Development and Validation of a Risk Prediction Model for Atherosclerotic Cardiovascular Disease in Japanese Adults: The Hisayama Study

Takanori Honda[1], Sanmei Chen[1], Jun Hata[1, 2, 3], Daigo Yoshida[1], Yoichiro Hirakawa[1, 3], Yoshihiko Furuta[1, 3, 4], Mao Shibata[1, 2], Satoko Sakata[1, 2, 3], Takanari Kitazono[2, 3] and Toshiharu Ninomiya[1, 2]

Takanori Honda and Sanmei Chen contributed equally to this work.

[1] Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
[2] Center for Cohort Studies, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
[3] Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
[4] Department of Medical-Engineering Collaboration for Healthy Longevity, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

Aim: To develop and validate a new risk prediction model for predicting the 10-year risk of atherosclerotic cardiovascular disease (ASCVD) in Japanese adults.

Methods: A total of 2,454 participants aged 40–84 years without a history of cardiovascular disease (CVD) were prospectively followed up for 24 years. An incident ASCVD event was defined as the first occurrence of coronary heart disease or atherothrombotic brain infarction. A Cox proportional hazards regression model was used to construct the prediction model. In addition, a simplified scoring system was translated from the developed prediction model. The model performance was evaluated using Harrell's C statistics, a calibration plot with the Greenwood-Nam-D'Agostino test, and a bootstrap validation procedure.

Results: During a median of a 24-year follow-up, 270 participants experienced the first ASCVD event. The predictors of the ASCVD events in the multivariable Cox model included age, sex, systolic blood pressure, diabetes, serum high-density lipoprotein cholesterol, serum low-density lipoprotein cholesterol, proteinuria, smoking habits, and regular exercise. The developed models exhibited good discrimination with negligible evidence of overfitting (Harrell's C statistics: 0.786 for the multivariable model and 0.789 for the simplified score) and good calibrations (the Greenwood-Nam-D'Agostino test: *P*=0.29 for the multivariable model, 0.52 for the simplified score).

Conclusion: We constructed a risk prediction model for the development of ASCVD in Japanese adults. This prediction model exhibits great potential as a tool for predicting the risk of ASCVD in clinical practice by enabling the identification of specific risk factors for ASCVD in individual patients.

See editorial vol. 29: 320-321

Introduction

Atherosclerotic cardiovascular diseases (ASCVDs) are diseases caused by the formation of cholesterol plaque in the arterial walls and its disruption by thrombosis, such as coronary heart disease (CHD) and stroke events, which are attributed to atherothrombosis. However, there is no consensus on the definition[1]. In preventive medicine, it is reasonable to consider ASCVD events as a single clinical entity[1] as patients with ASCVD often have the same risk factors and have concomitant systematic

atherosclerosis in different arterial beds and overlapping mechanisms influencing their prognosis[2-4]. To date, numerous risk prediction models have been developed worldwide[5], especially in Japan[6-11], to guide the prevention of cardiovascular disease (CVD). However, the data on risk prediction models that consider ASCVD events as a single clinical entity are scarce[12, 13]. Such specific prediction models are clinically significant as predictive risk factors—and the magnitude of effects of the same risk factors—may vary depending on the etiologic subtypes of stroke[14]. For instance, the findings from the Hisayama Study have demonstrated that standard CVD risk factors, such as serum non-high-density lipoprotein (non-HDL) cholesterol and low-density lipoprotein (LDL) cholesterol levels, are associated with the risk of atherothrombotic brain infarction and CHD but not with other stroke subtypes, including lacunar infarction, cardioembolic infarction, and hemorrhagic stroke[15, 16]. Thus, the development of a risk prediction tool that comprises risk factors relevant to atherothrombosis and integrates algorithms (i.e., scoring methods) reflecting the magnitude of their effects on ASCVD is important to better predict the risk of this clinical entity.

Different algorithms have been developed to predict the risk of ASCVD or death due to ASCVD[13, 17, 18]. However, none have distinguished the subtype of atherothrombotic brain infarction from total ischemic stroke in order to accurately define the outcome of interest, probably due to the lack of detailed data on the diagnosis of ischemic stroke subtypes. The 2013 American College of Cardiology and the American Heart Association Cardiovascular Risk Assessment Guideline developed a risk prediction tool for ASCVD[13]. In the construction of this tool, however, all CHD and total ischemic stroke (even lacunar and cardioembolic infarction) events were combined as the outcome of interest[13]. The model developed in the Prediction for ASCVD Risk in China Project has the same limitation[17]. In Japan, prediction models have been developed for individual ischemic CVD components, i.e., CHD or total ischemic stroke alone[19]. Thus far, to the best of our knowledge, no algorithms have combined atherothrombotic brain infarction with CHD as the outcome of interest to predict the summated risk of ASCVD events.

Aim

In this study, using reliable diagnoses of ischemic stroke subtypes from a cohort of general Japanese adults with a sufficiently long follow-up of 24 years, we developed a new tool to estimate the 10-year summated risk of ASCVD events comprising CHD and atherothrombotic brain infarction.

Methods

Participants

The Hisayama Study, which started in 1961, is an ongoing cohort study of cerebro- and cardiovascular diseases among men and women aged ≥ 40 years residing in Hisayama[20]. Hisayama is a suburb of Fukuoka City in southern Japan. The town's population was approximately 7500 in 1988, and full community health surveys have been conducted every 1–2 years since 1961. In 1988, a total of 2,742 Hisayama residents aged ≥ 40 years consented to participate in the health survey (participation rate: 80.9%). Of these participants, 2,589 who were aged ＜ 85 years and free of any CVD (CHD or any ischemic or hemorrhagic stroke) at baseline were included in this study. We excluded 1 participant who died before the start of the follow-up, 89 participants who had not fasted overnight before blood sample collection, 43 participants who had no estimates of serum LDL cholesterol, and 2 participants who had missing information on exercise. The final sample included 2,454 participants (1,026 men and 1,428 women) with a mean (±SD) age of 58 (±11) years **(Supplemental Fig.1)**. The study protocol was approved by the Kyushu University Institutional Review Board for Clinical Research. All participants provided informed consent.

Follow-Up Survey

The participants were prospectively followed up from December 1988 to November 2012 through annual health examinations and a daily surveillance system[21, 22]. The daily surveillance system comprised the study team, local physicians, and members of the town's Health and Welfare Office. The physicians of the study team regularly visited hospitals, clinics, and the town's office to update information on the incidence of endpoints. For participants who did not undergo the annual health survey or those who moved away from the town, their health conditions were checked annually *via* postal or telephone surveys. When a participant died, an autopsy was performed at the Department of Pathology of Kyushu University, if consent for autopsy was obtained. During the follow-up, a total of 969 participants died, and 664 of the decedents underwent autopsy; no participants were lost to follow-up.

Ascertainment of ASCVD Events

The incidence of an ASCVD event was defined

as the first occurrence of CHD and atherothrombotic brain infarction. Previously, the procedures for the diagnosis of CVD subtypes have been reported in detail[21, 22]. The criteria for the diagnosis of CHD events included acute myocardial infarction, silent myocardial infarction, sudden cardiac death within 1 h after the onset of acute illness, percutaneous transluminal coronary angioplasty, or coronary artery bypass graft surgery. Acute myocardial infarction was diagnosed when a participant fulfilled at least two of the following criteria: 1) typical symptoms of prolonged severe anterior chest pain, 2) abnormal cardiac enzymes over twice the upper limit of the normal range, 3) evolving diagnostic changes of electrocardiogram (ECG), or 4) morphological changes (local asynergy of the cardiac wall motion on echocardiography, persistent perfusion defect on cardiac scintigraphy, or myocardial necrosis or scars > 1 cm, accompanied by coronary atherosclerosis at autopsy). Silent myocardial infarction was diagnosed for participants who had myocardial scarring but no historical indication of clinical symptoms or abnormal cardiac enzyme changes, as detected by ECG, echocardiography, and cardiac scintigraphy or at autopsy. Stroke was defined as the sudden onset of nonconvulsive and focal neurological deficit persisting for >24 h. The diagnosis and classification of stroke were adjudicated by reviewing the relevant clinical medical records of the patients, including brain imaging, cerebral angiography, echocardiography, carotid duplex imaging, or autopsy findings. The diagnoses of ischemic stroke subtypes (atherothrombotic infarction, lacunar infarction, cardioembolic infarction, and undetermined subtype) were made in accordance with the guidelines of the Classification of Cerebrovascular Disease III proposed by the National Institute of Neurological Disorders and Stroke[23]. The diagnoses were also made based on the diagnostic criteria of the Trial of Org 10172 in Acute Stroke Treatment Study[24] and the Cerebral Embolism Task Force[25], according to our previously published diagnostic procedure[21].

Potential Risk Factors

For the present study, we evaluated the following potential risk factors that were previously found to be associated with ASCVD risk in Japanese adults: age, sex, systolic blood pressure, use of antihypertensive medication, diabetes, serum HDL cholesterol, serum LDL cholesterol, serum non-HDL cholesterol, body mass index (BMI), ECG abnormality, proteinuria, smoking habits, alcohol drinking, and regular exercise. After at least 5 min of rest, sitting blood pressure was measured three times at the right upper arm using a mercury sphygmomanometer; the average of the three measurements was used for the analysis. Fasting plasma glucose levels were measured using the glucose oxidase method. Diabetes was defined as presenting either fasting plasma glucose ≥ 7.0 mmol/L or casual or 2-h 75 g oral glucose post-load glucose ≥ 11.1 mmol/L, or taking antidiabetic medication. The levels of serum total cholesterol, HDL cholesterol, and triglyceride were determined enzymatically. The serum LDL cholesterol levels were estimated using the Friedewald formula. Conversely, the serum non-HDL cholesterol levels were calculated by subtracting the serum HDL cholesterol level from the serum total cholesterol level. Height and weight were measured in light clothes and without shoes; BMI was also calculated (kg/m^2). Obesity was defined as a BMI of ≥ 25.0 kg/m^2. ECG abnormalities were defined by left ventricular hypertrophy (3-1) or ST depression (4-1, 2, 3). Proteinuria was measured using the test paper method and defined as 1+ or more. Information on the following risk factors was obtained *via* a standard questionnaire administered by trained interviewers: the use of antihypertensive medication, smoking status, alcohol drinking status, and regular exercise. Smoking and drinking habits were classified as currently smoking or drinking or not. Current smokers were those who were currently smoking at least one cigarette per day regularly. Current drinkers were those who were currently drinking at least one alcoholic beverage per month regularly. Regular exercise was defined as engaging in sports or other forms of athletic activity at least three times per week in leisure time[14].

Statistical Analyses

The details of the statistical method are presented in the Supplemental Methods. Briefly, we constructed a multivariable prediction model to predict the risk of ASCVD based on a Cox's proportional hazards regression model. The backward elimination at a criterion of *P*-value <0.10 was employed to select the risk factors for the final risk prediction model. Given the long follow-up period and the potential bias due to the competing risk of death in the prediction of long-term risk[26, 27], we modified the baseline survival function of the prediction model by accounting for the competing risk of deaths without ASCVD events during follow-up. Then, we calculated the 10-year survival probability of developing ASCVD using the following formula:

$$\hat{p} = 1 - S_0(t=10)_{CR}^{\exp(\Sigma \beta_i x_i - \Sigma \beta_i \bar{x}_i)},$$

where $S_0(t=10)_{CR}$ denotes the baseline survival function accounting for the competing risk of death, and $\Sigma \beta_i x_i$ and $\Sigma \beta_i \bar{x}_i$ were computed by summing up

the products of the regression coefficients of each predictor (β_i) and individual values (x_i) and the mean values ($\bar{x}_i$), respectively.

We translated the developed multivariable prediction model into a simplified scoring system according to a procedure proposed by Sullivan *et al.*[28] The predicted 10-year probability of developing ASCVD was presented according to the score and age group. The agreement between the 10-year probability of developing ASCVD predicted by the multivariable model and that by the simplified score was evaluated using Spearman's rank correlation and a bivariate linear regression of the model-predicted probability on the score-predicted probability.

The discrimination of the developed model was evaluated by using Harrell's concordance index (C index). The internal validity was evaluated by estimating the model optimism using 200 bootstrap samples. The optimism-corrected C index was calculated according to the procedure proposed by Harrell *et al.*[29]. The calibration was assessed by plotting the model-based predicted 10-year probabilities against the actual observed probabilities over 10 years. We also statistically tested the calibration using the Greenwood-Nam-D'Agostino (GND) test, where deciles with few events were collapsed into the next decile as appropriate[30].

In addition, we developed an alternative model that included serum non-HDL cholesterol instead of serum LDL cholesterol, as the Japan Atherosclerosis Society guideline recommends the assessment of serum non-HDL cholesterol instead of serum LDL cholesterol when using non-fasting samples or samples with serum TG levels ≥ 400 mg/dL[14]. We evaluated the agreement between the 10-year probability of developing ASCVD predicted by the original multivariable model and that predicted by the alternative non-HDL cholesterol model using Spearman's rank correlation coefficient. We also evaluated the agreement using a linear regression of the predicted probability by the final multivariable model on that by the alternative model. Similarly, we evaluated the agreement between the 10-year probability predicted by the original simplified score and that predicted by the alternative non-HDL cholesterol simplified score. All statistical analyses were conducted using SAS version 9.4 (SAS Institute, Cary, NC). A two-sided *P*-value of <0.05 was considered statistically significant.

Results

The baseline characteristics of the study participants are presented in **Table 1**. During a

Table 1. Baseline characteristics of study participants (*n* = 2,454)

	Mean (SD), median (IQR), or percentage
Age, years	58.2 (11.1)
Sex, % men	41.8
Systolic blood pressure, mmHg	132.4 (20.4)
Diastolic blood pressure, mmHg	77.5 (11.2)
Use of antihypertensive drugs, %	14.1
Diabetes, %	11.8
Serum total cholesterol, mg/dL	207.4 (42.0)
Serum HDL cholesterol, mg/dL	50.6 (11.7)
Serum non-HDL cholesterol, mg/dL	156.8 (41.0)
Serum LDL cholesterol, mg/dL	134.0 (39.5)
Serum triglycerides, mg/dL	97 (70-138)
Body mass index, kg/m²	22.9 (3.1)
Obesity, %	23.5
ECG abnormalities, %	15.8
Proteinuria, %	5.7
Current drinker, %	30.4
Current smoker, %	25.0
Regular exercise, %	10.1

Abbreviations: SD, standard deviation; IQR, interquartile range; HDL cholesterol, high-density lipoprotein cholesterol; LDL cholesterol, low-density lipoprotein cholesterol; ECG, electrocardiogram.

median of 24 years (interquartile range: 14–24 years) of follow-up, 270 participants developed the first ASCVD event (crude incidence rate: 5.8 per 1000 person-years). This event consisted of 216 incident cases of CHD (crude incidence rate: 4.6 per 1000 person-years) and 62 incident cases of atherothrombotic brain infarction (crude incidence rate: 1.3 per 1000 person-years). Eight participants developed both CHD and atherothrombotic brain infarction and were thus treated as events in the analysis for each outcome. The event numbers of individual CHD subtypes were as follows: 95 cases of acute myocardial infarction, 58 cases of silent myocardial infarction, 20 cases of sudden cardiac death, 34 cases of percutaneous transluminal coronary angioplasty, and 9 cases of coronary artery bypass graft surgery.

Table 2 presents the hazard ratios and beta regression coefficients of each predictor of the 10-year risk of ASCVD in the final multivariable prediction model. In the final prediction model, the following variables were retained: age, sex, systolic blood pressure, diabetes, serum HDL cholesterol, serum LDL cholesterol, proteinuria, smoking habits, and regular exercise. The modified survival function that accounted for the competing risk of death at 10 years was as follows: $S_0(t=10)_{CR} = 0.9696$ (see **Supplemental**

Table 2. Multivariable model for predicting the risk of atherosclerotic cardiovascular disease ($n=2,454$)*

	HR (95% CI)	β	P-value
Age (per year)	1.08 (1.07-1.10)	0.077	<0.001
Men (vs women)	2.68 (2.01-3.57)	0.984	<0.001
Systolic blood pressure (per 1 mmHg)	1.01 (1.00-1.02)	0.010	<0.001
Diabetes (vs nondiabetic)	1.58 (1.17-2.14)	0.459	0.003
Serum HDL cholesterol (per 1 mg/dL)	0.99 (0.98-1.00)	-0.012	0.03
Serum LDL cholesterol (per 1 mg/dL)	1.01 (1.00-1.01)	0.005	0.002
Proteinuria (vs absent)	1.88 (1.27-2.79)	0.632	0.002
Current smoker (vs non-smoker)	1.40 (1.05-1.87)	0.336	0.02
No or irregular exercise (vs regular)	1.40 (0.94-2.09)	0.339	0.10
C statistics (95% CI)	0.786 (0.758-0.813)		
Optimism-corrected C statistics	0.776		

Abbreviations: HR, hazard ratio; 95% CI, 95% confidence interval; HDL cholesterol, high-density lipoprotein cholesterol; LDL cholesterol, low-density lipoprotein cholesterol.
*The risk factors selected in the final prediction model are presented. Cox proportional hazards regression models with a backward selection method were used to select predictors (P-value <0.1). Optimism-corrected C statistics were calculated based on 200 bootstrapping samples.

Table 1 for the cumulative incidence functions of ASCVD incidence accounting for the competing risk of death from year 0 to year 10). Thus, the final multivariable model for predicting the 10-year probability of developing the first ASCVD event was as follows:

$\hat{p} = 1 - 0.9696^{\exp(\Sigma \beta_i x_i - 6.7963)}$,

where $\Sigma \beta_i x_i = (0.077 \times \text{age in years})$
+ (0.984 if men)
+ (0.010 × systolic blood pressure in mmHg)
+ (0.459 if diabetic)
+ (–0.012 × serum HDL cholesterol levels in mg/dL)
+ (0.005 × serum LDL cholesterol levels in mg/dL)
+ (0.632 if presenting with proteinuria)
+ (0.336 if a current smoker)
+ (0.339 if lacking a regular exercise habit)

All the predictors included in the final multivariable model were recurrently selected in > 40% of the repeated backward selections in the 200 bootstrapping samples **(Supplemental Fig. 2)**. These predictors selected in the final multivariable prediction model demonstrated similar effect sizes when separately fitted for the risk of CHD and atherothrombotic brain infarction **(Supplemental Table 2)**.

The final multivariable prediction model was translated into a simplified risk score. The scoring method is presented in **Fig. 1**, and the determination of points is presented in **Supplemental Table 3**. The total score, calculated by the predictors except for age,

ranged from 0 to 27 points. The formula of 10-year probability of developing ASCVD was as follows:

$\hat{p} = 1 - 0.9696^{\exp([\text{total score} + \text{points for age}] \times 0.144 - 2.4767)}$,

where the points for age were 0 for 40–49 years, 5 for 50–59 years, 11 for 60–69 years, 16 for 70–79 years, and 20 for ≥ 80 years, respectively.

The distribution of 10-year probabilities of developing ASCVD predicted by the simplified risk score is presented in **Supplemental Fig. 3**. The 10-year probability of developing ASCVD predicted by the multivariable model and that by the simplified score were highly linearly correlated (Spearman's correlation coefficient r=0.974; intercept=0.000; and regression coefficient β =1.054 [95% CI 1.041–1.067] in a bivariate linear regression model that regressed the model-predicted probability on the score-predicted probability).

The Harrell's C statistics of the final multivariable prediction model indicated good discrimination (0.786, 95% CI: 0.758–0.813) in the original cohort. In the 200 bootstrapping samples, the multivariable model exhibited good internal validity (optimism-corrected C statistics 0.776). The simplified risk score also offered good discrimination (Harrell's C statistics of 0.789, 95% CI: 0.762–0.817). The calibration plot indicates that the predicted and observed percentages of ASCVD events were highly linearly correlated both in the multivariable model **(Fig. 2A)** and the simplified scoring system **(Fig. 2B)**, with the GND tests indicating good calibration (P=0.29 for the multivariable model, 0.52 for the simplified score).

The regression coefficients and HRs (95%CIs) for the alternative non-HDL cholesterol model are

Add up points (left panels), then look up the predicted 10-year ASCVD risk (right panel)

1. Sex	Points
Women	0
Men	7

2. Systolic blood pressure	Points
<120 mmHg	0
120-129 mmHg	1
130-139 mmHg	2
140-159 mmHg	3
160 mmHg -	4

3. Diabetes	Points
No	0
Yes	3

4. Serum HDL cholesterol	Points
60 mg/dL -	0
40-59 mg/dL	1
<40 mg/dL	2

5. Serum LDL cholesterol *	Points
<120 mg/dL	0
120-139 mg/dL	1
140-159 mg/dL	2
160 mg/dL -	3

6. Proteinuria	Points
No	0
Yes	4

7. Current smoker	Points
No	0
Yes	2

8. Regular exercise	Points
No	2
Yes	0

Total score (sum of #1 to #8) _____

Total score	40-49 years	50-59 years	60-69 years	70-79 years	80-84 years
0	<1.0%	<1.0%	1.4%	3.1%	5.6%
1	<1.0%	<1.0%	1.7%	3.6%	6.5%
2	<1.0%	<1.0%	1.9%	4.1%	7.6%
3	<1.0%	<1.0%	2.2%	4.8%	8.8%
4	<1.0%	1.0%	2.6%	5.6%	10.2%
5	<1.0%	1.2%	3.1%	6.5%	11.7%
6	<1.0%	1.4%	3.6%	7.6%	13.6%
7	<1.0%	1.7%	4.1%	8.8%	15.7%
8	<1.0%	1.9%	4.8%	10.2%	18.0%
9	1.0%	2.2%	5.6%	11.7%	20.7%
10	1.2%	2.6%	6.5%	13.6%	23.8%
11	1.4%	3.1%	7.6%	15.7%	27.2%
12	1.7%	3.6%	8.8%	18.0%	30.9%
13	1.9%	4.1%	10.2%	20.7%	35.1%
14	2.2%	4.8%	11.7%	23.8%	39.6%
15	2.6%	5.6%	13.6%	27.2%	44.5%
16	3.1%	6.5%	15.7%	30.9%	49.7%
17	3.6%	7.6%	18.0%	35.1%	>50.0%
18	4.1%	8.8%	20.7%	39.6%	>50.0%
19	4.8%	10.2%	23.8%	44.5%	>50.0%
20	5.6%	11.7%	27.2%	49.7%	>50.0%
21	6.5%	13.6%	30.9%	>50.0%	>50.0%
22	7.6%	15.7%	35.1%	>50.0%	>50.0%
23	8.8%	18.0%	39.6%	>50.0%	>50.0%
24	10.2%	20.7%	44.5%	>50.0%	>50.0%
25	11.7%	23.8%	49.7%	>50.0%	>50.0%
26	13.6%	27.2%	>50.0%	>50.0%	>50.0%
27	15.7%	30.9%	>50.0%	>50.0%	>50.0%

Fig. 1. Simplified point-based scoring system for atherosclerotic cardiovascular disease

The predicted probability was determined using the following formula: $\hat{p} = 1 - 0.9696^{\exp([\text{total score} + \text{points for age}] \times 0.144 - 2.4767)}$, where the points of 0, 5, 11, 16, and 20 for age were assigned to the age ranges of 40–49, 50–59, 60–69, 70–79, and ≥ 80 years, respectively. Probabilities are presented in green (low risk: <2.0% of the 10-year atherosclerotic cardiovascular disease risk, corresponding to the lowest 35% of the distribution in the population), yellow (middle risk: 2.0%–10.0%), and red (high risk: ≥ 10%, corresponding to the highest 20% of the distribution in the population). *In the alternative simplified score that included serum non-HDL cholesterol instead of serum LDL cholesterol, the points for the predefined categories of serum non-HDL cholesterol (<150, 150–169, 170–189, and ≥ 190 mg/dL) were 0, 1, 2, and 3, respectively.
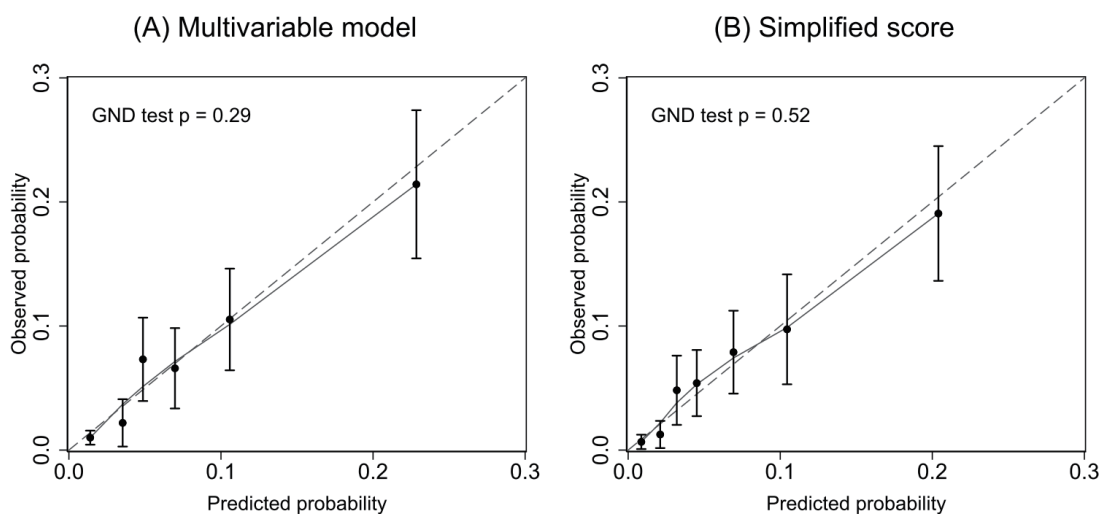
Fig. 2. Calibration plots of the predicted 10-year probability of atherosclerotic cardiovascular disease predicted by the multivariable model (A) and the simplified scoring system (B)

The dotted lines indicate the case of perfect calibration, corresponding to an intercept of zero and a slope of one for the calibration plot. The solid curves indicate the calibration curve fit by the loess smoother. Bars indicate the 95% confidence intervals of the observed probability in each group of the predicted probability. Abbreviation: GND, Greenwood-Nam-D'Agostino.

presented in **Supplemental Table 3**. By using this alternative model, the 10-year probability of developing the first ASCVD event can be calculated as follows:

$$\hat{p} = 1 - 0.9696^{\exp(\Sigma\beta_i x_i - 6.9895)},$$

where $\Sigma\beta_i x_i$ can be obtained from **Supplemental Table 4**.

The alternative prediction model also demonstrated a good satisfactory performance, with a Harrell's C statistics of 0.786 (95% CI: 0.758–0.814) and an optimism-corrected C statistics of 0.777. The predicted probability by the alternative non-HDL model exhibited good agreements with that by the original LDL cholesterol model **(Supplemental Fig. 4)**. After calculating the simplified score by assigning 0, 1, 2, or 3 points to the predefined categories of serum non-HDL cholesterol (<150, 150–169, 170–189, and ≥ 190 mg/dL) instead of the points for the LDL cholesterol levels, the probability of developing ASCVD predicted by the alternative non-HDL cholesterol simplified score exhibited good agreements with that by the LDL cholesterol simplified score **(Supplemental Fig. 5)**.

Discussion

In the present study, we constructed a risk prediction model to estimate the risk of developing ASCVD, which is defined as CHD and/or atherothrombotic brain infarction, in a general Japanese population, based on prospective follow-up data collected over 24 years. Using well-established, modifiable cardiovascular risk factors as predictors, our prediction model demonstrated good discrimination, good calibration, and satisfactory internal validity. In addition, the predictors included in this prediction model exhibited similar effect sizes on the development of each outcome when CHD and atherothrombotic brain infarction were separately analyzed, which further supports the combination of CHD and atherothrombotic brain infarction as a single clinical entity. Our prediction model and simplified risk calculator may be effective in helping individuals become aware of their own risk of developing ASCVD events and in preventing this clinical entity.

Currently, several risk prediction tools have been recommended by prevention guidelines to predict the risk of ischemic CVD, such as the Pooled Cohort Equation of ACC/AHA[13], the CHINA-PAR of China[17], the JBS3 of Joint British Societies[31], and the SCORE chart of the European Society of Cardiology[32]. However, none of these studies have defined ASCVD events by distinguishing the subtype

of atherothrombotic brain infarction from other stroke subtypes. To the best of our knowledge, this is the first study to develop a new prediction model for the summated risk of ASCVD events defined as CHD or atherothrombotic brain infarction using data on the reliable diagnosis of stroke subtypes.

The predictors included in our prediction model were generally concordant with, but slightly different from, those in previous prediction models for overall CVD, CHD, or ischemic stroke alone. For instance, the risk prediction tools developed in the abovementioned studies[13, 17, 31, 32] included serum total cholesterol or serum non-HDL cholesterol as a predictor. However, in the present study, we selected serum LDL cholesterol for the main analysis as LDL cholesterol has been considered as the major risk factor for ASCVD according to the latest guideline in Japan[14]. Moreover, we constructed an alternative model that used serum non-HDL cholesterol levels instead of serum LDL cholesterol levels. Both models demonstrated good performance, indicating their comparable predictive abilities. As for the difference between our prediction model and the existing models, the ECG abnormality did not remain in our ASCVD prediction model, which is inconsistent with the recently published Suita risk prediction model for CHD and stroke[10]. This is probably because we were able to exclude the CVD subtypes that are strongly associated with ECG abnormalities (i.e., lacunar and cardioembolic stroke) from the endpoints. In terms of the predictor diabetes, it is notable that in the present study, we used OGTT for the diagnosis rather than fasting plasma glucose and HbA1c. This facilitated the minimization of the possibility of missed diagnosis of diabetes and consequently the improvement of the accuracy of our algorithm. However, in clinical practice, data on fasting plasma glucose and HbA1c are more accessible than OGTT. Thus, it is important for clinical practitioners to be aware that they are likely to underestimate a patient's predicted ASCVD risk if missed diagnosis of diabetes occurred in the case of using fasting plasma glucose and/or HbA1c.

In addition to the high participation rate and the long follow-up period, the major strength of the present study is the reliable diagnosis of ischemic stroke subtypes. In the Hisayama Study, all CVD events were adjudicated by a panel of study physicians, and the presence of CVD lesions was morphologically confirmed via autopsy in most of the deceased subjects, which minimized the possibility of missed diagnoses and misclassification of diagnosis of ischemic stroke and CHD subtypes. Moreover, the rigorous surveillance for the endpoints along with the perfect follow-up rate significantly improved the

accuracy of estimating the absolute risk of ASCVD. However, this study has limitations. First, we only used a single measurement of the risk factors at baseline without considering their changes during the follow-up period, which could have resulted in misclassification, thus biasing the results toward the null. However, it may be more practical to use information collected at a single time point than the information obtained from repeated measurements for the risk assessment in the primary care and clinical settings. Second, due to the relatively small number of ASCVD cases at the low end of the predicted probability, we were unable to divide the participants into more groups of the predicted probabilities when assessing the model performance. Third, the applicability of the prediction model in other populations was not elucidated. Lastly, when applying our prediction model to the recent Japanese population, the predicted absolute risk of ASCVD could be overestimated, as there is a possibility that the incidence rate of ASCVD in the Japanese population has been decreasing with the recent improvement of treatment and screening methods. Thus, our prediction model for ASCVD should be externally validated in other populations, especially in recently established cohorts. Nevertheless, the estimation from our prediction model could be helpful for health instruction for the prevention of ASCVD in the general population and patients.

In conclusion, we developed a prediction model specifically for ASCVD events in Japanese adults. This model demonstrated good discrimination and calibration as well as satisfactory internal validity. This prediction model exhibits great potential as a tool for predicting ASCVD risk in routine clinical practice by enabling the identification of specific risk factors for ASCVD in individual patients. This, in turn, could facilitate practitioner–patient communications and improve targeted and personalized management of risk factors. In addition, our simplified risk calculator enables patients to quantify their own absolute ASCVD risk and could thereby promote effective self-monitoring and self-management[33]. Further studies are warranted to evaluate the effectiveness of the model-guided screening and interventions.

Conflict of Interest

We have no conflict of interest to declare.

References

1) Viles-Gonzalez JF, Fuster V, Badimon JJ. Atherothrombosis: a widespread disease with unpredictable and life-threatening consequences. Eur Heart J, 2004; 25: 1197-1207
2) Hoshino T, Sissani L, Labreuche J, Ducrocq G, Lavallée PC, Meseguer E, Guidoux C, Cabrejo L, Hobeanu C, Gongora-Rivera F, Touboul PJ, Gabriel Steg P, Amarenco P. Prevalence of systemic atherosclerosis burdens and overlapping stroke etiologies and their associations with long-term vascular prognosis in stroke with intracranial atherosclerotic disease. JAMA Neurol, 2018; 75: 203-211
3) Davis KL, Meyers J, Zhao Z, McCollam PL, Murakami M. High-risk atherosclerotic cardiovascular disease in a real-world employed japanese population: prevalence, cardiovascular event rates, and costs. J Atheroscler Thromb, 2015; 22: 1287-1304
4) Bhatt DL, Gabriel Steg P, Magnus Ohman E, Hirsch AT, Ikeda Y, Mas JL, Goto S, Liau CS, Richard AJ, Röther J, Wilson PWF. International prevalence, recognition, and treatment of cardiovascular risk factors in outpatients with atherothrombosis. JAMA, 2006; 295: 180-189

5) Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, Tzoulaki I, Lassale CM, Siontis GCM, Chiocchia V, Roberts C, Schlüssel MM, Gerry S, Black JA, Heus P, Van Der Schouw YT, Peelen LM, Moons KGM. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ, 2016; 353: i2416

6) Arima H, Yonemoto K, Doi Y, Ninomiya T, Hata J, Tanizaki Y, Fukuhara M, Matsumura K, Iida M, Kiyohara Y. Development and validation of a cardiovascular risk prediction model for Japanese: the Hisayama Study. Hypertens Res, 2009; 32: 1119-1122

7) Ishikawa S, Matsumoto M, Kayaba K, Gotoh T, Nago N, Tsutsumi A, Kajii E, Ishikawa S, Kayaba K, Gotoh T, Nago N, Tsutsumi A, Kajii E. Risk charts Illustrating the 10-year risk of myocardial infarction among residents of Japanese rural communities: the JMS Cohort Study. J Epidemiol, 2009; 19: 94-100

8) Yatsuya H, Iso H, Yamagishi K, Kokubo Y, Saito I, Suzuki K, Sawada N, Inoue M, Tsugane S. Development of a point-based prediction model for the incidence of total stroke: Japan Public Health Center Study. Stroke, 2013; 44: 1295-1302

9) Nishimura K, Okamura T, Watanabe M, Nakai M, Takegami M, Higashiyama A, Kokubo Y, Okayama A, Miyamoto Y. Predicting coronary heart disease using risk factor categories for a Japanese urban population, and comparison with the Framingham Risk Score: the Suita Study. J Atheroscler Thromb, 2014; 21: 784-798

10) Nakai M, Watanabe M, Kokubo Y, Nishimura K, Higashiyama A, Takegami M, Nakao YM, Okamura T, Miyamoto Y. Development of a cardiovascular disease risk prediction model using the Suita Study, a population-based prospective cohort study in Japan. J Atheroscler Thromb, 2020; 27: 1160-1175

11) Honda T, Yoshida D, Hata J, Hirakawa Y, Ishida Y, Shibata M, Sakata S, Kitazono T, Ninomiya T. Development and validation of modified risk prediction models for cardiovascular disease and its subtypes: the Hisayama Study. Atherosclerosis, 2018; 279: 38-44

12) Lackland DT, Elkind MS V, Agostino RD, Dhamoon MS, Goff DC, Higashida RT, Mcclure LA, Mitchell PH, Sacco RL, Sila CA, Smith SC, Tanne D, Tirschwell DL, Touze E. Inclusion of stroke in cardiovascular risk prediction instruments: a statement for healthcare professionals from the American Heart Association/ American Stroke Association. Stroke, 2012; 43: 1998-2027

13) Goff DCJ, Lloyd-jones DM, Bennett G, Coady S, Agostino RBD, Gibbons R, Lackland DT, Levy D, Donnell CJO, Robinson JG, Schwartz JS, Shero ST, Smith SC, Sorlie P, Stone NJ, Wilson PWF, Anderson JL, Halperin JL, Albert NM, Bozkurt B, Brindis RG, Curtis LH, Demets D, Hochman JS, Kovacs RJ, Ohman EM, Pressler SJ, Sellke FW, Shen W, Smith SCJ, Tomaselli GF. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation, 2014; 129: s49-73

14) Kinoshita M, Yokote K, Arai H, Iida M, Ishigaki Y, Ishibashi S, Umemoto S, Egusa G, Ohmura H, Okamura T, Kihara S, Koba S, Saito I, Shoji T, Daida H, Tsukamoto K, Deguchi J, Dohi S, Dobashi K, Hamaguchi H, Hara M, Hiro T, Biro S, Fujioka Y, Maruyama C, Miyamoto Y, Murakami Y, Yokode M, Yoshida H, Rakugi H, Wakatsuki A, Yamashita S, Okamura T, Ninomiya T, Hozawa A, Yokode M, Iida M, Hasegawa K, Hara M, Yamato H, Shoji T, Abe T, Egusa G, Emoto M, Kashihara N, Saiki A, Tomiyama H, Hashimoto JI, Hosomi N, Yamazaki Y, Bujo H, Inoue I, Inoue S, Koba S, Sawada S, Sone H, Fujioka Y, Hiro S, Funabashi T, Maruyama C, Matsumoto M, Ohtsuki T, Kitagawa K, Kitazono T, Terayama Y, Tomimoto H, Toyoda K, Harada-Shiba M, Arai H, Ishibashi S, Nohara A, Ohta T, Ogura M, Ohtake A, Yokote K. Japan Atherosclerosis Society (JAS) guidelines for prevention of atherosclerotic cardiovascular diseases 2017. J Atheroscler Thromb, 2018; 25: 846-984

15) Imamura T, Doi Y, Ninomiya T, Hata J, Nagata M, Ikeda F, Mukai N, Hirakawa Y, Yoshida D, Fukuhara M, Kitazono T, Kiyohara Y. Non-high-density lipoprotein cholesterol and the development of coronary heart disease and stroke subtypes in a general Japanese population: the Hisayama Study. Atherosclerosis, 2014; 233: 343-348

16) Imamura T, Doi Y, Arima H, Yonemoto K, Hata J, Kubo M, Tanizaki Y, Ibayashi S, Iida M, Kiyohara Y. LDL cholesterol and the development of stroke subtypes and coronary heart disease in a general Japanese population: the Hisayama study. Stroke, 2009; 40: 382-388

17) Yang X, Li J, Hu D, Chen J, Li Y, Huang J, Liu X, Liu F, Cao J, Shen C, Yu L, Lu F, Wu X, Zhao L, Wu X, Gu D. Predicting the 10-year risks of atherosclerotic cardiovascular disease in Chinese population: the China-PAR Project (Prediction for ASCVD Risk in China). Circulation, 2016; 134: 1430-1440

18) De Backer G, Ambrosioni E, Borch-Johnsen K, Brotons C, Cifkova R, Dallongeville J, Ebrahim S, Faergeman O, Graham I, Mancia G, Cats VM, Orth-Gomér K, Perk J, Pyörälä K, Rodicio JL, Sans S, Sansoy V, Sechtem U, Silber S, Thomsen T, Wood D. European guidelines on cardiovascular disease prevention in clinical practice: Third Joint Task Force of European and other Societies on Cardiovascular Disease Prevention in clinical practice. Eur Heart J, 2003; 24: 1601-1610

19) Yatsuya H, Iso H, Li Y, Yamagishi K, Kokubo Y, Saito I, Sawada N, Inoue M, Tsugane S. Development of a risk equation for the incidence of coronary artery disease and ischemic stroke for middle-aged Japanese–Japan Public Health Center-Based Prospective Study–. Circ J, 2016; 80: 1386-1395

20) Ninomiya T. Japanese legacy cohort studies: the Hisayama Study. J Epidemiol, 2018; 28: 444-451

21) Tanizaki Y, Kiyohara Y, Kato I, Iwamoto H, Nakayama K, Shinohara N, Arima H, Tanaka K, Ibayashi S, Fujishima M. Incidence and risk factors for subtypes of cerebral infarction in a general population: the Hisayama study. Stroke, 2000; 31: 2616-2622

22) Hata J, Ninomiya T, Hirakawa Y, Nagata M, Mukai N, Gotoh S, Fukuhara M, Ikeda F, Shikata K, Yoshida D, Yonemoto K, Kamouchi M, Kitazono T, Kiyohara Y. Secular trends in cardiovascular disease and its risk factors in Japanese: half-century data from the Hisayama Study (1961-2009). Circulation, 2013; 128: 1198-1205

23) Special report from the National Institute of Neurological

Disorders and Stroke. Classification of cerebrovascular diseases III. Stroke, 1990; 21: 637-676

24) Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke, 1993; 24: 35-41

25) Cerebral Embolism Task Force. Cardiogenic brain embolism. Arch Neurol, 1986; 43: 71-84

26) Pencina MJ, D'Agostino RB, Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease: the Framingham Heart Study. Circulation, 2009; 119: 3078-3084

27) Tai BC, Machin D, White I, Gebski V. Competing risks analysis of patients with osteosarcoma: a comparison of four different approaches. Stat Med, 2001; 20: 661-684

28) Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. Stat Med, 2004; 23: 1631-1660

29) 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

30) Demler O V., Paynter NP, Cook NR. Tests of calibration and goodness-of-fit in the survival setting. Stat Med, 2015; 34: 1659-1680

31) Deanfield J, Sattar N, Simpson I, Wood D, Bradbury K, Fox K, Boon N, Winocour P, Feher M, Doherty P, Buckley J, Jennings C, Keenan J, Weissberg P, Knapton M, Williams B, Poulter N, Stevens P, Taal M, Grant P, Home P, Rees A, Neely D, Wheeler D, Goldsmith D, Potter J, Jackson S. Joint British Societies' consensus recommendations for the prevention of cardiovascular disease (JBS3). Heart, 2014; 100: ii1-ii67

32) Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, Cooney MT, Corrà U, Cosyns B, Deaton C, Graham I, Hall MS, Hobbs FDR, Løchen ML, Löllgen H, Marques-Vidal P, Perk J, Prescott E, Redon J, Richter DJ, Sattar N, Smulders Y, Tiberi M, Van Der Worp HB, Van Dis I, Verschuren WMM, Binno S, De Backer G, Roffi M, Aboyans V, Bachl N, Carerj S, Cho L, Cox J, De Sutter J, Egidi G, Fisher M, Fitzsimons D, Franco OH, Guenoun M, Jennings C, Jug B, Kirchhof P, Kotseva K, Lip GYH, Mach F, Mancia G, Bermudo FM, Mezzani A, Niessner A, Ponikowski P, Rauch B, Stauder A, Turc G, Wiklund O, Windecker S, Zamorano JL, Achenbach S, Badimon L, Barón-Esquivias G, Baumgartner H, Bax JJ, Dean V, Erol Ç, Gaemperli O, Kolh P, Lancellotti P, Nihoyannopoulos P, Torbicki A, Carneiro AV, Metzler B, Najafov R, Stelmashok V, De Maeyer C, Dilić M, Gruev I, Miličić D, Vaverkova H, Gustafsson I, Attia I, Duishvili D, Ferrières J, Kostova N, Klimiashvili Z, Hambrecht R, Tsioufis K, Szabados E, Andersen K, Vaughan C, Zafrir B, Novo S, Davletov K, Jashari F, Kerimkulova A, Mintale I, Saade G, Petrulioniene Z, Delagardelle C, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J, 2016; 37: 2315-2381

33) Riegel B, Moser DK, Buck HG, VaughanDickson V, B.Dunbar S, Lee CS, Lennie TA, Lindenfeld JA, Mitchell JE, Treat-Jacobson DJ, Webber DE. Self-care for the prevention and management of cardiovascular disease and stroke: a scientific statement for healthcare professionals from the American heart association. J Am Heart Assoc, 2017; 6: e006997

Supplemental Methods

Statistical Analysis

We constructed a prediction model based on Cox's proportional hazards regression analyses. Potential predictors for the main analysis included age, sex, systolic blood pressure, use of antihypertensive medication, diabetes, serum HDL cholesterol, serum LDL cholesterol, body mass index, ECG abnormality, proteinuria, smoking, drinking, and regular exercise. To select risk factors from those potential predictors, a backward selection was performed with a P-value $<$ 0.10 as the variable elimination criterion. The stability of variable selection was checked by using a bootstrap resampling[1]. Analyses were also performed separately for predicting coronary heart disease and atherothrombotic brain infarction to confirm the consistency in the associations of predictors with these two outcomes.

The 10-year probability of developing ASCVD was computed based on the Cox regression analysis as follows:

$$\hat{p} = 1 - S_0(t)^{\exp(\Sigma \beta_i x_i - \Sigma \beta_i \bar{x}_i)},$$

where $S_0(t)$ was the average event-free survival at the time point of t (i.e., at year 10), and $\Sigma \beta_i x_i$ and $\Sigma \beta_i \bar{x}_i$ were computed by summing the products of the regression coefficients of each predictor (β_i) and individual values (x_i), and the mean values ($\bar{x}_i$), respectively.

A potential issue when using the standard Cox regression model to develop a risk prediction model, especially in the case of lifetime risk prediction, is the lack of consideration of competing risk. The Kaplan-Meier survival function tends to result in an overestimation of the incidence of the outcome of interest due to the competing risk of death. To account for the competing risk of death without ASCVD events during the follow-up, we modified the baseline survival function ($S_0[t]$) of the above prediction equation by a cumulative incidence function that accounts for the competing risk[2, 3]. First, we estimated the mean death- and ASCVD event-free survival probabilities ($S_{joint}(t)$) and ASCVD-free survival probabilities ($S_{event}(t)$) by using standard Cox regression models. The instantaneous hazard function of experiencing ASCVD events at time t ($h_{event}(t)$) was calculated as $h_{event}(t) = -\log(S_{event}(t_i) / S_{event}(t_i\text{-}1))$. The cumulative incidence function of ASCVD ($I_{event}(t)$) was then computed by summing the products of the instantaneous hazard function by the survival estimates at the time preceding t ($\Sigma h_{event}(t_i) \times S_{joint}(t_i\text{-}1)$). The cumulative incidence function at 10-year follow-up was subtracted from one to obtain the survival function that accounts for competing risk

($S_0(t=10)_{CR} = 1 - I_{event}(t=10)$), which was used as the baseline survival function (i.e., $S_0(t=10)$) instead of the naïve estimate from the Cox regression model.

The developed model was translated into a simplified risk score based on the coefficients of each predictor following the instruction by Sullivan et al.[4].

Continuous predictors were converted to categories to determine the risk scores. Age was categorized into 5 groups (40–49 years, 50–59 years, 60–69 years, 70–79 years, and 80–84 years). The youngest group was set as a reference category, and the values of 45, 55, 65, 75, and 82 were assigned to the respective groups. The levels of systolic blood pressure, serum HDL cholesterol, and serum LDL cholesterol were divided into predefined categories ($<$120, 120–129, 130–139, 140–159, $\geq$ 160 mmHg for SBP; $<$40, 40–59, $\geq$ 60 mg/dL for serum HDL cholesterol; $<$120, 120–139, 140–159, $\geq$ 160 mg/dL for serum LDL cholesterol). For each category of these variables, the nearest multiple of 5 to the median values was assigned. For the remaining categorical variables, the healthier of the dichotomous categories were set as a reference (assigned a value of 0), and the value of 1 was assigned to the unhealthier ones (i.e., male sex, presence of diabetes, proteinuria, current smoking, no regular exercise). The points for each category (j) of each predictor (i) were determined as follows:

$$\text{Point}_{ij} = \beta_i (W_{ij} - W_{iREF}) / \text{Constant} ,$$

where β_i is the β estimate for the predictor i in the risk prediction model described above. W and W_{REF} are the assigned values of each category and the reference category, respectively. Therefore, $W_{ij} - W_{iREF}$ is the distance of each category of each predictor from its reference category in their original units. We assigned a value of 0.144 to the constant, which was the β estimate for a 30 mg/dL increment in LDL cholesterol (i.e., the lowest value across the values of β_i [$W_{ij} - W_{iREF}$]). The determination of points is shown in **Supplemental Table 4**.

The simplified risk score-based estimates of 10-year probabilities of ASCVD are presented according to age groups. Thus, the points of variables except for age are summed up as a total score, and the points for age are added when calculating the probability of ASCVD incidence as follows:

$$\hat{p} = 1 - S_0(t=10)_{CR}^{\exp([\text{total score} + \text{points for age}] \times 0.144 - 2.4767)},$$

where the constant 2.4767 was calculated as:

$$\Sigma \beta_i \bar{x}_i - (\beta_i \times W_{iREF}) = 6.7963 - 4.3196.$$
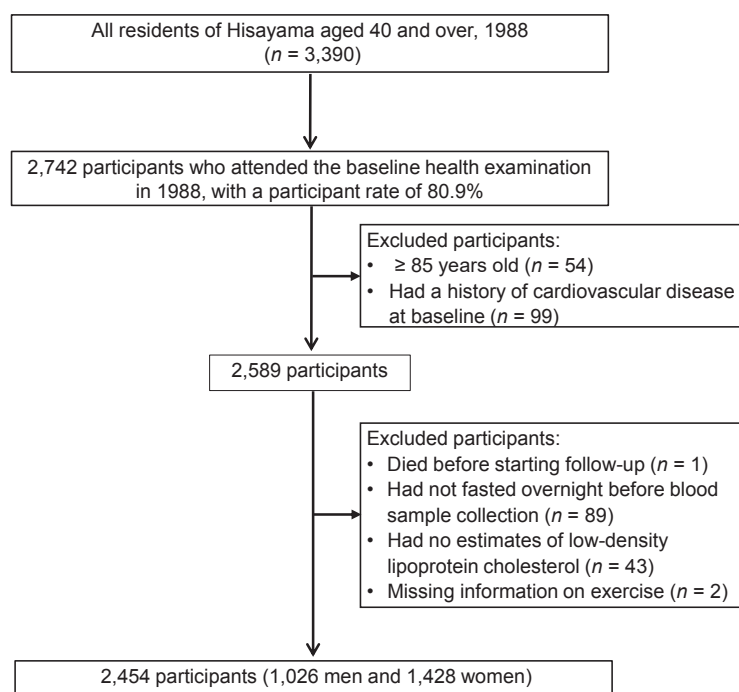
The agreement between the 10-year ASCVD probability predicted by the multivariable model and by the simplified score was assessed by using Spearman's rank correlation and a bivariate linear regression of the model-predicted probability on the score-predicted probability.

The discrimination of the developed model was assessed by the Harrell's concordance index (C index). The optimism was estimated by using 200 bootstrap samples. The optimism-corrected C index was calculated following the procedure proposed by Harrell *et al.*[5]. The calibration was assessed graphically by plotting the average of model-predicted probabilities against the observed probabilities over 10 years according to deciles of predicted 10-year probabilities. The Greenwood-Nam-D'Agostino (GND) test was performed to statistically test the calibration[6]. In this test, the cumulative incidence accounted for the competing risk of non-ASCVD death at the time of year 10 was substituted for the observed probabilities that were originally estimated by using the Kaplan Meier estimator. Because of the low numbers of ASCVD cases in the lower probability categories, deciles with less than five cases were collapsed with the next decile. We thus combined the lower 5 deciles into one group, and a GND test with 5 degrees of freedom was performed. The GND test for the simplified score was performed with 6 degrees of freedom, by combining the first through the third deciles into one category and the fourth and the fifth deciles into another category.

We developed an alternative model that included serum non-HDL cholesterol instead of serum LDL cholesterol, because the Japan Atherosclerosis Society guideline recommends the assessment of serum non-HDL cholesterol instead of serum LDL cholesterol when using non-fasting samples or samples with serum TG levels $\geq$ 400 mg/dL. In addition, an alternative simplified score, in which 0, 1, 2, or 3 points were assigned to predefined categories of serum non- HDL cholesterol ($<$150, 150–169, 170–189, $\geq$ 190 mg/dL) instead of the points for LDL cholesterol levels, was also evaluated. We assessed the agreement between the 10-year probability of ASCVD predicted by the original multivariable model (including LDL cholesterol levels) and that predicted by the alternative non-HDL cholesterol model using Spearman's rank correlation coefficient and a linear regression of the predicted probability by the original multivariable model on that by the alternative model. Similarly, we assessed the agreement between the 10-year ASCVD probability predicted by the original simplified score and that predicted by the alternative non-HDL cholesterol simplified score that substituted points of serum non-HDL cholesterol for those of LDL cholesterol. All statistical analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC). A two-sided *P*-value of $<$0.05 was considered statistically significant.

References for the Supplemental Methods

1) Richter B, Koller L, Hohensinner PJ, Zorn G, Brekalo M, Berger R, Mörtl D, Maurer G, Pacher R, Huber K, Wojta J, Hülsmann M, and Niessner A. A multi- biomarker risk score improves prediction of long-term mortality in patients with advanced heart failure. Int J Cardiol, 2013; 168: 1251-1257

2) Pencina MJ, D'Agostino RB, Larson MG, Massaro JM, and Vasan RS. Predicting the 30-year risk of cardiovascular disease: The framingham Heart Study. Circulation, 2009; 119: 3078-3084

3) Tai B, Machin D, White I, and Gebski V. Competing risks analysis of patients with osteosarcoma: a comparison of four different approaches. Stat Med, 2001; 20: 661-684

4) Sullivan LM, Massaro JM, and D'Agostino RB. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. Stat Med, 2004; 23: 1631-1660

5) Harrell FE, Lee KL, and Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med, 1996; 15: 361-387

6) Demler OV, Paynter NP, and Cook NR. Tests of calibration and goodness-of- fit in the survival setting. Stat Med, 2015; 34: 1659-1680

```
┌─────────────────────────────────────────────────┐
│   All residents of Hisayama aged 40 and over, 1988 │
│                  (n = 3,390)                       │
└─────────────────────────────────────────────────┘
                        │
                        ▼
┌─────────────────────────────────────────────────┐
│  2,742 participants who attended the baseline      │
│  health examination in 1988, with a participant   │
│  rate of 80.9%                                     │
└─────────────────────────────────────────────────┘
          │
          │         ┌──────────────────────────────────────────┐
          │────────▶│ Excluded participants:                    │
          │         │ •  ≥ 85 years old (n = 54)                 │
          │         │ •  Had a history of cardiovascular disease │
          │         │    at baseline (n = 99)                    │
          ▼         └──────────────────────────────────────────┘
┌───────────────────────────┐
│   2,589 participants       │
└───────────────────────────┘
          │
          │         ┌──────────────────────────────────────────┐
          │         │ Excluded participants:                    │
          │         │ •  Died before starting follow-up (n = 1) │
          │────────▶│ •  Had not fasted overnight before blood  │
          │         │    sample collection (n = 89)              │
          │         │ •  Had no estimates of low-density         │
          │         │    lipoprotein cholesterol (n = 43)        │
          │         │ •  Missing information on exercise (n = 2) │
          ▼         └──────────────────────────────────────────┘
┌─────────────────────────────────────────────────┐
│  2,454 participants (1,026 men and 1,428 women)   │
└─────────────────────────────────────────────────┘
```

Supplemental Fig. 1. Flow chart of the study sample

Supplemental Table 1. Calculation of the cumulative incidence function and survival of ASCVD events that accounted for competing risk of death

Follow-up, years	Survival function for joint events, $S_{joint}(t)$	Survival function for ASCVD events, $S_{event}(t)$	Hazard function for ASCVD events, $h_{event}(t)$	Cumulative incidence function, $I_{event}(t)$	Survival accounting for competing risk of death, $S_0(t)_{CR} = 1 - I_{event}(t)$
0	1.0000	1.0000	0.0000	0.0000	1.0000
1	0.9956	0.9988	0.0012	0.0012	0.9988
2	0.9889	0.9962	0.0026	0.0038	0.9962
3	0.9813	0.9937	0.0025	0.0062	0.9938
4	0.9733	0.9914	0.0023	0.0086	0.9915
5	0.9660	0.9887	0.0026	0.0112	0.9888
6	0.9551	0.9850	0.0037	0.0149	0.9851
7	0.9444	0.9810	0.0039	0.0187	0.9813
8	0.9348	0.9779	0.0030	0.0217	0.9783
9	0.9214	0.9735	0.0042	0.0259	0.9741
10	0.9039	0.9688	0.0045	0.0304	0.9696

The denotations are as described in the Supplemental methods. Follow-up was done for 24 years but the values for >10-year follow-up are not presented in the table. Joint events were defined as either ASCVD incidence or death without an ASCVD event. Survival functions ($S_{joint}(t)$ and $S_{event}(t)$) were estimated based on the Cox regression model. The (instantaneous) hazard function of ASCVD ($h_{event}(t)$) was calculated as $-\log$(survival at t_i /survival at t_{i-1}). The cumulative incidence function at time t ($I_{event}(t)$) was calculated as $\Sigma(h_{event}(t_i) \times S_{joint}(t_{i-1}))$.
Abbreviation: ASCVD, atherosclerotic cardiovascular disease.
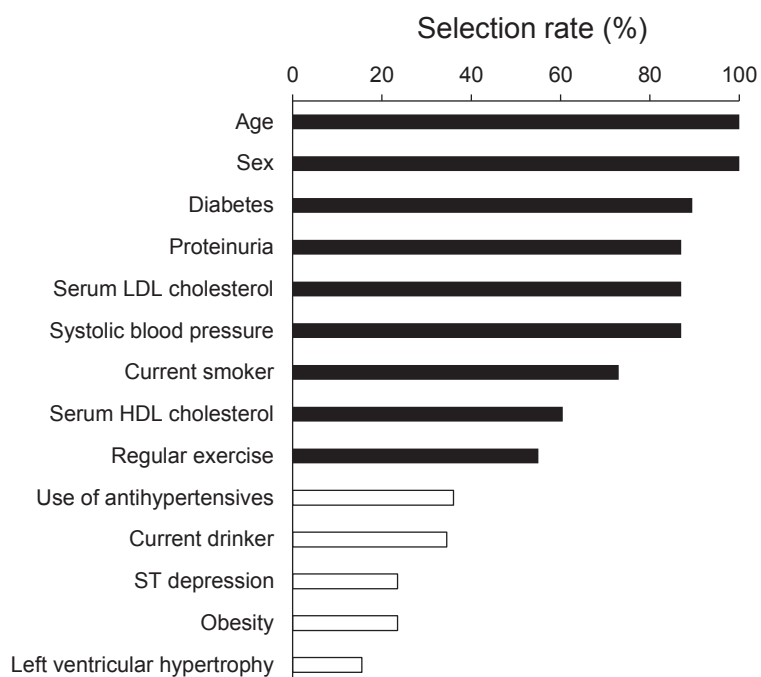
Selection rate (%)

Supplemental Fig. 2. Selection rate of the candidate predictors in 200 bootstrapping samples

The solid black bar indicates variables that selected for >40% (i.e., >80 times) of the bootstrapping resampling procedure.
Abbreviations: HDL cholesterol, high-density lipoprotein cholesterol; LDL cholesterol, low-density lipoprotein cholesterol.

Supplemental Table 2. Separate Cox's multivariable models for coronary heart disease and atherosclerotic brain infarction

	Coronary heart disease (*n* of events = 216)			Atherothrombotic brain infarction (*n* of events = 62)		
	HR (95% CI)	β	*P*-value	HR (95% CI)	β	*P*-value
Age (per year)	1.08 (1.07-1.10)	0.079	<0.001	1.08 (1.05-1.11)	0.075	<0.001
Men (vs women)	2.78 (2.01-3.85)	1.023	<0.001	2.75 (1.51-5.02)	1.011	0.001
Systolic blood pressure (per 1 mmHg)	1.01 (1.00-1.02)	0.011	0.001	1.01 (1.00-1.02)	0.008	0.21
Diabetes (vs nondiabetic)	1.61 (1.15-2.25)	0.478	0.005	1.52 (0.80-2.88)	0.417	0.20
Serum HDL cholesterol (per 1 mg/dL)	0.99 (0.97-1.00)	-0.015	0.02	1.00 (0.98-1.02)	-0.003	0.81
Serum LDL cholesterol (per 1 mg/dL)	1.01 (1.00-1.01)	0.005	0.003	1.00 (1.00-1.01)	0.002	0.47
Proteinuria (vs absent)	1.74 (1.11-2.73)	0.553	0.02	2.12 (0.95-4.74)	0.750	0.07
Current smoker (vs non-smoker)	1.44 (1.04-1.98)	0.362	0.03	1.38 (0.76-2.52)	0.323	0.30
No or irregular exercise (vs regular)	1.33 (0.86-2.04)	0.282	0.20	2.29 (0.82-6.38)	0.828	0.11
C statistics (95% CI)	0.786 (0.758-0.814)			0.780 (0.751-0.808)		

Abbreviations: HR, hazard ratio; 95% CI, 95% confidence interval; HDL cholesterol, high-density lipoprotein cholesterol; LDL cholesterol, low-density lipoprotein cholesterol.

Supplemental Table 3. Determination of points for the simplified risk score calculation

Variable	Levels	Median in the sample	Assigned value	Difference from reference value (A)	β coefficients (B)	Weight in regression units (C): (A) × (B)	Point ([C]/0.144)
Age	40-49 years		45 (reference)	0		0.000	0
	50-59 years		55	10	0.077	0.771	5
	60-69 years		65	20		1.543	11
	70-79 years		75	30		2.314	16
	80-84 years		82	37		2.854	20
Sex	Women		0	0		0.000	0
	Men		1	1	0.984	0.984	7
Systolic blood pressure	<120 mmHg	112	110 (reference)	0		0.000	0
	120-129 mmHg	124	125	15	0.010	0.155	1
	130-139 mmHg	135	135	25		0.259	2
	140-159 mmHg	147	150	40		0.414	3
	160 mmHg -	169	170	60		0.621	4
Diabetes	No		0 (reference)	0		0.000	0
	Yes		1	1	0.459	0.459	3
Serum HDL cholesterol	60 mg/dL -	65	65 (reference)	0		0.000	0
	40-59 mg/dL	49	50	-15	-0.012	0.178	1
	<40 mg/dL	35	35	-30		0.356	2
Serum LDL cholesterol	<120 mg/dL		100 (reference)	0		0.000	0
	120-139 mg/dL	129	130	30	0.005	0.144	1
	140-159 mg/dL	149	150	50		0.240	2
	160 mg/dL -	181	180	80		0.384	3
Proteinuria	No		0 (reference)	0		0.000	0
	Yes		1	1	0.632	0.632	4
Current smoker	No		0 (reference)	0		0.000	0
	Yes		1	1	0.336	0.336	2
Regular exercise	No		1 (reference)	1	0.339	0.339	2
	Yes		0	0		0.000	0

Abbreviations: HDL cholesterol, high-density lipoprotein cholesterol; LDL cholesterol, low-density lipoprotein cholesterol.
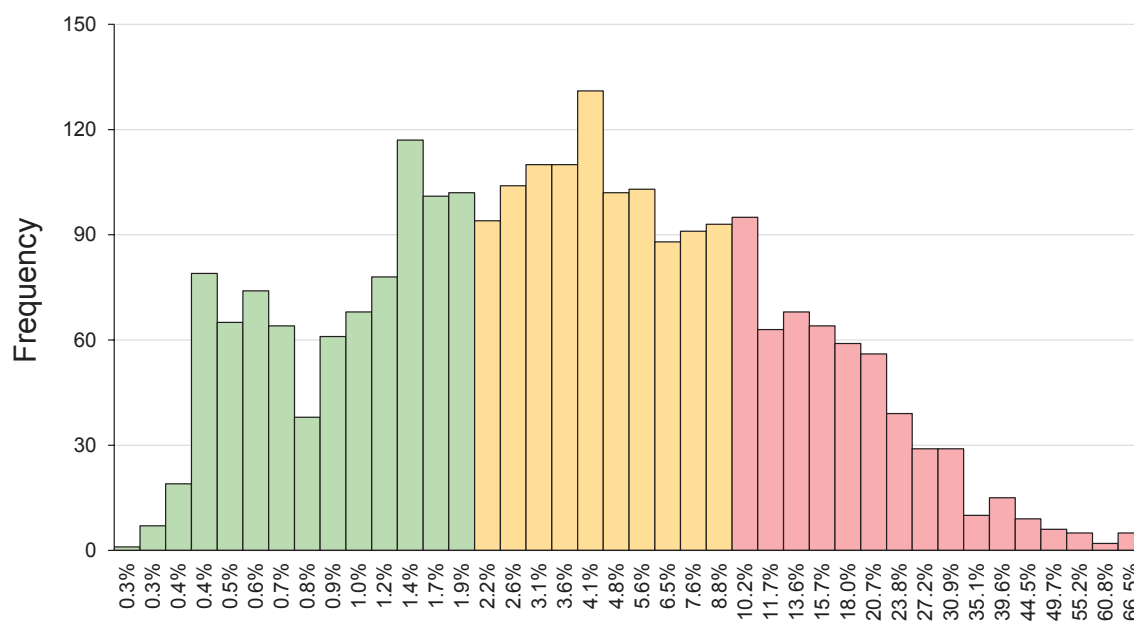
The 10-year ASCVD probabilities predicted by the simplified risk score

Supplemental Fig. 3. Histogram of the 10-year ASCVD probabilities predicted by the simplified risk score

Bars were color-coded as green (low-risk: <2.0% of 10-year atherosclerotic cardiovascular disease risk, corresponding to the lowest 35% of distribution in the population), yellow (middle-risk: 2.0%–10.0%), and red (high-risk: ≥ 10%, corresponding to the highest 20% of distribution in the population)

Supplemental Table 4. Multivariable model using serum non-HDL cholesterol, instead of LDL cholesterol, as a predictor for the risk of atherosclerotic cardiovascular disease

	HR (95% CI)	β	*P*-value
Age (per year)	1.08 (1.07-1.10)	0.078	<0.001
Men (vs women)	2.66 (2.00-3.55)	0.979	<0.000
Systolic blood pressure (per 1 mmHg)	1.01 (1.00-1.02)	0.010	<0.001
Diabetes (vs nondiabetic)	1.55 (1.14-2.09)	0.436	0.005
Serum HDL cholesterol (per 1 mg/dL)	0.99 (0.98-1.00)	-0.010	0.07
Serum non-HDL cholesterol (per 1 mg/dL)	1.01 (1.00-1.01)	0.005	0.001
Proteinuria (vs absent)	1.88 (1.26-2.79)	0.629	0.002
Current smoker (vs non-smoker)	1.41 (1.05-1.88)	0.341	0.02
No or irregular exercise (vs regular)	1.40 (0.94-2.08)	0.334	0.10
C statistics (95% CI)	0.786 (0.758-0.814)		

Abbreviations: HR, hazard ratio; 95% CI, 95% confidence interval; HDL cholesterol, high-density lipoprotein cholesterol; LDL cholesterol, low-density lipoprotein cholesterol.
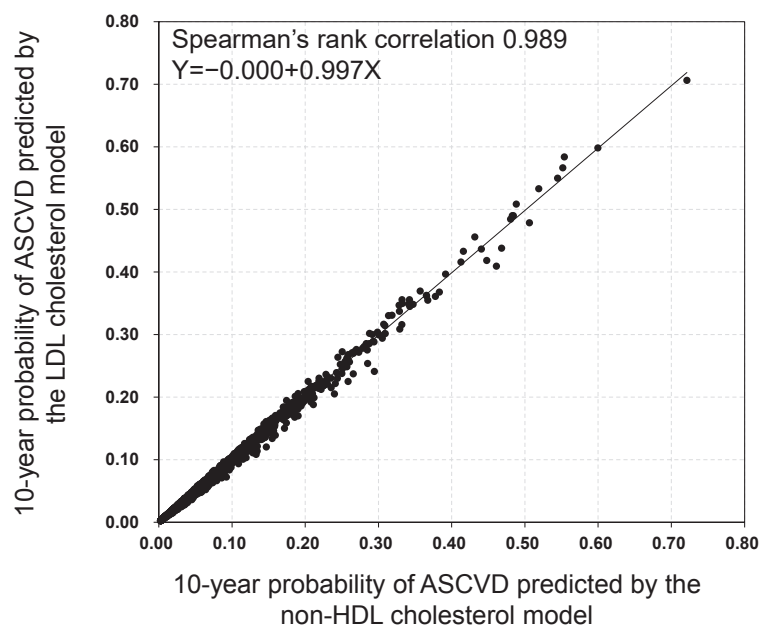
Supplemental Fig. 4. Agreement between the final multivariable model using serum LDL cholesterol and the alternative model using serum non-HDL cholesterol as a predictor
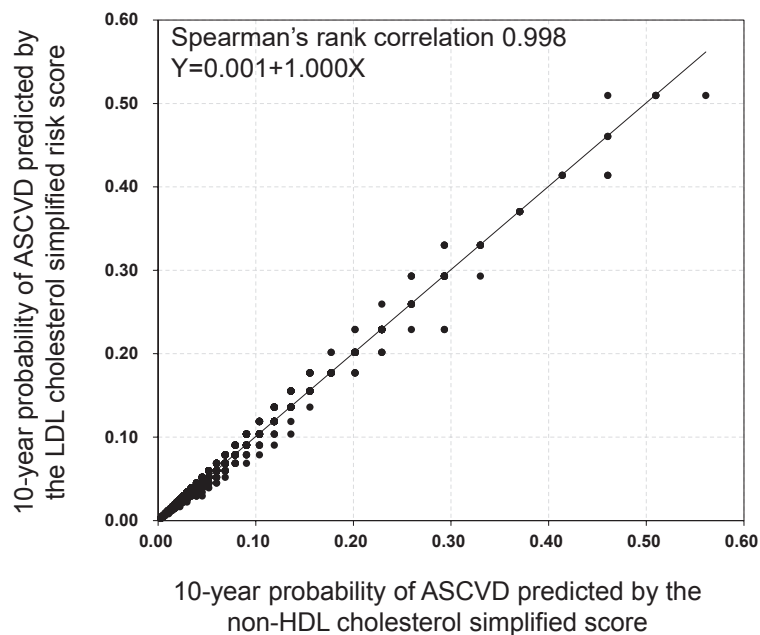
Supplemental Fig. 5. Agreement between the simplified risk score using serum LDL cholesterol and the alternative simplified score using serum non-HDL cholesterol as a predictor