

# BMJ Open Is there sufficient evidence for tuning fork tests in diagnosing fractures? A systematic review

Kayalvili Mugunthan,<sup>1</sup> Jenny Doust,<sup>1</sup> Bodo Kurz,<sup>2</sup> Paul Glasziou<sup>1</sup>

**To cite:** 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**:e005238. doi:10.1136/bmjopen-2014-005238

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2014-005238>).

Received 14 March 2014  
Accepted 11 July 2014

## ABSTRACT

**Objective:** To determine the diagnostic accuracy of tuning fork tests for detecting fractures.

**Design:** Systematic review of primary studies evaluating the diagnostic accuracy of tuning fork tests for the presence of fracture.

**Data source:** We searched MEDLINE, CINAHL, AMED, EMBASE, Sports Discus, CAB Abstracts and Web of Science from commencement to November 2012. We manually searched the reference lists of any review papers and any identified relevant studies.

**Study selection and data extraction:** Two reviewers independently reviewed the list of potentially eligible studies and rated the studies for quality using the QUADAS-2 tool. Data were extracted to form 2×2 contingency tables. The primary outcome measure was the accuracy of the test as measured by its sensitivity and specificity with 95% CIs.

**Data synthesis:** We included six studies (329 patients), with two types of tuning fork tests (pain induction and loss of sound transmission). The studies included patients with an age range 7–60 years. The prevalence of fracture ranged from 10% to 80%. The sensitivity of the tuning fork tests was high, ranging from 75% to 100%. The specificity of the tests was highly heterogeneous, ranging from 18% to 95%.

**Conclusions:** Based on the studies in this review, tuning fork tests have some value in ruling out fractures, but are not sufficiently reliable or accurate for widespread clinical use. The small sample size of the studies and the observed heterogeneity make generalisable conclusion difficult.

## Strength and limitations of this study

- Based on the studies in this review, tuning fork tests have value in ruling out some fractures, but current evidence is insufficient to state the circumstances when it is reliable.
- Quantification of the degree and causes of heterogeneity of the studies was not feasible, because of small sample size and varying methods of the studies.
- Therefore, this review does not support the current clinical use of tuning forks as a triage test for the diagnosis of fractures.

periosteum is heavily innervated, mechanical vibration over a fracture site stimulates the overlying periosteum, causing pain.<sup>2</sup> The pain stops or decreases with the removal of the tuning fork. The second method uses a vibrating tuning fork placed over a bony prominence distal to the fracture site. Using a stethoscope to listen to the sound over a bony prominence proximal to the fracture site, the fracture is detected by a reduction in the sound conducted along the bone compared to the unaffected limb.<sup>1</sup>

The aim of this review was to identify the techniques used to diagnose fractures using a tuning fork and assess all studies of the diagnostic accuracy of tuning fork tests for the presence of fracture.

## INTRODUCTION

Although imaging for suspected fractures is generally cheap and readily accessible, there are situations such as remote settings, where imaging is not readily available. Other clinical tests for fracture may then assist in decision making. One test which was proposed at least 60 years ago is the use of a tuning fork.<sup>1</sup>

Two methods of using tuning forks to detect fracture(s) have been developed. The first method uses a vibrating tuning fork placed directly over, or closely proximal to the suspected fracture site. Because the

## METHODS

The inclusion criteria for the review were primary studies that assessed the diagnostic accuracy of tuning forks, using either pain or reduction of sound as the index test, measured against a recognised reference standard, such as X-ray, MRI or bone scan for the diagnosis of fractures. We included studies that enrolled patients of all ages and in all clinical settings with no exclusion by the language of publication. We excluded case series, case-control studies and narrative review papers.



CrossMark

<sup>1</sup>Centre for Research in Evidence Based Practice, Faculty of Health Sciences & Medicine, Bond University, Robina, Australia

<sup>2</sup>Anatomisches Institut, CAU zu Kiel, Olshausenstrasse, Kiel, Germany

### Correspondence to

Dr Kayalvili Mugunthan; [kmugunth@bond.edu.au](mailto:kmugunth@bond.edu.au)

**Box 1** Ovid MEDLINE (<1948 to November Week 3 2012>)

## Search Strategy

1. tuning fork\*.tw. (302)
2. barford test\*.tw. (1)
3. tf test\*.tw. (79)
4. auscultation\*.tw. (2953)
5. or/1–4 (3334)
6. exp Fractures, Bone/(133424 )
7. fracture\*.tw. (149937)
8. or/6–7 (187939)
9. 5 and 8 (20)

### Search strategy

We searched MEDLINE, CINAHL, AMED, EMBASE, Sports Discus, CAB Abstracts and Web of Science from commencement to November 2012. We also searched the reference lists of any identified studies or review papers. We also searched for any systematic reviews or meta-analyses carried out on this diagnostic test.

The Medline search strategy is shown in **box 1**, and was run without a methodological filter.

### Data extraction and management

We selected studies in a two-stage process. The titles and abstracts of all search results were screened by two

authors (KM and JD) and full manuscripts for all potential relevant papers were obtained. Two review authors (KM and JD) independently reviewed each paper for inclusion according to the predefined inclusion criteria, rated the study quality and then extracted relevant data. In the case of duplicate publication, we selected the most complete version of the study. We resolved disagreements through discussion with the third author (PG).

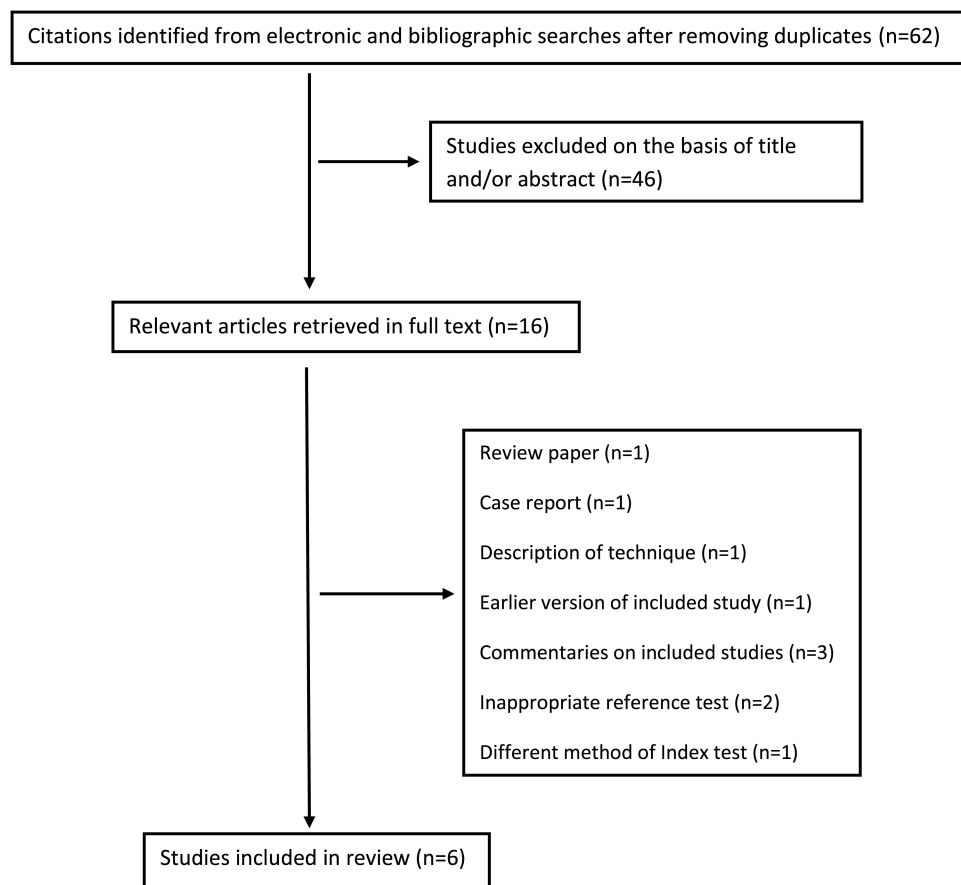
The primary outcome measure of interest was the accuracy of the test as measured by its sensitivity and specificity. Wherever possible, we used the raw data to construct 2×2 tables. 95% CIs for sensitivity and specificity were calculated with the Wilson score method and 95% CIs for positive and negative likelihood ratios were calculated with the method described by Simel *et al.*<sup>3 4</sup> We appraised each article using the QUADAS-2 tool.<sup>5</sup>

## RESULTS

### Literature identification and study quality

We identified 62 citations from the electronic and bibliographic searches. Sixteen articles in full text were obtained for further scrutiny. Six primary studies (329 patients) were included in the final review (**figure 1**).

The characteristics of the participants and the methods of testing are shown in **table 1**. Most studies included only adults; one study included paediatric



**Figure 1** Flow chart of studies included in the review.

**Table 1** Characteristics of the included studies

Index test	Sound conduction			Pain from vibration		
	Bache and Cross <sup>7</sup>	Moore <sup>6</sup>	Lesho <sup>9</sup>	Kazemi and Roscoe <sup>2</sup>	Dissmann and Han <sup>8</sup>	Wilder <i>et al</i> <sup>10</sup>
Number of participants	100	37	52	46	49	45
Age (years)	Mean 79	Range 7–60	Mean 25	Mean 30	Range 12–84	Mean 31
Setting	Emergency department	University sports clinic/orthopaedic centre	Army medical centre	Emergency department	Emergency department	Runners clinic
Suspected fracture type	Femoral neck fracture	Any fracture	Tibial stress fracture	Any fracture	Ankle inversion injuries*	Stress fractures in legs and feet
Reference test	X-ray	X-ray	Bone scan	Bone scan	X-ray	X-ray and bone scan
Time since symptom onset	Not reported	<7 days old	Not reported	0–10 days	Not reported	Not reported

\*Patients had tested positive to the 'Ottawa ankle rule'.

patients. The prevalence of fracture ranged from 10% to 80%. Two studies used the tuning fork test to investigate any suspected fracture,<sup>2 6</sup> one suspected femoral neck fracture,<sup>7</sup> one ankle inversion injury<sup>8</sup> and two stress fractures.<sup>9 10</sup> The studies investigating any fracture, femoral or ankle fractures used X-ray as a reference standard and the studies of stress fractures used either bone scan or X-ray and bone scan as a reference standard. The study of patients with ankle inversion injuries included patients who had tested positive to the 'Ottawa ankle rule'.

Four studies detected fractures using pain induced by the vibrating tuning fork,<sup>2 8–10</sup> while two studies used reduced sound conduction.<sup>7 6</sup> Four studies used a 128 Hz tuning fork alone,<sup>6–9</sup> but two studies compared the diagnostic accuracy of different frequency tuning forks within the studies.<sup>2 10</sup>

The methodological quality of the included studies was modest, with important elements that may indicate a

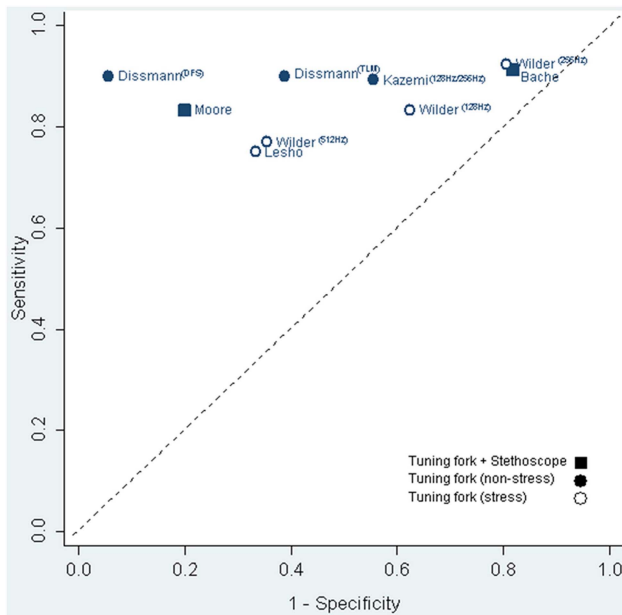
risk of bias being unclear or not reported. For example, in most studies it was either unclear or not stated whether the comparison between the tuning fork test and the reference test had been blind and independent of the reference standard (table 2).

Figure 2 shows sensitivity versus 1-specificity (receiver operating characteristic plot) for the six included studies. The sensitivity of the tuning fork tests was generally high, ranging from 75% to 100%. In the study to rule out fracture in patients who had tested positive to the 'Ottawa ankle rule', the use of the tuning fork on either the tip of the lateral malleolus or the distal fibula shaft gave a sensitivity of 100%, albeit there were only five patients with fractures.<sup>8</sup> However, the specificity of the test in the six studies was highly heterogeneous, ranging from 18% to 95%.

Two studies showed reasonable overall diagnostic accuracy with diagnostic ORs >10, but other studies showed only modest values (table 3). The two studies

**Table 2** Methodological quality of the included studies

Criterion	Bache and Cross <sup>7</sup>	Moore <sup>6</sup>	Lesho <sup>9</sup>	Kazemi and Roscoe <sup>2</sup>	Dissmann and Han <sup>8</sup>	Wilder <i>et al</i> <sup>10</sup>
Consecutive or random sample	Yes	Yes	Yes	Yes	Yes	Yes
Case-control study design avoided	Yes	Yes	Yes	Yes	Yes	Yes
Inappropriate exclusions avoided	Yes	Yes	Yes	Yes	Yes	Yes
Index test interpreted blind and independent of reference standard/ prespecified threshold	Unclear	Yes	Yes	Yes	Yes	Unclear
Appropriate reference standard	Yes	Yes	Yes	Yes	Yes	Yes
Reference standard interpreted blind and independent of index test	Unclear	Unclear	Unclear	Yes	Yes	Unclear
Appropriate interval between index test and reference standard	Not reported	Not reported	Within 30 days	Not reported	Not reported	Not reported
All patients received a reference standard/same reference standard	Yes	Yes	No	Yes	Yes	No
All patients included in the analysis	Yes	Yes	No	Yes	Yes	No



**Figure 2** Sensitivity versus 1-specificity (receiver operating characteristic) plot of included studies.

that compared the diagnostic accuracy of different frequency tuning forks on the same patients found no differences between frequencies.<sup>2, 10</sup> One study assessed the differences between pain ratings but differences were small. The study that assessed inter-tester reliability showed only low reliability.<sup>9</sup>

**DISCUSSION**

Two forms of tuning fork test, one based on pain induction and the other on sound transmission, showed modest diagnostic accuracy with some ability to rule out fractures. However, the estimated sensitivity (ranging from 75% to 100%) is not sufficient to be relied on to rule out fractures based on a negative test. The specificity is particularly heterogeneous, potentially resulting in a high proportion of false-positive test results. The reasons for this variation in accuracy are unclear, but may be related to both the way the test is performed or to characteristics of the injuries and fractures.

The low inter-tester reliability suggests that the techniques would benefit from standardisation and training. Wilder *et al*<sup>10</sup> compared different frequencies and found a higher induction of fracture pain using 256 Hz, but pain also occurred in patients without fractures resulting in a low specificity.

Based on the results in this review, the tuning fork test was less accurate for stress fractures than other types of fractures, but a number of features of this type of injury may modify the accuracy. Lesho<sup>9</sup> suggests that in the early stages, stress fractures might not be identified by the tuning fork test, because the bone shell is still more or less intact. A bone scan, however, would show an

**Table 3** Overview of the results of the included studies

Results of testing Type of tuning fork	Bache and Cross <sup>7</sup>		Moore <sup>6</sup>		Lesho <sup>9</sup>		Kazemi and Roscoe <sup>2</sup>		Dissmann and Han <sup>8</sup>		Wilder <i>et al</i> <sup>10</sup>		
	128 Hz	256 Hz	128 Hz	256 Hz	128 Hz	256 Hz	128 Hz	256 Hz	128 Hz	DFS	128 Hz	256 Hz	512 Hz
Prevalence of fractures	56%		32%	80%	80%	80%	80%	80%	10%	10%	27%	27%	27%
Sensitivity (%; 95% CI)	91 (81 to 96)		83 (55 to 95)	89 (75 to 96)	89 (75 to 96)	89 (75 to 96)	89 (75 to 96)	89 (75 to 96)	92 (52 to 99)	92 (52 to 99)	83 (55 to 95)	92 (67 to 99)	77 (49 to 92)
Specificity (%; 95% CI)	18 (9 to 32)		80 (61 to 91)	44 (19 to 73)	44 (19 to 73)	44 (19 to 73)	44 (19 to 73)	44 (19 to 73)	61 (46 to 74)	94 (84 to 98)	37 (23 to 55)	19 (9 to 36)	64 (47 to 79)
Diagnostic OR	2.3 (0.7 to 7.5)		20.0 (3.3 to 122)	6.6 (1.2 to 35.2)	6.6 (1.2 to 35.2)	6.6 (1.2 to 35.2)	6.6 (1.2 to 35.2)	6.6 (1.2 to 35.2)	17.3 (0.9 to 332)	187.0 (7.9 to 4424)	3.0 (0.6 to 16.1)	2.9 (0.3 to 26.7)	6.1 (1.4 to 26.7)
Positive likelihood ratio (95% CI)	1.1 (0.94 to 1.3)		4.2 (1.8 to 9.5)	1.6 (0.89 to 2.9)	1.6 (0.89 to 2.9)	1.6 (0.89 to 2.9)	1.6 (0.89 to 2.9)	1.6 (0.89 to 2.9)	2.4 (1.5 to 3.7)	16.5 (4.8 to 56)	1.3 (0.92 to 1.9)	1.1 (0.91 to 1.4)	2.2 (1.2 to 3.8)
Negative likelihood ratio (95% CI)	0.49 (0.17 to 1.4)		0.21 (0.06 to 0.75)	0.24 (0.08 to 0.79)	0.24 (0.08 to 0.79)	0.24 (0.08 to 0.79)	0.24 (0.08 to 0.79)	0.24 (0.08 to 0.79)	0.14 (0.01 to 2.0)	0.09 (0.01 to 1.3)	0.45 (0.12 to 1.7)	0.39 (0.05 to 3.0)	0.36 (0.13 to 0.99)

TLM, tip of lateral malleolus; DFS, distal fibula shaft.

increased activity in the fractured area. Timing may also affect the accuracy of the test.

A mineralised callus where fracture healing has been initiated might not be identified by these tests. It is unclear whether a discontinuity of the cortical bone is required in order to give a positive test result. Both types of tuning fork tests seem to be more accurate in diagnosing transverse fractures than other types of fractures. It is also unclear whether swelling or bruising in the area of the injury might affect the results.

A systematic review,<sup>11</sup> which examined a variety of methods for the diagnosis of stress fractures, included only two of the six studies we used in this review.

In conclusion, both tuning fork methods have some discrimination ability, but current techniques are not sufficiently reliable or accurate to rule in or out fractures and currently should have only limited use in clinical practice. The small sample size of the studies and the observed heterogeneity make generalisable conclusion difficult. However, the clinical usefulness of these tests might be in remote areas or athletic fields with no easy access to other options.

**Acknowledgements** The authors extend their gratitude to Sarah Thorning (Trial search coordinator, Bond University) for valuable help in literature search, and Elaine Beller (Statistician, Bond University) for statistical support.

**Contributors** KM, JD, BK and PG contributed to the concepts of the work and acquisition, analysis and interpretation of data. KM drafted the work. JD, BK and PG revised the work critically for important intellectual content. All authors approved the final version.

**Funding** Kayalvili Mugunthan was supported by a Primary Health Care Research Evaluation & Development (PHCRED) fellowship, Bond University.

**Competing interests** None.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

1. Weisberg JL. Evaluation of bone continuity by sound conduction. *Bull U S Army Med Dept* 1945;4:471–4.
2. Kazemi M, Roscoe MW. Is the tuning fork test a reliable tool in detecting acute simple fractures? *Int Sports J* 2000;4:1–8.
3. <http://www.stats.org.uk/statistical-inference/Newcombe1998.pdf> (accessed on 5 Jul 2014)
4. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991;44:763–70.
5. Whiting P, Rutjesw AW, Westwood M, *et al.* Updating QUADAS: evidence to inform the development of QUADAS-2. 2010. [www.bris.ac.uk/quadas/resources/quadas2reportv4.pdf](http://www.bris.ac.uk/quadas/resources/quadas2reportv4.pdf) (accessed 5 Jul 2014).
6. Moore MB. The use of a tuning fork and stethoscope to identify fractures. *J Athl Train* 2009;44:272–4.
7. Bache JB, Cross AB. The Barford test. A useful diagnostic sign in fractures of the femoral neck. *Practitioner* 1984;228:305–8.
8. 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;23:788–90.
9. Lesho EP. Can tuning forks replace bone scans for identification of tibial stress fractures? *Mil Med* 1997;162:802–3.
10. Wilder RP, Vincent HK, Stewart J, *et al.* Clinical use of tuning forks to identify running-related stress fractures. *Athletic Training Sports Health Care* 2009;1:12–18.
11. 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;42:760–71.