Evaluating medical tests and biomarkers
Wang, J.

Citation for published version (APA):
Wang, J. (2017). Evaluating medical tests and biomarkers: In primary studies and systematic reviews

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Download date: 30 Dec 2018
Computed tomography versus magnetic resonance imaging versus bone scintigraphy for clinically suspected scaphoid fractures in patients with negative plain radiographs

Wouter Mallee, Junfeng Wang, Rudolf Poolman, Peter Kloen, Mario Maas, Henrica de Vet, Job Doornberg

Cochrane Database of Systematic Reviews 2015, Issue 6.
Abstract

Background

In clinically suspected scaphoid fractures, early diagnosis reduces the risk of non-union and minimises loss in productivity resulting from unnecessary cast immobilisation. Since initial radiographs do not exclude the possibility of a fracture, additional imaging is needed. Computed tomography (CT), magnetic resonance imaging (MRI) and bone scintigraphy (BS) are widely used to establish a definitive diagnosis, but there is uncertainty about the most appropriate method.

Objectives

The primary aim of this study is to identify the most suitable diagnostic imaging strategy for identifying clinically suspected fractures of the scaphoid bone in patients with normal radiographs. Therefore we looked at the diagnostic performance characteristics of the most used imaging modalities for this purpose: computed tomography, magnetic resonance imaging and bone scintigraphy.

Search methods

In July 2012, we searched the Cochrane Register of Diagnostic Test Accuracy Studies, MEDLINE, EMBASE, the Database of Abstracts of Reviews of Effects, the Cochrane Central Register of Controlled Trials, the NHS Economic Evaluation Database. In September 2012, we searched MEDION, ARIF, Current Controlled Trials, the World Health Organization (WHO) International Clinical Trials Registry Platform, conference proceedings and reference lists of all articles.

Selection criteria

We included all prospective or retrospective studies involving a consecutive series of patients of all ages that evaluated the accuracy of BS, CT or MRI, or any
combination of these, for diagnosing suspected scaphoid fractures. We considered the use of one or two index tests or six-week follow-up radiographs as adequate reference standards.

**Data collection and analysis**

Two review authors independently screened titles and abstracts and assessed full-text reports of potentially eligible studies. The same authors extracted data from full-text reports and assessed methodological quality using the QUADAS checklist. For each index test, estimates of sensitivity and specificity from each study were plotted in ROC space; and forest plots were constructed for visual examination of variation in test accuracy. We performed meta-analyses using the HSROC model to produce summary estimates of sensitivity and specificity.

**Main results**

We included 11 studies that looked at diagnostic accuracy of one or two index tests: four studies (277 suspected fractures) looked at CT, five studies (221 suspected fractures) looked at MRI and six studies (543 suspected fractures) looked at BS. Four of the studies made direct comparisons: two studies compared CT and MRI, one study compared CT and BS, and one study compared MRI and BS. Overall, the studies were of moderate to good quality, but relevant clinical information during evaluation of CT, MRI or BS was mostly unclear or unavailable.

As few studies made direct comparisons between tests with the same participants, our results are based on data from indirect comparisons, which means that these results are more susceptible to bias due to confounding. Nonetheless, the direct comparisons showed similar patterns of differences in sensitivity and specificity as for the pooled indirect comparisons.

Summary sensitivity and specificity of CT were 0.72 (95% confidence interval (CI) 0.36 to 0.92) and 0.99 (95% CI 0.71 to 1.00); for MRI, these were 0.88 (95% CI 0.64 to 0.97) and 1.00 (95% CI 0.38 to 1.00); for BS, these were 0.99 (95% CI
0.69 to 1.00) and 0.86 (95% CI 0.73 to 0.94). Indirect comparisons suggest that diagnostic accuracy of BS was significantly higher than CT and MRI; and CT and MRI have comparable diagnostic accuracy. The low prevalence of a true fracture among suspected fractures (median = 20%) means the lower specificity for BS is problematic. For example, in a cohort of 1000 patients, 112 will be over-treated when BS is used for diagnosis. If CT is used, only 8 will receive unnecessary treatment. In terms of missed fractures, BS will miss 2 fractures and CT will miss 56 fractures.

**Authors’ conclusions**

Although quality of the included studies is moderate to good, findings are based on only 11 studies and the confidence intervals for the summary estimates are wide for all three tests. Well-designed direct comparison studies including CT, MRI and BS could give valuable additional information.

Bone scintigraphy is statistically the best diagnostic modality to establish a definitive diagnosis in clinically suspected fractures when radiographs appear normal. However, physicians must keep in mind that BS is more invasive than the other modalities, with safety issues due to level of radiation exposure, as well as diagnostic delay of at least 72 hours. The number of overtreated patients is substantially lower with CT and MRI.

Prior to performing comparative studies, there is a need to raise the initially detected prevalence of true fractures in order to reduce the effect of the relatively low specificity in daily practice. This can be achieved by improving clinical evaluation and initial radiographical assessment.
2.1 Plain language summary

2.1.1 Comparing different types of scan (CT, MRI, bone scan) for diagnosis of clinically suspected scaphoid fractures, when initial radiographs are negative

This summary of a Cochrane review presents what we know from research about the accuracy of imaging tests to detect true scaphoid fractures among suspected fractures.

When a patient presents to the emergency department with wrist injury and clinical signs of a scaphoid fracture, normal initial radiographs do not exclude a fracture. Approximately 20% of them do have a true scaphoid fracture and need additional imaging to establish a definitive diagnosis. Because of the low healing potential of the scaphoid bone, adequate diagnosis and treatment is vital to prevent complications such as non-union. If a patient is clinically suspected for a scaphoid fracture, their wrist will be immobilised in a cast until definitive diagnosis is obtained. This fear of under-treatment results in a large amount of over-treated wrist injuries. Computed tomography (CT), magnetic resonance imaging (MRI) and bone scintigraphy (BS; bone scan) are all imaging modalities that can be chosen at this stage. The aim of this systematic review was to establish which is the superior technique for identifying a true fracture and preventing unnecessary treatment. A high sensitivity reduces the risk of missing fractures; a low specificity increases the number of unnecessary treatments.

We conducted a thorough search of electronic databases, trial registers and conference proceedings up to July 2012. We included 11 studies in our analysis. The studies were moderate to good quality. Four studies (277 suspected fractures) looked at CT, five studies (221 suspected fractures) looked at MRI and six studies (543 suspected fractures) looked at BS. Four of these studies directly compared two modalities, such as both CT and MRI. When we compared the pooled data for the different imaging tests from all studies, we found that BS has the highest sensitivity,
but specificity was lower than CT and MRI. All three imaging tests were found to be highly accurate for definitive diagnosis. CT and MRI were comparable in diagnostic accuracy (the correct diagnosis is made). Although BS had significantly better accuracy than CT and MRI, it could lead to more people receiving unnecessary treatment. Moreover, BS is an invasive technique and is believed to be inappropriate for use in some populations, especially children.

Future studies should focus on improving clinical evaluation to raise the prevalence of true fractures. In addition, more direct comparison studies could add valuable data to determine which modality is superior in diagnosis of suspected scaphoid fractures.

2.2 Background

2.2.1 Target condition being diagnosed

The scaphoid bone is one of the carpal wrist bones and is located in the proximal row. Its surface mainly consists of cartilage and it articulates with the distal radius, and with four other carpal bones: the lunate, trapezium, trapezoid and capitate. When flexing and extending the wrist, the scaphoid rotates forwards and backwards. The same movements can be found when twisting the wrist from the radial to the ulnar side. Owing to the scaphoid’s anatomy, position and kinematics, it serves a key role in the function of the wrist.

Sustaining a fall on an outstretched hand (FOOSH) is the typical mechanism for fracturing the scaphoid. ’Axial fist’ trauma, involving transmission of an external force through the second metacarpal when the fist is clenched, as when punching, is another, less common, cause. These types of trauma are most common in young and active males performing sports. Scaphoid fractures constitute approximately 2% to 3% of all fractures [4]. The scaphoid is the most commonly fractured carpal bone [4–7].

One of the problems with fracturing the scaphoid is its low healing potential.
The scaphoid’s blood circulation mainly derives from small branches of the radial artery entering the bone from the distal part. The blood supply is fragile and can be interrupted when fractured [8, 9]. If untreated, this can lead to non-union, with or without avascular necrosis, and finally carpal collapse and disability [8, 10]. Early detection and adequate treatment can provide predictable and satisfactory rates of healing [11]. In contrast, delay of diagnosis and failure to recognise displacement are important risk factors for non-union of scaphoid wrist fractures [12, 13].

When someone with a FOOSH or 'axial fist’ trauma presents to the emergency department, certain clinical findings can lead to suspecting a scaphoid fracture. The most important physical examinations are pressing the anatomical snuffbox and applying longitudinal thumb compression [14–16]. If either of these result in pain in the scaphoid area, radiographs of the wrist and the scaphoid are necessary. Usually x-rays are then obtained in four views: postero-anterior, true lateral, semipronated oblique, and posteroanterior with the wrist in ulnar deviation [17]. Most scaphoid fractures will be identified with this imaging technique, but up to 16% are missed on initial radiographs [18, 19]. These missed fractures are also known as occult fractures. When clinical and radiographic findings do not match, we speak of a 'clinically suspected scaphoid fracture’ and additional imaging (second-line imaging) is needed.

In cases of inadequate or delayed diagnosis, possible problems in union (bone healing) can lead to functional wrist problems [9, 10]. Therefore, despite the normal radiographs, current clinical practice is to immobilise the scaphoid in a cast or splint until further imaging is established. The fear of under-treatment results in over-treatment of five out of six patients [19, 20].

Difficulties in detecting occult scaphoid fractures have been addressed in many radiological studies, aiming at exploring the value of novel imaging techniques or updates of already known techniques such as computed tomography (CT), magnetic resonance imaging (MRI), bone scintigraphy (BS) and ultrasound (US) [21–24]. However, there is currently no consensus regarding which modality is best to
detect an occult scaphoid fracture. Several worldwide and national studies showed considerable variation in the management of occult scaphoid fractures [25, 26]. This is partly attributed to the availability of the imaging tools and differences in costs, but also to the controversies regarding the best method to detect true scaphoid fractures. The international questionnaire-based survey of Groves (2006) revealed equivalent imaging strategies for suspected scaphoid fractures in only 6.7% of the, mainly university, hospitals [26]. Groves reported that the most commonly used second-line imaging modality in Europe was CT, whereas it was BS in Australasia and MRI in North America. This variation shows that there is a lack of agreed standard diagnostic practice, which amplifies the need for this review. Furthermore, the increase in availability of CT scanning in emergency (radiology) departments and dedicated MRI equipment, such as tailored sequences and dedicated wrist coils, enables earlier use of these techniques in daily clinical practice. Yet clear evidence of optimal scaphoid conventional imaging protocols is lacking, especially concerning cost effectiveness and patient safety (radiation protection).

Besides detecting a fracture, the location of the scaphoid fracture is important too. The proximal pole of the scaphoid is prone to complications after fracture owing to its limited vascularity. It has been proposed that these fractures need to be treated operatively because cast immobilisation will not ensure adequate healing. This differs from undisplaced fractures through the waist of the scaphoid for which union rates of up to 95% have been reported after cast immobilisation [27].

In general, the key to evaluating the performance of a diagnostic test is an agreed-upon reference standard that is used to define the presence or absence of a disease. We know that an important caveat in the interpretation of studies of the diagnostic performance characteristics of various imaging modalities for triage of suspected scaphoid fractures is the lack of an agreed-upon reference standard for the diagnosis of a true fracture of the scaphoid. The most commonly applied test is the six-week follow-up set of radiographs. This is generally considered to be the most valid reference test [19]. When we examine some of the prospective trials studying
one or more index tests, lists of reference standards are often given. Other methods used are:

- if two of the index tests are positive (MRI, CT, BS), the diagnosis is a fracture;
- if two of the index tests are negative (MRI, CT, BS), the diagnosis is 'no fracture’;
- clinical follow-up and radiographs after two weeks;
- clinical follow-up and MRI;
- single use of an index test (MRI, CT, BS);
- single use of clinical follow-up.

These methods are sometimes used in research as reference standards but some are considered suboptimal. These differences in approach hamper the interpretation of the scaphoid imaging literature because most of the results found are not checked with an optimal reference test. We consider the single use of an index test (MRI, CT, BS) and the clinical follow-up with radiographs after two weeks as a 'suboptimal’ reference test. The use of clinical follow-up alone is even more unsatisfactory as a reference standard.

Treatment of a non-displaced or minimally displaced fractured scaphoid can be operative or non-operative and is mainly based on the location of the fracture. The majority of the fractures are located in the waist of the scaphoid [27]. Whereas waist and distal pole fractures seem to heal with acceptable rates with cast treatment, it is a fracture of the proximal pole that is prone to non-union. Therefore, these fractures are considered unstable and require operative treatment [28]. The non-operative method is with use of a cast or splint that prevents the scaphoid’s movements. Healing of a scaphoid fracture to union is a time-consuming process that results in the need for a long period of immobilisation, ranging from 6 to 12 weeks [11, 29, 30]. To avoid this burden, operative fixation with a headless compression screw can be performed [31]. Surgical treatment is favourable in terms of time off work and functional outcome, but can lead to more (minor) complications [32].
2.2.2 Index test(s)

The tests evaluated in this review are multi-slice CT, MRI and BS.

CT creates axial images of the wrist that can be reconstructed in different planes, such as anatomical coronal and sagittal series. Several studies show preferable use of reconstructions in planes defined by the long axis of the scaphoid [19, 33, 34]. Image reconstruction in CT is a mathematical process that generates images from X-ray projection data acquired at many different angles around the patient. Image reconstruction has a fundamental impact on image quality and therefore on radiation dose. No literature could be found comparing different types of image reconstruction; we will therefore evaluate all types in this review.

MRI generates a strong magnetic field to align the hydrogen atoms in the body. This alignment is altered with use of radiofrequency pulses and can be detected to build the images. MRI was the first non-invasive method to create high-resolution images of the musculoskeletal system. In scaphoid injury, bone bruising or bone marrow oedema consists mainly of liquid with hydrogen atoms, and thus is well visualised. Cortical involvement of the fracture can, therefore, be less obvious. The exact value of bone marrow oedema in the clinical spectrum of scaphoid injury is unclear; as is its relationship to patient outcome.

BS is widely described for scaphoid disorders. After an intravenous injection with radioactive isotopes, the osteoblastic activity can be visualised. A gamma camera can detect the radiation emitted by the isotopes. Where there is a fracture, osteoblastic activity is high at the fracture site indicating the natural healing process of the bone. This activity is displayed as a dense spot in the bone. BS provides a radiation burden and is thus potentially harmful, especially to the younger age group.

When we consider the negative aspects of the additional imaging methods, we find that:

MRI:

• is known for its low availability and generally higher costs compared with CT;
• produces images in which bone bruising can be difficult to distinguish from a
fracture [19]. No clear criteria for a bruise or a fracture are established. When bone bruising is detected, the possibility of fracture development must be remembered [35]; and thus follow-up is important.

CT:

- is one of the modalities that uses radiation. Although the dose of 0.03 mSv for imaging the wrist is very low [36], its use in the younger patient group is debatable.

BS:

- uses radiation. With 4 mSv, the dose is much higher than CT, but still only the same as two years of natural background radiation [20]. BS is not recommended for children;
- needs radioactive isotopes that must be injected intravenously, which makes BS the most invasive procedure of all;
- can only be performed with an interval of 72 hours after injury. This delay is needed to capture osteoblastic activity at the fracture site in all patients [37];
- in the lead author’s hospital, the costs of BS are comparable with those for MRI.

### 2.2.3 Alternative test(s)

Ultrasound (US) can be used to diagnose suspected scaphoid fractures. The literature evaluating its performance characteristics is scarce and the latest review including US shows inferior results compared with MRI, CT or BS [38]. In addition, an international survey of imaging strategies among hospitals revealed no use of US for these injuries [26]. This review therefore does not consider US.

Another test, six-week follow-up radiographs, is extensively used in literature as a reference standard [19, 39]; but its accuracy is being questioned [19]. One of the main disadvantages is the time interval before this test can be performed, given the need for immobilisation. The importance of immediate diagnosis rules out the use
of the follow-up radiographs as an adequate diagnostic tool. Moreover, a positive CT, MRI or bone scan can be accompanied by normal x-rays after six weeks. These disadvantages make the quality and clinical applicability of this test questionable.

### 2.2.4 Rationale

In clinically suspected scaphoid fractures, early diagnosis reduces the risk of non-union and minimises any loss in productivity resulting from unnecessary cast immobilisation [40]. This means improvement of short-term management (avoid unnecessary immobilisation) and long-term outcome (risk of non-union, avascular necrosis). The value of an imaging tool with the highest accuracy is of great importance for both the patient and economically in terms of healthcare costs and productivity loss.

There are many controversies surrounding the choice of imaging modality; this is reflected in the considerable variation in practice [26]. All three imaging modalities (CT, MRI and BS) are widely used and reviews of these have reported that all show high sensitivity and specificity rates [17, 38]. The most recent review searched up to October 2008, but did not include non-English studies even though there were three potentially eligible reports in foreign languages [17]. Since 2000, several articles evaluating one or two tests have been published. Hence, an update of the evidence was warranted.

With this review, we evaluated the diagnostic performance characteristics of BS, MRI and CT with an updated search for diagnostic accuracy studies and the inclusion of non-English literature.

### 2.3 Objectives

The primary aim of this study is to identify the most suitable diagnostic imaging strategy for identifying clinically suspected fractures of the scaphoid bone in patients with normal radiographs. Therefore we looked at the diagnostic performance
characteristics (Appendix 1) of the most used imaging modalities for this purpose: computed tomography, magnetic resonance imaging and bone scintigraphy.

2.3.1 Secondary objectives

To investigate which imaging technique is the best for determining the location of the fracture (proximal, waist or distal).

2.3.2 Investigation of sources of heterogeneity

We assessed the potential influence of sources of heterogeneity on the diagnostic accuracy of the tests, especially the type of reference standard and blinded evaluation of the reference test (if reported).

2.4 Methods

2.4.1 Criteria for considering studies for this review

Types of studies

All prospective or retrospective studies involving a consecutive series of patients. We only included trials using reference standards that we considered optimal or adequate. Randomised controlled trials would have been included if these had been found.

Participants

People of all ages who presented at hospital or clinic within one week of trauma with a clinically suspected scaphoid fracture and negative post-trauma radiographs. Clinical suspicion of a scaphoid fracture is based on pain in the anatomical snuffbox or by longitudinal compression of the thumb, or both. The radiographs generally include two images of the wrist (postero-anterior and lateral views) and at least one of two additional scaphoid views.
Index tests

CT, MRI or BS, or a combination of two of these tests. Because the criteria for a fracture may differ (especially in MRI), we report all study characteristics, including 'fracture criteria’, in 'Characteristics of included studies’.

Target conditions

Clinically suspected scaphoid fractures (which could be proximal, waist or distal) with negative plain radiographs.

Reference standards

Various reference standards were included.

1. A scaphoid plain radiograph series, conducted six to 14 weeks after the initial injury, consisting of the following four views: posteroanterior with the wrist in neutral position; lateral; semipronated oblique scaphoid; and radial oblique scaphoid. An abnormal lucent line within the scaphoid is considered evidence of a fracture.
2. The use of two index tests. If both are positive or negative, a final diagnosis is obtained.
3. In addition, clinical findings are often combined with an index test or repeated radiographs obtained after six weeks to formulate a reference standard.
4. The use of only one of the second-line modalities has been described; this is somewhat unsatisfactory because these diagnostic techniques are still under study.

We considered six-week follow-up radiographs (1) the most suitable reference standard. Next we considered the use of two index tests with the same outcome and one index test including clinical findings (2 and 3). Although we considered the fourth option to be suboptimal, it was included in the review.
We did not include studies using clinical findings only six to 14 weeks after trauma or the single use of one- to two-week follow-up radiographs as a reference standard as we consider these inadequate.

2.4.2 Search methods for identification of studies

Electronic searches

We searched the Cochrane Register of Diagnostic Test Accuracy Studies (July 2012), MEDLINE (1946 to July Week 1 2012) and EMBASE (1974 to 2012 Week 27). We also searched the Database of Abstracts of Reviews of Effects (The Cochrane Library 2012 Issue 7), MEDION (Meta-analyses van Diagnostisch Onderzoek) (September 15th 2012) and the Aggressive Research Intelligence Facility (ARIF) reviews database (15 September 2012) for relevant diagnostic reviews. In addition, we searched the Cochrane Central Register of Controlled Trials (The Cochrane Library 2012 Issue 7) and the NHS Economic Evaluation Database (The Cochrane Library 2012 Issue 7) for comparative and cost-effectiveness studies looking at different diagnostic modalities. We searched Current Controlled Trials (15 September 2012) and the WHO International Clinical Trials Registry Platform (15 September 2012) for ongoing studies.

We developed a sensitive search strategy for MEDLINE (Ovid Web), EMBASE (Ovid Web) and The Cochrane Library (Wiley Online Library) as recommended in Chapter 7 of the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (de Vet 2008). The search strategies for all databases are shown in Appendix 2.

There were no restrictions based on language or publication status.

Searching other resources

We checked the reference lists of all articles, including reviews, for relevant primary diagnostic studies and systematic or narrative reviews.
We handsearched the abstracts of the conference proceedings of two societies: the American Society for Surgery of the Hand annual meetings (2000 to 2012); and the American Academy of Orthopaedic Surgeons annual meetings (2011 to 2013). If potentially eligible abstracts were found, we searched for the full reports.

We also contacted experts in the field and main investigators of relevant ongoing studies for additional information.

2.4.3 Data collection and analysis

Selection of studies

Two review authors (WHM and JND) independently screened the titles and abstracts of retrieved publications to identify potentially eligible studies for inclusion. WHM and JND assessed full-text reports of potentially eligible studies and independently determined study inclusion or exclusion. Any disagreement was either resolved by discussion; or, if necessary, by an arbiter (RWP). When WHM and JND were involved in one of the studies, two other authors (RWP and PK) were asked to assess eligibility. Only results of full reports were evaluated.

Data extraction and management

Two review authors (WHM and JND) independently extracted data from full-text reports. If studies had been published more than once, only data from the latest or most suitable report were included. (In cases of overlapping patient data, we only used the data once.) Any disagreement was discussed, either until consensus was achieved, or, if necessary, with an arbiter (RWP). When WHM and JND were involved in one of the included studies, two other authors were asked to extract data. Where necessary, we contacted study authors for additional information or data.

The following data were collected:

1. general information: title, journal, year, publication status, country of study, period of study, primary objective and study design (prospective versus
2. sample size (screened and included);
3. baseline characteristics: age, sex, side of injury, trauma mechanism, time of presentation, inclusion and exclusion criteria;
4. target condition, as reported;
5. index test: description of technique, criteria for a fracture, timing of test and expertise of the tester;
6. reference standard test: description of technique, criteria for a fracture, time from trauma to reference test and expertise of the tester;
7. sensitivity and specificity;
8. number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

Assessment of methodological quality

Two review authors (WHM and JND) independently assessed the methodological quality of the included studies using a slightly modified version of the QUADAS checklist [41]. Both review authors had prior knowledge of the methodological aspects of diagnostic accuracy studies. Where any disagreement on the quality assessment occurred, a third review author (RWP) was asked to arbitrate. When WHM and JND were involved in one of the included studies, two other authors were asked to assess the methodological quality. We used the QUADAS checklist with previously set criteria specific to the review topic (Table 2.1).

Table 2.1: QUADAS checklist and assessment criteria.

<table>
<thead>
<tr>
<th>Item question</th>
<th>Item answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was the spectrum of patients representative of the patients who will receive the test in practice? (representative spectrum)</td>
<td>Yes: 1) presentation to the emergency department within 72 hours; 2) all included patients were suspected of having a scaphoid fracture with normal radiographs; 3) prospective study design; and 4) consecutive series registration. Unclear: if insufficient information is presented on study design or inclusion criteria. No: 1) patients presented after 72 hours; 2) retrospective study design; or 3) not a consecutive series of patients.</td>
</tr>
<tr>
<td></td>
<td>Question</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2.</td>
<td>Is the reference standard likely to classify the target condition correctly? (acceptable reference standard)</td>
</tr>
<tr>
<td>3.</td>
<td>Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the 2 tests? (acceptable delay between tests)</td>
</tr>
<tr>
<td>4.</td>
<td>Did the whole sample or a random selection of the sample receive verification using the intended reference standard? (partial verification avoided)</td>
</tr>
<tr>
<td>5.</td>
<td>Did patients receive the same reference standard irrespective of the index test result? (differential verification avoided)</td>
</tr>
<tr>
<td>6.</td>
<td>Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)? (incorporation avoided)</td>
</tr>
<tr>
<td>7.</td>
<td>Were the reference standard results interpreted without knowledge of the results of the index test? (index test results blinded)</td>
</tr>
<tr>
<td>8.</td>
<td>Were the index test results interpreted without knowledge of the results of the reference standard? (reference standard results blinded)</td>
</tr>
</tbody>
</table>
To inform our assessment of overall methodological quality we established the following general 'rules'. We considered the methodological quality was 'excellent' if all QUADAS items were met; and 'good' if at least item 2 (acceptable reference standard?) was scored as 'yes', with the other items open for discussion between the two review authors (WHM and JND). We considered quality was 'moderate' if either item 1 (representative spectrum?) or item 2 was scored as 'unclear' or 'no'; again with the other items open for discussion. We considered quality was 'poor' if both items 1 and 2 were scored 'no'.

**Statistical analysis and data synthesis**

The main target was to identify the index test with the highest diagnostic accuracy for diagnosing suspected scaphoid fractures. With the outcomes of each primary study, we generated $2 \times 2$ tables (with TPs, TNs, FPs and FNs) for each diagnostic test according to the presence or absence of a true fracture. With these data, sensitivity and specificity fractions are presented. Where results were reported as 'inconclusive' (as in [42]), we treated these as negative findings. If the data presented in trials
had been uninterpretable in that $2 \times 2$ tables could not be generated, we planned to contact the original authors of the study for clarification, and otherwise present the data only descriptively.

The two main parameters of diagnostic test accuracy are sensitivity and specificity. As there is a trade-off between these parameters, they should not be analysed separately. For descriptive purposes, coupled forest plots are presented showing the pairs of sensitivities and specificities with 95% confidence intervals (CIs). Sensitivity and specificity are displayed in the receiver operating characteristic (ROC) space.

Diagnostic accuracy was first evaluated for each index test individually. For pooling sensitivities and specificities, we assume there is at least one common criterion for test positivity used across studies for a given test. Given the fact that different studies may have slightly different criteria for test positivity, and individual observers within a study may interpret the criteria a little differently, the bivariate random effects model was used to get the summary estimates of sensitivity and specificity. A separate model was fitted for each index test with bivariate approach except CT. For CT, the estimation from the bivariate model did not converge. This may be due to the small number of studies (four studies for CT) included in the meta-analysis. So we used the HSROC model as an alternative, which could give mathematical equivalent estimates of bivariate approach. Both models produced summary estimates of the mean sensitivity and specificity with corresponding 95% CIs. Summary estimates of sensitivity, specificity, positive likelihood ratio, negative likelihood ratio and their 95% CIs were calculated by using 'estimate' command in SAS.

Pairwise comparisons between CT, MRI and BS were based on the overall performance, measured by diagnostic odds ratio (DOR). We added test type as covariate into the HSROC model and tested the statistical significance of the covariate effects on the test accuracy. The strategy of comparison was as follows: first, we had model (a) (Table 2.2), which included covariates for shape (beta), accuracy (alpha) and threshold (theta); then covariate for shape was dropped and we got model (b),
and a $\chi^2$ test was performed on the change in the -2 log likelihood from model (a) to model (b). If the curves had different shapes, it indicated that the differences in test accuracy depended on threshold. Otherwise, we continued to drop the covariate for accuracy and got model (c), and then compared -2 log likelihood with model (b) using the $\chi^2$ test. If the likelihood test showed a significant change from model (b) to model (c), then we can say there is a significant difference in the accuracy between the tests being compared.

**Table 2.2:** -2 Log Likelihood of models in each pairwise comparison

<table>
<thead>
<tr>
<th></th>
<th>-2 Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model (a)</td>
</tr>
<tr>
<td>CT vs MRI</td>
<td>48.0</td>
</tr>
<tr>
<td>CT vs BS</td>
<td>63.8</td>
</tr>
<tr>
<td>MRI vs BS</td>
<td>72.3</td>
</tr>
</tbody>
</table>

Model (a) assumed different shape (beta), accuracy (alpha) and threshold effect (theta)
Model (b) assumed different shape (beta) and accuracy (alpha)
Model (c) assumed different shape (beta) only

Our second target was to identify the accuracy of fracture location detection (proximal, waist, distal). This was not done for the current version of the review. Should there be sufficient studies containing adequate information about fracture location in future, we plan to include only the fractured scaphoids and generate 2 × 2 tables for each diagnostic test. We plan to present sensitivity, specificity and predictive values and calculate these in the same way as our main target. We also intend to consider a second option, which is to keep the entire dataset (i.e. including people with no fracture), and compute the relative sensitivity and specificity for fractures in different locations; and thereby compare the accuracy to detect the presence of a fracture at each location.

**Investigations of heterogeneity**

Heterogeneity in diagnostic test accuracy reviews is expected. Aside from analyses in which the different index tests are presented as subgroups, none of the
planned subgroup analyses to investigate heterogeneity were performed. Should there be sufficient data available in future, we will conduct subgroup analyses based on the assessment of methodological quality (yes versus no or unclear) from items 2 (acceptable reference standard?), 3 (acceptable delay between tests?), 4 (partial verification avoided?), 5 (differential verification avoided?) and 6 (incorporation avoided?) of the QUADAS criteria. Additionally, if there are sufficient studies, we will perform a meta-regression analysis. Characteristics of the index test, study population (adults/children), and judgements for the five QUADAS items will be added to the model as covariates, to analyse their influence on diagnostic accuracy. Heterogeneity will be judged on the scatter of points and from the prediction ellipse. This graphical information will also be used to decide about subgroups. We will present pooled estimates per clinical relevant subgroups. The possibilities of performing meta-regression analyses will depend on the number of studies available for a specific index test providing sufficient information.

**Sensitivity analyses**

During the review, a number of subjective choices were made with regard to eligibility, methodological quality and clinical similarity. The influence of these decisions on the outcome of the review should ideally be explored in sensitivity analyses (e.g. QUADAS item 12 (clearly described fracture criteria for index test)), but this was not possible since there were too few studies for proper analyses.

Our planned sensitivity analysis based on indirect comparison versus direct comparison was also hindered because of the small numbers of studies making direct comparisons. In order to compare the accuracy of the index tests, two strategies could be applied. We could include all studies examining one or more index test or we could include only studies that presented a direct comparison between two or more index tests. Although the first analysis is based on all available data, the second analysis potentially gives more valid data for the comparison. These two strategies may lead to different conclusions, so, while we decided to include all studies, we also
checked the results of the few direct comparison studies. If there had been sufficient data, we would also have examined whether the results of the meta-analyses would have changed if we had included only direct comparison studies.

2.5 Results

2.5.1 Results of the search

For this search (main search date July 2012), we screened a total of 2900 records from the following databases: Cochrane Register of Diagnostic Test Accuracy Studies (14 records); MEDLINE (1226); EMBASE (1586), the Database of Abstracts of Reviews of Effects (2); MEDION (0); ARIF (3), the Cochrane Central Register of Controlled Trials (34); the NHS Economic Evaluation Database (8); the WHO International Clinical Trials Registry Platform (13) and Current Controlled Trials (14). We did not identify potentially eligible studies from other sources.

The search resulted in the identification of 64 potentially eligible articles, for which (where possible) full reports were obtained. Upon study selection, we included 11 studies [19, 21, 24, 39, 42–48]; and 45 studies were excluded, one of which was published in two reports [49]. There were no ongoing trials or studies awaiting classification. All studies were written in English. Five studies were conducted in The Netherlands, two in Austria and one in each of Turkey, Ireland, Norway and Denmark. All studies included patients that presented to the emergency department with clinical suspicion of a scaphoid fracture, but with normal initial radiographs.

A flow diagram summarising the study selection process is shown in Figure 2.1.
Included studies

The characteristics of the individual studies are reported in the Characteristics of included studies. Four studies evaluated CT [19, 39, 44, 45]; five studies evaluated MRI [19, 21, 39, 43, 48]; and six studies evaluated BS [24, 42–44, 46, 47]. Of these
studies, two compared CT with MRI [19, 39]; one study compared BS with CT [44];
and one compared MRI with BS [43].

The main objective for all studies was the detection of a true scaphoid fracture
among clinically suspected scaphoid fractures. A total of 717 patients with 719
clinically suspected scaphoid fractures were assessed. For CT, 276 patients with 277
suspected fractures provided data; 221 patients for MRI; and 542 patients with 543
suspected fractures for BS. The sample size ranged from 16 to 159, with a mean
of 65 patients. The weighted mean age of the studies was 36.5 years (range 10 to
88 years). Five studies included children, one of which evaluated MRI [21]; and
the other four of which evaluated BS [24, 42, 46, 47]. The gender distribution was
available for 10 studies, in which the proportion of men ranged from 49.7% [44], to
100% [45].

Seven studies assessed patients within 72 hours of the patient injuring their wrist;
four studies did not report the timing of presentation to the emergency department
[21, 42, 46, 47]. In seven studies, the index test was performed within 10 days
of trauma [19, 21, 39, 42–45]. Tenderness in the anatomical snuffbox was clearly
incorporated in clinical evaluation in six studies [19, 21, 24, 43–45]. One study
reported 'pain over the scaphoid' as being clinically suspected [39]. Four studies
did not define the content of clinical evaluation [42, 46–48]. Images of BS were
evaluated by a consultant clinical nuclear physician in four studies [24, 43, 44, 48];
three studies (two when BS was an index test, one when BS was a reference standard)
did not provide the expertise of the observer(s) [42, 46, 47]. For MRI and CT,
evaluation was performed by at least one experienced radiologist.

This review focused on true scaphoid fractures among clinically suspected
scaphoid fractures. In addition, all studies reported on the diagnosis of other wrist
fractures (see ‘Characteristics of included studies’).
Excluded studies

We excluded 45 studies; the characteristics of these studies are presented in the 'Characteristics of excluded studies'. The most common reasons for exclusion were that no reference standard was used or that it was inadequate (21 studies), or that patients were included after a second clinical evaluation after one to two weeks (eight studies). Inadequate reference tests included repeating the radiographs after 10 days or using only clinical evaluation after one to two weeks. Some studies did not perform any other test besides initial clinical and radiographic evaluation.

2.5.2 Methodological quality of included studies

The included studies were diverse but all were of moderate to good quality (Figure 2.2 and Figure 2.3). Five studies were considered 'good quality' [19, 21, 24, 39, 42]; and six studies were considered 'moderate quality'. Of these, two studies had three items scored as low quality [43, 44] and one study had five items scored as unclear and one item scored as low quality [46].

All studies recruited patients consecutively as per our inclusion criteria. A prospective study design was clearly reported in eight studies. In three studies this was unclear, but due to the use of a reference standard, we assumed these were prospective as well. In only one study [47] was the spectrum of patients not clear; since the timing of presentation and precise aspects of clinical evaluation were not reported, we judged this study to be low quality for this item. Participants in nine studies received an acceptable reference standard: seven studies used follow-up radiographs in four or more views after at least six weeks [19, 21, 24, 39, 42, 46, 47]; and two studies used a mixed reference standard (same outcome in two index tests or six-week follow-up radiographs) [43, 44]. Two studies were judged at lower quality as they used suboptimal reference standards: one used MRI [45]; and the other used BS [48]. Because of the mixed use of at least one index test as a reference test, differential verification and incorporation bias could not be avoided in these two studies [43, 44].

Only one study reported the use of clinically relevant information during evaluation.
Figure 2.2: Methodological quality graph: review authors’ judgements about each methodological quality item presented as percentages across all included studies

of the images [39]; five studies excluded this information intentionally [42–45, 48]. The criteria for diagnosing a fracture was not defined in [46] for BS; for CT in [44] and [45]; and for MRI in [43] and [48]. However, we rated the latter four studies as unclear for this item because of other information and that the evaluation of test results was performed by at least two observers.
Figure 2.3: Methodological quality summary: review authors’ judgements about each methodological quality item for each included study

2.5.3 Findings

Indirect comparisons

The forest plots of the diagnostic performance characteristics of CT, MRI and BS are presented in Figure 2.4. The median prevalence of a true scaphoid fracture among clinically suspected scaphoid fractures with normal radiographs is 20% (range 11% to 44%). For CT, sensitivity estimates ranged from 0.67 (95% CI 0.22 to 0.96) to 0.88 (95% CI 0.62 to 0.98) and specificity estimates from 0.96 (95% CI 0.82 to 1.00) to 1.00 (95% CI 0.81 to 1.00). For MRI, sensitivity estimates ranged from 0.67 (95% CI 0.37 to 0.97) to 0.88 (95% CI 0.62 to 0.98) and specificity estimates from 0.96 (95% CI 0.82 to 1.00) to 1.00 (95% CI 0.81 to 1.00). For BS, sensitivity estimates ranged from 0.67 (95% CI 0.37 to 0.97) to 0.88 (95% CI 0.62 to 0.98) and specificity estimates from 0.96 (95% CI 0.82 to 1.00) to 1.00 (95% CI 0.81 to 1.00).
CI 0.22 to 0.96) to 1.00 (95% CI 0.72 to 1.00) and specificity estimates from 0.89
(95% CI 0.72 to 0.98) to 1.00 (95% CI 0.66 to 1.00). For BS, sensitivity estimates
ranged from 0.95 (95% CI 0.75 to 1.00) to 1.00 (95% CI 0.54 to 1.00) and specificity
from 0.52 (95% CI 0.41 to 0.63) to 1.00 (95% CI 0.82 to 1.00).

The study specific and pooled estimates and 95% confidence regions are displayed
in a scatter plot for CT, MRI and BS (Figure 2.5). The pooled estimates for CT
sensitivity and specificity were 0.72 (95% CI 0.36 to 0.92) and 0.99 (95% CI 0.71 to
1.00), respectively; the pooled estimates for MRI sensitivity and specificity were
0.88 (95% CI 0.64 to 0.97) and 1.00 (95% CI 0.38 to 1.00), respectively; and the
pooled estimates for BS sensitivity and specificity were 0.99 (95% CI 0.69 to 1.00)
and 0.86 (95% CI 0.73 to 0.94), respectively.

Pairwise comparisons were performed using HSROC model (see Statistical
analysis and data synthesis section above). The -2 log likelihood of model (a),
(b) and (c) of comparisons between each pair of tests are shown in Table 2.2. By
comparing -2 log likelihood between model (a) and model (b), we found that the
differences in test accuracy do not depend on threshold (since CT, MRI and BS do
not have thresholds), thus we could continue to compare the overall accuracy (DOR)
between tests. When comparing the overall accuracy (comparing model (b) and
model (c)) of these tests, significant differences were found in 'CT versus BS' (χ² =
50.3, df = 1, P value < 0.01) and 'MRI versus BS' (χ² = 29.7, df = 1, P value < 0.01),
which indicates that the overall accuracy of BS is higher than CT and MRI; while no
evidence was found for a difference in accuracy between CT and MRI (χ² = 1.9,
df = 1, P value = 0.17). These results may be explained by the findings from the
summary estimates (see 'Summary of findings'): BS has a slightly lower specificity
but a much higher sensitivity than CT and MRI, which leads to higher DOR for BS.

**Direct comparisons**

The separate findings of the four studies providing direct comparisons between
tests are shown in Table 2.3. The direct comparisons showed similar patterns of
differences in sensitivity and specificity as for the indirect comparisons.

The two studies directly comparing CT and MRI found comparable sensitivities.
Table 2.3: Direct comparisons (CT versus MRI; CT versus BS; MRI versus BS) for detection of scaphoid fractures

<table>
<thead>
<tr>
<th>Study</th>
<th>Cases (95% CI)</th>
<th>Non-cases (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Difference in sensitivity (95% CI)</th>
<th>Difference in specificity (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mallee 2011</td>
<td>6 (0.03 to 0.07)</td>
<td>20 (0.78 to 0.96)</td>
<td>0.67 (0.22 to 0.96)</td>
<td>0.96 (0.82 to 1.00)</td>
<td>-0.53 (-0.06 to 0.21)</td>
<td>0.07 (-0.06 to 0.21)</td>
<td>P = 1.0</td>
</tr>
<tr>
<td>Memarsadeghi 2006</td>
<td>11 (0.10 to 0.22)</td>
<td>11 (0.75 to 1.00)</td>
<td>1.00 (0.72 to 1.00)</td>
<td>1.00 (0.81 to 1.00)</td>
<td>-0.27 (-0.54 to -0.01)</td>
<td>0.00 (-0.06 to 0.21)</td>
<td>P = 0.214</td>
</tr>
<tr>
<td>De Zwart 2012</td>
<td>20 (0.75 to 1.00)</td>
<td>0.94 (0.88 to 0.98)</td>
<td>0.70 (0.67 to 0.73)</td>
<td>0.94 (0.88 to 0.98)</td>
<td>-0.25 (-0.47 to -0.03)</td>
<td>0.05 (0.01 to 0.10)</td>
<td>P = 0.092</td>
</tr>
<tr>
<td>Beeres 2008</td>
<td>20 (0.83 to 1.00)</td>
<td>0.90 (0.81 to 0.96)</td>
<td>0.80 (0.56 to 0.94)</td>
<td>0.90 (0.81 to 0.96)</td>
<td>-0.20 (-0.38 to -0.03)</td>
<td>0.10 (0.03 to 0.17)</td>
<td>P = 0.106</td>
</tr>
</tbody>
</table>

Sensitivity, specificity and their CIs are recalculated with Review Manager.
Figure 2.5: Study specific and pooled estimates of test performance for CT, MRI and BS with 95% confidence regions
and specificities for the two tests [19, 39], with neither trial finding statistically significant differences between tests (reported P values > 0.05). The study directly comparing CT with BS [44] reported a lower sensitivity, which was not statistically significant (reported P = 0.13) and a higher specificity (reported P = 0.02) for CT, but no statistically significant difference in the percentage of correct predictions (accuracy) (reported P = 0.63). The study directly comparing MRI with BS [43], which found a lower sensitivity and higher specificity for MRI, reported no statistically significant difference in the percentage of correct predictions with MRI and bone scintigraphy (P = 0.388).

Secondary objectives

There was no information about the diagnostic accuracy of the tests for identifying the location of the fracture (proximal, waist, distal).

2.5.4 Summary of findings

2.6 Discussion

2.6.1 Summary of main results

Early diagnosis and treatment of patients with a clinically suspected scaphoid fracture minimises the risk of complications and prevents unnecessary cast immobilisation. If initial radiographs appear normal, approximately 20% will still have a true fracture. In clinical practice, a definitive diagnosis is established by using CT, MRI or BS. This systematic review summarised the evidence and compared the diagnostic accuracies of these three imaging modalities. Eleven studies, four which evaluated two index tests, were included in the comparison: four studies for CT, five studies for MRI and six for BS.

We found evidence that BS has a significantly higher diagnostic accuracy (DOR) than CT and MRI; which reflects the higher sensitivity for BS. The summary
### Table 2.4: Summary of findings

<table>
<thead>
<tr>
<th>Comparing the diagnostic accuracies of computed tomography versus magnetic resonance imaging versus bone scintigraphy for clinically suspected scaphoid fractures in patients with negative plain radiographs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient population</strong></td>
</tr>
<tr>
<td><strong>Prior testing</strong></td>
</tr>
<tr>
<td><strong>Setting</strong></td>
</tr>
<tr>
<td><strong>Index tests</strong></td>
</tr>
<tr>
<td><strong>Reference standard</strong></td>
</tr>
<tr>
<td><strong>Target condition</strong></td>
</tr>
<tr>
<td><strong>Importance</strong></td>
</tr>
<tr>
<td><strong>Included studies</strong></td>
</tr>
<tr>
<td><strong>Number of suspected fractures (patients) studied</strong></td>
</tr>
<tr>
<td><strong>Quality concerns</strong></td>
</tr>
<tr>
<td><strong>Limitations</strong></td>
</tr>
</tbody>
</table>

Continued in the next page...
Table 2.4: Summary of findings (continued)

<table>
<thead>
<tr>
<th>Test</th>
<th>Number of studies</th>
<th>Number of suspected fractures</th>
<th>Summary sensitivity (95% CI)</th>
<th>Summary specificity (95% CI)</th>
<th>Summary LR+1 (95% CI)</th>
<th>Summary LR-2 (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT</td>
<td>4</td>
<td>277</td>
<td>0.72 (0.36 to 0.92)</td>
<td>0.99 (0.71 to 1.00)</td>
<td>119.98 (1.49 to 9655.66)</td>
<td>0.28 (0.10 to 0.85)</td>
</tr>
<tr>
<td>MRI</td>
<td>5</td>
<td>221</td>
<td>0.88 (0.64 to 0.97)</td>
<td>1.00 (0.38 to 1.00)</td>
<td>826.64 (0.51 to 1334596)</td>
<td>0.12 (0.03 to 0.42)</td>
</tr>
<tr>
<td>BS</td>
<td>6</td>
<td>543</td>
<td>0.99 (0.69 to 1.00)</td>
<td>0.86 (0.73 to 0.94)</td>
<td>7.35 (3.51 to 15.37)</td>
<td>0.01 (0.00 to 0.49)</td>
</tr>
</tbody>
</table>

Comparisons of the imaging tests

- Overall diagnostic accuracy (DOR) of BS was significantly higher compared with CT ($\chi^2 = 50.3$, df = 1, $P < 0.01$) and MRI ($\chi^2 = 29.7$, df = 1, $P < 0.01$).
- CT and MRI were comparable in diagnostic accuracy ($\chi^2 = 1.9$, df = 1, $P = 0.17$).
- The direct comparisons had similar patterns of differences in sensitivity and specificity to the indirect comparisons.

Conclusions

- The meta-analyses showed that DOR of BS is significantly better than CT ($P < 0.01$) and MRI ($P < 0.01$). This is based on a large difference in sensitivity. Conclusively, excellent of CT and MRI are much higher than for BS. CT and MRI have comparable diagnostic accuracy.
- Quality of included studies was moderate to good. There was only one report of formal quality assessment for BS compared to two for CT.

Direct comparison showed similar patterns of differences in sensitivity and specificity. Reflecting the small number of studies, the confidence intervals for summary estimates are wide for all three tests. There is a concern that number missed fractures on CT.


The median prevalence was 20%, calculated by using all studies. Missed fractures and overtreated patients were calculated using the median prevalence.
sensitivity and specificity of BS were 0.99 and 0.86, respectively. For CT, summary sensitivity and specificity were 0.72 and 0.99. For MRI, summary sensitivity and specificity were 0.88 and 1.00. Specificities of CT and MRI were both higher than BS. The single studies that directly compared CT and BS and MRI and BS found a similar pattern of the differences in sensitivity and specificity; however, both studies reported a lack of significant difference in the percentage of correct predictions. No differences were found between the diagnostic accuracies of CT and MRI. This finding applied also to the data from the two studies directly comparing CT and MRI. A summary of all results is presented in 'Summary of findings’.

Quality assessment showed moderate quality (six studies) to good quality (five studies). All patients were consecutive cohorts and at least eight (though the methodology suggests all) studies were explicitly prospective research. ‘Relevant clinical information’ was often not available during evaluation of index tests and is therefore a possible risk of bias. This should be included in future studies as omitting it is not representative of clinical practice. The other 11 items were mainly scored as ‘Yes’, implying good quality.

We could not find any information on which imaging technique is best for determining the location of the fracture (proximal, waist or distal). Some articles presented the location of a scaphoid fracture when presenting results for an index test; however, diagnostic accuracy calculations were not performed. In scaphoid fractures, healing is believed to be more problematic when fractures occur in the proximal part since blood supply is interrupted.

2.6.2 Strengths and weaknesses of the review

The evidence provided by this review is based on a comprehensive and sensitive literature search with the aim of identifying all relevant studies. All major electronic databases were searched and articles were selected with clear inclusion and exclusion criteria. Only studies with consecutive series of patients were included, which mimics clinical practice.
Another strength of this review is the usage of a well-regarded assessment tool to evaluate the quality of included studies: QUADAS. This tool provides detailed evaluation of quality and enables a simple and clear presentation of the assessment (Figure 2.2; Figure 2.3).

A key issue in diagnostic accuracy studies is the application of an adequate reference standard to test for true disease status. This issue is much debated in scaphoid literature and the lack of evidence and consensus on the right reference standard limits evaluation of diagnostic accuracy. Even though it is debated, follow-up radiographs at six weeks is generally considered to be the most suitable reference standard. The timing of visualisation of a lucent line on a radiograph is unknown but believed to be two to six weeks. This supports our choice to exclude reference standards that only consisted of repeated (radiographical) evaluation after one to two weeks, as this has been shown to be inadequate.

In our decision to pool data from studies, the similarity or equivalence in the criteria for test positivity is a critical issue. Thus the failure of some studies to report clear fracture criteria, which is a vital aspect for the interpretation of images, is clearly a problem. Where the criteria were not described, we considered the evaluation of images by two observers provided some assurance of an appropriate process. Since when reported, the criteria for CT, MRI and BS were sufficiently similar to merit pooling, we decided that it was a reasonable assumption that similar criteria would have been applied in all studies. Clearly, more precise criteria would be desirable in all future studies.

Another weakness of the review is the lack of direct comparison studies that include all three index tests. In addition, only four direct comparison studies including two index tests were evaluated. This means that comparison of CT, MRI and BS is mainly based on studies testing diagnostic accuracy of only one index test, i.e. indirect evidence. Another limitation of the review is that the findings derive from only a few studies. Therefore, sensitivity analyses could not be performed and potential sources for heterogeneity could not be investigated formally.
Our secondary objective for the review, accuracy of determining the location of the fracture, could not be answered and is therefore a weakness of the review. To date, we know of no studies that present these results.

A key limitation is the date of the search, July 2012; however, we are not aware of any new studies or current research on this topic.

2.6.3 Applicability of findings to the review question

The quality of the included studies was moderate to good and the data from these suggest that BS is the most sensitive modality to use in diagnosis of suspected scaphoid fractures. Direct comparison studies were few as indeed were the numbers of indirect comparative studies for each test. The low number of included studies for data analyses lowers the precision of the data. There are several other aspects that also need attention or additional research in order to determine the most suitable diagnostic method. The low prevalence of true scaphoid fractures among suspected fractures must be emphasised. The relatively low specificity of BS means that the number of over-treated patients would be much higher than with CT or MRI.

The effect of the low prevalence (20%) of true fractures among suspected scaphoid fractures is clearer when we apply the diagnostic accuracies in a cohort of 1000 patients (‘Summary of findings’). BS has a higher sensitivity and would lead to only 2 missed fractures in a cohort of 1000 patients, compared with 56 and 24 missed fractures by CT and MRI, respectively. The relatively low specificity of BS would result in unnecessary treatment of 112 patients, compared with only 8 over-treated patients when diagnosis is performed using CT and none when diagnosis is performed using MRI. Although we could not detect statistically significant differences between the specificities of all three modalities, the clinical impact of lower specificity combined with the low prevalence of a fracture is substantial. This shows the challenges in the diagnostic management of scaphoid fractures. A possible way to improve the diagnostic accuracy and lower the impact on clinical practice is by raising the prevalence of true fractures among suspected fractures. This can be
achieved by improving clinical evaluation or initial radiographic assessment, or both.

An interesting finding was the number of other fractures reported by all three imaging modalities. This review is focused on the scaphoid, but carpal and distal radius fractures were frequently found. The clinical significance for detecting these fractures is unknown, but does emphasise the questionable accuracy of current initial diagnostic methods.

Moreover, BS is the most invasive method to use with the intravenous application of radioactive isotopes and, compared with CT, gives a much higher dose of radiation. Therefore, BS is generally not recommended for children. BS also requires a delay of at least 72 hours to capture the osteoblastic activity at the fracture site and is therefore not applicable for instant diagnosis. Therefore, while BS might be the imaging modality with the highest sensitivity, it may not be the most suitable in practice.

2.7 Authors’ conclusions

2.7.1 Implications for practice

The diagnostic accuracy (DOR) of all three modalities studied in this review is considered good. However, we found evidence that BS has a significantly higher diagnostic accuracy than CT and MRI. In the meta-analysis, BS shows better sensitivity than CT and MRI. However, BS is also characterised by a lower specificity than either CT or MRI. The number of studies included is small and the confidence intervals for summary estimates are wide for all three tests. Even fewer studies directly compared index tests. This reduces the precision and generalisability of our results. The more invasive aspects of BS need also to be borne in mind. This test is less favourable compared with CT and MRI in terms of timing and safety due to a diagnostic delay of more than 72 hours and the intravenous administration of radioactive isotopes. It is debatable whether sensitivity or specificity is more important in this scenario. With the big impact of over-treatment due to the relatively low
specificity and with the invasive character of BS in mind, we would not recommend performing BS. CT and MRI both have good and comparable diagnostic accuracies, as shown in both meta-analyses and direct comparative studies. Given the data do not discriminate between the use of these tests, either of these tests can be used where available.

2.7.2 Implications for research

Prospective studies, perhaps involving randomisation of diagnostic tests, with direct comparisons of CT and MRI in the same patient population would add valuable data. We question the need for further research evaluating BS because of its limited use and invasive character. It would be useful if such studies incorporated economic (direct and indirect costs) and patient-related outcome measures (e.g. Disabilities of the Arm Shoulder and Hand, Patient Related Wrist Evaluation). Given the debate on the current best available reference standard (six week radiographs), consideration should be given to the practicalities of a check radiological follow-up, perhaps at one year, to examine for missed fractures. Prior to these, studies looking at ways to improve initial diagnostic management are needed to increase the identification of true scaphoid fractures among clinically suspected fractures.

Acknowledgements

We would like to thank Helen Handoll, the editors of the Cochrane Diagnostic Test Accuracy Working Group (particularly Petra Macaskill (review) and Gianni Virgili (protocol)), Peter Amadio, Lindsey Elstub, Laura MacDonald and Steven Rhemrev for their constructive feedback, comments and support in drafting the protocol and review. We also thank Joanne Elliott and the librarian of the VUMC of Amsterdam, Marijke Mol, for their support in developing the search strategies and running searches.

This project was supported by the National Institute for Health Research (NIHR)
via Cochrane Infrastructure funding to the Cochrane Bone, Joint and Muscle Trauma Group, and the Cochrane Diagnostic Test Accuracy Working Group. The views and opinions expressed herein are those of the review authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR or the UK National Health Service or Department of Health.
Appendices

Appendix 1. Diagnostic test performance characteristics

Sensitivity; specificity; accuracy

- Sensitivity: the proportion of patients who had a scaphoid fracture according to the reference standard and who were classified as having a positive test.
- Specificity: the proportion of patients who had no scaphoid fracture according to the reference standard and who were classified as having a negative test.
- Accuracy: the proportion of patients who were correctly classified by the test.

Positive predictive value (PPV); negative predictive value (NPV)

- PPV: the probability that a patient with a positive test has a scaphoid fracture.
- NPV: the probability that a patient with a negative test has no scaphoid fracture.

Appendix 2. Search strategies

MEDLINE (Ovid Web)

1 exp Magnetic Resonance Imaging/ (269663)
2 ((magnetic resonance or MR or NMR) adj2 (imag* or tomograph* or scan*)).tw. (149610)
3 (MRI or MRIs or NMRI).tw. (112187)
4 (diffusion weighted imag* or DWI or T2-weighted imag*).tw. (12506)
5 or/1-4 (321609)
6 exp Tomography, X-Ray Computed/ (267102)
7 (comput* adj3 tomograph*).tw. (154076)
8 (CT or CAT).tw. (246660)
9 micro-computed tomog*.tw. (1454)
10 or/6-9 (453788)
11 Radionuclide Imaging/ (24045)
12 (scintigra* or radioscinghra*).tw. (39873)
13 (bone adj3 scan).tw. (4431)
14 scintiscan*.tw. (1491)
15 or/11-14 (63106)
16 or/5,10,15 (746644)
17 Scaphoid bone/ or Wrist injuries/ or Wrist Joint/ (11542)
18 exp Fractures, Bone/ (129394)
19 and/17-18 (3618)
20 ((Scaphoid* or wrist or navicular) adj3 (fracture* or injur* or trauma)).tw. (2892)
21 or/19-20 (5363)
22 and/16,21 (761)
23 Diagnostic Imaging/ or Diagnosis, Differential/ or exp "Sensitivity and Specificity"/ or "Predictive Value of Tests"/ (723481)
24 scaphoid.mp. (3223)
25 and/23-24 (231)
26 Fractures, Bone/di, ra, ri [Diagnosis, Radiography, Radionuclide Imaging] (10875)
27 and/24,26 (586)
28 or/22,25,27 (1226)

EMBASE (Ovid Web)

1 exp Nuclear Magnetic Resonance Imaging/ (434436)
2 ((magnetic resonance or MR or NMR) adj2 (imag* or tomograph* or scan*)).tw. (188295)
3 (MRI or MRIs or NMRI).tw. (166239)
4 (diffusion weighted imag* or DWI or T2-weighted imag*).tw. (17259)
5 or/1-4 (476776)
6 exp Computer Assisted Tomography/ (512291)
7 (comput* adj3 tomograph*).tw. (193911)
8 (CT or CAT).tw. (336635)
9 micro-computed tomog*.tw. (1660)
10 or/6-9 (702231)
11 Bone Scintiscanning/ or Radiodiagnosis/ (50657)
12 (scintigra* or radioskintigra*).tw. (54696)
13 (bone adj3 scan).tw. (6365)
14 scintiscan*.tw. (2105)
15 or/11-14 (103388)
16 or/5,10,15 (1102315)
17 Scaphoid Fracture/ (1339)
18 Scaphoid Bone/ or Wrist Injury/ or Wrist/ or Wrist Radiography/ (20278)
19 exp Fracture/ (183932)
20 and/18-19 (3626)
21 ((scaphoid* or wrist or navicular) adj3 (fracture* or injur* or trauma)).tw. (3638)
22 or/17,20-21 (6434)
23 and/16,22 (1208)
24 "Sensitivity and Specificity"/ or Diagnostic Imaging/ or Receiver Operating Characteristic/ or Diagnostic Accuracy/ or Diagnostic Test/ or Diagnostic Value/ or Diagnostic Procedure/ (557240)
25 scaphoid.mp. (4476)
26 and/24-25 (382)
27 Fracture/di [Diagnosis] (3601)
28 and/25,27 (112)
29 Scaphoid Fracture/di [Diagnosis] (497)
30 or/23,26,28-29 (1586)
Cochrane Database of Abstracts of Reviews of Effects, Cochrane Central Register of Controlled Trials, NHS Economic Evaluation Database (Wiley Online Library)

1 MeSH descriptor Magnetic Resonance Imaging explode all trees 4459
2 ((magnetic resonance or MR or NMR) NEAR/2 (imag* or tomograph* or s-can*)):ti,ab,kw 5453
3 (MRI or MRIs or NMRI):ti,ab,kw 2723
4 (diffusion weighted imag* or DWI or T2-weighted imag*):ti,ab,kw 404
5 (#1 OR #2 OR #3 OR #4) 6382
6 MeSH descriptor Tomography, X-Ray Computed explode all trees 3159
7 (comput* NEAR/3 tomograph*):ti,ab,kw 6556
8 (CT or CAT):ti,ab,kw 19978
9 (micro-computed tomog*):ti,ab,kw 12
10 (#6 OR #7 OR #8 OR #9) 24123
11 MeSH descriptor Radionuclide Imaging, this term only 213
12 (scintigra* or radioscintigra*):ti,ab,kw 1550
13 (bone NEAR/3 scan):ti,ab,kw 177
14 (scintiscan*):ti,ab,kw 201
15 (#11 OR #12 OR #13 OR #14) 1933
16 (#5 OR #10 OR #15) 31013
17 MeSH descriptor Scaphoid Bone, this term only 41
18 MeSH descriptor Wrist Injuries, this term only 103
19 MeSH descriptor Wrist Joint, this term only 164
20 (#17 OR #18 OR #19) 284
21 MeSH descriptor Fractures, Bone explode all trees 3390
22 (#20 AND #21) 121
23 (Scaphoid* or wrist or navicular) NEAR/3 (fracture* or injur* or trauma):ti,ab,kw 226
24 (#22 OR #23) 243
25 (#16 AND #24) 31
26 MeSH descriptor Diagnostic Imaging, this term only 187
27 MeSH descriptor Diagnosis, Differential, this term only 1330
28 MeSH descriptor Sensitivity and Specificity explode all trees 13581
29 MeSH descriptor Predictive Value of Tests explode all trees 5043
30 (#26 OR #27 OR #28 OR #29) 14595
31 (scaphoid) 100
32 (#30 AND #31) 7
33 MeSH descriptor Fractures, Bone, this term only with qualifiers: DI,RA,RI 116
34 (#31 AND #33) 23
35 (#25 OR #32 OR #34) 34
36 (diagnos* and scaphoid and fractur*):ti,ab,kw in Trials 17
37 (#35 OR #36) 2 (DARE), 34 (CENTRAL), 8 (NHS EED)

**Other databases**

We searched the following databases for the term "scaphoid": MEDION (0 records), ARIF (3), WHO International Clinical Trials Registry Platform (13), and Current Controlled Trials (14).

**Contributions of authors**

All authors contributed to the development of the review and commented on and approved the final version. The guarantor of this review is Wouter Mallee.

**Declarations of interest**

Wouter H Mallee: was involved in the publication of a potentially eligible study. Therefore, this trial was assessed by other review authors.
Job N Doornberg: was involved in the publication of a potentially eligible study. Therefore, this trial was assessed by other review authors.

Rudolf W Poolman: Dr Poolman’s institution was supported for several research activities by several implant and drug companies, none of which are related to this review (for example, Link, AMGEN, AstraTech).

Peter Kloen: none known

Mario Maas: has provided consultancy to several companies (Genzyme, Shire, Protalix, Actelion) in an unrelated area of "Gaucher Disease and Imaging".

Henrica CW de Vet: none known

J Wang: none known

Sources of support

Internal sources

- Academic Medical Center of Amsterdam, Netherlands.
- Onze Lieve Vrouwe Gasthuis Amsterdam, Netherlands.
- Vrije Universiteit Medical Center of Amsterdam, Netherlands.
- Massachusetts General Hospital, USA.

External sources

- No sources of support supplied

Differences between protocol and review

To improve clarity, minor edits were made to Background and Methods.

In ’Assessment of methodological quality’, we added in a paragraph describing the key rules for rating overall methodological quality.
In ‘Data extraction and management’, we clarified on our usage of overlapping data where studies had been published more than once. We clarified that in cases of overlapping patient data, we only used the data once. To compare CT, MRI and BS, we used the HSROC model instead of the bivariate model. When we tried to use the bivariate model, the parameter estimates did not converge in SAS PROC NLMIXED. This might be caused by the small number of studies.

**Characteristics of studies**

**Characteristics of included studies [ordered by study ID]**

- ASB = Anatomical snuffbox
- PA = Posteroanterior
- Lat = Lateral
- ED = Emergency department
- BS = Bone Scintigraphy
- MRI = Magnetic Resonance Imaging
- CT = Computed Tomography
- FOOSH = Fall on outstretched hand
Clinical features and settings

**Inclusion criteria:** presentation to the ED within 48 hours after trauma. Clinically suspected for scaphoid fracture: tenderness ASB and painful ASB on longitudinal compression of thumb or index fingers.

**No fracture on initial radiographs in three views (PA, Lat, oblique with ulnar deviation)**

**Exclusion criteria:** polytrauma patients, patients under the age of 18 years, and those in whom MRI was contraindicated

**Participants**

- **Study location:** The Hague, The Netherlands
- **Study period:** March 2004 to January 2007
- **Participants enrolled:** 100; sex: 50 men and 50 women; mean age 42 years (range 18 to 84)
- **Participants included in analyses:** 100

**Study design**

- **Prospective, consecutive cohort**

**Target condition and reference standard(s)**

- **Target condition(s):** true scaphoid fracture among clinically suspected scaphoid fractures
- **Reference test:** MRI & BS (-) = no fracture; MRI & BS (+) = fracture. If discrepancy between MRI & BS, 6-week follow-up radiographs including physical examination

**Index and comparator tests**

- **Index test(s):** MRI and BS
  - **Magnetic resonance imaging**
    - **Timing:** performed within 24 hours after presentation to the ED
    - **Description of imaging technique (including patient position):** 1.5 Tesla with flexible surface coil around the wrist.
    - **Patient in prone position with the injured hand extended forward, palm down and overhead. T1 and T2 images were obtained. A slice thickness of 3 mm and distance factor of 10% was used
    - **Interpretation:** images were evaluated by a trainee and a consultant radiologist
  - **Criteria for a positive test:** not given
  - **Bone scintigraphy**
    - **Timing:** performed within 3 to 5 days after presentation to the ED
    - **Description of imaging technique (including patient position):** palmar and dorsal images of both wrists between 2.5 and 4 hours after injection of 500 MBq 99mTc-HDP
    - **Interpretation:** images were evaluated by a consultant clinical nuclear physician
    - **Criteria for a positive test:** osteoblast activity

**Follow-up**

- **None reported after final reference standard test was performed**

**Notes**

- **Prevalence of scaphoid fracture:** 20% (20/100)
- **Reporting of other fractures:** MRI reported 24 other fractures, BS reported 40 other fractures, reference standard reported 32 other fractures

**Table of Methodological Quality**

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative spectrum?</td>
<td>Yes</td>
<td>All tests</td>
</tr>
<tr>
<td>Acceptable reference standard?</td>
<td>Yes</td>
<td>All tests</td>
</tr>
<tr>
<td>Acceptable delay between tests?</td>
<td>Yes</td>
<td>All tests</td>
</tr>
<tr>
<td>Partial verification avoided?</td>
<td>Yes</td>
<td>All tests</td>
</tr>
<tr>
<td>Differential verification avoided?</td>
<td>No</td>
<td>All tests</td>
</tr>
<tr>
<td>Incorporation avoided?</td>
<td>No</td>
<td>All tests</td>
</tr>
<tr>
<td>Index test results blinded?</td>
<td>Yes</td>
<td>All tests</td>
</tr>
<tr>
<td>Reference standard results blinded?</td>
<td>Yes</td>
<td>All tests</td>
</tr>
<tr>
<td>Relevant clinical information?</td>
<td>No</td>
<td>All tests</td>
</tr>
<tr>
<td>Uninterpretable results reported?</td>
<td>Yes</td>
<td>All tests</td>
</tr>
<tr>
<td>Withdrawals explained?</td>
<td>Yes</td>
<td>All tests</td>
</tr>
<tr>
<td>Fracture criteria defined?</td>
<td>Unclear</td>
<td>All tests</td>
</tr>
</tbody>
</table>

**Prevalence of scaphoid fracture:** 20% (20/100)

**Reporting of other fractures:** MRI reported 24 other fractures, BS reported 40 other fractures, reference standard reported 32 other fractures
Breitenseher 1997

| Clinical features and settings | Inclusion criteria: presentation to the ED after acute trauma. Clinically suspected for scaphoid fracture: pain, swelling and tenderness of ASB during evaluation in the ED. No fracture on initial radiographs in 6 views (AP, 2 x Lat, PA with ulnar deviation, 2 x oblique). Exclusion criteria: fracture on initial radiographs. |
| Study design | Prospective, consecutive cohort. |
| Target condition and reference standard(s) | Target condition(s): true scaphoid fracture among clinically suspected scaphoid fractures. Reference test: 6-week follow-up radiographs compared with initial radiographs. |
| Index and comparator tests | Index test: MRI. Magnetic resonance imaging. Timing: performed within 7 days (mean 3.8 days) after trauma. Description of imaging technique (including patient position): 1 Tesla with circular surface coil. T1, STIR (both section thickness of 2.4 mm and intersection gap of 0.3 mm) and T2 (section thickness of 1.5 mm) images were performed. Patient position unknown. Interpretation: images were evaluated by two experienced radiologists. Criteria for a positive test: evidence of cortical or trabecular fracture line, diffuse area of bone marrow oedema or combination of all. |
| Follow-up | None reported after final reference standard test was performed. |
| Notes | Prevalence of scaphoid fracture: 33% (14/42). Reporting of other fractures: MRI reported 7 other fractures, reference standard reported none. |

<p>| Table of Methodological Quality |</p>
<table>
<thead>
<tr>
<th>Item</th>
<th>Authors’ judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative spectrum? All tests</td>
<td>Unclear</td>
<td>Unclear what the delay between wrist trauma and presentation to the ED was.</td>
</tr>
<tr>
<td>Acceptable reference standard? All tests</td>
<td>Yes</td>
<td>Reference standard was adequate according to the requirements mentioned in the protocol: 6-week follow-up radiographs.</td>
</tr>
<tr>
<td>Acceptable delay between tests? All tests</td>
<td>Yes</td>
<td>6 weeks for radiographs is an acceptable delay.</td>
</tr>
<tr>
<td>Initial verification avoided? All tests</td>
<td>Yes</td>
<td>All patients received the intended reference standard.</td>
</tr>
<tr>
<td>Differential verification avoided? All tests</td>
<td>Yes</td>
<td>All patients received the same reference standard: 6-week follow-up radiographs.</td>
</tr>
<tr>
<td>Incorporation avoided? All tests</td>
<td>Yes</td>
<td>The index test, MRI, was not part of the reference standard.</td>
</tr>
<tr>
<td>Index test results blinded? All tests</td>
<td>Yes</td>
<td>All images were separately evaluated in a blinded fashion.</td>
</tr>
<tr>
<td>Reference standard results blinded? All tests</td>
<td>Yes</td>
<td>All images were separately evaluated in a blinded fashion.</td>
</tr>
<tr>
<td>Relevant clinical information? All tests</td>
<td>Unclear</td>
<td>It was unclear if clinical information was present during evaluation of MRI.</td>
</tr>
<tr>
<td>Uninterpretable results reported? All tests</td>
<td>Yes</td>
<td>Some outcomes of index tests and reference standard were described, no uninterpretable test results were present.</td>
</tr>
<tr>
<td>Withdrawals explained? All tests</td>
<td>Yes</td>
<td>Consecutive series without loss to follow-up. All patients included in the study underwent both MRI and radiographs.</td>
</tr>
<tr>
<td>Fracture criteria defined? All tests</td>
<td>Yes</td>
<td>Clear fracture criteria for both radiographs and MRI were given.</td>
</tr>
</tbody>
</table>
Inclusion criteria: presentation to the ED within 48 hours after trauma. Clinically suspected for scaphoid fracture: tenderness ASB and painful ASB on longitudinal compression of thumb or index finger. No fracture on initial radiographs in 3 views (PA, Lat, oblique with ulnar deviation).

Exclusion criteria: polytrauma patients, patients younger than 18 years and those with contraindications for bone scintigraphy or CT.

Participants
Study location: The Hague, The Netherlands
Study period: November 2007 to January 2011
Participants enrolled: 159: 79 men and 80 women. Mean age 41 (range 17 to 88)
Participants included in analyses: 159

Study design
Prospective, consecutive cohort

Target condition and reference standard(s)
Target condition(s): true scaphoid fracture among clinically suspected scaphoid fractures
Reference test: CT & BS (-) = no fracture; CT & BS (+) = fracture. If discrepancy between CT & BS: 6-week follow-up radiographs including physical examination

Index and comparator tests
Index test: CT and BS
Computed tomography
Timing: performed within 24 hours after presentation to the ED
Description of imaging technique (including patient position): prone position with the hand extended forward palm down over the patient’s head with the wrist in neutral flexion and neutral radial-ulnar deviation.
Scout images were obtained to ensure that the scanning plane corresponded with the scans that provided a lateral view of the scaphoid bone as defined by the central longitudinal axis of the scaphoid.
Coronal plane images defined as images that provided a posteroanterior view of the scaphoid in the anatomic plane and in line with the axis of the scaphoid were obtained by supinating the forearm 90° keeping the wrist in a neutral position. Slice thickness was 0.625 mm with reconstructions every 0.4 mm.
For multiplanar reformatted images, parameters were 2 mm slice thickness, 2 mm interval
Interpretation: images were evaluated by a resident and a consultant radiologist
Criteria for a positive test: none given

Bone scintigraphy
Timing: performed within 3 to 5 days after presentation to the ED
Description of imaging technique (including patient position): palmar and dorsal images of both wrists were performed between 2.5 and 4 hours after the intravenous injection of 500 MBq of Tc-99m-HDP visualizing the osteoblastic activity with a planar collimator
Interpretation: images were evaluated by a consultant clinical nuclear physician
Criteria for a positive test: osteoblastic activity

Follow-up
None reported after final reference standard test was performed

Notes
Prevalence of scaphoid fracture: 13% (20/159) Reporting of other fractures: CT reported 35 other fractures, BS reported 57 other fractures, reference standard unknown.

Table of Methodological Quality

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative spectrum?</td>
<td>All tests</td>
</tr>
<tr>
<td>Acceptable reference standard?</td>
<td>All tests</td>
</tr>
<tr>
<td>Acceptable delay between tests?</td>
<td>All tests</td>
</tr>
<tr>
<td>Partial verification avoided?</td>
<td>All tests</td>
</tr>
<tr>
<td>Differential verification avoided?</td>
<td>All tests</td>
</tr>
<tr>
<td>Incorporation avoided?</td>
<td>All tests</td>
</tr>
<tr>
<td>Index test results blinded?</td>
<td>All tests</td>
</tr>
<tr>
<td>Reference standard results blinded?</td>
<td>All tests</td>
</tr>
<tr>
<td>Relevant clinical information?</td>
<td>All tests</td>
</tr>
<tr>
<td>Uninterpretable results reported?</td>
<td>All tests</td>
</tr>
<tr>
<td>Withdrawals explained?</td>
<td>All tests</td>
</tr>
<tr>
<td>Fracture criteria defined?</td>
<td>All tests</td>
</tr>
</tbody>
</table>

60
Inclusion criteria: presentation to the ED within 72 hours after trauma. Clinically suspected for scaphoid fracture: tenderness ASB and scaphoid’s tubercle.

No fracture on initial radiographs in 3 views (PA with ulnar deviation, Lat, oblique with wrist in semi-supination).

Exclusion criteria: Patients admitted more than 72 hours after the trauma or less than 18 years of age.

Participants
Study location: Ankara, Turkey
Study period: December 2007 to November 2008
Participants enrolled: 54 (55 wrists): all men. Mean age 22 (range 20 to 40).
Participants included in analyses: 55 (wrists)

Study design
Prospective, consecutive cohort

Target condition and reference standard(s)
Target condition(s): True scaphoid fracture among clinically suspected scaphoid fractures.
Reference test: MRI

Index and comparator tests
Index test: CT
Computed tomography
Timing: performed within 7 days after trauma
Description of imaging technique (including patient position): 64-detector multislice system. Prone position with the hand above the head and the wrist placed flat.
Interpretation: images were evaluated by two experienced radiologists
Criteria for a positive test: none given

Magnetic resonance imaging:
Timing: performed within 7 days after trauma
Description of imaging technique (including patient position): 1.5 Tesla with dedicated wrist coil. Prone position with the affected arm above the body.
Coronal and axial T1 images (3 to 5 mm slice thickness with a 0.5 mm gap); coronal and axial fat-saturated proton density-weighted images with fat saturation (3 to 5 mm slice thickness with 0.5 to 1.0 mm gap) and coronal T2 images (3 mm slice thickness) were obtained.
Interpretation: 2 radiologists
Criteria for a positive test: evidence of a cortical fracture line, a trabecular fracture line, or a combination of these abnormalities

Follow-up
None reported after final reference standard test was performed

Notes
Fracture prevalence: 29% (16/55)
Reporting of other fractures: MRI reported 6 other fractures, CT reported 5 other fractures

Table of Methodological Quality

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors’ judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Represented spectrum? All tests</td>
<td>Yes</td>
<td>Consecutive series of patients with acute wrist injury, clinical signs for a scaphoid fracture and normal initial radiographs</td>
</tr>
<tr>
<td>Acceptable reference standard? All tests</td>
<td>Unclear</td>
<td>Reference standard was one index test: MRI</td>
</tr>
<tr>
<td>Acceptable delay between tests? All tests</td>
<td>Yes</td>
<td>Both CT and MRI were performed within 1 week after wrist injury</td>
</tr>
<tr>
<td>Partial verification avoided? All tests</td>
<td>Yes</td>
<td>All patients received the intended reference standard</td>
</tr>
<tr>
<td>Differential verification avoided? All tests</td>
<td>Yes</td>
<td>All patients received the same reference standard: MRI</td>
</tr>
<tr>
<td>Incorporation avoided? All tests</td>
<td>Yes</td>
<td>The index test, CT, was not part of the reference standard</td>
</tr>
<tr>
<td>Index test results blinded? All tests</td>
<td>Yes</td>
<td>All images were separately evaluated in a blinded fashion</td>
</tr>
<tr>
<td>Reference standard results blinded? All tests</td>
<td>Yes</td>
<td>All images were separately evaluated in a blinded fashion</td>
</tr>
<tr>
<td>Relevant clinical information? All tests</td>
<td>No</td>
<td>Clinical information was not present during evaluation of the images. Observers were blinded from all other data</td>
</tr>
<tr>
<td>Uninterpretable results reported? All tests</td>
<td>Yes</td>
<td>All outcomes of index tests and reference standard were described, no uninterpretable test results were present</td>
</tr>
<tr>
<td>Withdrawals explained? All tests</td>
<td>Yes</td>
<td>Patients who did not return for MRI were reported including explanations</td>
</tr>
<tr>
<td>Fracture criteria defined? All tests</td>
<td>Unclear</td>
<td>Fracture criteria given for initial radiographs and MRI. However, these were not given for CT although it was well defined and evaluation was performed by 2 observers</td>
</tr>
</tbody>
</table>
Mallee 2011

Clinical features and settings

Inclusion criteria: presentation to the ED within 24 hours after trauma. Clinically suspected for scaphoid fracture: fall on outstretched hand, tenderness ASB.

Exclusion criteria: age of less than 18 years; any concurrent distal ulnar, radial, or carpal fracture; previous scaphoid fracture; rheumatoid arthritis and cognitive dysfunction that would limit physical examination.

Participants

Study location: Amsterdam, The Netherlands

Study period: April 2008 to October 2008

Participants enrolled: 40: 25 men and 15 women. Mean age 44.3 years (range 20 to 84)

Participants included in analyses: 34

Study design

Prospective, consecutive cohort

Target condition and reference standard(s)

Target condition(s): true scaphoid fracture among clinically suspected scaphoid fractures.

Reference test: 6-week follow-up radiographs compared with initial radiographs

Index and comparator tests

Index tests: MRI and CT

Magnetic resonance imaging

Timing: performed within 10 days (mean 3.6 days) after trauma

Description of imaging technique (including patient position): 1.0 Tesla with dedicated wrist coil. slice thickness of 3 mm and a 0.6 mm gap and included the following series:

a localizer image, a coronal slice of a short tau inversion recovery (STIR) sequence, and a coronal slice of a spin-echo T1-weighted sequence, in coronal views.

The patient was positioned supine, with the forearm and wrist alongside the body. The open MRI allowed for central placement of the hand relative to the magnetic field

Interpretation: evaluated by a consensus panel: a musculoskeletal radiologist, a trauma surgeon and an orthopaedic surgeon

Criteria for a positive test: presence of a cortical fracture line, a trabecular fracture line, or a combination of both. Extensive focal zone of oedema without a clear cortical fracture line, comparable with that seen with a stress fracture, was discussed to decide if the findings represented a fracture or not

Computed tomography

Timing: performed within 10 days (mean 3.6 days) after trauma

Description of imaging technique (including patient position): multidetector 64-slice, high-resolution 0.5 mm slice section thickness. Prone position with the affected arm above the body and palm down.

Reconstructions in planes defined by the long axis of the scaphoid were made

Interpretation: evaluated by a consensus panel: a musculoskeletal radiologist, a trauma surgeon and an orthopaedic surgeon

Criteria for a positive test: presence of a sharp lucent line within the trabecular bone pattern, a break in the continuity of the cortex, a sharp step in the cortex, or a dislocation of bone fragments

Follow-up

None reported after final reference standard test was performed

Notes

Fracture prevalence: 18% (6/34)

Reporting of other fractures: MRI reported 12 other fractures, CT reported 15 other fractures, reference standard reported 4 other fractures.

Table of Methodological Quality

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative spectrum?</td>
<td>All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Acceptable reference standard?</td>
<td>All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Acceptable delay between tests?</td>
<td>All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Partial verification avoided?</td>
<td>All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Differential verification avoided?</td>
<td>All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Incorporation avoided?</td>
<td>All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Index test results blinded?</td>
<td>All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Reference standard results blinded?</td>
<td>All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Relevant clinical information?</td>
<td>All tests</td>
<td>Unclear</td>
</tr>
<tr>
<td>Uninterpretable results reported?</td>
<td>All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Withdrawals explained?</td>
<td>All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Fracture criteria defined?</td>
<td>All tests</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Fracture criteria for radiographs, CT and MRI were well defined.
**Clinical features and settings**

Inclusion criteria: presentation to the ED within 72 hours. Clinically suspected for scaphoid fracture: pain over the scaphoid.

Exclusion criteria: none reported

**Participants**

Study location: Vienna, Austria

Study period: June 2000 to July 2002

Participants enrolled: 29: 17 men and 12 women. Mean age 34 years (range 17 to 62)

Participants included in analyses: 29

**Study design**

Prospective, consecutive cohort

**Target condition and reference standard(s)**

Target condition: True scaphoid fracture among clinically suspected scaphoid fractures

Reference test: 6-week follow-up radiographs

**Index and comparator tests**

Index tests: MRI and CT

MRI

Timing: performed on the same day within 6 days (mean 4.1 days) after trauma

Description of imaging technique (including patient position): 1.0 Tesla with appropriate receive-only surface coil.

Coronal and transverse STIR images (2.4 mm section thickness), coronal and transverse T1 images (2.4 mm section thickness), and coronal three-dimensional T2 images (1.5 mm section thickness) were made. Prone position with the affected arm above the body.

Interpretation: evaluated by a resident and an experienced musculoskeletal radiologist

Criteria for a positive test: presence of a cortical fracture line, a trabecular fracture line, or a combination of both

Computed tomography

Timing: performed on the same day within 6 days (mean 4.1 days) after trauma

Description of imaging technique (including patient position): four detector with detector configuration of two sections at 0.5 mm section thickness.

Multiplanar reformations of 0.7 mm thickness were calculated in coronal, sagittal, and transverse planes relative to the wrist. Prone position with the affected arm above the body.

Interpretation: evaluated by a resident and an experienced musculoskeletal radiologist

Criteria for a positive test: presence of a sharp lucent line within the trabecular bone pattern, a break in the continuity of the cortex, a sharp step in the cortex, or a dislocation of bone fragments

**Follow-up**

None reported after final reference standard test was performed

**Notes**

Fracture prevalence: 38% (11/29)

Reporting of other fractures: MRI reported 11 other fractures, CT reported 5 other fractures, reference standard reported 13 other fractures

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative spectrum? All tests</td>
<td>Yes</td>
<td>Consecutive series of patients with acute wrist injury, clinical signs for a scaphoid fracture and normal initial radiographs</td>
</tr>
<tr>
<td>Acceptable reference standard? All tests</td>
<td>Yes</td>
<td>Reference standard was adequate according to the requirements mentioned in the protocol: 6-week follow-up radiographs</td>
</tr>
<tr>
<td>Acceptable delay between tests? All tests</td>
<td>Yes</td>
<td>The interval between injury and reference test was acceptable: 6 weeks. CT and MRI were performed on the same day within 6 days</td>
</tr>
<tr>
<td>Partial verification avoided? All tests</td>
<td>Yes</td>
<td>All patients received the intended reference standard</td>
</tr>
<tr>
<td>Differential verification avoided? All tests</td>
<td>Yes</td>
<td>All patients received the same reference standard: 6-week follow-up radiographs</td>
</tr>
<tr>
<td>Interpretation avoided? All tests</td>
<td>Yes</td>
<td>Index tests, CT and MRI, were not part of the reference standard</td>
</tr>
<tr>
<td>Index test results blinded? All tests</td>
<td>Unclear</td>
<td>Not reported if observers were blinded for results of CT and MRI during evaluation of the 6-week follow-up radiographs</td>
</tr>
<tr>
<td>Reference standard results blinded? All tests</td>
<td>Yes</td>
<td>Evaluation of CT and MRI was performed without knowledge of the reference test results</td>
</tr>
<tr>
<td>Relevant clinical information? All tests</td>
<td>Yes</td>
<td>Site of clinical symptoms were known during evaluation of the CT and MRI</td>
</tr>
<tr>
<td>Uninterpretable results reported? All tests</td>
<td>Yes</td>
<td>All outcomes of index tests and reference standard were described; no uninterpretable test results were present</td>
</tr>
<tr>
<td>Withdrawals explained? All tests</td>
<td>Yes</td>
<td>All included patients were reported in the results without withdrawals</td>
</tr>
<tr>
<td>Fracture criteria defined? All tests</td>
<td>Yes</td>
<td>Fracture criteria for radiographs, CT and MRI were well defined</td>
</tr>
</tbody>
</table>
Inclusion criteria: clinically suspected for scaphoid fracture. No fracture, or inconclusive, on initial radiographs in 4 views (2x unknown, 2x oblique). Timing of presentation to the ED not reported.

Exclusion criteria: none reported.

Participants

Study location: Denmark, Randers.

Study period: 1980

Participants enrolled: 100 (101 wrists): 61 men and 39 women. Mean age 33 years (range 10 to 80)

Participants included in analyses: 101 (wrists)

Study design

Prospective, consecutive cohort

Target condition and reference standard(s)

Target condition(s): true scaphoid fracture among clinically suspected scaphoid fractures.

Reference test: 2-week and 2-month follow-up radiographs

Index and comparator tests

Index test: BS

Bone scintigraphy

Timing: performed within 10 days after trauma.

Description of imaging technique (including patient position): scanner was equipped with a parallel hole collimator, 3 hours after intravenous administration of 99m-Tc-MDP

Interpretation: not reported who evaluated the images

Criteria for a positive test: “negative”: homogeneous and symmetrical tracer uptake in both wrists; “positive”: focal uptake radially or centrally in the wrist; “inconclusive”: diffuse tracer uptake in the wrist or focal uptake in the ulnar part of the wrist

Follow-up

None reported after final reference standard test was performed

Notes

Fracture prevalence: 11% (11/101)

Reporting of other fractures: BS reported 43 other fractures, reference standard reported 9 other fractures

Table of Methodological Quality

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors’ judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative spectrum?</td>
<td>Unclear</td>
<td>Delay between wrist injury and presentation to the ED was not reported</td>
</tr>
<tr>
<td>Acceptable reference standard?</td>
<td>Yes</td>
<td>Reference standard was adequate according to the requirements mentioned in the protocol: both 2-week and 2-month follow-up radiographs</td>
</tr>
<tr>
<td>Acceptable delay between tests?</td>
<td>Yes</td>
<td>The interval between injury and reference test was acceptable: 2 months</td>
</tr>
<tr>
<td>Partial verification avoided?</td>
<td>Yes</td>
<td>All patients received the intended reference standard</td>
</tr>
<tr>
<td>Differential verification avoided?</td>
<td>Yes</td>
<td>All patients received the same reference standard: 2-week and 2-month follow-up radiographs</td>
</tr>
<tr>
<td>Incorporation avoided?</td>
<td>Yes</td>
<td>Index test, BS, was not part of the reference standard</td>
</tr>
<tr>
<td>Index test results blinded?</td>
<td>Unclear</td>
<td>It was not reported if BS results were present during evaluation of the follow-up radiographs</td>
</tr>
<tr>
<td>Reference standard results blinded?</td>
<td>Yes</td>
<td>BS was already evaluated prior to the follow-up radiographs</td>
</tr>
<tr>
<td>Relevant clinical information?</td>
<td>No</td>
<td>Clinical and radiographic information was not present during evaluation of BS</td>
</tr>
<tr>
<td>Uninterpretable results reported?</td>
<td>Yes</td>
<td>All outcomes of index tests and reference standard were described; inconclusive test results were present and reported</td>
</tr>
<tr>
<td>Withdrawals explained?</td>
<td>Yes</td>
<td>All included patients underwent the entire study protocol</td>
</tr>
<tr>
<td>Fracture criteria defined?</td>
<td>Yes</td>
<td>Fracture criteria for BS were defined</td>
</tr>
</tbody>
</table>
O'Carroll 1982

Clinical features and settings
Inclusion criteria: timing of presentation to the ED not reported. Clinically suspected for scaphoid fracture. No fracture on initial radiographs, number of views unknown (AP, Lat, oblique).
Exclusion criteria: none reported.

Participants
Study location: Dublin, Ireland
Study period: not reported
Participants enrolled: 30: 21 men/9 women. Mean age 32 (range 11-72)
Participants included in analyses: 30

Study design
Prospective, consecutive cohort

Target condition and reference standard(s)
Target condition(s): true scaphoid fracture among clinically suspected scaphoid fractures.
Reference test: 2-week, 4-week and 6-week follow-up radiographs

Index and comparator tests
Index test: BS
Bone scintigraphy
Timing: performed within 10 days (range 3 to 32 days) after trauma.
Description of imaging technique (including patient position): Tc-99m-HDP and a large field of view Gamma camera. Ventral views of both wrists were obtained 3 hours following intravenous injection of the radiopharmaceutical and 500,000 counts were obtained in each picture.
Interpretation: not reported who evaluated the images
Criteria for a positive test: none reported

Follow-up
None reported after final reference standard test was performed

Notes
Fracture prevalence: 20% (6/30)
Reporting of other fractures: none reported.

Table of Methodological Quality

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative spectrum? All tests</td>
<td>Unclear</td>
<td>Delay between wrist injury and presentation to ED was not reported</td>
</tr>
<tr>
<td>Acceptable reference standard? All tests</td>
<td>Yes</td>
<td>Reference standard was adequate according to the requirements mentioned in the protocol: 2-, 4-, 6-week follow-up radiographs</td>
</tr>
<tr>
<td>Acceptable delay between tests? All tests</td>
<td>Unclear</td>
<td>BS was performed up to 32 days after wrist injury</td>
</tr>
<tr>
<td>Partial verification avoided? All tests</td>
<td>Yes</td>
<td>All patients received the intended reference standard</td>
</tr>
<tr>
<td>Differential verification avoided? All tests</td>
<td>Yes</td>
<td>All patients received the same reference standard: 2-, 4-, 6-week follow-up radiographs</td>
</tr>
<tr>
<td>Incorporation avoided? All tests</td>
<td>Yes</td>
<td>Index test, BS, was not part of the reference standard</td>
</tr>
<tr>
<td>Index test results blinded? All tests</td>
<td>Unclear</td>
<td>Not reported if result of BS was present during evaluation of radiographs</td>
</tr>
<tr>
<td>Reference standard results blinded? All tests</td>
<td>Unclear</td>
<td>Not reported if reference test results were present during evaluation of BS</td>
</tr>
<tr>
<td>Relevant clinical information? All tests</td>
<td>Unclear</td>
<td>Not reported if clinical information was present during evaluation of the images</td>
</tr>
<tr>
<td>Uninterpretable results reported? All tests</td>
<td>Yes</td>
<td>All outcomes of index tests and reference standard were described. Inclusive test results were not presented</td>
</tr>
<tr>
<td>Withdrawals explained? All tests</td>
<td>Yes</td>
<td>All included patients received both index test and reference standard, no withdrawals</td>
</tr>
<tr>
<td>Fracture criteria defined? All tests</td>
<td>No</td>
<td>Fracture criteria for BS were not well defined</td>
</tr>
</tbody>
</table>
Stordahl 1984

Clinical features and settings

Inclusion criteria: timing of presentation to the ED not reported. Clinically suspected for scaphoid fracture. No fracture, or inconclusive, on initial radiographs in 4 views (PA, Lat, 2x oblique).

Exclusion criteria: none reported.

Participants

Study location: Kristiansand, Norway

Study period: not reported

Participants enrolled: 30: 18 men and 12 women. Mean age 31 years (range 10 to 69)

Participants included in analyses: 28 (2 patients were excluded because, upon "revision of the material", their fractures were "already evident on the initial x-rays")

Study design

Not reported, consecutive

Target condition and reference standard(s)

Target condition(s): true scaphoid fracture among clinically suspected scaphoid fractures

Reference test: 2-week, 4-week and 6-week follow-up radiographs

Index and comparator tests

Index test: BS

Bone scintigraphy

Timing: performed 2 weeks after trauma. Description of imaging technique (including patient position): radionuclide imaging was performed three hours after the administration of Tc-99m-HDP. A Pho/Gamma 4 Camera with divergent low energy collimator, or pinhole collimator was used

Interpretation: two examiners of unknown expertise evaluated the images

Criteria for a positive test: increased radioactivity, focal or diffuse

Follow-up

None reported after final reference standard test was performed

Notes

Fracture prevalence: 32% (9/28)

Reporting of other fractures: BS reported 3 other fractures,

Table of Methodological Quality

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative spectrum? All tests</td>
<td>No</td>
</tr>
<tr>
<td>Delay between wrist injury and presentation was not reported and inconclusive radiographs were included</td>
<td></td>
</tr>
<tr>
<td>Acceptable reference standard? All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Reference standard was adequate according to the requirements mentioned in the protocol: 2-, 4-, 6-week follow-up radiographs</td>
<td></td>
</tr>
<tr>
<td>Acceptable delay between tests? All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>The interval between injury and reference test was acceptable: 6 weeks. BS was performed within 2 weeks</td>
<td></td>
</tr>
<tr>
<td>Partial verification avoided? All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>All patients received the intended reference standard</td>
<td></td>
</tr>
<tr>
<td>Differential verification avoided? All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>All patients received the same reference standard: 2-, 4-, 6-week follow-up radiographs</td>
<td></td>
</tr>
<tr>
<td>Incorporation avoided? All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Index test, BS, was not part of the reference standard</td>
<td></td>
</tr>
<tr>
<td>Index test results blinded? All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>All images were separately evaluated in a blinded fashion</td>
<td></td>
</tr>
<tr>
<td>Reference standard results blinded? All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>All images were separately evaluated in a blinded fashion</td>
<td></td>
</tr>
<tr>
<td>Relevant clinical information? All tests</td>
<td>Unclear</td>
</tr>
<tr>
<td>Not well reported if clinical findings were present during evaluation of radiographs and BS</td>
<td></td>
</tr>
<tr>
<td>Uninterpretable results reported? All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>All outcomes of index tests and reference standard were described; no uninterpretable test results were present</td>
<td></td>
</tr>
<tr>
<td>Withdrawals explained? All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>All included patients received both index test and reference standard. 2 patients were excluded since they had fractures on initial radiographs</td>
<td></td>
</tr>
<tr>
<td>Fracture criteria defined? All tests</td>
<td>Yes</td>
</tr>
<tr>
<td>Fracture criteria for BS were well defined</td>
<td></td>
</tr>
</tbody>
</table>
Clinical features and settings

Inclusion criteria: presentation to ED within 24 hours. Clinically suspected for scaphoid fracture: FOOSH, tenderness in ASB. No fracture initial radiographs in four views (PA with ulnar deviation, Lat, 2x oblique).

Exclusion criteria: none reported.

Participants

Study location: Amsterdam, The Netherlands

Study period: September 1987 to September 1990

Participants enrolled: 160: 82 men and 78 women. Mean age 38.6 years (range 12 to 84)

Participants included in analyses: 125

Study design

Not reported, consecutive

Target condition and reference standard(s)

Target condition(s): true scaphoid fracture among clinically suspected scaphoid fractures.

Reference test: 2-week and 6-week follow-up radiographs

Index and comparator tests

Index test: BS

Bone scintigraphy

Timing: performed 3 to 34 days (mean 12.3) after injury

Description of imaging technique (including patient position): after administration of 200 MBq Tc-99m-HDP, dynamic phase (after 2 to 5 minutes) and static phase (after 2 to 3 hours) were recorded.

Anterior and lateral static views were obtained

Interpretation: evaluated by an experienced nuclear physician. All radiographs were judged blindly by a panel consisting of a bone radiologist, a general radiologist and a senior traumatologist

Criteria for a positive test: presence of focally increased activity in both dynamic and static phase

Follow-up

Additional follow-up after a minimum of 1 year

Notes

Fracture prevalence: 17% (21/125)

Reporting of other fractures: BS reported 49 other fractures, reference standard reported 24 other fractures

Table of Methodological Quality

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors’ judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative spectrum? All tests</td>
<td>Yes</td>
<td>Consecutive series of patients with acute wrist injury, clinical signs for a scaphoid fracture and normal initial radiographs</td>
</tr>
<tr>
<td>Acceptable reference standard? All tests</td>
<td>Yes</td>
<td>Reference standard was adequate according to the requirements mentioned in the protocol: 2- and 6-week follow-up radiographs</td>
</tr>
<tr>
<td>Acceptable delay between tests? All tests</td>
<td>Yes</td>
<td>The interval between injury and reference test was acceptable: 6 weeks. BS was performed within 2 weeks</td>
</tr>
<tr>
<td>Partial verification avoided? All tests</td>
<td>Yes</td>
<td>All patients received the intended reference standard</td>
</tr>
<tr>
<td>Differential verification avoided? All tests</td>
<td>Yes</td>
<td>All patients received the same reference standard: 2- and 6-week follow-up radiographs</td>
</tr>
<tr>
<td>Incorporation avoided? All tests</td>
<td>Yes</td>
<td>Index test, BS, was not part of the reference standard</td>
</tr>
<tr>
<td>Index test results blinded? All tests</td>
<td>Yes</td>
<td>All images were separately evaluated in a blinded fashion</td>
</tr>
<tr>
<td>Reference standard results blinded? All tests</td>
<td>Yes</td>
<td>BS was performed and evaluated prior to the reference standard</td>
</tr>
<tr>
<td>Relevant clinical information? All tests</td>
<td>Unclear</td>
<td>Not reported if clinical information was present during evaluation of BS</td>
</tr>
<tr>
<td>Uninterpretable results reported? All tests</td>
<td>Yes</td>
<td>All outcomes of index tests and reference standard were described, no uninterpretable test results were present</td>
</tr>
<tr>
<td>Withdrawals explained? All tests</td>
<td>Yes</td>
<td>Withdrawals were reported including explanation</td>
</tr>
<tr>
<td>Fracture criteria defined? All tests</td>
<td>Yes</td>
<td>Fracture criteria for BS were well-defined</td>
</tr>
</tbody>
</table>
Clinical features and settings

Inclusion criteria: presentation to ED within 72 hours after trauma. Clinically suspected for scaphoid fracture. No fracture initial radiographs in four views (PA, oblique with ulnar deviation, Lat, PA with tilted beam).

Exclusion criteria: claustrophobia

Participants

Study location: Amsterdam, The Netherlands.
Study period: not reported
Participants enrolled: 16: 11 men and 5 women. Mean age 36 years (range 24 to 60)
Participants included in analyses: 16

Study design

Not reported, consecutive

Target condition and reference standard(s)

Target condition(s): true scaphoid fracture among clinically suspected scaphoid fractures
Reference test: BS performed in 3 to 14 days (mean 10)

Index and comparator tests

Index test: MRI and BS

Magnetic resonance imaging
Timing: performed in 3 to 14 days (mean 10)
Description of imaging technique (including patient position): 1.5 Tesla with a circular surface coil. Prone position with arm extended overhead with wrist fixed in neutral position. T1 and T2 images were obtained with 3.0 mm slice thickness
Interpretation: images were evaluated by an experienced radiologist

Criteria for a positive test: none reported

Bone scintigraphy
Timing: performed in 3 to 14 days (mean 10)
Description of imaging technique (including patient position): after administration of 200 MBq Tc-99m-HDP, dynamic phase and static phase were recorded. Anterior and lateral static views were obtained
Interpretation: all carpal radiographs, MR images and bone scans were reviewed by a panel of the same radiologists and nuclear physician
Criteria for a positive test: presence of focally increased activity in both dynamic and static phase

Follow-up

None reported after final reference standard test was performed

Notes

Fracture prevalence: 44% (7/16)
Reporting of other fractures: BS reported one lunate fracture

Table of Methodological Quality

<table>
<thead>
<tr>
<th>Item</th>
<th>Authors' judgement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative spectrum? All tests</td>
<td>Yes</td>
<td>Consecutive series of patients with acute wrist injury, clinical signs for a scaphoid fracture and normal initial radiographs</td>
</tr>
<tr>
<td>Acceptable reference standard? All tests</td>
<td>Unclear</td>
<td>Reference standard was one index test: BS</td>
</tr>
<tr>
<td>Acceptable delay between tests? All tests</td>
<td>Yes</td>
<td>The interval between injury and reference test was acceptable: both tests were performed within 14 days</td>
</tr>
<tr>
<td>Partial verification avoided? All tests</td>
<td>Yes</td>
<td>All patients received the intended reference standard</td>
</tr>
<tr>
<td>Differential verification avoided? All tests</td>
<td>Yes</td>
<td>All patients received the same reference standard: BS</td>
</tr>
<tr>
<td>Incorporation avoided? All tests</td>
<td>Yes</td>
<td>Index test, MRI, was not part of the reference standard</td>
</tr>
<tr>
<td>Index test results blinded? All tests</td>
<td>Yes</td>
<td>All images were separately evaluated in a blinded fashion</td>
</tr>
<tr>
<td>Reference standard results blinded? All tests</td>
<td>Yes</td>
<td>All images were separately evaluated in a blinded fashion</td>
</tr>
<tr>
<td>Relevant clinical information? All tests</td>
<td>No</td>
<td>BS was evaluated blinded from other data</td>
</tr>
<tr>
<td>Uninterpretable results reported? All tests</td>
<td>Yes</td>
<td>All outcomes of index tests and reference standard were described, no uninterpretable test results were present</td>
</tr>
<tr>
<td>Withdrawals explained? All tests</td>
<td>Yes</td>
<td>Withdrawals were reported including explanation</td>
</tr>
<tr>
<td>Fracture criteria defined? All tests</td>
<td>Unclear</td>
<td>Fracture criteria for BS were well defined but those for MRI were lacking</td>
</tr>
</tbody>
</table>
### Characteristics of excluded studies [ordered by study ID]

<table>
<thead>
<tr>
<th>Study</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adey 2007</td>
<td>The series of patients was not consecutive</td>
</tr>
<tr>
<td>Akdemir 2004</td>
<td>Patients were only included if they were still clinically suspected to have a scaphoid fracture after 1 to 2 weeks</td>
</tr>
<tr>
<td>Asche 1982</td>
<td>Inadequate reference standard</td>
</tr>
<tr>
<td>Beeres 2005</td>
<td>No reference standard</td>
</tr>
<tr>
<td>Beeres 2007</td>
<td>Inadequate reference standard</td>
</tr>
<tr>
<td>Breederveld 2004</td>
<td>Patients were only included if they were still clinically suspected after 1 to 2 weeks</td>
</tr>
<tr>
<td>Bretlau 1999</td>
<td>Patients were only included if they were still clinically suspected after 1 to 2 weeks</td>
</tr>
<tr>
<td>Brismar 1988</td>
<td>Inadequate reference standard</td>
</tr>
<tr>
<td>Bury 2010</td>
<td>Letter to editor</td>
</tr>
<tr>
<td>Cook 1997</td>
<td>Inadequate patient selection since patients were not excluded when they had a fracture on initial radiographs</td>
</tr>
<tr>
<td>Coupland 1996</td>
<td>Inadequate reference standard</td>
</tr>
<tr>
<td>Cruickshank 2007</td>
<td>Inadequate reference standard</td>
</tr>
<tr>
<td>Fowler 1998</td>
<td>Patients were only included if they were still clinically suspected after 1 to 2 weeks</td>
</tr>
<tr>
<td>Fusetti 2008</td>
<td>Letter to editor</td>
</tr>
<tr>
<td>Ganel 1979</td>
<td>No reference standard</td>
</tr>
<tr>
<td>Geijer 2011</td>
<td>No index test</td>
</tr>
<tr>
<td>G?bler 2001</td>
<td>Patients were only included if they were still clinically suspected after 1 to 2 weeks</td>
</tr>
<tr>
<td>Hobby 2001</td>
<td>Review</td>
</tr>
<tr>
<td>Hunter 1997</td>
<td>Inadequate reference standard</td>
</tr>
<tr>
<td>Jenkins 2008</td>
<td>No reference standard</td>
</tr>
<tr>
<td>Jonsson 1992</td>
<td>Patients were only included if they were still clinically suspected after 7 to 18 days</td>
</tr>
<tr>
<td>Keim 2003</td>
<td>No reference standard</td>
</tr>
<tr>
<td>Klein 2002</td>
<td>Inadequate patient selection since all patients with wrist pain were included; this study did not focus on the diagnostic test accuracy for clinically suspected scaphoid fractures</td>
</tr>
<tr>
<td>Kristensen 1983</td>
<td>Inadequate reference standard</td>
</tr>
<tr>
<td>Kumar 2005</td>
<td>No reference standard</td>
</tr>
<tr>
<td>Kusano 2002</td>
<td>No reference standard</td>
</tr>
<tr>
<td>Lanik 1986</td>
<td>Inadequate reference standard</td>
</tr>
<tr>
<td>Lapa 2000</td>
<td>Case report</td>
</tr>
<tr>
<td>Lepage 2004</td>
<td>No reference standard</td>
</tr>
<tr>
<td>Lindequist 1998</td>
<td>No reference standard</td>
</tr>
<tr>
<td>Low 2005</td>
<td>The series of patients was not consecutive</td>
</tr>
<tr>
<td>Lozano-Calderon 2006</td>
<td>Inadequate patient selection since a convenience sample was used</td>
</tr>
<tr>
<td>Moller 2004</td>
<td>No reference standard</td>
</tr>
<tr>
<td>Murphy 1995</td>
<td>Patients were only included if they were still clinically suspected after 1 to 2 weeks</td>
</tr>
<tr>
<td>Querellou 2009</td>
<td>Not focused on scaphoid fractures</td>
</tr>
<tr>
<td>Raby 2001</td>
<td>No reference standard</td>
</tr>
<tr>
<td>Schick 1999</td>
<td>Not focused on scaphoid fractures</td>
</tr>
<tr>
<td>Tennoe 2000</td>
<td>Letter to editor</td>
</tr>
<tr>
<td>Thorpe 1996</td>
<td>Patients were only included if they were still clinically suspected after 1 to 2 weeks</td>
</tr>
<tr>
<td>Tiel-van Buul 1993b</td>
<td>Inadequate reference standard</td>
</tr>
<tr>
<td>Tiel-van Buul 1995</td>
<td>Letter to editor</td>
</tr>
<tr>
<td>Tiel-van Buul 1998</td>
<td>Letter to editor</td>
</tr>
<tr>
<td>Vrettos 1996</td>
<td>Inadequate reference standard</td>
</tr>
<tr>
<td>Wilson 1986</td>
<td>Inadequate reference standard</td>
</tr>
<tr>
<td>You 2007</td>
<td>No reference standard</td>
</tr>
</tbody>
</table>
Data

Presented below are all the data for all of the tests entered into the review.

Table 2.5: Data tables by test

<table>
<thead>
<tr>
<th>Test</th>
<th>No. of studies</th>
<th>No. of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CT</td>
<td>4</td>
<td>277</td>
</tr>
<tr>
<td>2 MRI</td>
<td>5</td>
<td>221</td>
</tr>
<tr>
<td>3 BS</td>
<td>6</td>
<td>543</td>
</tr>
</tbody>
</table>

Test 1. CT

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CT</td>
<td>4</td>
<td>3</td>
<td>20</td>
<td>0.73 (0.43, 0.89)</td>
<td>0.59 (0.52, 0.65)</td>
</tr>
<tr>
<td>2 MRI</td>
<td>5</td>
<td>1</td>
<td>20</td>
<td>0.67 (0.42, 0.88)</td>
<td>0.50 (0.42, 0.60)</td>
</tr>
</tbody>
</table>

Test 2. MRI

<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 BS</td>
<td>6</td>
<td>3</td>
<td>21</td>
<td>0.71 (0.56, 0.84)</td>
<td>0.60 (0.51, 0.69)</td>
</tr>
</tbody>
</table>

Test 3. BS

70