noisier and have a higher percentage of missingness [36]. Another explanation might be that many CVDR scores were developed on populations observed between 170 and 2000, and thus not be well-adjusted to changes in treatment and interventions of the last two decades. As observed from Table 2, our study population generally had good control of their blood pressure (SBP <140 mmHg), cholesterol (total cholesterol <4.5 mmol/L), and HbA1c (% <8). This is contrasted with the ARIC cohort, where the median cholesterol level was between 5.18 and 6.21 mmol/L [12], and the DARTS cohort, where the average HbA1C was 8% and SBP was 144 mmHg [21]. This explanation is also supported by slightly better discrimination power of Advance, DCS, DMCx, HKDR, RECODe, and UKPDS OM2, which were developed on more recent cohorts.

Ability to differentiate between different CVD complications can help guide treatment and inform patients of their risks. However, to be useful in practice, CVDR models should provide accurate and well-calibrated estimates of risk. While discrimination ability is moderate, several of the common predictor variables were missing for our study population, thus making them less practical. Future studies can focus on replicating results in other healthcare systems to determine viability of implementing risk calculations on EHR systems for real-time use by clinical providers.

4.3. Conclusions

This extensive study showed a comparable and moderate ability to discriminate between PT2DM who will experience a specific CVD complication in the following year based on EHR data. Simple recalibration of CVDR models to our study population resulted in improved estimates of actual risk. However, there was not a single existing prediction model that outperformed all others on all CVD endpoints. The results also suggest that detailed sub-analysis of risk scores should be performed to determine impact arising from the race as it yielded unexpected results for the MI outcome. Models that relied on less data imputation such as Advance, Framingham, and PCE can potentially be used by healthcare providers in a clinic setting to identify patients who are at low or high risk of developing CVD complications. In addition, multiple CVDR scores can be used simultaneously as they offer different performance across the individual outcomes. Moreover, the calculation can be performed in real-time from the current measurements collected during the subject's clinic visit, as the cvdm package can integrate to existing EHR systems. However, to improve risk estimates and develop more accurate prediction models, practitioners should collect measurements that are commonly used in these models to better assess CVDR.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

All authors contributed to the interpretation of the findings, manuscript revisions, and approved the final version of the manuscript. J.C.H carried out the data preparation and statistical analyses. The study was conceived by J.C.H., L.R.S., V.S.H., R.L.S., and K.M.V.N. L.O.M. contributed to the assessments of calibration and imputation. J.C.H. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. We would like to acknowledge the contributions of Emma Wang and Jeff Weaver, colleagues from the Emory Library and Information Technology Data Solutions group, for their assistance in extracting data from the Emory Healthcare Clinical Data Warehouse.

Sources of funding

This work was supported by funding from the National Library of Medicine grant 1K01LM012924 and grant P30DK111024 from the Georgia Center for Diabetes Translation Research funded by the National Institute of Diabetes and Digestive and Kidney Diseases. The content is solely the responsibility of the authors and does not necessarily represent the official views of this organization.

References

[1]. Joseph JJ, Deedwania P, Acharya T, Aguilar D, Bhatt DL, Chyun DA, Di Palo KE, Golden SH, Sperling LS, on behalf of the AHA diabetes Committee of the Council on lifestyle and cardiometabolic health; council on arteriosclerosis thrombosis and vascular biology; council on clinical cardiology; and council on hypertension. Comprehensive management of cardiovascular risk factors for adults with type 2 diabetes: a scientific statement from the American Heart Association, Circulation 145 (2022) e722–e759. [PubMed: 35000404]

[2]. 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. (2022) 184.

[3]. Jvd Leeuw, Sv Dieren, Beulens JWJ, Boeing H, Spijkerman AMW, Graaf Yvd, DLvd A, Nöthlings U, Visseren FLJ, Rutten GEHM, Moons KGM, Schouw YTvd, Peelen LM, The validation of cardiovascular risk scores for patients with type 2 diabetes mellitus, Heart 101 (2015) 222. [PubMed: 25256148]

[4]. Elaheh Y, Rokhsareh A, Alireza K, Mohammad Ebrahim K, Cardiovascular Events in People With Type 2 Diabetes: performance of Framingham, UKPDS, and ADVANCE Risk Equations, Acta Med. Iran. (2021) 59.

[5]. Read SH, Diepen Mv HM Colhoun N Halbesma RS Lindsay JA McKnight D McAllister A, Pearson ER, Petrie JR, Philip S, Sattar N, Woodward M, Wild HS, On behalf of the Scottish Diabetes Research Network Epidemiology Group, Performance of cardiovascular disease risk scores in people diagnosed with type 2 diabetes: external validation using data from the National Scottish Diabetes Register, Diabetes Care 41 (2018), dc180578.

[6]. Szymonifka J, Conderino S, Cigolle C, Ha J, Kabeto M, Yu J, Dodson JA, Thorpe L, Blaum C, Zhong J, Cardiovascular disease risk prediction for people with type 2 diabetes in a population-based cohort and in electronic health record data, JAMIA Open (3) (2020) 583–592. [PubMed: 33623893]

[7]. Dziopa K, Asselbergs FW, Gratton J, Chaturvedi N, Schmidt AF, Cardiovascular risk prediction in type 2 diabetes: a comparison of 22 risk scores in primary care settings, Diabetologia 65 (2022) 644–656. [PubMed: 35032176]

[8]. DeFilippis AP, Young R, Carrubba CJ, McEvoy JW, Budoff MJ, Blumenthal RS, Kronmal RA, McClelland RL, Nasir K, Blaha MJ, An analysis of calibration and discrimination among multiple cardiovascular risk scores in a modern multiethnic cohort, Ann. Intern. Med. 162 (2015) 266–275. [PubMed: 25686167]

[9]. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, Boehme AK, Buxton AE, Carson AP, Commodore-Mensah Y, Elkind MSV, Evenson KR, Eze-Nliam C, Ferguson JF, Generoso G, Ho JE, Kalani R, Khan SS, Kissela BM, Knutson KL, Levine DA, Lewis TT, Liu J, Loop MS, Ma J, Mussolino ME, Navaneethan SD, Perak AM, Poudel R, Rezk-Hanna M, Roth GA, Schroeder EB, Shah SH, Thacker EL, VanWagner LB, Virani SS, Voecks JH, Wang N-Y, Yaffe K, Martin SS, On behalf of the American Heart association council on epidemiology and prevention statistics committee and stroke statistics subcommittee. Heart Disease and Stroke Statistics-2022 Update: A Report From the American Heart Association. Circulation 2022;145. [PubMed: 35861771]

[10]. Fishbein HA, Birch RJ, Mathew SM, Sawyer HL, Pulver G, Poling J, Kaelber D, Mardon R, Johnson MC, Pace W, Umbel KD, Zhang X, Siegel KR, Imperatore G, Shrestha S, Proia K, Cheng Y, Bullard K, Gregg EW, Rolka D, Pavkov ME, The longitudinal epidemiologic assessment of diabetes risk (LEADR): unique 1.4M patient electronic health record cohort, Healthcare 8 (2020), 100458. [PubMed: 33011645]

[11]. Wolfson J, Vock DM, Bandyopadhyay S, Kottke T, Vazquez-Benitez G, Johnson P, Adomavicius G, J O'Connor P, Use and customization of risk scores for predicting cardiovascular events using electronic health record data, J Am Heart Assoc (2017) 6.

[12]. Folsom AR, Chambless LE, Duncan BB, Gilbert AC, Pankow JS, Atherosclerosis risk in communities study investigators. Prediction of coronary heart disease in middle-aged adults with diabetes, Diabetes Care 26 (2003) 2777–2784. [PubMed: 14514579]

[13]. Berkelmans GFN, Gudbjörnsdottir S, Visseren FLJ, Wild SH, Franzen S, Chalmers J, Davis BR, Poulter NR, Spijkerman AM, Woodward M, Pressel SL, Gupta AK, Schouw YTvd A-M Svensson, Graaf Yvd, Read SH, Eliasson B, Dorresteijn JAN, Prediction of individual life-years gained without cardiovascular events from lipid, blood pressure, glucose, and aspirin treatment based on data of more than 500 000 patients with Type 2 diabetes mellitus, Eur. Heart J. 40 (2019) 2899–2906. [PubMed: 30629157]

[14]. Wan EYF, Fong DYT, Fung CSC, Yu EYT, Chin WY, Chan AKC, Lam CLK, Development of a cardiovascular diseases risk prediction model and tools for Chinese patients with type 2 diabetes mellitus: a population-based retrospective cohort study, Diabetes Obes. Metab. 20 (2018) 309–318. [PubMed: 28722290]

[15]. Yang X, Ma RC, So WY, Kong AP, Ko GT, Ho CS, Lam CW, Cockram CS, Tong PC, Chan JC, Development and validation of a risk score for hospitalization for heart failure in patients with Type 2 Diabetes Mellitus, Cardiovasc. Diabetol. 7 (2008) 9. [PubMed: 18430204]

[16]. Yang X, So WY, Kong APS, Ma RCW, Ko GTC, Ho CS, Lam CWK, Cockram CS, Chan JCN, Tong PCY, Development and validation of a total coronary heart disease risk score in type 2 diabetes mellitus, Am J Cardiol 101 (2008) 596, 601. [PubMed: 18308005]

[17]. Yang X, So WY, Kong APS, Ho CS, Lam CWK, Stevens RJ, Lyu RR, Yin DD, Cockram CS, Tong PCY, Wong V, Chan JCN, Development and validation of stroke risk equation for Hong Kong Chinese patients with type 2 diabetes: the Hong Kong Diabetes Registry, Diabetes Care 30 (2007) 65–70. [PubMed: 17192335]

[18]. Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM, 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 56 (1925) (2013) 1933.

[19]. Kakadiaris IA, Vrigkas M, Yen AA, Kuznetsova T, Budoff M, Naghavi M, Machine learning outperforms ACC/AHA CVD risk calculator in MESA, J. Am. Heart Assoc. 7 (2018), e009476. [PubMed: 30571498]

[20]. Kengne AP, Patel A, Marre M, Travert F, Lievre M, Zoungas S, Chalmers J, Colagiuri S, Grobbee DE, Hamet P, Heller S, Neal B, Woodward M, Contemporary model for cardiovascular risk prediction in people with type 2 diabetes, Eur. J. Cardiovasc. Prev. Rehabil. 18 (2011) 393–398. [PubMed: 21450612]

[21]. Donnan PT, Donnelly L, New JP, Morris AD, Derivation and validation of a prediction score for major coronary heart disease events in a U.K. type 2 diabetic population, Diabetes Care 29 (2006) 1231–1236. [PubMed: 16732001]

[22]. Elley CR, Robinson E, Kenealy T, Bramley D, Drury PL, Derivation and validation of a new cardiovascular risk score for people with type 2 diabetes, Diabetes Care 33 (2010) 1347–1352. [PubMed: 20299482]

[23]. Davis WA, Knuiman MW, Davis TME, An Australian cardiovascular risk equation for type 2 diabetes: the Fremantle Diabetes Study, Intern. Med. J. 40 (2010) 286–292. [PubMed: 19323700]

[24]. Zethelius B, Eliasson B, Eeg-Olofsson KA-M Svensson S Gudbjörnsdottir J Cederholm, A new model for 5-year risk of cardiovascular disease in type 2 diabetes, from the Swedish National Diabetes Register (NDR), Diabetes Res. Clin. Pract. 93 (2011) 276–284. [PubMed: 21719139]

[25]. Hippisley-Cox J, Coupland C, Development and validation of risk prediction equations to estimate future risk of heart failure in patients with diabetes: a prospective cohort study, BMJ Open 5 (2015), e008503.

[26]. 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 randomised trials, Lancet Diabetes Endocrinol. 5 (2017) 788–798. [PubMed: 28803840]

[27]. Stevens RJ, Kothari V, Adler AI, Stratton IM, Holman RR, The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56), Clin Sci 101 (671) (2001) 679.

[28]. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, Greenland P, Lackland DT, Levy D, O'Donnell CJ, Robinson JG, Schwartz JS, Shero ST, Smith SC, Sorlie P, Stone NJ, F Wilson PW, ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Circulation 129 (2013) S49–S73, 2014. [PubMed: 24222018]

[29]. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, General cardiovascular risk profile for use in primary care: the Framingham Heart Study, Circulation 117 (743) (2008) 753.

[30]. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, Backer GD, Bacquer DD, Ducimetière P, Jousilahti P, Keil U, Njølstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham IM, Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project, Eur. Heart J. 24 (2003) 987–1003. [PubMed: 12788299]

[31]. McKinney W, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, O'Reilly Media, Inc., 2017.

[32]. Walsh CG, Sharman K, Hripcsak G, Beyond discrimination: a comparison of calibration methods and clinical usefulness of predictive models of readmission risk, J. Biomed. Inform. 76 (2017) 9–18. [PubMed: 29079501]

[33]. Zadrozny B, Elkan C, Transforming classifier scores into accurate multiclass probability estimates, in: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2022, pp. 694–699.

[34]. Echouffo-Tcheugui JB, Kengne AP, Comparative performance of diabetes-specific and general population-based cardiovascular risk assessment models in people with diabetes mellitus, Diabetes Metab. 39 (2013) 389–396. [PubMed: 24080092]

[35]. Weiskopf NG, Bakken S, Hripcsak G, Weng C, A data quality assessment guideline for electronic health record data reuse, Egems (2017) 5.

[36]. Haneuse S, Arterburn D, Daniels MJ, Assessing missing data assumptions in EHR-based studies: a complex and underappreciated task, JAMA Netw. Open 4 (2021) e210184–e210184. [PubMed: 33635321]

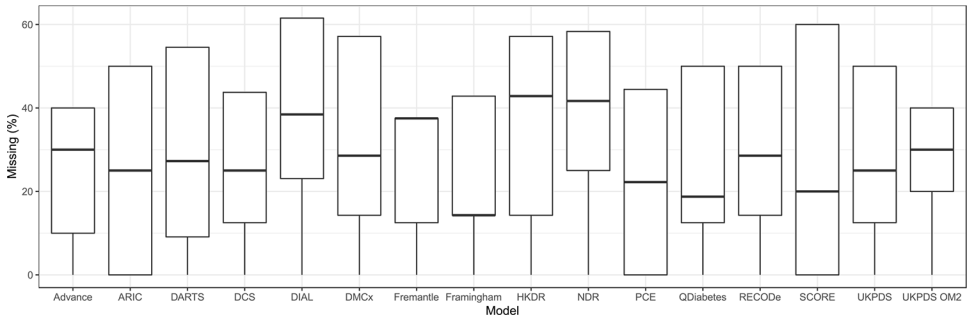

Fig. 1.

A boxplot of the missing predictor variables across the 16 risk prediction models for the CVD endpoint.

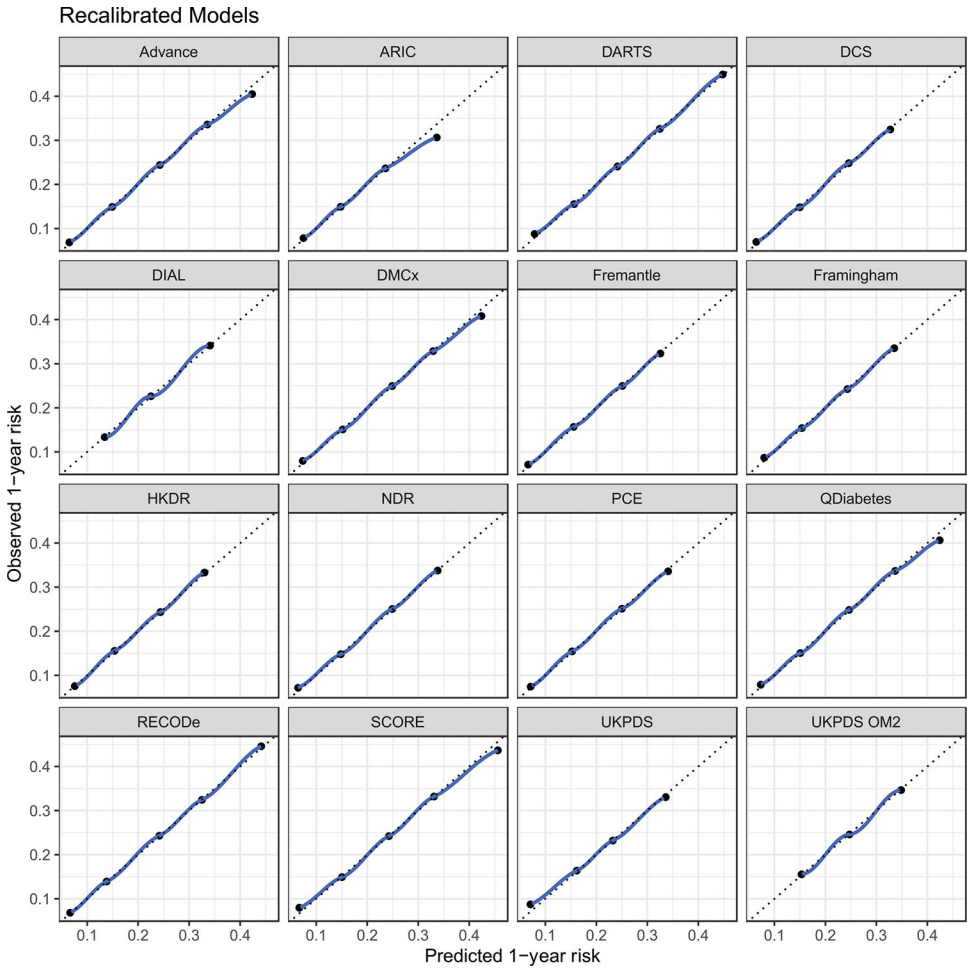

Fig. 2.

Calibration plots for the recalibrated risk score models on the CVD outcome. The dashed gray line reflects perfect agreement between observed and predicted risk.

Table 1

Characteristics of the 16 CVD risk prediction models.

Score Name	Cohort Description	Cohort type	Time frame	Outcome (years)	Model type	CHD	CHF	MI	Stroke	Internal C-Statistic	Missing Data
Advance [20]	7168 subjects from 20 countries in Asia, Australasia, Europe and Canada	Randomized Controlled Trial	2010–2014	4	Cox			X	X	0.70	Not reported
ARIC [12]	1273 patients from 4 United States communities	Observational cohort study	1987–1989	10	Cox	X		X		0.76–0.78	Excluded
DARTS [21]	4569 patients from Scotland	Regional Diabetes Registry	1995–2004	5	Weibull	X		X		0.71	Excluded
DCS [22]	36,127 patients from New Zealand	Observational cohort study	2000–2009	5	Cox	X		X		0.68	Excluded
DIAL [13]	292,024 patients from Swedish	National Diabetes Registry	2002–2012	1/10	Cox			X	X	0.83	Predictive mean matching
DMCx [14]	137,935 patients in Hong Kong	Retrospective cohort study	2010	5	Cox	X	X	X	X	0.71–0.72	Multiple imputation with chained equations
Framingham [29]	8491 patients from Framingham, MA	Observational cohort study	1968–1987	10	Cox	X	X	X	X	0.76–0.79	Excluded
Fremantle [23]	1240 patients from Western Australia	Observational cohort study	1993–1996	5	Cox			X	X	0.80	Not reported
HKDR [15–17]	7067/7209 patients in Hong Kong	National Diabetes Registry	1995–2005	5	Cox	X	X		X	0.70–0.85	Multiple imputation with chained equations
NDR [24]	24,288 patients in Sweden	National Diabetes Registry	2002–2007	5	Cox	X			X	0.72	Not reported
PCE [28]	24,626 subjects from 4 United States cohort studies	Observational cohort study	1968–1993	5/10	Cox	X		X	X	0.70–0.81	Not reported
QDiabetes [25]	437,806 patients in general practice	Observational cohort study	1998–2014	10	Cox		X			0.76–0.78	Multiple imputation with chained equations
RECODe [26]	9635 patients from ACCORD study	Randomized Controlled Trial	2001–2009	10	Cox		X	X	X	0.69–0.75	Multiple imputation with chained equations
SCORE [30]	205,178 persons from 12 European cohort studies	Observational cohort study	1967–1988	10	Weibull	X	X	X	X	0.71–0.84	Not reported
UKPDS [27]	4540 subjects from the United Kingdom	Randomized Controlled Trial	1977–1991	variable	Gompertz	X		X		Not reported	Excluded

Score Name	Cohort Description	Cohort type	Time frame	Outcome (years)	Model type	CHD	CHF	MI	Stroke	Internal C-Statistic	Missing Data
UKPDS OM2 [18]	5102 subjects from the United Kingdom	Randomized Controlled Trial	1977–1991	variable	Weibull		X	X	X	Not reported	Excluded

Table 2

Baseline characteristics of study population and subgroups of interest. For the vital signs, the number in the parenthesis (x%) denotes the percentage of patients with missing predictor values. HTN, hypertension; CHD, coronary heart disease; CHF, congestive heart failure; AFib, atrial fibrillation; SBP, systolic blood pressure; BMI, body mass index; HbA1c, Hemoglobin A1c; eGFR, estimated glomerular filtration rate.

	CHD CASE (*n* = 5923)	**CONTROL (*n* = 33,264)**	**CHF CASE (*n* = 4355)**	**CONTROL (*n* = 39,252)**	**CVD CASE (*n* = 7840)**	**CONTROL (*n* = 28,780)**	**MI CASE (*n* = 2304)**	**CONTROL (*n* = 43,626)**	**STROKE CASE (*n* = 2390)**	**CONTROL (*n* = 43,120)**
AGE (YEARS)	63.82 ± 10.26	58.53 ± 12.69	63.72 ± 11.33	58.93 ± 12.37	63.22 ± 10.86	58.08 ± 12.73	63.99 ± 10.44	60.25 ± 12.39	64.20 ± 11.10	60.13 ± 12.35
SEX (%)										
FEMALE	43.9	57.5	49.7	52.8	49.3	57.3	44.5	52.8	52.8	52.1
MALE	56.1	42.4	50.3	47.2	50.7	42.7	55.5	47.2	47.2	47.9
RACE(%)										
CAUCASIAN	51.1	46.4	46.7	51.3	49.0	47.8	51.3	49.8	52.6	50.5
BLACK	42.9	47.2	48.6	42.2	45.2	45.4	42.4	44.0	42.2	43.3
ASIAN	2.5	2.4	1.4	2.5	2.2	2.5	2.7	2.4	1.9	2.4
OTHER	0.8	0.8	0.6	0.8	0.8	0.8	0.9	0.7	0.8	0.8
UNKNOWN	2.7	3.2	2.7	3.2	2.8	3.5	2.7	3.1	2.5	3.0
SMOKING (%)										
PREVIOUS	2.4	15.9	25.1	18.3	20.1	15.1	27.9	19.4	22.4	19.9
CURRENT	7.2	6.0	7.5	6.3	6.9	5.7	9.7	6.2	7.5	6.3
NEVER	40.2	44.6	43.5	43.1	41.0	43.4	42.5	43.9	44.9	43.5
UNKNOWN	30.2	33.5	23.8	33.2	32.0	35.8	19.9	30.5	25.2	30.2
BMI (KG/M2)	31.9 ± 9.0 (76.2%)	33.0 ± 12.9 (79.0%)	32.3 ± 11.5 (67.4%)	32.2 ± 10.6 (78.6%)	31.7 ± 8.7 (77.9%)	32.7 ± 12.7 (82.0%)	31.7 ± 9.5 (67.2%)	32.7 ± 11.6 (75.9%)	32.0 ± 12.1 (70.4%)	32.7 ± 11.2 (75.2%)
HBA1C (%)	7.4 ± 2.0 (71.9%)	7.4 ± 2.0 (70.3%)	7.5 ± 2.1 (70.1%)	7.4 ± 2.0 (71.3%)	7.4 ± 2.0 (73.1%)	7.5 ± 2.2 (72.3%)	7.4 ± 1.9 (69.3%)	7.4 ± 1.9 (69.6%)	7.6 ± 2.1 (68.5%)	7.4 ± 2.0 (69.6%)
EGFR (ML/MIN/ 1.73 M2)	52.1 ± 22.6 (57.7%)	60.6 ± 17.4 (50.8%)	49.4 ± 23.5 (51.4%)	61.0 ± 16.9 (51.9%)	52.2 ± 23.0 (56.8%)	62.1 ± 15.9 (52.9%)	49.7 ± 23.3 (48.9%)	59.1 ± 18.5 (50.1%)	52.8 ± 21.8 (46.1%)	59.1 ± 18.6 (50.0%)
SBP (MMHG)	140 ± 24 (25.2%)	137 ± 22 (31.9%)	142 ± 25 (18.4%)	137 ± 21 (30.6%)	141 ± 24 (26.5%)	137 ± 21 (34.1%)	140 ± 26 (16.0%)	137 ± 22 (28.8%)	141 ± 25 (20.8%)	137 ± 22 (28.4%)
TOTAL CHOL. (MMOL/L)	4.3 ± 1.3 (72.2%)	4.5 ± 1.2 (68.0%)	4.2 ± 1.3 (70.0%)	4.5 ± 1.2 (67.8%)	4.4 ± 1.3 (73.6%)	4.5 ± 1.2 (68.7%)	4.3 ± 1.3 (70.0%)	4.4 ± 1.2 (67.3%)	4.4 ± 1.3 (69.9%)	4.4 ± −1.2 (66.9%)
AFIB (%)	9.4 (0%)	5.0 (0%)	13.1 (0%)	5.0 (0%)	7.2 (0%)	3.3 (0%)	10.5 (0%)	7.4 (0%)	10.8 (0%)	7.3 (0%)

	CHD CASE (*n* = 5923)	CONTROL (*n* = 33,264)	CHF CASE (*n* = 4355)	CONTROL (*n* = 39,252)	CVD CASE (*n* = 7840)	CONTROL (*n* = 28,780)	MI CASE (*n* = 2304)	CONTROL (*n* = 43,626)	STROKE CASE (*n* = 2390)	CONTROL (*n* = 43,120)
CHD (%)	0.0 (0%)	0.0 (0%)	35.0 (0%)	17.8 (0%)	0 (0%)	0 (0%)	49.7 (0%)	19.2 (0%)	32.7 (0%)	21.2 (0%)
CHF (%)	16.4 (0%)	7.9 (0%)	0.0 (0%)	0.0 (0%)	0 (0%)	0 (0%)	27.0 (0%)	12.2 (0%)	22.4 (0%)	12.8 (0%0
STROKE (%)	5.7 (0%)	3.8 (0%)	7.5 (0%)	5.0 (0%)	0 (0%)	0 (0%)	8.3 (0%)	5.5 (0%)	6.7 (0%)	3.4 (0%)
HTN (%)	13.5 (0%)	9.2 (0%)	18.0 (0%)	8.6 (0%)	12.3 (0%)	6.7 (0%)	18.2 (0%)	11.1 (0%)	18.0 (0%)	11.0 (0%)

Table 3

Risk calculation comparison based on the C-statistic and the 95% CI and mean imputation.

MODEL	CHD	CHF	CVD	MI	STROKE
ADVANCE	0.66 (0.66–0.67)	0.63 (0.63–0.64)	0.65 (0.65–0.66)	0.62 (0.62–0.63)	0.62 (0.61–0.62)
ARIC	0.61 (0.61–0.61)	0.55 (0.55–0.56)	0.59 (0.58–0.59)	0.57 (0.57–0.58)	0.53 (0.53–0.54)
DARTS	0.64 (0.63–0.64)	0.58 (0.58–0.59)	0.63 (0.62–0.63)	0.60 (0.59–0.60)	0.59 (0.59–0.60)
DCS	0.64 (0.64–0.65)[a]	0.62 (0.62–0.63)[a]	0.64 (0.64–0.64)[a]	0.61 (0.60–0.62)[d]	0.63 (0.62–0.63)[a]
DIAL	0.59 (0.59–0.59)	0.59 (0.59–0.60)	0.59 (0.59–0.59)	0.60 (0.60–0.61)	0.57 (0.57–0.58)
DMCX	0.67 (0.66–0.67)	0.65 (0.65–0.66)	0.66 (0.65–0.66)	0.65 (0.64–0.66)	0.62 (0.62–0.63)
FREMANTLE	0.63 (0.63–0.64)	0.62 (0.61–0.62)	0.63 (0.63–0.63)	0.60 (0.60–0.61)	0.63 (0.62–0.63)
FRAMINGHAM	0.65 (0.64–0.65)	0.60 (0.60–0.61)	0.63 (0.63–0.64)	0.61 (0.61–0.62)	0.60 (0.60–0.60)
HKDR	0.65 (0.65–0.65)[b]	0.65 (0.65–0.66)[c]	0.64 (0.64–0.64)[c]	0.64 (0.63–0.65)[d]	0.62 (0.62–0.62)[e]
NDR	0.65 (0.65–0.65)	0.61 (0.61–0.62)	0.64 (0.64–0.64)	0.61 (0.60–0.62)	0.61 (0.60–0.61)
PCE	0.64 (0.63–0.64)	0.59 (0.59–0.59)	0.63 (0.62–0.63)	0.60 (0.59–0.60)	0.59 (0.59–0.60)
QDIABETES	0.65 (0.64–0.65)	0.62 (0.62–0.62)	0.64 (0.64–0.64)	0.61 (0.60–0.62)	0.60 (0.60–0.60)
RECODE	0.65 (0.64–0.65)[c]	0.67 (0.67–0.68)[c]	0.66 (0.66–0.66)[c]	0.64 (0.63–0.65)[d]	0.61 (0.60–0.61)[e]
SCORE	0.63 (0.63–0.64)	0.58 (0.58–0.59)	0.62 (0.62–0.63)	0.59 (0.58–0.60)	0.60 (0.59–0.60)
UKPDS	0.62 (0.61–0.62)	0.55 (0.54–0.55)	0.60 (0.60–0.60)	0.58 (0.57–0.59)	0.55 (0.55–0.55)
UKPDS OM2	0.58 (0.58–0.59)[c]	0.60 (0.59–0.60)[c]	0.61 (0.61–0.62)[c]	0.64 (0.63–0.65)[d]	0.61 (0.61–0.61)[e]

For scores with multiple equations

[a]=CVD

[b]=CHD

[c]=CHF

[d]=MI

[e]=Stroke.