Validation of Clinical Classification Schemes for Predicting Stroke
Results From the National Registry of Atrial Fibrillation

Brian F. Gage, MD, MSc
Amy D. Waterman, PhD
William Shannon, PhD
Michael Boechler, PhD
Michael W. Rich, MD
Martha J. Radford, MD

The atrial fibrillation (AF) population is heterogeneous in terms of ischemic stroke risk. Subpopulations have annual stroke rates that range from less than 2% to more than 10%.1-3 Because the relative risk reductions from warfarin sodium (62%) and aspirin (22%) therapy are consistent across these subpopulations,2,6-8 the absolute benefit of antithrombotic therapy depends on the underlying risk of stroke. Although there has been agreement that warfarin therapy is favored when the risk of stroke is high and that aspirin is favored when the risk of stroke is low,9,10 there has been little agreement about how to predict the risk of stroke.11-13 Thus, an accurate, objective scheme to estimate the risk of stroke in the AF population would allow physicians and patients to choose antithrombotic therapy more judiciously.

The Atrial Fibrillation Investigators (AFI) pooled data from several trials to form a unified stroke classification scheme. Among trial participants who did not receive antithrombotic therapy, these researchers found that the risk of stroke increased by a factor of 1.4 per decade of age and by 3 clinical risk factors: hypertension, prior cerebral ischemia (either stroke or transient ischemic attack [TIA]), and diabetes mellitus (DM).2,8 There were 5.9 to 10.4 strokes per 100 patient-years among participants randomized to no antithrombotic therapy who had at least 1 of the 3 clinical risk factors.2,8 In contrast to...
these high-risk participants, Medicare-aged participants without any of these risk factors were at moderate risk of stroke, averaging 2.7 to 4.3 strokes per 100 patient-years. Participants younger than age 65 years who had none of the 3 risk factors were at low risk for stroke, averaging 1.0 to 1.8 strokes per 100 patient-years, and if they also lacked 2 equivocal stroke risk factors—coronary artery disease and congestive heart failure (CHF)—they had 0.0 to 1.6 strokes per 100 patient-years.12,16

Second, the schemes conflict: many participants lacked hypertension. Based on data from their first 2 trials, the SPAF investigators identified 4 independent risk factors for stroke: blood pressure higher than 160 mm Hg, prior cerebral ischemia, recent heart failure (ie, active within the past 100 days) or documented by echocardiography, or the combination of 75 years or older and being female.13 In SPAF III, participants with hypertension lacking these risk factors had an annual stroke rate of 3.2 to 3.6 per 100 patient-years of aspirin therapy13 and the rate was only 1.1 in those who also lacked hypertension.

The Stroke Prevention and Atrial Fibrillation (SPAF) investigators reported their classification scheme from SPAF participants who were treated with aspirin therapy. Based on data from their first 2 trials, the SPAF investigators identified 4 independent risk factors for stroke: blood pressure higher than 160 mm Hg, prior cerebral ischemia, recent heart failure (ie, active within the past 100 days) or documented by echocardiography, or the combination of 75 years or older and being female.13 In SPAF III, participants with hypertension lacking these risk factors had an annual stroke rate of 3.2 to 3.6 per 100 patient-years of aspirin therapy13 and the rate was only 1.1 in those who also lacked hypertension.

The promulgation of 2 stroke-risk classification schemes (AFI and SPAF), each with cautions about how equivocal risk factors influenced the risk of stroke in low-risk participants, complicates the estimation of stroke risk. First, the 2 schemes conflict: many patients classified as low risk by one scheme are classified as moderate or high risk by the other.15-18,19 Second, the classification schemes are sometimes ambiguous. For example, into which SPAF risk group should one classify a patient who initially presented with a systolic blood pressure higher than 160 mm Hg but whose blood pressure is controlled on follow-up evaluation? Third, the development of both classification schemes was data driven, and therefore the schemes could have captured apparent risk factors that represented idiosyncrasies in the data set rather than true associations.17 Fourth, because the original schemes were based on trial participants whose average age was 69 years, their performance in older and frailer populations is not well characterized.12 Given these limitations, we decided to validate the 2 existing classification schemes and their variations10,18,19 in an independent sample.

Our goal was to find a convenient and accurate classification scheme to estimate stroke risk in a national registry of Medicare-aged patients who have nonrheumatic AF and were not prescribed warfarin at the time of hospital discharge.

**METHODS**

**Formation of CHADS2**

We amalgamated the 2 classification schemes to form a new stroke-risk index, CHADS2. We then assessed the predictive accuracy of the AFI, SPAF, and CHADS2 schemes using data from a registry of Medicare beneficiaries who had AF. To create the CHADS2 index, we included independent risk factors that were identified in either the AFI or SPAF schemes: prior cerebral ischemia, history of hypertension, DM, CHF, and age 75 years or older. We included a history of hypertension, rather than having blood pressure higher than 160 mm Hg, because even well-controlled hypertension is an independent risk factor for stroke.20 We included age 75 years or older rather than the combination of age 75 years or older plus female sex, because there is an age-related increase in stroke in both women and men.21,22 We included recent CHF exacerbation, rather than any CHF, because the former is similar to the CHF definition used in the SPAF scheme.

To create CHADS2, we assigned 2 points to a history of prior cerebral ischemia and 1 point for the presence of other risk factors because a history of prior cerebral ischemia increases the relative risk (RR) of subsequent stroke commensurate to 2 other risk factors combined.2,4,5,23,24 We calculated CHADS2 by adding 1 point each for any of the following—recent CHF, hypertension, age 75 years or older, and DM—and 2 points for a history of stroke or TIA. Thus, CHADS2 is an acronym for the risk factors and their scoring. For example, an 82-year-old (+1) patient who had hypertension (+1) and a prior stroke (+2) would have a CHADS2 score of 4.

**State-Specific Cohorts**

The NRAF data set contained anonymous patient records gathered by 5 quality improvement/peer review organizations (QIO/PROs) that serve 7 states (California, Connecticut, Louisiana, Maine, Missouri, New Hampshire, and Vermont). These QIO/PROs had assembled state-specific cohorts of patients with AF for quality improvement projects under the Health Care Quality Improvement Initiative of the Health Care Financing Administration.25 Using Medicare Part A claims records (MEDPAR), the QIO/PRO analysts used the appropriate International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code 427.31 in either a principal or secondary diagnosis to identify Medicare beneficiaries who may have had AF.

Through record review, including electrocardiographic and physician documentation, QIO/PRO reviewers confirmed the presence of chronic or recurrent AF during the index hospitalization: Medicare beneficiaries who had acute AF and beneficiaries who died during hospitalization were excluded. During their chart abstractions, QIO/PRO reviewers documented stroke risk factors, other comorbid conditions, and the antithrombotic therapy prescribed at hospital discharge. No additional charts were abstracted to create the NRAF dataset. Abstractors used standardized abstraction forms and statewide sampling techniques that had been adapted by each QIO/PRO participant, but QIO/PRO protocols were sufficiently similar to allow the statewide data sets to be combined. The 2 QIO/PRO reviewers who calculated the inter-rater reliability for the chart abstraction found agreement between abstractors of more than 90%.
Outcomes Assessment
The study outcome was hospitalization for ischemic stroke as determined by Medicare claims. To identify stroke from the MEDPAR data, we used the following ICD-9-CM codes in the primary position: 434 (occlusion of cerebral arteries), 435 (transient cerebral ischemia), and 436 (acute, but ill-defined, cerebrovascular disease). We did not use the ICD-9-CM code of 433 (occlusion and stenosis of precerebral arteries) because that code is used for asymptomatic carotid artery disease.  
We had a minimum of 365 days of follow-up claims for all Medicare beneficiaries, and we censored beneficiaries at the time of nonstroke death or at a maximum of 1000 days after the index hospitalization. For beneficiaries who experienced multiple strokes, we excluded events and patient-days of follow-up that occurred after the initial stroke.

Statistical Analyses
To calculate the stroke rate as a function of CHADS2, we used an exponential survival model.  
We used the survival model to measure how the hazard rate for stroke was affected by each 1-point increase in CHADS2 and by prescription of aspirin. We also used the model to predict the annual rate of stroke as a function of CHADS2 and of aspirin use. We confirmed the appropriateness of using an exponential survival model graphically (by plotting the negative of the logarithm of the survival curve vs time).  
We performed our survival analyses in SAS (Version 6.12; SAS Institute Inc; Cary, NC) using the LIFEREG and LIFETEST procedures.  
We calculated the RR reduction from aspirin therapy as 1 minus the relative hazard of prescribing aspirin (as obtained from the exponential model).

To determine the predictive validity of each of the 3 classification schemes, we performed additional time-to-event analyses. We censored deaths that were not accompanied by a hospitalization for stroke and then calculated the stroke rate based on patient-years of follow-up data for each risk group identified by each classification scheme. We calculated the 95% confidence interval (CI) of these rates using the binomial approximation. We quantified the predictive validity of the classification schemes by using the c statistic to test the hypothesis that these classification schemes performed significantly better than chance (indicated by a c statistic of 0.5).

To determine the predictive accuracy of the classification schemes, we used the bootstrap analysis to generate 95% CIs using the percentile method.  
We calculated the c statistic on a random sample of all patients 1000 times and then noted the 2.5% and 97.5% percentiles. We declared the classification schemes statistically significantly different if these CIs did not overlap.

RESULTS
Baseline Characteristics of the NRAF Cohort
Compared with participants in the AF trials, the 1733 patients in the NRAF cohort were more likely to be women and elderly and more often had stroke risk factors: history of CHF (56%), CHF as the reason for the index hospitalization (14%); history of hypertension (56%); DM (23%); and history of cerebral ischemia (25%) (TABLE 1). The mean CHADS2 score was 2.1 for the 1204 members of the NRAF cohort who were not prescribed any antithrombotic therapy and 2.3 for the 529 members who were prescribed aspirin.

Stroke Rate in the NRAF Cohort
The 1733 patients were followed up for a mean (median) of 1.2 (1.0) years. During the 2121 patient-years of follow-up, 94 NRAF patients were readmitted for an ischemic event (rate, 4.4 per 100 patient-years), 71 patients were admitted for a stroke as indicated by ICD-9-CM codes 434 or 436, and 23 patients were admitted for transient cerebral ischemia as indicated by ICD-9-CM code 435. We refer to all of these events as stroke for simplicity and because 8 of the 23 patients had a subsequent hospitalization with ICD-9-CM...
code 434 or 436. Of the 94 patients admitted for a stroke, 25 (27%) died within 30 days of the hospital admission.

The stroke rate was lowest among the 120 patients in the NRAF cohort who had a CHADS2 score of 0, a crude stroke rate of 1.2, and an adjusted rate of 1.9 per 100 patient-years without antithrombotic therapy (TABLE 2). The stroke rate increased by a factor of 1.5 (95% CI, 1.3-1.7) for each 1-point increase in the CHADS2 score (P<.001). Aspirin was associated with a hazard rate of 0.80 (95% CI, 0.5-1.3), corresponding to a nonsignificant 20% RR reduction in the rate of stroke (P = .27).

The AFI and the SPAF classification schemes also identified patients at low risk for stroke. The 303 patients (17%) in the NRAF cohort identified as low risk according to the SPAF scheme had a stroke rate of 0.74 (95% CI, 0.71-0.76); and CHADS2 score is calculated by adding 1 point for each of the following conditions: recent congestive heart failure, hypertension, age at least 75 years, or diabetes mellitus and adding 2 points for having had a prior stroke or transient ischemic attack. CI indicates confidence interval.

Accuracy of the Stroke Classification Schemes

The AFI scheme had a c statistic of 0.68 (95% CI, 0.65-0.71); the SPAF scheme, 0.74 (95% CI, 0.71-0.76); and CHADS2 score is calculated by adding 1 point for each of the following conditions: recent congestive heart failure, hypertension, age at least 75 years, or diabetes mellitus and adding 2 points for having had a prior stroke or transient ischemic attack. CI indicates confidence interval.

We included all cohort members aged 75 years or older from the AFI moderate-risk cohort, the stroke rate was 1.1 per 100 patient-years of follow-up, but only 130 patients (8%) of the NRAF population were in this cohort.

Accuracy of the Stroke Classification Schemes

The AFI scheme had a c statistic of 0.68 (95% CI, 0.65-0.71); the SPAF scheme, 0.74 (95% CI, 0.71-0.76); and CHADS2 score is calculated by adding 1 point for each of the following conditions: recent congestive heart failure, hypertension, age at least 75 years, or diabetes mellitus and adding 2 points for having had a prior stroke or transient ischemic attack. CI indicates confidence interval.

We included all cohort members aged 75 years or older from the AFI moderate-risk cohort, the stroke rate was 1.1 per 100 patient-years of follow-up, but only 130 patients (8%) of the NRAF population were in this cohort.

Accuracy of the Stroke Classification Schemes

The AFI scheme had a c statistic of 0.68 (95% CI, 0.65-0.71); the SPAF scheme, 0.74 (95% CI, 0.71-0.76); and CHADS2 score is calculated by adding 1 point for each of the following conditions: recent congestive heart failure, hypertension, age at least 75 years, or diabetes mellitus and adding 2 points for having had a prior stroke or transient ischemic attack. CI indicates confidence interval.

We included all cohort members aged 75 years or older from the AFI moderate-risk cohort, the stroke rate was 1.1 per 100 patient-years of follow-up, but only 130 patients (8%) of the NRAF population were in this cohort.

Accuracy of the Stroke Classification Schemes

The AFI scheme had a c statistic of 0.68 (95% CI, 0.65-0.71); the SPAF scheme, 0.74 (95% CI, 0.71-0.76); and CHADS2 score is calculated by adding 1 point for each of the following conditions: recent congestive heart failure, hypertension, age at least 75 years, or diabetes mellitus and adding 2 points for having had a prior stroke or transient ischemic attack. CI indicates confidence interval.

We included all cohort members aged 75 years or older from the AFI moderate-risk cohort, the stroke rate was 1.1 per 100 patient-years of follow-up, but only 130 patients (8%) of the NRAF population were in this cohort.

Accuracy of the Stroke Classification Schemes

The AFI scheme had a c statistic of 0.68 (95% CI, 0.65-0.71); the SPAF scheme, 0.74 (95% CI, 0.71-0.76); and CHADS2 score is calculated by adding 1 point for each of the following conditions: recent congestive heart failure, hypertension, age at least 75 years, or diabetes mellitus and adding 2 points for having had a prior stroke or transient ischemic attack. CI indicates confidence interval.

We included all cohort members aged 75 years or older from the AFI moderate-risk cohort, the stroke rate was 1.1 per 100 patient-years of follow-up, but only 130 patients (8%) of the NRAF population were in this cohort.

Accuracy of the Stroke Classification Schemes

The AFI scheme had a c statistic of 0.68 (95% CI, 0.65-0.71); the SPAF scheme, 0.74 (95% CI, 0.71-0.76); and CHADS2 score is calculated by adding 1 point for each of the following conditions: recent congestive heart failure, hypertension, age at least 75 years, or diabetes mellitus and adding 2 points for having had a prior stroke or transient ischemic attack. CI indicates confidence interval.

We included all cohort members aged 75 years or older from the AFI moderate-risk cohort, the stroke rate was 1.1 per 100 patient-years of follow-up, but only 130 patients (8%) of the NRAF population were in this cohort.
...years without warfarin therapy in 66 patients classified as low risk according to the SPAF definition and observed a stroke rate of 2.4 per 100 patient-years without warfarin therapy in 47 patients classified as moderate risk according to the AFI scheme. Hellemans et al. found an annual rate of stroke of approximately 2.1 per 100 patient-years of aspirin therapy in an AF trial that excluded patients who had a prior stroke or TIA or age greater than 77 years. During long-term follow-up of 55 elderly patients with lone AF, Kopecky et al. found a stroke rate of 0.9 per 100 patient-years, despite a low use of warfarin. These studies support the premise that by applying a classification scheme, clinicians can identify AF patients who are at low risk for stroke even without warfarin therapy.

Because the net benefit of antithrombotic therapy correlates with the underlying risk of stroke, CHADS2 may be helpful in several clinical settings. For example, in identifying low-risk patients, a CHADS2 score of 0 defines an AF population who should be offered the option of aspirin therapy. In addition, CHADS2 could aid in decision making for patients with AF and who are undergoing surgical or dental procedure because perioperative management depends on their risk of hemorrhage from the procedure compared with the underlying thrombotic risk. Also, in patients for whom taking warfarin would be burdensome, CHADS2 could facilitate risk stratification and selection of antithrombotic therapy based on a patient-specific risk of stroke.

Our study has several strengths. First, we used chart review, rather than ICD-9-CM claims, to document the presence of AF and to identify the stroke risk factors. These chart reviews also identified patients who were discharged from the hospital and received aspirin, enabling adjustment for the prescription of aspirin in our calculations of the CHADS2-specific stroke rates. Because of the number of strokes (94), we were able to calculate stroke rates with precision (Table 2). The NRAF cohort included Medicare beneficiaries from 7 states that represented all geographic regions of the United States. Because we formulated CHADS2 based on previous studies, rather than on the NRAF data set, our study validates CHADS2. In addition, because we validated CHADS2 in Medicare beneficiaries who were recently hospitalized, rather than in healthier trial participants, CHADS2 should be generalizable to frail or elderly patients who have nonrheumatic AF.

The CHADS2 scheme with either definition of CHF that we tested in CHADS2—any CHF as identified on chart review and CHF identified as the principal diagnosis by ICD-9-CM code—had a c statistic of 0.82. We used the later definition in CHADS2 because it was closer to the definition of CHF that was an independent predictor of stroke in other studies: CHF that caused symptoms within the past 100 days was an independent predictor of stroke in SPAF, and moderate or severe left-ventricular systolic dysfunction on echocardiography was an independent predictor of stroke in both SPAF and AFI. We did not have access to echocardiographic results, which would have allowed us to assess how they could have improved the predictive accuracy of CHADS2 and SPAF.

Our study had several important limitations. First, NRAF cohort members were older and sicker than participants in clinical trials (Table 1), and the AFI and SPAF schemes may have performed better in a healthier population that included patients younger than 65 years. Because CHADS2 was based on the AFI and SPAF schemes, it too might have performed better in a younger or healthier population. Second, we used a single chart review to assess most of the stroke risk factors and had no way of capturing new risk factors if they developed. Third, we studied patients who had been hospitalized and who were not prescribed warfarin from our analyses. Future analyses are needed to evaluate CHADS2 and the other classification schemes in other populations. Fourth, we used Medicare claims to ascertain ischemic events and have no way to verify these events. Because Medicare claims...
cannot be expected to capture all strokes, our estimates of the CHADS2-specific stroke rates may be biased downward. Comparison of stroke rates that we observed with published rates (Table 3) suggests that this ascertainment bias is modest. In contrast, the CHADS2-specific stroke rates may significantly underestimate the stroke rate in patients who have high-risk conditions that we did not consider, such as mitral stenosis, cardiac thrombus, mechanical heart valve, recent anterior myocardial infarction, or high-grade carotid artery stenosis. Likewise, CHADS2 may underestimate stroke risk in patients with hypertension that exceeds 160 mm Hg,7,15 or in patients with a prior stroke or TIA that occurred in the previous 3 months.7,18 Although the 20% risk reduction for aspirin effectiveness in preventing stroke was not statistically significant in this study, it does have clinical significance when combined with other research. Our finding is consistent with a recent meta-analysis that estimated a 22% risk reduction for aspirin therapy but differed from a subgroup analysis that found no effectiveness of aspirin in AF populations aged 75 years or older.8,11 Thus, although our nonrandomized study could not determine the effectiveness of aspirin, our results suggest that aspirin therapy should be prescribed for elderly patients with AF who are not suitable candidates for warfarin.

In summary, CHADS2 is an easy-to-use classification scheme that estimates the risk of stroke in elderly patients with AF. Physicians and patients could use CHADS2 to make decisions about antithrombotic therapy based on patient-specific risk of stroke.

Author Contributions: Study concept and design: Gage, Rich, Radford. Acquisition of data: Gage, Waterman, Boechler, Radford. Analysis and interpretation of data: Gage, Waterman, Boechler, Rich, Radford. Drafting of the manuscript: Gage, Waterman. Critical revision of the manuscript for important intellectual content: Gage, Waterman, Shannon, Boechler, Rich, Radford. Statistical expertise: Gage, Shannon, Boechler. Obtained funding: Gage, Waterman, Radford. Administrative, technical, or material support: Waterman, Boechler.

Study supervision: Rich, Radford.

Funding/Support: This project was supported by grant R01-HS10133 from the Agency for Healthcare Research and Quality. The authors are grateful to the Health Care Financing Administration (HCFA) of the Department of Health and Human Services and to the 5 peer review organizations who provided the deidentified data set that made this research possible. The chart abstractions were performed as part of the Health Care Quality Improvement Initiative that was initiated and sponsored by HCFA under HCFA contracts 900-99-C402 and Utilization and Quality Control Peer Review Organization for the state of California, California Medical Review Inc provided data on 549 Medicare beneficiaries from the period of 1993-1998, under HCFA contract 500-96-P612, Utilization and Quality Control Peer Review Organization for the state of Connecticut, Qualidigm provided data on 1998 Connecticut Medicare beneficiaries from the period of 1994-1997, under HCFA contract 500-99-L402, Utilization and Quality Control Peer Review Organization for the state of Louisiana, Louisiana Health Care Review Inc provided data on 531 Medicare beneficiaries from the period of 1996-1998, under HCFA contract 500-99-P549, Utilization and Quality Control Peer Review Organization for the state of Maine, New Hampshire, and Vermont, Northeast Health Care Quality Foundation provided data on 657 Medicare beneficiaries from the period of 1996-1998.

Disclaimer: The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of commercial products or organizations imply endorsement of them by the US government. The authors assume full responsibility for the accuracy and completeness of the data presented. The conclusions presented herein are solely of the authors and do not represent those of the 5 peer review organizations, AHRQ, or HCFA. Ideas and contributions to the authors engaging with the issues presented are welcome.

Acknowledgment: We thank David Nilasena, Robert A. Felberg, Lyn Dupré Oppenheimer, the Missouri Patient Care Review Foundation, and the Writer’s Group of the Division of General Medical Sciences at Washington University for their thoughtful review of a prior version of the manuscript. We are indebted to Jill Duncan for her statistical assistance.

An active field of science is like an immense anthill; the individual almost vanishes into the mass of minds tumbling over each other, carrying information from place to place, passing it around at the speed of light.
—Lewis Thomas (1913-1994)