Clinical scores used for the prediction of negative events in patients undergoing catheter ablation for atrial fibrillation

Falco Kosich, Katja Schumacher, Tatjana Potpara, Gregory Y. Lip, Gerhard Hindricks, Jelena Kornej

Atrial fibrillation (AF) is the most prevalent sustained cardiac arrhythmia in adults. Catheter ablation (CA) is one of the most important management strategies to reduce AF burden and AF-associated complications. In order to stratify the risk of adverse events and to predict treatment success in AF patients undergoing CA, several risk stratification scores had been developed during the last decade. The aim of this review is to provide an overview of the most important clinical risk scores predicting rhythm outcomes, electro-anatomical substrate and mortality in AF.

KEYWORDS
atrial fibrillation, biomarkers, mortality, electro-anatomical remodeling, recurrences, scores

1 | INTRODUCTION

Atrial fibrillation (AF) is the most prevalent sustained cardiac arrhythmia in adults. Catheter ablation (CA) is an important management strategy to reduce AF burden and AF-associated complications. In general, AF is associated with several cardio- and cerebrovascular complications, such as heart failure, stroke, and death. Furthermore, AF recurrence rates after single CA range from 30% to 50%, often requiring repeated CA and leading to increased treatment costs. Pathophysiological, electrical, and structural atrial remodeling plays an important role in AF pathogenesis and is associated with endothelial damage, inflammation, and fibrosis. Several pro-fibrotic blood biomarkers as well as electro-anatomical mapping during CA and magnetic resonance imaging (MRI) techniques had been shown to predict these remodeling processes.

In order to stratify the risk of negative outcomes in AF patients undergoing CA, several risk prediction scores had been developed. These risk scores were established on the basis of various clinical factors, such as age, gender, body mass index (BMI), AF type, left atrial (LA) size, heart failure, chronic kidney disease, and early AF recurrences (ERAF). Clinical risk scores can be categorized by their selection of predictive factors (ie, biomarker-based or clinical variable-based risk scores). First, there are several scores, such as CHADS2, CHA2DS2-VASc, and R2CHADS2 that were originally developed to predict thromboembolic events in AF patients. Later on, rhythm
outcome-specific prediction scores, such as ALARMEc, BASE-AF, APPLE, CAAP-AF, and MB-LATER, were introduced. Recently, the prediction of mortality (2MACE) or electro-anatomical substrate (DR-FLASH) had been investigated (Table 1). Furthermore, it had been shown that electro-anatomical substrate measured during CA or through MRI was associated with arrhythmia recurrences after CA in AF patients. Finally, blood biomarkers had become a promising tool for risk stratification and were included into several risk stratification tools (eg, ABC, AEQ).

The aim of this review is to provide an overview of the most important clinical risk scores predicting rhythm outcomes, electro-anatomical substrate, and mortality in AF patients undergoing CA for AF.

2 | SEARCH STRATEGY

Comprehensive electronic searches for relevant publications were performed in the PubMed database. For structural purposes, the literature research had been categorized according to different adverse events considered in this article. Major search terms were generated and combined with specific search terms for each event. The major search terms included “atrial fibrillation OR AF” AND “score OR risk OR index OR scheme OR ratio”. As specific search terms the following list has been used:

1. AND “recurrence”
2. AND “LVA OR low voltage area OR substrate OR AF nest OR atrial foci OR atrial premature depolarization (APD)”
3. AND “death OR mortality”

Studies were included when they reported the prediction of an outcome in AF patients using risk assessment tools. Two authors (F.K. and J.K.) screened all retrieved publications for qualifications by title and abstract screening and full text reviewing. By applying this search strategy, we considered a number of publications and clinical scores to be relevant references for this article.

3 | SCORES (TABLE 1)

3.1 | CHADS2/CHA2DS2-VASc

Both scores had been originally developed for stroke prediction in AF patients. The CHADS2 consisted of five variables: one point for congestive heart failure, hypertension, age ≥ 75, diabetes mellitus, and 2 points for previous stroke. The scoring range is from 0 to 6 points. The predictive value for different adverse events in AF patients was tested in several studies. Congestive heart failure, hypertension, age ≥ 75 (2 points), diabetes mellitus, previous stroke (2 points), vascular disease, age 65 to 74 years, and female sex were included to the CHA2DS2-VASc score. The scoring range is from 0 to 9 points.
3.2 | ALARMEc
The score was developed for the prediction of the arrhythmia recurrences after CA for AF, and ranges from 0 to 4 points; AF clinical type, left atrium size, renal insufficiency, metabolic syndrome, and cardiomyopathy were considered for the ALARMEc score (1 point each).

3.3 | BASE-AF2
The BASE-AF2 score was developed to predict recurrences in AF patients after cryoballoon ablation (CBA). BMI > 28 kg/m², atrial dilatation >40 mm, current smoking, early AF recurrence post-CA, duration of AF history of >6 years, and non-paroxysmal type of AF were included, each weighing 1 point.

3.4 | APPLE
The APPLE score includes age ≥ 65 years, persistent AF, impaired eGFR (<60 mL/min/1.73 m²), LA diameter ≥ 43 mm, EF < 50% (1 point for each variable). Therefore, a maximum of five points could be achieved. The APPLE score can be used for the prediction of electro-anatomical substrate and recurrences after first and repeated CA in AF patients.

3.5 | DR-FLASH
The DR-FLASH score was originally developed for the prediction of low voltage area (LVA). The clinical variables diabetes mellitus, renal dysfunction (assessed by using the Cockcroft-Gault formula), persistent form of AF, LA diameter > 45 mm, age > 65 years, female sex and hypertension were used in this score. Each variable scores 1 point, so the score's range is from 0 to 7 points.

3.6 | CAAP-AF
This score was developed to predict AF freedom after CA and ranges from 0 to 13 points. Coronary artery disease, left atrial diameter, age, presence of persistent, or long-standing AF, antiarrhythmics failed and female sex were included to CAAP-AF score.

3.7 | MB-LATER
Male sex, bundle brach block, left atrium ≥ 47 mm, clinical type of AF, and early recurrent AF (ERAF) were included in the MB-LATER score. Each variable scores 1 point. The MB-LATER score was developed to predict very late recurrences of AF (VLRAF) >12 months after CA.

3.8 | ATLAS
Age > 60 years (1 point), type of AF—non-paroxysmal (2 points), indexed left atrial volume (1 point for each 10 mL/m²) female sex (4 points) and current smoking (7 points) were detected as independent predictors for arrhythmia recurrences after CA and they were included in the ATLAS score.

3.9 | ABC death risk score
Age, heart failure in the clinical history, N-terminal pro-B-type natriuretic peptide, troponin-T, and growth differentiation factor-15 levels were included in the ABC death risk score.

3.10 | 2MACE
The 2MACE scoring ranges from 0 to 7 points and it is composed of 2 points each for metabolic syndrome and age > 75 years. The remaining variables (myocardial infarction [MI]/revascularization; congestive heart failure [EF < 40%] and thromboembolic events) are rated with 1 point each.

3.11 | LAGO
Five clinical items had been included to the LAGO score: AF phenotype, structural heart disease, CHA2DS2-Vasc score < 1, LA diameter and LA sphericity. Each item scores 1 point.

4 | RHYTHM OUTCOMES AFTER CATHETER ABLATION
Arrhythmia recurrences after medical treatment (invasive or pharmacological) can be categorized into early recurrences (ERAF), late recurrences (LRAF) or very late recurrences of AF (VLRAF). ERAF is defined as any atrial tachyarrhythmia occurring within 3 months after index procedure, LRAF refers to the recurrences between 3 and 12 months post-CA, and VLRAF denotes any atrial tachyarrhythmia recurrence, which occurs after 12 months after index procedure.

We identified eight risk assessment tools that had been applied for prediction of arrhythmia recurrences after radiofrequency ablation (RFA) or CBA for AF: CHADS2, ALARMEc, BASE-AF2, CAAP-AF, APPLE, MB-LATER, ATLAS, and LAGO (Table 2).

Of note, there were substantial differences in the respective derivation cohort size among the scores. Generally, the scores developed in the large cohorts (eg, ATLAS, APPLE, and CAAP-AF) could have better generalizability than scores derived in smaller cohorts (ie, BASE-AF2, ALARMEc, MB-LATER), but each score would require further validation. Although an external validation is an important quality criterion of the risk stratification tools, it was not performed for every score. Indeed, the ALARMEc, APPLE, and MB-LATER scores were the only scores that had been validated in several external cohorts.

The recurrence prediction using well-known stroke risk scores (ie, the CHADS2, CHA2DS2-VASc, and R2CHADS2 scores) has been tested in several studies. All three scores were compared to each other and showed only modest ability to predict the arrhythmia recurrence after CA for AF. However, CHADS2, CHA2DS2-VASc, and R2CHADS2 were inferior compared to scores originally developed for arrhythmia recurrence, such as the APPLE and MB-LATER scores, for example.

Of note, most of the arrhythmia outcomes-specific scores were developed to predict the ERAF and/or LRAF after AF CA, while the BASE-AF2 and MB-LATER scores use ERAF as a variable for the LRAF/VLRAF prediction and, therefore, cannot be used at baseline.
Chao et al\textsuperscript{11} investigated the predictive value of the CHADS\textsubscript{2} score based on a complex calculation,\textsuperscript{32} the introduced scores are mainly calculation are basic necessities. Except for the ALARME\textsubscript{c} score, which is defined variables, such as LA size, sex, or AF type, and easy score calculation is very important factors for its clinical relevance. The use of clearly before index CA. Finally, the simplicity and practicality of a score are very important factors for its clinical relevance. The use of clearly defined variables, such as LA size, sex, or AF type, and easy score calculation are basic necessities. Except for the ALARME\textsubscript{c} score, which is based on a complex calculation,\textsuperscript{32} the introduced scores are mainly feasible.

Of note, the BASE-AF\textsubscript{2} score had been developed in an AF cohort undergoing CBA. The reliability of this score in the recurrence prediction after RFA is not proven so far.

### 4.1 CHADS\textsubscript{2}

Chao et al\textsuperscript{11} investigated the predictive value of the CHADS\textsubscript{2} score for arrhythmia recurrences after catheter ablation. Two-hundred thirty-eight patients with paroxysmal AF undergoing radiofrequency catheter ablation had been included into analysis. This study demonstrated that LA diameter (hazard ratio [HR] 1.057, \(P < 0.001\)) and high CHADS\textsubscript{2} score (cut-off \(\geq 3\)) (HR 1.372, \(P < 0.001\)) were independent predictors for arrhythmia recurrence after CA which occurred after 12 months follow-up (VLRAF).\textsuperscript{11}

Several studies compared the CHADS\textsubscript{2} and CHA\textsubscript{2}DS\textsubscript{2}-VASc scores in terms of recurrence prediction after CA. Letsas et al\textsuperscript{33} demonstrated that the difference in receiver operating characteristics (ROC) between CHADS\textsubscript{2} (0.644) and CHA\textsubscript{2}DS\textsubscript{2}-VASc (0.627) did not reach significance (\(P > 0.05\)). However, even though their study had been based on a relatively small cohort (128 patients; median FU 16 months), both scores were able to predict recurrence after ablation effectively. In a cohort with 2179 patients (non-paroxysmal and paroxysmal), Jacobs et al had demonstrated that the CHA\textsubscript{2}DS\textsubscript{2}-VASc score is superior to the CHADS\textsubscript{2} score regarding the prediction of recurrence after catheter ablation.\textsuperscript{34}

### 4.1 CHADS\textsubscript{2}

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Participants</th>
<th>Scores</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chao et al</td>
<td>2012</td>
<td>Overall: 238 PAF: 238</td>
<td>CHADS\textsubscript{2}</td>
<td>A high CHADS\textsubscript{2} score (cut-off (\geq 3)) was an independent predictor for arrhythmia recurrence after CA</td>
</tr>
<tr>
<td>Wójcik et al</td>
<td>2013</td>
<td>Overall: 213 PAF: 99</td>
<td>ALARME\textsubscript{c}</td>
<td>ALARME\textsubscript{c} demonstrated better predictive value than CHADS\textsubscript{2} and CHA\textsubscript{2}DS\textsubscript{2}-VASc scores for the prediction of arrhythmia recurrences after repeated catheter ablation (CBA and RFA)</td>
</tr>
<tr>
<td>Canpolat et al</td>
<td>2013</td>
<td>Overall: 236 PAF: 188</td>
<td>BASE-AF\textsubscript{2}</td>
<td>BASE-AF\textsubscript{2} (\geq 3) points was an independent predictor for AF recurrences after CBA (AUC = 0.94; 95% CI: 0.89-0.97, (P &lt; 0.001))</td>
</tr>
<tr>
<td>Letsas et al</td>
<td>2014</td>
<td>Overall: 128 VPAF: 128</td>
<td>CHADS\textsubscript{2}, CHA\textsubscript{2}DS\textsubscript{2}-VASc</td>
<td>CHADS\textsubscript{2} (AUC = 0.644) and CHA\textsubscript{2}DS\textsubscript{2}-VASc (AUC = 0.627) scores (cut-off (&gt; 2)) reached moderate predictive accuracy</td>
</tr>
<tr>
<td>Wójcik et al</td>
<td>2014</td>
<td>Overall: 911 PAF: 528</td>
<td>ALARME\textsubscript{c}</td>
<td>ALARME\textsubscript{c} score was able to predict the outcome after multiple ablation procedures. Success Rate after procedure was increased in patients with low ALARME\textsubscript{c}</td>
</tr>
<tr>
<td>Kornej et al</td>
<td>2015</td>
<td>Overall: 1391 PAF: 710</td>
<td>APPLE</td>
<td>APPLE demonstrated better c-indices for the prediction of LRAF after RCA than CHADS\textsubscript{2} and CHA\textsubscript{2}DS\textsubscript{2}-VAS</td>
</tr>
<tr>
<td>Wójcik et al</td>
<td>2015</td>
<td>Overall: 378 PAF: 320</td>
<td>ALARME\textsubscript{c}</td>
<td>ALARME\textsubscript{c} score could be used for the selection of patients before CBA. In patients with (\geq 3) points (high risk) CBA should be avoided</td>
</tr>
<tr>
<td>Jacobs et al</td>
<td>2015</td>
<td>Overall: 2179 PAF: 1246</td>
<td>CHADS\textsubscript{2}, CHA\textsubscript{2}DS\textsubscript{2}-VASc</td>
<td>Both scores can be used for recurrence prediction after first catheter ablation. CHA\textsubscript{2}DS\textsubscript{2}-VASc score reached slightly better results</td>
</tr>
<tr>
<td>Paylos et al</td>
<td>2016</td>
<td>Overall: 128 Only PAF</td>
<td>ALARME\textsubscript{c}</td>
<td>Patients (PAF) with a low ALARME\textsubscript{c} score had an excellent long-term outcome after CBA (second generation)</td>
</tr>
<tr>
<td>Winkle et al</td>
<td>2016</td>
<td>Overall: 2062 PAF: 644</td>
<td>CAAP-AF</td>
<td>CAAP-AF score could be used for the prediction of AF freedom after ablation. A correlation between CAAP-AF score and amount of LA-scar (detected by magnetic resonance imaging) could be demonstrated</td>
</tr>
<tr>
<td>Kornej et al</td>
<td>2017</td>
<td>Overall: 379 PAF: 265</td>
<td>APPLE</td>
<td>APPLE score can be used for the prediction of arrhythmia recurrences after repeated catheter ablation. The APPLE score demonstrated better c-indices compared to CHADS\textsubscript{2} and CHA\textsubscript{2}DS\textsubscript{2}-VASc</td>
</tr>
<tr>
<td>Mujović et al</td>
<td>2017</td>
<td>Overall: 133 PAF: 92</td>
<td>MB-LATER</td>
<td>MB-LATER achieved the highest c-index for the prediction of VLRAF compared to APPLE, ALARME\textsubscript{c}, BASE-AF\textsubscript{2}, CHA\textsubscript{2}DS\textsubscript{2}-VASc, CHADS\textsubscript{2}, and HATCH</td>
</tr>
<tr>
<td>Mesquita et al</td>
<td>2017</td>
<td>Overall: 1934 AF: 1488</td>
<td>ATLAS</td>
<td>ATLAS score was able to identify high-risk patients for AF-relapse after first PVI, despite of AF type. There was n’t any comparison with other risk scores performed</td>
</tr>
<tr>
<td>Deng et al</td>
<td>2018</td>
<td>Overall: 1410 PAF: 1089</td>
<td>MB-LATER</td>
<td>MB-LATER was compared to six scores (HATCH, CHADS\textsubscript{2}, CHA\textsubscript{2}DS\textsubscript{2}-VASc, BASE-AF\textsubscript{2}, CAAP-AF, and APPLE. MB-LATER had been reached largest net reclassification improvement</td>
</tr>
<tr>
<td>Potpara et al</td>
<td>2018</td>
<td>Overall: 226 PAF: 142</td>
<td>MB-LATER</td>
<td>MB-LATER was compared with CAAP-AF, CHADS\textsubscript{2}, and CHA\textsubscript{2}DS\textsubscript{2}-VASc. MB-LATER and CAAP-AF reached modest predictive value for LRAF</td>
</tr>
<tr>
<td>Bisbal et al</td>
<td>2018</td>
<td>Overall: 243 PAF: 160</td>
<td>LAGO</td>
<td>Different cardiovascular imaging parameters had been included to the LAGO score. The LAGO score can be used for recurrence prediction after RFA and CBA</td>
</tr>
</tbody>
</table>

Abbreviations: AF, atrial fibrillation; CBA, cryoballoon ablation; PAF, paroxysmal AF; PVI, pulmonary vein isolation; ERAF, early recurrence AF; LRAF, late recurrence AF; VLRAF, very late recurrence AF; RFA, radiofrequency ablation.
4.2 | ALARMEc

The ALARMEc score has been developed in a cohort of 213 patients (73 RFA and 140 CBA patients). The authors reported the predictive value of the ALARMEc score for recurrences after repeat CA in comparison to the CHADS2 and CHA2DS2VASc scores. In the ROC curve analysis, ALARMEc (area under the curve [AUC] 0.657; P < 0.001) was superior to CHADS2 (AUC 0.533; P = 0.413) and CHA2DS2-VASc (AUC 0.519; P = 0.641). Furthermore, the same group of authors showed in a larger cohort of 911 patients with a longer follow-up of 60 months that patients with moderate and high ALARMEc strata benefited from multiple procedures or more extensive substrate modification. Of note, there are also controversial data regarding ALARMEc score showing that CBA should be avoided in patients with a high ALARMEc score (>2 points) because they have poor outcomes.

Interestingly, in another small study of 128 CBA patients, it was shown that a low ALARMEc score correlated with freedom from recurrent arrhythmia. Although ALARMEc is the only study addressing prediction of recurrences both within 3 to 12 months and > 12 months, the results of this study are difficult to interpret because of non-standardized definitions of renal dysfunction and metabolic syndrome used in that study.

4.3 | BASE-AF2

The BASE-AF2 score has been developed in a cohort of 238 patients with a follow-up of 30 months (median 20 months) after CBA. A higher BASE-AF2 score (≥3 points) was significantly associated with AF recurrences (HR 3.34, P = 0.001). Since ERAF is used as an independent predictor in this score, the BASE-AF2 score cannot be used for baseline prediction, before index CA. Similar to the ALARMEc study, some variables included in the BASE-AF2 score such as BMI were not in accordance with current definitions. Also, the unclear cut-off of AF duration >6 years complicates the assessment due to the fact that in some patients AF may begin with asymptomatic episodes.

4.4 | CAAP-AF

The CAAP-AF score has been developed in a large cohort of 1125 AF patients. Of 14 tested clinical variables, six factors (coronary artery disease, left atrial diameter, age, presence of persistent or longstanding AF, antifibrhythmics failed, and female sex) were significantly associated with AF recurrence after CA. A low CAAP-AF score (<4 points) was associated with a better long-term outcome after CA. The predictive value of CAAP-AF score was significantly superior to CHADS2 and CHA2DS2-VASc on the ROC curve analysis (AUC 0.634 vs 0.538 and 0.542; P < 0.001, respectively). Similar results (AUC 0.624, P < 0.001) were reported in an external validation cohort of 261 patients from the Vanderbilt University.

Moreover, the APPLE score has been shown to be useful for the prediction of rhythm outcome after repeat CA in The Leipzig Heart Center Ablation Registry. In comparison to CHADS2 (AUC 0.577, P = 0.037) and CHA2DS2-VASc (AUC 0.590, P = 0.015), the APPLE score showed significantly better prediction of arrhythmia recurrences (AUC 0.617, P = 0.002) than other scores. So far, the APPLE score has been validated in several external cohorts showing similar results as in the development cohort. The third external validation was performed by Mujović et al comparing APPLE with the MB-LATER score. Both scores showed reasonably good predictive ability in the ROC curve analysis (AUC 0.716, P = 0.002 vs AUC 0.782, P < 0.001) for the prediction of VLRAF.

Another comparison of the APPLE, MB-LATER, and DR-FLASH scores has been published recently. Kornej et al used data from two study groups: the BioAF cohort (Heart Center Leipzig), which consisted of 241 patients, and The Leipzig Heart Center AF Ablation Registry, which provided 873 patients for the validation cohort. Beside LVA prediction, the predictive value for LRAF had been analyzed. The APPLE score (OR 1.550; P < 0.001) was significantly associated with arrhythmia recurrence within 1 year after CA in the validation cohort. Of note, on multivariable analysis only the MB-LATER score (OR 1.747; P < 0.001) achieved slightly higher values within the validation cohort.

4.6 | MB-LATER

The MB-LATER score has been introduced to predict VLRAF after RFA in 133 patients who were free from recurrent arrhythmia within first 12 months after CA. After the development and internal validation (cohort of 39 patients), the score was compared to APPLE (AUC 0.716); BASE-AF2 (AUC 0.648), ALARMEc (AUC 0.671), HATCH (AUC 0.582), CHADS2 (AUC 0.555), and CHA2DS2-VASc (AUC 0.510) scores. According to this comparison, MB-LATER showed better predictive accuracy for VLRAF (AUC 0.782, P < 0.001) than the other scores. Of note, the APPLE score (AUC 0.716, P = 0.002) showed almost similar prediction as the MB-LATER score.

Recently, two external MB-LATER score validation studies have been published. First, Deng et al used a Chinese cohort of >1400 patients. They compared seven risk stratification scores (HATCH, CHADS2, CHA2DS2-VASc, BASE-AF2, APPLE, CAAP-AF, MB-LATER) regarding their predictive ability for LRAF after CA. The MB-LATER, APPLE, BASE-AF2, and CAAP-AF reached good predictive values with an AUC of 0.73, 0.74, 0.75, and 0.71, respectively (all P < 0.01). These scores were superior to HATCH (AUC 0.58, P < 0.01), CHADS2 (AUC 0.57, P < 0.01), and CHA2DS2-VASc (AUC 0.57, P < 0.01). However,
the MB-LATER score had the largest net reclassification index (NRI, for 30%-82.6%) and integrated discrimination index (IDI, for 2.6%-18.6%) in comparison to other scores.

Of note, APPLE score had been mentioned as an alternative to MB-LATER.\textsuperscript{31} It is noteworthy that a score value of ≥2 of both MB-LATER (HR 1.52, \(P < 0.01\)) and APPLE score (HR 1.35, \(P < 0.01\)) has been significantly associated with an increased risk (52.1% and 35.3%, respectively) for AF recurrences.\textsuperscript{31} The second external validation study had been performed by Potpara et al\textsuperscript{38} in a cohort of 226 patients. The MB-LATER score was compared to CHADS\textsubscript{2}, CHA\textsubscript{2}DS\textsubscript{2}-VASc, and CAAP-AF scores regarding their predictive ability for LRAF. Only MB-LATER (AUC 0.62, \(P = 0.003\)) and CAAP-AF (0.59, \(P = 0.024\)) significantly predicted arrhythmia recurrences, and MB-LATER showed the largest net benefit compared to the other scores\textsuperscript{38} in ROC analysis. In addition, also as shown in Dengs study, the MB-LATER cut-off value of ≥2 had reached the highest predictive ability for LRAF. In both external validation cohorts ERAF, which has been included in the MB-LATER score, was shown to be an independent predictor for LRAF, which may partly explain the good predictive values of the MB-LATER. Interestingly, MB-LATER included male sex as a risk factor for recurrences in contrast to the CAAP-AF\textsuperscript{17} and ATLAS\textsuperscript{28} scores which included female sex as a risk factor. The MB-LATER score good predictive ability for VLRAF could be partly explained by higher AF prevalence in men.\textsuperscript{40,41}

### 4.7 ATLAS

The ATLAS score has been developed in a cohort of 1934 AF patients undergoing first CA which were divided into a development and a validation cohort (50% each).\textsuperscript{28} ATLAS score classified patients into low (<6 points), intermediate (6-10 points) and high risk (>10 points) for arrhythmia recurrences. Patients were followed-up for 4.2 ± 2.7 years, and recurrent arrhythmia occurred in 22% of patients during follow-up. In the development group, AF recurrence rates were 8, 11, and 17%/year for low (<6 points), intermediate (6-10 points) and high-risk patients (>10 points), respectively (\(P < 0.001\)). In the validation group, AF recurrence rates were 8, 11, and 18%/year, respectively (\(P < 0.001\)). There were significant differences in hazard ratio (HR) between intermediate (1.10, \(P = 0.35\)) and high (1.6, \(P < 0.001\)) risk groups. The score showed good discriminative power (censored c-statistic of 0.75 in both cohorts). Comparisons among other risk stratification scores were not performed.\textsuperscript{28}

### 4.8 LAGO

Different cardiovascular imaging parameters, such as LA sphericity had been shown to be associated with AF. Based on this knowledge, Bisbal et al\textsuperscript{29} developed the left atrial geometry and outcome (LAGO) score in a multicenter study including 243 patients after first RFA or CBA. So far, the score is not validated in a larger cohort, further investigations are needed. Furthermore, the study cohort includes both RFA and CBA, the predictive value of the LAGO score for each ablation technique in particular is not shown.\textsuperscript{29}

### 5 PREDICTION OF LOW-VOLTAGE AREA/ELECTRO-ANATOMICAL SUBSTRATE

AF progression is related to electro-anatomical changes in atrial myocardium and indicates advanced atrial remodeling. LVA can be detected during CA and through MRI\textsuperscript{3} and was defined as any region with <0.5 mV\textsuperscript{23} during electro-anatomical voltage mapping. Voltage-guided substrate modification by targeting LVA in addition to circumferential pulmonary vein isolation (PVI) is more effective than conventional PVI ablation approaches concerning arrhythmia freedom after the ablation.\textsuperscript{42-44} Recently, Yagishita et al showed that a LA voltage cut-off of <1.1 mV for electro-anatomic voltage mapping in sinus rhythm was an independent predictor for recurrences in patients without LVA (<0.5 mV).\textsuperscript{45} Although LVA is an important risk factor for post-procedural AF, there are no standardized methods to predict LVA non-invasively before CA.\textsuperscript{33,44}

The predictive ability of CHADS\textsubscript{2}, DR-FLASH, APPLE, and MB-LATER for LVA are discussed in this section (Table 3).

#### 5.1 CHADS\textsubscript{2}

Chao et al demonstrated that electrophysiological properties of the atrium differ between CHADS\textsubscript{2} strata, in a cohort of 247 patients with paroxysmal AF. In this study, the authors compared atrial voltage and total activation time of right and left atrium within CHADS\textsubscript{2} strata.\textsuperscript{46} In this relatively small cohort, it could be demonstrated that a higher CHADS\textsubscript{2} score (>3 pts.) is associated with LVA.

#### 5.2 DR-FLASH

The DR-FLASH score is currently the only score developed specifically for the prediction of LVA. The derivation cohort included 238 patients (153 with persistent AF). LVAs were found in 66 (28%) patients, and the score showed a good predictive ability for LVA with cut-off of 3 points (c-statistic 0.801; \(P < 0.001\)). The DR-FLASH score has been validated in an external cohort and showed similar results (AUC was 0.767, \(P < 0.001\)). Furthermore, DR-FLASH showed also a predictive ability for AF recurrence (1.3-fold increase per 1 point, \(P = 0.020\)) post-CA. Patients with LRAF had also a high DR-FLASH score (cut-off >3).\textsuperscript{22} Of note, female sex—as a component of DR-FLASH score—was recently considered as a risk factor for AF substrate. Indeed, females have a 2-fold risk for LVA\textsuperscript{47} and an almost 3-fold increased risk for AF recurrence following CA.\textsuperscript{2} Females could present with clinical AF in a later stage of fibro-fatty infiltration, which could explain a higher presence of electro-anatomical substrate and worse rhythm outcomes after CA in female patients.\textsuperscript{27}

#### 5.3 APPLE

Recently, it has been shown that the APPLE score—originally developed for arrhythmia recurrences—could be also used to predict LVA.\textsuperscript{39} In a population of 214 patients, we showed that the APPLE score (OR 1.921, \(P < 0.001\)) and female sex (OR 2.283, \(P = 0.005\)) were independent predictors for LVA.\textsuperscript{39} Interestingly, although atrial natriuretic peptide (NT-proANP) was an independent predictor on
univariable analysis, there was no increase in the predictive value by adding NT-proANP to the APPLE score. Importantly, the APPLE score can be used for baseline prediction of recurrent AF post-CA and contribute to an individualized AF therapy. Its components such as an impaired ejection fraction (EF) and renal dysfunction were also associated with an electro-anatomical substrate.48,49

This correlation had also been shown in our recent study where the APPLE, DR-FLASH, and MB-LATER scores were compared regarding their predictive ability for LVA and recurrences.50 First, we analyzed this prediction in the BioAF cohort of 214 AF patients and then validated the results in a retrospective cohort from The Leipzig Heart Center AF Ablation Registry. While on univariable analysis all scores were significantly associated with LVA, on multivariable analysis only the APPLE (OR 1.789, \( P < 0.001 \)) and DR-FLASH scores (OR 2.144, \( P < 0.001 \)) remained significant predictors. However, the MB-LATER score (OR 1.445, \( P = 0.034 \)) and ERAF (OR 5.078, \( P < 0.001 \)), but not the APPLE score, were associated with LRAF on multivariable analysis.50 All scores were significantly associated with recurrences, but ERAF was the most powerful predictor for later rhythm outcomes. In summary, on multivariable analysis the APPLE score was associated with prediction of both LVA and arrhythmia recurrences, whereas, as expected, DR-FLASH score (“a substrate score”) showed the best prediction for LVA, but not for rhythm outcomes, and MB-LATER was significantly associated with rhythm outcomes, but not LVA.

6 | MACE AND DEATH

Several studies demonstrated that AF is associated with an increased risk of stroke, heart failure, or sudden cardiac death.51,52 Consequently, AF patients have a higher mortality rate compared to patients without AF regardless of gender and age.51,53 In addition, it had been demonstrated that an increased mortality rate in anticoagulated AF patients is mostly because of cardiovascular causes other than ischemic stroke.54,55

In this section, the CHADS2/CHA2DS2-VASc, ABC death risk score and the 2MACE score (Table 4) are discussed in terms of their relevance for death or major adverse cardiovascular events (MACE) prediction including fatal/non-fatal myocardial infarction, cardiac revascularization, and cardiovascular death.

6.1 | CHADS2/CHA2DS2-VASc

The value of CHA2DS2-VASc and CHADS2 score for death and stroke prediction in AF patients after catheter ablation had been investigated by Chao et al.52 Both scores had been compared in a cohort of 565 AF patients. CHA2DS2-VASc score had been shown slightly better results (AUC 0.830) than the CHADS2 score (AUC 0.785) in ROC analysis. However, the difference between both curves did not reach significance (\( P = 0.116 \)).52 Both scores had been proven their ability to predict death and thromboembolic events in AF patients after CA. Similar findings had been published by Jacobs et al.34 in 2015. The predictive ability of CHA2DS2-VASc and CHADS2 score for death and MACE after first CA had been investigated in a cohort of 2179 AF patients. After FU of 5 years, it had been shown that CHA2DS2-VASc (HR 1.16; \( P = 0.04 \)) and CHADS2 (HR 1.30; \( P = 0.02 \)) score were associated with MACE.

6.2 | 2MACE

The 2MACE score includes five clinical variables, and it has been developed for the prediction of major adverse cardiovascular events (MACE) in a derivation cohort of 1019 AF patients with oral

## Table 3

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Participants</th>
<th>Scores</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chao et al</td>
<td>2011</td>
<td>Overall: 247 PAF: 247</td>
<td>CHADS2</td>
<td>Higher CHADS2-score is associated with LVA prediction</td>
</tr>
<tr>
<td>Kosiuk et al</td>
<td>2015</td>
<td>Overall: 902 PAF: 545</td>
<td>DR-FLASH</td>
<td>The optimal cut-off value for LVA prediction was 3 points. DR-FLASH score was also associated with the prediction of arrhythmia recurrences after PVI</td>
</tr>
<tr>
<td>Kornej et al</td>
<td>2018</td>
<td>Overall: 214 PAF: 88</td>
<td>APPLE</td>
<td>APPLE score and NT-proANP were independent predictors for LVA before catheter ablation</td>
</tr>
<tr>
<td>Kornej et al</td>
<td>2018</td>
<td>Overall: 1114 PAF: 621</td>
<td>APPLE, DR-FLASH, MB-LATER</td>
<td>APPLE and DR-FLASH demonstrated robust predictive value for LVA in both study groups</td>
</tr>
</tbody>
</table>

Abbreviations: AF, atrial fibrillation; PAF, paroxysmal AF; PVI, pulmonary vein isolation.

## Table 4

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Participants</th>
<th>Scores</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jacobs et al</td>
<td>2015</td>
<td>Overall: 2179 PAF: 1246</td>
<td>CHADS2, CHA2DS2- VASc</td>
<td>Both scores were associated with MACE prediction after catheter ablation (FU 5 years) CHA2DS2-VASc (HR 1.16; ( P = 0.04 )) and CHADS2 (HR 1.30; ( P = 0.02 ))</td>
</tr>
<tr>
<td>Pastori et al</td>
<td>2016</td>
<td>Overall: 2108 et al PAF:n.a.</td>
<td>2MACE</td>
<td>The highest specificity and sensitivity for MACE could be reached by an value of three in the 2MACE score</td>
</tr>
<tr>
<td>Hijazi et al</td>
<td>2017</td>
<td>Overall: 23159 PAF: 4115</td>
<td>ABC death score</td>
<td>ABC death score was compared to CHA2DS2-VASc score and it achieved higher c-indices. Three biomarkers (growth differentiation factor 15, high sensitivity cardiac troponin T and N-terminal pro B-type natriuretic peptide) were used in the ABC death score</td>
</tr>
</tbody>
</table>

Abbreviations: AF, atrial fibrillation; PAF, paroxysmal AF; MACE, major adverse cardiovascular events.
antiocoagulation (OAK). The median follow-up was 24 months (IQR 13.9-46.3) or 2287 person-years. The MACE incidence rate was 3.4%/year (111 cases). The external validation cohort included 1089 AF patients who were treated with vitamin K antagonists. The cut-off value of 3 points in the 2MACE score showed the best combination of sensitivity and specificity to predict MACE, and the score (AUC value of 3 points in the 2MACE score showed the best combination of sensitivity and specificity to predict MACE, and the score (AUC 0.660) was superior to CHADS2 (AUC 0.660) and CHA2DS2-VASc (AUC 0.667). In an external validation cohort, the predictive ability of 2MACE score has been confirmed (AUC 0.66, P < 0.001).

Indeed, all three investigated scores (ie, the CHADS2, CHA2DS2-VASc, and 2MACE) showed similar ability for the prediction of MACE. In addition, it was shown that an optimal anticoagulation level also reduced the risk of MACE.

While number of participants within the study groups was relatively large, the incidence of MACE was only ~180 out of 1,019 patients. Therefore, the predictive value of the 2MACE score needs to be validated in further studies.

### 6.3 ABC death risk score

Recently, the ABC death risk score had been introduced to predict death in AF patients without CA treatment. The development (ARISTOTLE) and external validation (RE-LY trial) cohorts included over 23,000 patients. The average follow-up in derivation cohort was 1.9 years or 28,396 person-years. The external validation cohort was based on 16,794 person-years of follow-up. The incidence rate of cardiovascular death was 3.69 per 100 person-years (1047 events) in the development cohort and 3.54 per 100 person-years (594 events) in the external validation cohort.

The ABC score was compared to CHA2DS2-VASc score in both cohorts and different subgroups. The ABC score showed a better predictive ability for mortality risk (AUC 0.75) than CHA2DS2-VASc (AUC 0.58) in all cohorts and subgroups.

Presently, these are the largest analyses using blood biomarkers (ie, cardiac troponin T, growth differentiation factor-15, and NT-proBNP) in AF patients. The usefulness of biomarkers as an important tool for risk prediction was also shown in several other studies.

### 7 CONCLUSION AND FUTURE DIRECTIONS

The major purpose of this review article was to provide an overview of the current scores for diverse negative events in AF patients undergoing catheter ablation and to discuss them critically. The development of the “ideal score” for prediction negative outcomes in AF patients still remains a clinical unmet need. Therefore, choosing only one optimal scoring system seems impossible, regarding that all of the presented scores in this review have their strengths and limitations. We therefore recommend to consider several important quality criteria for risk stratification scores: (a) size of development cohort, (b) an external validation, (c) clinical relevance, and (d) simplicity and practicality of each score. Moreover, the possibility of baseline prediction upfront the catheter ablation plays an important role due to a better feasibility.

However, there are several scores useful for the prediction of at least one adverse outcome. While DR-FLASH and APPLE scores are useful for prediction of both LVA and recurrences, the MB-LATER and ABC death scores predict recurrences within first year after ablation as well as ≥12 months (very late outcomes). Nevertheless, further investigations are needed to develop a universal score for patients undergoing AF catheter ablation. Biomarkers as well as cardiovascular imaging could improve the existing scores leading to better predictive values.

### CONFLICTS OF INTEREST

The authors declare no potential conflict of interests.

### ACKNOWLEDGMENT

We acknowledge support from the German Research Foundation (DFG) and Universität Leipzig within the program of Open Access Publishing

### ORCID

Falco Kosich [https://orcid.org/0000-0002-8552-8888](https://orcid.org/0000-0002-8552-8888)

### REFERENCES


