Validity and internal consistency of the American Orthopaedic Foot and Ankle Society Midfoot Scale in patients with Lisfranc injury

Ponkilainen, Ville; Uimonen, Mikko; Repo, Jussi; Mattila, Ville; Haapasalo Heidi

Abstract

Background: The American Orthopaedic Foot and Ankle Society (AOFAS) Midfoot Scale is an extensively used outcome measure instrument for evaluating outcomes after foot and ankle surgery or trauma.

Methods: In total, 117 patients with Lisfranc injury completed the AOFAS Midfoot Scale and the Visual Analogue Scale Foot and Ankle (VAS-FA) instruments. Internal consistency (correlation between different items), floor and ceiling values, convergent validity, item threshold distribution, and the coverage (item difficulty) of the AOFAS Midfoot Scale were tested.

Results: AOFAS Midfoot Scale had high convergent validity and acceptable internal consistency (Cronbach’s alpha >0.70). The ceiling effect was confirmed. The person-item distribution indicated that the scale had a lack of coverage and targeting in our sample.

Conclusions: Our data suggests that the AOFAS Midfoot Scale has acceptable validity and internal consistency. However, due to the lack of coverage and targeting, it should not be the primary outcome measure to be used to evaluate the outcomes after Lisfranc injury in the future studies.

Introduction

Injuries affecting the tarsometatarsal joint, also known as the Lisfranc joints, are relatively rare injuries (9/100 000/person-years) which can lead to pain and loss of function if inadequately treated [1, 2]. To date, there have only been two randomized controlled studies that have investigated the operative treatment of Lisfranc injury [3, 4]. The results of both of these studies suggest that primary arthrodesis might be a better long-term treatment option than open reduction and internal fixation (ORIF) [3, 4]. However, the problem with these studies is that various patient-reported outcome measures (PROMs) that were not specific to the foot were used to evaluate treatment outcomes.

The evaluation of clinical outcomes with rating scales has become common in the field of surgery [5-7]. The potential benefits of using outcome rating scales include benchmarking, comparing the outcomes between patients with similar foot and ankle conditions, and evaluating the
assessment of clinical outcomes after treatment [10-12]. For these reasons, at least 140 PROMs are used in foot and ankle surgery to provide the patient perspective [6, 8].

The American Orthopaedic Foot & Ankle Society (AOFAS) Clinical Rating Systems are one of the most widely used outcome measures for foot and ankle patients [6, 13]. Although the minimal important changes of the AOFAS Clinical Rating Systems have been defined, their validity and reliability have been questioned [11, 14-16]. Validity refers to the extent to which the scale measures what it is designed to measure, whereas reliability indicates the general consistency of the scale [17-20]. Hence, the Visual Analogue Scale Foot and Ankle (VAS-FA) was developed in 2006 to correct the flaws in the validity of the widely used AOFAS Midfoot Scale [16]. The VAS-FA has been validated and psychometrically tested for evaluating outcomes after foot and ankle surgery [16, 21].

In addition to the foot-specific PROMs, general health-related quality of life instruments, Main and Jowett criteria, radiographic evaluation, reoperation rate, return to sports and surgeons’ opinion have all been used to evaluate outcomes after a Lisfranc injury [3, 4, 22-25]. However, it may be advantageous to evaluate the outcomes with properly validated instruments developed for the specific clinical situation [17, 20]. In terms of practical use (if the patient has clinically significantly improved) as well as improving the quality of the studies (calculating the correct sample size), knowing the minimal important change of the instrument would be crucial [17, 26]. The aim of this study is therefore to test the validity and internal consistency of the AOFAS Midfoot Scale in patients treated for a Lisfranc injury.

Materials and Methods

The patients in this study were collected during a 5-year period (January 1, 2012 to December 31, 2016) in a Level One Trauma Center serving a catchment population of 500 000. The data used in this study was gathered from two studies: one retrospective and one prospective. The retrospective data were collected by reviewing all CT-scans that were performed due to an acute injury of the foot and ankle. All patients with a CT-verified Lisfranc joint injury (N=233) were included in the study. These patients were contacted via postal mail between 2 and 6 years after the injury. The prospective data were collected from a prospective trial, where patients were recruited directly from the emergency room. The PROMs used in the prospective study were completed at 12-month and 24-month follow-up visits. The demographic data of the study population are provided in Table 1. The recruited patients provided a written consent form for participation in the study according to the Declaration of Helsinki. The patients completed two foot and ankle-specific PROMs: the AOFAS Midfoot Scale and the VAS-FA [13, 16]. The study protocol was approved by the Regional Ethics Committee of the Hospital District.
Outcome measures

American Orthopaedic Foot and Ankle Society Midfoot Scale

The AOFAS Midfoot Scale is a hybrid outcome measure that can be reported either by clinician or patient and it has been developed to evaluate the pain and function of the foot [13]. The scale comprises 7 items, and each item has either three or four answer categories with various scorings [13]. The total score is calculated as a sum of all 7 items. If any of the items are missing, the total score cannot be calculated [13]. The total score ranges from 0 to 100 with a higher score indicating a better outcome [13]. The AOFAS scale is one of the most widely used outcome measure instruments in foot and ankle research [6, 8, 27]. The scale has not, however, been validated for midfoot-specific conditions. The Cronbach’s alpha has previously been found to be 0.59 [11].

Visual Analogue Scale Foot and Ankle

The VAS-FA is a foot and ankle-specific PROM that has been validated to assess pain, function, and other complaints [16, 21, 28, 29]. The scale contains 20 items scaled on a visual analog scale from 0 to 100 mm, with 0 indicating the worst, and 100 indicating the best result. The VAS-FA allows the items to be divided into three modules: Pain (4 items), Function (11 items), and Other complaints (5 items) [16]. The overall score and the scores of the modules are computed as the mean scores of the completed items of the instrument or its modules [16]. The normative VAS-FA scores for normal and various foot pathologies have been previously presented [30].

Statistical analysis

Clinical and demographic data are presented as medians and interquartile ranges (IQR) or as counts and percentages based on the distribution of the data. Hypotheses of the measured features were defined beforehand in accordance with the Consensus-based Standards for the selection of health status Measurement Instruments (COSMIN) checklist (Table 2) [18]. Floor and ceiling effects were assessed, and if more than 15% of the patients scored the minimum or the maximum points, the threshold was considered to have been achieved [31].

Convergent validity was evaluated by calculating Spearman correlation coefficients between the AOFAS Midfoot Scale and the VAS-FA. The correlation coefficients were interpreted according to the previous literature: 0.00 to 0.30 negligible, 0.30 to 0.50 low, 0.50 to 0.70 moderate, 0.70 to 0.90 high, and 0.90 to 1.00 very high correlation [32]. Linear regression analyses were used to evaluate the strength of the relationship between the instruments. Age-, and gender-standardized regression coefficient β indicates how strongly the score of the AOFAS Midfoot Scale predicts the total score of
the VAS-FA. The β values of .1, .3 and .5 were interpreted as small, moderate, and strong relationship, respectively.

Thresholds between the response categories of each item were investigated. The thresholds of the response category represent the location where there is a similar (50%) chance for the answer to end up in an adjacent response category.

To investigate scale targeting and coverage, a person-item distribution map was constructed to see how well the distribution of item difficulty matched with the coverage of the study sample within the AOFAS Midfoot Scale. The results of this analysis provided information on how well the scale performs in a distinct group of patients. The statistical analyses were performed using R (version 1.1.453) and SPSS (IBM® version 25.0) statistics software.

Results

The sample comprised 117 patients. The questionnaires were completed on average (SD) 3.9 (1.5) years after the injury. Altogether, 58 (50%) patients were treated nonoperatively and 59 (50%) operatively. The distribution of the AOFAS Midfoot Scale was skewed towards higher scores Figure 1. The ceiling effect was confirmed for the AOFAS Midfoot Scale because 30 (28%) of the patients scored maximum points (Table 2 and 3). For the VAS-FA Score, the ceiling effect was not confirmed because only 10 (9%) patients scored the maximum points. None of the patients scored the minimum points in either of the instruments, and therefore the floor effect was not confirmed. The VAS-FA and its subscales had high Cronbach’s alpha: 0.90 (Pain), 0.96 (Function), 0.82 (Other complaints), and 0.97 (Overall). The results indicate high internal consistency for the VAS-FA total score and its subscales. The AOFAS Midfoot Scale had Cronbach alpha of 0.75 (>0.70), indicating acceptable internal consistency.

There was a high correlation between the total scores of the instruments (r= 0.89) indicating good correspondence between the scores of the instruments (Figure 2A-C.). The correlations were also high between the Pain (r= 0.86) and Function (r= 0.77) subscales. All correlations were statistically significant (P< 0.001). The correlation between follow-up time and the AOFAS Midfoot Scale total score was negligible (Figure 3). The age- and sex-adjusted regression coefficient β of the VAS-FA subscales (Pain, Function, Other complaints, and Overall) against AOFAS Midfoot Scale total score were .83, .82, .80, and .87, respectively (Figure 4). The coefficients indicate a strong relationship between the VAS-FA and the AOFAS Midfoot Scale.

All items had ordered thresholds between the response categories (Figure 5A-C). Item 3 (“Maximum walking distance, blocks”) had only a narrow gap between the thresholds between the response categories 2 (“4-6”) and 3 (“1-3”). None of the patients gave the worst answers to items 1, 2, or 4.
The person-item distribution map shows that item difficulty matched well with the coverage of the study sample of the AOFAS Midfoot Scale (Figure 6). Many of the patients scored high scores, which was not covered by the instrument, and indicates that the instrument has deficiencies in its coverage and targeting for this patient group.

**Discussion**

High correlations and relationship with the VAS-FA indicated the strong convergent validity of the subscale (Pain, Function) scores of the AOFAS Midfoot Scale. In addition, the AOFAS Midfoot scale has acceptable internal consistency (Cronbach’s alpha, 0.75), which diverges from the results of a previous study that investigated patients with hallux valgus (Cronbach’s alpha, 0.59) [11]. In contrast, the ceiling effect suggests potential flaws in the coverage of the AOFAS Midfoot Scale among patients with Lisfranc injury. Similarly, the person-item distribution map also showed inappropriate coverage and targeting. In addition, there was negligible correlation between the follow-up time and the AOFAS Midfoot Scale total score, and the VAS-FA did not have the ceiling effect despite the long follow-up time (2-6 years). Therefore, the long follow-up time did not explain the ceiling effect. The main result of our study was that the AOFAS Midfoot Scale has high convergent validity and acceptable internal consistency, but the instrument had a notable drawback (ceiling effect and person-item distribution) concerning its coverage and targeting in the assessment of the outcomes after Lisfranc injury.

The differences between the VAS-FA and the AOFAS Ankle-Hindfoot Score in patients with ankle fractures has been evaluated in a previous study [33]. The finding of this study was that both instruments have a similar pattern to extract the functional outcome scores. However, they did not compare the psychometric properties with regard to Classical Test Theory (CTT) or Item Response Theory (IRT), which are the two methods used to compare the validity and reliability of the instruments [17]. The COSMIN checklist requires that the validity, reliability, and responsiveness of the PROM are assessed prior to applying the PROM in practice [17, 20]. Furthermore, once the PROM has been tested with the CTT methods, it should be further assessed with IRT methods [17, 20]. In our study, the principles of both CTT (internal consistency) and IRT (person-item distribution, thresholds between response categories) were combined.

The use of the AOFAS Clinical Rating Systems has been questioned, since their psychometric properties do not fulfill the acceptable criteria set for PROMs [8, 14, 34]. In addition, the score cannot be obtained if even one answer is missing [14]. Even the developers of the AOFAS Clinical Rating Systems suggest that the scale is not reliable, and that other outcome measures, such as the PROMIS Physical Function Computerized Adaptive Test (CAT) or Lower Extremity CAT combined with an additional pathology-specific instrument, should be considered [34]. In addition, it has been pointed out that the AOFAS Midfoot Scale uses non-specific expressions [16]. For example, gait abnormality
is assessed as “none, slight”, “obvious”, or “marked”, and alignment is defined as “good, plantigrade, well-aligned”, “fair, plantigrade, some degree of malalignment”, or “poor, non-plantigrade, severe malalignment”. In the present study, the thresholds between the response categories of each item were ordered and did not show significant malfunctions. Item 3 (“Maximum walking distance, blocks”) had relatively narrow thresholds for the responses 2 (“4-6”) and 3 (“1-3”), and therefore the answers could be united. Additionally, items 1, 2, and 4 did not receive any worst responses. This may have been due to the relatively long follow-up time of the patients. However, the properly ordered categories do not solve the problems of the non-specific explanations of the answer categories.

Despite these flaws, many of the previous studies investigating Lisfranc injuries have used the AOFAS Midfoot Scale as a primary outcome measure [3, 23, 35-39]. Based on the findings of the present study, it would seem that the AOFAS Midfoot Scale has an imbalance of difficult and easy items, and therefore it does not differentiate the patients well enough. The term “difficult items” refers to those items that need higher levels of the latent trait to achieve high scores, whereas the “easy items”, in contrast, can provide high scores even at lower levels of the latent trait. Since we observed deficiencies concerning the scale’s coverage and targeting, the results of this study suggest that the previous studies that used the AOFAS Midfoot Scale might have missed some information on less symptomatic patients due to the outcome measure used [3, 23, 35-39]. Other foot and ankle specific PROMs, such as the VAS-FA [16], the Lower Extremity Functional Scale (LEFS) [40, 41], the Foot and ankle ability measure (FAAM) [42], the Self-reported Foot and Ankle Score (SEFAS) [43], and the European Foot and Ankle Society (EFAS) score [44], might have psychometric properties that could potentially fill the gap that the AOFAS has in assessing outcomes in the treatment of foot and ankle injury. Future studies should therefore focus on assessing the measurement properties and minimal important change for the validated foot and ankle PROMs.

The strength of our study was the large group of patients with Lisfranc injury treated both nonoperatively and operatively. The limitations of the study were the cross-sectional study design, the use of only one reference outcome measure, and the lack of reproducibility testing (test-retest).

Conclusions

As a conclusion, the present study found that the AOFAS Midfoot Scale has high convergent validity and acceptable internal consistency when used to evaluate the long-term outcomes after treatment of Lisfranc injury. The scale seems to have deficiencies regarding its coverage and targeting, and there are flaws with the non-specific expressions of the responses. Based on the relatively high ceiling effect, the scale seems to be inappropriately targeted when assessing long-term outcomes in the treatment of Lisfranc injury. Because it is the most frequently used instrument in the published literature, this study provides information that can be used when interpreting the results of these previous studies. However,
it should not be the preferred instrument to be used as the primary outcome measure in patients with Lisfranc injuries in the future studies.
Table 1. Clinical information and distributions of the patient reported outcome measure scores of patients with Lisfranc injuries.

<table>
<thead>
<tr>
<th></th>
<th>N = 117</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean</td>
<td>41±17</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>75 (64)</td>
</tr>
<tr>
<td>Treatment, n (%)</td>
<td></td>
</tr>
<tr>
<td>Non-operative</td>
<td>58 (50)</td>
</tr>
<tr>
<td>ORIF</td>
<td>21 (18)</td>
</tr>
<tr>
<td>Arthrodesis</td>
<td>23 (20)</td>
</tr>
<tr>
<td>Multiple operations</td>
<td>12 (10)</td>
</tr>
<tr>
<td>Closed reduction with K-wire fixation</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Follow-up, mean months</td>
<td>46±18</td>
</tr>
<tr>
<td>AOFAS</td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>88 (73 - 100)</td>
</tr>
<tr>
<td>Floor, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Ceiling, n (%)</td>
<td>30 (28)</td>
</tr>
<tr>
<td>VAS-FA</td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>89 (72 - 98)</td>
</tr>
<tr>
<td>Floor, n (%)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Ceiling, n (%)</td>
<td>10 (9)</td>
</tr>
</tbody>
</table>

ORIF: Open Reduction and Internal Fixation  
±: Standard deviation  IQR: Interquartile range  
VAS-FA: Visual Analogue Scale Foot and Ankle  
AOFAS: The American Orthopaedic Foot & Ankle Society Midfoot Scale  
Floor: The number of patients who reached the minimum score  
Ceiling: The number of patients who reached the maximum score
Table 2. Predefined hypotheses for the validation of the American Orthopaedic Foot & Ankle Society Midfoot Scale.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Hypothesis</th>
<th>Result</th>
<th>Confirmed/ Rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Cronbach alpha is &gt; 0.70</td>
<td>0.75</td>
<td>Confirmed</td>
</tr>
<tr>
<td>Validity Coverage</td>
<td>Floor effect &lt; 15%</td>
<td>0%</td>
<td>Confirmed</td>
</tr>
<tr>
<td></td>
<td>Ceiling effect &lt; 15%</td>
<td>28%</td>
<td>Rejected</td>
</tr>
<tr>
<td>Convergent validity</td>
<td>Correlation with VAS-FA is ≥0.50</td>
<td>r=0.89</td>
<td>Confirmed</td>
</tr>
<tr>
<td></td>
<td>Correlation with VAS-FA Pain is ≥0.50</td>
<td>r=0.86</td>
<td>Confirmed</td>
</tr>
<tr>
<td></td>
<td>Correlation with VAS-FA Function is ≥0.50</td>
<td>r=0.79</td>
<td>Confirmed</td>
</tr>
<tr>
<td></td>
<td>Item difficulty matches with the coverage of the study sample</td>
<td>Good coverage</td>
<td>Rejected</td>
</tr>
</tbody>
</table>

VAS-FA: Visual Analogue Scale Foot and Ankle
Table 3. The mean scores and floor and ceiling values of the American Orthopaedic Foot & Ankle Society Midfoot Scale.

<table>
<thead>
<tr>
<th>Item</th>
<th>Response categories (points)</th>
<th>Mean (SD)</th>
<th>Floor (%)</th>
<th>Ceiling (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pain</td>
<td>4 (0-40)</td>
<td>32 (8)</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>2. Activity limitations, support</td>
<td>4 (0-10)</td>
<td>9 (2)</td>
<td>0</td>
<td>63</td>
</tr>
<tr>
<td>3. Maximum walking distance</td>
<td>4 (0-10)</td>
<td>9 (2)</td>
<td>2</td>
<td>79</td>
</tr>
<tr>
<td>4. Footwear requirements</td>
<td>3 (0-5)</td>
<td>5 (1)</td>
<td>0</td>
<td>76</td>
</tr>
<tr>
<td>5. Walking surfaces</td>
<td>3 (0-10)</td>
<td>7 (3)</td>
<td>6</td>
<td>55</td>
</tr>
<tr>
<td>6. Gait abnormality</td>
<td>3 (0-10)</td>
<td>9 (2)</td>
<td>1</td>
<td>76</td>
</tr>
<tr>
<td>7. Alignment</td>
<td>3 (0-15)</td>
<td>13 (4)</td>
<td>3</td>
<td>66</td>
</tr>
</tbody>
</table>
Figure 1. Distribution of the total scores of the American Orthopaedic Foot & Ankle Society Midfoot Scale for patients with Lisfranc injury.
A

Spearman rho = 0.89, p < 0.001

B

Spearman rho = 0.86, p < 0.001
Figure 2A-C. A: Correlation between the Visual Analogue Scale Foot and Ankle (VAS-FA) and the American Orthopaedic Foot & Ankle Society (AOFAS) Midfoot Scale among patients with Lisfranc injury. B: Correlation between the VAS-FA and AOFAS Midfoot Scale Pain subscales. C: Correlation between the VAS-FA and AOFAS Midfoot Scale Function subscales.
AOFAS: The American Orthopaedic Foot & Ankle Society Midfoot Scale

Figure 3. Correlation between the follow-up time and the American Orthopaedic Foot & Ankle Society Midfoot Scale total score was negligible.
Figure 4. Relationships between the Visual Analogue Scale Foot and Ankle subscales and the American Orthopaedic Foot & Ankle Society Midfoot Scale total score. Cohen’s standard for β-values above .10 for small, .30 for moderate and .50 for large relationships. Boxes represent the mean scores (VAS-FA: Pain, Function, Other complaints, and Overall) with 95% CIs.
Figure 5A-C. Thresholds of response categories for items 2 (A), 3 (B), and 4 (C) of the American Orthopaedic Foot & Ankle Society Midfoot Scale. All response categories are ordered correctly. Item 2 (A) has evenly distributed response categories. Response categories of item 3 (B) showed misfunction as there is only a narrow gap between the thresholds between the response categories 2 and 3. Item 4 (C) had ordered threshold values, yet none of the patients answered the worst response category.
Figure 6. Person-Item distribution of the seven items of the American Orthopaedic Foot & Ankle Society Midfoot Scale. Bars represent the location of the patients and circles represent the difficulty of the items.
References

18. Prinsen, C.A.C., et al., COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of life research : an international journal of quality of life aspects of


44. Richter, M., et al., *EFAS score—Development and validation by the score committee of the European foot and ankle society (EFAS).* Foot and Ankle Surgery, 2018.