Prospective Assessment of Clinical Tests Used to Evaluate Tibial Stress Fracture

Michael D. Rosenthal,* MPT, DSc, Mitchell J. Rauh,[†‡] PT, PhD, MPH, and James E. Cowan,[§] MPT, DSc

Investigation performed at Naval Special Warfare Center, Coronado, California, USA

Background: Tibial stress fracture (SFx) is the most common SFx of the lower extremity. Presently, diagnostic accuracy of clinical examination techniques for tibial SFx remains suboptimal.

Purpose: To assess the diagnostic effectiveness of 5 clinical tests for tibial SFx individually versus a test item cluster.

Study Design: Cohort study (diagnosis); Level of evidence, 3.

Methods: A total of 50 patients with tibial pain (17 with bilateral symptoms) were assessed with 5 clinical examination tests (tibial fulcrum test, focal tenderness to palpation, heel percussion test, therapeutic ultrasound test, and 128-Hz tuning fork test) before they underwent diagnostic imaging (radionuclide bone scan). The application of the clinical tests was counterbalanced to minimize the likelihood of carryover effects. Patients provided a pain rating immediately before and after the application of each clinical test.

Results: The prevalence of tibial SFx among the study participants was 52.2%. High levels of specificity were produced by the therapeutic ultrasound test (93.8%), tuning fork test (90.6%), and percussion test (90.6%). The fulcrum test had moderate to high specificity (84.4%). All tests demonstrated low levels of sensitivity, with the highest levels found for focal tenderness to palpation (48.6%) and fulcrum (45.7%). The fulcrum test provided the highest positive likelihood ratio (2.93), followed by the therapeutic ultrasound test (2.30). The fulcrum test had the lowest negative likelihood ratio (0.64), with the focal tenderness to palpation and tuning fork tests having negative likelihood ratios >1.0. Combinations of these clinical tests did not improve the prediction of tibial SFx above that observed among the individual tests.

Conclusion: The clinical tests evaluated were generally highly specific, but all had low sensitivity. The fulcrum test provided the highest level of diagnostic accuracy; however, it was inadequate for definitive clinical management. Combining tests did not improve the diagnostic accuracy of tibial SFx.

Keywords: stress fracture; therapeutic ultrasound; fulcrum test; tuning fork test; heel percussion test

Stress fracture (SFx) of the tibia is among the most common SFxs encountered in athletes and military populations.[5,14,20,28,31,33,37] An accurate diagnosis of tibial SFx can be difficult because of the myriad of potential overuse musculoskeletal conditions, variable presentation and report of symptoms, lack of definitive clinical examination tests, and delays in supporting imaging studies.

Accurate and prompt clinical diagnosis is desired for optimal patient management. Decision making for return to preinjury activities often incorporates imaging studies, as clinical examination components have demonstrated low levels of diagnostic sensitivity and/or specificity.[25,39] The use of various special (provocation) tests has been reported in the literature with unacceptable and inconsistent levels of diagnostic accuracy. The percussion test, percussing the heel with the palm of the examiner's hand, can induce pain at the bone stress injury site.[3,19] Use of the fulcrum test, applying a bowing stress (eg, varus/valgus or posterior-to-anterior directed force) to long bones, has proven useful in the assessment of femoral shaft bone stress injury.[4,15,36] Increased pain with application of a vibrating tuning fork has also been purported to aid in the differential diagnosis of bone stress injury.[9,18,34] Lesho[18] found the 128-Hz tuning fork test (TFT) to have 75% sensitivity and 67% specificity. Several reports suggest that pain with application of therapeutic ultrasound (TUS) to the cutaneous surface overlying bone (eg, tibia) may improve identification of bone stress injuries.[2,7,12,22,23,27,29] Conversely, TUS has also been reported to lack reliability in the diagnosis of tibial SFx.[30] Tenderness to palpation (TTP) is also suggested as an important factor in the assessment of bone stress injury.[1,21,26] Prior studies on the diagnostic accuracy of these clinical tests for SFx have been performed in isolation and have not evaluated the use of multiple clinical tests as a test item cluster.

The absence of a clinical test with high diagnostic accuracy for bone stress injuries of the leg has led to reliance upon advanced imaging (eg, radionuclide bone scan or

The Orthopaedic Journal of Sports Medicine, 10(9), 23259671221122356
DOI: 10.1177/23259671221122356

magnetic resonance imaging [MRI]), which is costly and not always readily accessible. Identification of a clinical test, or cluster of tests, with a high predictive value for detecting the presence or absence of tibial SFx would optimize early management, minimize lost training time, and reduce expenditures for imaging studies.

The purpose of this study was to (1) assess the diagnostic accuracy of 5 clinical tests as compared with bone scan for the diagnosis of tibial SFx in military trainees and (2) evaluate whether a combination of clinical test results was more predictive of tibial SFx than a singular test.

METHODS

Participants

The study protocol received institutional review board approval. Our study population consisted of 50 male participants with leg (tibial) pain. The mean age of the study sample was 22 ± 3.0 years (range, 18-30 years). All participants were active-duty military personnel enrolled in US Navy Basic Underwater Demolition/Sea, Air, Land (SEAL) training.

Data Collection

Included were patients who were evaluated at a single medical clinic for tibial leg pain for which imaging was determined necessary for further diagnosis and to aid in the determination of duty (training) status. Anteroposterior and lateral radiographs of the leg and a radionuclide bone scan were ordered for patients who were unable to walk without a limp or could not perform a single-leg hop because of leg pain. Potential participants were excluded from the study if they had a prior diagnosis of tibial SFx (within the past 6 months) on the side that was currently symptomatic.

Testing Procedure

Five clinical tests were performed on each patient within 48 hours before the imaging studies. The 5 clinical tests we examined were chosen because they are readily available in orthopaedic and physical therapy clinics. They were as follows: (1) tibial fulcrum test, (2) focal TTP examination, (3) heel percussion test, (4) TFT (128 Hz), and (5) TUS test (3 MHz). The order of clinical testing was counterbalanced to minimize the likelihood of carryover effects. For example, patient 1 underwent the tests in the order stated above, patient 2 underwent the tests in the order of tests 2 to 5, followed by test 1, and the rotation of test order continued through patient 5; the order of testing with patient 6 was the same as that for patient 1.

Patients 1 through 37 were evaluated by 1 examiner (M.D.R.), and patients 38 through 50 were evaluated by another examiner (J.E.C.). Both examiners were physical therapists with more than 10 years of orthopaedic clinical experience and board certified in sports or orthopaedic physical therapy. To establish consistency in test performance, the examiners reviewed the test performance used on the first 30 patients and then assessed interrater agreement through blinded testing on 10 patients. Interrater agreement ranged from 80% (percussion and tuning fork) to 100% (fulcrum).

Before performing each test, the participant rated his pain using a visual analog scale (VAS; range, 0-10, with 10 indicating worst pain) by placing a vertical line across a 10-cm horizontal line. The initial rating was used as the baseline pain measure. Immediately after completion of each test, the patient once again rated his pain level. The patient notified the examiner when the pain level had returned to the baseline pain measure, at which time the next clinical test was performed. None of the study patients reported discomfort leading to the inability to proceed with subsequent clinical testing.

For all tests except the TTP examination, a positive result for tibial SFx was defined solely as an increase in the VAS pain score >2 points compared with baseline.

Clinical Tests

Tibial Fulcrum Test. Before initiating study enrollment, the lead investigator (M.D.R.) performed the test with a handheld dynamometer to assess the magnitude of force application and establish consistency in performance of the test. Approximately 11 kg force was applied on the side opposite of the location of pain (eg, lateral if pain was on the medial side) and directed perpendicular to the patient's reported location of pain. For example, to stress the posteromedial aspect of the tibia, force was applied through the palmar aspect of the examiner's hand at the patient's anterolateral tibia in a posteromedial (valgus) direction with the examiner stabilizing the leg at the ankle (Figure 1).

Focal Tenderness to Palpation. The TTP examination, using the index and middle fingers, was performed to assess for tenderness along the posteromedial shaft of the tibia (Figure 2). Patients were positioned supine, with the hip and knee flexed to approximately 45° and 90°,

†Address correspondence to Mitchell J. Rauh, PT, PhD, MPH, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA (email: mrauh@sdsu.edu).

*University of Nebraska Medical Center, Omaha, Nebraska, USA.

‡San Diego State University, San Diego, California, USA.

§US Army–Baylor University, Fort Sam Houston, San Antonio, Texas, USA.

Final revision submitted May 28, 2022; accepted June 9, 2022.

The authors declared that they have no conflicts of interest in the authorship and publication of this contribution. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

Ethical approval for this study was obtained from Naval Medical Center San Diego.

Figure 1. Positioning and application of the tibial fulcrum test. The arrow depicts the direction of force application on an individual with pain over the medial tibia (star).

Figure 2. Positioning for performing the focal tenderness to palpation test.

respectively. This allowed for relaxation of the posterior compartment musculature and more direct palpation along the posteromedial aspect of the tibia. As palpation was performed along the posteromedial tibial shaft, pen marks were made to delineate the length of the area(s) of TTP. The test was considered positive if an increase in the VAS pain score >2 points was present compared with baseline and if the area of tenderness was ≤2.5 cm in length; the test was negative if it was >2.5 cm.[1]

Heel Percussion Test. The percussion test was performed with the patient supine. The limb was passively lifted approximately 6 inches off the table and with support distal to the calcaneus. Direct impact was then applied to the plantar aspect of the calcaneus by striking it with the palm of the hand[3,16] (Figure 3).

Figure 3. Heel percussion test.

Figure 4. Tuning fork test.

Tuning Fork Test. The TFT was assessed with the patient supine, the hip and knee flexed (to approximately 45° and 90°, respectively), and the foot resting on the examination table. A 128-Hz tuning fork was then struck and placed proximal to the location where the patient identified the point of greatest pain. While maintaining contact with the tibia, the tuning fork was moved distally to a point 2.5 cm distal to the area of pain. A second application of the TFT was performed by placing the vibrating tuning fork at the location of greatest pain as identified by the patient and maintaining contact between the tuning fork and the tibia at that point until the vibration ceased (Figure 4).

Therapeutic Ultrasound Test. The 3-MHz TUS test was performed spanning the patient-reported area of pain and

Figure 5. Therapeutic ultrasound test.

using an area the size of three 5-cm^2 ultrasound heads[29] (Figure 5). Ultrasound was applied continuously at an intensity of 0.5 W/cm^2 for the first minute, increased to 1.0 W/cm^2 during the second minute, and then increased to 1.5 W/cm^2 for the period from 3 to 5 minutes. The TUS test was stopped before 5 minutes had elapsed if the patient reported an increase in pain >2 points on the VAS.

Imaging Studies

Within 48 hours of completing the battery of clinical tests, each patient underwent a standard imaging protocol of radiographs followed by radionuclide bone scan. The presence or absence of a tibial SFx was documented based on the radiologist's official report after completion of imaging studies.

Statistical Analysis

Specificity, sensitivity, and positive and negative predictive values, along with the positive likelihood ratio (PLR) and negative likelihood ratio (NLR), as well as overall accuracy were calculated for each clinical test.[11] Radionuclide bone scan test results served as the standard for comparison. Likelihood ratios were determined, as they have been considered the best analysis for indicating the usefulness of a test and how the test result (positive or negative) changes the probability of the presence (positive likelihood) or absence (negative likelihood) of the condition.[11]

Using bivariate and multivariate logistic regression analysis, a post hoc receiver operating characteristic (ROC) curve was constructed for each test and test combination. All positive tests were given a score of 1 (evaluator indicated test as positive), and a score of 0 was given if the test was negative. When assessing multiple tests in combination, a 1 was awarded if any single test was positive (eg, if the fulcrum was positive and TTP was negative, the combination of fulcrum and TTP was scored as a 1, as compared with a 0 if all tests were negative). The point on the ROC curve that maximized both the sensitivity and the specificity of the test was identified as the reach distance. Finally, the area under each ROC curve (AUC) was calculated for each test and test combination.

RESULTS

Of the 50 participants in the study, 26 had at least 1 tibial SFx for a prevalence rate of 52.2%. Of the 26 participants who incurred a SFx, 17 had bilateral symptoms resulting in clinical testing on 67 extremities. The duration of reported leg pain was 6 to 365 days, with an average symptom duration of 47 days. Most patients reported their leg pain to be along the posteromedial tibia, which is consistent with prior literature.[16] Each tibial SFx was diagnosed by bone scan; none were diagnosed by radiographs. The location of all tibial SFxs was along the posteromedial aspect, with 3 in the proximal third, 1 in the distal third, and the remainder in the midthird. All sensitivity values were considered low (<50%), with a range from 48.6% (focal TTP) to 5.7% (TFT) (Table 1). High levels of specificity were present for the percussion test (90.6%), TFT (90.6%), and TUS test (93.8%). Moderate to high specificity was demonstrated by the fulcrum test (84.4%), while focal TTP demonstrated low specificity (40.6%). The fulcrum test (2.93; 95% CI, 1.21 -7.07) and TUS test (2.30; 95% CI, 0.48-11) demonstrated the highest PLRs (Table 1). The lowest NLR was produced by the fulcrum test (0.64; 95% CI, 0.46-0.90), with the other tests approaching (percussion and TUS both at 0.91) or exceeding (TFT, 1.04; focal TTP, 1.27) 1.0.

The AUCs ranged from 0.52 to 0.65 for individual tests (Table 2). The highest AUC for any combination of tests was 0.68 for the fulcrum and TUS (Figure 6).

DISCUSSION

The fulcrum test provided superior overall diagnostic accuracy (64%) in comparison with the other clinical tests and was the only test with an overall diagnostic accuracy >52%. None of the other clinical tests were better than a coin toss at predicting the presence of a tibial stress injury. The only combination of clinical tests that improved diagnostic accuracy above that provided by the fulcrum test in isolation was the combination of fulcrum and TUS. The magnitude of difference between those 2 AUCs was small (0.65 vs 0.68), suggesting a low impact of combining the 2 tests. This is the first study to include the fulcrum test in a series of tests for the assessment of exercise-related leg pain.

Despite relatively high incidence rates, clinical diagnosis of tibial SFx remains elusive. The continuum of bone stress injury along with a multitude of other musculoskeletal conditions of the leg contribute to the diagnostic challenge. Imaging studies result in a significant loss of training time for military trainees. In this study, the total procedure time (including transportation and return to the medical clinic) for radiographs was 2 to 3 hours and 6 to 8 hours for a radionuclide bone scan. In some cases, a delay in definitive diagnosis results in a decision to withhold the trainee from

TABLE 1
Diagnostic Statistics for the Clinical Tests[a]

Test	Sensitivity, %	Specificity, %	PPV, %	NPV, %	PLR (95% CI)	NLR (95% CI)	Overall Accuracy, %
Fulcrum	45.7	84.4	76.2	58.9	2.93 (1.21-7.07)	0.64 (0.46-0.90)	64
Percussion	17.1	90.6	66.7	50.0	1.83 (0.50-6.71)	0.91 (0.76-1.10)	52
TFT	5.7	90.6	40.6	46.8	0.61 (0.11-3.42)	1.04 (0.91-1.19)	46
TTP	48.6	40.6	47.2	41.9	0.82 (0.52-1.28)	1.27 (0.75-2.15)	45
TUS	14.3	93.8	71.4	50.0	2.30 (0.48-11.00)	0.91 (0.78-1.08)	52

[a]NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value; TFT, tuning fork test; TTP, focal tenderness to palpation; TUS, therapeutic ultrasound.

TABLE 2
Post Hoc AUC Values for Diagnostic Tests[a]

Test	AUC (95% CI)	SE
Fulcrum	0.65 (0.55-0.76)	0.05
TTP	0.55 (0.43-0.67)	0.06
TUS	0.54 (0.47-0.61)	0.04
Percussion	0.54 (0.46-0.62)	0.04
TFT	0.52 (0.45-0.58)	0.03

[a]AUC, area under the receiver operating characteristic curve; TFT, tuning fork test; TTP, focal tenderness to palpation; TUS, therapeutic ultrasound.

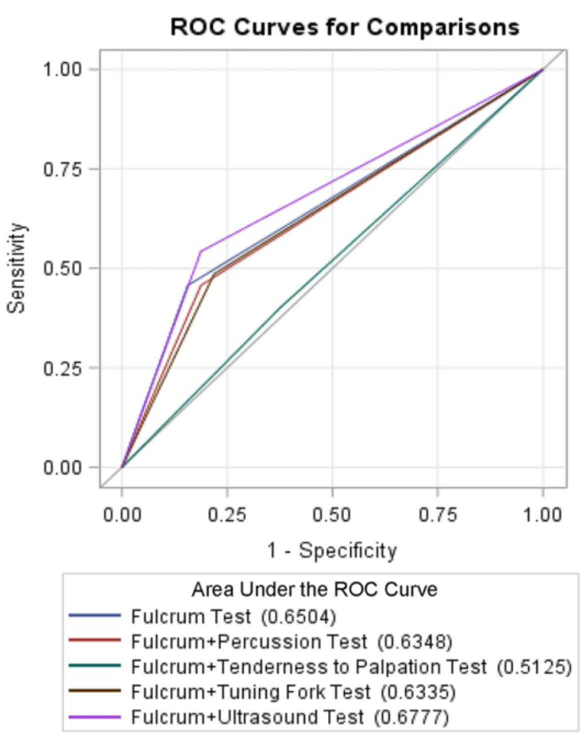

Figure 6. Post hoc receiver operating characteristic (ROC) curves for the fulcrum test and related test combinations.

activity until the next training class begins. This delay can be multiple months, or it may result in a decision to remove the individual from the training program altogether. Thus, the presence of a valid and reliable clinical diagnostic test would mitigate these challenges.

Tibial Fulcrum Test

As most participants reported pain primarily along the posteromedial aspect of the tibia, the force was applied from anterolateral to posteromedial. For participants who reported pain along other regions (eg, anterior shaft of the tibia), the force was applied perpendicular to that area (eg, from posterior to anterior). The fulcrum test demonstrated the best overall accuracy, highest positive and negative predictive values, and best PLR and NLR (Table 1). Furthermore, it was the only test with an AUC >0.6. In this sample, a positive fulcrum test increased the likelihood of a tibial SFx from 52% to 76%. By applying an a priori likelihood of 20%, based on prior research,[35] the shift in diagnosis produced by a PLR of 2.93 would result in a posttest probability of tibial SFx being 42%. The fulcrum test was the only test, albeit minimally, to improve diagnostic accuracy in ruling out a tibial SFx. The posttest probability of a patient not having a tibial SFx with a negative fulcrum test improved from 48% to 41%. Although this was the best test in this study for ruling out tibial SFx, the negative predictive value, NLR, and level of sensitivity clearly indicate it should not be used to confidently rule out tibial SFx.

TUS Test

A review of previously published studies on TUS for the evaluation of tibial bone stress injury indicated variation in test parameters of TUS intensity, frequency, sound head size, rate of sound head movement across the site of symptoms or static application to the site, and the tissue area covered during ultrasound application.[2,22,23,27,29,30] Such variability may have contributed to the wide range of sensitivity, specificity, likelihood ratios, and diagnostic accuracy reported (Table 3). Three-Megahertz TUS was selected based on the depth of penetration to the affected bony surface being <2.5 cm and its being a commonly available frequency on ultrasound units in the United States.[8,13] The TUS test produced the highest level of specificity (93.8%) and second highest PLR. The likelihood of a tibial SFx increased from 52% to 71% when the TUS test was positive. A negative ultrasound test was not useful for ruling out

TABLE 3
Diagnostic Values of Therapeutic Ultrasound in the Evaluation of Tibial Stress Fracture[a]

Study	Sensitivity	Specificity	NPV	PPV	PLR	NLR	Overall Accuracy, %
Boam[2]	0.43	0.49	0.51	0.41	0.84	1.2	46
Devereaux[7]	0.53	—	—	0.73	0.53	—	44
Moss[22]	0.91	1.0	0.89	1.0	—	0.09	95
Nitz[23]	1.0	0.80	1.0	0.89	5.0	—	93
Papalada[27]	0.82	0.67	0.13	0.99	2.5	0.26	81
Romani[29]	0.67	0.59	0.91	0.22	1.63	0.56	60
Present study	0.14	0.94	0.50	0.71	2.30	0.91	52

[a]Dashes indicate values unable to be calculated. NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value.

tibial SFx, with a shift in likelihood from 48% (no SFx) to 50%.

Tuning Fork Test

Aside from TUS, the TFT is the only other clinical test that has been critically assessed in more than a single research study on tibial SFx.[9,18,24] This study used the 128-Hz tuning fork, which was also used in the research by Fatima et al[9] and Lesho.[18] Wilder et al[38] reported that the highest diagnostic sensitivity was produced with the 256-Hz tuning fork and greater specificity with the 128- or 512-Hz tuning forks. Systematic reviews have advised caution in the use of the TFT because of the less than acceptable sensitivity.[30,34] While the TFT produced a high level of specificity (90.6%), the PLR (0.61) and NLR (1.04) did not result in an improvement in posttest probability of the presence or absence of tibial SFx. There was a greater chance of a positive TFT in a patient without a tibial SFx and a negative TFT in a patient with the presence of a tibial SFx.

Focal TTP

The focal TTP test has been recommended to distinguish between a SFx and shin splints.[1,16] Bony tenderness has been reported to provide a high level of sensitivity and positive and negative predictive values.[26] However, specific details on performing the palpatory examination were not provided. Diagnostic accuracy values of focal TTP in the assessment of tibial SFx have not been previously reported. Focal TTP, similar to the TFT, also produced PLRs and NLRs on the opposite sides of 1.0. The NLR of 1.27 indicated that there was a greater chance of a negative finding in a patient with a tibial SFx as compared with a patient without a tibial SFx.

Heel Percussion Test

To our knowledge, no detailed report of administration or critical analysis of the heel percussion test in the evaluation of individuals with leg pain is available in the literature. The magnitude of percussion to the calcaneus has not been reported in the literature, and the investigators did not standardize the magnitude of impact applied to the heel in this study. The percussion test had the second highest level of specificity and the third highest PLR. A positive percussion test shifted the probability of having a tibial SFx from 52% to 67%.

Combination of Tests

The highest level of sensitivity, 0.67, was attained with a test item cluster of the fulcrum, TUS, and focal TTP tests. Thus, the presence of negative fulcrum, TUS, and TTP tests provides a 67% likelihood of the patient's not having a tibial SFx. Positive fulcrum and ultrasound tests in combination yielded a slightly higher probability of correct diagnosis with an AUC of 0.68 compared with the fulcrum test in isolation. Evaluation of test item clusters did not significantly improve the ability to identify which patients had a tibial SFx compared with the fulcrum test in isolation.

Study Limitations

Some limitations of this study are noteworthy. Clinical examination tests were counterbalanced to reduce the potential of carryover effects. Randomizing the order of tests may have yielded different results. Military duty station assignment resulted in a need to have a second examiner, which introduced heterogeneity and may have influenced the findings. The research team implemented training to standardize application of the tests and minimize the possible confounding influence of multiple examiners. The use of radionuclide bone scan as the gold standard imaging study was a limitation of this study. Radionuclide bone scan was used based on its high level of sensitivity, moderate specificity, and the rapid clinical availability to get such examinations within <48 hours after clinical assessment for leg pain. Research has demonstrated a comparable level of sensitivity between radionuclide bone scan and MRI.[39] While MRI provides the highest level of specificity in the evaluation of bone stress injuries, delays >1 week for routine MRI of tibial pain made its use unrealistic in time-critical patient management. Although radiographs are recognized as providing a low level of sensitivity in the evaluation of lower extremity SFx, they were part of the imaging sequence protocol used at the institution where patients underwent imaging. Imaging of leg pain using MRI and the grading scale proposed by Fredericson et al[10] may have yielded differences in diagnostic accuracy results for these clinical tests. Another possible limitation of this study is that 34% (17/50) of the sample was identified as having radionuclide bone scan findings consistent with an SFx on an asymptomatic side or location. Of those 17 individuals, 11 (65%) did have imaging findings consistent with a tibial SFx. Kiuru et al[17] reported that 44% of military trainees with positive findings on MRI did not report any symptoms, but it should be noted that those with an absence of symptoms were identified as having less severe (<grade 3) bone stress injury. This increased uptake in the absence of reportable symptoms may indicate a preclinical condition or could also be the result of an individual's underreporting musculoskeletal symptoms, which has been noted to be common in military personnel.[6,17,32]

Additional limitations of this study include its small sample size and wide range of the 95% CIs. The study sample included only male patients; thus, it is unknown if female patients, recognized as having a higher incidence of SFx, would have the same response to the clinical tests. Future research may benefit from consideration of other methods to stress the tibia (eg, evaluation of change in VAS during single-leg stance as compared with nonweightbearing), further clarification of physical examination components (eg, palpation technique to include magnitude of pressure application), the inclusion of women, and the incorporation of more details from patients' subjective examination (eg, report of night pain or pain pattern postactivity).

CONCLUSION

When managing patients with exercise-related leg pain and suspicion of tibial SFx, all elements of the patient evaluation need to be incorporated into the development of a working diagnosis. Although the results of this study indicate superiority of the tibial fulcrum test over more commonly reported clinical tests, the diagnostic accuracy is unacceptably low in comparison with advanced imaging (eg, MRI). Further research incorporating the fulcrum test and additional examination components may further improve diagnostic accuracy. Currently, however, clinicians should continue to rely on advanced imaging modalities when definitive diagnosis of suspected bone stress injuries is necessary for patient management.

REFERENCES

1. Batt ME, Ugalde V, Anderson MW, Shelton DK. A prospective controlled study of diagnostic imaging for acute shin splints. *Med Sci Sports Exerc*. 1998;30(11):1564-1571.
2. Boam WD, Miser WF, Yuill SC, Delaplain CB, Gayle EL, MacDonald DC. Comparison of ultrasound examination with bone scintiscan in the diagnosis of stress fractures. *J Am Board Fam Pract*. 1996;9(6): 414-417.
3. Brukner P, Bennell K, Mattheson G. *Stress Fractures*. Blackwell Science; 1999.
4. Casterline M, Osowski S, Ulrich G. Femoral stress fracture. *J Athl Train*. 1996;31(1):53-56.
5. Changstrom BG, Brou L, Khodaee M, Braund C, Comstock RD. Epidemiology of stress fracture injuries among US high school athletes, 2005-2006 through 2012-2013. *Am J Sports Med*. 2015;43(1):26-33.
6. Cohen BS, Pacheco BM, Foulis SA, et al. Surveyed reasons for not seeking medical care regarding musculoskeletal injury symptoms in US Army trainees. *Mil Med*. 2019;184(5-6):e431-e439.
7. Devereaux MD, Parr GR, Lachmann SM, Page-Thomas P, Hazleman BL. The diagnosis of stress fractures in athletes. *JAMA* 1984;252(4): 531-533.
8. Draper DO, Castel JC, Castel D. Rate of temperature increase in human muscle during 1 MHz and 3 MHz continuous ultrasound. *J Orthop Sports Phys Ther*. 1995;22(4):142-150.
9. Fatima ST, Jeilani A, Duha M, et al. Validation of tuning fork test in stress fractures and its comparison with radionuclide bone scan. *J Ayub Med Coll Abbottabad*. 2012;24(3-4):180-182.
10. Fredericson M, Bergman AG, Hoffman KL, Dillingham MS. Tibial stress reaction in runners: correlation of clinical symptoms and scintigraphy with a new magnetic resonance imaging grading system. *Am J Sports Med*. 1995;23(4):472-481.
11. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Phys Ther*. 2001;81(9):1546-1564.
12. Giladi M, Nili E, Ziv Y, Danon YL, Aharonson Z. Comparison between radiography, bone scan, and ultrasound in the diagnosis of stress fractures. *Mil Med*. 1984;149(8):459-461.
13. Hayes BT, Merrick MA, Sandrey MA, Cordova ML. Three-MHz ultrasound heats deeper into the tissues than originally theorized. *J Athl Train*. 2004;39(3):230-234.
14. Iwamoto J, Takeda T. Stress fractures in athletes: review of 196 cases. *J Orthop Sci*. 2003;8(3):273-278.
15. Johnson AW, Weiss CB, Wheeler DL. Stress fractures of the femoral shaft in athletes—more common than expected: a new clinical test. *Am J Sports Med*. 1994;22(2):248-256.
16. Kahanov L, Eberman LE, Games KE, Wasik M. Diagnosis, treatment, and rehabilitation of stress fractures in the lower extremity in runners. *Open Access J Sports Med*. 2015;6:87-95.
17. Kiuru MJ, Niva M, Reponen A, Pihlajamaki HK. Bone stress injuries in asymptomatic elite recruits: a clinical and magnetic resonance imaging study. *Am J Sports Med*. 2005;33(2):272-276.
18. Lesho EP. Can tuning forks replace bone scans for identification of tibial stress fractures? *Mil Med*. 1997;162(12):802-803.
19. Maitra RS, Johnson DL. Stress fractures: clinical history and physical examination. *Clin Sports Med* 1997;16(2):259-274.
20. Matheson GO, Clement DB, Mckenzie DC Taunton JE, Lloyd-Smith JE, MacIntyre JG. Stress fractures in athletes: a study of 320 cases. *Am J Sports Med*. 1987;15(1):46-58.
21. Moen MH, Tol JL, Weir A, Steunebrink M, DeWinter TC. Medial tibial stress syndrome: a critical review. *Sports Med*. 2009:39(7):523-546.
22. Moss A, Mowat AG. Ultrasonic assessment of stress fractures. *Br Med J (Clin Res Ed)*. 1983;286(6376):1479-1480.
23. Nitz AJ, Scoville CR. Use of ultrasound in early detection of stress fractures of the medial tibial plateau. *Mil Med*. 1980;145(12):844-846.
24. Nussbaum ED, Gatt CJ, Bjornarra J, Yang C. Evaluating the clinical tests for adolescent tibial bone stress injuries. *Sports Health*. 2021; 13(5):502-510.
25. Nye NS, Covey CJ, Sheldon L, et al. Improving diagnostic accuracy and efficiency of suspective bone stress injuries: algorithm and clinical prediction rule. *Sports Health*. 2016;8(3):278-283.
26. Nye NS, Covey CJ, Pawlak M, Olsen C, Boden BP, Beutler AI. Evaluating and algorithm and clinical prediction rule for diagnosis of bone stress injuries. *Sports Health*. 2020;12(5):449-455.
27. Papalada A, Malliaropoulos N, Tsitas K, et al. Ultrasound as a primary evaluation tool of bone stress injury in elite track and field athletes. *Am J Sports Med*. 2012;40(4):915-919.
28. Rauh MJ, Macera CA, Trone DW, Shaffer RA, Brodine SK. Epidemiology of stress fracture and overuse injury in women recruits. *Med Sci Sports Exerc*. 2006;38(9):1571-1577.
29. Romani WA, Perrin DH, Dussault RG, Ball DW, Kahler DM. Identification of tibial stress fractures using therapeutic continuous ultrasound. *J Orthop Sports Phys Ther*. 2000;30(8):444-452.
30. Schneiders AG, Sullivan AJ, Hendrick PA, et al. The ability of clinical tests to diagnose stress fractures: a systematic review and meta-analysis. *J Orthop Sports Phys Ther*. 2012;42(9):760-771.

31. Shwayhat AF, Linenger JM, Hofherr LK, Slyment DJ, Johnson CW. Profiles of exercise history and overuse injuries among United States Navy Sea, Air, and Land (SEAL) recruits. *Am J Sports Med*. 1994; 22(6):835-840.
32. Smith L, Westrick R, Sauers S, et al. Underreporting of musculoskeletal injuries in the US Army: findings from an infantry Brigade Combat Team survey study. *Sports Health*. 2016;8(6):507-513.
33. Snyder RA, Koester MC, Dunn WR. Epidemiology of stress fractures. *Clin Sports Med*. 2006:25(1):37-52.
34. Toney CM, Games KE, Winkelmann ZK, Eberman LE. Using tuning-fork tests in diagnosing fractures. *J Athl Train*. 2016;51(6):498-499.
35. Trone DW, Villasenior A, Macera CA. Stress fracture and attrition in Basic Underwater Demolition SEAL trainees. *J Spec Ops Med*. 2006; 6(1):32-40.
36. Weishaar MD, McMillian DJ, Moore JH. Identification and management of 2 femoral shaft stress injuries. *J Orthop Sports Phys Ther*. 2005:35(10):665-673.
37. Wentz L, Liu PY, Haymes E, Ilich JZ. Females have a greater incidence of stress fractures than males in both military and athletic populations: a systematic review. *Mil Med*. 2011:176(4): 420-430.
38. Wilder RP, Vincent HK, Stewart J, Pack C, Vincent KR. Clinical use of tuning forks to identify running-related stress fractures: a pilot study. *Athl Train Sports Health Care*. 2009;1(1):12-18.
39. Wright AA, Hegedus EJ, Lenchik L, Kuhn KJ, Santiago L, Smoliga JM. Diagnostic accuracy of various imaging modalities for suspected lower extremity stress fractures. *Am J Sports Med*. 2015;44(1): 255-263.