



Published in final edited form as:

Cardiology. 2017 ; 138(3): 133–140. doi:10.1159/000476030.

A Clinical Score for Predicting Atrial Fibrillation in Patients with Cryptogenic Stroke or Transient Ischemic Attack

Calvin Kwong, MD^{a,*}, Albee Y. Ling^{b,*}, Michael H. Crawford, MD^c, Susan X. Zhao, MD^{d,#}, and Nigam H. Shah, MBBS, Ph.D^{e,#}

^aDepartment of Medicine, Santa Clara Valley Medical Center, San Jose, California

^bBiomedical Informatics Training Program, Stanford University, Stanford, California

^cUniversity of California San Francisco, San Francisco, California

^dDivision of Cardiology, Santa Clara Valley Medical Center, San Jose, California

^eCenter for Biomedical Informatics Research, Stanford University School of Medicine, Stanford, California

Abstract

Objectives—Detection of atrial fibrillation (AF) in post cryptogenic stroke (CS) or transient ischemic attack (TIA) patients carries important therapeutic implications.

Methods—To risk stratify CS/TIA patients for later development of AF, we conducted a retrospective cohort study using data from 1995–2015 in Stanford Translational Research Integrated Database Environment (STRIDE).

Results—Of the 9,589 adult patients (age 40) with CS/TIA included, 482 (5%) patients developed AF post CS/TIA. Of these patients, 28.4%, 26.3%, and 45.3% were diagnosed with AF 1–12 months, 1–3 years, and >3 years post index CS/TIA, respectively. Age (>75), obesity, congestive heart failure, hypertension, coronary artery disease, peripheral vascular disease, and valve disease are significant risk factors with respective odds ratios (95% CI) 1.73 (1.39–2.16), 1.53 (1.05–2.18), 3.34 (2.61–4.28), 2.01 (1.53–2.68), 1.72 (1.35–2.19), 1.37 (1.02–1.84), 2.05 (1.55–2.69). A risk scoring system, the HAVOC score, was constructed using these 7 clinical variables that successfully stratifies patients into 3 risk groups, with good model discrimination (AUC=0.77).

Conclusions—Findings from this study support the strategy of looking longer and harder for AF in post CS/TIA patients. The HAVOC score identifies different levels of AF risk and may be used to select patients for extended rhythm monitoring.

Corresponding Author: Susan Zhao, MD; Tel: (+1) 408-885-2345; fax: (+1) 408-885-4387, 751 S. Bascom Avenue, Suite # 340, San Jose, CA 95128, USA, susan.zhao@hhs.sccgov.org.

*These authors contributed equally to this work.

#These authors contributed equally to this work. Dr. Zhao is the medical expert and Dr. Shah is the informatics expert.

Disclosures

The authors have no conflicts of interest to disclose.

Keywords

atrial fibrillation; cryptogenic stroke; rhythm monitoring; cardiovascular risk and prevention

Cryptogenic stroke (CS) is defined as stroke of unknown etiology. Survivors of a cryptogenic stroke (CS) or transient ischemic attack (TIA) have an increased risk of another CS/TIA which is a major source of increased mortality and morbidity [1,2]. Atrial fibrillation (AF) has been shown to be an independent risk factor for CS/TIA [3]. Approximately 10% of patients with acute ischemic stroke or TIA will have new AF detected during their hospital admission, while an additional 11% may be found to have new AF if tested within 30 days of discharge by continuous electrocardiographic monitoring [4]. Diagnosis of AF in CS/TIA patients carries significant therapeutic implications in that current practice favors antiplatelet agents alone for CS/TIA patients without known risk for cardioembolism, while oral anticoagulants have been shown by a large body of clinical evidence to be superior in stroke prevention in those with proven AF [5]. Despite the recommendation by the American Heart Association/American Stroke Association joint statement in 2014 of prolonged rhythm monitoring (30 days) for AF detection within 6 months of index CS/TIA [4], the optimum monitoring duration and method of AF detection after CS/TIA are unknown. The EMBRACE study [6] and the CRYSTAL AF study [7] underscore the importance of prolonged monitoring for the detection of AF and reclassification of ischemic stroke subtype. However, it is impossible to apply extended cardiac monitoring to all stroke patients in clinical practice. Therefore, there is an unmet need to risk stratify patients for both clinical and cost-benefit purposes. With the advent of large practice-based electronic health records (EHR), we set out to assess the clinical risk factors that are associated with diagnosis of AF following CS/TIA to identify patients for whom prolonged rhythm monitoring and high clinical vigilance must be maintained. In this retrospective cohort study, we hypothesized that common clinical risk factors at time of index of CS/TIA can predict incident AF rate.

Methods

We used data from the Stanford Translational Research Integrated Database Environment (STRIDE) which contains clinical information of over 2 million pediatric and adult patients cared for at Stanford Health Care and Stanford Children's Health from 1995 to 2015, including 20 million patient encounters with transcriptions of all inpatient and outpatient clinical notes, pathology and radiology reports, medication lists, lab results, and vitals data. This data source was accessed under approved Institutional Review Board protocols.

Through a previously validated and implemented text-processing pipeline to analyze clinical data [8–10], we used Unitex [11] as an annotator and over 10 clinical ontologies to extract positive present mentions of disease concepts from all clinical notes. We excluded uninformative phrases based on the term frequency analysis [12] and kept only terms with more than 4 characters to avoid ambiguity. We also flagged negative mentions (e.g. “ruled out stroke”) and determined if a term was from patients' history or family history sections of

a note [13]. The product of this pipeline is a list of present, positive mentions of biomedical concepts in each patient note.

We identified all patients who had their first ICD-9 documentation of CS/TIA at age 40 or older in either inpatient and outpatient encounters. The inclusion criteria using ICD-9 diagnosis codes are stroke (434 and 436) and TIA (435.9). These ICD-9 codes were selected because have been previously shown to have high specificity and sensitivity for ischemic stroke when confirmed with chart review[14].

From the CS/TIA patients identified using these codes, some were removed from the cohort based on both ICD-9 and clinical text evidence that meets specific exclusion criteria to increase specificity for patients without these conditions. Patients who had ICD-9 diagnosis of carotid artery occlusion or stenosis (433.1), intracranial hemorrhage (431), and atrial septal defects (745.5) were excluded as these are identifiable etiologies of stroke. Patients with rheumatic heart disease (433.1) or prosthetic valve(s) (V43.3) were excluded as AF in these contexts belong to a separate entity, valvular AF, separate from AF of the general population. Those with hyperthyroid disease (242.9) were also excluded as this is a known reversible cause of AF. Patients who had clinical text evidence of rheumatic heart disease, prosthetic valve(s), and/or patent foramen ovale were also excluded.

The outcome of interest in this study is diagnosis of AF after CS/TIA. All patients who had history of AF were identified by ICD-9 code (427.31 and 427.32). Those positive for AF were defined as patients over 40 years old with CS/TIA whose first ICD-9 documentation of AF was at least 30 days after first episode of CS/TIA. We used a 30-day cutoff to exclude patients who may have had delayed documentation of AF related to their hospitalization for initial stroke. Those negative for AF were defined as patients with CS/TIA with no ICD-9 documentation of AF during the extent of their follow-up as documented in their records.

Basic demographic information such as age at time of CS/TIA and sex were obtained from the structured fields of their records. Risk factors were extracted based on ICD-9 documentation at any time point of patients' records to enable us to better capture those patients' chronic conditions. Risk factors assessed were hypertension (HTN), diabetes (DM), obesity defined by BMI>30, systolic and/or diastolic heart failure (CHF), coronary artery disease (CAD), peripheral vascular disease (PVD), chronic kidney disease (CKD) stage III, IV, V, aortic valve disease, mitral valve disease, tricuspid valve disease, and pulmonary valve disease. Such clinical factors are well known to be risk factors for AF [15,16]. A comprehensive list of conditions covered by each respective ICD-9 used is detailed in the supplemental table.

We randomly split the total cohort into 2 groups: the first for model derivation (80%) and the second for model validation (20%). Candidate predictor variables include age, sex and all the risk factors described above. Univariable logistic regression was first applied to identify the association between each of the predictor variables and diagnosis of AF in the derivation cohort. A multivariable logistic regression model with stepwise variable selection was then trained on the derivation cohort to identify predictors of AF and to estimate their relative predictive power. A simplified risk stratification system was developed based on the beta

coefficients of the multivariable logistic regression model as validated by previously published methods [17]. Points assigned to each significant risk factor were obtained by dividing each by the lowest coefficient and rounding to the nearest integer [18]. We then calculated a patient's risk score by summing up all points that correspond to the risk factors present in the given patient's record.

We assessed model discrimination by using the c-statistic, or area under the curve (AUC) of the Receiver Operating Characteristic (ROC) Curve, which defines how well a model or prediction rule can discriminate between patients who are and are not positive for an event. We then used the Cochran-Armitage trending statistic to assess the ability of the risk score system to differentiate low-risk from high-risk patients. The scoring system was applied to and evaluated in the derivation cohort to assess its applicability. The performance of HAVOC and CHA₂DS₂-Vasc was compared at a score of 4, which was the cutoff value between the low and medium risk strata for both scoring systems. McNemar's Chi-squared tests were used to compare the sensitivity, specificity, and accuracy. Positive predictive values (PPV) and negative predictive values (NPV) were compared using a test score developed by Leisenring et al. [19] All analyses were performed using open source statistical program R Version 3.2.2 [20].

Results

9,589 patients met the inclusion and exclusion criteria, 482 (5%) of whom had a new diagnosis of AF >30 days post diagnosis of CS/TIA (Figure 1). Of these patients, 28.4% received a new diagnosis of AF 1 to 12 months post index CS/TIA, 26.3% between 1 to 3 years post index CS/TIA, and 45.3% after 3 years post index neurological event (Figure 2). 7,671 (80%) patients were set aside for the derivation cohort and 1,918 (20%) patients for the validation cohort.

Within the derivation cohort, comparing between the AF positive and AF negative patients, univariable logistic regression revealed that the following risk factors were significantly associated with development of AF ($p < 0.05$): Age ≥ 75 , HTN, DM, Obesity, CHF, CAD, PVD, CKD, Mitral Valve Disease, Tricuspid Valve Disease, Pulmonary Valve Disease, and Aortic Valve Disease. A combined risk factor "Valve Disease" (aortic, mitral, tricuspid, and/or pulmonary valve disease) was also found to be significant (Table 1). Mean age (standard deviation) in the AF positive and AF negative groups were 68.14 (13.41) and 67.50 (13.47), respectively. Numeric age values were converted to binary values using an age cutoff of 75 years old, as distribution above and below cutoff is statistically different in AF positive and AF negative patients using a chi-squared test ($p = 3.96E-17$). Thus, the final input variables into logistic regression models were all binary.

A multivariable logistic regression model with stepwise feature selection was applied to data from the derivation cohort. Given that the combined variable "valve disease" was significant in our univariable logistic regression model, it was entered in the multivariable analysis instead of the 4 individual conditions. Age ≥ 75 , CHF, HTN, CAD, PVD, obesity, and valve disease were found to be statistically significant in the multivariable logistic regression model (Table 2). The predictive model developed using these risk factors had good

discrimination in the derivation cohort (c-statistic 0.77) with very similar results when applied to the validation cohort (c-statistic 0.77, $p=0.79$ using DeLong's test).

The HAVOC score (abbreviate for **h**ypertension, **a**ge, **v**alvular heart disease, peripheral **v**ascular disease, **o**besity, **c**ongestive heart failure, and **c**oronary artery disease) was developed by assigning respective points for each risk predictor based on the corresponding regression coefficients (Table 2). The regression coefficients were transformed by dividing each coefficient by the smallest coefficient in the model and then rounding to the nearest integer to obtain a respective point value. After points were summed, possible total scores ranged from 0 to 14.

The scores were then categorized into 3 risk levels, low risk (scores 0–4), medium risk (5–9), and high risk (10–14) [17]. Of the derivation cohort, 78.8% patients were in the low risk group, 16.4% in the medium risk group, and 4.8% in the high-risk group with a similar trend in the validation cohort. AF rate in the derivation and validation cohorts increased significantly with risk score strata ($p<.0001$ by Cochran-Armitage trending test for both derivation and validation cohorts). In the derivation cohort, those with score of 0–4 had 2.5% risk of developing AF >30 days post stroke. In contrast, those with score of 10–14 had 24.9% risk. A similar trend was observed in the validation cohort (Figure 3).

Given the overlapping nature between HAVOC and CHA₂DS₂-VASc [21], we applied CHA₂DS₂-VASc scores to our cohort of patients as well. The range of CHA₂DS₂-VASc scores in our cohort of patients was from 2 to 9. Similar to HAVOC, the CHA₂DS₂-VASc scores were further divided into 3 risk categories: low risk (scores 2–4), medium risk (5–6), and high risk (7–9). Results from Cochran Armitage Test show that the rate of AF positive patients also increase with CHA₂DS₂-VASc score strata (p-value <0.001). Comparing HAVOC and CHA₂DS₂-VASc using the cutoff values between low and medium risk strata (4 points in both scoring system), HAVOC has higher specificity and accuracy (both p values < 0.001) (Table 3).

Discussion

Diagnosing AF after CS/TIA is a clinically significant event since eligible patients will start oral anticoagulants in lieu of antiplatelet agents that are standard of care in this patient group. Guideline recommendations have evolved from at least 24 hours of ECG monitoring [22] to 30 days of rhythm monitoring [4]. Prolonged monitoring post index neurological event has been shown to improve AF diagnosis rate [6,7]. However, cost-effective analysis is currently lacking to support the wide-spread use of these expensive devices, particularly implanted cardiac monitors, in the post stroke population. Scoring systems have been developed to predict post-stroke atrial fibrillation [15,23–27]; yet these studies are inconclusive due to small sample size, short monitoring period, difficulty in data acquisition, or poor applicability to the CS/TIA population. Our study was conducted to address this pressing need to properly triage resources in stroke patients without manifest AF beyond the initial 30-day window.

Through a large EHR database, a cohort of 9,589 CS/TIA patients was identified, 5% of whom was diagnosed with AF during a median of 2.6 years follow-up, a percentage that is comparable to some [28] but lower than others [29]. Previous studies have estimated that AF can be detected in about 10% of patients with stroke by cardiac monitoring. However, cohorts in these studies had different proportions of patients with different types of stroke and different means of detection methods. Using the proposed classification [29], post stroke cardiac monitoring was stratified into 4 consecutive phases: phase 1 (emergency room); phase 2 (in hospital); phase 3 (1st ambulatory period) and phase 4 (second ambulatory period). Our study's focus was on phase 4 when uncertainty about the need for further rhythm monitoring in post-stroke patient is the greatest. Significant independent risk factors associated with diagnosis of AF at least 30 days after CS/TIA were age ≥ 75 , obesity, history of CHF, HTN, CAD, PVD, and non-rheumatic, non-prosthetic valve disease. These factors closely resemble those identified by the Framingham Heart study in which age, sex, body-mass index, treatment for hypertension, PR interval, clinically significant cardiac murmur, and heart failure were strongly associated with AF in an epidemiological, non-stroke cohort [30]. A risk scoring system, the HAVOC score, was developed using multivariable regression coefficients and patients could be further assigned to 1 of 3 strata with varying risk of AF. The HAVOC score, with good discrimination and calibration, independently identified 4 components of the CHA₂DS₂-VASc system (CHF, HTN, age, vascular disease) as risk factors for AF when there is preceding stroke. These predisposing conditions represent clusters of common cardiovascular risk factors and play major role in various atherosclerotic/thrombotic processes. While the CHA₂DS₂-VASc system has been reported to be associated with cardiovascular events in general population or non-AF patient population [31,32], it has not been validated in predicting AF in the post-stroke population. Two unique features included in the HAVOC score (and not part of the CHA₂DS₂-VASc score) are obesity and non-rheumatic, non-prosthetic valvular disease. Obesity, as defined here with BMI over 30, is a well recognized contributor to the genesis and maintenance of AF [33]. Similarly, almost any valvular lesion with significant stenosis or regurgitation is associated with AF. Inclusion of obesity and valvular heart disease into an AF prediction tool makes biological sense and is well supported by clinical as well as epidemiological data.

The HAVOC score can successfully stratify CS/TIA patients into low, medium, and high risk for having AF. It is particularly powerful at identifying low risk patients in whom expensive, prolonged rhythm monitoring post CS/TIA may not be necessary and cost-effective. In this regard, the HAVOC score outperforms CHA₂DS₂-VASc score in both test specificity and overall accuracy. The HAVOC score therefore provides a means to triage valuable resources to those who will benefit the most.

Our study demonstrates the need for prolonged rhythm monitoring as evidenced by Figure 2, which shows that a large proportion of patients were diagnosed with AF after 1 year post index CS/TIA (26.3% in 1 to 3 years and 45.3% over 3 years). It is important to note that with prolonged rhythm monitoring, delayed (1 to 3 years post) or very late (over 3 years) AF may simply reflect the increased propensity to develop AF with advancing age rather than a true causative relationship between AF diagnosed years later and the initial index event. The very late group (>3 years), which represented nearly half (at 45.3%) of all AF diagnoses,

was particularly alarming because this is even beyond the standard monitoring window of currently available implantable devices and underscores the importance of maintaining high clinical vigilance for AF surveillance in this group of patients.

Our study also demonstrates the value of EHR in clinical risk stratification applications. Compared to studies using primary data sources such as survey data, our study was conducted on a relatively large sample, with long patient follow-up time. The primary limitation of the study is the fact that the analysis is retrospective and dependent on the reliability of ICD-9 coding for determining diagnosis of CS/TIA and AF; however, our methods of data mining have been validated and implemented in other studies [10,34]. Risk factors explored are based on widely accepted risk factors for AF and/or cardiovascular disease. Echocardiographic parameters were not included in the modeling as it has been shown that they did not improve risk reclassification [30] yet may introduce unnecessary complexity to hinder easy applicability. Future studies could further take advantage of the wealth of information in EHR to learn more potential risk factors. Characteristics such as laboratory values, although they could have identified additional significant risk factors, were not evaluated given relatively incomplete documentation in our database in its current state. Additional efforts to ensure data quality would make those types of data more useful as well. Lastly, our new scoring system requires independent validation from cohorts similar to that of CRYSTAL AF and other prospective studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

None.

Grant Support: Shah acknowledges NIGMS grant R01 GM101430, Bethesda, MD, USA, and infrastructure to carry out the project was funded in part by Janssen Research and Development. Ling acknowledges support from the Stanford Graduate Fellowship, Stanford, CA, USA.

References

1. Mohan KM, Wolfe CDA, Rudd AG, Heuschmann PU, Kolominsky-Rabas PL, Grieve AP. Risk and cumulative risk of stroke recurrence: a systematic review and meta-analysis. *Stroke J Cereb Circ.* 2011 May;42:1489–1494.
2. Burn J, Dennis M, Bamford J, Sandercock P, Wade D, Warlow C. Long-term risk of recurrent stroke after a first-ever stroke. The Oxfordshire Community Stroke Project. *Stroke J Cereb Circ.* 1994 Feb; 25:333–337.
3. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke J Cereb Circ.* 1991 Aug;22:983–988.
4. Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, et al. Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke J Cereb Circ.* 2014 Jul;45:2160–2236.
5. Medi C, Hankey GJ, Freedman SB. Stroke risk and antithrombotic strategies in atrial fibrillation. *Stroke J Cereb Circ.* 2010 Nov;41:2705–2713.
6. Gladstone DJ, Spring M, Dorian P, Panzov V, Thorpe KE, Hall J, et al. Atrial Fibrillation in Patients with Cryptogenic Stroke. *N Engl J Med.* 2014 Jun 26;370:2467–2477. [PubMed: 24963566]

7. Sanna T, Diener H-C, Passman RS, Di Lazzaro V, Bernstein RA, Morillo CA, et al. Cryptogenic Stroke and Underlying Atrial Fibrillation. *N Engl J Med*. 2014 Jun 26;370:2478–2486. [PubMed: 24963567]
8. Jung K, LePendu P, Iyer S, Bauer-Mehren A, Percha B, Shah NH. Functional evaluation of out-of-the-box text-mining tools for data-mining tasks. *J Am Med Inform Assoc JAMIA*. 2015 Jan;22:121–131. [PubMed: 25336595]
9. LePendu P, Iyer SV, Bauer-Mehren A, Harpaz R, Mortensen JM, Podchiyska T, et al. Pharmacovigilance using clinical notes. *Clin Pharmacol Ther*. 2013 Jun;93:547–555. [PubMed: 23571773]
10. Cole TS, Frankovich J, Iyer S, LePendu P, Bauer-Mehren A, Shah NH. Profiling risk factors for chronic uveitis in juvenile idiopathic arthritis: a new model for EHR-based research. *Pediatr Rheumatol Online J*. 2013; 11:45. [PubMed: 24299016]
11. Paumier S. UNITEX 3.0 User Manual. 2015
12. Wu ST, Liu H, Li D, Tao C, Musen MA, Chute CG, et al. Unified Medical Language System term occurrences in clinical notes: a large-scale corpus analysis. *J Am Med Inform Assoc JAMIA*. 2012 Jun;19:e149–156. [PubMed: 22493050]
13. Harkema H, Dowling JN, Thornblade T, Chapman WW. Context: An Algorithm for Determining Negation, Experiencer, and Temporal Status from Clinical Reports. *J Biomed Inform*. 2009 Oct; 42:839–851. [PubMed: 19435614]
14. Goldstein LB. Accuracy of ICD-9-CM coding for the identification of patients with acute ischemic stroke: effect of modifier codes. *Stroke J Cereb Circ*. 1998 Aug;29:1602–1604.
15. Bugnicourt J-M, Flament M, Guillaumont M-P, Chillon J-M, Leclercq C, Canaple S, et al. Predictors of newly diagnosed atrial fibrillation in cryptogenic stroke: a cohort study. *Eur J Neurol*. 2013 Oct;20:1352–1359. [PubMed: 23121252]
16. Psaty BM, Manolio TA, Kuller LH, Kronmal RA, Cushman M, Fried LP, et al. Incidence of and Risk Factors for Atrial Fibrillation in Older Adults. *Circulation*. 1997 Oct 7;96:2455–2461. [PubMed: 9337224]
17. Lipsky BA, Weigelt JA, Sun X, Johannes RS, Derby KG, Tabak YP. Developing and validating a risk score for lower-extremity amputation in patients hospitalized for a diabetic foot infection. *Diabetes Care*. 2011 Aug;34:1695–1700. [PubMed: 21680728]
18. Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med*. 2004 May 30;23:1631–1660. [PubMed: 15122742]
19. Leisenring W, Alonzo T, Pepe MS. Comparisons of predictive values of binary medical diagnostic tests for paired designs. *Biometrics*. 2000 Jun;56:345–351. [PubMed: 10877288]
20. R Core Team R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. Available from: <https://www.R-project.org/>
21. Lip GYH, Nieuwlaat R, Pisters R, Lane DA, Crijns HJGM. 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 Feb;137:263–272. [PubMed: 19762550]
22. Jauch EC, Saver JL, Adams HP, Bruno A, Connors JJB, Demaerschalk BM, et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke J Cereb Circ*. 2013 Mar;44:870–947.
23. Malik S, Hicks WJ, Schultz L, Penstone P, Gardner J, Katramados AM, et al. Development of a scoring system for atrial fibrillation in acute stroke and transient ischemic attack patients: the LADS scoring system. *J Neurol Sci*. 2011 Feb 15;301:27–30. [PubMed: 21130468]
24. Suissa L, Bertora D, Lachaud S, Mahagne MH. Score for the Targeting of Atrial Fibrillation (STAF) A New Approach to the Detection of Atrial Fibrillation in the Secondary Prevention of Ischemic Stroke. *Stroke*. 2009 Aug 1;40:2866–2868. [PubMed: 19461041]
25. Favilla CG, Ingala E, Jara J, Fessler E, Cucchiara B, Messé SR, et al. Predictors of finding occult atrial fibrillation after cryptogenic stroke. *Stroke J Cereb Circ*. 2015 May;46:1210–1215.

26. Brunner KJ, Bunch TJ, Mullin CM, May HT, Bair TL, Elliot DW, et al. Clinical predictors of risk for atrial fibrillation: implications for diagnosis and monitoring. *Mayo Clin Proc.* 2014 Nov. 89:1498–1505. [PubMed: 25444486]
27. Suzuki S, Sagara K, Otsuka T, Kano H, Matsuno S, Takai H, et al. Usefulness of frequent supraventricular extrasystoles and a high CHADS2 score to predict first-time appearance of atrial fibrillation. *Am J Cardiol.* 2013 Jun 1.111:1602–1607. [PubMed: 23499279]
28. Seet RCS, Friedman PA, Rabinstein AA. Prolonged rhythm monitoring for the detection of occult paroxysmal atrial fibrillation in ischemic stroke of unknown cause. *Circulation.* 2011 Jul 26.124:477–486. [PubMed: 21788600]
29. Sposato LA, Cipriano LE, Saposnik G, Ruiz Vargas E, Riccio PM, Hachinski V. Diagnosis of atrial fibrillation after stroke and transient ischaemic attack: a systematic review and meta-analysis. *Lancet Neurol.* 2015 Apr.14:377–387. [PubMed: 25748102]
30. 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 Lond Engl.* 2009 Feb 28.373:739–745.
31. Melgaard L, Gorst-Rasmussen A, Lane DA, Rasmussen LH, Larsen TB, Lip GYH. Assessment of the CHA2DS2-VASc Score in Predicting Ischemic Stroke, Thromboembolism, and Death in Patients With Heart Failure With and Without Atrial Fibrillation. *JAMA.* 2015 Sep 8.314:1030–1038. [PubMed: 26318604]
32. Mitchell LB, Southern DA, Galbraith D, Ghali WA, Knudtson M, Wilton SB, et al. Prediction of stroke or TIA in patients without atrial fibrillation using CHADS2 and CHA2DS2-VASc scores. *Heart Br Card Soc.* 2014 Oct.100:1524–1530.
33. Goudis CA, Korantzopoulos P, Ntalas IV, Kallergis EM, Ketikoglou DG. Obesity and atrial fibrillation: A comprehensive review of the pathophysiological mechanisms and links. *J Cardiol.* 2015 Nov.66:361–369. [PubMed: 25959929]
34. Nead KT, Gaskin G, Chester C, Swisher-McClure S, Dudley JT, Leeper NJ, et al. Androgen Deprivation Therapy and Future Alzheimer's Disease Risk. *J Clin Oncol Off J Am Soc Clin Oncol.* 2016 Feb 20.34:566–571.

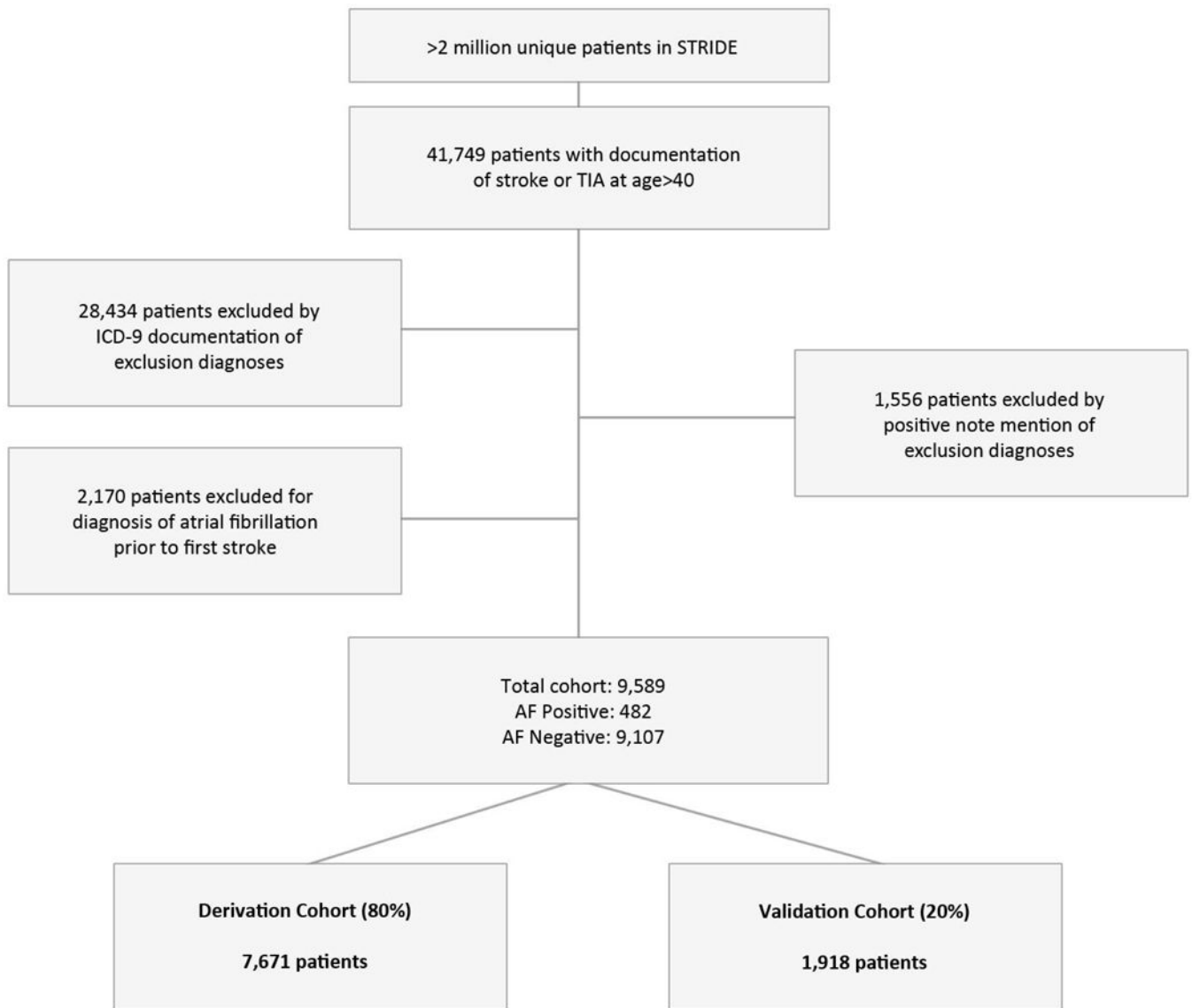


Figure 1. Flow chart of cohort selection using both ICD-9 codes and processed clinical notes. AF was defined by having ICD-9 diagnosis of 427.31 and 427.32 at least 30 days after CS/TIA.

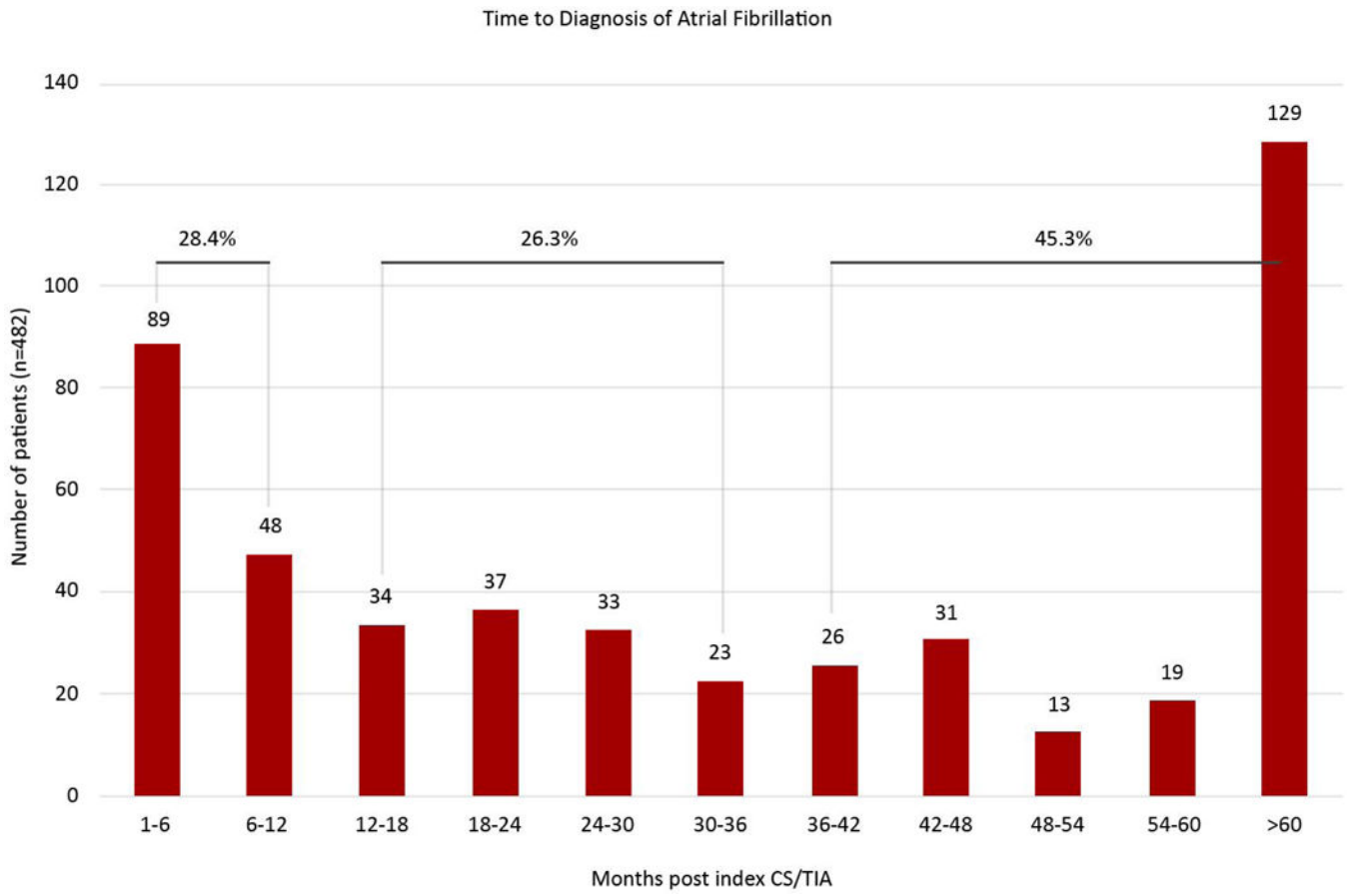


Figure 2. Distribution of when CS/TIA patients receiving diagnosis of AF by months. 28.4% of the patients received a new diagnosis of AF between 1–12 months post CS/ITA, 26.3% of patients at 1–3 years, and 45.3% at >3 years.

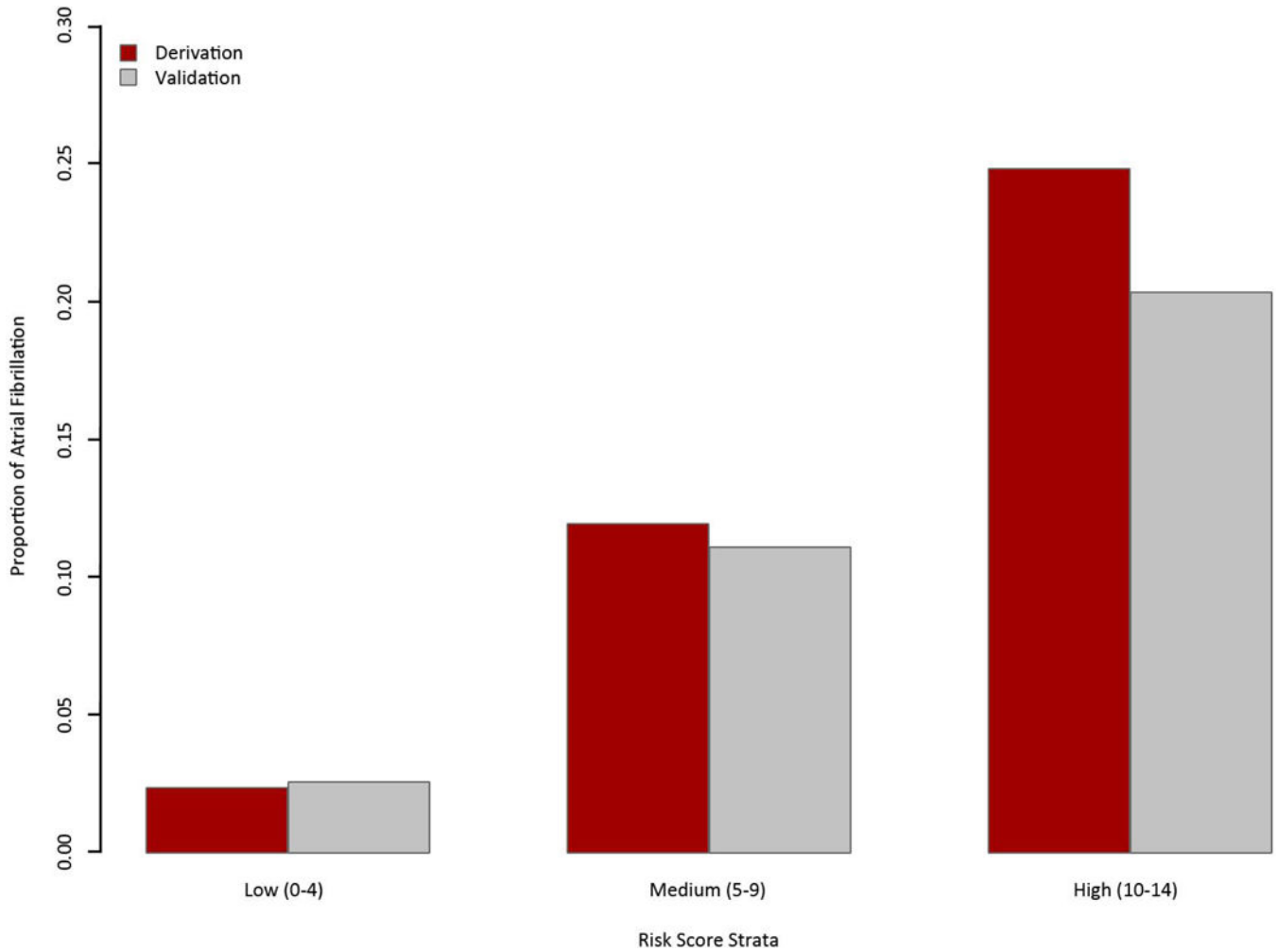


Figure 3.

Three risk strata based on HAVOC score. In the derivation cohort, AF rate in the low, medium, and high risk strata was 2.5%, 11.8%, and 24.9% respectively. In the validation cohort, AF rate in the low, medium, and high risk strata was 2.6%, 11.1%, and 20.3% respectively. There is a significant increase in risk between each stratum ($p < .0001$) by Cochran-Armitage trending test for both derivation and validation cohorts.

Table 1

Patient characteristics (derivation cohort only)

Characteristics	AF Positive (n = 390)	AF Negative (n = 7,281)	p-value
Age 75	320 (82.05%)	4031 (55.38%)	2.30E-22 *
Male	206 (52.82%)	3,703 (50.87%)	0.45
Hypertension	320 (82.05%)	4,031 (55.38%)	2.3E-22 *
Diabetes	128 (32.82%)	1,531 (21.03%)	5.4E-08 *
Obesity (BMI>30)	41 (10.51%)	391 (5.37%)	2.6E-05 *
Congestive Heart Failure	170 (43.59%)	743 (10.21%)	6.0E-69 *
Coronary Artery Disease	190 (48.72%)	1,391 (19.11%)	1.2E-39 *
Peripheral Vascular Disease	72 (18.46%)	585 (8.04%)	4.3E-12 *
Chronic Kidney Disease (III, IV, V)	23 (5.90%)	159 (2.18%)	6.9E-06 *
Valve Disease	94 (24.10%)	481 (6.61%)	4.9E-32 *
Aortic Valve Disease	49 (12.56%)	232 (3.19%)	9.7E-19 *
Mitral Valve Disease	55 (14.10%)	269 (3.70%)	4.0E-20 *
Tricuspid Valve Disease	11 (2.82%)	48 (0.66%)	1.3E-05 *
Pulmonary Valve Disease	4 (1.03%)	17 (0.23%)	7.7E-03 *

AF = atrial fibrillation; BMI = body mass index.

* P value <0.05 was considered significant

Table 2Significant ($p < 0.05$) risk factors from multivariable analysis in the derivation cohort

Predictor	Coefficient	OR (95% CI)	<i>p</i> -value	Score
Hypertension	0.70	2.01 (1.53–2.68)	1.10E-06	2
Age 75	0.55	1.73 (1.39–2.16)	8.32E-07	2
Valve Disease	0.72	2.05 (1.55–2.69)	3.25E-07	2
Vascular Disease (Peripheral)	0.32	1.37 (1.02–1.84)	3.49E-02	1
Obesity	0.42	1.53 (1.05–2.18)	2.24E-02	1
Congestive Heart Failure	1.21	3.34 (2.61–4.28)	1.70E-21	4
Coronary Artery Disease	0.54	1.72 (1.35–2.19)	1.08E-05	2

OR = odds ratio.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3Comparing low risk categories (≤ 4 points) of HAVOC vs. CHA₂DS₂-Vasc Scores

	HAVOC	CHA ₂ DS ₂ -Vasc
Sensitivity	0.55	0.77 *
Specificity	0.82	0.55 *
PPV	0.14	0.096
NPV	0.97	0.98
Accuracy	0.80	0.56 *

PPV = Positive Predictive Value, NPV = Negative Predictive Value

* p value < 0.001

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript