The validity and reliability of the Turkish version of the University of Toronto Atrial Fibrillation Severity Scale

Nihan KAHYA EREN1,*, Selcen YAKAR TÜLÜCE1, Barış KILIÇASLAN2, Cem NAZLI1, Oktay ERGENE3
1Cardiology Department, İzmir Kâtip Çelebi University Atatürk Education and Research Hospital, İzmir, Turkey
2Cardiology Department, İzmir Tepecik Education and Research Hospital, İzmir, Turkey
3Cardiology Department, Faculty of Medicine, Dokuz Eylül University, İzmir, Turkey

Background/aim: There are various instruments to assess quality of life (QoL) in patients with atrial fibrillation (AF). The aim of this study is to determine the reliability and validity of the Turkish version of the University of Toronto Atrial Fibrillation Severity Scale (AFSS).

Materials and methods: The AFSS and Short Form-36 (SF-36) were completed by 130 patients with documented AF. The Canadian Cardiovascular Society Severity in Atrial Fibrillation (SAF) scale and European Heart Rhythm Association (EHRA) scale were also utilized by the attending physicians. To assess test–retest reliability, the AFSS was readministered to 47 clinically stable patients at a 1-month follow-up visit. Internal consistency reliability, test–retest reproducibility, and construct validity were evaluated.

Results: The mean age of the patients was 63.1 ± 10.9 years and 58.5% of patients were male. The outcome scores of the Turkish version of the AFSS showed good correlations with theoretically related SF-36 domains. Additionally, AFSS outcome scores showed a linear correlation with the SAF and EHRA scores. Cronbach’s alpha values for internal consistency were consistent and similar with the English language version of the AFSS. Intraclass correlation coefficients for reproducibility exceeded 0.80 for every item.

Conclusion: Convergent–divergent and known-groups validity and reliability were established for the Turkish version of the University of Toronto AFSS.

Key words: Atrial fibrillation, health-related quality of life, validation, disease-specific questionnaire

1. Introduction

Atrial fibrillation is the most common chronic arrhythmia, affecting 1%–2% of the general population (1). Health-related quality of life (HRQoL) is significantly impaired in patients with AF when compared with healthy controls (2). Rate versus rhythm control trials have demonstrated that both of the treatment strategies have a similar effect on mortality and morbidity in patients with AF (3–5). Therefore, controlling the symptoms of AF and improving HRQoL are important determinants in making decisions for the appropriate treatment strategy.

Quality of life (QoL) is a subjective phenomenon; however, standardized questionnaires have been developed for accurate quantification of the perceived QoL (2). Two main types of questionnaires have been defined: a) generic, which are used to evaluate HRQoL in different populations, and b) disease-specific, which are used to evaluate patients with specific conditions and focus on items or symptoms related to the condition under study (6).

The University of Toronto Atrial Fibrillation Severity Scale (AFSS) is a disease-specific HRQoL questionnaire designed for patients with AF (7–9). It consists of 19 items combined into 3 parts to capture total AF burden, health care utilization, and the severity of AF-related symptoms. The AFSS was tested according to the standardized psychometric parameters of content validity, reliability, and variability (7,8).

The aim of this study was to determine the reliability and validity of the Turkish version of the University of Toronto AFSS in patients with AF.

2. Materials and methods

Patients with AF attending a cardiology outpatient clinic were invited to participate in the study; 130 patients who were willing to provide informed consent were recruited. Patients with AF of any classification were eligible for enrollment as long as their AF was documented by electrocardiogram (ECG) or Holter monitoring. We
excluded patients who were unwilling to participate in the study or were illiterate and unable to self-administer the questionnaires. Data on sociodemographic characteristics (age, sex, educational level, and working status), as well as clinical data including AF classification (paroxysmal, persistent, or permanent), treatment strategy (rate vs. rhythm control), current medications, underlying heart diseases, and other comorbidities were recorded. Twelve-lead ECGs and transthoracic echocardiography were performed.

The attending physicians provided a Severity in Atrial Fibrillation (SAF) classification and European Heart Rhythm Association (EHRA) classification for every patient. Subsequently, all patients completed the Short Form-36 (SF-36) generic HRQoL instrument and the disease-specific AFSS. The study protocol was approved by the local ethics committee.

2.1. Outcome measures
The Canadian Cardiovascular Society Severity in Atrial Fibrillation (SAF) is a scale utilized by physicians to assess the functional consequences of symptoms and to quantify the effect of AF on a patient's quality of life (10). SAF class ratings range from 0 (asymptomatic) to 4 (severe impact of symptoms on HRQoL and activities of daily living).

The EHRA classification is score for AF symptoms (11). It provides a simple quantification of symptoms that are attributable to functional consequences of AF. The 4 EHRA classes are defined from I (no symptoms) to IV (disabling symptoms).

The SF-36 is a widely used generic HRQoL scale with 36 items combined into 8 domains to measure physical functioning, role functioning, social functioning, mental health, vitality, pain, and general health perceptions (12). The Turkish version of the SF-36 has been previously validated (13).

The University of Toronto AFSS is a disease-specific QoL questionnaire. It is composed of 3 parts, A, B, and C. Part A includes questions regarding overall well-being (scored on a Likert scale from 0 to 10) and the frequency, duration, and overall severity of AF episodes. Part B includes questions regarding the presence and the frequency of the cardioversions, specialist appointments, emergency room visits, and hospitalizations within the past year, and part C is composed of questions regarding the presence and severity of individual symptoms attributable to AF (such as palpitations, dyspnea, dizziness, weakness, or chest pain). A measure of total AF burden is obtained by combining the measures of frequency, duration, and overall severity of AF episodes. Each of the 3 measures contributes equally, and each measure ranges from 1 to 10 to yield total AF burden scores ranging from 3–30. Higher scores indicate greater AF burden. Symptom severity is measured by summing the values of the questions in part C to yield a total score ranging from 0 to 35. Higher scores indicate more severe symptoms (7,9,14). The AFSS was professionally translated into Turkish with back-translations into English for verification for this study. Accepted translation strategies were used (15,16).

2.2. Validity and reliability
Convergent–divergent and known-groups validities were examined to establish the construct validity. Convergent validity establishes the correlation between 2 measures of different constructs that are theoretically related to each other. Divergent validity was used to demonstrate the poor correlation of theoretically unrelated constructs. For convergent and divergent validity, the SF-36 domains were correlated with AFSS outcome scores. Additionally, AFSS outcome scores were tested for differences among SAF and EHRA classes and were correlated with them for known-groups validity.

The reliability was assessed in terms of internal consistency, which measures the extent to which the items in the same construct are interrelated and test–retest reproducibility. Test–retest reproducibility was evaluated in clinically stable AF patients whose therapy was not changed. Clinically stable patients completed the AFSS at baseline and at a 1-month follow-up visit. Test–retest reliability was measured using these scores.

2.3. Statistical analysis
Patient sociodemographic and clinical characteristics were described as frequencies, means ± standard deviations, or median (minimum–maximum). Scores from each domain of the SF-36 and AFSS were reported as median (minimum–maximum). The correlations between AFSS outcome scores and the other outcome measures (SF-36, EHRA, and SAF classes) were determined by using Spearman correlations to test convergent, divergent, and known-group validity. An r value of >0.60 was considered to indicate a strong correlation, whereas r values between 0.35 and 0.60 indicated a moderate correlation. If the r value was <0.35, the correlation was considered to be weak (17,18).

The internal consistency of the questionnaire was measured by Cronbach's alpha and test–retest reliability was measured using intraclass correlation coefficient. The relationship between the AFSS outcome scores and the other variables (sociodemographic and clinical characteristics) was compared with the Mann–Whitney U test or the Kruskal–Wallis test.

3. Results
The mean age of the study group was 63.1 ± 10.9 years, and 58.5% of the patients (n = 76) were male. Sociodemographic and clinical characteristics of the patients are presented in Table 1. Paroxysmal AF was present in 30% of patients, and 16.2% (n = 22) were on antiarrhythmic drugs to maintain
sinus rhythm. Sixty-nine percent of the patients were on oral anticoagulation therapy (Table 1).

All patients completed AFSS in <5 min. Missing response rate was 2.3% for item 4 (global well-being), 7% for items 5 (AF frequency) and 6 (AF duration), and 3% for item 8 (perceived severity of the first AF episode). Response rate was 100% for all other items.

The relation between AFSS outcome scores and demographic and clinical variables was investigated. AFSS outcome scores (global well-being, total AF burden, symptom severity) were not related to sociodemographic and clinical variables (sex, family status, educational level, employment, AF pattern, comorbidities), except for age and presence of coronary artery disease. Older patients reported lower AF burdens (frequency, duration, and AF severity) than younger patients (r = –0.2; P = 0.04). Patients with coronary artery disease reported higher AF burdens compared to patients without coronary artery disease (17.0 ± 7.4 vs. 13.3 ± 6.7; P = 0.02).

The correlation between SF-36 domain scores and AFSS outcome scores is shown in Table 2. The global well-being subscale in the AFSS was correlated with all domains of the SF-36, but the correlation was stronger with the general health domain. There was a weak–moderate correlation between total AF burden and domains of the SF-36. On the other hand, symptom severity correlated at a moderate–high level with all domains of the SF-36 but had the strongest correlation with physical function and limitations in physical functioning.

AFSS outcome scores for every EHRA and SAF class are shown in Tables 3 and 4. The EHRA and SAF are similar scales, both of which quantify symptoms and their impact on patient's daily living and functionality. As expected, we observed a strong positive correlation between the SAF and EHRA classes rated by the physicians (r = 0.89, P < 0.001). AF frequency, duration, severity, and symptom severity scores increased in accordance with the increases in SAF and EHRA classes (Tables 3 and 4). Total AF burden and symptom severity were strongly correlated with the patient's SAF or EHRA class. There was a negative linear relationship between the global well-being subscale and EHRA and SAF classes. The number of emergency room visits and hospitalizations within the last year, which are theoretically related to disease burden, were also correlated with EHRA and SAF classes.

This scale has internal consistency (Cronbach α) of 0.88 for symptom severity, α = 0.75 for AF severity, α = 0.70 for AF burden, and α = 0.62 for health care utilization. The intraclass correlation coefficients were higher than 0.80 for every item, demonstrating a good test–retest reliability.

4. Discussion

Previous trials have failed to demonstrate the mortality benefit of a specific therapy in patients with AF, with the
exception of anticoagulation therapy (1). On the other hand, patients with AF have an impaired QoL that is comparable to postmyocardial infarction patients (8). Therefore, symptom reduction and improvement of the patient's well-being have become important objectives in the management of AF (1). A patient's well-being is usually expressed as QoL, which is a subjective phenomenon and usually does not correlate with the objective measures of disease severity (8). The most common approach to assess HRQoL is to apply standardized questionnaires. The University of Toronto AFSS has been validated to assess the impact of AF on quality of life (7) and has been used in clinical research (8,9,14,19). In this study, we have shown the validity and reliability of the Turkish version of the AFSS.

In our study, the first objective was the translation of the AFSS questionnaire, on which a consensus was obtained. The Turkish version of the AFSS was well perceived by Turkish patients with no need for any changes. Although the educational level of our study population was relatively low, response rates were acceptable. The response rate in part B and part C was 100%, while the response rate was 97.7% for item 4 (global well-being), 93% for items 5 (AF frequency) and 6 (AF duration), and 97% for item 8 (perceived severity of the first AF episode). The patients who did not respond to items 5 and 6 had permanent AF. These patients had been previously informed about their rhythm problem but had never been symptomatic due to AF. The patients who did not respond to item 8 reported that they could not remember the severity of the first AF episode.

Construct validity was obtained by correlating the AFSS outcome scores with the SF-36 and physician-estimated SAF and EHRA classes. The global well-being subscale showed consistency with the general health domain of the SF-36. The SF-36 has performed reasonably well in AF in previous studies, with the largest changes seen in scales related to

**Table 2. Correlation between AFSS scores and SF-36 scores.**

<table>
<thead>
<tr>
<th></th>
<th>Physical functioning</th>
<th>Role physical</th>
<th>Bodily pain</th>
<th>General health</th>
<th>Vitality</th>
<th>Social functioning</th>
<th>Role emotional</th>
<th>Mental health</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global well being</td>
<td>0.41**</td>
<td>0.38**</td>
<td>0.24**</td>
<td>0.49**</td>
<td>0.45**</td>
<td>0.37**</td>
<td>0.33**</td>
<td>0.44**</td>
</tr>
<tr>
<td>AF frequency†</td>
<td>–0.42**</td>
<td>–0.47**</td>
<td>–0.19*</td>
<td>–0.26**</td>
<td>–0.39**</td>
<td>–0.35**</td>
<td>–0.31**</td>
<td>–0.33**</td>
</tr>
<tr>
<td>AF duration†</td>
<td>–0.39**</td>
<td>–0.36**</td>
<td>–0.19*</td>
<td>–0.17</td>
<td>–0.26**</td>
<td>–0.32**</td>
<td>–0.23**</td>
<td>–0.25**</td>
</tr>
<tr>
<td>AF severity†</td>
<td>–0.36**</td>
<td>–0.41**</td>
<td>–0.28**</td>
<td>–0.19*</td>
<td>–0.34**</td>
<td>–0.36**</td>
<td>–0.35**</td>
<td>–0.35**</td>
</tr>
<tr>
<td>Total AF burden</td>
<td>–0.50**</td>
<td>–0.51**</td>
<td>–0.26**</td>
<td>–0.25**</td>
<td>–0.39**</td>
<td>–0.42**</td>
<td>–0.32**</td>
<td>–0.34**</td>
</tr>
<tr>
<td>Symptoms</td>
<td>–0.72**</td>
<td>–0.69**</td>
<td>–0.58**</td>
<td>–0.52**</td>
<td>–0.66**</td>
<td>–0.56**</td>
<td>–0.53**</td>
<td>–0.55**</td>
</tr>
</tbody>
</table>

*Correlation is significant at 0.05 level. **Correlation is significant at 0.01 level. The increasing scores for each AFSS subscale indicate increasing symptoms and severity except for global well-being, where higher scores indicate better well-being. For SF-36 domains, higher scores indicate better perceived health.

†AF frequency, AF duration, and AF severity are components of total AF burden.

**Table 3. AFSS outcome scores by EHRA classification and the correlation between the AFSS outcome scores and EHRA classification.**

<table>
<thead>
<tr>
<th></th>
<th>EHRA 1 (n = 32)</th>
<th>EHRA 2 (n = 53)</th>
<th>EHRA 3 (n = 31)</th>
<th>EHRA 4 (n = 14)</th>
<th>P-value</th>
<th>Correlation coefficient (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global well-being (1–10)*</td>
<td>8 (5–10)</td>
<td>7 (2–10)</td>
<td>5 (3–10)</td>
<td>6 (3–10)</td>
<td>&lt;0.001</td>
<td>–0.37 (&lt;0.001)</td>
</tr>
<tr>
<td>AF frequency (1–10)*</td>
<td>1 (1–5)</td>
<td>4 (1–10)</td>
<td>7 (1–10)</td>
<td>7 (2–10)</td>
<td>&lt;0.001</td>
<td>0.57 (&lt;0.001)</td>
</tr>
<tr>
<td>AF duration (1–10)*</td>
<td>1 (1–7.5)</td>
<td>2.5 (1–10)</td>
<td>6.2 (1.2–10)</td>
<td>6.2 (3.2–10)</td>
<td>&lt;0.001</td>
<td>0.62 (&lt;0.001)</td>
</tr>
<tr>
<td>AF severity (1–10)*</td>
<td>1.2 (1–10)</td>
<td>6 (1–10)</td>
<td>7 (1–10)</td>
<td>8.7 (4.5–10)</td>
<td>&lt;0.001</td>
<td>0.57 (&lt;0.001)</td>
</tr>
<tr>
<td>Total AF burden (3–30)*</td>
<td>3 (2–17.5)</td>
<td>14.2 (2–28.5)</td>
<td>17.9 (9.2–27)</td>
<td>21.9 (16.7–30)</td>
<td>&lt;0.001</td>
<td>0.70 (&lt;0.001)</td>
</tr>
<tr>
<td>Symptoms (0–35)*</td>
<td>2 (0–17)</td>
<td>10 (0–21)</td>
<td>16 (3–31)</td>
<td>24.5 (6–34)</td>
<td>&lt;0.001</td>
<td>0.74 (&lt;0.001)</td>
</tr>
</tbody>
</table>

Health care utilization

<table>
<thead>
<tr>
<th></th>
<th>Cardioversion (0–7)*</th>
<th>Emergency room visit (0–7)*</th>
<th>Hospitalization (0–7)*</th>
<th>Specialist visit (0–7)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 (0–3)</td>
<td>0 (0–4)</td>
<td>0 (0–5)</td>
<td>0 (0–5)</td>
</tr>
<tr>
<td></td>
<td>0 (0–1)</td>
<td>0 (0–5)</td>
<td>1 (0–6)</td>
<td>3 (0–7)</td>
</tr>
<tr>
<td></td>
<td>0 (0–1)</td>
<td>0 (0–3)</td>
<td>1 (0–5)</td>
<td>1 (0–5)</td>
</tr>
<tr>
<td></td>
<td>1 (0–5)</td>
<td>2 (0–7)</td>
<td>3 (0–7)</td>
<td>3 (0–7)</td>
</tr>
</tbody>
</table>

Data are presented as median (minimum–maximum). *The numbers indicate the range of scores.

Increasing scores indicate increasing symptoms and severity, except for global well-being, where increasing scores indicate better perceived well-being.
In accordance with this, we found a strong correlation between symptom severity and the SF-36 domains related to physical function. However, there was weak–moderate correlation between AF frequency, AF duration, and AF severity and SF-36 domains. This finding is also in line with previous studies (9). Dorian et al. reported that there is a subjective distinction between the AF burden of the AFSS and estimates of general quality of life measures (9).

Known-groups validity was demonstrated by relating the AFSS outcome scores with the EHRA and SAF classes. We observed a positive linear correlation with AFSS subscales (AF frequency, duration, severity, and symptom severity) and EHRA and SAF classes, whereas a negative correlation was found between global well-being and EHRA and SAF classes. EHRA and SAF classifications were strongly correlated with the total AF burden and symptom severity. Similarly, Dorian et al. demonstrated a positive linear correlation with symptom severity and SAF classification (10). The SAF scale evaluates data about severity, as well as frequency of subjective symptoms related to AF, to determine the impact of AF on patient’s well-being (20). Therefore, it is not an unexpected result that patient-perceived AF frequency and duration (21), different patient populations may be responsible for the disparity regarding the relation of AF frequency or duration and SAF classification.

Cronbach alpha values in our study are consistent and similar to those of the English version of the AFSS (7). Older age was related with a lesser AF burden in our study, which is in accordance with previous studies. Older patients are often less symptomatic than younger patients (22). Howes et al. showed that elderly patients with chronic AF have similar QoL and exercise capacity compared with age-matched controls in sinus rhythm (23). In our study, the presence of coronary artery disease was related with a higher AF burden. Previous studies found that the New York Heart Association (NYHA) functional class was an independent predictor for worse QoL in patients with AF (9,22). In this study we did not evaluate the NYHA classes of the patients, but left ventricular systolic dysfunction was significantly more prevalent in patients with coronary artery disease. Accordingly, the higher AF burden observed in patients with coronary artery disease might be related to lower ejection fraction and lower functional capacity.

The results of our study indicate that Turkish version of the AFSS is a reliable and valid instrument to assess impact of AF on patients’ QoL.

**Acknowledgments**

The authors gratefully acknowledge the statistical support of Bülent Özkan and the assistance of M Erdinç Arıkan for language editing.
References


