MRI as a reference standard for suspected scaphoid fractures

[1]A D DE ZWART, MS, [1,4]F J P BEERES, MD, PhD, [3]D RING, MD, PhD, [5]L M KINGMA, MD, PhD, [2]E G COERKAMP, MD, [1]S A G MEYLAERTS, MD, PhD and [1]S J RHEMREV, MD

[1]Department of Surgery, Medical Center Haaglanden, The Hague, Netherlands, [2]Department of Radiology, Medical Centre Haaglanden, The Hague, Netherlands, [3]Department of Orthopaedic Surgery, Massachusetts General Hospital, Boston, MA, USA, [4]Department of Surgery Leiden University Medical Centre, Leiden, Netherlands, and [5]Department of Radiology, University Medical Centre Groningen, Groningen, Netherlands

Objectives: Some have suggested that MRI might be the best reference standard for a true fracture among patients with suspected scaphoid fractures. The primary aim of this study was to determine the rate of false-positive diagnosis of an acute scaphoid fracture in a cohort of healthy volunteers.
Methods: In a prospective study, 33 healthy volunteers were recruited and both wrists of each were scanned, except for 2 volunteers for whom only one wrist was scanned. To simulate the usual clinical context the 64 scans of healthy volunteers were mixed with 60 MRI scans of clinically suspected scaphoid fractures but normal scaphoid radiographs. These 124 MRI scans were blinded and randomly ordered. Five radiologists evaluated the MRI scans independently for the presence or absence of a scaphoid fracture and other injuries according to a standard protocol.
Results: To answer the primary question, only the diagnoses from the 64 scans of healthy volunteers were used. The radiologists diagnosed a total of 13 scaphoid fractures; therefore, specificity for diagnosis of scaphoid fracture was 96% (95% confidence interval: range 94–98%). The 5 observers had a moderate interobserver agreement regarding diagnosis of scaphoid fracture in healthy volunteers (multirater $\kappa=0.44$; $p<0.001$).
Conclusions: The specificity of MRI for scaphoid fractures is high (96%), but false-positives do occur. Radiologists have only moderate agreement when interpreting MRI scans from healthy volunteers. MRI is not an adequate reference standard for true fractures among patients with suspected scaphoid fractures.

The American College of Radiologists recommends MRI for diagnosis of true fractures among suspected scaphoid fractures [1]. A number of published studies cite sensitivities and specificities approaching 99% [2–9], but other studies have reported a lower sensitivity (80%) and substantial interobserver variation ($\kappa=0.67$) for diagnosis of a scaphoid fracture [2, 8]. It has been difficult to agree upon a reliable reference standard for true fractures among suspected scaphoid fractures, and these studies often use only repeated radiographs 6 weeks after trauma as reference standard. However, it is also known that not all occult scaphoid fractures become apparent on repeated radiographs.

It is not clear how to distinguish true fracture from other changes in bone signal that are detected with MRI. We propose that MRI of the wrists of healthy volunteers with no history of wrist or hand injury represents a reliable reference standard for the absence of an acute fracture of the scaphoid waist. By evaluating MRI scans of healthy volunteers, we may learn more about the diagnostic performance characteristics of MRI for suspected scaphoid fracture. A set of MRI scans with a reliable reference standard would also provide useful information about the reproducibility of the interpretation of MRI for suspected scaphoid fracture.

The primary study question was to investigate the occurrence of false-positive diagnosis of an acute scaphoid fracture on MRI using a reliable reference standard (healthy volunteers). Secondarily we also investigated the interobserver variation of diagnosis of scaphoid fracture on MRI in healthy volunteers.

Methods and materials

This is a prospective cohort study approved by our institutional review board.

Healthy volunteers

Healthy volunteers were recruited from acquaintances of the main study investigators. Before inclusion one of the investigators assessed an interview concerning whether there was:

(1) any history of wrist or hand injury
(2) any history of wrist pain or arthritis
(3) any contraindication for an MRI scan.

If any of these questions were positive the volunteer was excluded. All healthy volunteers were aware of the primary aim of the study and gave oral informed

Address correspondence to: Mr Andele de Zwart, Department of Surgery, Medical Centre Haaglanden, Postbus 432, 2501 CK, the Hague, Netherlands. E-mail: andeledezwart@gmail.com

consent. They volunteered and there was no form of compensation.

62 MRI scans were made of both wrists of 31 healthy volunteers. Two healthy volunteers had an MRI scan of one wrist because they had a history of wrist trauma on one side. There were 44 MRI scans of males and 20 MRI scans of females, with a mean age of 28 years (range 19–53 years).

Suspected scaphoid fractures

To simulate the clinical context, we mixed the MRI scans of the healthy volunteers with MRI scans from a cohort of patients with a suspected scaphoid fracture from a database of a previous study [2]. All patients with a suspected scaphoid fracture had a recent trauma, a tender anatomical snuffbox and pain when applying axial pressure on the thumb or index finger. Of the group of 60 scans of patients with suspected scaphoid fractures, there were 32 males and 28 females with a mean age of 38 years (range 16–70 years). According to the reference standard used in this study, 15 patients were diagnosed with a scaphoid fracture, 20 with other fractures, 2 with contusions of os triquetrum, 3 with contusion of os scaphoid and 20 without any injury [2]. This study was performed in the same clinic with an identical MRI protocol to our current study, but with the participation of different radiologists.

In total there were 124 MRI scans, of which 64 were from healthy volunteers and 60 from the used cohort study.

Evaluation

All MRI scans were stripped of patient identifiers, and presented in random order to five radiologists that specialise in musculoskeletal radiology. Five radiologists participated in the study. Radiologist 1 was a resident in the USA with two years of experience. Radiologists 2 and 3 were fellows in skeletal radiology in the USA. Radiologists 4 and 5 were established radiologists in the Netherlands with 14 years and 18 years of experience, respectively.

The radiologists were aware that the MRI scans were obtained from both healthy volunteers and patients with suspected scaphoid fracture; however, they were not aware about the size of each group. Each radiologist independently rated each MRI scan using a standardised scoring sheet containing the following four items:

(1) scaphoid fracture (yes/no)
(2) other fracture (yes/no; if yes, metacarpal, other carpal or distal radius fracture)
(3) other lesions
(4) no injury (yes/no).

There were no specific criteria for diagnosis of a fracture determined by the investigators before the start of the study.

MRI protocol

A 1.5 T MRI scanner (Symphony^{TM}; Siemens, Erlangen, Germany) was used. The patient lies prone on the scanner table with the hand extended forward, palm down, over the patient's head. A flexible surface coil was wrapped around the wrist. The MRI protocol included coronal T_1 weighted turbo spin-echo images with a repetition time (TR) of 450 ms, an echo time (TE) of 13 ms, a field of view of 180×115.2 mm (64%), a base resolution of 512, two averages, a slice thickness of 3 mm with a distance factor of 10% and a scan time of 2.17 min. The parameters for the coronal fat-suppressed T_2 weighted fast spin-echo images were a TR of 5220 ms, a TE of 73 ms, a field of view of 220×141.46 mm (64.3%), a base resolution of 448, 3 averages, a slice thickness of 3 mm with a distance factor of 10%, and a scan time of 4.33 min. Both wrists of the healthy volunteers were scanned.

Statistical methods

For data analysis we excluded 60 scans of the suspected scaphoid fractures as this was not the main goal of this study.

Statistics are based on the 64 scans of the healthy volunteers. As one of the radiologists did not rate one of the healthy scans, the data set contained 319 diagnoses/ratings (64 healthy scans rated by 4 radiologists and 63 scans rated by 1 radiologist).

Among the healthy volunteers, we calculated the specificity (the proportion of healthy volunteers correctly diagnosed as having no scaphoid fracture) based on a binomial-based robust estimator (method of Clopper and Pearson). To get a robust estimate of specificity with 95% confidence interval, we used repeated-measures terms in the logistic regression model using generalised linear models. Here we accounted for multiple wrists from the same patient as a repeated measure, and different MRI scans read by the same radiologist as another repeated-measures parameter [10].

In addition, we calculated the interobserver agreement. A multirater macro was used to determine interobserver agreement while accounting for five independent observers. Left and right wrists were evaluated separately and then pooled together. To account for multiple raters and wrists from the same patient, the independent working correlation structure worked best, and model fit was better treating both healthy volunteer and radiologist as subject effects, as opposed to either one alone [11].

Results

Healthy volunteers

Among the 319 rated MRI scans in healthy volunteers, 247 were diagnosed with no injury, 13 with scaphoid fracture, 23 with other fracture and 36 as a ''bone bruise'' (Figure 1). Of the 36 bone bruises, 10 involved the scaphoid and 26 involved other bones (carpus, metacarpus or distal radius).

4 out of 5 radiologists diagnosed a total of 13 scaphoid fractures (Table 1). On average the 5 radiologists diagnosed 2.6 (range 0–5) scaphoid fractures.

The 13 scaphoid fractures were diagnosed in 6 scans. For three scans the diagnosis was supported by only one
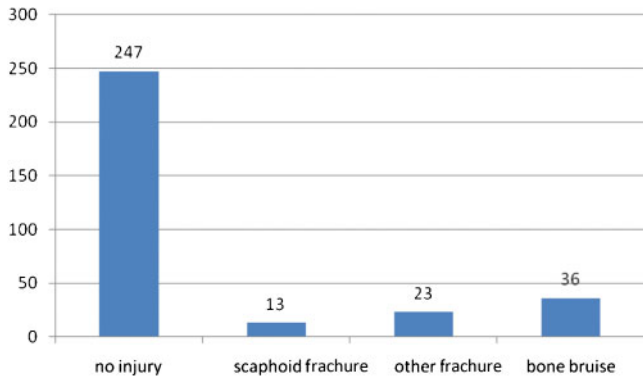
Figure 1. All diagnoses of the 64 healthy volunteers scored by 5 radiologists (*n*=319).

radiologist; two scans were supported by three radiologists; and one scan was supported by four radiologists.

In these six scans ten bone bruises of the scaphoid were diagnosed by radiologists who did not diagnose a fracture. Three scaphoids were diagnosed by all five radiologists as either fractured or bruised (Table 2).

Based on these data, the specificity of MRI was estimated as 95.9% (95% confidence interval: range 93.8–98.1%).

The multirater kappa value was 0.47 for all the right wrists and 0.359 for all the left wrists. The kappa value based on presence or absence of scaphoid fracture of the 5 observers was 0.44 (*p*<0.001), which is considered moderate interobserver agreement [10].

Discussion

MRI cannot be used as the reference standard for true fractures among suspected scaphoid fractures because, even in healthy volunteers with no history of wrist trauma, there are signal changes that radiologists sometimes interpret as a fracture or a bone bruise (Figure 2). The use of a reliable reference standard for no fracture (healthy volunteers) allows us to estimate the specificity of MRI for scaphoid fracture as approximately 96%, which is good, but not perfect. Given that the prevalence of true fractures among suspected scaphoid fractures is low, even small imperfections in diagnostic tests are magnified [8].

False-positive diagnosis in our cohort may be due to: (1) scoring a bone bruise as a fracture (some authors do this, while others require an interruption of the cortex or a clear fracture line); (2) the use of an abbreviated scanning protocol with only coronal plane images; and (3) training and experience (our impression is that radiologists with greater specialisation in musculoskeletal MRI have fewer false-positives). With respect to the influence of the MRI scanning protocol, a limited and fast (<7 min) protocol was used in order to evaluate a protocol that is easily implemented in a busy daily clinic (this protocol is routinely used in our clinic). Additional views or additional sequences might influence the diagnostic performance. Partial volume artefact within the limited scanning protocol could have contributed to false-positive findings.

Additional research is needed to determine optimal diagnostic criteria for a scaphoid fracture on MRI and what defines an adequate MRI scan for diagnosis of scaphoid fracture.

Given that our most sophisticated imaging techniques are imperfect, it must be accepted that there is no reference standard for true fracture of the scaphoid among patients with a suspected fracture. It is conceivable that certainty regarding the diagnosis of scaphoid fracture may be elusive. It may be more appropriate to treat patients and perform research based on the probability of a fracture. Latent class analysis—an alternative statistical method for calculating diagnostic performance statistics in the absence of a consensus reference standard—may be more appropriate in this context [12].

Table 1. Agreement between 5 radiologists in the presence or absence of a scaphoid fracture in 64 MRI scans of healthy volunteers

Radiologist	Scaphoid fracture (*n*)	No scaphoid fracture (*n*)	Specificity (95% confidence interval)
1	3	61	95% (87–99%)
2	1	63	98% (91–100%)
3	0	63	100% (94–100%)
4	5	59	92% (84–98%)
5	4	60	94% (91–100%)
Total	13	306	96% (93–98%)

Table 2. MRI findings in 6 healthy volunteers, in whom 1 or more radiologists diagnosed scaphoid fracture (13/319 scans), scaphoid bone bruise (*n*=10) or no abnormality (*n*=7)

Volunteer	Radiologist 1	Radiologist 2	Radiologist 3	Radiologist 4	Radiologist 5
1	No abnormality	No abnormality	No abnormality	Fracture	Bone bruise
2	No abnormality	No abnormality	No abnormality	Fracture	No abnormality
3	Fracture	Fracture	Bone bruise	Fracture	Fracture
4	Fracture	Bone bruise	Bone bruise	Fracture	Fracture
5	Fracture	Bone bruise	Bone bruise	Fracture	Fracture
6	Bone bruise	Bone bruise	Bone bruise	Bone bruise	Fracture

(a)   (b)
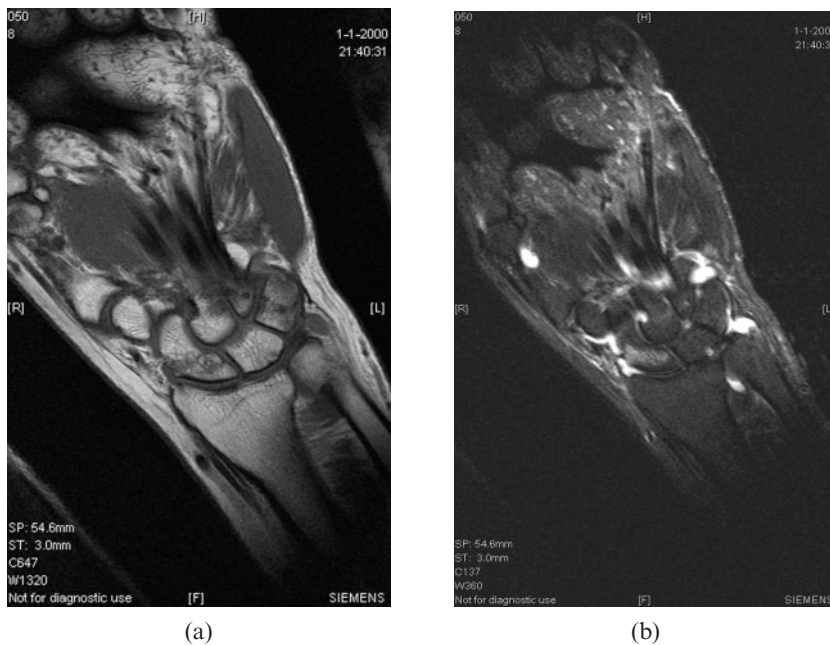
Figure 2. Four radiologists scored this MRI scan of a healthy volunteer as a scaphoid fracture. (a) T_1 sequence; (b) T_2 sequence.

Acknowledgments

References

1. Rubin DA, Daffner RH, Weissman BN, Bennett DL, Blebea JS, Jacobson JA, et al. ACR Appropriateness Criteria® acute hand and wrist trauma. [online publication]. Expert Panel on Musculoskeletal Imaging. Reston, VA: American College of Radiology (ACR); 2008. Available from: http://www.guideline.gov/content.aspx?id=13662AUQ4

2. Beeres FJP, Rhemrev SJ, den Hollander P, Kingma LM, Meylaerts SAG, le Cessie S, et al. Early magnetic resonance imaging compared with bone scintigraphy in suspected scaphoid fractures. J Bone Joint Surg Br 2008;90:1205–9.

3. Breitenseher MJ, Metz VM, Gilula LA, Gaebler C, Kukla C, Fleischmann D, et al. Radiographically occult scaphoid fractures: value of MR imaging in detection. Radiology 1997;203:245–50.

4. Fowler C, Sullivan B, Williams LA, McCarthy G, Savage R, Palmer A. A comparison of bone scintigraphy and MRI in the early diagnosis of the occult scaphoid waist fracture. Skeletal Radiol 1998;27:683–7.

5. Gaebler C, Kukla C, Breitenseher M, Trattnig S, Mittlboeck M, Vecsei V. Magnetic resonance imaging of occult scaphoid fractures. J Trauma 1996;41:73–6.

6. Hunter JC, Escobedo EM, Wilson AJ, Hanel DP, Zink-Brody GC, Mann FA. MR imaging of clinically suspected scaphoid fractures. AJR Am J Roentgenol 1997;168:1287–93.

7. Memarsadeghi M, Breitenseher MJ, Schaefer-Prokop C, Weber M, Aldrian S, Gabler C, et al. Occult scaphoid fractures: comparison of multidetector CT and MR imaging—initial experience. Radiology 2006;240:169–76.

8. Ring D, Lozano-Calderon S. Imaging for suspected scaphoid fracture. J Hand Surg Am 2008;33:954–7.

9. Tiel-van Buul MM, Roolker W, Verbeeten BW, Broekhuizen AH. Magnetic resonance imaging versus bone scintigraphy in suspected scaphoid fracture. Eur J Nucl Med 1996;23:971–5.

10. Julious SA. Two-sided confidence intervals for the single proportion: comparison of seven methods by Robert G. Newcombe, Statistics in Medicine 1998; 17:857–872. Stat Med 2005;24:3383–4.

11. Kundel HL, Polansky M. Measurement of observer agreement. Radiology 2003;228:303–8.

12. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. Biometrics 1980;36:167–71.