BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005238
Article Type:	Research
Date Submitted by the Author:	11-Mar-2014
Complete List of Authors:	Mugunthan, Kayalvili; Bond University, Center for Research in Evidence Based Practice Doust, Jenny; Bond University, Center for Research in Evidence Based Practice, Faculty of Health Sciences and Medicine Kurz, Bodo; CAU zu Kiel, Anatomisches Institut Glasziou, Paul; Bond University, Center for Research in Evidence Based Practice, Faculty of Health Sciences and Medicine
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Qualitative research
Keywords:	Epidemiology < ONCOLOGY, Orthopaedic sports trauma < ORTHOPAEDIC & TRAUMA SURGERY, QUALITATIVE RESEARCH

Title Page

1. **Title of the article**

Is there sufficient evidence for tuning fork tests in diagnosing fractures?

A systematic review

2. **Corresponding author** Kayalvili Mugunthan

14,University Drive, Robina, QLD 4229 , Australia.

kmugunth@bond.edu.au, + 61755955521

3. **Co- authors**

a) Jenny Doust

Professor of Public Health, Centre for Research in Evidence Based Practice

Faculty of Health Sciences & Medicine, Bond University, Robina, QLD 4229 ,

Australia

b) Prof. Dr. Bodo Kurz

Anatomisches Institut ,

CAU zu Kiel,

Olshausenstrasse 40-60
24098 Kiel

c) Paul Glasziou

Director, Centre for Research in Evidence Based Practice,

Faculty of Health Sciences & Medicine, Bond University, Robina, QLD 4229 ,

Australia

Key words

Tuning fork, bone fractures, diagnostic procedures and techniques, diagnosis, sensitivity specificity

Word count

1318

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 was extracted to form 2x2 contingency tables. The primary outcome measure was the accuracy of the test as measured by its sensitivity and specificity with 95% confidence intervals.

Data synthesis

We included 6 studies (329 patients), with 2 types of tuning fork tests (pain induction, loss of sound transmission). The studies included patients with an age range of 7 to 60. 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%.

Conclusion

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 numbers included in previous studies and the observed heterogeneity make generalizable conclusions difficult.

Strength and limitations of the 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 sizes 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.

Is there sufficient evidence for tuning fork tests in diagnosing fractures?

A systematic review

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.[9]

Two methods of using tuning forks to detect fractures have been developed. The first method uses a vibrating tuning fork placed directly over, or closely proximal to the suspected fracture site. Because the periosteum is heavily innervated, mechanical vibration over a fracture site stimulates the overlying periosteum, causing pain.[4] 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.[9]

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.

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 language of publication. We excluded case series, case-control studies, and narrative review papers.

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 done 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, JD) and full manuscripts for all potential relevant papers were obtained. Two review authors (KM, JD) independently reviewed each paper for inclusion according to the predefined inclusion criteria, then rated study quality and extracted relevant data. In the case of duplicate publication, we selected the most complete version of the study. We resolved disagreements through discussion with a 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 2x2 tables.95% confidence intervals for sensitivity and specificity were calculated with the Wilson score method and 95% confidence intervals for positive and negative likelihood ratios were calculated with the method described by Simel et al (1991).[3,8]We appraised each article using the QUADAS 2 tool.[10]

Results

Literature identification and study quality

We identified 62 citations from the electronic and bibliographic searches. 16 articles in full text were obtained for further scrutiny. Six primary studies (329 patients) were included in the final review (Figure1).

The characteristics of the participants and the methods of testing are shown in Table 1. Most studies included only adults; one study included paediatric patients. The prevalence of fracture ranged from 10% to 80%.Two studies used the tuning fork test to investigate any suspected fracture [4,6] one suspected femoral neck fractures,[1]one ankle inversion injuries [2] and two stress fractures.[5,11] 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 by using pain induced by the vibrating tuning fork, [2,4,5,11] while two studies used reduced sound conduction .[1,6] Four studies used a 128Hz tuning fork alone[1,2,5,6] but two studies compared the diagnostic accuracy of different frequency tuning forks within the studies .[4,11]

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 reference test had been blind and independent of the reference standard (Table 2).

Figure 2 shows sensitivity versus 1- specificity (ROC plot) for the 6 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 5 patients with fractures [9].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 odds ratios > 10,but other studies showed only modest values (Table 3). The two studies that compared the diagnostic accuracy of different frequency tuning forks on the same patients found no differences between frequencies.[4,11] One study assessed the differences between pain ratings but differences were small. The one study that assessed inter-tester reliability showed only low reliability.[5]

Discussion

Two forms of tuning fork test, one based on pain induction and one 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 upon 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 done or to characteristics of the injuries and fractures.

The low inter-tester reliability suggests that the techniques would benefit from standardization and training. Wilder et al [11] compared different frequencies and found a higher induction of

fracture pain using 256Hz, 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[5] 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 increased activity in the fractured area. Timing may also affect the accuracy of the test.

A mineralized 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 also unclear whether swelling or bruising in the area of the injury might affect the results.

A systematic review[7] which examined a variety of methods for the diagnosis of stress fractures included only 2 of the 6 studies we used in this review.

In conclusion, both tuning fork methods have some discrimination ability, but current techniques are not sufficient reliable or accurate to rule in or out fractures and currently should have only limited use in clinical practice. The clinical usefulness of these tests might be in remote areas with no easy access to other options.

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)

Table 1: Characteristics of the included studies

Criterion	Bache[1] (1984)	Moore[6] (2009)	Lesho[5] (1997)	Kazemi[4] (2000)	Dissmann[2] (2006)	Wilder[11] (2008)
Index test	**Sound conduction**		**Pain from vibration**			
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/ orthopedic center	Army medical center	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 leg 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'

Table 2: Methodological quality of the included studies

Criterion	Bache[1] (1984)	Moore[6] (2009)	Lesho[5] (1997)	Kazemi [2] (2000)	Dissmann[4] (2006)	Wilder[11] (2008)
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/ Pre-specified 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

Table 3: Overview of the results of the included studies

Results of testing	Bache[1] (1984)	Moore[6] (2009)	Lesho[5] (1997)	Kazemi[4] (2000)		Dissmann[2] (2006)		Wilder[11] (2008)		
Type of tuning fork	128Hz	128Hz	128Hz	128Hz	256Hz	128Hz TLM	128Hz DFS	128Hz	256Hz	512Hz
Prevalence of fractures	56%	32%	61%	80%	80%	10%	10%	27%	27%	27%
Sensitivity (%, 95% CI)	91 (81-96)	83 (55-95)	75 (57-87)	89 (75-96)	89 (75-96)	92 (52-99)	92 (52-99)	83 (55-95)	92 (67-99)	77 (49-92)
Specificity (%, 95% CI)	18 (9-32)	80 (61–91)	67 (44-84)	44 (19-73)	44 (19-73)	61 (46-74)	94 (84-98)	37 (23-55)	19 (9-36)	64 (47-79)
Diagnostic Odd ratio	2.3 (0.7-7.5)	20.0 (3.3-122)	6.0 (1.6-22)	6.6 (1.2-35.2)	6.6 (1.2-35.2)	17.3 (0.9-332)	187.0 (7.9-4424)	3.0 (0.6-16.1)	2.9 (0.3-26.7)	6.1 (1.4-26.7)
Positive likelihood ratio: (95% CI)	1.1 (0.94-1.3)	4.2 (1.8–9.5)	2.2 (1.1-4.5)	1.6 (0.89-2.9)	1.6 (0.89-2.91)	2.4 (1.5–3.7)	16.5 (4.8-56)	1.3 (0.92-1.9)	1.1 (0.91-1.4)	2.2 (1.2-3.8)
Negative likelihood ratio: (95% CI)	0.49 (0.17-1.4)	0.21 (0.06–0.75)	0.37 (0.18-0.77)	0.24 (0.08-0.79)	0.24 (0.08-0.79)	0.14 (0.01-2.0)	0.09 (0.01-1.3)	0.45 (0.12-1.7)	0.39 (0.05-3.0)	0.36 (0.13-0.99)

TLM- tip of lateral malleolus ; DFS- distal fibula shaft

Acknowledgements

We extend our gratitude to SarahThorning (Trial search coordinator, Bond University) for valuable help in literature search and Elaine Beller (Statistician, Bond University) for statistical support.

Competing interest

The authors have declared no competing interest

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

Contributorship Statement:

1)substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data;
Kayalvili Mugunthan
Jenny Doust
Paul Glaziou

2)drafting the article or revising it critically for important intellectual content;
Kayalvili Mugunthan
Jenny Doust
Bodo Kurz

3)final approval of the version to be published

Jenny Doust
Paul Glaziou

D=1)substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data;
Kayalvili Mugunthan
Jenny Doust
Paul Glaziou

2)drafting the article or revising it critically for important intellectual content;
Kayalvili Mugunthan

Jenny Doust
Bodo Kurz

3)final approval of the version to be published

Jenny Doust
Paul Glaziou

Data Sharing Statement

N/A

References;

1.Bache JB, Cross AB. The Barford test. A useful diagnostic sign in fractures of the femoral neck. *Practitioner*.1984 Mar;**228**(1389):305-8.

2. 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):788–90. http://dx.doi.org /10.1136/emj.2006.035519

3..http://www.stats.org.uk/statistical-inference/Newcombe1998.pdf.

4. Kazemi M, Roscoe MW. Is the tuning fork test a reliable tool in detecting acute simple fractures? *International Sports Journal*. 2000Summer;**4**(2):1-8.

5. Lesho EP. Can tuning forks replace bone scans for identification of tibial stress fractures? *Mil Med*.1997 Dec;**162**(12):802-3.

6. Moore MB. The use of a tuning fork and stethoscope to identify fractures. *Journal of Athletic Training*. 2009;**44**(3):272-4. http://dx.doi.org /10.4085/1062-6050-44.3.272.

7. 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*. 2012Sep;42(9):760-71. http://dx.doi.org / 10.2519/jospt.2012.4000

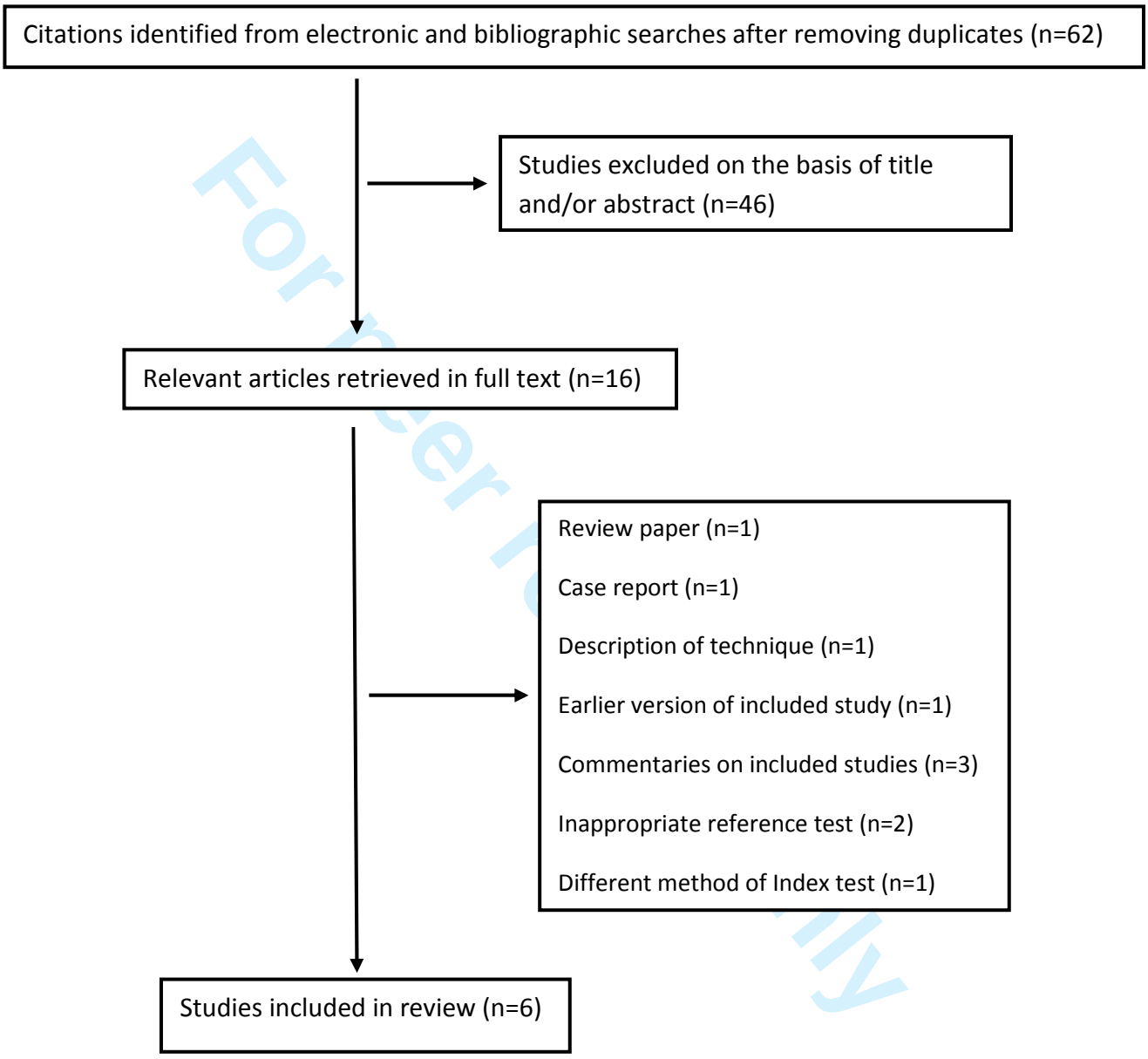
8, Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *Journal of Clinical Epidemiology*.1991;**44:**763-70.

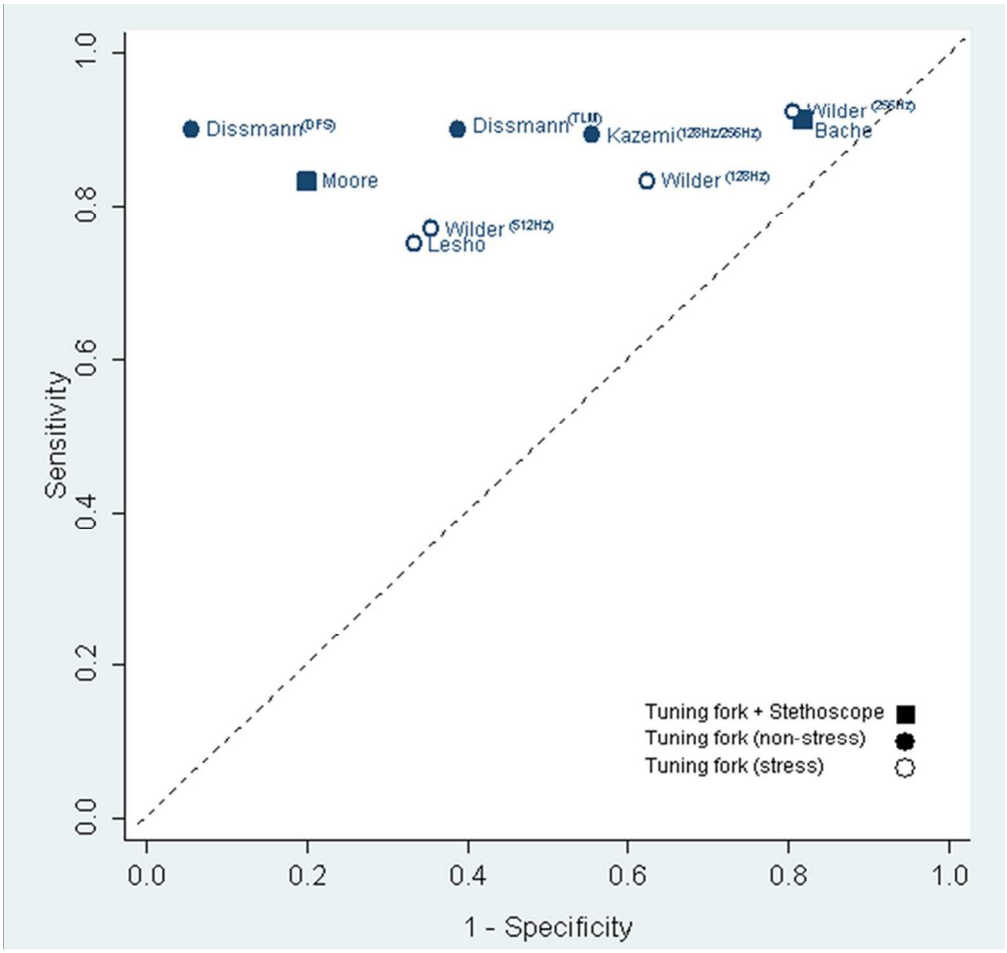
9. Weisberg JL Evaluation of bone continuity by sound conduction. *Bulletin of the United States Army Medical Department*.1945;**4:**471-4

10. Whiting P, Rutjesw AW, Westwood M, et al.Updating QUADAS: evidence to inform the development of QUADAS-2. 2010.Accessed at www.bris.ac.uk/quadas/resources/quadas2reportv4.pdf on 5 September 2011.

11.Wilder RP, Vincent HK, Stewart J et al. Clinical Use of Tuning Forks to Identify Running-Related Stress Fractures. Athletic Training & Sports Health Care: *The Journal for the Practicing Clinician*. 2009;**1**(1):12-8. http://dx.doi.org /10.3928/19425864-20090101-10

Figure 1: Flowchart of studies included in the review





155x146mm (300 x 300 DPI)

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	nil
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	3
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	3-4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	3-4
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	4, QUADAS-2
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9, QUADAS-2
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	20

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	4, 5 and Fig2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9 , TABLE 2
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	TABLE 3,Fig2
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	4, Fig 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	4-5, TABLE 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	9, TABLE 2
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	TABLE 3, Fig 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10, TABLE 3, FIG 2
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	9, TABLE 2
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Fig 2
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	5-6
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	5-6
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	6
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	**Primary Health Care Research Evaluation & Development**

PRISMA 2009 Checklist

			(PHCRED)

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	nil
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	3
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	3-4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	3-4
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	4, QUADAS-2
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9, QUADAS-2
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	20
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	4, 5 and Fig2

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9 , TABLE 2
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	TABLE 3,Fig2
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	4, Fig 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	4-5, TABLE 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	9, TABLE 2
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	TABLE 3, Fig 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10, TABLE 3, FIG 2
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	9, TABLE 2
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Fig 2
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	5-6
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	5-6
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	6
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Primary Health Care Research Evaluation & Development (PHCRED)

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005238.R1
Article Type:	Research
Date Submitted by the Author:	11-Jul-2014
Complete List of Authors:	Mugunthan, Kayalvili; Center for Research in Evidence Based Practice,Faculty of Health Sciences and Medicine, Bond University Doust, Jenny; Center for Research in Evidence Based Practice, Faculty of Health Sciences and Medicine, Bond UniversityCenter for Kurz, Bodo; CAU zu Kiel, Anatomisches Institut Glasziou, Paul; Center for Research in Evidence Based Practice, Faculty of Health Sciences and Medicine, Bond University
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Qualitative research
Keywords:	Epidemiology < ONCOLOGY, Orthopaedic sports trauma < ORTHOPAEDIC & TRAUMA SURGERY, QUALITATIVE RESEARCH

Title Page

1. **Title of the article**

Is there sufficient evidence for tuning fork tests in diagnosing fractures?

A systematic review

2. **Corresponding author** Kayalvili Mugunthan

14,University Drive, Robina, QLD 4229 , Australia.

kmugunth@bond.edu.au, + 61755955521

3. Co- authors

a)Jenny Doust

Professor of Public Health, Centre for Research in Evidence Based Practice

Faculty of Health Sciences & Medicine, Bond University, Robina, QLD 4229 ,

Australia

b)Prof. Dr. BodoKurz

AnatomischesInstitut ,

CAU zu Kiel,

Olshausenstrasse 40-60
24098 Kiel

c) Paul Glasziou

Director,Centre for Research in Evidence Based Practice,

Faculty of Health Sciences & Medicine, Bond University, Robina, QLD 4229 ,

Australia

Key words

Tuning fork, bone fractures, diagnostic procedures and techniques, diagnosis, sensitivity specificity

Word count

1322

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 was extracted to form 2x2 contingency tables. The primary outcome measure was the accuracy of the test as measured by its sensitivity and specificity with 95% confidence intervals.

Data synthesis

We included 6 studies (329 patients), with 2 types of tuning fork tests (pain induction, loss of sound transmission). The studies included patients with an age range of 7 to 60. 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%.

Conclusion

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 sizes of the studies and the observed heterogeneity make generalizable conclusion difficult.

Strength and limitations of the 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 sizes 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.

Is there sufficient evidence fortuning fork test in diagnosing fractures?

A systematic review

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.[1]

Two methods of using tuning forks to detect fractures have been developed. The first method uses a vibrating tuning fork placed directly over, or closely proximal to the suspected fracture site. Because the periosteum is heavily innervated, mechanical vibration over a fracture site stimulates the overlying periosteum, causing pain.[2] 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.[1]

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.

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 language of publication. We excluded case series, case-control studies, and narrative review papers.

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 done 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, JD) and full manuscripts for all potential relevant papers were obtained. Two review authors (KM, JD) independently reviewed each paper for inclusion according to the predefined inclusion criteria, then rated study quality and extracted relevant data. In the case of duplicate publication, we selected the most complete version of the study. We resolved disagreements through discussion with a 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 2x2 tables.95% confidence intervals for sensitivity and specificity were calculated with the Wilson score method and 95% confidence intervals for positive and negative likelihood ratios were calculated with the method described by Simel et al (1991).[3,4]We appraised each article using the QUADAS 2 tool.[5]

Results

Literature identification and study quality

We identified 62citations from the electronic and bibliographic searches. 16 articles in full text were obtained for further scrutiny. Six primary studies (329 patients) were included in the final review (Figure1).

The characteristics of the participants and the methods of testing are shown in Table 1. Most studies included only adults; one study included paediatric patients. The prevalence of fracture ranged from 10% to 80%.Two studies used the tuning fork test to investigate any suspected fracture ,[2,6] one suspected femoral neck fractures,[7]one ankle inversion injuries,[8] and two stress fractures.[9,10] 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 by using pain induced by the vibrating tuning fork,[2,8,9,10]while two studies used reduced sound conduction .[7,6] Four studies used a 128Hz tuning fork alone,[6,7,8,9] but two studies compared the diagnostic accuracy of different frequency tuning forks within the studies.[2,10]

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 reference test had been blind and independent of the reference standard (Table 2).

Figure 2 shows sensitivity versus 1- specificity (ROC plot) for the 6 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 5 patients with fractures,[1].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 odds ratios > 10,but other studies showed only modest values (Table 3). The two studies that compared the diagnostic accuracy of different frequency tuning forks on the same patients found no differences between frequencies.[2,10]One study assessed the differences between pain ratings but differences were small. The one study that assessed inter-tester reliability showed only low reliability.[9]

Discussion

Two forms of tuning fork test, one based on pain induction and one 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 upon 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 done or to characteristics of the injuries and fractures.

The low inter-tester reliability suggests that the techniques would benefit from standardization and training. Wilder et al,[10]compared different frequencies and found a higher induction of

fracture pain using 256Hz, 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[9] 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 increased activity in the fractured area. Timing may also affect the accuracy of the test.

A mineralized 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 also unclear whether swelling or bruising in the area of the injury might affect the results.

A systematic review,[11]which examined a variety of methods for the diagnosis of stress fractures included only 2 of the 6 studies we used in this review.

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

Acknowledgements

We extend our gratitude to SarahThorning(Trial search coordinator, Bond University) for valuable help in literature search and Elaine Beller(Statistician, Bond University) for statistical support.

Contributorship Statement

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

Competing interest

The authors have declared no competing interest

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

Data Sharing No additional data available

References:

1 Weisberg JL. Evaluation of bone continuity by sound conduction. *Bulletin of the United States Army Medical Department*.1945;4**:**471-4

2 Kazemi M, Roscoe MW. Is the tuning fork test a reliable tool in detecting acute simple fractures?*Internaional Sports Journal*.2000 Summer;4(2):1-8.

3 http://www.stats.org.uk/statistical-inference/Newcombe1998.pdf Accessed on 5th July 2014

4 Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *Journal of Clinical Epidemiology*.1991;44**:**763-70.

5 Whiting P, Rutjesw AW, Westwood M, et al**.**Updating QUADAS: evidence to inform the development of QUADAS-2. 2010.Accessed at www.bris.ac.uk/quadas/resources/quadas2reportv4.pdf on 5 July 2014.

6 Moore MB. The use of a tuning fork and stethoscope to identify fractures.*Journal of Athletic Training*. 2009;44(3):272-4.

7 Bache JB, Cross AB. The Barford test. A useful diagnostic sign in fractures of the femoral neck. *Practitioner*.1984;228(1389):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(10):788– 90.

9 Lesho EP. Can tuning forks replace bone scans for identification of tibial stress fractures? *Mil Med*.1997;162(12):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: *The Journal for the Practicing Clinician*. 2009;1(1):12-8.

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(9):760-71.

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)

Table 1: Characteristics of the included studies

	Bache[7] (1984)	Moore[6] (2009)	Lesho[9] (1997)	Kazemi[2] (2000)	Dissmann[8] (2006)	Wilder[10] (2008)
Index test	**Sound conduction**		**Pain from vibration**			
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/ orthopedic center	Army medical center	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 leg 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'

Table 2: Methodological quality of the included studies

Criterion	Bache[7] (1984)	Moore[6] (2009)	Lesho[9] (1997)	Kazemi [8] (2000)	Dissmann[2] (2006)	Wilder[10] (2008)
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/ Pre-specified 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

Table 3: Overview of the results of the included studies

Results of testing	Bache[7] (1984)	Moore[6] (2009)	Lesho[9] (1997)	Kazemi[2] (2000)		Dissmann[8] (2006)		Wilder[10] (2008)		
Type of tuning fork	128Hz	128Hz	128Hz	128Hz	256Hz	128Hz TLM	128Hz DFS	128Hz	256Hz	512Hz
Prevalence of fractures	56%	32%	61%	80%	80%	10%	10%	27%	27%	27%
Sensitivity (%, 95% CI)	91 (81-96)	83 (55-95)	75 (57-87)	89 (75-96)	89 (75-96)	92 (52-99)	92 (52-99)	83 (55-95)	92 (67-99)	77 (49-92)
Specificity (%, 95% CI)	18 (9-32)	80 (61–91)	67 (44-84)	44 (19-73)	44 (19-73)	61 (46-74)	94 (84-98)	37 (23-55)	19 (9-36)	64 (47-79)
Diagnostic Odd ratio	2.3 (0.7-7.5)	20.0 (3.3-122)	6.0 (1.6-22)	6.6 (1.2-35.2)	6.6 (1.2-35.2)	17.3 (0.9-332)	187.0 (7.9-4424)	3.0 (0.6-16.1)	2.9 (0.3-26.7)	6.1 (1.4-26.7)
Positivelike lihood ratio: (95% CI)	1.1 (0.94-1.3)	4.2 (1.8–9.5)	2.2 (1.1-4.5)	1.6 (0.89-2.9)	1.6 (0.89-2.91)	2.4 (1.5–3.7)	16.5 (4.8-56)	1.3 (0.92-1.9)	1.1 (0.91-1.4)	2.2 (1.2-3.8)
Negativelik elihood ratio: (95% CI)	0.49 (0.17-1.4)	0.21 (0.06–0.75)	0.37 (0.18-0.77)	0.24 (0.08-0.79)	0.24 (0.08-0.79)	0.14 (0.01-2.0)	0.09 (0.01-1.3)	0.45 (0.12-1.7)	0.39 (0.05-3.0)	0.36 (0.13-0.99)

TLM- tip of lateral malleolus ; DFS- distal fibula shaft

Figures and Tables

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

Figure 2: Sensitivity versus 1-specificity (ROC) plot of included studies.

Table 1: Characteristics of the included studies

Table 2: Methodological quality of the included studies

Table 3: Overview of the results of the included studies

Title Page

1. **Title of the article**

Is there sufficient evidence for tuning fork tests in diagnosing fractures?

A systematic review

2. **Corresponding author** KayalviliMugunthan

14,University Drive, Robina, QLD 4229 , Australia.

kmugunth@bond.edu.au, + 61755955521

3. Co- authors

a)Jenny Doust

Professor of Public Health, Centre for Research in Evidence Based Practice

Faculty of Health Sciences & Medicine, Bond University, Robina, QLD 4229 ,

Australia

b)Prof. Dr. BodoKurz

AnatomischesInstitut ,

CAU zu Kiel,

Olshausenstrasse 40-60
24098 Kiel

c) Paul Glasziou

Director,Centre for Research in Evidence Based Practice,

Faculty of Health Sciences & Medicine, Bond University, Robina, QLD 4229 ,

Australia

Key words

Tuning fork, bone fractures, diagnostic procedures and techniques, diagnosis, sensitivity specificity

Word count

1322

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 was extracted to form 2x2 contingency tables. The primary outcome measure was the accuracy of the test as measured by its sensitivity and specificity with 95% confidence intervals.

Data synthesis

We included 6 studies (329 patients), with 2 types of tuning fork tests (pain induction, loss of sound transmission). The studies included patients with an age range of 7 to 60. 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%.

Conclusion

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 sizes of the studies and the observed heterogeneity make generalizable conclusion difficult.

Strength and limitations of the 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 sizes 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.

Is there sufficient evidence fortuning fork test in diagnosing fractures?

A systematic review

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.[1]

Two methods of using tuning forks to detect fractures have been developed. The first method uses a vibrating tuning fork placed directly over, or closely proximal to the suspected fracture site. Because the periosteum is heavily innervated, mechanical vibration over a fracture site stimulates the overlying periosteum, causing pain.[2] 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.[1]

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.

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 language of publication. We excluded case series, case-control studies, and narrative review papers.

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 done 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, JD) and full manuscripts for all potential relevant papers were obtained. Two review authors (KM, JD) independently reviewed each paper for inclusion according to the predefined inclusion criteria, then rated study quality and extracted relevant data. In the case of duplicate publication, we selected the most complete version of the study. We resolved disagreements through discussion with a 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 2x2 tables.95% confidence intervals for sensitivity and specificity were calculated with the Wilson score method and 95% confidence intervals for positive and negative likelihood ratios were calculated with the method described by Simel et al (1991).[3,4]We appraised each article using the QUADAS 2 tool.[5]

Results

Literature identification and study quality

We identified 62citations from the electronic and bibliographic searches. 16 articles in full text were obtained for further scrutiny. Six primary studies (329 patients) were included in the final review (Figure1).

The characteristics of the participants and the methods of testing are shown in Table 1. Most studies included only adults; one study included paediatric patients. The prevalence of fracture ranged from 10% to 80%.Two studies used the tuning fork test to investigate any suspected fracture ,[2,6] one suspected femoral neck fractures,[7]one ankle inversion injuries,[8] and two stress fractures.[9,10] 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 by using pain induced by the vibrating tuning fork,[2,8,9,10]while two studies used reduced sound conduction .[7,6] Four studies used a 128Hz tuning fork alone,[6,7,8,9] but two studies compared the diagnostic accuracy of different frequency tuning forks within the studies.[2,10]

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 reference test had been blind and independent of the reference standard (Table 2).

Figure 2 shows sensitivity versus 1- specificity (ROC plot) for the 6 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 5 patients with fractures,[1].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 odds ratios > 10,but other studies showed only modest values (Table 3). The two studies that compared the diagnostic accuracy of different frequency tuning forks on the same patients found no differences between frequencies.[2,10]One study assessed the differences between pain ratings but differences were small. The one study that assessed inter-tester reliability showed only low reliability.[9]

Discussion

Two forms of tuning fork test, one based on pain induction and one 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 upon 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 done or to characteristics of the injuries and fractures.

The low inter-tester reliability suggests that the techniques would benefit from standardization and training. Wilder et al,[10]compared different frequencies and found a higher induction of

fracture pain using 256Hz, 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[9] 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 increased activity in the fractured area. Timing may also affect the accuracy of the test.

A mineralized 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 also unclear whether swelling or bruising in the area of the injury might affect the results.

A systematic review,[11]which examined a variety of methods for the diagnosis of stress fractures included only 2 of the 6 studies we used in this review.

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

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)

Table 1: Characteristics of the included studies

	Bache[7] (1984)	Moore[6] (2009)	Lesho[9] (1997)	Kazemi[2] (2000)	Dissmann[8] (2006)	Wilder[10] (2008)
Index test	**Sound conduction**		**Pain from vibration**			
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/ orthopedic center	Army medical center	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 leg 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'

Table 2: Methodological quality of the included studies

Criterion	Bache[7] (1984)	Moore[6] (2009)	Lesho[9] (1997)	Kazemi [8] (2000)	Dissmann[2] (2006)	Wilder[10] (2008)
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/ Pre-specified 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

Table 3: Overview of the results of the included studies

Results of testing	Bache[7] (1984)	Moore[6] (2009)	Lesho[9] (1997)	Kazemi[2] (2000)		Dissmann[8] (2006)		Wilder[10] (2008)		
Type of tuning fork	128Hz	128Hz	128Hz	128Hz	256Hz	128Hz TLM	128Hz DFS	128Hz	256Hz	512Hz
Prevalence of fractures	56%	32%	61%	80%	80%	10%	10%	27%	27%	27%
Sensitivity (%, 95% CI)	91 (81-96)	83 (55-95)	75 (57-87)	89 (75-96)	89 (75-96)	92 (52-99)	92 (52-99)	83 (55-95)	92 (67-99)	77 (49-92)
Specificity (%, 95% CI)	18 (9-32)	80 (61–91)	67 (44-84)	44 (19-73)	44 (19-73)	61 (46-74)	94 (84-98)	37 (23-55)	19 (9-36)	64 (47-79)
Diagnostic Odd ratio	2.3 (0.7-7.5)	20.0 (3.3-122)	6.0 (1.6-22)	6.6 (1.2-35.2)	6.6 (1.2-35.2)	17.3 (0.9-332)	187.0 (7.9-4424)	3.0 (0.6-16.1)	2.9 (0.3-26.7)	6.1 (1.4-26.7)
Positivelike lihood ratio: (95% CI)	1.1 (0.94-1.3)	4.2 (1.8–9.5)	2.2 (1.1-4.5)	1.6 (0.89-2.9)	1.6 (0.89-2.91)	2.4 (1.5–3.7)	16.5 (4.8-56)	1.3 (0.92-1.9)	1.1 (0.91-1.4)	2.2 (1.2-3.8)
Negativelik elihood ratio: (95% CI)	0.49 (0.17-1.4)	0.21 (0.06–0.75)	0.37 (0.18-0.77)	0.24 (0.08-0.79)	0.24 (0.08-0.79)	0.14 (0.01-2.0)	0.09 (0.01-1.3)	0.45 (0.12-1.7)	0.39 (0.05-3.0)	0.36 (0.13-0.99)

TLM- tip of lateral malleolus ; DFS- distal fibula shaft

Acknowledgements

We extend our gratitude to SarahThorning(Trial search coordinator, Bond University) for valuable help in literature search and Elaine Beller(Statistician, Bond University) for statistical support.

Contributorship Statement

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

Competing interest

The authors have declared no competing interest

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

Data Sharing No additional data available

Figures and Tables

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

Figure 2: Sensitivity versus 1-specificity (ROC) plot of included studies.

Table 1: Characteristics of the included studies

Table 2: Methodological quality of the included studies

Table 3: Overview of the results of the included studies

References:

1 Weisberg JL. Evaluation of bone continuity by sound conduction. *Bulletin of the United States Army Medical Department*.1945;4**:**471-4

2 Kazemi M, Roscoe MW. Is the tuning fork test a reliable tool in detecting acute simple fractures?*International Sports Journal*.2000 Summer;4(2):1-8.

3 http://www.stats.org.uk/statistical-inference/Newcombe1998.pdf Accessed on 5th July 2014

4 Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *Journal of Clinical Epidemiology*.1991;44**:**763-70.

5 Whiting P, Rutjesw AW, Westwood M, et al.Updating QUADAS: evidence to inform the development of QUADAS-2. 2010.Accessed at www.bris.ac.uk/quadas/resources/quadas2reportv4.pdf on 5 July 2014.

6 Moore MB. The use of a tuning fork and stethoscope to identify fractures.*Journal of Athletic Training*. 2009;44(3):272-4.

7 Bache JB, Cross AB. The Barford test. A useful diagnostic sign in fractures of the femoral neck. *Practitioner*.1984;228(1389):305-8.

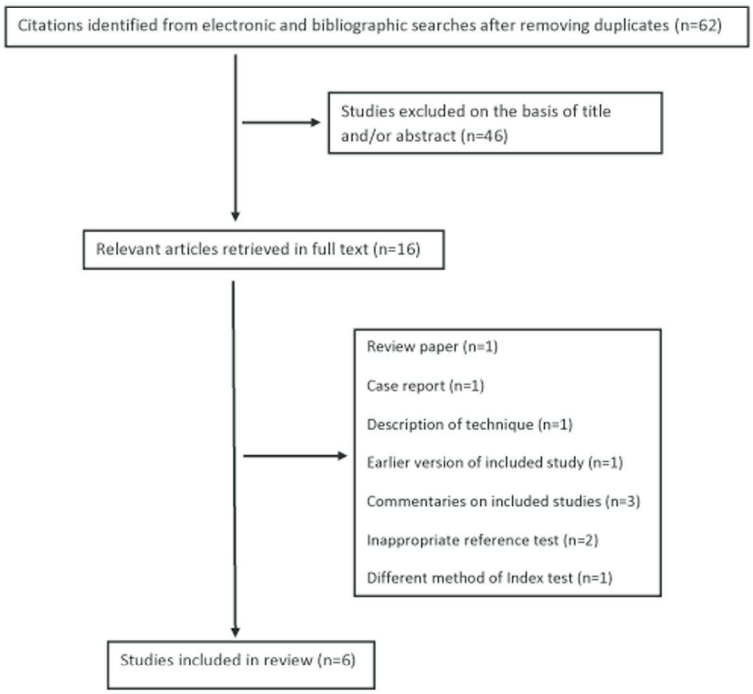
8 Dissmann PD, Han KH. The tuning fork test: a useful tool for improving specificity in ‘‘Ottawa positive’’ patients after ankle inversion injury. *Emerg Med J*. 2006;23(10):788– 90.

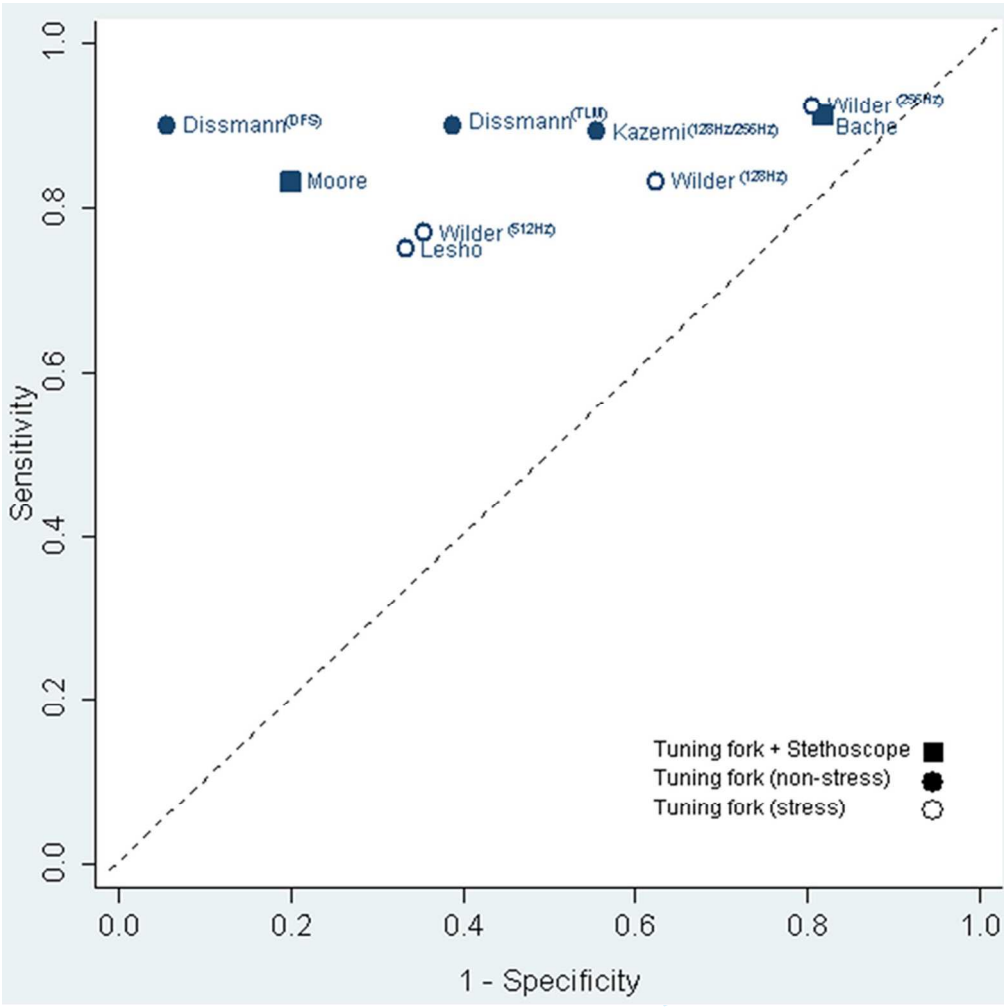
9 Lesho EP. Can tuning forks replace bone scans for identification of tibial stress fractures? *Mil Med*.1997;162(12):802-3.

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: *The Journal for the Practicing Clinician*. 2009;1(1):12-8.

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(9):760-71.

Figure 1: Flowchart of studies included in the review




Dissmann (DFS)
Dissmann (TLH)
Kazemi (128Hz/256Hz)
Wilder (256Hz)
Bache
Moore
Wilder (128Hz)
Wilder (512Hz)
Lesho
Tuning fork + Stethoscope
Tuning fork (non-stress)
Tuning fork (stress)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	nil
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	3
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	3-4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	3-4
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	4, QUADAS-2
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9, QUADAS-2
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	20

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency ($e.g., I^2$) for each meta-analysis.	4, 5 and Fig2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9 , TABLE 2
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	TABLE 3,Fig2
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	4, Fig 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	4-5, TABLE 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	9, TABLE 2
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	TABLE 3, Fig 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10, TABLE 3, FIG 2
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	9, TABLE 2
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Fig 2
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	5-6
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	5-6
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	6
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	**Primary Health Care Research Evaluation & Development**

PRISMA 2009 Checklist

			(PHCRED)

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	nil
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	3
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	3-4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	3-4
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	4, QUADAS-2
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	9, QUADAS-2
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	20
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	4, 5 and Fig2

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	9 , TABLE 2
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	TABLE 3,Fig2
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	4, Fig 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	4-5, TABLE 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	9, TABLE 2
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	TABLE 3, Fig 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10, TABLE 3, FIG 2
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	9, TABLE 2
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Fig 2
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	5-6
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	5-6
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	6
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Primary Health Care Research Evaluation & Development (PHCRED)