

The Lateral Fibular Stress Test: High Variability of Force Applied by Orthopaedic Surgeons in a Biomechanical Model

Eitan M. Ingall, MD¹ , Philip Kaiser, MD², Soheil Ashkani-Esfahani, MD³ , John Zhao, MD¹, and John Y. Kwon, MD⁴

Abstract

Background: The lateral fibular stress test (LFST), also known as the hook or Cotton test, is commonly performed to assess syndesmotoc instability intraoperatively. Several studies have used 100 N as the force applied when performing the LFST to detect syndesmotoc instability, though no evidence-based requisite force has been described for the test. We hypothesize that surgeons do not apply force uniformly or consistently when performing the LFST and that substantial variation exists. Fundamentally, this could lead to inconsistent diagnosis of syndesmotoc instability as surgeons may not be applying the force in a consistent manner.

Methods: A biomechanical ankle model consisting of an industrial force gauge attached through a SawBones model was fashioned. Orthopaedic attending surgeons and trainees were asked to perform a series of LFSTs and to simulate the force they typically apply intraoperatively. Basic demographic data were collected on each participant.

Results: Thirty-three surgeons participated in the study, including 18 trainees. The median (IQR) force applied during the LFST was 96.42 (71.42–126.33), 87.49 (69.19–117.40), 99.99 (79.91–137.49), for the pooled group, attendings, and trainees respectively. More than half (54.5%) of all trials were less than 100 N (57.8% of surgeons, 51.8% trainees). Intraobserver correlation was excellent within the overall cohort (0.92, $P < .001$), trainees (0.90, $P < .001$), and attendings (0.94, $P < .001$), respectively. Interobserver reliability was fair among the overall cohort ($\kappa = 0.28$, $P = .49$), and poor between the attendings ($\kappa = 0.11$, $P = .69$) and the trainees ($\kappa = 0.05$, $P = .82$), respectively.

Conclusion: Our study demonstrates that the amount of force applied by typical surgeons when performing the LFST test is highly variable. Variable force application when performing the LFST may lead to inconsistent detection of syndesmotoc instability, which may portend a poorer outcome.

Clinical Relevance: In this study, we demonstrate the wide variability in the amount of force used during a lateral fibular stress test. High variability of force application when performing the LFST may lead to inconsistent diagnosis of syndesmotoc instability, which may portend a poorer outcome. Our findings suggest the need for further investigation into the technical aspects of syndesmotoc testing that will permit more reproducible and valid interrogation of the syndesmosis.

Keywords: syndesmosis, Cotton test, stress test

Introduction

Syndesmotoc injuries are common in both Weber B and Weber C ankle fractures. Recognition and anatomic stabilization of syndesmotoc disruption correlates with improved clinical outcomes.^{1,4,7,12} The lateral fibular stress test (LFST), also referred to as the hook or Cotton test,³ is commonly performed to diagnose syndesmotoc instability intraoperatively. A bone hook or clamp is placed on the fibula

near the level of the superior border of the syndesmosis with the foot in a neutral position. A lateral distraction force is applied while evaluating for fluoroscopic widening of the medial clear space (MCS), tibiofibular overlap (TFOL), and tibiofibular clear space (TFCS). In a cadaveric model, Stoffel et al¹³ demonstrated that 100 N of force applied in an LFST to an uninjured ankle specimen was the point after which no further MCS, TFOL, or TFCS increases were



appreciated, and the clamp was noted to crush into the bone. Additional studies have similarly used 100 N as the applied force for detecting syndesmotic instability.^{6,8} Despite this, there does not appear to be any conclusive evidence that 100 N is the requisite minimum amount of force.

Additionally, it remains unknown how much force is actually applied by surgeons when performing the LFST in the operating room. Underdetection of syndesmotic instability may occur if surgeons are not applying the requisite force in a consistent manner. Likewise, pulling too hard may cause the clamp to subside in osteoporotic bone or even iatrogenically widen the syndesmosis secondary to supraphysiologic loading. Therefore, the purpose of this study was to evaluate the amount of force orthopaedic surgeons apply during an LFST in a simulated ankle fracture model. Our hypothesis is that (1) there is substantial variation among surgeons in the amount of force applied, (2) surgeons do not consistently apply 100 N of force, and (3) the amount of force pulled is independent of level of training or subspecialty.

Methods

This study was conducted after approval by our institutional review board. A biomechanical SawBones lower leg model including simulated soft tissue envelope (SawBones Inc, Vashon Island, Washington) was mounted to a board and a 1-cm-diameter hole was drilled across the tibia and fibula in the area of the syndesmosis. An industrial force gauge (Nidec-Shimpo, Kyoto, Japan) was then mounted. A metal extension piece was passed from the force gauge through the hole and positioned so that the tip was exposed on the lateral side of the model (Figure 1). A commonly utilized reduction clamp was affixed to the force gauge. The construct allowed applied forces to be directly transmitted to the force gauge, increasing measurement accuracy and obviating any potential issues such as fracture of the SawBones model, clamp pull-out, or variability of clamp placement had the reduction clamp been applied directly to the fibula. Validation of the model was performed with a second strain gauge with reproducible force readings consistently within 0.5 lb of force. This calibration was repeated after every 10 participants to ensure continued validity and accuracy of the model.

Attending orthopaedic surgeons and orthopaedic trainees (residents and fellows) from two institutions who had previously performed an intraoperative LFST were eligible



Figure 1. Biomechanical model. An industrial force gauge was fashioned to a sawbones model for simulation.

for inclusion in the study. Surgeons who did not know how to perform an LFST and those who had not previously treated an ankle fracture were excluded. Participants were grouped as either attending surgeons or trainees for analysis. Basic demographic data were collected on each participant including years in practice for attending surgeons. Participant gender, handedness, fellowship subspecialty training, and the number of ankle fractures treated operatively within the previous year (prior to testing) were also recorded. Participants were asked to perform a series of LFSTs and to simulate the force they typically applied intraoperatively. The display of the gauge was covered such that participants were unable to see the amount of force they were exerting. After a demonstration of the system and 1 practice attempt (with the force gauge covered), 3 trials were recorded for each participant with a 1-minute break between tests.

Statistical Analysis

Data are exhibited as median and interquartile range (IQR). The Shapiro Wilk test demonstrated non normally distributed data ($P > .05$). Basic demographic data between

¹Harvard Combined Orthopaedic Residency Program, Massachusetts General Hospital, Boston, MA, USA

²Department of Orthopaedic Surgery, Massachusetts General Hospital, Boston, MA, USA

³Foot & Ankle Research and Innovation Laboratory (FARIL), Massachusetts General Hospital, Boston, MA, USA

⁴Department of Orthopaedic Surgery, Division of Foot and Ankle Surgery, Massachusetts General Hospital, Boston, MA, USA

Corresponding Author:

Eitan M. Ingall, MD, Harvard Combined Orthopaedic Residency Program, Massachusetts General Hospital, 55 Fruit St, Boston, MA 02114, USA
Email: eingall@partners.org

Table 1. Individual Characteristics of the Participants.

	Attendings (n = 15; 45.5%)	Trainees (n = 18; 54.5%)	P Value ^a
Gender, n (%)			.41
Male	14 (93.3)	15 (83.3)	
Female	1 (6.7)	3 (16.7)	
Handedness, n (%)			.81
Right	14 (93.3)	18 (100)	
Left	1 (6.7)	0 (0)	
Years in practice/PGY, median (IQR)	9 (3-17)	4 (2-5)	NA
Ankle fixations/year, %			.09
0-10	6.7	22.2	
11-20	13.3	16.7	
>20	80	61.1	
Routine LFST use	11 (73.3)	18 (100)	.07

Abbreviations: IQR, interquartile range; LFST, lateral fibular stress test; NA, not applicable; PGY, postgraduate year.
^a χ^2 test where applicable.

groups including gender, handedness, subspecialty, and number of ankle fractures fixed per year, and routine use of the LFST were compared with chi-squared tests. The amount of force applied in consecutive pulls were compared within groups using the Friedman test and between groups using the Wilcoxon signed-rank test. *P* value < .05 was considered statistically significant. Intraclass correlation coefficient (ICC) estimation and Fleiss multirater Kappa test were performed for the entire cohort and among each group to assess the intra- and interrater reliability, respectively. Kappa index was interpreted as poor if less than 0.20, fair if 0.20 to 0.40, moderate if 0.40 to 0.60, good if 0.60 to 0.80, and very good if 0.80 to 1.00. ICC below 0.50 was considered poor; 0.50 to 0.75, moderate; 0.75 to 0.90, good; and above 0.90, excellent. Target enrollment was 20 subjects as we estimated the study would have 90% power (alpha 0.05) to detect a 15-N difference in pull strength between attendings and residents with a sample size of 10 subjects per group.

Results

Basic demographic data of the participants in each group are depicted in Table 1. Thirty-three surgeons participated in the study, of which 4 were female (12.1%). Among the attendings, mean years in practice was 9 (IQR 3-17). Trainees were mostly postgraduate year 4 and 5 (62.5%) residents. Eighty percent of the attendings stated they fixed >20 ankles in the previous year (vs 61% of trainees). A majority (73%) of attendings were routinely using the LFST to evaluate the syndesmosis intraoperatively.

The median (IQR) force applied during the LFST was 96.42 (71.42-126.33), 87.49 (69.19-117.40), and 99.99 (79.91-137.49) for the pooled group, attendings, and trainees, respectively. There was no significant difference between the attendings and trainees with respect to the first

(*P* = .42), second (*P* = .49), or third (*P* = .49) trials. There was no difference in the amount of force between those with foot and ankle subspecialty training vs other subspecialties in any of the 3 trials (*P* = .74, .78, .69, respectively). More than half (54.5%) of all LFSTs were less than 100 N (57.8% of surgeons, 51.8% trainees), with the distribution depicted in Figure 2.

The ICC was excellent within the overall cohort (0.92, *P* < .001), trainees (0.90, *P* < .001), and attendings (0.94, *P* < .001), respectively. Interobserver reliability was fair among the overall cohort (κ = 0.28, *P* = .49) and poor between the attendings (κ = 0.11, *P* = .69) and the trainees (κ = 0.05, *P* = .82), respectively.

Discussion

Significant challenges in diagnosing syndesmotom instability exist even when performing intraoperative stress tests. The accuracy of these tests relies heavily on technical expertise, applying a reproducible amount of adequate force, and the ability to discern small but meaningful fluoroscopic changes in the MCS, TFCS, and TFOL. The findings of this investigation demonstrate that a wide variability in force is applied by orthopaedic surgeons during simulated LFST testing using a biomechanical ankle model. Although there is excellent intraobserver reproducibility of force applied during an LFST, our simulation suggests fair to poor interobserver reliability. Furthermore, the amount of force applied does not appear to be related to level of training, subspecialty training nor other demographic factors. Finally, more than 50% of all trials were below 100 N.

Despite considerable recent research evaluating various parameters to identify syndesmotom instability, little research has been performed examining the technique surgeons' use to stress the syndesmosis in vivo. Stoffel et al¹³ demonstrated in an uninjured cadaveric ankle that after 100

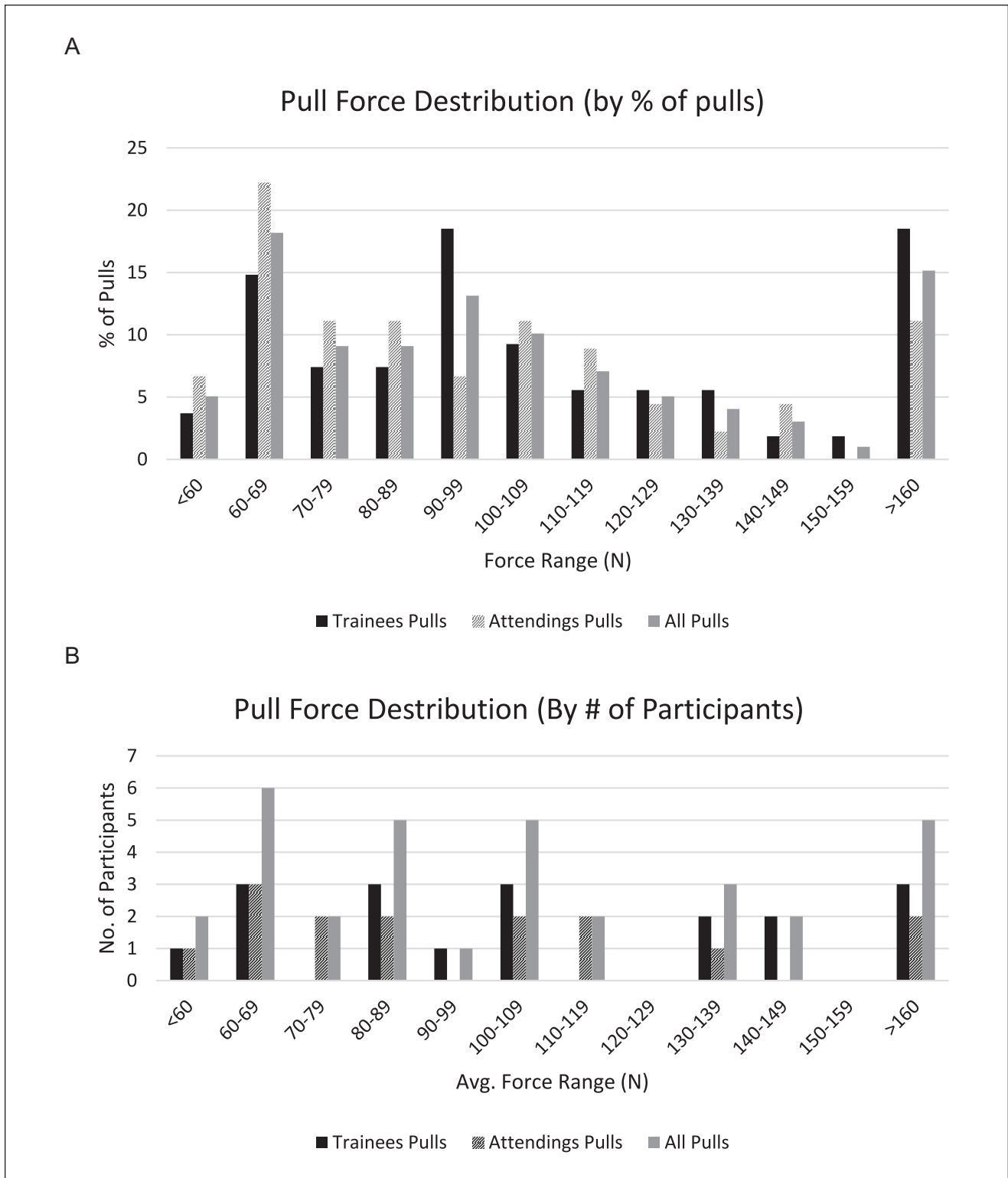


Figure 2. Distribution of force applied during LFST by trainees, attendings and the total cohort, respectively, displayed as (A) percentage of total pulls and (B) number of participants (based on the average^a pull per participant). (LFST, lateral fibular stress test)
^aAverage force range: Each participant performed 3 LFSTs in this investigation. The x axis is grouped according to each participants' respective average of the 3 pulls.

N of applied force, no further widening of MCS, TFCS, or TFOI was noted. This amount of quantified force has been used by several investigators^{1,8,11} to detect syndesmotic instability in cadaveric models, even though it was not intended or investigated as a “requisite force” for the LFST by Stoffel. In fact, it seems logical that accurate diagnosis in vivo is likely based on multiple patient and injury characteristics. Anatomic factors such as bone strength, soft tissue tension, and/or intact parts of the ligamentous complex may all factor into how much the fibula translates with an applied force. To our knowledge, no study has investigated the “correct” amount of force to use for an LFST. Intraoperatively, a force gauge is not typically used during the LFST, so surgeons remain unaware of the actual force they are using to stress the syndesmosis. Although it may seem intuitive that attending orthopaedic surgeons who commonly treat ankle fractures have adequate clinical experience to apply the diagnostic requisite force, our results show that wide variability exists between even experienced surgeons.

Diagnosis of syndesmotic injury with radiographic stress testing continues to be a challenge in the care of ankle fractures. In a 2-surgeon comparison of the external rotation (ER) stress test vs the Cotton test on 140 unstable ankle fractures undergoing surgery, Pakarinen et al¹⁰ found that the LFST had a sensitivity of just 0.25%. Although intra- and interobserver reliability in their study were high, these results should be interpreted in the context of a 2-surgeon comparison. Our simulation challenges their finding of high interobserver reliability of the LFST with testing of more than 30 subjects. Jiang et al⁶ performed a cadaveric study, which demonstrated that the Cotton test increased the TFCS most reliably after syndesmotic injury. Although their biomechanical model used 100 N of force for the LFST, our study highlights that in clinical practice, collectively, surgeons do not routinely apply this amount of force and that significant variability exists between surgeons. The findings of our study therefore question several foundational assumptions underpinning such studies.

Several studies have directly compared the LFST with the ER stress test.^{9,13} Common to all of these studies, however, is the lack of a standard method of LFST when being compared to the ER stress test. In previous cadaveric studies, 100 N of force is most often applied. However, in the clinical studies, the amount of force pulled is not delineated. In fact, the authors were unable to find a single study that quantified the in vivo force used by surgeons during the LFST. This is in contrast to the ER stress test, where a standardized torque of 7.5 Nm can be applied with an F-Tool.⁵ Given the amount of variability identified between subjects in our study, we would be concerned that the clinical variability in the force applied during the Cotton test would alter measurement of syndesmotic widening and may have implications on treatment.

Standardization of the LFST goes beyond simply the amount of force pulled. Direction of distraction has also been proposed as a factor for consideration.² Furthermore, where the surgeon places his or her hand on the leg to apply countertraction and where the clamp is placed on the fibula may all impact assessment of radiographic parameters when stressing the mortise. Further study of these variables in an effort to further standardize testing would be warranted to optimize accuracy of diagnosing syndesmotic injury.

There are several limitations to our study. Foremost, this is a biomechanical study using an ankle fracture model and does not fully replicate in vivo situations. However, using a SawBones model reduced variability compared to cadaveric or in vivo study as we were able to solely examine the force surgeons apply during a simulated LFST without other potential confounding factors. In vivo or cadaveric tissue, however, would inherently better mimic the biofeedback experienced by surgeons in the operating theatre and may have led to differential force generation. Additionally, this study was performed under ideal conditions to quantify force generation without common clinical concerns such as iatrogenic osseous fibular injury when applying a reduction clamp or disrupting a concomitantly repaired fibular fracture. Based on our own experiences, surgeons likely apply variable amounts of force depending on the specific clinical scenario and type of ankle fracture. Additionally, although our surgeon cohort likely represents typical abilities generalizable to most practicing orthopaedic surgeons, this study was performed at two institutions in a common geographic area. Although not a weakness specific to this investigation, it should be noted that many syndesmotic injuries in clinical practice are readily evident on initial radiographs. LFST or other intraoperative stressing testing is not always required to diagnose syndesmotic instability. Finally, we want to emphasize that 100 N for performing an LFST may in fact not be the requisite amount of force for all patients. Although this amount of force has been used several times throughout the literature, its questionable use as a methodologic predicate is derived from one study that was not aiming to determine the “correct” amount of force to use in an LFST. The appropriate amount of force required in an LFST to detect syndesmotic injury should be an area of further investigation. Furthermore, our findings regarding the wide variability of force pulled and low interobserver reliability of the test, highlight the need for improvement and standardization of the LFST.

In conclusion, our study demonstrates that the amount of lateral force applied by surgeons in a biomechanical ankle model when performing the LFST is variable. Variability in force application when performing the LFST may impact consistent detection of syndesmotic instability, which may portend a poorer outcome. Either the intraoperative use of force gauges and/or specific practice outside of the operating theatre (to become familiarized with the proprioceptive feel

of generating the requisite force) may permit surgeons to consistently apply the test in a manner that is clinically reproducible. Finally, these results suggest that further investigation into the technical reproducibility and accuracy of intraoperative syndesmotic testing, specifically the LFST, on both cadaveric specimens and in vivo is warranted.

Ethical Approval

Ethical approval for this study was obtained from Beth Israel Deaconess Medical Center IRB (Protocol no. 2021P000178).



Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. ICMJE forms for all authors are available online.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Eitan M. Ingall, MD,  <https://orcid.org/0000-0003-4842-5801>
Soheil Ashkani-Esfahani, MD,  <https://orcid.org/0000-0003-2299-6278>

References

1. Bejarano-Pineda L, Guss D, Waryasz G, DiGiovanni CW, Kwon JY. The syndesmosis, part I: anatomy, injury mechanism, classification, and diagnosis. *Orthop Clin North Am.* 2021;52(4):403-415. doi:10.1016/j.ocl.2021.05.010
2. Candal-Couto JJ, Burrow D, Bromage S, Briggs PJ. Instability of the tibio-fibular syndesmosis: have we been pulling in the wrong direction? *Injury.* 2004;35(8):814-818. doi:10.1016/j.injury.2003.10.013
3. Cotton FJ. *Dislocations and Joint-Fractures.* W. B. Saunders Company; 1910:549-550. <http://resource.nlm.nih.gov/36210580R>
4. Heifner JJ, Kilgore JE, Nichols JA, Reb CW. Syndesmosis injury contributes a large negative effect on clinical outcomes: a systematic review. *Foot Ankle Spec.* Published online January 20, 2022. doi:10.1177/19386400211067865
5. Jenkinson RJ, Sanders DW, Macleod MD, Domonkos A, Lydestadt J. Intraoperative diagnosis of syndesmosis injuries in external rotation ankle fractures. *J Orthop Trauma.* 2005;19(9):604-609. doi:10.1097/01.bot.0000177114.13263.12
6. Jiang KN, Schulz BM, Tsui YL, Gardner TR, Greisberg JK. Comparison of radiographic stress tests for syndesmotic instability of supination-external rotation ankle fractures: a cadaveric study. *J Orthop Trauma.* 2014;28(6):123-127. doi:10.1097/BOT.0000000000000010
7. Kaiser PB, Bejarano-Pineda L, Kwon JY, DiGiovanni CW, Guss D. The syndesmosis, part II: surgical treatment strategies. *Orthop Clin North Am.* 2021;52(4):417-432. doi:10.1016/j.ocl.2021.05.011
8. Massri-Pugin J, Lubberts B, Vopat BG, Wolf JC, DiGiovanni CW, Guss D. Role of the deltoid ligament in syndesmotic instability. *Foot Ankle Int.* 2018;39(5):598-603. doi:10.1177/1071100717750577
9. Matuszewski PE, Dombroski D, Lawrence JTR, Esterhai JL, Mehta S. Prospective intraoperative syndesmotic evaluation during ankle fracture fixation. *J Orthop Trauma.* 2015;29(4):e157-e160. doi:10.1097/bot.0000000000000247
10. Pakarinen H, Flinkkilä T, Ohtonen P, et al. Intraoperative assessment of the stability of the distal tibiofibular joint in supination-external rotation injuries of the ankle sensitivity, specificity, and reliability of two clinical tests. *J Bone Joint Surg Am.* 2011;93(22):2057-2061. doi:10.2106/JBJS.J.01287
11. Pakarinen HJ, Flinkkilä TE, Ohtonen PP, Ristiniemi JY. Stability criteria for nonoperative ankle fracture management. *Foot Ankle Int.* 2011;32(2):141-147. doi:10.3113/FAI.2011.0141
12. Stenquist DS, Ye MY, Kwon JY. Acute and chronic syndesmotic instability: role of surgical stabilization. *Clin Sports Med.* 2020;39(4):745-771. doi:10.1016/j.csm.2020.06.002
13. Stoffel K, Wysocki D, Baddour E, Nicholls R, Yates P. Comparison of two intraoperative assessment methods for injuries to the ankle syndesmosis: a cadaveric study. *J Bone Joint Surg Am.* 2009;91(11):2646-2652. doi:10.2106/JBJS.G.01537