ESC European Society of Cardiology

European Heart Journal - Digital Health (2025) **6**, 466–475
https://doi.org/10.1093/ehjdh/ztaf023

ORIGINAL ARTICLE

A hybrid algorithm-based ECG risk prediction model for cardiovascular disease

Pan Zhou[1,2,3,4,5,†], Zhao Yang[1,2,3,4,5,†], Yiming Hao[1,2,3,4,5,†], Fangfang Fan[6,7,†], Wenlang Zhao[1,2,3,4,5], Ziyu Wang[1,2,3,4,5], Qiuju Deng[1,2,3,4,5], Yongchen Hao[1,2,3,4,5], Na Yang [1,2,3,4,5], Lizhen Han[1,2,3,4,5], Pingping Jia[1,2,3,4,5], Yue Qi[1,2,3,4,5], Yan Zhang[6,7,*], and Jing Liu [1,2,3,4,5,*]

[1]Center for Clinical and Epidemiologic Research, Beijing Anzhen Hospital, Capital Medical University, Beijing, China; [2]Beijing Institute of Heart, Lung and Blood Vessel Diseases, Beijing, China; [3]National Clinical Research Center for Cardiovascular Diseases, Beijing, China; [4]The Key Laboratory of Remodeling-Related Cardiovascular Diseases, Ministry of Education, Beijing, China; [5]The Beijing Municipal Key Laboratory of Clinical Epidemiology, Beijing, China; [6]Institute of Cardiovascular Disease, Peking University First Hospital, No.8 Xishiku Road, Xicheng District, Beijing 100034, China; and [7]Department of Cardiology, Peking University First Hospital, No.8 Xishiku Road, Xicheng District, Beijing 100034, China

Received 4 December 2024; revised 8 February 2025; accepted 21 February 2025; online publish-ahead-of-print 19 March 2025

Aims Little is known about the role of electrocardiography (ECG) in the community population independent of physical and laboratory examinations. Thus, this study developed and validated several ECG-based models for cardiovascular disease (CVD) risk assessment, with or without simple questionnaire-based variables.

Methods and results Using a derivation cohort of 3734 Chinese participants aged ≥40 years, we developed the ECG-based models to predict the risk of developing CVD (comprising fatal and non-fatal coronary heart disease, unstable angina, stroke, and heart failure). Candidate predictors associated with CVD were screened from hundreds of ECG characteristics using a hybrid algorithm. By incorporating the questionnaire-based predictors, we constructed the ECG–questionnaire model. All models were tested in an external validation cohort ($n = 1224$) to determine their discrimination and calibration. Over a maximum follow-up of 7 years, 433 CVD events occurred in the derivation cohort. The ECG model with 37 selected features achieved comparable performance concerning the clinical model using traditional cardiovascular risk factors (*C*-statistic: 0.690, 95% confidence interval [CI]: 0.638–0.743) in the external validation cohort. Such performance significantly improved when the questionnaire-based predictors were added (*C*-statistic: 0.734, 95% CI: 0.685–0.784; calibration χ^2: 3.334, $P = 0.950$). Compared with the clinical model, 17.4% of the participants were correctly assigned to the corresponding risk groups, with an absolute integrated discrimination index of 0.048 (95% CI: 0.016–0.080).

Conclusion The ECG model with/without questionnaire-based variables can accurately predict future CVD risk independent of physical and laboratory examinations, suggesting its great potential in routine clinical practice.

Lay summary New electrocardiography (ECG)-based risk prediction models with or without questionnaire-based variables were developed to predict 7-year cardiovascular risk in the community population. Independent of physical and laboratory tests, the ECG model has comparable performance to the traditional cardiovascular risk factors and outperformed predictive accuracy when questionnaire-based information is added. This study emphasizes the role of ECG in cardiovascular risk assessment, as a low-cost, non-invasive, and ubiquitous clinical test. The ECG-based model can be embedded in ECG examination to support quick cardiovascular risk assessment without having to wait for laboratory or imaging tests, which may be more suitable for clinical practitioners when initiating risk-based lifestyle changes and pharmaceutical interventions.

Keywords Electrocardiography • Cardiovascular disease • Predictive value • Risk assessment

* Corresponding authors. Tel: + 86 10 64456710, Fax: + 86 10 64419738, Email: jingliu@ccmu.edu.cn (J.L.), Tel: + 86 10 83575260, Fax: + 86 10 66551211, Email: drzhy1108@163.com (Y.Z.)
†The first four authors contributed equally to the study.

Introduction

Cardiovascular disease (CVD) remains the leading cause of death, both in China and worldwide.[1,2] Accurately assessing CVD risk is paramount for identifying individuals who would benefit most from lifestyle or pharmaceutical interventions.[3–5] Nowadays, several traditional cardiovascular risk factor-based tools for predicting CVD have been recommended in current prevention guidelines issued by the Chinese Society of Cardiology,[3] European Society of Cardiology,[4] and American College of Cardiology/American Heart Association,[5] aiming to guide clinical decision-making around the initiation of intervention strategies and actions, such as the management of dyslipidaemia[6–8] and hypertension.[9–11]

Electrocardiography (ECG), a low-cost, non-invasive, and ubiquitous standard clinical test, is used for the auxiliary diagnosis of cardiac diseases.[4] Changes in the electrical activity of the heart captured by ECG can reflect cardiac hypertrophy-[12] or ischaemia-oriented[13] changes in myocardial cells during depolarisation and repolarization. Some classic ECG features that characterized ischaemic abnormalities in ECG (e.g. QRS complex and T waves) are associated with various adverse CVD events, supporting their potential in CVD risk assessment.[14–16] However, previous studies focusing on single[17–22] or multiple[23–25] well-established ECG features have shown mild-to-moderate improvements in CVD risk prediction when added to traditional models, resulting in an existing debate.[26]

Notably, rich digital ECG signals can be automatically collected alongside hundreds of ECG-derived features during examination. However, using these ECG features remains challenging under conventional analytical frameworks, exemplified by the high correlation between ECG features, the implicitly non-linear association, and the high dimensionality of ECG parameters. In such cases, machine learning algorithms may provide a more appropriate avenue for feature selection and model development, given the enormity and complexity of ECG data.[27] For example, excellent performance has been noted when using hundreds of ECG features to predict all-cause death[28,29] or CVD[30–32] risk in disease-specific populations, particularly after incorporating physical examinations, laboratory tests, or imaging information. Clinical guidelines recommend long-term risk assessment for CVD as the cornerstone for primary prevention. Consequently, taking advantage of the non-invasive, rapid, simple, and reproducible characteristics of ECG, an algorithm fully utilising these digital ECG signals may improve CVD risk prediction and provide a better risk assessment tool for the primary prevention of CVD.

In this study, we developed and validated a novel ECG model to predict 7-year CVD risk merely using selected ECG features *via* machine learning algorithms. We also evaluated whether adding easy-to-obtain questionnaire-based variables improved the predictive capability of the ECG model, based on two prospective community-based cohorts.

Methods

Study design and population

Chinese multi-provincial cohort study-Beijing project 2012 visit

The Chinese multi-provincial cohort study (CMCS)-Beijing Project is an ongoing community-based cohort study that initially recruited participants free of CVD from the Beida and Shougang communities during 1992–1993 and from the Beida community during 1996–1999 in Beijing. The study design and enrolment criteria have been described elsewhere.[33,34] In the present study, the 2012 visit of the CMCS-Beijing Project was considered the baseline visit. The data of all participants at baseline were collected *via* face-to-face interviews or standardized questionnaires modified from the World Health Organization's MONICA protocol,[35] in-person physical examinations, ECG examination, or laboratory testing using overnight blood samples. All participants were actively followed up to record the onset of incident CVD or death every 1–2 years up to the end of 2020, supplemented by the local disease surveillance systems. This study was approved by the ethics committee of Beijing Anzhen Hospital, Capital Medical University, and all participants provided written informed consent.

Shougang community-based cohort 2014 visit

The design of the Shougang community-based cohort has been reported previously.[36] Briefly, participants aged ≥40 years were enrolled from 2011 to 2012 in the Shougang area in Beijing. In the present study, we considered the 2014 visit as the baseline visit. Individual-level data, including demographic characteristics, lifestyle factors, medical history, and medication use, were collected by trained researchers using standardized questionnaires after obtaining written informed consent, with clinical measurements and ECG tests being performed during the physical examinations. An overnight blood sample was also collected from each participant and tested in the laboratory. All participants were followed up from the time of the survey to the end of 2021 to record the onset of CVD or death. The study was approved by the ethics committee of Peking University First Hospital.

In this study, 5014 participants aged ≥40 years who lived in the Shougang community from both cohorts were considered the derivation cohort. We excluded 687 participants with prevalent CVD, 572 without 12-lead ECG records, two with pacemakers, and 19 with missing values for critical predictors, leaving 3734 participants in the derivation cohort. The external validation cohort comprised 1449 participants from the CMCS-Beijing Project in the Beida community. We excluded 186 individuals with prevalent CVD, 34 without ECG records, and five with pacemakers, leaving 1224 participants in the external validation cohort, as depicted in *Figure 1*.

Risk factor measurements

Demographics (age and sex), smoking status, medical history (hypertension, diabetes mellitus, and dyslipidaemia), and medication use (anti-hypertensive and glucose-lowering medications and statins) were collected using standardized questionnaires by trained researchers. Blood pressure was measured during physical examinations. Current smoker was defined as ≥1 cigarette per day for at least 3–6 months. The medical history of hypertension, diabetes mellitus, or dyslipidemia was self-reported. The use of anti-hypertensive and glucose-lowering medications, or statins within the past 4 weeks was ascertained by document records. Diabetes mellitus was defined as fasting blood glucose (FBG) ≥7.0 mmol/L, a self-reported physician diagnosis of diabetes mellitus, or documented glucose-lowering drug use within the past 4 weeks.

Laboratory measurements in this study followed well-established procedures, as detailed previously.[34,37,38] Briefly, overnight fasting venous blood samples were used to measure FBG, total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), and low-density lipoprotein cholesterol (LDL-C). FBG and TC were measured using standard enzymatic methods, and HDL-C and LDL-C were measured using a homogeneous method.

Electrocardiography

Standard 12-lead ECGs were collected using the FX-8322 ECG machine (Beijing Fukuda Denshi Medical Instruments Co., Ltd.) in the CMCS-Beijing Project and the GY-5000 digital multi-channel ECG machine (HeNan HuaNan Medical Science and Technology Co., Ltd.) in the Shougang community-based cohort. During the physical examinations, the sample rate was 500 Hz, and the duration was 10 s. Raw digital ECG files were stored in the ECG machines.

All recorded ECG signals were preprocessed and decompressed to extract the ECG features in all leads using the manufacture-specific commercial software EFS-200C (proprietary digital format, Beijing Fukuda Denshi Medical Instruments Co., Ltd.) in the CMCS-Beijing Project and the software 12DECG (proprietary digital format, HeNan HuaNan Medical Science and Technology Co., Ltd.) in the Shougang community-based cohort. Then, 12 features measuring signal noise were removed from the set of ECG features in the CMCS-Beijing Project, focusing on 418 shared temporal–spatial ECG features (e.g. amplitude, duration, and area measures) in both cohorts. Supplementary material online, *Table S1* presents the details of all extracted ECG features in the derivation and external

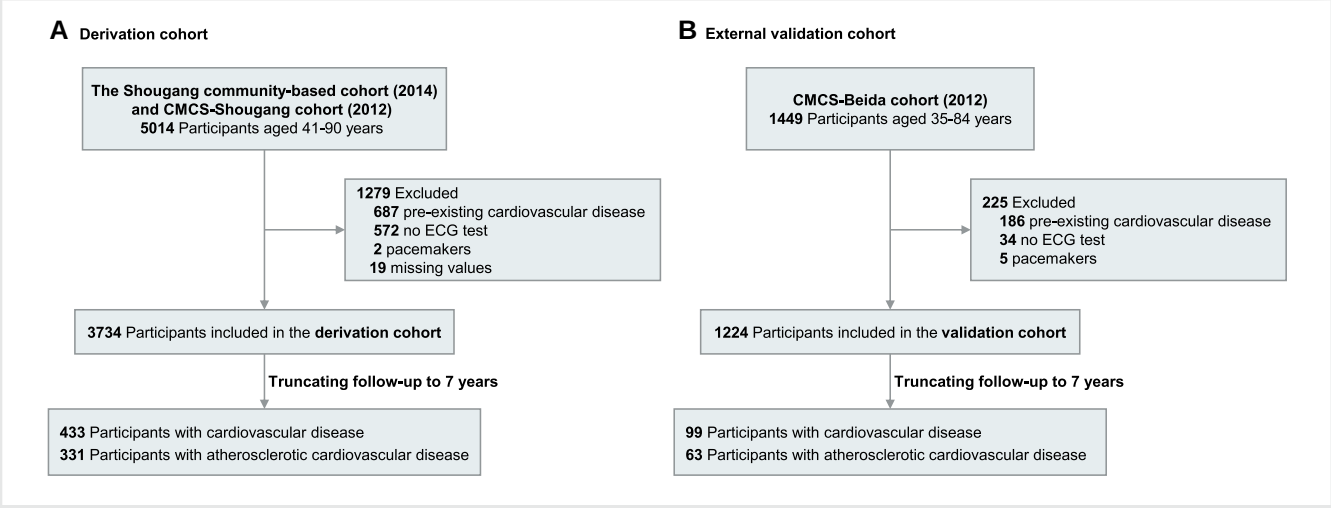

Figure 1 Flowcharts of the derivation and external validation cohorts used to develop and validate the CVD risk prediction models.

validation cohorts, including atrial waves (e.g. P-wave or PR interval), ventricular waves (e.g. QRS complex, T-wave, U-wave, ST segment, QT interval, or ventricular activation time), cardiac cycle waves (e.g. heart rate, RR interval, or PP duration), and arrhythmic waves (e.g. fascicular ventricular tachycardia time, number of premature ventricular contraction, or number of premature atrial contraction).

Outcome ascertainment

The primary outcome was incident CVD events, defined as a composite of acute coronary heart disease (CHD), stroke, and hospitalisation for heart failure. Acute CHD events included acute myocardial infarction, fatal CHD, coronary revascularization, and unstable angina. Acute stroke events included hemorrhage, ischemic, and uncategorized stroke. Heart failure was defined as a hospital discharge diagnosis with International Classification of Disease, 10th Revision, code I50 based on the electronic medical records. The secondary outcome was incident atherosclerotic CVD (ASCVD), comprising CHD and ischaemic stroke events. In the CMCS-Beijing Project, all CHD, stroke, and death events were supplemented *via* the Beijing Hospital Discharge Information System and the Beijing Vital Registration Monitoring System, whereas in the Shougang community-based cohort, the Chinese Center for Disease Control and Prevention-National Mortality Surveillance System and the Beijing Municipal Health Commission-Beijing inpatient medical record home page system were used and then adjudicated by a panel of physicians.

ECG feature selection

A comprehensive data-driven procedure was developed to select candidate predictors among the 415 shared non-zero ECG features. Specifically, we removed 66 ECG features for which the proportion of zero values was 95% or higher,[39] leaving 349 available ECG features that were *z*-score normalized. Then, we selected predictors by removing highly correlated ECG features using the following algorithm: (i) select highly correlated feature pairs (defined as those with a Pearson's correlation $\rho \geq r$, i.e. $r \in [0.6, 0.9]$, with $P < 0.05$) based on pairwise Pearson's correlations among the 349 ECG features; (ii) retain the feature with the highest feature-outcome correlation (e.g. Pearson's, Spearman's, Kendall's, or distance's correlation) among those highly correlated pairs; and (iii) repeat (ii) iteratively until all highly correlated pairs have been filtered. After that, candidate predictors were further reduced using elastic net or Least Absolute Shrinkage and Selection Operator (LASSO) penalized Cox regression alongside a stepwise selection procedure.[40]

Finally, after removing 191 features highly correlated features for pairwise correlations, 95 with LASSO regression, and 26 with backward selection, we identified 37 ECG features as predictors for the model that achieved the best predictive performance under 10-fold cross-validation in the derivation cohort.

Model development and validation

We developed four risk models for various purposes in the derivation cohort. We first constructed an ECG model with 37 selected ECG features to determine whether only using ECG features could accurately predict CVD risk. Then, we constructed an ECG–questionnaire model to evaluate whether incorporating questionnaire-based variables (i.e. age, sex, smoking status, medical history of hypertension, diabetes mellitus, and dyslipidaemia) improved the predictive capability of the ECG model. For comparison, we also developed an ECG–clinical model by adding traditional cardiovascular risk factors (i.e. age, sex, smoking status, systolic blood pressure [SBP], HDL-C, LDL-C, diabetes mellitus, and use of anti-hypertensive medications) recommended by the Chinese Society of Cardiology for the primary prevention of ASCVD.[3] Moreover, we developed sex-specific models for predicting 7-year CVD risk using these traditional cardiovascular risk factors as the reference model (clinical model) for all ECG-based models to determine any improvement in prediction. Of note, all ECG-based models were non-sex-specific to ensure an adequate sample size. We verified the proportional-hazards assumption for all models using the Schoenfeld residual test.

All models were internally validated using 10-fold cross-validation in the derivation cohort and externally validated in the external validation cohort, to determine their overall performance (explained variance [R^2]), discrimination (Harrell's *C* statistic), and calibration (calibration χ^2 and slope). Specifically, R^2 quantifies the degree to which the models explain the CVD risk variation, in which a higher R^2 value signifies a better model. Harrel's *C*-statistic quantifies the probability of the correct order concerning shorter time-to-event outcomes for participants with higher predicted risks in a randomly selected pair, with a higher value indicating better model discrimination.[41] Calibration was evaluated by comparing the mean predicted 7-year CVD risk with the observed risk and assessed using the modified Greenwood-Nam-D'Agostino test.[42] The calibration slope was used to assess the agreement between the predicted and observed risks, with a value of one suggesting perfect calibration. Given the distinct incidence of cardiovascular events between the derivation and external validation cohorts, all models were recalibrated for the baseline survival rate in the external validation cohort.

The categorical net reclassification improvement (NRI)[43] and absolute integrated discrimination index (IDI)[44] were computed to quantify the improvement in the predictive capability of all ECG-based models compared with the clinical model. Decision curve analysis was conducted to evaluate the clinical utility of the ECG-based models by assessing the trade-off between the benefits from true positives and the harms from false positives across a range of threshold probabilities.[45]

Table 1 Baseline characteristics of the participants in the derivation and external validation cohorts

Characteristics, *N*. (%)	Derivation cohort ($n = 3734$)	External validation cohort ($n = 1224$)	*P* value
Age, years	59.2 ± 8.3	67.3 ± 9.0	<0.001
Male	1367 (36.6)	527 (43.1)	<0.001
Current smoker	604 (16.2)	113 (9.2)	<0.001
SBP, mmHg	127.9 ± 16.8	137.6 ± 16.2	<0.001
SBP if treated, mmHg	135.0 ± 16.0	143.3 ± 14.7	<0.001
SBP if not treated, mmHg	125.7 ± 16.4	132.6 ± 15.7	<0.001
DBP, mmHg	75.7 ± 9.8	79.3 ± 8.9	<0.001
FBG, mmol/L	5.78 ± 1.44	5.68 ± 1.24	0.024
TC, mmol/L	5.13 ± 0.96	5.22 ± 1.03	0.010
HDL-C, mmol/L	1.24 ± 0.29	1.36 ± 0.31	<0.001
LDL-C, mmol/L	3.02 ± 0.79	3.06 ± 0.86	0.084
History of hypertension	1208 (32.4)	617 (50.4)	<0.001
History of diabetes mellitus[a]	511 (13.7)	209 (17.1)	0.004
History of dyslipidemia	1135 (30.4)	601 (49.1)	<0.001
Diabetes mellitus[b]	654 (17.5)	234 (19.1)	0.220
Anti-hypertensive drugs	881 (23.6)	570 (46.6)	<0.001
Glucose-lowering drugs	307 (8.2)	173 (14.1)	<0.001
Statin therapies	170 (4.6)	233 (19.0)	<0.001

DBP, diastolic blood pressure; FBG, fast blood glucose; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; SBP, systolic blood pressure; TC, total cholesterol.
[a]Self-reported diabetes mellitus previously diagnosed by physicians.
[b]Defined by FBG ≥7.0 mmol/L, a self-reported physician diagnosis of diabetes mellitus, and documented use of the glucose-lowering drug within the past 4 weeks.

Performance of all CVD models for predicting ASCVD risk

To demonstrate the general predictive capability of all of ECG-based CVD risk prediction models, we calculated the ASCVD risk *via* directly multiplying the predicted CVD probability by the proportion of ASCVD events (ratio of Kaplan–Meier event rates). Then we evaluated the performance of these models for predicting ASCVD in the derivation and external validation cohorts.

Feature importance

The feature importance of all predictors in the ECG–questionnaire model was quantified using the decreased explained variance for CVD risk with the leave-one-out approach in the derivation cohort. Moreover, scaled importance was used to give information on the direct comparison between each feature's importance and the most important one based on the ratio between the importance of each feature and the highest importance. We also computed the contribution of all features to the ECG–questionnaire model for predicting CHD and any stroke to characterize the feature importance of each predictor for the different CVD components.

Sensitivity analysis

We performed a sensitivity analysis to evaluate whether using fewer ECG predictors (i.e. 5 or 10) had similar predictive performance as the model using 37 ECG features for 7-year cardiovascular risk. Based on the rank from the leave-one-out R^2 for CVD in the ECG model, the top 5 and 10 ECG predictors were included in the ECG model to predict CVD and ASCVD risk in the derivation and external validation cohorts.

Statistical analysis

The baseline characteristics are described as the mean (SD) or frequency (percentage), as appropriate. The cumulative incidence rate of CVD was calculated using the Kaplan–Meier method. Follow-up years was defined as the duration from the baseline survey to the occurrence of CVD events, a maximum follow-up of 7 years, or administrative censoring, whichever came first. All statistical analyses were performed using R software (version 4.2.3, https://cran.r-project.org/). Sample codes are available at http://github.com/yangzhao98/riskModel.

The study adhered to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis statement.[46]

Result

Baseline characteristics

Table 1 summarizes the baseline characteristics of the study participants in the derivation ($n = 3734$) and external validation ($n = 1224$) cohorts. The mean age in the derivation cohort was 59.2 (8.3) years, with 36.6% of males. In the external validation cohort, the mean age was 67.9 (9.0) years, and 43.1% of the participants were male. The external validation cohort had a lower prevalence of smoking but a higher prevalence of disease history and medication use, alongside higher blood pressure and cholesterol levels. Over a 7-year truncated follow-up period, 433 CVD and 331 ASCVD events occurred in the derivation cohort, yielding cumulative incidence rates of 11.7% (95% CI: 10.6–12.7%) and 9.0% (8.0–9.9%), respectively. In the external validation cohort, 99 CVD and 63 ASCVD events were recorded, with cumulative incidence rates of 8.2% (6.7–9.8%) and 5.3% (4.0–6.5%).

Associations between ECG features and CVD

Supplementary material online, *Table S2* shows the associations between the 37 ECG features (ECG model) and CVD risk in the derivation cohort. Similar results were noted even after adjustment for

Table 2 Summary performance of all derived models for CVD in the derivation and external validation cohorts

Performance metrics	Derivation cohort	External validation cohort	
		Without recalibration	After recalibration[a]
Clinical model[b]			
R^2	0.171 (0.132, 0.211)	0.126 (0.055, 0.223)	0.126 (0.055, 0.223)
C-statistic	**0.715 (0.692, 0.738)**	**0.698 (0.651, 0.745)**	**0.695 (0.647, 0.742)**
Calibration slope	1.049 (1.048, 1.050)	0.561 (0.559, 0.563)	0.901 (0.897, 0.904)
Calibration χ^2	9.295[#]	147.490***	19.378*
Categorical NRI	Reference	Reference	Reference
Absolute IDI	Reference	Reference	Reference
ECG model[c]			
R^2	0.298 (0.236, 0.364)	0.237 (0.129, 0.324)	0.237 (0.129, 0.324)
C-statistic	**0.712 (0.688, 0.737)**	**0.690 (0.638, 0.743)**	**0.690 (0.638, 0.743)**
Calibration slope	1.031 (1.028, 1.033)	0.394 (0.385, 0.404)	0.523 (0.506, 0.540)
Calibration χ^2	4.895[#]	29.648**	4.877[#]
Categorical NRI	−0.069 (−0.129, −0.008)*	0.101 (0.002, 0.201)*	0.022 (−0.114, 0.157)[#]
Absolute IDI	0.020 (0.005, 0.035)**	0.010 (−0.028, 0.047)[#]	0.014 (−0.015, 0.042)[#]
ECG–questionnaire model[d]			
R^2	0.355 (0.308, 0.399)	0.317 (0.206, 0.396)	0.317 (0.206, 0.396)
C-statistic	**0.745 (0.723, 0.767)**	**0.734 (0.685, 0.784)**	**0.734 (0.685, 0.784)**
Calibration slope	0.951 (0.942, 0.959)	0.529 (0.527, 0.532)	0.802 (0.793, 0.812)
Calibration χ^2	11.394[#]	71.813***	3.334[#]
Categorical NRI	0.070 (0.019, 0.122)**	0.174 (0.099, 0.250)***	0.146 (0.024, 0.268)*
Absolute IDI	0.042 (0.029, 0.055)***	0.048 (0.016, 0.080)**	0.032 (0.009, 0.056)**
ECG–clinical model[e]			
R^2	0.370 (0.325, 0.422)	0.320 (0.231, 0.406)	0.320 (0.231, 0.406)
C-statistic	**0.748 (0.726, 0.770)**	**0.738 (0.690, 0.786)**	**0.738 (0.690, 0.786)**
Calibration slope	0.987 (0.983, 0.991)	0.496 (0.491, 0.500)	0.686 (0.674, 0.699)
Calibration χ^2	8.969[#]	63.791***	5.579[#]
Categorical NRI	0.097 (0.050, 0.144)***	0.190 (0.110, 0.271)***	0.190 (0.073, 0.307)**
Absolute IDI	0.050 (0.038, 0.062)***	0.053 (0.022, 0.084)***	0.040 (0.016, 0.064)**

CVD, cardiovascular disease; ECG, electrocardiography; IDI, integrated discrimination improvement; NRI, net reclassification improvement.
[a]Recalibration of the predicted CVD risk for the baseline survival rate in the external validation cohort.
[b]Sex-specific models including age, smoking status, treated or untreated systolic blood pressure, HDL-C, LDL-C, and diabetes mellitus, with continuous variables being natural log-transformed.
[c]Including screened 37 ECG features merely.
[d]Additionally including questionnaire-based variables (i.e. age, sex, smoking status, medical history of hypertension, diabetes mellitus, and dyslipidemia).
[e]Additionally including traditional cardiovascular risk factors (i.e. age, sex, smoking status, systolic blood pressure, HDL-C, LDL-C, diabetes mellitus, and use of anti-hypertensive medications).
[#]$P > 0.05$, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

questionnaire-based variables (ECG–questionnaire model) or traditional cardiovascular risk predictors (ECG–clinical model). The regression coefficients and the baseline survival rates without and after recalibration for each ECG-based model were also provided to calculate the absolute risk of developing CVD within the next 7 years.

Performance of the ECG-based models for CVD

Table 2 presents the performance of all models in the derivation and external validation cohorts. With the inclusion of 37 selected features, the ECG model exhibited good discrimination (*C*-statistic: 0.712, 95% CI: 0.688–0.737) and calibration (slope: 1.031, 1.028–1.033; calibration χ^2: 4.895, $P = 0.843$; *Figure 2A*), which was comparable to the traditional cardiovascular risk factor model (clinical model) in internal validation. Similar performance was observed in the external validation cohort, with a *C*-statistic of 0.690 (95% CI: 0.638–0.743). Notably, the calibration slope was 0.394 (0.385–0.404; calibration χ^2: 29.648, $P = 0.001$), suggesting risk overestimation (*Figure 2B*). After recalibration for the baseline survival rate in the external validation cohort, the ECG model exhibited good calibration (*Figure 2C*).

When incorporating the questionnaire-based variables, the ECG–questionnaire model outperformed the clinical and ECG models, with an explained variance of around one third and a *C*-statistic of 0.734 (0.685–0.784) in the external validation cohort (*Table 2*). The calibration was adequate upon internal validation (*Figure 2A*), but the risk was overestimated upon external validation (*Figure 2B*), with a calibration slope of 0.529 (0.527–0.532) and calibration χ^2 of 71.813 ($P < 0.001$). The recalibrated ECG–questionnaire model was well calibrated (χ^2: 3.334, $P = 0.950$, *Table 2*) in the external validation cohort, as shown in *Figure 2C*. Moreover, compared with the clinical model, the ECG–questionnaire model correctly categorized 17.4% of the

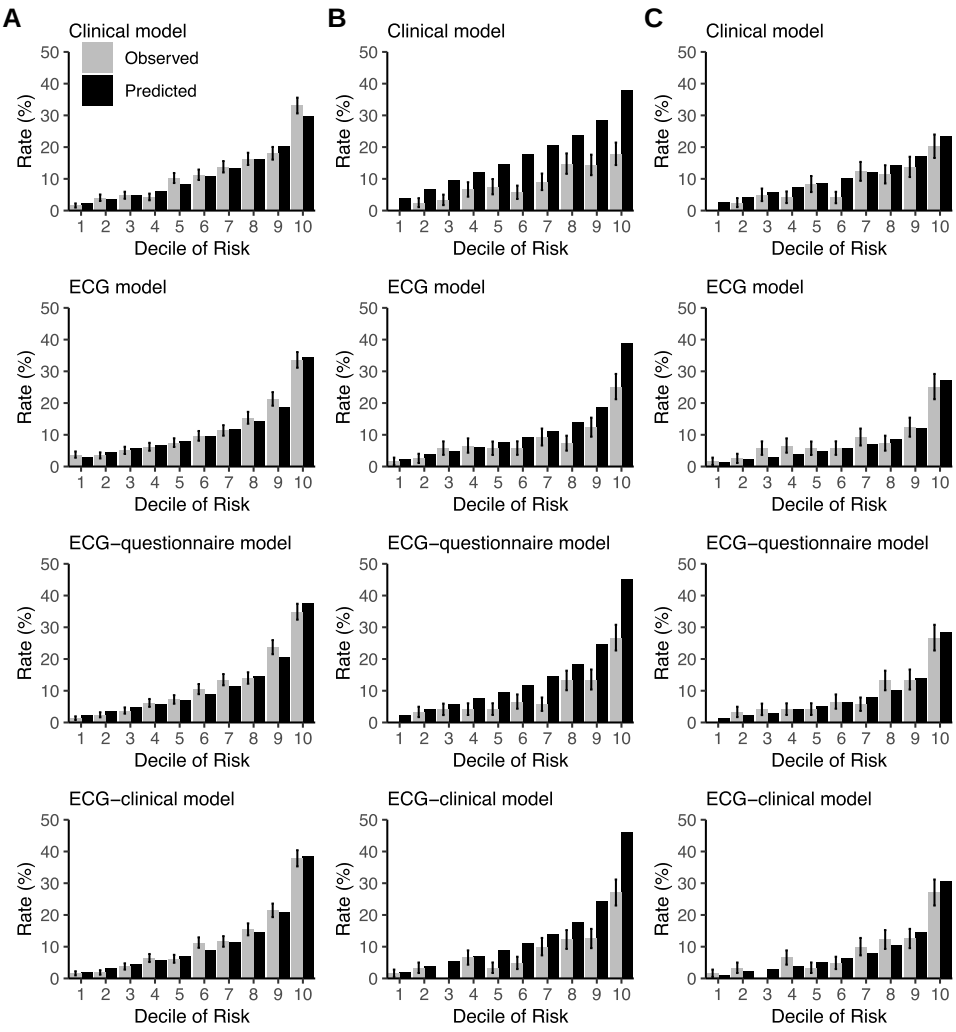

Figure 2 Model calibration for CVD risk prediction in the derivation and external validation cohorts. (*A*) Derivation cohort; (*B*) External validation cohort; (*C*) External validation cohort after recalibration. The gray and black bars indicated observed and predicted risks of CVD by decile, respectively. CVD, cardiovascular disease; ECG, electrocardiography.

participants (9.9–25.0%) participants into the corresponding risk groups, with an absolute IDI of 0.048 (0.016–0.080) in the external validation cohort. Similar risk stratification was noted for the recalibrated ECG–questionnaire model (*Table 2*).

The ECG–clinical model performed similarly in predicting 7-year CVD risk compared with the ECG–questionnaire model, with a *C*-statistic of 0.738 (0.690–0.787) in the external validation cohort. However, the 7-year CVD risk was overestimated in the external validation cohort. Fortunately, the performance improved after recalibration (*Figure 2*). Compared with the clinical model, incorporating traditional cardiovascular risk factors improved the risk stratification of the ECG model, with the categorical NRI of 19.0% (11.0–27.1%) and absolute IDI of 0.053 (0.022–0.084) in the external validation cohort, similar to the recalibrated ECG–clinical model (*Table 2*).

Clinical utility

Figure 3 illustrates the decision curves of the four models by assessing their clinical utility (net benefit) for predicting CVD in the derivation and external validation cohorts. As expected, the ECG model had similar clinical utility to the clinical model. Moreover, the net benefit was higher for the original and recalibrated ECG–questionnaire models in the external validation cohort, similar to the ECG–clinical model.

Performance of the ECG-based models for ASCVD

Similar results were observed when applying the developed ECG-based models for predicting ASCVD. The ECG model had good discrimination (*C*-statistic: 0.657, 0.590–0.734, Supplementary material online, *Table S3*) and calibration after recalibration for ASCVD in the external validation cohort. The inclusion of questionnaire-based predictors improved the predictive performance of the ECG model for ASCVD, with a *C*-statistic of 0.703 (0.637–0.769) in the external validation cohort, particularly after recalibration (see Supplementary material online, *Figure S1*). Moreover, the recalibrated ECG–questionnaire model correctly categorized 15.9% (0.5–31.2%) of participants into their corresponding risk groups concerning the clinical model. Including traditional cardiovascular risk factors yielded no significant

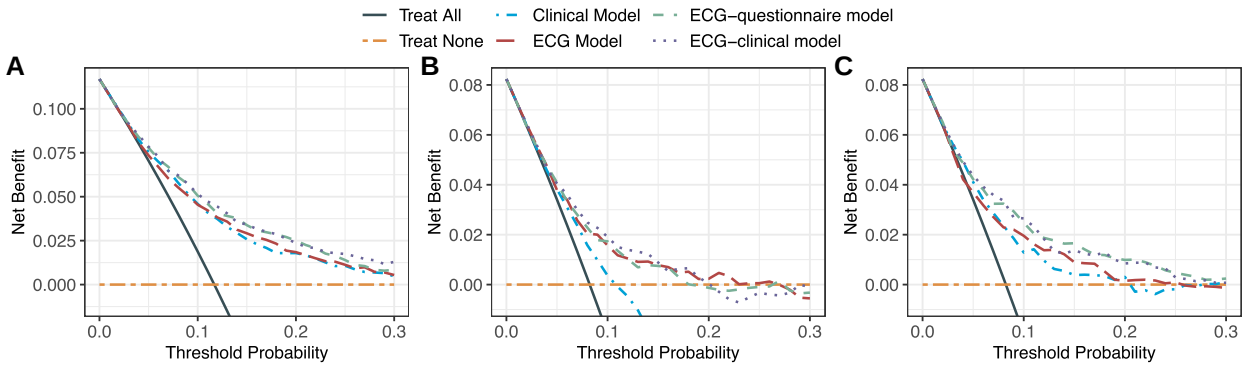

Figure 3 Decision curves to assess the clinical utility (net benefit) of each model for predicting CVD in the derivation and external validation cohorts. (*A*) Derivation cohort; (*B*) External validation cohort; (*C*) External validation cohort after recalibration. CVD, cardiovascular disease; ECG, electrocardiography.

improvement in predicting ASCVD risk (see Supplementary material online, *Table S3*). In terms of clinical utility, the ECG–questionnaire model exhibited great net benefit for ASCVD, similar to the ECG–clinical model, while it outperformed the clinical and ECG models (see Supplementary material online, *Figure S2*).

Predictors importance

Supplementary material online, *Table S4* presents the 20 leading predictors in the ECG–questionnaire model for predicting the risks of developing CVD, CHD, and stroke events. Age was the most important predictor of CVD, followed by P-wave duration in lead I, S-wave amplitude in lead V4, R-wave duration in lead V5, and ST3 amplitude in lead V2. Moreover, S-wave amplitude in lead V4, P-wave amplitude in lead II, R-wave duration in lead V5, Q-wave area in lead aVL, and R-wave amplitude in lead aVL were the leading predictors of CHD. The leading predictors of stroke were age, T-wave amplitude in lead V1, T'-wave amplitude in lead II, STj amplitude in lead V1, and ST3 amplitude in lead V2. Of these, age, smoking, ST3 amplitude in lead V2, R-wave amplitude in lead V4, and R-wave area in lead II were the common leading predictors of all cardiovascular events. Moreover, disease-specific predictors were noted. For example, P-wave duration in lead I was the leading predictor of CVD but not CHD, and S-wave amplitude in lead V4 was the leading predictor of CVD and CHD, but not stroke events.

Sensitivity analysis

Supplementary material online, *Figure S3* presents the performance comparison of the top ECG predictors with different sizes for predicting CVD risk in the derivation and external validation cohorts. We found that the performance when using fewer ECG features was inferior to the final model incorporating 37 features. Specifically, the *C*-statistic of the ECG model for CVD in the derivation cohort decreased from 0.712 for 37 predictors to 0.660 for 10 ($P < 0.001$), and 0.646 for 5 ($P < 0.001$). The use of fewer ECG predictors led to poorer risk stratification, with the top 10 features exhibiting a categorical NRI of −0.098 (95% CI: −0.201–0.004) in the external validation cohort. Similar findings were noted also when the top ECG predictors were used for ASCVD (see Supplementary material online, *Figure S3*).

Future application

To illustrate the process of incorporating our ECG-based models into clinical practice, we have provided a demonstration of the ECG machine used for automatic cardiovascular risk assessment during ECG examination by embedding the questionnaire and our built-in algorithm (see Supplementary material online, *Figure S4* and Supplementary material online, Appendix).

Discussion

Principal findings

In this study, we developed three novel ECG-based risk prediction models for CVD in Chinese adults, including the ECG model (using 37 selected features), ECG–questionnaire model (additionally incorporating questionnaire-based variables, i.e. age, sex, smoking status, and medical history of hypertension, diabetes mellitus, and dyslipidaemia), and ECG–clinical model (additionally incorporating traditional cardiovascular risk factors, i.e. age, sex, smoking status, SBP, LDL-C, HDL-C, diabetes mellitus, and use of anti-hypertensive medications), and further validated their good-to-excellent performance in another independent community-based cohort. Our results demonstrated that the ECG model could accurately predict 7-year CVD risk and was comparable to the clinical model. More importantly, by incorporating the questionnaire-based variables, the ECG–questionnaire model outperformed the clinical and ECG models, yielding competitive performance with the ECG–clinical model. These findings support the use of ECG in CVD risk prediction, particularly when questionnaire-based variables are available.

Comparison with other ECG-based risk equations

There has been much debate regarding the role of ECG in CVD risk prediction beyond traditional cardiovascular risk factors,[26] primarily due to the small to moderate or negligible increase in discrimination (improvement in *C*-statistics or area under the receiver operating characteristic curves ranging from 0.001 to 0.05[18,19,23,47–52]) in previous studies focusing on single or multiple ECG features. Notably, Shah *et al.*[23] and others[30–32] have shown that ECG might have comparable predictive value for CVDs (*C*-statistics ranging from 0.70 to 0.80) independent of traditional risk factors, particularly when using full ECG features. In line with these findings, we showed that the selected ECG features could accurately predict future cardiovascular risk and exhibit comparable predictive capability to the existing recommended models, particularly when incorporating questionnaire-based predictors. Moreover, unlike existing ECG-based CVD risk models that included

information from physical examinations, laboratory tests, and imaging beyond hundreds of ECG features,[30–32] our developed ECG and ECG–questionnaire models only use questionnaire predictors that are easy to obtain by routine communication during clinical practice.

The ECG and ECG–questionnaire models may have important advancement in clinical practice. Traditional clinical predictor-based risk estimation is often considered time-consuming and is largely dependent on its feasibility in the primary care setting.[53–55] A nationwide survey of primary care physicians in the United States showed that 20% rarely used risk calculators, whereas 42% never used them, with time being cited as the greatest barrier to their use.[55] Physicians' inertia should be considered in model development to select appropriate predictors. The ECG-based models proposed in the present study would circumvent this issue because all predictors can be obtained during ECG measurement, and an estimated CVD risk can be reported automatically using a built-in risk assessment algorithm in the digital ECG machine. This simple tool could alleviate the time burden associated with the use of risk prediction models, facilitating their application in clinical practice and primary care.

Finally, we demonstrated the disease-specific patterns of the ECG predictors. Similar to the ECG interpretation of cardiac and non-cardiac diseases[56] that the S-wave amplitude in lead V4 was closely associated with left ventricular hypertrophy, particularly when the deepest S-wave available,[57] which may improve CVD risk prediction,[58] our findings confirmed its predictive value for CVD, particularly CHD events. Moreover, consistent with previous studies,[59,60] our findings showed that P-wave parameters were particularly useful for assessing cardiovascular risk, such as, ischaemic stroke and cardiovascular risk.

Strengths and limitations

The strengths of this study include the large-scale population-based prospective cohorts, availability of a wide range of candidate risk factors, use of digital ECG data, accurate CVD case ascertainment, and rigorous internal and external validation.

However, there are also several limitations. First, only shared ECG features between both cohorts were used, affecting their representation of digital ECG signals. Notably, the impact on the model's predictive capability is expected to be limited due to the highly correlated temporal–spatial nature of ECG data. Additionally, all features that were selected for inclusion in our ECG-based models were standard (PQRST complex) and can be easily extracted with various ECG machines, directly facilitating their application in routine clinical practice regardless of the primary care district and setting.

Second, although all ECG features can be automatically extracted from the ECG machines by the manufacturers, incorporating our risk assessment algorithms into the ECG machines requires authorization and support from the manufacturer with an additional cost for built-in software. Therefore, the practical utility and the cost-efficacy of this risk prediction method await future investigation.

Third, the conventional Cox model failed to handle the non-linear effects and interactions among the ECG data. Artificial intelligence might be helpful in this regard,[61,62] as shown previously in the context of heart failure[63] and sudden cardiac death.[64] However, artificial intelligence-based risk prediction raises another question about model interpretation,[65] which is critical for clinicians, who need to understand how these models work before entrusting them to augment their practice. In this study, we provided the leading predictors and assessed their disease-specific roles in predicting cardiovascular risk, thus facilitating their use by practitioners in clinical practice. However, further investigations using artificial intelligence-based approaches are needed.

Finally, the maximum follow-up period in the CMCS-Beijing Project (~10 years) differed from that of the Shougang community-based cohort (~7.6 years), leading to a limited predictive horizon of 7 years for CVD instead of either 10-year or lifelong risk prediction.[3,45] Therefore, external validation in a prospective cohort with a longer follow-up period is warranted.

Conclusions

We developed novel ECG-based models for predicting CVD risk and validated their performance independent of traditional cardiovascular risk factors in Chinese adults, both with and without the incorporation of easy-to-obtain questionnaire-based predictors. We conclude that embedding the ECG-based CVD risk prediction model in ECG examination may provide valuable information to assist clinicians in initiating risk-based lifestyle changes and pharmaceutical interventions.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Acknowledgements

We thank the participants and staffs of the CMCS Beijing project and Shougang community-based cohort for their important participation and contribution.

Author contributions

Conceptualization: J.L. and Y.Z. Data curation: P.Z., Yi.H., W.Z., Z.W. Formal analysis: P.Z., Z.Y., Yi.H. Funding acquisition: Q.D., J.L., and Y.Z. Investigation: P.Z., Z.Y., Yi.H. Methodology: P.Z., Z.Y., Yi.H., Yo.H, N.Y., L.H., P.J., Y.Q., J.L. Project administration: P.Z., Z.Y., J.L., and Y.Z. Resources: J.L. and Y.Z. Software: P.Z., Z.Y., and Yi.H. Supervision: J.L. and Y.Z. Validation: P.Z., Z.Y., and Yi.H. Writing – original draft: P.Z., Z.Y., Yi.H., and F.F. Writing – review & editing: P.Z., Z.Y., F.F., J.L., and Y.Z.

Funding

This study was supported by the National Key Research and Development Program of China (grant 2022YFC3602501, 2021YFC2500503, 2021YFC2500600, and 2021YFC2500601), the National Natural Science Foundation of China (grant 82073635, 82103962, and 12226005), the Beijing Municipal Commission of Health (grant 2021-07 and 2023-09), and the Capital Funds for Health Improvement and Research (grant CFH 2024-2-1052). The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Conflict of interest: none declared.

Data availability

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *Lancet* 2020;**396**:1204–1222.
2. Zhao D. Epidemiological features of cardiovascular disease in Asia. *JACC Asia* 2021;**1**:1–13.
3. Chinese Society of Cardiology of Chinese Medical Association, Cardiovascular Disease Prevention and Rehabilitation Committee of Chinese Association of Rehabilitation Medicine, Cardiovascular Disease Committee of Chinese Association of Gerontology and Geriatrics, Thrombosis Prevention and Treatment Committee of Chinese Medical Doctor Association. Chinese guideline on the primary prevention of cardiovascular diseases. *Cardiol Discov* 2021;**1**:70–104.
4. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, *et al.* 2021 ESC guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J* 2021;**42**:3227–3337.

5. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, *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. *Circulation* 2019;**140**:e596–e646.
6. Joint Committee on the Chinese Guidelines for Lipid Management. Chinese guidelines for lipid management (2023). *Zhonghua Xin Xue Guan Bing Za Zhi* 2023;**51**:221–255.
7. Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, *et al.* 2019 ESC/EAS guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur Heart J* 2020;**41**:111–188.
8. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, *et al.* 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. *Circulation* 2019;**139**:e1082–e1143.
9. Wang J. Chinese guidelines for the prevention and treatment of hypertension (2024 revision). *J Geriatr Cardiol* 2025;**22**:1–149.
10. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, *et al.* 2018 ESC/ESH guidelines for the management of arterial hypertension. *Eur Heart J* 2018;**39**:3021–3104.
11. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, *et al.* 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. *J Am Soc Hypertens* 2018;**12**:579.e1–579.e73.
12. Hancock EW, Deal BJ, Mirvis DM, Okin P, Kligfield P, Gettes LS, *et al.* AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part V: electrocardiogram changes associated with cardiac chamber hypertrophy: a scientific statement from the American Heart Association electrocardiography and arrhythmias committee, council on clinical cardiology; the American College of Cardiology foundation; and the heart rhythm society: endorsed by the international society for computerized electrocardiology. *Circulation* 2009;**119**:e251–e261.
13. Wagner GS, Macfarlane P, Wellens H, Josephson M, Gorgels A, Mirvis DM, *et al.* AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part VI: acute ischemia/infarction: a scientific statement from the American Heart Association electrocardiography and arrhythmias committee, council on clinical cardiology; the American College of Cardiology foundation; and the heart rhythm society: endorsed by the international society for computerized electrocardiology. *Circulation* 2009;**119**:e262–e270.
14. De Bacquer D, De Backer G, Kornitzer M, Blackburn H. Prognostic value of ECG findings for total, cardiovascular disease, and coronary heart disease death in men and women. *Heart* 1998;**80**:570–577.
15. Ström Möller C, Zethelius B, Sundström J, Lind L. Persistent ischaemic ECG abnormalities on repeated ECG examination have important prognostic value for cardiovascular disease beyond established risk factors: a population-based study in middle-aged men with up to 32 years of follow-up. *Heart* 2007;**93**:1104–1110.
16. Yang Y, Li W, Zhu H, Pan X-F, Hu Y, Arnott C, *et al.* Prognosis of unrecognised myocardial infarction determined by electrocardiography or cardiac magnetic resonance imaging: systematic review and meta-analysis. *BMJ* 2020;**369**:m1184.
17. Maheshwari A, Norby FL, Roetker NS, Soliman EZ, Koene RJ, Rooney MR, *et al.* Refining prediction of atrial fibrillation-related stroke using the P(2)-CHA(2)DS(2)-VASc score. *Circulation* 2019;**139**:180–191.
18. Jørgensen PG, Jensen JS, Marott JL, Jensen GB, Appleyard M, Mogelvang R. Electrocardiographic changes improve risk prediction in asymptomatic persons age 65 years or above without cardiovascular disease. *J Am Coll Cardiol* 2014;**64**:898–906.
19. Auer R, Bauer DC, Marques-Vidal P, Butler J, Min LJ, Cornuz J. Association of major and minor ECG abnormalities with coronary heart disease events. *JAMA* 2012;**307**:1497–1505.
20. Shin J, Lee Y, Park JK, Shin J-H, Lim Y-H, Ran H, *et al.* Prognostic value of myocardial injury-related findings on resting electrocardiography for cardiovascular risk in the asymptomatic general population: the 12-year follow-up report from the Ansan-Ansung cohort. *Ann Med* 2020;**52**:215–224.
21. Okwuosa TM, Soliman EZ, Lopez F, Williams KA, Alonso A, Ferdinand KC. Left ventricular hypertrophy and cardiovascular disease risk prediction and reclassification in blacks and whites: the atherosclerosis risk in communities study. *Am Heart J* 2015;**169**:155–161.e5.
22. Graversen P, Abildstrøm SZ, Jespersen L, Borglykke A, Prescott E. Cardiovascular risk prediction: can systematic coronary risk evaluation (SCORE) be improved by adding simple risk markers? Results from the Copenhagen city heart study. *Eur J Prev Cardiol* 2016;**23**:1546–1556.
23. Shah AJ, Vaccarino V, Janssens AC, Flanders WD, Kundu S, Veledar E, *et al.* An electrocardiogram-based risk equation for incident cardiovascular disease from the national health and nutrition examination survey. *JAMA Cardiol* 2016;**1**:779–786.
24. Segar MW, Vaduganathan M, Patel KV, McGuire DK, Butler J, Fonarow GC, *et al.* Machine learning to predict the risk of incident heart failure hospitalization among patients with diabetes: the WATCH-DM risk score. *Diabetes Care* 2019;**42**:2298–2306.
25. Segar MW, Jaeger BC, Patel KV, Nambi V, Ndumele CE, Correa A, *et al.* Development and validation of machine learning-based race-specific models to predict 10-year risk of heart failure: a multicohort analysis. *Circulation* 2021;**143**:2370–2383.
26. U. S. Preventive Services Task Force; Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, *et al.* Screening for cardiovascular disease risk with electrocardiography: US preventive services task force recommendation statement. *JAMA* 2018;**319**:2308–2314.
27. Goecks J, Jalili V, Heiser LM, Gray JW. How machine learning will transform biomedicine. *Cell* 2020;**181**:92–101.
28. Gorodeski EZ, Ishwaran H, Kogalur UB, Blackstone EH, Hsich E, Zhang Z-M, *et al.* Use of hundreds of electrocardiographic biomarkers for prediction of mortality in postmenopausal women: the women's health initiative. *Circ Cardiovasc Qual Outcomes* 2011;**4**:521–532.
29. Herman R, Vanderheyden M, Vavrik B, Beles M, Palus T, Nelis O, *et al.* Utilizing longitudinal data in assessing all-cause mortality in patients hospitalized with heart failure. *ESC Heart Fail* 2022;**9**:3575–3584.
30. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, *et al.* Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circ Res* 2017;**121**:1092–1101.
31. Siva Kumar S, Al-Kindi S, Tashtish N, Rajagopalan V, Fu P, Rajagopalan S, *et al.* Machine learning derived ECG risk score improves cardiovascular risk assessment in conjunction with coronary artery calcium scoring. *Front Cardiovasc Med* 2022;**9**:976769.
32. Fan Z, Du Z, Fu J, Zhou Y, Zhang P, Shi C, *et al.* Comparing the performance of machine learning and conventional models for predicting atherosclerotic cardiovascular disease in a general Chinese population. *BMC Med Inform Decis Mak* 2023;**23**:134.
33. Liu J, Hong Y, D'Agostino RB Sr, Wu Z, Wang W, Sun J, *et al.* Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese multi-provincial cohort study. *JAMA* 2004;**291**:2591–2599.
34. Qi Y, Han X, Zhao D, Wang W, Wang M, Sun J, *et al.* Long-term cardiovascular risk associated with stage 1 hypertension defined by the 2017 ACC/AHA hypertension guideline. *J Am Coll Cardiol* 2018;**72**:1201–1210.
35. Wu Z, Yao C, Zhao D, Wu G, Wang W, Liu J, *et al.* Sino-MONICA project: a collaborative study on trends and determinants in cardiovascular diseases in China, part I: morbidity and mortality monitoring. *Circulation* 2001;**103**:462–468.
36. Fan F, Qi L, Jia J, Xu X, Liu Y, Yang Y, *et al.* Noninvasive central systolic blood pressure is more strongly related to kidney function decline than peripheral systolic blood pressure in a Chinese community–based population. *Hypertension* 2016;**67**:1166–1172.
37. Liu S, Zhao D, Wang M, Qi Y, Sun J, Liu J, *et al.* Association of serum potassium levels with mortality and cardiovascular events: findings from the Chinese multi-provincial cohort study. *J Gen Intern Med* 2022;**37**:2446–2453.
38. Liu M, Jia J, Fan F, Sun P, Wu Z, Jiang Y, *et al.* Association of depressive symptoms with the progression of carotid intima-media thickness in a community-based cohort in Beijing, China. *BMJ Open* 2021;**11**:e050254.
39. Al-Zaiti S, Besomi L, Bouzid Z, Faramand Z, Frisch S, Martin-Gill C, *et al.* Machine learning-based prediction of acute coronary syndrome using only the pre-hospital 12-lead electrocardiogram. *Nat Commun* 2020;**11**:3966.
40. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. *Fam Med Community Health* 2020;**8**:e000262.
41. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**15**:361–387.
42. Demler OV, Paynter NP, Cook NR. Tests of calibration and goodness-of-fit in the survival setting. *Stat Med* 2015;**34**:1659–1680.
43. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011;**30**:11–21.
44. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;**27**:157–172; discussion 207–112.
45. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;**26**:565–574.
46. Moons KG, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;**162**:W1–W73.
47. Badheka AO, Patel N, Tuliani TA, Rathod A, Marzouka GR, Zalawadiya S, *et al.* Electrocardiographic abnormalities and reclassification of cardiovascular risk: insights from NHANES-III. *Am J Med* 2013;**126**:319–326.e2.
48. Badheka AO, Patel NJ, Grover PM, Shah N, Singh V, Deshmukh A, *et al.* ST-T wave abnormality in lead aVR and reclassification of cardiovascular risk (from the national health and nutrition examination survey-III). *Am J Cardiol* 2013;**112**:805–810.
49. Denes P, Larson JC, Lloyd-Jones DM, Prineas RJ, Greenland P. Major and minor ECG abnormalities in asymptomatic women and risk of cardiovascular events and mortality. *JAMA* 2007;**297**:978–985.

50. 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* 2003;**26**:2777–2784.
51. Ishikawa J, Ishikawa S, Kario K. Prolonged corrected QT interval is predictive of future stroke events even in subjects without ECG-diagnosed left ventricular hypertrophy. *Hypertension* 2015;**65**:554–560.
52. Tereshchenko LG, Henrikson CA, Sotoodehnia N, Arking DE, Agarwal SK, Siscovick DS, *et al.* Electrocardiographic deep terminal negativity of the P wave in V(1) and risk of sudden cardiac death: the atherosclerosis risk in communities (ARIC) study. *J Am Heart Assoc* 2014;**3**:e001387.
53. Shillinglaw B, Viera AJ, Edwards T, Simpson R, Sheridan SL. Use of global coronary heart disease risk assessment in practice: a cross-sectional survey of a sample of U.S. physicians. *BMC Health Serv Res* 2012;**12**:20.
54. Bakhai S, Bhardwaj A, Sandhu P, Reynolds JL. Optimisation of lipids for prevention of cardiovascular disease in a primary care. *BMJ Open Qual* 2018;**7**:e000071.
55. Eaton CB, Galliher JM, McBride PE, Bonham AJ, Kappus JA, Hickner J. Family physician's knowledge, beliefs, and self-reported practice patterns regarding hyperlipidemia: a national research network (NRN) survey. *J Am Board Fam Med* 2006;**19**:46–53.
56. de Jong J. ECGpedia. https://www.ecgpedia.org/A4/ECGpedia_on_1_A4En.pdf, 2010.
57. Peguero JG, Lo Presti S, Perez J, Issa O, Brenes JC, Tolentino A. Electrocardiographic criteria for the diagnosis of left ventricular hypertrophy. *J Am Coll Cardiol* 2017;**69**:1694–1703.
58. Iribarren C, Round AD, Lu M, Okin PM, McNulty EJ. Cohort study of ECG left ventricular hypertrophy trajectories: ethnic disparities, associations with cardiovascular outcomes, and clinical utility. *J Am Heart Assoc* 2017;**6**:e004954.
59. He J, Tse G, Korantzopoulos P, Letsas KP, Ali-Hasan-Al-Saegh S, Kamel H, *et al.* P-Wave indices and risk of ischemic stroke: a systematic review and meta-analysis. *Stroke* 2017; **48**:2066–2072.
60. Chen LY, Ribeiro ALP, Platonov PG, Cygankiewicz I, Soliman EZ, Gorenek B, *et al.* P wave parameters and indices: a critical appraisal of clinical utility, challenges, and future research—a consensus document endorsed by the international society of electrocardiology and the international society for holter and noninvasive electrocardiology. *Circ Arrhythm Electrophysiol* 2022;**15**:e010435.
61. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol* 2021;**18**: 465–478.
62. Attia ZI, Harmon DM, Behr ER, Friedman PA. Application of artificial intelligence to the electrocardiogram. *Eur Heart J* 2021;**42**:4717–4730.
63. Akbilgic O, Butler L, Karabayir I, Chang PP, Kitzman DW, Alonso A, *et al.* ECG-AI: electrocardiographic artificial intelligence model for prediction of heart failure. *Eur Heart J Digit Health* 2021;**2**:626–634.
64. Shiraishi Y, Goto S, Niimi N, Katsumata Y, Goda A, Takei M, *et al.* Improved prediction of sudden cardiac death in patients with heart failure through digital processing of electrocardiography. *Europace* 2023;**25**:922–930.
65. Somani S, Russak AJ, Richter F, Zhao S, Vaid A, Chaudhry F, *et al.* Deep learning and the electrocardiogram: review of the current state-of-the-art. *Europace* 2021;**23**: 1179–1191.