Research Article

External Validation of SAFE Score to Predict Atrial Fibrillation Diagnosis after Ischemic Stroke: A Retrospective Multicenter Study

Miguel Quesada López, Laura Amaya Pascasio, Sara Blanco Madera, Jorge Pagola, Diana Vidal de Francisco, Elena de Celis Ruiz, Inmaculada Villegas Rodríguez, Joaquín Carneado-Ruíz, Juan Antonio García-Carmona, Diana Vidal de Francisco, Elena de Celis Ruiz, Inmaculada Villegas Rodríguez, Ana López Ferreiro, Álvaro Vázquez Parra, Blanca Fuentes Gimeno, Ricardo Jaime Rigual Bobillo, María Isabel López López, Íñigo Esain González, María Dolores Ortega Ortega, Marina Blanco Ruiz, Irene Pérez Ortega, Joaquín Carneado-Ruíz, and Patricia Martínez Sánchez

1Neurology, Hospital Universitario Torrecárdenas, Almería, Spain
2Neurology, Hospital Universitario Virgen de las Nieves, Granada, Spain
3Neurology, Hospital Universitari Vall d’Hebron, Barcelona, Spain
4Neurology, Hospital Universitario de León, León, Spain
5Neurology, Hospital La Paz Institute for Health Research-IdiPaz (La Paz University Hospital-Universidad Autónoma de Madrid, Spain
6Neurology, Hospital Universitario de San Cecilio, Granada, Spain
7Neurology, Hospital Universitario Puerta de Hierro, Madrid, Spain
8Neurology, Hospital Universitario Santa Lucía, Cartagena, Spain
9Emergency and Research Unit, Torrecárdenas University Hospital, 04009 Almería, Spain
10CIBER de Epidemiología y Salud Pública (CIBERESP), 28029 Madrid, Spain
11Instituto de Investigación Biosanitaria Ibs, 18012 Granada, Spain
12Faculty of Health Science, Health Research Center (CEINSA), University of Almería, Spain

Correspondence should be addressed to Patricia Martínez Sánchez; patrinda@ua.es

Received 26 May 2023; Revised 13 November 2023; Accepted 27 November 2023; Published 7 December 2023

Academic Editor: Domenico Maria Mezzapesa

Copyright © 2023 Miguel Quesada López et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction. The screening for atrial fibrillation (AF) scale (SAFE score) was recently developed to provide a prediction of the diagnosis of AF after an ischemic stroke. It includes 7 items: age ≥ 65 years, bronchopathy, thyroid disease, cortical location of stroke, intracranial large vessel occlusion, NT-ProBNP ≥ 250 pg/mL, and left atrial enlargement. In the internal validation, a good performance was obtained, with an AUC = 0.88 (95% CI 0.84-0.91) and sensitivity and specificity of 83% and 80%, respectively, for scores ≥ 5. The aim of this study is the external validation of the SAFE score in a multicenter cohort. Methods. A retrospective multicenter study, including consecutive patients with ischemic stroke or transient ischemic attack between 2020 and 2022 with at least 24 hours of cardiac monitoring. Patients with previous AF or AF diagnosed on admission ECG were excluded. Results. Overall, 395 patients were recruited for analysis. The SAFE score obtained an AUC = 0.822 (95% CI 0.778-0.866) with a sensitivity of 87.2%, a specificity of 65.4%, a positive predictive value of 44.1%, and a negative predictive
value of 94.3% for a SAFE score ≥ 5, with no significant gender differences. Calibration analysis in the external cohort showed an absence of significant differences between the observed values and those predicted by the model (Hosmer-Lemeshow’s test 0.089).

Conclusions. The SAFE score showed adequate discriminative ability and calibration, so its external validation is justified. Further validations in other external cohorts or specific subpopulations of stroke patients might be required.

1. Introduction

Despite an extensive etiological study, 26-40% of ischemic strokes remain without a well-defined cause [1–3], so they are called cryptogenic strokes (CS). Occult atrial fibrillation (AF), which is often paradoxical and asymptomatic, is an important underlying cause of CS [4] and, if detected, the risk of stroke recurrence would be drastically reduced with anticoagulant treatment [5]. Moreover, AF-related strokes are generally more extensive, causing higher health costs, higher lethality, and disability, as well as a higher recurrence rate [6].

The diagnosis of occult AF is currently a challenge, with different diagnostic devices available, from external recorders to subcutaneous insertable cardiac monitors (ICMs). In patients with CS, prolonged ECG monitoring (up to 36 months) with ICMs shows an incidence of occult AF of 30%, much higher than with conventional devices [7]. However, the number of patients who benefit from them is small due to the invasive procedure and the limited availability of technical, human, and economic resources. Patient selection based on well-known AF-risk factors may improve diagnostic yield and opportunities for treatment and prevention of stroke recurrence [8]. Recently, an ESC guideline on screening for subclinical AF after stroke or transient ischemic attack (TIA) of undetermined origin [9] strongly recommends extending cardiac monitoring time to more than 48 hours, although with low certainty of the evidence of the upper monitoring limit. In addition, it finds weak evidence at the moment to use biomarkers predictive of AF to identify patients at higher risk, recognizing this area as important for future research.

In this context, it is necessary to optimize tools that allow clinicians to predict which ischemic stroke/TIA patients are more likely to have occult AF. A predictive model of AF (property registration number 2106148086772) has been developed in a single-center retrospective cohort of patients with ischemic stroke treated at a tertiary stroke center, the screening for AF scale (SAFE score) [10]. This score combines the majority of parameters that have been described as being associated with AF (clinical, echocardiographic, analytical, and neuroimaging) and includes age, bronchopathy, thyroid disease, N-terminal pro-B-type natriuretic peptide (NT-ProBNP), left atrial enlargement (LAE), cortical topography of stroke, and intracranial large vessel occlusion (LVO) (Table 1). The score obtained a good performance in the validation cohort, with an AUC (area under the ROC curve) of 0.88 (95% confidence interval 0.84–0.91). Punctuation ≥ 5 was related to patients with paroxysmal AF with a sensitivity of 83%, a specificity of 80%, and a negative predicted value of 94%.

The present study is aimed at carrying out the external validation of the SAFE score in an independent multicenter cohort of patients, including a gender perspective.

2. Methods

2.1. Patient Selection. This multicenter retrospective study was conducted across eight comprehensive stroke centers in Spain. Regarding inclusion criteria, consecutive patients over 18 years of age who were admitted for ischemic stroke or TIA between January 2020 and February 2022 were selected. Patients from Torrecardenas University Hospital included in the internal validation cohort (those between January 2020 and May 2021) were excluded from this external validation cohort. Moreover, the enrollment period for each center was determined according to the availability of the necessary complementary studies to fulfill the inclusion criteria of the study.

All included patients underwent an etiological assessment, which involved an evaluation of cerebral arteries by angio-CT or neurosonological study, a minimum cardiac monitoring of 24 hours, a determination of NT-ProBNP levels during the admission, and an echocardiography/focused cardiac ultrasound either during admission or within the first 6 months after the index event.

Patients with known AF or those diagnosed with AF on the initial electrocardiogram (ECG), patients with other major cardioembolic sources (severe ventricular dysfunction, mechanical prosthetic valve, and rheumatic mitral stenosis), and patients with hemorrhagic strokes were excluded.

2.2. Data Collection and Variable Definition. The database of the study has been designed on the REDCap (Research Electronic Data Capture) platform [11, 12], an encrypted web platform for managing databases.

A diagnosis of AF was considered during admission or within 12 months after the ischemic event, with a tracing of at least 30 seconds of a heart rhythm without P waves and irregular RR intervals. The diagnosis of AF could have been made by the treating neurologist, by a cardiologist, or by any physician attending to the patient during the period under consideration.

The SAFE scale variables were defined as follows: [10] (a) age was scored when it was greater than or equal to 65 years; (b) for the diagnosis of bronchopathy, comorbidity with chronic obstructive pulmonary disease (COPD), obstructive sleep apnea (OSA), or bronchial hyperresponsiveness without meeting the criteria for COPD has been considered; (c) thyroid disease was noted when the patient presented with hyper- or hypothyroidism, excluding subclinical forms; (d) NT-ProBNP was scored when it was greater than or equal to 250 pg/mL (limits established in the internal validation); (e) LAE was defined as atrial enlargement of any degree, following the recommendations of the American Society of Echocardiography and the European Association of Cardiovascular Imaging; [13] (f) the cortical topography of stroke was defined as the nonlacunar involvement of the
From a total of 439 patients included in the database, 395 finally selected for the analysis (Figure 1). Patient characteristics, laboratory data, echocardiographic, and neuroimaging features are detailed in Table 2. These characteristics have also been assessed according to gender in Table S1 of the Supplementary material file. Considering the data obtained in the internal validation study [10], the main difference was a greater age in our population (patients with age ≥ 65 years, 68.1% versus 50.4% in the internal validation study). Of all the variables included in the SAFE score, only thyroid disease did not show a significant difference in this external cohort.

In terms of performance, a SAFE score of ≥5 points achieved a sensitivity of 87.2%, a specificity of 65.4%, a positive predictive value (PPV) of 44.1%, and a negative predictive value (NPV) of 94.3% for AF diagnosis. These parameters were similar between genders for the abovementioned cut-off point (sensitivity males 83%, females 91%; specificity males 66%, females 65%; PPV males 38%, females 52%; NPV males 94%, females 95%).

When the model was retested in the internal validation cohort using the repeated cross-validation method, the confusion matrix reported high values for sensitivity (0.87), precision (0.87), and accuracy (0.837) (Figure 2(a)). Further, the model was trained on the external multicenter cohort, achieving values in the confusion matrix that were again remarkable in terms of sensitivity (0.841), precision (0.861), and accuracy (0.775) (Figure 2(b)).

The discriminative ability, measured by the AUC, reported a value of 0.879 (95% CI 0.844-0.915) in the internal cohort versus 0.822 (95% CI 0.778-0.866) in the external cohort (Figure 3). The AUC in this external cohort did not differ significantly between genders (Figure S1, Supplementary material file). Concerning the calibration of the initial model (internal cohort) and that applied to the external cohort, in both cases, a nonsignificant result was obtained in the Hosmer-Lemeshow test (0.295 and 0.089, respectively) (Figure 4).

4. Discussion

This retrospective multicenter study provides evidence that the SAFE score is a valid tool for predicting the risk of hidden AF in ischemic stroke patients. The model is highlighted for its sensitivity and precision, as well as a very high NPV for a cutoff point of ≥5 points. Thus, it could be useful for optimizing resources for the diagnosis of occult AF in these patients.

Left atrial cardiomyopathy (LACM) was defined in 2016 as "any complex of structural, architectural, contractile, or electrophysiological changes affecting the atria with the potential to produce clinically relevant manifestations" [14]. A bidirectional relationship between LACM and AF is postulated, assuming in many cases that LACM precedes the onset of AF, and conversely, AF may act as a trigger for atrial remodeling [15]. Although the diagnostic criteria are not completely defined, factors such as NT-ProBNP, obstructive sleep apnea, or LAE are included. On the other hand, "lone AF" is described when no apparent explanation or underlying comorbidity can be identified, with a very low embolic risk (only a 1-2% cumulative 15-year risk of stroke) [16]. However, with aging or the concurrence of vascular risk factors, the embolic risk associated with this "lone AF" increases [14]. These facts emphasize the great importance of having a reliable risk scale that includes factors...
associated with atrial disease, such as those contained in the SAFE score.

Different scores described in the literature can be found with the same purpose as the present one [17], but only two of them reflect all the types of variables included in the SAFE score. The model published by Seo et al. [18] reports clinical (age), echocardiographic (left atrial size), laboratory (free fatty acid and triglycerides), and neuroimaging (susceptibility vessel sign, hemorrhagic transformation, and cortical stroke involvement) variables. It is noteworthy that no brain natriuretic peptide is included in the score since it is the analytical parameter most closely related to AF [19]. Furthermore, it has not been externally validated in an independent cohort in addition to the one initially published. In the internal validation, the performance of this model was described with a C-index of 0.908 (95% CI 0.887-0.930) [18]. Kneihsl et al. have recently published the Graz AF risk score [20], which also includes clinical (age), echocardiographic (left ventricular ejection fraction, LAE), laboratory (NT-ProBNP), neuroimaging (recurrent stroke under antiplatelet treatment, multiterritory brain infarct, and prior cortical/cerebellar infarction) factors, and electrocardiographic variables (supraventricular premature beats or atrial runs > 20 beats). This scale was not developed in a cohort of stroke patients but was based on a bibliographic search by the authors. The Graz AF risk score was validated in a single-center prospective cohort of CS, involving only 24 patients with AF, and achieved an AUC of 0.85 (95% CI 0.78-0.92).

Not all the scales published in the literature report external validations. In a recent systematic review [17], twelve of the seventeen scores included were externally validated. The STAF score [21] is probably the one that accumulates the most external validations, being a scale that incorporates four parameters (age, baseline NIHSS score, LAE, and absence of established vascular etiology). Although they have positive validations [22, 23], some of them show limited utility [24, 25] or show a reduced yield in a cohort of patients with CS [26]. Indeed, it is common to find a worse performance of the scales in external cohorts other than those of the original validation. The same applies to the HAVOC score [27], a scale with 7 items (age, hypertension, valve disease, peripheral vascular disease, obesity, congestive heart failure, and coronary artery disease), which also performed worse in a cohort of embolic strokes of undetermined source (ESUS) [28]. The same effect was found in the study by Chen et al. [29] with LADS [30] and iPAB [31] scores or in the study by Kneihsl et al. [20] with the CHADS2 [32] and ASSF [33] scores.

There are several limitations related to this study. First is the retrospective design with a moderate number of patients. Secondly, there could be significant heterogeneity in the etiological study performed in the different participant centers. Finally, other than the minimum required 24 h cardiac monitoring, only 23% of patients had an additional 24 h Holter and another 17.7% had a 28-day Holter (Table 2). This unavailability of prolonged cardiac monitoring in all patients may have resulted in the misdiagnosis of some cases of AF.

Figure 1: Flow diagram of study participants. CS: cryptogenic stroke; EKG: electrocardiogram; LVO: large vessel occlusion; NT-ProBNP: N-terminal pro-B-type natriuretic peptide.
Table 2: Patient characteristics (between brackets, data of the internal validation cohort [11]).

<table>
<thead>
<tr>
<th>Variable</th>
<th>All patients * (n = 395)</th>
<th>Atrial fibrillation * (n = 94)</th>
<th>Nonatrial fibrillation * (n = 301)</th>
<th>P value **</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median age (IQR), years</td>
<td>72 (20) [65]</td>
<td>77.5 (16) [74]</td>
<td>69 (22) [62]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age ≥ 65 years, n (%)</td>
<td>269 (68.1%) [50.4%]</td>
<td>87 (92.6%) [86.3%]</td>
<td>182 (60.5%) [40.2%]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>231 (58.5%) [67%]</td>
<td>47 (50%) [52%]</td>
<td>184 (61.7%) [71.2%]</td>
<td>0.044</td>
</tr>
<tr>
<td>Arterial hypertension, n (%)</td>
<td>246 (62.3%) [60%]</td>
<td>73 (77.7%) [74.5%]</td>
<td>173 (57.5%) [55.9%]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes mellitus, n (%)</td>
<td>117 (29.6%) [24.8%]</td>
<td>31 (33%) [21.6%]</td>
<td>86 (28.7%) [25.7%]</td>
<td>0.425</td>
</tr>
<tr>
<td>Dyslipidemia, n (%)</td>
<td>186 (47.1%) [41.5%]</td>
<td>52 (55.3%) [47.1%]</td>
<td>134 (44.7%) [39.9%]</td>
<td>0.071</td>
</tr>
<tr>
<td>Ischemic heart disease, n (%)</td>
<td>44 (11.2%) [7.2%]</td>
<td>18 (19.1%) [6.9%]</td>
<td>26 (8.7%) [7.3%]</td>
<td>0.005</td>
</tr>
<tr>
<td>Chronic renal failure, n (%)</td>
<td>38 (9.6%) [8.9%]</td>
<td>11 (11.7%) [13.7%]</td>
<td>27 (9%) [7.6%]</td>
<td>0.433</td>
</tr>
<tr>
<td>Bronchopathy, n (%)</td>
<td>50 (12.7%) [16.7%]</td>
<td>19 (20.2%) [29.4%]</td>
<td>31 (10.3%) [13.1%]</td>
<td>0.012</td>
</tr>
<tr>
<td>Thyroid disease, n (%)</td>
<td>35 (8.9%) [7.2%]</td>
<td>9 (9.6%) [13.7%]</td>
<td>26 (8.6%) [5.3%]</td>
<td>0.780</td>
</tr>
<tr>
<td>Previous ischemic stroke, n (%)</td>
<td>57 (14.6%) [13.5%]</td>
<td>16 (17.2%) [19.6%]</td>
<td>41 (13.8%) [11.7%]</td>
<td>0.411</td>
</tr>
<tr>
<td>NIHSS score on admission, median (IQR)</td>
<td>3 (6) [4]</td>
<td>5.5 (11.5) [7]</td>
<td>3 (6) [3]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Laboratory measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NT-ProBNP levels, median (IQR), pg/mL</td>
<td>214 (679) [186]</td>
<td>735 (1661) [784]</td>
<td>156 (436) [125]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>NT-ProBNP ≥250 pg/mL, n (%)</td>
<td>188 (47.6%) [41.7%]</td>
<td>71 (75.5%) [79.4%]</td>
<td>117 (38.9%) [31%]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Echocardiographic features</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left atrial enlargement, n (%)</td>
<td>159 (40.3%) [27.4%]</td>
<td>67 (71.3%) [59.8%]</td>
<td>92 (30.6%) [18.2%]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Left ventricular hypertrophy, n (%)</td>
<td>157 (40.3%) [39.3%]</td>
<td>47 (50.5%) [36.3%]</td>
<td>110 (37%) [40.2%]</td>
<td>0.021</td>
</tr>
<tr>
<td>Diastolic dysfunction, n (%)</td>
<td>89 (23%) [36%]</td>
<td>17 (18.7%) [34.3%]</td>
<td>72 (24.3%) [36%]</td>
<td>0.263</td>
</tr>
<tr>
<td><strong>Neuroimaging features</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cortical topography of stroke, n (%)</td>
<td>223 (56.5%) [60%]</td>
<td>73 (77.7%) [81.4%]</td>
<td>150 (49.8%) [53.9%]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Intracranial large vessel occlusion, n (%)</td>
<td>124 (31.4%) [32%]</td>
<td>48 (51.1%) [51%]</td>
<td>76 (25.2%) [26.5%]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Chronic cortical stroke, n (%)</td>
<td>92 (23.3%) [8.9%]</td>
<td>29 (30.9%) [17.6%]</td>
<td>63 (20.9%) [6.4%]</td>
<td>0.047</td>
</tr>
<tr>
<td><strong>AF assessment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median telemetry duration (IQR), days</td>
<td>2 (2) [4]</td>
<td>3 (2) [4]</td>
<td>2 (2) [4]</td>
<td>0.468</td>
</tr>
<tr>
<td>Median time until diagnosis (IQR), days</td>
<td>3 (13.5) [2]</td>
<td>3 (13.5) [2]</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>24 h Holter monitoring, n (%)</td>
<td>91 (23%) [17.2%]</td>
<td>15 (16%) [16.7%]</td>
<td>76 (25.2%) [17.3%]</td>
<td>0.062</td>
</tr>
<tr>
<td>28 days Holter monitoring, n (%)</td>
<td>70 (17.7%) [17.2%]</td>
<td>24 (25.5%) [18.6%]</td>
<td>46 (15.3%) [16.8%]</td>
<td>0.023</td>
</tr>
</tbody>
</table>

IQR: interquartile range; N.A.: not applicable; NT-ProBNP: N-terminal pro-B-type natriuretic peptide. *Brackets indicate absolute frequency or median, as appropriate. **Comparisons between the atrial fibrillation (AF) and non-AF groups of the external validation cohort. Significant values are highlighted in bold.

![Confusion matrix](a)

![Confusion matrix](b)

Figure 2: Repeated cross-validation method with k = 5 and 10 repetitions. (a) Confusion matrix of the model retested on the internal validation cohort. (b) Confusion matrix of the model trained on the external multicenter cohort. AF: atrial fibrillation diagnosis; no AF: absence of atrial fibrillation diagnosis.
The SAFE score showed adequate discriminative ability in the external cohort assessed by AUC, with no significant differences between the values obtained and those predicted in terms of calibration, so its external validation is justified. This score could be a useful tool for the identification of patients with ischemic stroke at high risk of occult AF, thus allowing to optimize the resources needed for its diagnosis. Further validations of the SAFE score in other external cohorts are warranted.

**5. Conclusion**

Figure 3: Receiver operating characteristic (ROC) curve showing the atrial fibrillation (AF) prediction by SAFE in the internal and external cohorts.

Figure 4: Graphic representation of the Hosmer-Lemeshow test showing nonsignificant differences between observed and predicted values in both cohorts.
cohorts or specific subpopulations of stroke patients, such as CS or ESUS, as well as with prolonged cardiac monitoring, might be required.

**Data Availability**

The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

**Ethical Approval**

The study protocol is reviewed and approved by the Ethics Committee of Almeria (Almeria Provincial Research Ethics Committee, ethics approval reference 122/2021).

**Disclosure**

This study is part of the Spanish Health Outcomes-Oriented Cooperative Research Networks (RICORS-ICTUS), Instituto de Salud Carlos III (Carlos III Health Institute), Ministerio de Ciencia e Innovación (Ministry of Science and Innovation), RD21/0006/0010 (Torrecardenas University Hospital), and RD21/0006/0012 (La Paz University Hospital).

**Conflicts of Interest**

The group of authors of the Torrecardenas University Hospital who proposed the present work (MQL, LAP, JGT, AAP, and PMS) overlaps with those who developed the SAFE score. However, these authors have deliberately avoided participating in the statistical analysis of the results, focusing instead on the design of the study, data collection, and drafting of the manuscript.

**Authors’ Contributions**

MQL, LAP, AAP, and PMS conceived and designed the methodology of the study. MQL, LAP, SBM, JP, DVF, ECR, IVR, JC-R, JGC, ALF, IEB, RRB, MLL, IEG, MOO, MBR, IPO, CLH, and BFG contributed to the collection of data. MQL, LAP, JGT, AAP, and PMS drafted the manuscript. All authors reviewed and approved the final version of the manuscript. Miguel Quesada López and Laura Amaya Pascasio contributed equally to this paper and shared the first authorship.

**Acknowledgments**

We acknowledge FIBAO (Biomedical Research Foundation) and Adrián Aparicio Mota for their assistance with statistical analysis. Adrián Aparicio (a FIBAO statistician) analyzed the collected data. The project was the winner of the IV research grant “STROKE PROJECT 2020” from the Spanish Society of Neurology.

**Supplementary Materials**

In the Supplementary material, there is a gender analysis, with Table S1 showing the baseline characteristics of the patients by gender and Figure S1 showing the ROC curves by gender in the external cohort and the comparison of both AUCs. (Supplementary Materials)

**References**


