A simple clinical risk score (C_2HEST) for predicting incident atrial fibrillation in Asian subjects

Li, Yan-guang; Pastori, Daniele; Farcomeni, Alessio; Yang, Pil-Sung; Jang, Eunsun; Joung, Boyoung; Wang, Yu-Tang; Guo, Yu-Tao; Lip, Gregory Y. H.

DOI:
10.1016/j.chest.2018.09.011

Document Version
Peer reviewed version

Citation for published version (Harvard):
Li, Y, Pastori, D, Farcomeni, A, Yang, P-S, Jang, E, Joung, B, Wang, Y-T, Guo, Y-T & Lip, GYH 2018, 'A simple clinical risk score (C_2HEST) for predicting incident atrial fibrillation in Asian subjects', *Chest*. https://doi.org/10.1016/j.chest.2018.09.011

Accepted Manuscript

A simple clinical risk score (C_2HEST) for predicting incident atrial fibrillation in Asian subjects: Derivation in 471,446 Chinese subjects, with internal validation and external application in 451,199 Korean subjects

Yan-Guang Li, PhD, Daniele Pastori, MD, PhD, Alessio Farcomeni, PhD, Pil-Sung Yang, MD, Eunsun Jang, MD, Boyoung Joung, MD, PhD, Yu-Tang Wang, MD, PhD, Yu-Tao Guo, MD, PhD, Gregory Y.H. Lip, MD

PII: S0012-3692(18)32492-9

DOI: 10.1016/j.chest.2018.09.011

Reference: CHEST 1971

To appear in: *CHEST*

Received Date: 29 May 2018

Revised Date: 21 August 2018

Accepted Date: 6 September 2018

Please cite this article as: Li YG, Pastori D, Farcomeni A, Yang PS, Jang E, Joung B, Wang YT, Guo YT, Lip GYH, A simple clinical risk score (C_2HEST) for predicting incident atrial fibrillation in Asian subjects: Derivation in 471,446 Chinese subjects, with internal validation and external application in 451,199 Korean subjects, *CHEST* (2018), doi: https://doi.org/10.1016/j.chest.2018.09.011.

A simple clinical risk score (C_2HEST) for predicting incident atrial fibrillation in Asian subjects: Derivation in 471,446 Chinese subjects, with internal validation and external application in 451,199 Korean subjects

Yan-Guang Li, PhD (1,2) *, Daniele Pastori, MD, PhD (1,3)*, Alessio Farcomeni, PhD (4), Pil-Sung Yang, MD (5), Eunsun Jang, MD (6), Boyoung Joung, MD, PhD (6), Yu-Tang Wang, MD, PhD (2)**, Yu-Tao Guo, MD, PhD (2)**, Gregory Y.H. Lip, MD (1,2,6,7,8)**

*equal contribution; ** Joint senior authors

(1) Institute of Cardiovascular Sciences, University of Birmingham, Birmingham, United Kingdom; (2) Department of Cardiology, Chinese PLA General Hospital, Chinese PLA Medical School, Beijing, China; (3) I Clinica Medica, Atherothrombosis Center, Department of Internal Medicine and Medical Specialties, Sapienza University, Rome, Italy; (4) Department of Public Health and Infectious Diseases, Sapienza University, Rome, Italy; (5) Department of Cardiology, CHA Bundang Medical Center, CHA University, Seongnam, Republic of Korea; (6) Division of Cardiology, Department of Internal Medicine, Yonsei University Health System, Seoul, Republic of Korea; (7) Aalborg Thrombosis Research Unit, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark; (8) Liverpool Centre for Cardiovascular Sciences, University of Liverpool, United Kingdom.

Corresponding Authors

Professor Gregory Y.H. Lip

Institute of Cardiovascular Sciences, University of Birmingham, Birmingham, England (UK)

Email: g.y.h.lip@bham.ac.uk

Professor Yu-Tao Guo

Chinese PLA General Hospital, Chinese PLA Medical School, Department of Cardiology, Beijing, China; Email: dor_guoyt@hotmail.com

Source of Funding

This work was supported by China Scholarship Council (201708110232).

Declarations of interest

None directly related to this paper.

GYHL: Consultant for Bayer/Janssen, BMS/Pfizer, Medtronic, Boehringer Ingelheim, Novartis, Verseon and Daiichi-Sankyo. Speaker for Bayer, BMS/Pfizer, Medtronic, Boehringer Ingelheim, and Daiichi-Sankyo. No fees are directly received personally.

Other authors: None declared.

Keywords: atrial fibrillation; risk factors; risk score; prediction model; cohort study; Asian.

Abbreviations list

AF = Atrial fibrillation

ARIC = Atherosclerosis Risk In Communities Study

AUC = area under curve

CAD = coronary artery disease

CHARGE-AF = Cohorts for Heart and Aging Research in Genomic Epidemiology-Atrial Fibrillation

CI = confidence interval

COPD = chronic obstructive pulmonary disease

FHS = Framingham Heart Study

HF = heart failure

HR = hazard ratio

ICD = International Classification of Disease

IR = incidence rate

IS = ischemic stroke

NHIS = National Health Insurance Service

PLA = People's Liberation Army

ROC = receiver operating characteristic

SD = standard deviation

SHD = structural heart disease

Abstract

Background: The incidence of atrial fibrillation (AF) is increasing, conferring a major healthcare issue in Asia. No risk score for predicting incident AF has been specifically developed in Asian subjects. Our aim was to investigate risk factors for incident AF in Asian subjects and to combine them into a simple clinical risk score.

Methods: Risk factors for incident AF were analyzed in 471,446 subjects from the Chinese Yunnan Insurance Database (internal derivation cohort), and then combined into a simple clinical risk score. External application of the new score was performed in 451,199 subjects from the Korean National Health Insurance Service (external cohort).

Results: In the internal cohort, structural heart disease (SHD), heart failure (HF), age≥75 years, coronary artery disease (CAD), hyperthyroidism, chronic obstructive pulmonary disease (COPD) and hypertension were associated with incident AF. Given the low prevalence and the strong association of SHD with incident AF (hazard ratio: 26.07 [18.22-37.30], p<0.001), these patients should be independently considered as 'high-risk' for AF and were excluded from the analysis. The remaining predictors were combined into the new simple C_2HEST score: C_2: CAD / COPD (*1 point each*); H: Hypertension; E: Elderly (Age≥75, *doubled*); S: Systolic HF (*doubled*); T: Thyroid disease (hyperthyroidism). The C_2HEST score showed good discrimination with AUC 0.75 [0.73-0.77] and had good calibration (p=0.774). The score was internally validated by bootstrap sampling procedure, giving an AUC of 0.75 [0.73-0.77]. External application gave an AUC of 0.65 [0.65-0.66]. The C_2HEST score was superior to $CHADS_2$ and CHA_2DS_2VASc scores in both cohorts in predicting incident AF.

Conclusions: We have developed and validated the C_2HEST score as a simple clinical tool to assess the individual risk of developing AF in the Asian population *without SHD*.

Introduction

Atrial fibrillation (AF) is the most prevalent arrhythmia worldwide and is associated with an increased risk of ischemic stroke (IS), heart failure (HF), coronary artery disease (CAD) and mortality[1-3]. The 2010 Global Burden of Disease Study demonstrated that the worldwide age-adjusted prevalence of AF is 596 per 100,000 men and 373 per 100,000 women[4]. The incidence of AF increases dramatically with the development of incident risk factors for AF, such as ageing, hypertension, HF, CAD and chronic obstructive pulmonary disease (COPD). Indeed, this arrhythmia confers a major healthcare burden in Asia. For example, Japan will have 1 million AF patients, and China will have 9 million AF patients by 2050[5,6].

Considering that many patients with incident AF are asymptomatic, a considerable number of patients are only diagnosed with AF when presenting with the AF-related complications. Because of this under-diagnosis, the actual prevalence of AF could be considerably higher[7,8]. Hence, risk evaluation and early identification through screening would be important to allow the targeting of early prevention strategies and improve prognosis[9,10]. Therefore, effective and cost-effective strategies should focus on identifying patients who are at higher risk of incident AF.

Several clinical risk scores have been developed to predict incident AF, including the Framingham Heart Study (FHS) score[11], the Atherosclerosis Risk In Communities Study (ARIC) score[12] and the Cohorts for Heart and Aging Research in Genomic Epidemiology-Atrial Fibrillation (CHARGE-AF) score[13]. These three scoring systems were all derived from large follow-up cohorts and have good

predictive ability for incident AF[11-13]. However, they require many instrumental and laboratory variables to be calculated, being generally too complicated for everyday clinical application. Moreover, many of the previous scores were derived in Western populations, and a predictive score for incident AF in an Asian population has been never been developed. This is relevant, given that risk factors for AF may be different between Western and Asian populations[14,15].

Our primary objective was to analyze risk factors for incident AF in a large cohort of Asian subjects, which were combined into a simple clinical risk stratification score, which would be easy and practical to apply in everyday practice.

Methods

Derivation cohort

Details of the derivation database, the Yunnan Medical Insurance Database, have been described previously[16]. In brief, this is a medical insurance database in Yunnan Province, China, including >10 million individuals from January 1, 2001, through December 30, 2012[16]. This medical insurance scheme covers urban residents in Yunnan Province, located in the far southwest of China, spanning approximately 394,000 km^2 and with a population of 46.3 million (2011 population statistics), representing 3% of the total Chinese population. It was part of governmental medical insurance plan, ensuring all participants had a permanent and personal registration number, through which the information of medical history, drugs, and mortality data recorded could be collected. All medical information of participants was obtained from local 2A- and 3A-grade hospitals which guaranteed the reliability of records regarding any medical service. Random sampling was performed on the enrolled individuals biennially, based on the periods of 2001-2002, 2003-2004, 2005-2006, 2007-2008, 2009-2010, 2011-2012. A total of 1,228,539 people was selected. After excluding people with incomplete data (n=2,611 cases), readmission (n=754,582) and prevalent AF at baseline (n=30), 471,446 cases were entered in the final analysis[16].

This derivation study was approved by the medical ethics committee of Chinese PLA (People's Liberation Army) General Hospital (Approval No. 13BJZ40). The definition of AF and other risk factors were according to the International Classification of Diseases (ICD), Ninth Revision/10th revision (Supplementary data). AF was defined based on an ECG or Holter recording. The inclusion criteria for an AF case were limited to inpatients, with AF diagnosis confirmed on admission and

discharge. The accuracy of AF diagnosis was validated previously using sensitivity analysis[16]. Patients with paroxysmal AF were also included in the incident analysis. Patients with rheumatic heart disease were those who suffered from rheumatic fever and concomitant valvular heart disease (ie. mitral stenosis). Patients with structural heart disease (SHD) included those patients with rheumatic heart disease and dilated cardiomyopathy.

External application

The external cohort study was based on the Korean National Health Insurance Service (NHIS)-Health Screening (HealS) cohort released in 2015, including subjects who participated in health screening programs provided by the NHIS in the Republic of Korea in 2002 and 2003. This external cohort consisted of 514,764 Koreans subjects aged 40 to 80 years, who comprised a 10% simple random sample of all health screening participants. A total of 55.3% of the participants lived in non-metropolitan areas, which covers some urban areas and all rural areas. The follow up started from 2002 through 2013, with a mean duration of 87.3±17.6 months. Detailed information and profile of the NHIS-HealS Cohort has been described in a previous report[17]. The baseline demography for this external cohort is summarized in online Supplementary Table S1.

This study was approved by the Institutional Review Board of Yonsei University Health System, and the informed consents were waived. We excluded subjects by following criteria: (i) those diagnosed with AF before conducting the health check-ups (n=5,019), and (ii) those with valvular heart disease (n=935; with a diagnosis of mitral stenosis [ICD-10: I05.0, I05.2, and I34.2] or prosthetic heart valves [ICD-10: Z95.2–Z95.4], and insurance claims for valve replacement or valvuloplasty), and (iii)

those with cardiomyopathy (n=357) [ICD-10: I42.0-I42.2]. Finally, a population cohort of 451,199 patients was included for this analysis. Patients were defined as having new-onset AF as follows: i) if there was a discharge diagnosis of AF for inpatients with no prior history of AF'; or ii) when an episode of AF was detected in an outpatient and confirmed by a specialist. This strategy of AF diagnosis has previously been validated in the NHIS database with a positive predictive value of 94.1%[18,19].

Statistical analysis

Continuous variables were reported as mean with standard deviation (SD) and categorical variables as observed number of patients (percentage). Annual incident rates were defined as the number of patients with incident AF divided by the number of person-years free of AF within a 1-year period.

Multivariable Cox regression analysis was used for identifying independent risk factors for incident AF over 11 years of follow-up (mean 4.1 (SD = 3.5) years) in the derivation cohort. The new predictive score was obtained with a stepwise model selection procedure based on Akaike information criterion. The pool of variables was also confirmed by a removal approach with p-value threshold of 10%. Risk factors in the final model were tested for interaction, but no further improvement was possible by including interactions.

The score was evaluated by means of time-dependent (at follow-up time 11 years) area under the curve (AUC) for the receiver operating characteristic (ROC) curve for predicting incident AF. Internal validation was obtained by means of 1000 bootstrap replicates and calibration was assessed by means of Gronnesby and Borgan test[20]. Performance of the external application was also evaluated as AUC in the whole Korean cohort.

The Kaplan-Meier curves were computed to present the survival rates free from AF during follow-up, after dividing patients into 3 groups (low, medium, and high risk) according to the new score.

Given that the $CHADS_2$, CHA_2DS_2VASc and HATCH scores have been previously reported to be associated with incident AF[21-23], we also performed an exploratory analysis to investigate how the C_2HEST score compares against these scores.

All tests were two-tailed, a value of $p<0.05$ was considered as statistically significant. Analyses were performed using SPSS Statistics, version 23.0 (IBM) and R version 3.4.2.

Results

Baseline characteristics of the derivation cohort are presented in Table 1. A total of 471,446 subjects entered the final analysis, in which 921 subjects developed AF during 11 years of follow-up (mean 4.1, SD = 3.5 years) with an incidence of 0.5 per 1000 person-years. Compared with subjects without AF, the patients who developed AF were older (P<0.001), more frequently male (P<0.001), with higher rates of hypertension (p<0.001), CAD (p<0.001), COPD (p<0.001), previous IS, renal dysfunction, hyperthyroidism, HF (p<0.001) and SHD (p<0.001).

Risk factors for incident AF and score development

Univariate Cox regression analysis is shown in Table 2. On multivariable analysis, SHD, HF, age ≥75 years, CAD, hyperthyroidism, COPD, and hypertension were independent risk factors for incident AF (Table 2). Given the low prevalence and very strong association of SHD with incident AF (Hazard Ratio: 26.07, 95% Confidence Interval [CI] 18.22-37.30, p<0.001), this subgroup of patients should be considered as being at (very) 'high-risk' for incident AF independently from other risk factors. Thus, this variable was excluded from the score. The HR of other risk factors significantly associated with incident AF did not change after removal of SHD from the multivariable model (not shown).

We combined these independent risk factors into the new simple C_2HEST score (Table 3): C_2: CAD / COPD (*1 point each*); H: Hypertension (*1 point*); E: Elderly (Age ≥ 75, *2 points*); S: Systolic HF (*2 points*); T: Thyroid disease (*1 point*). Total score ranged from 0 to 8 points.

The score showed a good discrimination with AUC 0.750 (95%CI 0.730-0.771, Figure 1a) and a good calibration (p=0.774). The score was then internally validated by bootstrap sampling procedure, which gave an AUC of 0.749 (95% CI 0.729-0.769).

When applied to the external cohort, the score showed moderate discrimination with AUC 0.654 (95%CI 0.649-0.659, Figure 1b).

Incident rates of AF and the C_2HEST score

Table 4 shows incidence rates (IR) and HRs at each point of the C_2HEST score. We divided patients in the derivation cohort into three groups according to the C_2HEST score: low (0-1 points, IR 0.34%/year), medium (2-3 points, IR 2.60%/year) and high-risk (>3 points, IR 15.98%/year). Kaplan-Meier curves for risk categories (Figure 2) showed an increased risk of AF across the three groups (log-rank test p<0.001).

Comparison of the C_2HEST score with the $CHADS_2$, CHA_2DS_2VASc and HATCH scores

Based on time-dependent prognostic models, the C_2HEST score had significantly better predictivity for incident AF, as compared to the $CHADS_2$, CHA_2DS_2VASc and HATCH scores in the derivation cohort from China, as follows:

- AUC $CHADS_2$: 0.632, 95%CI: 0.604-0.660 (p<0.001 vs. C_2HEST).
- AUC CHA_2DS_2VASc: 0.687, 95%CI: 0.659-0.716 (p<0.001 vs. C_2HEST).
- AUC HATCH: 0.633, 95%CI: 0.598-0.667 (p<0.001 vs. C_2HEST).

In the external cohort from Korea, the C_2HEST score also had significantly better predictivity for incident AF, as compared to the $CHADS_2$, CHA_2DS_2VASc scores:

- AUC $CHADS_2$: 0.637, 95%CI: 0.632-0.642 ($p<0.001$ vs. C_2HEST).
- AUC CHA_2DS_2VASc: 0.637, 95%CI: 0.632-0.642 ($p<0.001$ vs. C_2HEST).

while this difference was only marginally significant for HATCH score:

- AUC HATCH: 0.646, 95%CI: 0.641-0.651 ($p=0.059$ vs. C_2HEST).

Discussion

To our knowledge, this is the largest cohort study from an Asian population, aimed at developing a simple risk assessment tool for incident AF. We investigated risk factors for incident AF and derived and validated the new C_2HEST score as a user-friendly clinical score to assess individuals' risk of developing incident AF.

In the multivariable analysis, we found that independent risk factors for incident AF were SHD, HF, ageing (≥75 years), CAD, hyperthyroidism, COPD, and hypertension; all these risk factors were also demonstrated to increase the risk of incident AF in previous studies[24-27]. Indeed, SHD dramatically increases the risk of incident AF such that patients with SHD *per se* are very high risk of incident AF [24]. Heart failure is also another significant and independent risk factor, with a 2 to 6-fold increase in the risk of incident AF[25,26]. The ARIC (Atherosclerosis Risk in Communities) study showed that stable CAD was an independent risk factor for incident AF (HR, 2.21; 95% CI, 1.71-2.84)[12].

As highlighted above, previous studies have proposed several predictive tools for incident AF, such as the FHS score, the ARIC score and the CHARGE-AF score[11-13]. All these scoring systems were derived from large cohorts and showed good predictive values (C statistics, 0.75-0.78). However, these scores require many instrumental and laboratory variables to be calculated. Recently, the Suita study in Japan has developed a risk score for incident AF with a good predictive power (C-statistics, 0.75; 95% CI, 0.72-0.77)[28]. However, this score is complex (more than 16 points) and we could not compare the C_2HEST score with the Suita study, as our dataset did not have reliable data on 'cardiac

murmur', which is could be regarded as subjective parameter based on physician auscultation.

As an exploratory analysis, we do show that C_2HEST score had significantly better predictivity for incident AF, as compared to the $CHADS_2$, CHA_2DS_2VASc and HATCH scores in the derivation cohort. Similarly, in the external cohort, the C_2HEST confirmed higher predictivity for incident AF than $CHADS_2$, and CHA_2DS_2VASc scores, while it was only marginally significant compared to the HATCH score. These comparisons are with the caveat that these scores were not derived nor designed for the prediction of incident AF but they were elaborated for stroke risk stratification[29,30] or arrhythmia progression from paroxysmal to persistent AF[31].

In contrast, the C_2HEST was specifically designed to predict incident AF, and we used established approach whereby a large Asian cohort was used to derive the score with internal validation using bootstrap and calibration methods– and we then externally applied it in an *independent Asian population sample*, for new onset AF. The risk factors included in the C_2HEST score may be promptly deduced by a careful clinical evaluation of a patient with no need for any laboratory evaluation to calculate the individual risk of incident AF.

Strengths and limitations

This is the first study aimed at developing a simple, user-friendly clinical risk score, the C_2HEST score, from a large (n=471,446) community-based Asian population with long follow-up. Also, we provided an external application using a nationwide cohort derived from another large Asian population dataset, the Korean NIHS dataset (n=451,199). The C_2HEST score is easy to calculate and to apply in clinical practice and allows a good identification of patients at risk for incident AF.

Nevertheless, the present study has some limitations. There were some differences in the demographic characteristics and comorbidities between the Chinese and Korean cohorts, reflecting the different populations, healthcare system settings and the nature of the cohorts per se, that is, health insurance scheme data from one single China province (Yunnan) from local 2A and 3A-grade hospitals, and a nationwide sample of Korean subjects.

The incidence of AF in the derivation cohort was also lower than that of Western populations which may result from under-diagnosis and ethnic differences. Asymptomatic AF may be under-represented, given that there was no opportunistic screening for AF performed in this study. Indeed, the lower incidence of AF in Asian populations compared with Western populations is in accordance with previous reports showing an incidence of AF of 0.50-1.37 per 1000 person-years in Asians vs. 3.04-3.68 per 1000 person-years in Western populations, respectively[32-35]. The detection of asymptomatic AF is also problematic in large population studies considering the random pattern of AF onset. Nevertheless, it would generally be impossible to screen *half a million people* for detecting asymptomatic AF, while such 'AF screening' work could be done only in small highly selected samples (and possibly underpowered), which would again diminish the power of a large population-targeted study. Therefore, in order to do such large population-based epidemiology study in developing countries, such as China and other Asian countries, we are using the most viable way – based on a simple 'clinical score' based approach. Another issue is that screening method selection could have major impact on results.

Although we have excluded those patients with known AF at baseline according to available medical

records, there may have some patients with unknown AF (or transient AF) for which we could not confirm thorough AF screening at the patient level. However, this would not defeat the meaning of evaluating the subject's risk of developing incident AF during follow-up. Indeed, patients with AF occurring transiently are associated with high incidence of another AF episode in the coming days[36], and these patients usually have poor AF-related outcomes[37,38].

The derivation cohort subjects of the present study came from Yunnan Provence in the southwest of China and may not represent the whole of China and Asia. We could not compare the C_2HEST score with other risk scores for incident AF, such as FHS and CHARGE-AF as some variables from these scores were not collected for this dataset. However, the use of such complicated scoring systems relies on easily accessible computing and information systems, which are not applicable in the relatively underdeveloped Asia area. The simplicity of C_2HEST, which could be calculated by every clinical practitioner without relying on advanced information technology, may address the current unmet medical needs in Asia and help address the burden of AF in this part of the world. While we have performed external application in another very large Asian cohort from Korea, further validation of the C_2HEST score is needed, especially in non-Asian cohorts (currently ongoing**).** Cultural health and environmental variations may be other factors which would have increased the complexity of any derived score (defeating the purpose of a simple clinical score for everyday practice), and are generally subordinated factors compared with clinical situations or disease conditions when making clinical decisions for the risk of incident AF.

Conclusion

We have developed and validated the C_2HEST score as a simple clinical score to assess the individual risk of developing AF in the Asian population. This novel score may help identify patients *without SHD* who are at risk of incident AF and may be targeted for prevention strategies.

Acknowledgements

Y.G.L, D.P., A.F. and G.Y.H.L. had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Y.G.L was principal author. Y.G.L, D.P., A.F., P.S.Y., E.J., B.J., and G.Y.H.L contributed to the study concept and data analyses and Y.G.L, D.P., Y.T.W, Y.T.G, and G.Y.H.L. contributed to drafting and revisions of the manuscript.

References

1. Pastori D, Pignatelli P, Angelico F, et al. Incidence of myocardial infarction and vascular death in elderly patients with atrial fibrillation taking anticoagulants: relation to atherosclerotic risk factors. *Chest.* 2015;147(6):1644-1650.
2. Lee HY, Yang PS, Kim TH, et al. Atrial fibrillation and the risk of myocardial infarction: a nation-wide propensity-matched study. *Scientific reports.* 2017;7(1):12716.
3. Gomez-Outes A, Lagunar-Ruiz J, Terleira-Fernandez AI, Calvo-Rojas G, Suarez-Gea ML, Vargas-Castrillon E. Causes of Death in Anticoagulated Patients With Atrial Fibrillation. *Journal of the American College of Cardiology.* 2016;68(23):2508-2521.
4. Chugh SS, Havmoeller R, Narayanan K, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation.* 2014;129(8):837-847.
5. Zhou Z, Hu D. An epidemiological study on the prevalence of atrial fibrillation in the Chinese population of mainland China. *J Epidemiol.* 2008;18(5):209-216.
6. Ohsawa M, Okayama A, Sakata K, et al. Rapid increase in estimated number of persons with atrial fibrillation in Japan: an analysis from national surveys on cardiovascular diseases in 1980, 1990 and 2000. *J Epidemiol.* 2005;15(5):194-196.
7. Friberg L, Engdahl J, Frykman V, Svennberg E, Levin LA, Rosenqvist M. Population screening of 75- and 76-year-old men and women for silent atrial fibrillation (STROKESTOP). *Europace : European pacing, arrhythmias, and cardiac electrophysiology : journal of the working groups on cardiac pacing, arrhythmias, and cardiac cellular electrophysiology of the European Society of Cardiology.* 2013;15(1):135-140.
8. Lowres N, Neubeck L, Redfern J, Freedman SB. Screening to identify unknown atrial fibrillation. A systematic review. *Thromb Haemost.* 2013;110(2):213-222.
9. Cosio FG, Aliot E, Botto GL, et al. Delayed rhythm control of atrial fibrillation may be a cause of failure to prevent recurrences: reasons for change to active antiarrhythmic treatment at the time of the first detected episode. *Europace : European pacing, arrhythmias, and cardiac electrophysiology : journal of the working groups on cardiac pacing, arrhythmias, and cardiac cellular electrophysiology of the European Society of Cardiology.* 2008;10(1):21-27.
10. Nattel S, Guasch E, Savelieva I, et al. Early management of atrial fibrillation to prevent cardiovascular complications. *Eur Heart J.* 2014;35(22):1448-1456.
11. Schnabel RB, Sullivan LM, Levy D, et al. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *The Lancet.* 2009;373(9665):739-745.
12. Chamberlain AM, Agarwal SK, Folsom AR, et al. A clinical risk score for atrial fibrillation in a biracial prospective cohort (from the Atherosclerosis Risk in Communities [ARIC] study). *The American journal of cardiology.* 2011;107(1):85-91.
13. Alonso A, Krijthe BP, Aspelund T, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *Journal of the American Heart Association.* 2013;2(2):e000102.
14. Tse HF, Wang YJ, Ahmed Ai-Abdullah M, et al. Stroke prevention in atrial fibrillation--an

Asian stroke perspective. *Heart rhythm : the official journal of the Heart Rhythm Society.* 2013;10(7):1082-1088.

15. Li Y, Pastori D, Guo Y, Wang Y, Lip GYH. Risk factors for new-onset atrial fibrillation: A focus on Asian populations. *International journal of cardiology.* 2018;261:92-98.
16. Guo Y, Tian Y, Wang H, Si Q, Wang Y, Lip GYH. 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.
17. Seong SC, Kim YY, Park SK, et al. Cohort profile: the National Health Insurance Service-National Health Screening Cohort (NHIS-HEALS) in Korea. *BMJ Open.* 2017;7(9):e016640.
18. Kim TH, Yang PS, Uhm JS, et al. CHA2DS2-VASc Score (Congestive Heart Failure, Hypertension, Age >/=75 [Doubled], Diabetes Mellitus, Prior Stroke or Transient Ischemic Attack [Doubled], Vascular Disease, Age 65-74, Female) for Stroke in Asian Patients With Atrial Fibrillation: A Korean Nationwide Sample Cohort Study. *Stroke.* 2017;48(6):1524-1530.
19. Kim TH, Yang PS, Kim D, et al. CHA2DS2-VASc Score for Identifying Truly Low-Risk Atrial Fibrillation for Stroke: A Korean Nationwide Cohort Study. *Stroke.* 2017;48(11):2984-2990.
20. Gronnesby JK, Borgan O. A method for checking regression models in survival analysis based on the risk score. *Lifetime data analysis.* 1996;2(4):315-328.
21. Suenari K, Chao TF, Liu CJ, Kihara Y, Chen TJ, Chen SA. Usefulness of HATCH score in the prediction of new-onset atrial fibrillation for Asians. *Medicine.* 2017;96(1):e5597.
22. Chao TF, Liu CJ, Chen SJ, et al. CHADS2 score and risk of new-onset atrial fibrillation: a nationwide cohort study in Taiwan. *International journal of cardiology.* 2013;168(2):1360-1363.
23. Hsieh CY, Lee CH, Wu DP, Sung SF. Prediction of new-onset atrial fibrillation after first-ever ischemic stroke: A comparison of CHADS2, CHA2DS2-VASc and HATCH scores and the added value of stroke severity. *Atherosclerosis.* 2018;272:73-79.
24. Zhou Z, Hu D. An Epidemiological Study on the Prevalence of Atrial Fibrillation in the Chinese Population of Mainland China. *Journal of Epidemiology.* 2008;18(5):209-216.
25. Bordignon. S, Corti. MC, Bilato. C. Atrial fibrillation associated with heart failure, stroke and mortality. *Journal of Atrial fibrillation.* 2012;5(1):467.
26. Benjamin. EJ, Levy. D, Vaziri. SM, D'Agostino. RB. Independent risk factors for atrial fibrillation in a population-based cohort. The Framingham Heart Study. *JAMA.* 1994;271(1994):840-844.
27. Heeringa J, van der Kuip DA, Hofman A, et al. Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. *European heart journal.* 2006;27(8):949-953.
28. Kokubo Y, Watanabe M, Higashiyama A, Nakao YM, Kusano K, Miyamoto Y. Development of a Basic Risk Score for Incident Atrial Fibrillation in a Japanese General Population- The Suita Study. *Circulation journal : official journal of the Japanese Circulation Society.* 2017;81(11).
29. Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest.* 2010;137(2):263-272.

30. Gage BF, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA : the journal of the American Medical Association.* 2001;285(22):2864-2870.
31. de Vos CB, Pisters R, Nieuwlaat R, et al. Progression from paroxysmal to persistent atrial fibrillation clinical correlates and prognosis. *Journal of the American College of Cardiology.* 2010;55(8):725-731.
32. Miyasaka Y, Barnes ME, Gersh BJ, et al. Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. *Circulation.* 2006;114(2):119-125.
33. Chien KL, Su TC, Hsu HC, et al. Atrial fibrillation prevalence, incidence and risk of stroke and all-cause death among Chinese. *International journal of cardiology.* 2010;139(2):173-180.
34. Ruigomez A, Johansson S, Wallander MA, Rodriguez LA. Incidence of chronic atrial fibrillation in general practice and its treatment pattern. *J Clin Epidemiol.* 2002;55(4):358-363.
35. Stewart S, Hart CL, Hole DJ, McMurray JJ. Population prevalence, incidence, and predictors of atrial fibrillation in the Renfrew/Paisley study. *Heart.* 2001;86(5):516-521.
36. Lubitz SA, Yin X, Rienstra M, et al. Long-term outcomes of secondary atrial fibrillation in the community: the Framingham Heart Study. *Circulation.* 2015;131(19):1648-1655.
37. Adderley NJ, Nirantharakumar K, Marshall T. Risk of stroke and transient ischaemic attack in patients with a diagnosis of resolved atrial fibrillation: retrospective cohort studies. *Bmj.* 2018;361:k1717.
38. Fauchier L, Clementy N, Bisson A, et al. Prognosis in patients with atrial fibrillation and a presumed "temporary cause" in a community-based cohort study. *Clinical research in cardiology : official journal of the German Cardiac Society.* 2017;106(3):202-210.

Table 1. Baseline characteristics of 471,446 subjects included in the internal validation cohort.

Characteristics	Subjects without AF (n=470,525)	Subjects with incident AF (n=921)	P value
Age, mean (SD), y	47 (16)	62 (12)	<0.001
Male sex, No. (%)	247,752 (52.7)	574 (62.3)	<0.001
Medical history, No. (%)			
Hypertension	45,444 (9.7)	236 (25.6)	<0.001
Diabetes	18,900 (4.0)	29 (3.1)	0.180
CAD	14,813 (3.1)	187 (20.3)	<0.001
Hyperlipidemia	6,610 (1.4)	17 (1.9)	0.198
Vascular disease	6,003 (1.3)	9 (1.0)	0.486
COPD	1,687 (0.4)	25 (2.7)	<0.001
Previous IS	1,608 (0.3)	7 (0.8)	0.023
Renal dysfunction	1,423 (0.3)	8 (0.9)	0.002
Hyperthyroidism	971 (0.2)	5 (0.5)	0.025
HF	700 (0.1)	43 (4.7)	<0.001
SHD	401 (0.1)	33 (3.6)	<0.001
Hypothyroidism	203 (0.0)	1 (0.1)	0.340
CHA_2DS_2VASc score, median (IQR)	1 [0-1]	2 [1-3]	<0.001

AF=atrial fibrillation; CAD=coronary artery disease; COPD=chronic obstructive pulmonary disease; HF=heart failure; IQR=interquartile range; IS=ischemic stroke; SD=standard deviation; SHD=structural heart disease; CHA_2DS_2VASc score=congestive heart failure, hypertension, age ≥75, diabetes, stroke, vascular disease, age 65-74, female.

Table 2 Hazard ratios of risk factors for incident atrial fibrillation.

Risk factors	Univariable analysis			Multivariable analysis		
	HR	95% CI	P value	HR	95% CI	P value
SHD	39.00	27.6-55.2	<0.001	26.1	18.2-37.3	<0.001
HF	34.66	25.5-47.1	<0.001	7.95	5.76-11.0	<0.001
Age ≥75 years	8.24	6.80-9.98	<0.001	5.83	4.80-7.09	<0.001
CAD	7.01	5.97-8.23	<0.001	4.14	3.50-4.90	<0.001
Hyperthyroidism	2.06	0.85-4.96	0.107	3.20	1.33-7.71	0.010
COPD	6.81	4.57-10.1	<0.001	3.01	1.33-7.71	<0.001
Hypertension	4.63	4.05-5.29	<0.001	3.24	2.82-3.73	<0.001
Renal dysfunction	2.91	1.45-5.83	0.003			
Previous IS	2.13	1.01-4.49	0.046			
Hyperlipidemia	1.12	0.71-1.79	0.620			
Male	1.22	1.07-1.40	0.003			
Hypothyroidism	1.79	0.25-12.7	0.561			
Diabetes	0.61	0.42-0.89	0.009			
Vascular Disease	1.36	0.38-1.41	0.355			

HR=hazard ratio; CI=confidence interval; for other abbreviations see Table 1.

Table 3 The C_2HEST score for incident atrial fibrillation

Acronym	Risk factor	Points
C_2	CAD (*1 point*) / COPD (*1 point*)	1-2
H	Hypertension (*1 point*)	1
E	Elderly (Age ≥ 75, *2 points*)	2
S	Systolic HF (*2 points*)	2
T	Thyroid disease (hyperthyroidism) (*1 point*)	1
	Total points	0-8
	AUC (c-index)	**95% CI**
C_2HEST score	0.749	0.729-0.769

AUC=area under the curve; CI=confidence interval; for other abbreviations see Table 1.

Table 4 Annual incidence of atrial fibrillation by C_2HEST score.

Score	Subjects (No.)	Incident AF (No.)	Incidence of AF *	Hazard Ratio	95% CI
0	310,117	246	0.18	1.00	-
1	88,825	378	0.82	4.31	3.67-5.06
2	19,270	148	2.31	12.8	10.4-15.6
3	8,253	68	3.73	22.6	17.2-29.6
4	1,373	68	16.1	97.0	74.1-127.0
5	90	6	28.7	187.4	83.3-421.6
≥6	45	7	59.8	332.0	156.6-704.0

* Per 1000 person-year; AF= atrial fibrillation; CI=confidence interval.

Figure Legends

Figure 1. Receiver operating characteristic (ROC) curves for the C_2HEST score in predicting incident AF. Panel A. Internal cohort, Panel B. External cohort.

Figure 2. Kaplan-Meier curves for risk categories according to the C_2HEST score. Patients were divided into three groups low (0-1 points), medium (2-3 points) and high-risk (>3 points).

Figure 1a

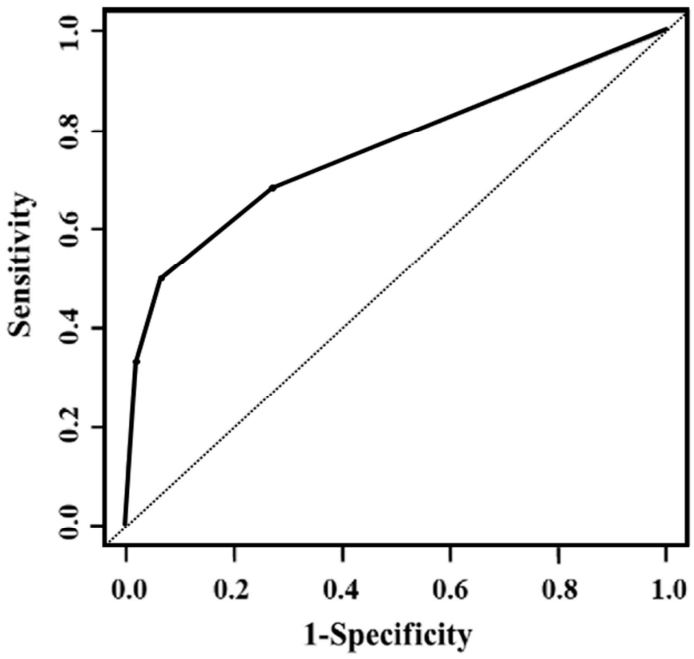


Figure 1b

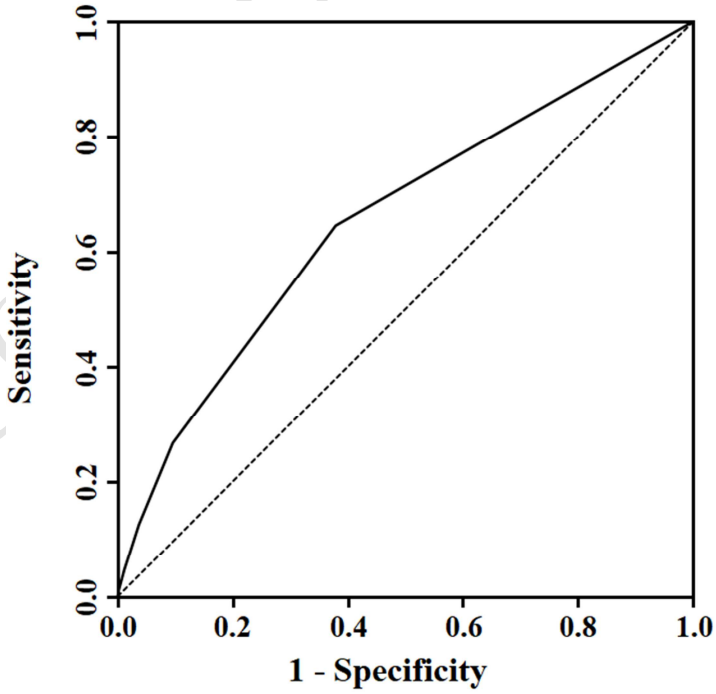
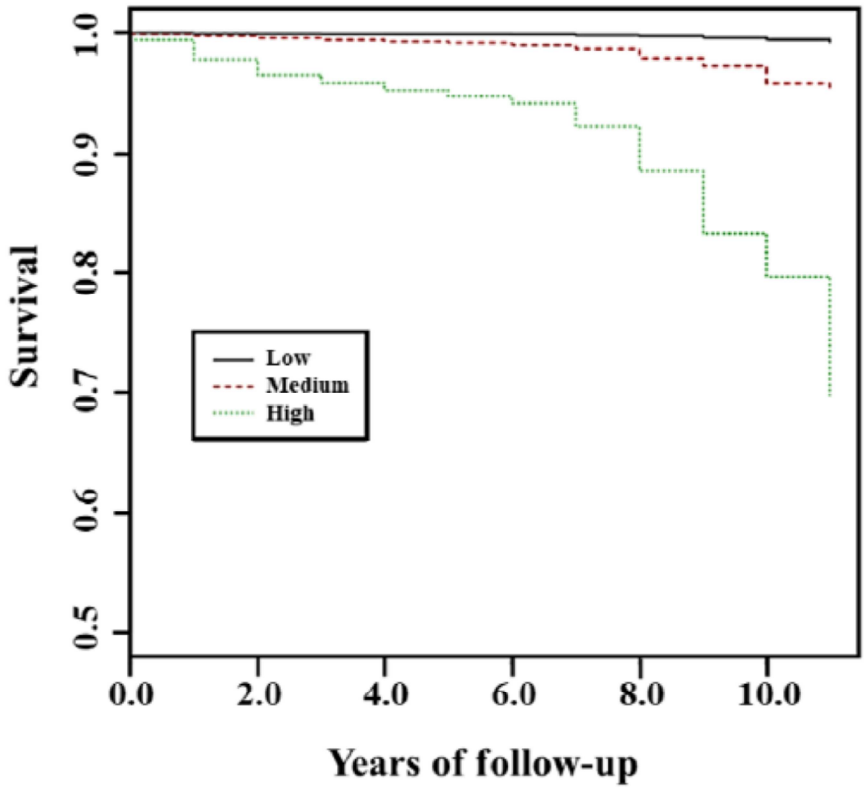


Figure 2.



Number at risk						
Low	398942	281464	219461	175221	107720	33995
Medium	27335	13528	9170	7112	4056	1148
High	1696	929	616	455	249	47

CHEST® Online Supplement

e-Appendix 1.

AF was diagnosed upon an ECG or Holter recording (ICD-9 codes: 427.31 or ICD10 codes: I48). Rheumatic valvular disease (ICD-9 codes: 393-398; ICD-10 codes: I05, I06, I07, I09.9), dilated cardiomyopathy (ICD-9 codes: 425.4; ICD-10 codes: I42.0), HF (ICD-9 codes: 428; ICD-10 codes: I42, I50, I110, I819), CAD (ICD-9 codes: 410-414; ICD-10 codes: I20-I25), hyperthyroidism (ICD-9 codes: 242; ICD-10 codes: E05), hypothyroidism (ICD-9 codes: 244; ICD-10 codes: E03), COPD (ICD-9 codes: 490-496; ICD-10 codes: J42, J44.0-9), hypertension (ICD-9 codes: 401-405; ICD-10 codes: I10-I15), renal dysfunction (ICD-9 codes: 585, 586; ICD-10 codes: M1A.3), IS (ICD-9 codes: 436; ICD-10 codes: I63), hyperlipidemia (ICD-9 codes: 272.4; ICD-10 codes: E78.0-3, E78.5), myocardial infarction (ICD-9 codes: 410; ICD-10 codes: I21, I22), peripheral vascular disease (ICD-9 codes: 440.2; ICD-10 codes: I65, I70-74), diabetes mellitus (ICD-9 codes: 249-250; ICD-10 codes: E10-E14).

e-Table 1. Baseline characteristics of subjects in the validation cohort

Characteristics	Total subjects (n=451,199)	Subjects without AF (n=439,056)	Subjects with incident AF (n=12,143)	P value
Age, mean (SD), y	56.1 (9.3)	55.9 (9.3)	62.3 (9.9)	<0.001
Male sex, No. (%)	243,503 (54.0)	236220 (53.8)	7283 (60.0)	<0.001
Comorbidities				
Hypertension	143,168 (31.7)	136586 (31.1)	6582 (54.2)	<0.001
Diabetes mellitus	37,372 (8.3)	35752 (8.1)	1620 (13.3)	<0.001
Coronary artery disease	9,946 (2.2)	9029 (2.1)	917 (7.6)	<0.001
Hyperlipidemia	45,012 (10.0)	43234 (9.8)	1778 (14.6)	<0.001
Vascular disease	13,656 (3.0)	12837 (2.9)	819 (6.7)	<0.001
COPD	44,470 (9.9)	42120 (9.6)	2350 (19.4)	<0.001
Previous ischemic stroke	14,980 (3.3)	14072 (3.2)	908 (7.5)	<0.001
Renal dysfunction	3,461 (0.8)	3267 (0.7)	194 (1.6)	<0.001
Systolic heart failure	5,515 (1.2)	5082 (1.2)	433 (3.6)	<0.001
Thyroid disease	2,477 (0.6)	2382 (0.5)	95 (0.8)	<0.001
CHA_2DS_2-VASc score, median (IQR)	0 (0-1)	0 (0-1)	1 (0-2)	<0.001