International Journal of Health Sciences and Research
www.ijhsr.org ISSN: 2249-9571

Review Article

A Comparison of Diagnostic Imaging and Vibrating Tuning Forks in the Detection of Fractures

Derek Charles[1*], Ashleigh Elkins[*], Andrew Kneeburg[*], Kelsey Nikkila[*]

[1]Assistant Professor,
[*]Department of Physical Therapy, Tennessee State University 3500 John A Merritt Boulevard Nashville TN.

Corresponding Author: Derek Charles

ABSTRACT

INTRODUCTION: The purpose of this review was to determine the effectiveness of tuning forks compared to diagnostic imaging in ruling out and ruling in fractures.
METHODS: Multiple databases including Ebscohost, Pub Med, and Sport Discus were utilized in the literature review. Keywords included tuning fork test, fracture detection, x-ray, ultrasound, MRI, bone scan, CT scan, and diagnostic imaging. Inclusion criteria included Oxford Level of Evidence 3 or higher, statistics on sensitivity, specificity, and reliability, peer-reviewed, and English only journals. 9 articles published between 1997 and 2016 were included in the synthesis of the results.
RESULTS: For the tuning fork test, sensitivity ranged from 75% to 92%, specificity ranged from 18% to 94%, positive likelihood ratios were between 1.1 and 16.5 and negative likelihood ratios were between .09 and 1.62. Either pain induction or reduction of sound transmission while listening with a stethoscope was used for fracture detection but there was not standardization in training or methodology.
CONCLUSIONS: MRI continues to be the preferred method of fracture detection due to its high sensitivity and specificity. Computed Tomography and bone scans also demonstrated high specificity and sensitivity. While a tuning fork is simple to administer and cheap, it has low diagnostic capabilities when used alone, and therefore other imaging is still necessary for confirmation even with a positive or negative test.
***KEY WORDS*:** tuning forks, fractures, x-rays, MRI, diagnostic imaging, diagnostic accuracy

INTRODUCTION

Stress fractures are a partial or incomplete fractures caused by repetitive weight bearing activities such as running, jumping, or marching and are classified into two types: fatigue and insufficiency. [1] A fatigue fracture is more common and is typically seen in an active population, such as runners or military recruits, where training intensity rapidly increases over a short period of time. Under stress a bone will generally deform from its normal shape and recoil without damage; but if a bone is loaded beyond its ability to withstand forces, it plastically deforms which increases the potential for a fracture. [1] Osteoclastic activity outpaces osteoblastic activity resulting in the gradual weakening and eventual failure of bone. This results in ischemia caused by repetitive loading and micro damage to the bone's capillary supply. [2] An insufficiency fracture is usually found in older adults whereby normal stress is placed on a mineral deficient bone due to osteoporosis and the bone fractures from loading, often during normal activities of daily living. [1] In a review of the literature pertaining to the etiology and diagnosis of stress fractures, Reeder and colleagues reported new bone ossification takes up to 90 days to complete, exposing weakened bone to stress for extended periods of time and ultimately resulting in a fracture. [3]

Stress fractures can occur at any location but certain areas are more susceptible, depending on the nature of the activity. For example, the navicular is small in diameter and exposed to repetitive bouts of loading during walking, and even more in running. [3] Fractures result from the coupling of a flexible foot where the arch is flattened and the navicular repeatedly comes into contact with the ground, increasing the potential for injury. Other susceptible sites include the tibia, talus, sesamoid of the 1st toe, and base of the 5th metatarsal. [1, 3, 4]

Diagnostic imaging such as radiographs, bone scintigraphy, ultrasound, computed tomography and MRI, is traditionally used in the detection of bony injuries. [5-9] Tuning forks are a diagnostic tool that can theoretically detect fractures by transmission of vibrations through a bone and if present, the patient should complain of pain due to the percussion. [4] The purpose of this review was to determine the effectiveness of tuning forks compared to diagnostic imaging in the detection of fractures.

METHODS

The review of the literature was conducted utilizing multiple databases, including Ebscohost, Pub Med, and Sport Discus. Keyword searches included tuning fork test, stress fractures, x-ray, ultrasound, MRI, bone scan, CT scan, and diagnostic imaging. Articles from the initial search were eliminated if it was determined the study did not examine fractures, tuning forks, diagnostic imaging, or include statistics on reliability, sensitivity, or specificity. The abstracts of the remaining articles were reviewed by one of the five investigators and the following inclusion criteria were applied: Oxford Level of Evidence 3 or higher, peer-reviewed, and English only journals. The method of detection was either reproduction of pain or reduction of sound transmission while listening through a stethoscope. 9 articles published between 1997 and 2016 were included in the synthesis of the results.

RESULTS

Sensitivity

Overall, the tuning fork test exhibited moderate to high sensitivity. According to a systemic review by Mugunthan et al., [10] sensitivity ranged from 75% to 92% with common lower extremity stress fractures when performed on homogenous populations. However, while a positive tuning fork test could be sufficient to warrant the initiation of treatment while waiting of the results of imaging, it does not display consistent or sufficient sensitivity to rule out a stress fracture if no pain was present during the test.

Specificity

The results yielded a wide spread in the specificity, or the ability of a tuning fork test to rule in a fracture. According to the same review by Mugunthan et al., [10] specificity varied from 18% to 94%, meaning there is high potential for false positive results. The significant fluctuations in specificity were likely due to the fact pain can occur even without an actual fracture, making the tuning fork method unreliable for ruling in a fracture. [11, 12]

Likelihood Ratios

The likelihood ratios provided additional information about swings in probability for potential fractures.[13] Positive likelihood ratios were between 1.1 and 16.5 while negative likelihood ratios were between .09 and 1.62. [10-12, 14] A positive LR greater than 10 indicated a strong shift towards the probability of a fracture while a negative LR less than .10 indicated a strong shift towards the nonexistence of a fracture. [13] The large range in likelihood ratios does not allow any definitive clinical associations. It should be noted the study that concentrated on the use of a 128 Hz tuning fork with suspected distal fibula fractures after an inversion ankle injury in individuals with positive Ottawa ankle rules had the highest positive and negative likelihood ratios, which indicated the tuning fork test was valid in the detection of fractures at that location. [15]

Table 1. Results of the studies reviewed related to the use of tuning forks

Reference	Participants	Exposure	Outcome	Key Findings	Comments
Fatima, Jeilani, Abbasi, et al.[26] *J Ayub Med Coll* 2012: 24(3-4): 180-182	55 subjects, ages 18-28, with negative x-rays to the tibia or fibula.	Each participant had a TF test to the painful site at 128 Hz followed by bone scintigraphy.	67 total stress fractures were located with scintigraphy. The TF test was positive for 53 participants.	Sensitivity of the TF test was 79% and specificity was 63%.	Caution should be taken when interpreting the results of a TF test to rule in or out a stress fracture.
Reference	**Participants**	**Exposure**	**Outcome**	**Key Findings**	**Comments**
Dissmann, Han[15] *Emerg Med J.* 2006; 23(10): 788-790.	Pilot study with 49 patients, ages 12-84, who were examined by a single instigator. Patients then received radiographs for a reference standard	Patients with ''Ottawa positive'' findings had a 128 Hz TF applied to the lateral malleolus or distal fibula	Sensitivity was 100% and specificity was 61% for the lateral malleolus. Sensitivity was 100% and specificity was 95% for the distal fibula.	Diagnostic accuracy was 65% for the lateral malleolus and 96% for the distal fibula. The +LR was 2.59 for the lateral malleolus and 22 for the distal fibula.	The addition of a TF test for ''Ottawa positive'' patients may lead to a reduction in ankle radiographs. This may be relevant when radiological facilities are not readily available or where access has to be prioritized.
Reference	**Participants**	**Exposure**	**Outcome**	**Key Findings**	**Comments**
Wilder, Vincent, Stewart, et al[12]. *Athletic Training and Sports Health Care* 2009;1(1):12-18	45 runners (31.2 +/- 13.1 years) with stress fracture symptoms who had already received radiographs, MRI, and a bone scan.	Participants had a TF at 128, 256, and 512 Hz placed over the painful site and rated symptom intensity on a 0-3 scale.	Higher fork-induced pain ratings were correlated with positive imaging (r = 0.156, P = .056). The odds risk of a positive image was 5.91 with a fork-induced pain rating of 3, compared with pain ratings ≤ 2.	The 256-Hz fork elicited the highest pain ratings and sensitivity for fracture detection (range, 77.7%–92.3%), and the 512-Hz fork elicited the lowest (range, 50%–76.9%) of all diagnostic tests.	While the 256-Hz tuning fork had a high sensitivity, the test had very poor specificity: only around 20%. So individuals without stress fractures will nevertheless have pain during a tuning fork test, making it difficult to use as a diagnostic aid.
Reference	**Participants**	**Exposure**	**Outcome**	**Key Findings**	**Comments**
Moore[24] *J Athl Train* 2009; 44(3):272-274	19 males and 18 females, ranging from 7-60 years old,were evaluated for possible fractures < 7 days old.	A 128 Hz TF was placed distal to the suspected fracture while a stethoscope was placed proximally	Results included 10 true positives, 20 true negatives, 5 false positives and 2 false negatives.	Sensitivity was 83% and specificity was 80%. Diagnostic accuracy was 81%. Use of a stethoscope did not alter the sensitivity, but specificity increased to 92% and diagnostic accuracy to 89%.	A TF test has increases specificity and diagnostic accuracy when combined with a stethoscope.
Reference	**Participants**	**Exposure**	**Outcome**	**Key Findings**	**Comments**
Mugunthan, Doust, Kurz, et al.[10] *BMJ Open.* 2014; 4(8):e005238.	6 studies were pooled with over 329 patients between the ages of 7-60.	Various LE locations were examined using pain induction or reduced sound conduction with a stethoscope. 4 studies used a 128 Hz TF while 2 used different frequencies.	Fracture prevalence ranged from 10% to 80%. Sensitivity ranged from 75% to 100% and specificity ranged from 18% to 95%.	The heterogeneous specificity resulted in a high proportion of false positives due to pain reproduction with a 256 Hz TF in subjects without fractures.	In the study with patients positive for the Ottawa ankle rules, sensitivity was 100%, although there were only five patients with fractures.
Reference	**Participants**	**Exposure**	**Outcome**	**Key Findings**	**Comments**
Schneiders, Sullivan, Hendrick, et al.[14] *JOSPT* 2012; *42*(9), 760-771	9 studies reviewed with 420 subjects, ranging from 18-45 years old. 6 studies involved both genders and 2 were men only. 5 studies examined multiple LE locations and 1 study focused on osteoporosis.	Inclusion criteria included at least 1 radiological reference test, computation of diagnostic values, no age restrictions, published between 1950 and 2011, and only LE stress fractures.	Sensitivity ranged from 43-100%, specificity 0-100%, +LR 1.04 to 3.67, -LR .06 to 1.62, DOR values 1.08-171.20, and QUADAS score from 7-23.	All studies examined athletes and/or military personnel, confirming these populations are at risk. But the results can't be generalized to other populations including those with underlying pathologies such as osteoporosis.	Other variances in the studies included time frames for detection, the definition of a stress fracture, methods for quantifying pain, and frequency of the TF, and type of reference test.

Table 1: Continued…

Reference	Participants	Exposure	Outcome	Key Findings	Comments
Toney, Games, Winkelmann, et al.[12] *J Athl Train* 2016; 51(6): 498-499.	6 studies reviewed the accuracy of a 128 Hz TF method for pain induction and reduction of sound transmission. Patient age range was 7 to 84.	All types of LE stress fracture were measured against a MRI, x-ray, or bone scan. Case series, case-controls studies, or narrative reviews weren't assessed.	Sensitivity ranged from 75% to 92% and specificity ranged from 18% to 94%. The +LR was 1.1 to 16.5 and the –LR was 0.09 to 0.49.	Overall the review demonstrated a high ability to rule out a fracture but there was much greater variability in the ability to rule in a fracture.	Further research is needed to determine if TF tests are only effective in certain physiologic adaptations, such as cortical bone discontinuity, or are affected by the timing of the injury. Additionally, larger sample sizes with various fracture types and locations using higher quality protocols, including blinding of testers to the reference standard are recommended.
Reference	**Participants**	**Exposure**	**Outcome**	**Key Findings**	**Comments**
Kazemi and Roscoe[4] *International Sports Journal.* 2000; 4(2): 1-8	46 consecutive patients, 2-91 years old, presented to an ambulatory center with acute UE, LE, or rib pain <10 days due to trauma	An orthopedist examined each patient without inclusion of a TF. The 2nd examiner placed a 128 and 256 HZ TF over the involved side and a 256 HZ TF on the uninvolved side. Radiographs were then performed.	37/46 patients had fractures. Sensitivity was 86.8% and specificity was 50% for the TF. The Kappa was .35	There was marginal reproducibility and poor specificity in the TF test, regardless of the frequency. Avulsion fractures and radial head fractures were especially difficult to detect.	Even in the presence of a negative TF test, follow up imaging is recommended due to the high rate of false negatives.
Reference	**Participants**	**Exposure**	**Outcome**	**Key Findings**	**Comments**
Lesho[17] *Mil Med.* 1997; 162(12): 802-803.	46 military personnel, average age 25 years old, with exercise related shin pain localized to the tibia.	128 Hz TF was applied to the anterior tibia. Pain reproduction was considered positive. Patients also had a bone scan.	The sensitivity and specificity of the TF test were 75% and 67%.	The TF test had poorer diagnostic capabilities compared to x-ray and especially to the MRI's 100% and 86% sensitivity and specificity.	The sensitivity of the TF test is not enough to rule out a stress fracture with a negative or pain free test.

Reliability

Overall, the results did not conclude a tuning fork test was more reliable for fracture detection based on anatomical location or type of fracture. Articles reviewed included tibial stress fractures, femoral neck fractures, ankle inversion injuries, foot and ankle stress fractures, rib fractures, and unspecified locations. [14] The fractures covered in the literature review were characterized numerous types, including stress fractures, avulsion fractures, greenstick fractures, comminuted fractures, and hairline fractures. [16] All studies included in this review either utilized pain induction and/or reduction of sound transmission as the method for fracture detection. There did not appear to be standardization in training or use of the tuning forks and future research should determine whether one method of detection is preferable. [10]

The clinical presentation of each fracture type may be unique with varying physiological features and this could have affected the accuracy of detection. [16] A tuning fork test is most accurate when the periosteum is completely disturbed or there is minimal soft tissue around the bone. [12] Generally, tuning fork tests were significantly less accurate (sensitivity = 75% [95% confidence interval {CI} = 57%, 87%]; specificity = 67% [95% CI = 44%, 84%]; positive likelihood ratio = 2.2 [95% CI = 1.1, 4.5]; negative likelihood ratio = 0.37 [95% CI = 0.18, 0.77]) for stress fractures, when the outer layer of bone was still intact. [17] The results of the studies reviewed are summarized in Table 1.

DISCUSSION

While MRI has the highest cost, [18] it displays the greatest detail and has the highest sensitivity and specificity of any imaging technique, [6,19] and is therefore the preferred choice for the detection of stress fractures. [20] CT scans are also more expensive than other forms of imaging but have a high level of specificity. Bone scans cost less than MRI or CT scans and have high sensitivity [18] however, bone scans have an increased rate of false positives in the presence of an infection or a tumor so it may not be suitable for all populations. [2, 6, 19] X-rays have a low cost and accessibility that make it an easy diagnostic tool to obtain but has less detail and low initial sensitivity than other techniques. [18, 21] While a tuning fork is simple to administer and extremely cheap, [22] it has low sensitivity and specificity when used alone, and therefore other imaging is still necessary for confirmation even with a positive test. [17] Table 2 compares the costs and accuracy of the various diagnostic tools. [4]

Table 2. Comparison of the cost and accuracy of various diagnostic tools

Test	Cost	Sensitivity/Specificity
MRI	$750-$2200	High sensitivity and specificity
Radiographs	$90-$950	Poor initial sensitivity
Bone Scans	$150-$600	High sensitivity
CT Scan	$950-$2000	Low sensitivity/high specificity
Tuning Fork	$5-$6	Low sensitivity/low specificity when used alone

It is proposed stress fractures have two different stages: a cortical bone weakening stage and a cortical bone discontinuity stage. [1] Due to this, use of a tuning fork should be studied in the different stages because this may explain the low specificity in previous studies. If results show the tuning fork test has a greater sensitivity and/or specificity once the fracture progresses to the discontinuity stage, then the tuning fork test may help clinicians recognize the signs and symptoms associated with a stress fracture earlier and therefore assist in the determination of additional imaging as appropriate.

Minimizing delays in detection has an impact on the ability to return to activities of daily living, occupational duties, and recreational activities in a timely manner. There are financial implications due to time spent away from work [21] and evidence suggests athletes who wait more than three weeks after the onset of symptoms ultimately have a delay and increased timeline for return to sport. [23] Ohta-Fukushima analyzed 222 stress fractures in 208 athletes, determining there was a statistically significant difference in time for return to competition between athletes who visit a hospital within three weeks of symptoms and those who waited longer. It took an average of 14.2 weeks to return to sport for those athletes who visited the hospital within three weeks of symptoms and sometimes over six months for those who waited longer than three weeks. [23] So regardless of the method of detection, early diagnosis and intervention are keys for return to activity.

CONCLUSION

Additional research is necessary to conclude if tuning-fork tests are beneficial under particular conditions, such as a lack of cortical bone continuity, [24] in conjunction with other tests, [25] or are influenced by the timing of the injury. [26] Also, investigators should use greater sample sizes with different types of fractures, different locations, standardized protocols, and blinding of testers. The method of pain induction using the 128 Hz tuning fork requires less training than use of a stethoscope to listen for sound reduction and therefore could be the focus of future research. In summary, caution should be taken when interpreting the results of a TF test to rule in or rule out a fracture. Diagnostic imaging should continue to be the standard for identifying fractures regardless of the suspected type or location.

REFERENCES

1. Romani WA, Gieck J H, Perrin DH, et al. Mechanisms and management of stress fractures in physically active persons. *J Athl Train*. 2002; 37(3): 306–314.
2. Otter MW, XY Q, Rubin CT, et al. Does bone perfusion/reperfusion initiate bone remodeling and the stress fracture syndrome? Med Hypotheses. 1999; 53(5): 363-368.

3. Reeder MT, Dick BH, Atkins JK, et al. Stress fractures. Current concepts of diagnosis and treatment. *Sports Med*. 1996; 22(3): 198-212.
4. Kazemi M, Roscoe M. Is the tuning fork test a reliable tool in detecting acute simple fractures? *International Sports Journal*. 2000; 4(2): 1-8.
5. Ishibashi Y, Okamura Y, Otsuka H, et al. Comparison of scintigraphy and magnetic resonance imaging for stress injuries of bone. *Clin J Sport Med*. 2002; 12: 79-84.
6. Zwart A, Beeres F, Ring D, et al. MRI as a reference standard for suspected scaphoid fractures. *Br J Radiol*. 2012; 1098-1101.
7. Bretlau T, Christensen OM, Edstrom P, et al. Diagnosis of scaphoid fracture and dedicated extremity MRI. *ACTA Orthopedics of Scandinavia*. 1999; 70(5): 504-508.
8. Cabarrus MC, Ambekar A, Lu Y, et al. MRI and CT of insufficiency fractures of the pelvis and the proximal femur. *Am J Roentgenol*. 2008; 191: 995-1001.
9. Gaeta M., Minutoli F, Scribano E, et al. CT and MR imaging findings in athletes with early tibial stress injuries: comparison with bone scintigraphy findings and emphasis on cortical abnormalities. *Radiology*. 2005; 235(2): 553-61.
10. Mugunthan K, Doust J, Kurz B, et al. Is there sufficient evidence for tuning fork tests in diagnosing fractures? A systematic review. *BMJ Open*. 2014; 4(8):e005238.
11. Wilder RP, Vincent HK, Stewart J, et al. Clinical use of tuning forks to identify running-related stress fractures: a pilot study.*Athletic Training and Sports Health Care*. 2009; 1(1): 12-18.
12. Toney CM, Games KE, Winkelmann ZK, et al. Using tuning-fork tests in diagnosing fractures. *J Athl Train*. 2016; 51(6):498-499.
13. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet*. 2005; 365(9469):1500–1505.
14. Schneiders AG, Sullivan SJ, 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.
15. Dissmann PD, Han KH. The tuning fork test—a useful tool for improving specificity in "Ottawa positive" patients after ankle inversion injury. *Emerg Med J*. 2006; 23(10): 788-790.
16. Müller ME, Koch P, Nazarian S, et al. The comprehensive classification of fractures of long bones. New York, NY: Springer Science and Business Media; 1990.
17. Lesho EP. Can Tuning Forks Replace Bone Scans For Identification of Tibial Stress Fractures? *Mil Med*. 1997; 162(12): 802-803.
18. Average Out-of-Pocket Costs for Uninsured and Self-Pay Patients | El Camino Hospital. 2015; Retrieved from http://www.elcaminohospital.org/Patient_Services/Billing_Insurance/Patient_Services_Pricing_and_Estimates/Average_Out-of-Pocket_Costs_for_Uninsured_Patients
19. Fredericson M, Bergman AG, Hoffman KL, et al. 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.
20. Liong SY, Whitehouse RW. Lower extremity and pelvic stress fractures in athletes. *Br J Radiol*. 2012; 85: 1148-1156.
21. Matheson GO, Clement DB, McKenzie DC, et al. Stress fractures in athletes. A study of 320 cases. *Am J Sports Med*. 1987; 15(1): 46-58.
22. TENSnet - Over 25,000 health and medical products at everyday low prices. 2015; Retrieved from http://TENSNET.COM
23. Fukushima-Ohta M, Mutoh Y, Takasugi S, et al. Characteristics of stress fractures in young athletes under 20 years. *J Sports Med Phys Fitness*. 2002; 42: 198-206.
24. Moore M. The use of a tuning fork and stethoscope to identify fractures. *J Athl Train*. 2009; 44(3): 272-274.
25. Bachmann LM, Kolb E, Koller MT, et al. Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review. *BMJ: British Medical Journal*. 2003; 326(7386): 417.
26. Fatima ST, Jeilani A, Abassi N, et al. Validation of tuning fork test in stress fractures and its comparison with radionuclide bone scan. *J Ayub Med Coll*. 2012; 24(3-4): 180-182.

How to cite this article: Charles D, Elkins A, Kneeburg A et al. A comparison of diagnostic imaging and vibrating tuning forks in the detection of fractures. Int J Health Sci Res. 2017; 7(8):473-478.

****