Reliability, Surgeon Preferences, and Eye-Tracking Assessment of the Stress Examination of the Ankle Syndesmosis

Sara L. Naguib, DPM1, Andrew J. Meyr, DPM, FACFAS2

1 Resident, Temple University Hospital Podiatric Surgical Residency Program, Philadelphia, PA
2 Clinical Associate Professor, Department of Podiatric Surgery, Temple University School of Podiatric Medicine, Philadelphia, PA

ARTICLE INFO

Level of Clinical Evidence: 4

Keywords:
diagnostic testing
high ankle sprain
rotational ankle fracture
syndesmotic ligament
tibiofibular articulation

ABSTRACT

The diagnosis and stabilization of ankle syndesmotic injuries after acute injury remains an area of controversy in the foot and ankle surgical literature, seemingly without universal consensus (1–13). Although much of this discussion has focused on hardware constructs and determination of reduction, one specific area within this broader topic that has particularly interested our group is the intraoperative stress examination of syndesmotic stability. Fracture radiographic characteristics and Lauge-Hansen classification might provide surgeons with some degree of pretest probability of syndesmotic disruption certainly (14–17), but it has been our clinical experience that many surgeons rely primarily on the intraoperative bone hook test, stress dorsiflexion-external rotation test, and/or a combination of these and other tests during ankle fracture open-reduction-internal fixation. However, we are unaware of any standardized method for the performance and/or interpretation of these commonly performed surgical techniques, despite the fact that they likely play a large role in surgical and functional outcomes (3,4,8,9,18,19). This introduces the possibility of unwanted subjectivity and variability in the performance of a purportedly objective diagnostic test.

Therefore the primary objective of this investigation was to determine the frequency of agreement and reliability of the stress examination of the ankle syndesmosis. Secondary objectives were to determine surgeon preferences and protocols with respect to the ankle syndesmosis and to use gaze recognition software to perform an eye-tracking assessment during performance of stress examinations. Twelve foot and ankle surgeons, 12 residents, and 12 students were shown 5 intraoperative fluoroscopic still images and videos of the stress examination of the ankle syndesmosis. They were asked to evaluate the result as being “positive” or “negative” for syndesmotic stability. The overall reliability of the interpretation of the stress examination of the ankle syndesmosis was a kappa of 0.123 (surgeons 0.087; residents 0.019; students 0.237), indicating “slight” agreement. Survey results indicated wide variability in the perioperative preferences and protocols of surgeons dealing with the evaluation and treatment of the ankle syndesmosis. Eye-tracking results also demonstrated variability in the anatomic structures of interest focused on during performance of this testing. The results of this investigation provide evidence of reliability well below what would be expected of a gold standard test during stress examination of the ankle syndesmosis. These results indicate that future scientific endeavors are required to standardize the performance and interpretation of this testing.

© 2018 by the American College of Foot and Ankle Surgeons. All rights reserved.

Patients and Methods

After this study received ethical approval from our institutional review board, 12 foot and ankle surgeons, 12 podiatric residents, and 12 podiatric medical students consented to participate. Surgeons were recruited from the faculty of a regional continuing medical education meeting of a national foot and ankle surgical organization. All were board-certified and reported feeling comfortable with the evaluation and treatment of ankle fractures. Residents were recruited from a local board examination review meeting hosted by the senior author (A.J.M.). All were senior-level residents from Council on Podiatric Education-approved programs with the added credential in rearfoot and ankle reconstructive surgery. Students were recruited from the Temple University School of Podiatric Medicine. All were fourth-year students who had completed their clinical externships, the residency interview process, and the Part 2 examination of the National Board of Podiatric Medical Examiners.
Participants were shown 5 intraoperative fluoroscopic images of the stress examination of the ankle syndesmosis. These included 3 still images and 2 videos and stress examinations performed with the dorsiflexion-external rotation test (maximal dorsiflexion and external rotation of the talus against the lateral ankle mortise) and Cotton hook test (tibiofibular distraction with bone hooks or clamps). All images contained a distal-lateral fibular hardware construct with a reduced fibular fracture, and 3 (60.0%) of 5 also contained evidence of medial malleolar fixation. No images had clear evidence of posterior malleolar involvement. Participants viewed these in an order produced by a random number generator. Participants evaluated each image/video and reported whether they thought that the test was “positive” (indicating the syndesmosis was unstable, and they would perform stabilization) or “negative” (indicating the syndesmosis was stable, and they would not perform stabilization). No time limit was enforced on examination of the images, and the videos could be reviewed multiple times if requested.

The primary outcome measure was the level of agreement between board-certified surgeons with respect to the interpretation of the stress examinations. This was measured with a percent count. However, since there was a 50% likelihood that participants would agree on the interpretation simply as a result of chance within this design, reliability was also measured using the Fleiss’ Kappa. This is a measure of agreement between more than 2 raters when data are categorical, in this case “positive” versus “negative.” An established value interpretation of the kappa statistic is as follows: kappas from 0.01 to 0.20 indicate “slight” agreement, from 0.21 to 0.40 indicate “fair” agreement, from 0.41 to 0.60 indicate “moderate” agreement, from 0.61 to 0.80 indicate “substantial” agreement, and from 0.81 to 1.00 indicate “almost perfect” agreement (20). Calculated kappa values less than 0.00 were considered 0.00.

Percent counts and kappa values were also calculated for residents, students, and the entire cohort; however, these were considered secondary outcomes since the residents and students would be expected to have less clinical experience and expertise in the interpretation of these tests.

The surgeons additionally completed a survey after completion of their stress syndesmotic evaluations that attempted to elucidate perioperative preferences dealing with ankle syndesmotic injuries, including their preferred methods of fixation and their impression on how they determine syndesmotic instability. We also had the residents complete the survey but modified the questions to elucidate what they thought would be their preferences and protocols next year when in practice. We chose not to have the students complete the survey. It was our opinion that their level of clinical experience would not be to the point of perioperative preferences and protocols.

Furthermore, eye-tracking and gaze recognition software (Gazepoint, Clemson, South Carolina) was used to provide a subjective measure of what specific anatomy participants were looking at during performance of the stress examinations. A 2-computer-monitor setup was used so that participants were unable to visualize their specific eye-tracking during performance of the testing (Fig. 1). These results were categorized into 5 groups by the senior author (A.J.M.): “exclusively lateral focus,” “primarily lateral focus but with glances medially,” “equal focus between lateral and medial,” “primarily medial focus but glances laterally,” and “exclusively medial focus” (Figs. 2–4). Lateral focus was defined as anatomy involving the tibiofibular overlap, lateral ankle gutter, and fibular fracture line. Medial focus was defined as anatomy involving the medial clear space and medial malleolus.

Data were stored in a password-protected personal computer for subsequent statistical analysis. All statistical analyses were performed using Statistical Analysis Systems software, version 9.2 (SAS Institute, Cary, North Carolina), by the senior author (A.J.M.).

Results

Level of Agreement and Reliability

Full study results are shown in Tables 1–5. With respect to the first still ankle image, 9 (75.0%) of the 12 surgeons, 11 (91.7%) of the 12 residents, and 10 (83.3%) of the 12 students considered it “positive.” With respect to the second still ankle image, 9 (75.0%) of the 12 surgeons, 7 (58.3%) of the 12 residents, and 6 (50.0%) of the 12 students considered it “positive.” With respect to the third still ankle image, 8 (66.7%) of the 12 surgeons, 10 (83.3%) of the 12 residents, and 12 (100.0%) of the 12 students considered it “positive.” The overall reliability of the interpretation of the still images was a kappa of 0.033 (surgeons 0.000; residents 0.036; students 0.182).

With respect to the first ankle video, 5 (41.7%) of the 12 surgeons, 7 (58.3%) of the 12 residents, and 6 (50.0%) of the 12 students considered it “positive.” With respect to the second ankle video, 3 (25.0%) of the 12 surgeons, 7 (58.3%) of the 12 residents, and 3 (25.0%) of the 12 students considered it “positive.” The overall reliability of the interpretation of the videos was a kappa of 0.00 (surgeons 0.00; residents 0.00; students 0.00).

The overall reliability of the interpretation of the stress examination of the ankle syndesmosis, including still images and videos, was a kappa of 0.123 (surgeons 0.087; residents 0.019; students 0.237).

Fig. 1. Example of eye-tracking setup and calibration. Participants sat comfortably in a chair in front of a computer monitor. A separate monitor (which the participant could not visualize) was used to capture specific eye-tracking data. After calibration, participants were shown 3 still images and 2 videos of the stress examination of the ankle syndesmosis. They were asked to evaluate the images/videos as either “positive” or “negative” with respect to syndesmotic stability.
Fig. 2. Example of eye-tracking demonstrating a primarily lateral focus on anatomic structures. We observed that 2 (16.7%) of the 12 surgeons had “exclusively lateral focus” (an example of which is shown in A, with near complete attention directed to the tibiofibular overlap), whereas 4 (33.3%) of the 12 surgeons had “primarily lateral focus but with glances medially” (an example of which is shown in B, with primary focus on the fracture line and glances at the medial clear space). Ten (83.3%) of the 12 surgeons reported looking to the tibiofibular overlap as part of their protocol during testing to assess for syndesmotic stability, with 6 surgeons reporting that they primarily looked to the tibiofibular overlap while making the determination of syndesmotic stability. The green circles indicate where the subjects were looking, with larger circles indicating consistent focus and gaze.

Fig. 3. Example of eye-tracking showing a primarily medial focus on anatomic structures. We observed that no surgeons had an “exclusively medial focus,” whereas 2 (16.7%) of 12 surgeons had “primarily medial focus but with glances laterally.” Figure 3 shows an example of near-complete focus on the medial and superior clear spaces. All 12 surgeons (100.0%) reported looking to the medial clear space as part of their protocol during testing to assess for syndesmotic stability, with 3 surgeons reporting that they primarily looked to the medial clear space while making the determination of syndesmotic stability.

Table 1
Level of agreement of the stress examination of the ankle syndesmosis

<table>
<thead>
<tr>
<th>Frequency of agreement with a “positive” result</th>
<th>Still image 1</th>
<th>Still image 2</th>
<th>Still image 3</th>
<th>Video 1</th>
<th>Video 2</th>
<th>Reliability (kappa value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeons (n = 12)</td>
<td>75.0%</td>
<td>75.0%</td>
<td>66.7%</td>
<td>41.7%</td>
<td>25.0%</td>
<td>0.087</td>
</tr>
<tr>
<td>Residents (n = 12)</td>
<td>91.7%</td>
<td>58.3%</td>
<td>83.3%</td>
<td>58.3%</td>
<td>58.3%</td>
<td>0.019</td>
</tr>
<tr>
<td>Students (n = 12)</td>
<td>83.3%</td>
<td>50.0%</td>
<td>100.0%</td>
<td>50.0%</td>
<td>25.0%</td>
<td>0.237</td>
</tr>
<tr>
<td>Total (n = 36)</td>
<td>83.3%</td>
<td>61.1%</td>
<td>83.3%</td>
<td>50.0%</td>
<td>36.1%</td>
<td>0.123</td>
</tr>
</tbody>
</table>

Table 2
Perioperative syndesmotic fixation preferences and protocols

<table>
<thead>
<tr>
<th>Prefer static (screw) fixation</th>
<th>Prefer dynamic (suture) fixation</th>
<th>Prefer combination of static and dynamic fixation</th>
<th>Prefer to always remove syndesmotic fixation</th>
<th>Prefer to never remove syndesmotic fixation</th>
<th>Prefer to sometimes remove syndesmotic fixation based on patient symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeons (n = 12)</td>
<td>83.3%</td>
<td>8.3%</td>
<td>8.3%</td>
<td>50.0%</td>
<td>41.6%</td>
</tr>
<tr>
<td>Residents (n = 12)</td>
<td>66.7%</td>
<td>25.0%</td>
<td>8.3%</td>
<td>50.0%</td>
<td>41.6%</td>
</tr>
</tbody>
</table>
Survey Results of Perioperative Preferences

The 12 surgeons reported 6 different syndesmotic stabilization construct preference patterns. Ten (83.3%) of 12 surgeons reported preferring exclusive static fixation with screws, 1 (8.3%) of 12 surgeons reported preferring exclusive dynamic fixation with a suture product, and 1 (8.3%) of 12 surgeons reported preferring a combination of static fixation with a screw and dynamic fixation with a suture product. Of those surgeons who preferred screws, 5 (50.0%) of 10 reported preferring either a single 3.5-mm or 4.0-mm screw inserted across either 3 or 4 cortices; 4 (40.0%) of 10 reported preferring 2 3.5-mm or 4.0-mm screws inserted across either 3 or 4 cortices; and 1 (12.5%) of 8 reporting thinking they would prefer a single 3.5-mm screw inserted across 3 cortices; 1 (12.5%) of 8 reporting thinking they would prefer two 3.5-mm cortical screws inserted across either 3 or 4 cortices.

Six (50.0%) of 12 surgeons reported to prefer to “always” remove syndesmotic fixation with timeframes ranging between 3 and 6 months; 5 (41.6%) of 12 surgeons reported to prefer to “sometimes” remove syndesmotic fixation based on patient symptoms or hardware failure; and 1 (8.3%) of the 12 surgeons reported to prefer to “never” remove syndesmotic fixation.

Six (50.0%) of the 12 residents reported thinking that they would prefer to “always” remove the syndesmotic fixation; 5 (41.6%) of the 12 residents reported thinking that they would prefer to “sometimes” remove their syndesmotic fixation; and 1 (8.3%) of the 12 residents reported thinking that they would prefer to “never” remove their syndesmotic fixation.

Seven (58.3%) of the 12 surgeons reported commonly using a bone hook or clamp to provide a distraction force as part of their protocol to assess for syndesmotic stability, with 6 surgeons reporting that this was their primary determinant of syndesmotic stability. Seven (58.3%) of the 12 surgeons reported commonly using the stress dorsiexion-external rotation test as part of their protocol to assess for syndesmotic stability, with 2 surgeons reporting that this was their primary determinant of syndesmotic stability. One (8.3%) of the 12 surgeons additionally reported physically palpating the anterior syndesmosis as part of their

---

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Perioperative preferences of stress syndesmosis examination (testing)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Routinely perform stress-dorsiflexion external rotation test to assess for syndesmotic stability (primary determinant)</td>
</tr>
<tr>
<td></td>
<td>Surgeons (n = 12)</td>
</tr>
<tr>
<td></td>
<td>81.3% (50.0%)</td>
</tr>
<tr>
<td></td>
<td>58.3% (25.0%)</td>
</tr>
<tr>
<td></td>
<td>33.3% (0.0%)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Perioperative preferences of stress syndesmosis examination (anatomy)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Look to medial clear space in determination (primary determinant)</td>
</tr>
<tr>
<td></td>
<td>Surgeons (n = 12)</td>
</tr>
<tr>
<td></td>
<td>100.0% (25.0%)</td>
</tr>
<tr>
<td></td>
<td>83.3% (50.0%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Categorized results of eye-tracking gaze focus during stress examination of the ankle syndesmosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exclusively lateral focus</td>
</tr>
<tr>
<td></td>
<td>Surgeons (n = 12)</td>
</tr>
<tr>
<td></td>
<td>16.7%</td>
</tr>
<tr>
<td></td>
<td>33.3%</td>
</tr>
<tr>
<td></td>
<td>33.3%</td>
</tr>
<tr>
<td></td>
<td>16.7%</td>
</tr>
</tbody>
</table>
protocol to assess for syndesmotic stability. One (8.3%) of the 12 surgeons reported that they did not have a primary test for determining syndesmotic stability. No other methods were reported.

Six (50.0%) of the 12 residents reported thinking that they would commonly use a bone hook or clamp to provide a distraction force as part of their protocol to assess for syndesmotic stability, with 4 residents reported thinking that this would be their primary determinant of syndesmotic stability. Eleven (91.7%) of the 12 residents reported thinking that they would commonly use the stress dorsiflexion-external rotation test as part of their protocol to assess for syndesmotic instability, with 8 residents reported thinking that this would be their primary determinant of syndesmotic stability. Four (33.3%) of the 12 residents reported thinking that they would use the mechanism of injury and preoperative radiographic findings as part of their protocol to assess for syndesmotic stability.

All 12 surgeons (100.0%) reported looking to the medial clear space as part of their protocol during testing to assess for syndesmotic stability, with 3 surgeons reporting that they primarily looked to the medial clear space while making the determination of syndesmotic stability. Ten (83.3%) of the 12 surgeons reported looking to the tibiofibular overlap as part of their protocol during testing to assess for syndesmotic stability, with 6 surgeons reporting that they primarily looked to the tibiofibular overlap while making the determination of syndesmotic stability. One (8.3%) of the 12 surgeons additionally reported looking to frontal plane rotation of the talus as part of their protocol during testing for syndesmotic stability. Three (25.0%) of the 12 surgeons reported that they did not have a primary anatomic area to look for during testing.

Eleven (91.7%) of the 12 residents reported thinking that they would look to the medial clear space as part of their protocol during testing to assess for syndesmotic stability, with 8 residents reporting that they thought they would primarily look to this area while making the determination of syndesmotic stability. Eight (66.7%) of the 12 residents reported thinking that they would look to the tibiofibular overlap as part of their protocol during testing to assess for syndesmotic stability, with 4 residents reporting that they thought they would primarily look to this area while making the determination of syndesmotic stability. One (8.3%) of the 12 residents additionally reported looking to frontal plane rotation of the talus as part of their protocol during testing to assess for syndesmotic stability.

**Eye-Tracking and Gaze Recognition Results**

We observed that 2 (16.7%) of the 12 surgeons had “exclusively lateral focus,” 4 (33.3%) of the 12 surgeons had “primarily lateral focus but with glances medially,” 4 (33.3%) of the 12 surgeons had “equal focus between lateral and medial,” 2 (16.7%) of 12 surgeons had “primarily medial focus but with glances laterally,” and no surgeons had “exclusively medial focus.”

We observed that 1 (8.3%) of the 12 residents had “exclusively lateral focus,” 2 (16.7%) of the 12 residents had “primarily lateral focus but with glances medially,” 3 (25.0%) of the 12 residents had “equal focus between lateral and medial,” 5 (41.7%) of the 12 residents had “primarily medial focus but with glances laterally,” and 1 (8.3%) of the 12 residents had “exclusively medial focus.”

We observed that 2 (16.7%) of the 12 students had “exclusively lateral focus,” 3 (25.0%) of the 12 students had “primarily lateral focus but with glances medially,” 7 (58.3%) of the 12 students had “equal focus between lateral and medial,” and no students had “primarily medial focus but with glances laterally” or “exclusively medial focus.”

**Discussion**

As with any scientific investigation, critical readers are encouraged to review and assess the study design and specific results to reach their own independent conclusions; however, the following are our conclusions based on the data. We also never consider data to be definitive, but we do think that several findings are worthy of clinical attention and future investigation. First, the observed levels of agreement were poor and not consistent with what would be expected from a gold standard diagnostic test. The observed “slight” levels of agreement (as measured by the kappa coefficient) were essentially the same as would be expected from agreement by chance. This indicates that the stress examination of the ankle syndesmosis might benefit from the creation of an objective definition and standardized interpretation. The survey results and eye-tracking/gaze recognition patterns observed in this study might be useful as a starting point in defining just such an objectification.

Second, the results of the study indicate variability in clinical practice and teaching with respect to the performance of these tests and perioperative protocols when dealing with the ankle syndesmosis. Although it should certainly not be considered epidemiologic data representing contemporary clinical practice in the United States, it does at least show that a wide variety of perioperative preferences exists. The similar variability observed in residents about to enter clinical practice demonstrates that this might extend from training and not just from clinical experience. Taken together, these 2 conclusions indicate that the stress examination of the syndesmosis might warrant as much scientific attention as issues such as fixation construct and reduction determination previously have.

We also embrace the fact that all investigations have limitations, and this study had several important ones to consider. First, data were collected from a limited number of subjects, and therefore these results may not be representative of a broader population sampling. Second, evaluating still images and intraoperative videos almost certainly does not recreate the intraoperative decision-making environment when surgeons typically have the ability to “feel” the performance of the tests. We are unable to infer how this might influence their results of interpretation of syndesmotic stability. Finally, we did not evaluate every form of stress examination of the ankle syndesmosis that surgeons might use in their practices. It is possible that other forms of assessment have different levels of agreement and reliability.

In conclusion, the results of this investigation provide evidence of reliability well below what would be expected of a gold standard during stress examination of the ankle syndesmosis. These results indicate that future research is required to standardize the performance and interpretation of this test. It is our hope that our survey questions and eye-tracking results might be used to do so.

**References**


