ORIGINAL ARTICLE

Incidence of atrial fibrillation and its risk prediction model based on a prospective urban Han Chinese cohort

L Ding[1], J Li[2], C Wang[3], X Li[1], Q Su[4], G Zhang[5] and F Xue[1]

Prediction models of atrial fibrillation (AF) have been developed; however, there was no AF prediction model validated in Chinese population. Therefore, we aimed to investigate the incidence of AF in urban Han Chinese health check-up population, as well as to develop AF prediction models using behavioral, anthropometric, biochemical, electrocardiogram (ECG) markers, as well as visit-to-visit variability (VVV) in blood pressures available in the routine health check-up. A total of 33 186 participants aged 45–85 years and free of AF at baseline were included in this cohort, to follow up for incident AF with an annually routine health check-up. Cox regression models were used to develop AF prediction model and 10-fold cross-validation was used to test the discriminatory accuracy of prediction model. We developed three prediction models, with age, sex, history of coronary heart disease (CHD), hypertension as predictors for simple model, with left high-amplitude waves, premature beats added for ECG model, and with age, sex, history of CHD and VVV in systolic and diabolic blood pressures as predictors for VVV model, to estimate risk of incident AF. The calibration of our models ranged from 1.001 to 1.004 (P for Hosmer Lemeshow test > 0.05). The area under receiver operator characteristics curve were 78%, 80% and 82%, respectively, for predicting risk of AF. In conclusion, we have identified predictors of incident AF and developed prediction models for AF with variables readily available in routine health check-up.

INTRODUCTION

Atrial fibrillation (AF), one of the most common cardiac arrhythmia in clinical practice, has emerged as a major public health problem and is closely related to stroke, mortality, decreased quality of life and a high healthcare cost burden.[1] In the last 20 years, the prevalence of AF was increasing worldwide[1,2] and affected ~ 5–8% of those older than 65 years in western countries.[3–6] It was estimated that the number of patients with AF in 2030 in Europe would be 14 to 17 million and the number of incident AF for 1 year would be 120 000 to 215 000.[1] In China, the updated information on the prevalence of AF is 3.5% for those above 65 years old and the age-standardized prevalence of AF in China (⩾30 years) is 0.65%.[7] Although the prevalence of AF is lower in China than in western countries, the prevalence of AF has increased 20-fold over an 11-year period in China,[8] and AF burden as well as the risk of AF-related stroke have increased significantly over the past decade.[8] The increasing trend of AF challenges the prevention and control of cardiovascular disease, and identifying high-risk individuals for targeted prevention of AF is useful in the primary prevention.

Various prediction models have been constructed to assist clinicians and epidemiologists in assessing individual's risk of AF. The most widely used model is the 10-year risk score in the Framingham Heart Study[9] and the more recent one for 5-year risk of AF was the Cohorts for Heart and Aging Research in Genomic Epidemiology-AF scores.[10] All these prediction models were derived from theUS population and validations in minorities in the United States have shown poor calibration and moderate discriminatory accuracy.[11] To our knowledge, no studies have

evaluated the applicability of these prediction models in Asia population and there is no AF prediction model derived from Asia population. As both the incidence and risk burden of AF vary by regions, races and social economic status,[2] it is necessary to investigate an applicable AF prediction model in Chinese population, to help estimate AF risks and identify high-risk individuals for primary prevention of AF in clinical and public health practice.

Therefore, in this study, we aimed to investigate the incidence of AF in the Chinese health check-up population, as well as to construct an AF prediction model using behavioral, anthropometric, biochemical and electrocardiogram (ECG) markers available in the routine health check-up.

SUBJECTS AND METHODS

Study population, cohort design and diagnosis of AF

The study was based on Shandong multi-center health check-up longitudinal study. The database of Shandong multi-center health check-up longitudinal study include individuals participated in routine health check-up at several hospitals from 2004 to 2014. The characteristics of participants in this cohort included Han Nationality, occupational or retired (having jobs or retirement) and living in urban area. A total of 95 874 participants who visited the health check-up for at least two times with ECG examination in the health check-up from 2004 to 2014 were included in this study. Among these participants, 252 persons were excluded for having AF or atrial flutter at baseline; therefore, 95 622 participants were included as original cohort to get age-specific incidence. As the incidence of AF was very low before 45 years (see Figure 1), 33 186 participants aged 45–85 years were included in the final analysis (see Figure 2 for the framework of this cohort). Incident AF cases were ascertained from a rest

[1]Department of Biostatistics, School of Public Health, Shandong University, Jinan, China; [2]Department of Cardiology, Shandong Provincial Hospital, Shandong University, Jinan, China; [3]Department of Health Management, The Affiliated Hospital of Jining Medical University, Jining, China; [4]Central Laboratory, Linyi People's Hospital, Shandong University, Linyi, China and [5]Health Management Center, Shandong Provincial QianFoShan Hospital, Jinan, China. Correspondence: Professor F Xue, Department of Biostatistics, School of Public Health, Shandong University, PO Box 100, Wenhua Xi Road 44, Jinan 250012, China.
E-mail: xuefzh@sdu.edu.cn
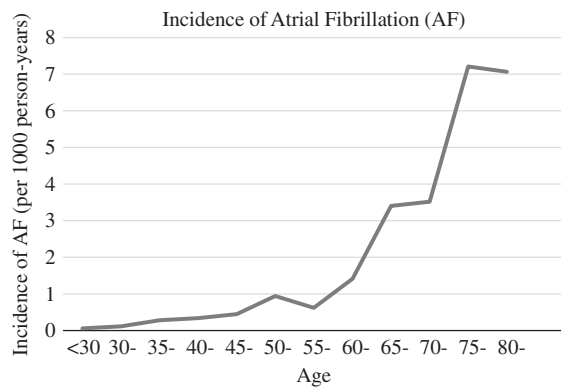
Age	No.	Number of cases	Follow-up (person-year)	Incidence (per 1000 person-years)
<30	21903	3	51243	0.06
30-	13812	4	34403	0.12
35-	13598	10	35793	0.28
40-	13123	12	35256	0.34
45-	9997	11	24675	0.45
50-	7221	18	19136	0.94
55-	5296	8	12922	0.62
60-	3195	12	8479	1.42
65-	2627	24	7046	3.41
70-	2491	22	6250	3.52
75-	1683	29	4022	7.21
80-85	676	10	1415	7.07

Figure 1. The incidence of atrial fibrillation by age in the original cohort ($n = 95\,622$).
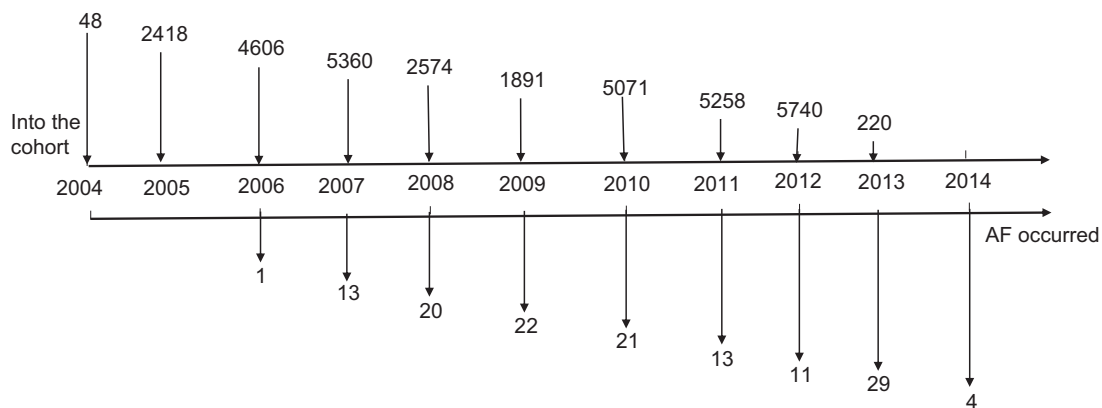

Figure 2. The framework of cohort including 33 186 participants aged 45–85 years for final analysis.

12-lead ECG in an annually regular clinical and laboratory examination, with atrial flutter and AF diagnosed by ECG as AF cases.

The study protocol was approved by ethics committee of School of Public Health, Shandong University. Written informed consent was obtained from all participants.

Measurements

All participants in the health check-up underwent a standardized routine health questionnaire, anthropometric measurements, laboratory tests and a rest 12-lead ECG. Behavioral factors such as smoking, drinking and disease history were obtained from self-reported questionnaire. Anthropometric measurements included weight, height and blood pressures. Body mass index was calculated by dividing weight in kilograms by height in meters squared. Systolic blood pressure (SBP), diastolic blood pressure (DBP) and heart rate were measured twice on the right arm after 5 min of rest in sitting position. We defined visit-to-visit variability (VVV) in blood pressure as the s.d. of blood pressure measurements during follow-up. Blood samples were taken after at least 12 h overnight fasting for laboratory tests. Red blood cell count, white blood cell count, hemoglobin, hematocrit, platelet count, fasting blood glucose, total cholesterol, triglycerides, high-density lipoprotein, low-density lipoprotein, serum creatinine, alanine transaminase, total bilirubin, uric acid, serum albumin, serum globulin, blood urea nitrogen were included in this analysis.

Smoking was coded as ever smoker and never smoker, and drinking was coded as current drinker and current non-drinker (never drinker and ex-drinker). History of coronary heart disease (CHD) was defined as having history of physician-diagnosed CHD. Hypertension was defined as having history of physician-diagnosed hypertension or a mean SBP $\geqslant$ 140 mm Hg, or a mean DBP $\geqslant$ 90 mm Hg.[12] Diabetes was defined as a history of physician-diagnosed diabetes or having a fasting plasma glucose $\geqslant$ 7.0 mmol/l (126 mg dl^{-1}).[13]

A rest 12-lead ECG was performed on all participates in the supine position with a paper speed of 25 mm s^{-1} and amplification of 0.1 mV mm^{-1}. All ECG reports were interpreted and coded according to Minnesota Code Classification system by two independent cardiologists (see Supplementary Information for details).[14] If there was any disagreement between two readers, a third expert would be involved to obtain a consensus interpretation.

Statistical analysis

The age-specific incidence was based on original cohort ($n = 95\,622$). For baseline characteristics of final analysis data ($n = 33\,186$), continuous variables were presented as mean (s.d.) and categorical variables were presented as frequency (percentage). We calculated the number of person-year as the sum of the follow-up times from the baseline to the onset of AF or the last health check-up. Differences in baseline characteristics between incident AF and non-AF were compared by t-test for continuous variables and χ^2-test for categorical variables.

Selection of eligible predictors of AF was based on prior reports. Information on self-reported questionnaire (smoking, drinking and history of CHD), anthropometric markers (body mass index, SBP, DBP, heart rate, VVV in SBP and VVV in DBP), biochemical examinations (fasting blood glucose, total cholesterol, triglyceride, high-density lipoprotein, low-density lipoprotein, uric acid, total bilirubin, serum creatinine, serum albumin, serum globulin, blood urea nitrogen, hemoglobin, white blood cell, red blood cell and platelet count), ECG markers and diagnosis of hypertension and diabetes were included as candidate predictors of AF. First, Cox proportional hazard models were used to estimate the age- and sex-adjusted hazard ratios of each risk factor. Then, significant risk factors were selected into multivariable cox proportional model and backward elimination method was used to deal with highly collinearities of covariates. For the validation of prediction model, receiver operating characteristic curve analysis was applied. The area under receiver

Table 1. Baseline characteristics of the cohort

	Total (n = 33 186)	AF (n = 134)	Non-AF (n = 33 042)	P-value
Age, years	56.69 (9.85)	66.34 (10.82)	56.65 (9.83)	< 0.0001
Male, no (%)	22 223 (66.96)	111 (82.84)	22 112 (66.90)	< 0.0001
Female, no (%)	10 963 (33.04)	23 (17.16)	10 940 (33.10)	
SBP, mm Hg	135.38 (20.75)	149.05 (21.38)	135.32 (20.73)	< 0.0001
DBP, mm Hg	81.07 (12.82)	82.68 (12.25)	81.06 (12.82)	0.1569
VVV of SBP (s.d.)	8.92 (6.73)	14.37 (10.01)	8.90 (6.70)	< 0.0001
VVV of DBP (s.d.)	5.88 (4.33)	8.99 (5.47)	5.86 (4.32)	< 0.0001
Heart rate (beat per minute)	73.60 (9.90)	72.64 (14.43)	73.61 (9.88)	0.5873
BMI, kg m^{-2}	24.34 (3.64)	23.92 (3.33)	24.35 (3.64)	0.2346
FBG, mmol l^{-1}	5.49 (0.94)	5.64 (1.00)	5.49 (0.94)	0.0752
TC, mmol l^{-1}	5.22 (0.92)	5.22 (0.87)	5.22 (0.92)	0.9283
HDL, mmol l^{-1}	1.41 (0.35)	1.42 (0.36)	1.41 (0.34)	0.6790
LDL, mmol l^{-1}	3.04 (0.71)	3 (0.74)	3.04 (0.71)	0.6090
TG, mmol l^{-1}	1.55 (0.89)	1.51 (0.97)	1.55 (0.89)	0.6962
ALT, mmol l^{-1}	21.52 (11.09)	19.43 (9.59)	21.53 (11.09)	0.0143
TBIL, mmol l^{-1}	11.55 (4.46)	11.27 (4.22)	11.55 (4.46)	0.6003
UA, mmol l^{-1}	318.47 (79.27)	343.15 (81.18)	318.39 (79.26)	0.0149
CREA, mmol l^{-1}	72.05 (16.52)	80.46 (16.12)	72.02 (16.51)	< 0.0001
ALB, mmol l^{-1}	45.75 (2.58)	45.62 (2.25)	45.75 (2.58)	0.5129
GLO, mmol l^{-1}	27.36 (3.99)	27.7 (4.20)	27.36 (3.99)	0.3430
BUN, mmol l^{-1}	5.24 (1.2)	5.65 (1.26)	5.24 (1.20)	0.0004
HB, g l^{-1}	146.45 (13.69)	148.52 (12.53)	146.45 (13.70)	0.1173
WBC, 10^9 l^{-1}	6.28 (1.48)	6.46 (1.67)	6.28 (1.48)	0.2022
HCT, %	43.7 (3.77)	44.28 (3.77)	43.7 (3.77)	0.1001
RBC, 10^{12} l^{-1}	4.79 (0.43)	4.8 (0.44)	4.79 (0.43)	0.7218
PLT, 10^9 l^{-1}	227.85 (51.75)	210.48 (54.82)	227.92 (51.72)	0.0004
History of CHD	2432 (7.33)	36 (26.87)	2396 (7.25)	< 0.0001
Hypertension	14 892 (44.87)	93 (69.40)	14 799 (44.77)	< 0.0001
Diabetes	4185 (12.61)	21 (15.67)	4164 (12.60)	0.2848
Smoking	16 648 (50.17)	66 (49.25)	16 582 (50.17)	0.8324
Drinking	14 889 (45.48)	54 (40.91)	14 835 (45.49)	0.2910
Major Q wave abnormalities (MC 1)	788 (2.37)	6 (4.48)	782 (2.37)	0.1091
QRS axis deviation (MC 2)	578 (1.74)	0	578 (1.75)	0.1225
Left high-amplitude R waves	959 (2.89)	13 (9.70)	946 (2.86)	< 0.0001
Right high-amplitude R waves	148 (0.45)	1 (0.75)	147 (0.44)	0.6011
Major ST–T abnormalities (MC 4)	2810 (8.47)	14 (10.45)	2796 (8.46)	0.4093
T-wave items (MC 5)	3663 (11.04)	21 (15.67)	3642 (11.02)	0.0863
A–V conduction defect (MC 6)	620 (1.87)	7 (5.22)	613 (1.85)	0.0040
Ventricular conduction defect (MC 7)	1520 (4.58)	8 (5.97)	1512 (4.57)	0.4406
Arrhythmias (MC 8)	4107 (12.38)	33 (24.63)	4074 (12.33)	< 0.0001
Premature beats (MC 8-1)	1009 (3.04)	16 (11.94)	993 (3.00)	< 0.0001
ST segment elevation and else (MC 9)	1156 (3.48)	6 (4.48)	1150 (3.48)	0.5294

Abbreviations: ALB, serum albumin; ALT, alanine transaminase; BMI, body mass index; BUN, blood urea nitrogen; CHD, coronary heart disease; CREA, serum creatinine; DBP, diastolic blood pressure; FBG, fasting blood glucose; GLO, serum globulin; HB, hemoglobin; HCT, hematocrit; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MC, Minnesota Code; PLT, platelet count; RBC, red blood cell count; SBP, systolic blood pressure; TBIL, total bilirubin; TC, total cholesterol; TG, triglyceride; UA, uric acid; VVV, visit-to-visit variability; WBC, white blood cell count. Values are mean (s.d.) or frequency (percentage).

operating characteristic curve (AUC) was used to estimate the discriminative accuracy of AF prediction model. The predicted 2-year risk of each individual was estimated and the ratio of observed risk with predicted risk as well as Hosmer Lemeshow test were used to evaluate the calibration.

P-values < 0.05 were considered statistically significant. All data analyses were performed by SAS statistical software version 9.4 (SAS Institute, Inc., Cary, NC, USA).

RESULTS

Figure 1 demonstrates the incidence of AF by different age groups for original cohort (n = 95 622). As the incidence of AF was in a very low level before 45 years (< 0.40 per 1000 person-years), the final analysis included participants aged 45–85 years for developing AF prediction model (n = 33 186, see Supplementary Table S1 for the characteristics of original health check-up database).

Table 1 presents baseline characteristics of the cohort (n = 33 186). Among the 33 186 participants aged 45–85 years, 66.96% were men and the mean age was 56.69 ± 9.85 years. A total of 2.73% had a history of CHD, 44.87% had hypertension, 12.61% had diabetes, 34.03% had dyslipidemia and 50.17% were ever smokers, whereas 45.48% were current drinkers. The means and s.d. of biomedical biomarkers are presented in Table 1. Age, sex, SBP, VVV in SBP, VVV in DBP, alanine transaminase, uric acid, serum creatinine, blood urea nitrogen, platelet count, history of CHD, hypertension, left high-amplitude waves, A–V conduction defect and arrhythmias including premature beats were significantly different between incident AF and non-AF groups (P < 0.05).

Among these 33 186 participants, during the 83 945 person-years of follow-up (median = 2.6 years, see Figure 2), 134 AF cases were observed and the crude incidence density of AF was 1.60 per 1000 person-years for individuals aged 45–85 years.

Table 2 shows the age- and sex-adjusted hazard ratios of AF. Several factors were associated with incident AF at P < 0.05 level after adjusting age and sex, including age, sex, history of CHD, hypertension, SBP, DBP, VVV in SBP, VVV in DBP, left high-amplitude R waves and premature beats. We observed a higher risk of incident AF in men, older subjects, those with a history of

Table 2. Age- and sex-adjusted Cox model for prediction of atrial fibrillation

Variable	Hazard ratio	95% CI		P-value
Age	1.09	1.07	1.11	< 0.0001
Female vs male	0.48	0.31	0.76	0.0016
SBP	1.34	1.14	1.59	0.0005
DBP	1.28	1.08	1.51	0.0048
VVV of SBP	1.05	1.04	1.07	< 0.0001
VVV of DBP	1.10	1.07	1.13	< 0.0001
Heart rate	1.00	0.97	1.02	0.7780
BMI [a]	0.94	0.77	1.14	0.5200
GLU [a]	1.04	0.89	1.22	0.5844
TC [a]	0.99	0.83	1.18	0.9006
HDL [a]	1.13	0.91	1.39	0.2749
LDL [a]	0.95	0.77	1.18	0.6471
TG [a]	1.00	0.83	1.20	0.9964
ALT [a]	0.89	0.72	1.10	0.2800
TBIL [a]	0.99	0.77	1.29	0.9632
UA [a]	1.28	0.98	1.66	0.0667
CREA [a]	1.11	0.89	1.38	0.3626
ALB [a]	1.02	0.84	1.22	0.8705
GLO [a]	0.98	0.82	1.17	0.8351
BUN [a]	1.15	0.96	1.39	0.1401
HB [a]	1.17	0.92	1.49	0.1996
WBC [a]	1.04	0.86	1.26	0.6779
HCT [a]	1.08	0.87	1.35	0.4732
RBC [a]	1.04	0.83	1.29	0.7539
PLT [a]	0.90	0.73	1.10	0.2948
History of CHDs	1.98	1.32	2.97	0.0010
Hypertension	1.64	1.12	2.40	0.0106
Diabetes	0.91	0.57	1.44	0.6739
Smoking	1.10	0.77	1.58	0.6024
Drinking	0.97	0.67	1.41	0.8713
ECG markers				
Major Q wave abnormalities (MC 1)	0.95	0.42	2.15	0.8920
QRS axis deviation (MC 2)	-	-	-	-
Left high-amplitude R waves (MC 3)	2.97	1.67	5.28	0.0002
Right high-amplitude R waves (MC 3)	1.40	0.20	10.04	0.7358
Major ST–T abnormalities (MC 4)	0.87	0.50	1.52	0.6257
T-wave items (MC 5)	1.39	0.87	2.21	0.1737
A–V conduction defect (MC 6)	1.67	0.78	3.60	0.1879
Ventricular conduction defect (MC 7)	0.76	0.37	1.57	0.4619
Arrhythmias (MC 8)	1.70	1.15	2.53	0.0083
Premature beats (MC 8–1)	2.38	1.40	4.05	0.0014
ST segment elevation and else (MC 9)	0.82	0.36	1.87	0.6361

Abbreviations: ALB, serum albumin; ALT, alanine transaminase; BMI, body mass index; BUN, blood urea nitrogen; CHD, coronary heart disease; CREA, serum creatinine; DBP, diastolic blood pressure; GLO, serum globulin; HB, hemoglobin; HCT, hematocrit; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MC, Minnesota Code; PLT, platelet count; RBC, red blood cell count; SBP, systolic blood pressure; TBIL, total bilirubin; TC, total cholesterol; TG, triglyceride; UA, uric acid; VVV, visit-to-visit variability; WBC, white blood cell count. [a]Hazard ratios are per 1 s.d. increase.

Table 3. Prediction models of atrial fibrillation

Variable	Simple model		ECG model		VVV model	
	HR	P-value	HR	P-value	HR	P-value
Age	1.074	< 0.001	1.071	< 0.001	1.069	< 0.001
Sex (male vs female)	2.099	0.001	1.987	0.003	2.065	0.002
History of CHD	1.883	0.002	1.840	0.003	1.844	0.003
Hypertension	1.566	0.022	1.502	0.038		
VVV of SBP					1.033	0.001
VVV of DBP					1.069	< 0.001
Left high-amplitude R waves			2.618	0.001	2.338	0.004
Premature beats			2.359	0.002	2.378	0.001

Abbreviations: CHD, coronary heart disease; HR, hazard ratio; VVV, visit-to-visit variability.

age and sex, history of CHD, hypertension, SBP, DBP, left high-amplitude R waves and premature beats as candidate predictors, hypertension was selected into the model, together with age, sex, history of CHD, left high-amplitude R waves and premature beats (ECG model). Then, we added VVV in SBP and VVV in DBP as candidate predictors and also using backward elimination, age, sex, history of CHD, VVV in SBP, VVV in DBP, left high-amplitude R waves and premature beats were independent predictors of AF and finally entered into the AF prediction model (VVV model). In the VVV model, the hazard ratios of age, male, history of CHD, VVV in SBP, VVV in DBP, left high-amplitude R waves and premature beats were 1.069, 2.065, 1.844, 1.033, 1.069, 2.338 and 2.378, respectively. Supplementary Figures S1–S3 present the score system of the simple, ECG and VVV models, respectively.

The AUC of three prediction models were 78%, 80% and 82%, respectively, and after 10-fold cross-validation, the AUC became 77%, 78% and 79% for predicting risk of AF. The AUC showed good discriminatory accuracy for the prediction of AF risk. The calibration (ratio of observed risk with predicted risk) of these three models were 1.001 (P for Hosmer Lemeshow test = 0.8032) for simple model, 1.002 (P = 0.8673) for ECG model and 1.004 (P = 0.5396) for VVV model.

DISCUSSION

Our analysis of the Shandong multi-center health check-up longitudinal cohort showed the AF incidence density of 1.60 per 1000 person-years. For incidence of AF in China, previously, only one cohort study was conducted in China to explore the incidence of AF. It revealed that the incidence of AF in China was 0.5 and 0.7 per 1000 person-years for participants aged ⩾20 and ⩾50 years, respectively, by using 2001 medical insurance database.[8] In our study, we observed 134 incident AF cases during 83 945 person-years follow-up in 2004–2014, with AF incidence density of 1.60 per 1000 person-years. As data on incidence of AF in China was limited, this study was valuable for providing updated AF incidence data for urban Han Chinese population based on health check-up database.

A series of behavioral, clinical, biochemical and ECG markers had been identified as risk factors of incident AF, such as markers associated with atrial stress, inflammation, kidney function, CHD and so on.[15–18] In this study, we tested all behavioral, anthropometric, biochemical and ECG factors available in the routine health check-up. After adjusting age and sex, history of CHD, hypertension, SBP, DBP, VVV in SBP, VVV in DBP, left high-amplitude R waves and premature beats were significantly associated with incident AF. The associations of age, sex, CHD and hypertension

CHD, higher baseline blood pressure, higher VVV in blood pressures and individuals with ECG features of left high-amplitude R waves and premature beats.

The hazard ratios and P-values corresponding to the prediction models of AF were presented in Table 3. We first developed a simple prediction model by using age, sex, history of CHD and hypertension as predictors (Simple model). Then, by using backward elimination of multivariable Cox regression, with

with risk of AF were well established and age, sex, history of CHD and hypertension had been included in various prediction models to estimate risk of incident AF.[19,20] Beyond hypertension, the VVV in blood pressure has emerged as a significant predictor of overall mortality and cardiovascular events, and has proven to be correlated with subclinical markers of vascular dysfunction.[21,22] For ECG markers, left high-amplitude R waves indicating left ventricular hypertrophy was a predictor for risk of several adverse cardiovascular events.[23] Atherosclerosis Risk in Communities study also included ECG-defined left ventricular hypertrophy as one predictor of incident AF.[24] Premature beats tested by ECG was a new predictor of AF found by our study. Premature beat was a common type of arrhythmias in clinical practice and had been identified to be an independent predictor of AF in patients with cryptogenic stroke.[25] By adopting multivariable Cox regression, we developed three prediction models, with age, sex, history of CHD and hypertension as independent predictors (simple model); with age, sex, history of CHD, hypertension, left high-amplitude R waves and premature beats as independent predictors (ECG model); and with age, sex, history of CHD, left high-amplitude R waves and premature beats, VVV in SBP and VVV in DBP as independent predictors (VVV model). All the predictors were independently associated with the risk of AF and finally included in the prediction models of AF.

To the best of our knowledge, this study was the first large prospective cohort study conducted in Asian countries, which aimed to develop prediction model of incident AF. Three predictive models had been developed to predict AF risk derived from US population. The first prediction model was Framingham Heart Study score, which was derived from White Americans, getting an AUC of 0.78 for predicting AF.[9] Then, it had been evaluated in White and non-White US population and developed the second score.[24] In 2013, three cohort studies in the United States were pooled and got the Cohorts for Heart and Aging Research in Genomic Epidemiology-AF model, with AUC of 0.747.[26] These prediction models were all based on US population and none had been validated in Asian population. It was well-known that prediction models developed in certain ethnic groups tend to overestimate or underestimate the risk in other ethnic groups.[27–30] As the epidemic of AF in China is quite different from the United States, we have reasons to suspect the applicability of these three AF prediction models in Chinese population. As there was no risk prediction model derived from Chinese population, in this study we aimed to develop risk prediction models by using Chinese health check-up database. The calibration of our models were good, with the ratio of observed risk with predicted risk range from 1.001 to 1.004 ($P > 0.05$). The AUC of our models were 78, 80, and 82%, indicating that the discriminatory accuracy was good in predicting incident AF.

There were several limitations to be considered. First, the participants of this cohort were urban inhabitants occupational or retired in Shandong province and only participants having at least two health check-ups were included in the cohort; the applicability of our results should be validated by further studies. Second, the follow-up duration was relatively short, so we could not predict long-term risk of incident AF. Third, lifestyle habits such as physical activity and novel biomarkers such as B-type natriuretic peptide were not measured in routine health check-up; therefore, we could not explore the predictive ability of these predictors in this study. Fourth, information on antihypertensive treatments was not available in our study, so we could not compare the estimated AF risks of our prediction model with previous developed models.

The strengths of this study included its large sample size, prospective design and abundant potential markers for predicting AF risk. In addiiition, our study was the first cohort study conducted in China aimed to develop prediction model to estimate risk of AF and also provided updated AF incidence data for urban Han Chinese population based on health check-up database.

In conclusion, our findings identified predictors of incident AF and developed risk prediction models for AF, which adequately predicted incident AF in health check-up population. The models could help to identify high-risk individuals and define targeted population for primary prevention of AF in clinical and public health practice.

What is known about topic?
- In the last 20 years, the prevalence and incidence of AF was increasing worldwide. The only cohort study in China revealed that the incidence of AF in China was 0.7 per 1000 person-years for participants aged ⩾ 50 years by using 2001 medical insurance database.
- Prediction models of AF have been developed from the US population and the applicability of these models in Chinese population is uncertain.

What this study adds?
- This study provided updated AF incidence for urban Han Chinese population based on health check-up database.
- We have identified predictors of incident AF and developed prediction models for AF with variables readily available in routine health check-up.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1 Zoni-Berisso M, Lercari F, Carazza T, Domenicucci S. Epidemiology of atrial fibrillation: European perspective. *Clin Epidemiol* 2014; **6**: 213–220.

2 Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation* 2014; **129**(8): 837–847.

3 Murphy NF, Simpson CR, Jhund PS, Stewart S, Kirkpatrick M, Chalmers J et al. A national survey of the prevalence, incidence, primary care burden and treatment of atrial fibrillation in Scotland. *Heart* 2007; **93**(5): 606–612.

4 Furberg CD, Psaty BM, Manolio TA, Gardin JM, Smith VE, Rautaharju PM. Prevalence of atrial fibrillation in elderly subjects (the Cardiovascular Health Study). *Am J Cardiol* 1994; **74**(3): 236–241.

5 Heeringa J, van der Kuip DA, Hofman A, Kors JA, van Herpen G, Stricker BH et al. Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. *Eur Heart J* 2006; **27**(8): 949–953.

6 Wheeldon NM, Tayler DI, Anagnostou E, Cook D, Wales C, Oakley GD. Screening for atrial fibrillation in primary care. *Heart* 1998; **79**(1): 50–55.

7 Chei CL, Raman P, Ching CK, Yin ZX, Shi XM, Zeng Y et al. Prevalence and risk factors of atrial fibrillation in Chinese elderly: results from the chinese longitudinal healthy longevity survey. *Chinese Med J* 2015; **128**(18): 2426–2432.

8 Guo Y, Tian Y, Wang H, Si Q, Wang Y, Lip GY. Prevalence, incidence, and lifetime risk of atrial fibrillation in China: new insights into the global burden of atrial fibrillation. *Chest* 2015; **147**(1): 109–119.

9 Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB et al. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *Lancet* 2009; **373**(9665): 739–745.

10 Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc* 2013; **2**(2): e000102.

11 Kolek MJ, Graves AJ, Xu M, Bian A, Teixeira PL, Shoemaker MB et al. Evaluation of a prediction model for the development of atrial fibrillation in a repository of electronic medical records. JAMA Cardiol 2016; 1(9): 1007–1013.

12 Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL Jr. et al. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension 2003; 42(6): 1206–1252.

13 Yuan X, Liu T, Wu L, Zou ZY, Li C. Validity of self-reported diabetes among middle-aged and older Chinese adults: the China Health and Retirement Longitudinal Study. BMJ Open 2015; 5(4): e006633.

14 Mb RJP, Crow RS, Zhang Z-M. The Minnesota Code Manual of Electrocardiographic Findings. Springer: London, UK, 2010.

15 O'Neal WT, Venkatesh S, Broughton ST, Griffin WF, Soliman EZ. Biomarkers and the prediction of atrial fibrillation: state of the art. Vasc Health Risk Manag 2016; 12: 297–303.

16 Vilchez JA, Roldan V, Hernandez-Romero D, Valdes M, Lip GY, Marin F. Biomarkers in atrial fibrillation: an overview. Int J Clin Pract 2014; 68(4): 434–443.

17 Lau YF, Yiu KH, Siu CW, Tse HF. Hypertension and atrial fibrillation: epidemiology, pathophysiology and therapeutic implications. J Hum Hypertens 2012; 26(10): 563–569.

18 Milan A, Caserta MA, Dematteis A, Naso D, Pertusio A, Magnino C et al. Blood pressure levels, left ventricular mass and function are correlated with left atrial volume in mild to moderate hypertensive patients. J Hum Hypertens 2009; 23(11): 743–750.

19 Vlachos K, Letsas KP, Korantzopoulos P, Liu T, Georgopoulos S, Bakalakos A et al. Prediction of atrial fibrillation development and progression: current perspectives. World J Cardiol 2016; 8(3): 267–276.

20 Anumonwo JM, Kalifa J. Risk factors and genetics of atrial fibrillation. Cardiol Clin 2014; 32(4): 485–494.

21 Sabayan B, Wijsman LW, Foster-Dingley JC, Stott DJ, Ford I, Buckley BM et al. Association of visit-to-visit variability in blood pressure with cognitive function in old age: prospective cohort study. BMJ (Clin Res Ed) 2013; 347: f4600.

22 Gavriilaki E, Gkaliagkousi E, Douma S. Visit-to-visit blood pressure variability: more to come. J Clin Hypertens (Greenwich) 2015; 17(2): 116–117.

23 Rautaharju PM, Soliman EZ. Electrocardiographic left ventricular hypertrophy and the risk of adverse cardiovascular events: a critical appraisal. J Electrocardiol 2014; 47(5): 649–654.

24 Chamberlain AM, Agarwal SK, Folsom AR, Soliman EZ, Chambless LE, Crow R et al. A clinical risk score for atrial fibrillation in a biracial prospective cohort (from the Atherosclerosis Risk in Communities [ARIC] study). Am J Cardiol 2011; 107(1): 85–91.

25 Gladstone DJ, Dorian P, Spring M, Panzov V, Mamdani M, Healey JS et al. Atrial premature beats predict atrial fibrillation in cryptogenic stroke: results from the EMBRACE trial. Stroke 2015; 46(4): 936–941.

26 Alonso A, Roetker NS, Soliman EZ, Chen LY, Greenland P, Heckbert SR. Prediction of atrial fibrillation in a racially diverse cohort: the Multi-Ethnic Study of Atherosclerosis (MESA). J Am Heart Assoc 2016; 5(2): e003077.

27 D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. Jama 2001; 286(2): 180–187.

28 Thomsen TF, McGee D, Davidsen M, Jorgensen T. A cross-validation of risk-scores for coronary heart disease mortality based on data from the Glostrup Population Studies and Framingham Heart Study. Int J Epidemiol 2002; 31(4): 817–822.

29 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(21): 2591–2599.

30 Hense HW, Schulte H, Lowel H, Assmann G, Keil U. Framingham risk function overestimates risk of coronary heart disease in men and women from Germany--results from the MONICA Augsburg and the PROCAM cohorts. Eur Heart J 2003; 24(10): 937–945.