Using Tuning-Fork Tests in Diagnosing Fractures

Chelsey M. Toney, LAT, ATC; Kenneth E. Games, PhD, LAT, ATC;
Zachary K. Winkelmann, MS, LAT, ATC; Lindsey E. Eberman, PhD, LAT, ATC

Department of Applied Medicine and Rehabilitation, Indiana State University, Terre Haute

Reference/Citation: Mugunthan K, Doust J, Kurz B, Glasziou P. Is there sufficient evidence for tuning fork tests in diagnosing fractures? A systematic review. *BMJ Open*. 2014;4(8):e005238.

Clinical Question: Does evidence support the use of tuning-fork tests in the diagnosis of fractures in clinical practice?

Data Sources: The authors performed a comprehensive literature search of AMED, CAB Abstracts, CINAHL, EMBASE, MEDLINE, SPORTDiscus, and Web of Science from each database's start to November 2012. In addition, they manually searched reference lists from the initial search result to identify relevant studies. The following key words were used independently or in combination: *auscultation, barford test, exp fractures, fracture, tf test, tuning fork*.

Study Selection: Studies were eligible based on the following criteria: (1) primary studies that assessed the diagnostic accuracy of tuning forks; (2) measured against a recognized reference standard such as magnetic resonance imaging, radiography, or bone scan; and (3) the outcome was reported using pain or reduction of sound. Studies included patients of all ages in all clinical settings with no exclusion for language of publication. Studies were not eligible if they were case series, case-control studies, or narrative review papers.

Data Extraction: Potentially eligible studies were independently assessed by 2 researchers. All relevant articles were included and assessed for inclusion criteria and value using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool, and relevant data were extracted. The QUADAS-2 is an updated version of the original QUADAS and focuses on both the risk of bias and applicability of a study through a series of

questions. A third researcher was consulted if the 2 initial reviewers did not reach consensus. Data for the primary outcome measure (accuracy of the test) were presented in a 2 × 2 contingency table to show sensitivity and specificity (using the Wilson score method) and positive and negative likelihood ratios with 95% confidence intervals.

Main Results: A total of 62 citations were initially identified. Six primary studies (329 patients) were included in the review. The 6 studies assessed the accuracy of 2 tuning-fork test methods (pain induction and reduction of sound transmission). The patients ranged in age from 7 to 84 years. The prevalence of fracture in these patients ranged from 10% to 80% using a reference standard such as magnetic resonance imaging, radiography, or bone scan. The sensitivity of the tuning-fork tests was high, ranging from 75% to 92%. The specificity of the tuning-fork tests had a wide range of 18% to 94%. The positive likelihood ratios ranged from 1.1 to 16.5; the negative likelihood ratios ranged from 0.09 to 0.49.

Conclusions: The studies included in this review demonstrated that tuning-fork tests have some value in ruling out fractures. However, strong evidence is lacking to support the use of current tuning-fork tests to rule in a fracture in clinical practice. Similarly, the tuning-fork tests were not statistically accurate in the diagnosis of fractures for widespread clinical use. Despite the lack of strong evidence for diagnosing all fractures, tuning-fork tests may be appropriate in rural and remote settings in which access to the gold standards for diagnosis of fractures is limited.

Key Words: sensitivity, specificity, sound transmission

COMMENTARY

When performing a diagnostic test, it is important for athletic trainers to understand the diagnostic accuracy of the test to help rationalize either ruling in or ruling out a specific condition. The increased costs of imaging have demonstrated a need for a cost-effective and reproducible method or tool that would aid athletic trainers in the physical examination of suspected fractures in clinical practice.¹ The current reference standard for diagnosing fractures is magnetic resonance imaging, radiography, or bone scan. Tuning forks may be a cost-effective screening tool for practitioners examining patients with suspected fractures.

Standard practice and training in using tuning forks for clinical practice are lacking. However, it has been suggested that a positive tuning-fork test is demonstrated either by way of increased pain from placing the vibrating

tuning fork over the fracture site or by an audible decrease in sound conduction (detected via stethoscope) when placing the vibrating tuning fork over a bony prominence distal to the fracture site and comparing it with the healthy limb.^{2,3}

In this systematic review, Mugunthan et al⁴ evaluated the accuracy of tuning-fork tests in clinical practice for diagnosing fractures. The 6 articles included in the review demonstrated high sensitivity, or the ability to rule out a fracture, with a range of 75% to 92%; however, there was a much wider range from 18% to 94% for specificity, or the ability to rule in a fracture. Based on these results, 75% to 92% of evaluations using tuning-fork tests completed on similar populations will accurately detect the presence of a fracture, whereas 18% to 94% of evaluations using tuning-fork tests will accurately detect the absence of a fracture. The data reflect a wide range of specificity, resulting in a high proportion of false-positive test results. Pain can occur

in patients without fractures, resulting in low specificity and the method being described as not reliable.⁵ Thus, tuning-fork tests should not be used in clinical practice if the clinician is attempting to rule in a fracture.⁴

Due to the publication policies of the *British Medical Journal*, we were able to investigate the reviewers' comments, which suggested that some of the reported likelihood ratios were inaccurately calculated. As such, the authors of the systematic review recalculated the data and reported that the positive likelihood ratio ranged from 1.1 to 16.5 and the negative likelihood ratio ranged from 0.09 to 0.49. These ratios give the clinician insight into how the probability shifts for suspected fractures.⁶ Positive likelihood ratios higher than 10 are a strong shift in the probability of a fracture, whereas a negative likelihood ratio less than 0.1 is a strong shift in the probable absence of a fracture.⁶ The wide range of likelihood ratios in this systematic review allows clinicians to use this information in ruling in or ruling out fractures. However, the results from this systematic review, when considering the positive and negative likelihood values, did not offer any conclusive clinical implications. The 2 strongest values for each likelihood ratio came from the same study that focused specifically on possible fractures at the distal fibular shaft.⁷ Therefore, these data suggest that clinicians should use the distal fibular shaft location with a 128-Hz tuning fork test to identify suspected fractures with ankle-inversion injuries.

Mugunthan et al⁴ explained the need for standardized guidelines and training in the use of tuning-fork tests. Although 2 methods have been described (pain induction and reduction of sound conduction), fractures are classified in several ways, including (but not limited to) greenstick fractures, stress fractures, avulsion fractures, and comminuted fractures.⁸ Each fracture classification has a different clinical presentation and different physiologic characteristics.⁸ Tuning-fork tests are effective only when the bony shell is fully disrupted or there is no excessive tissue surrounding the bone, which might have resulted in the demonstrated variations in accuracy.

Suspected fracture types analyzed in the Mugunthan et al⁴ systematic review included femoral neck fracture, any

fracture,² tibial stress fracture,¹ ankle-inversion injuries,⁷ and stress fractures in the legs and feet. The review provided no definitive conclusion as to whether the tuning-fork tests detected a fracture in 1 area more than another. The 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, which is potentially due to the bone shell being partially intact.¹

Further research is needed to determine if tuning-fork tests are only effective in certain physiologic adaptations, such as cortical bone discontinuity, or are affected by the timing of the injury. In addition, investigators should explore larger sample sizes with various fracture types and locations of fractures using higher-quality tuning-fork testing protocols, including blinding of testers to the reference standard.

In conclusion, tuning-fork tests have diagnostic accuracy in ruling out some fractures, but the current evidence is insufficient to identify which test locations are most accurate. Though the authors of the systematic review did not specifically describe best practices, the evidence suggests using a 128-Hz tuning fork to induce pain as a way of identifying a suspected fracture. This method requires less training than the technique that relies on reduction of sound conduction. The athletic training profession should seek to identify if a certain technique is more accurate for specific locations and fracture types and use tuning forks not only in clinical practice but also in the education of students. Although diagnostic imaging is superior, tuning-fork tests are clinically useful for identifying suspected fractures, particularly in rural or remote settings, where access to imaging technologies may be minimal. Regardless of the anatomic location, tuning-fork tests are helpful to the clinician in determining if a patient with a suspected fracture should be sent for referral and in making return-to-participation decisions.

REFERENCES

1. Lesho EP. Can tuning forks replace bone scans for identification of tibial stress fractures? *Mil Med*. 1997;162(12):802–803.
2. Moore MB. The use of a tuning fork and stethoscope to identify fractures. *J Athl Train*. 2009;44(3):272–274.
3. Weisburg JL. Evaluation of bone continuity by sound conduction. *Bull US Army Med Dept*. 1945;4:471–474.
4. Mugunthan K, Doust J, Kurz B, Glasziou P. Is there sufficient evidence for tuning fork tests in diagnosing fractures? A systematic review. *BMJ Open*. 2014;4(8):e005238.
5. 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.
6. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet*. 2005;365(9469):1500–1505.
7. 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.
8. Müller ME, Koch P, Nazarian S, Schatzker J. *The Comprehensive Classification of Fractures of Long Bones*. New York, NY: Springer Science and Business Media; 1990.

Address correspondence to Zachary K. Winkelmann, MS, LAT, ATC, Department of Applied Medicine and Rehabilitation, Indiana State University, 567 North 5th Street, Terre Haute, IN 47809. Address e-mail to zwinkelmann@sycamores.indstate.edu.