Diagnostic work-up of suspected scaphoid fractures

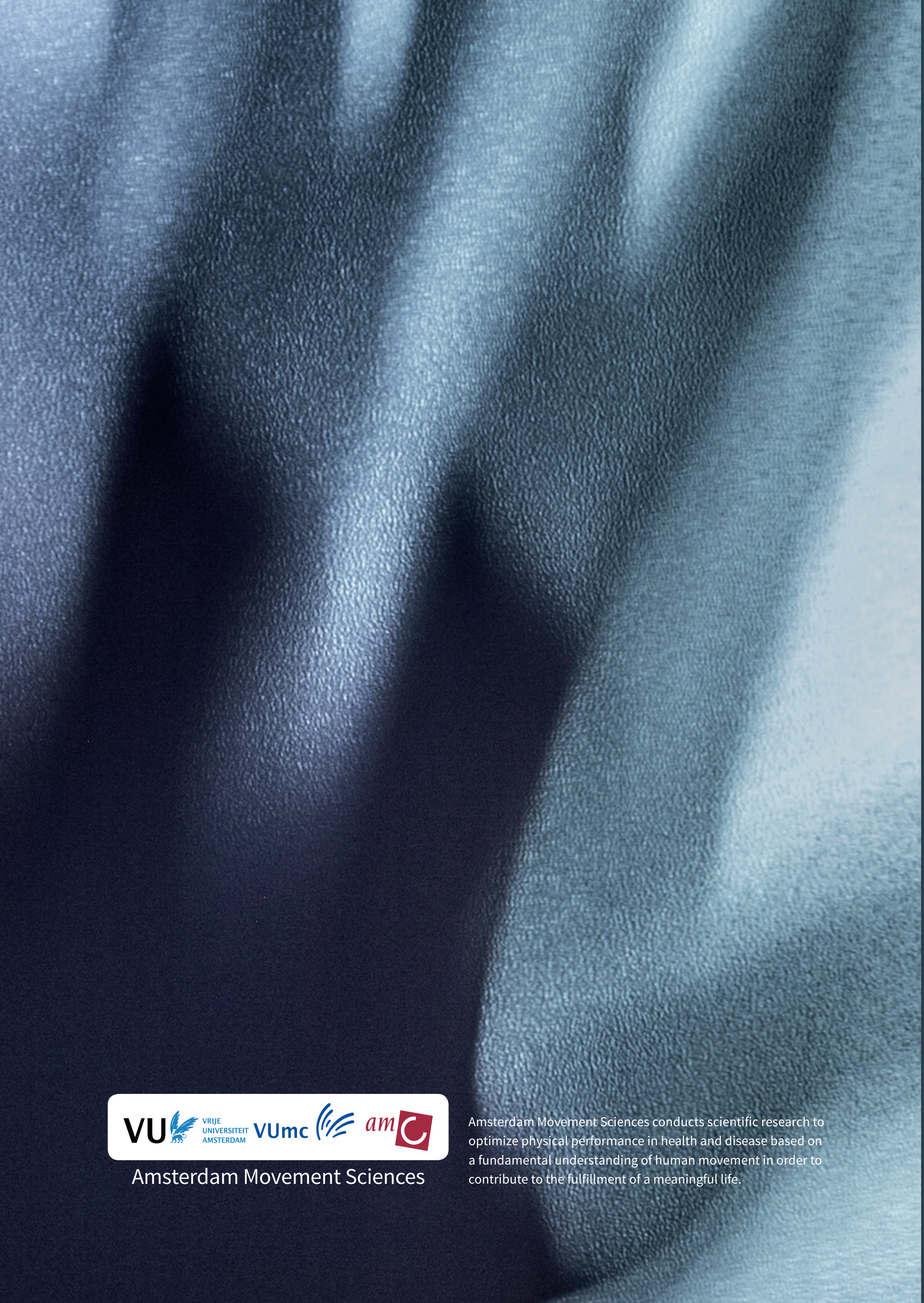
Mallee, W.H.

Diagnostic Work-Up of Suspected Scaphoid Fractures

Wouter H. Mallee

Diagnostic Work-Up of Suspected Scaphoid Fractures

Wouter H. Mallee



VU VRIJE UNIVERSITEIT AMSTERDAM VUmc amc

Amsterdam Movement Sciences

Amsterdam Movement Sciences conducts scientific research to optimize physical performance in health and disease based on a fundamental understanding of human movement in order to contribute to the fulfillment of a meaningful life.

DIAGNOSTIC WORK-UP OF SUSPECTED SCAPHOID FRACTURES

Wouter Harm Mallee

ISBN: 978-94-6182-938-2

Layout and printing: Off Page, Amsterdam

Cover design: Evelien Wajer

The research described and publication of this thesis was supported by a full AMC PhD Scholarship

This PhD thesis was embedded within Amsterdam Movement Sciences research institute, at the Department of Orthopaedic Surgery, Amsterdam UMC, University of Amsterdam, the Netherlands.

Printing of this thesis was financially supported by:

Amsterdam UMC, Department of Orthopaedic Surgery

Bauerfeind

Chipsoft

Anna Fonds

ETB-BISLIFE

Care10

Implantcast

Leuk Orthopedie

Probrace

Nederlandse Orthopaedische Vereniging

QMediq

Sectra

Traumaplatform

DIAGNOSTIC WORK-UP OF SUSPECTED SCAPHOID FRACTURES

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op donderdag 7 maart 2019, te 14:00 uur

door

Wouter Harm Mallee
geboren te Rotterdam

PROMOTIECOMMISSIE

Promotor(es)

Prof. dr. C.N. van Dijk	AMC-UvA
Prof. dr. J.C. Goslings	AMC-UvA

Copromotor(es)

Prof. dr. M. Maas	AMC-UvA
Dr. J.N. Doornberg	AMC-UvA

Overige leden

Prof. dr. G.M.M.J. Kerkhoffs	AMC-UvA
Prof. dr. R.J. Bennink	AMC-UvA
Prof. dr. R.J. Oostra	AMC-UvA
Prof. dr. F. Nollet	AMC-UvA
Dr. P.P.M. Rood	Erasmus MC
Prof. dr. J.H. Coert	UMC Utrecht
Dr. S.J. Rhemrev	MC Haaglanden

Faculteit der Geneeskunde

TABLE OF CONTENTS

PREFACE

The elusive scaphoid; a small bone in the proximal carpal row of the wrist.
To date, it has been the subject of an extensive amount of research over the past decades.
Much more than any other bone in the wrist.
Why?
Good question.
Here's why…

PART 1

INTRODUCING THE SCAPHOID: WHAT'S THE PROBLEM?

CHAPTER 1

GENERAL INTRODUCTION

With approximately 2-3% of all fractures and 90% of carpal fractures, scaphoid fractures are common(1-4). Moreover, patients with a suspected scaphoid fracture are even encountered 5 times as often in our Emergency Departments, resulting in a significant burden of -suspected-disease to society.

The scaphoid suffers from a low healing potential due to its fragile blood supply deriving from small radial artery branches that penetrate the bone distally and can be at risk in a fracture(5). Non-union (with avascular necrosis) can occur with potential carpal collapse and long term wrist arthritis if treatment is inadequate(6, 7).

Since mainly young and active adults suffer from this type of injury, its impact on both sport and work life is substantial. Establishing an early and adequate diagnosis is therefore crucial for a successful treatment(8-10). In addition, early and accurate diagnosis avoids unnecessary overtreatment, increases patient satisfaction and reduces both healthcare and societal costs.

However, early and accurate diagnosis is a challenge.

CURRENT ISSUES IN DIAGNOSTIC MANAGEMENT

Inaccurate and delayed diagnoses are still issues in scaphoid fracture management. Inadequate clinical evaluation and imaging are main contributors to the burden of costs and morbidity to our health care system. These should be optimalised to refrain patients from overtreatment initially in the 'suspected' scaphoid patient-group; and preventable wrist problems in the 'confirmed' scaphoid fracture patient-group.

Diagnostic Issues – Clinical Evaluation

A clinically suspected scaphoid fracture has considerable consequences for further management in terms of cast or splint immobilization -and thus time of work- and further follow-up visits and imaging. Despite its consequences, there is limited scientific evidence on the effectiveness of clinical evaluation for detection of scaphoid fractures in wrist trauma.

A painful anatomic snuffbox has been 'classically' described as the key test to suggest a scaphoid fracture(11, 12). Later, longitudinal compression of the thumb and scaphoid tubercle tenderness were added(12-14). A combination of these tests was believed to improve accuracy(15, 16). However, to date considerably more patients are clinically suspected for having a scaphoid fracture without having a true fracture. Up to 84% of patients receive unnecessary immobilization and follow-up visits(17).

In addition, the low prevalence of true fractures among suspected fractures form a statistical hazard in evaluating diagnostic performance characteristics of follow-up imaging strategies when Bayes' Theorem is applied. The Bayes' Theorem accounts for the a priori prevalence of the disease in calculating sensitivity and specificity of a test(18).

Clinical and Scientific Needs – Clinical Evaluation

The starting point is a clinically suspected scaphoid fracture. So, in order to…

- reduce unnecessary immobilization, hospital visits and use of imaging modalities for the 'suspected' scaphoid patient group;

- not miss scaphoid fractures initially for the 'confirmed' scaphoid fracture group

… there is a clinical- and scientific need to introduce a standardized evidence based clinical evaluation protocol, similar to the clinically relevant and scientifically based

Ottawa Ankle Rules(19). This type of clinical prediction rules could play an important role in detecting scaphoid fractures, and thus decrease the burden to society of the patient-group with suspected scaphoid fractures.

Diagnostic Issues – Imaging Strategy

If a scaphoid fracture is clinically suspected, radiographs are obtained. Radiographic series consist of standardized postero-anterior en lateral wrist radiographs and additional scaphoid views (semi-pronated oblique and posteroanterior with the wrist in ulnar deviation)(20). It is known that up to 38% of all patients with a clinically suspected scaphoid fracture and normal radiographs will still have a fracture(21, 22). In other words, radiographs fail to identify these occult scaphoid fractures; they are the reason for the defensive initial management that follows. These patients are initially treated with a cast or splint to immobilize the scaphoid. One or two weeks later, further evaluation is performed.

Incorporating advanced imaging modalities such as Bone Scintigraphy (BS), Computed Tomography (CT) and/or Magnetic Resonance Imaging (MRI) in the diagnostic work-up for patients with a clinically suspected scaphoid fracture and normal radiograph became more interesting: one could rule out a 'suspected' fracture and patients are able to participate in society. These modalities have all been described as being highly accurate in establishing definitive diagnosis(23, 24). However, all advanced imaging techniques come with a price and are not readily available. It is still unknown which of the three is superior in ruling a fracture in or out.

Another interesting aspect is timing of additional imaging. Initial splinting when a scaphoid fracture is suspected refrains people from physical activity in sports and work. In this young and active population, immediate diagnosis is strongly preferred. One could argue that obtaining immediate CT or MRI could be a cost-effective approach as patients may return to work 2 weeks earlier, however macroeconomic cost analyses of this 'early-imaging-strategy' are unknown.

For diagnostic test accuracy studies of imaging techniques, a reference standard is needed. Repeating radiographs after six weeks has long been the preferred reference standard for establishing definitive diagnosis since these are believed to show fracture healing(22, 25). The actual accuracy and reliability of this test however is unknown and since the substantial time interval is not beneficial for both research and time off work for patients, their value in scaphoid injury is questionable(26-28).

Clinical Evaluation – Imaging Strategy

If there is a lack of consensus in literature, a large variety in the diagnostic work-up and initial treatment of suspected scaphoid fractures is inevitable(29, 30). It amplifies the need for an evidence-based imaging protocol. In order to...

- ensure accurate definitive diagnosis in case of a clinically suspected scaphoid fracture
- create a beneficial approach for both patient and healthcare system in terms of costs and timing

… there is a clinical- and scientific need to identify the superior imaging technique to establish a definitive diagnosis and its timing and costs must be accounted for. In addition, the on-going debate on the questionable reference standard must be clarified.

THESIS AIMS

The overall aim of this thesis is to thoroughly review current diagnostic management of 'suspected' scaphoid fractures and to significantly improve the future work-up. This is divided into two parts: 1) improving Clinical Evaluation; and 2) improving Diagnostic Imaging.

The following study questions will be answered:

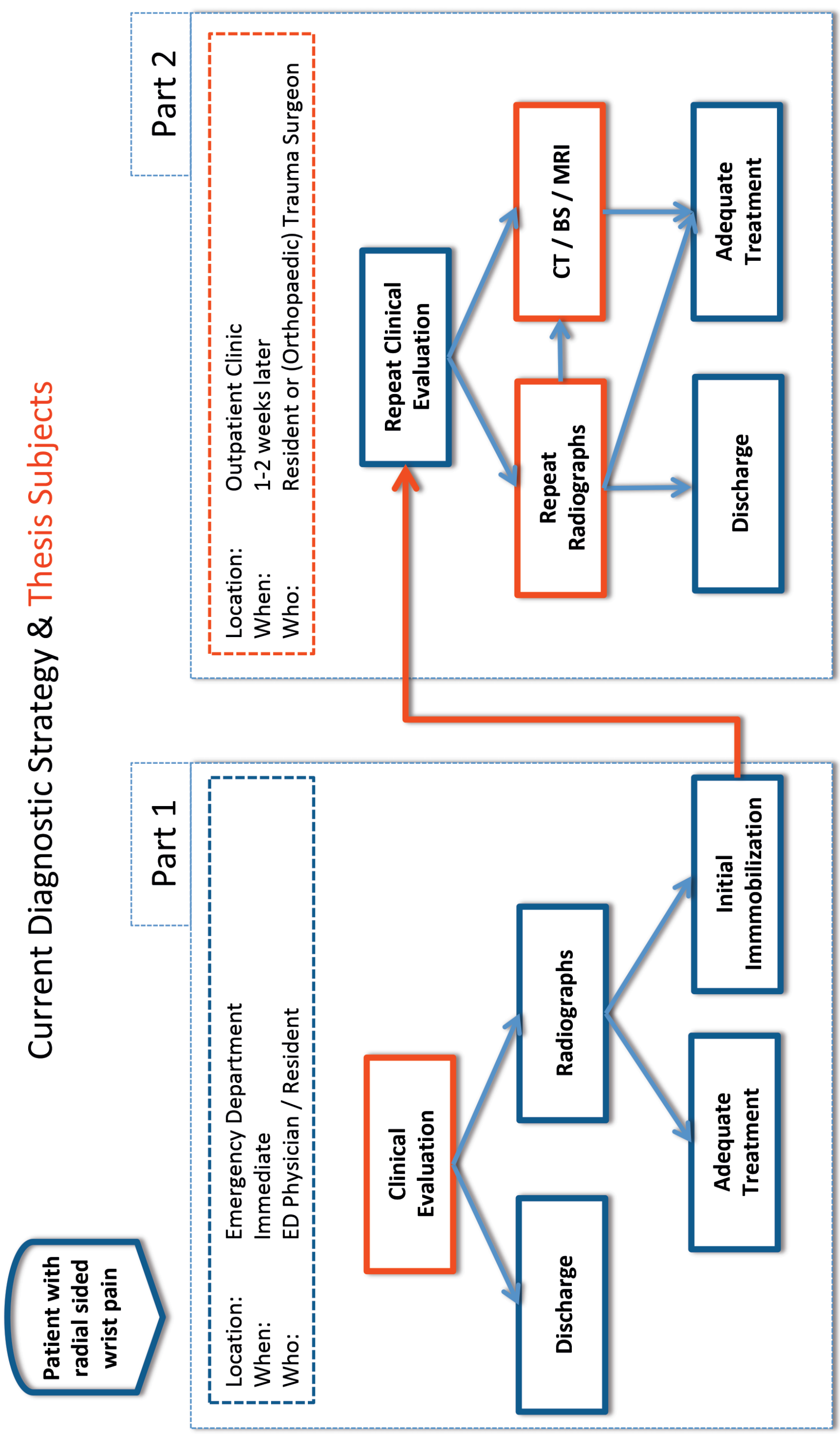
1. What are the current protocols and issues in management of suspected scaphoid fractures?
2. What are current tests in clinical evaluation to suspect a scaphoid injury and how accurate are these?
3. How can we improve the selection of patients with a clinically suspected scaphoid fracture?
4. Which imaging modality should be used to establish a definitive diagnosis in patients with a clinically suspected scaphoid fracture and normal radiographs?
5. What is the most efficient timing of applying advanced imaging in terms of costs?

The final goal is to present a cost efficient and protocolled work-up that leads to earlier and more accurate diagnosis, reduces overtreatment and follow-up imaging and outpatient clinic visits without an increased risk of missing a fracture.

THESIS OUTLINE (FLOWCHART 1)

Part 1 introduces diagnostic issues. In Chapter 2, current scaphoid fracture management in Dutch hospitals is presented based on a national survey among all 100 hospitals with an Emergency Department. Use of imaging, timing and treatment strategies were questioned and compared with current literature. Due to varying availability of imaging tools and lack of evidence-based medicine, we hypothesized large variety between hospitals in diagnosis and treatment of scaphoid fractures.

Part 2 focuses on clinical evaluation. In Chapter 3, a systematic review and meta-analysis is performed to evaluate all available clinical tests and their diagnostic performance characteristics in an attempt to identify the best possible tests for detection of a scaphoid fracture. This review provides a set of important predictors that will be used in Chapter 4. In this Chapter, a large prospective multicenter study is performed to develop and internally validate a clinical decision rule that applies to the patient group with a clinically suspected scaphoid fracture.


Current Diagnostic Strategy & Thesis Subjects
Patient with radial sided wrist pain
Part 1
Location: Emergency Department
When: Immediate
Who: ED Physician / Resident
Clinical Evaluation
Discharge
Radiographs
Adequate Treatment
Initial Immmobilization
Part 2
Location: Outpatient Clinic
When: 1-2 weeks later
Who: Resident or (Orthopaedic) Trauma Surgeon
Repeat Clinical Evaluation
Repeat Radiographs
CT / BS / MRI
Discharge
Adequate Treatment

Flowchart 1.

We hypothesize that a unique combination of tests identified in Chapter 3, will provide both a reduction in the need for follow-up imaging as well as a low risk of missing scaphoid fractures.

Part 3 focuses on additional imaging modalities in suspected scaphoid fractures: Bone Scinitigraphy (BS), Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). These advanced imaging techniques are used to establish a definitive diagnosis when radiographs are unclear. Chapter 5 presents a closer look at creating different reformations of CT scanning in order to accurately visualize the scaphoid bone. Fractures of the scaphoid after radiographs appear normal are smaller and more challenging to visualize. We will compare two different reformations and hypothesize that reformations in the long axis of the scaphoid are superior in fracture diagnosis.

In Chapter 6, a prospective comparison of CT and MRI is performed. All included patients underwent both CT and MRI together with the current best reference standard: repeated radiographs 6 weeks after trauma. We hypothesize that CT and MRI are similar in diagnostic performance characteristics.

Since the use of plain radiographs as the reference standard for advanced imaging techniques has been debated, the interobserver reliability and diagnostic performance characteristics are tested when compared to CT and MRI in Chapter 7. The hypothesis is that both accuracy and interobserver reliability are insufficient to warrant future use in diagnostic scaphoid studies.

The potential limitations of a debatable reference standard require an out-of-the-box solution: Chapter 8 presents a different statistical approach coined latent class analysis. This method is applicable to calculate the diagnostic test accuracy in the absence of a reference- or gold standard. The null-hypothesis is that applying latent class analysis will not differ from the standardized calculations of diagnostic performance characteristics.

In Chapter 9a, the diagnostic performance characteristics of all three advanced imaging modalities are systematically reviewed in a Cochrane Systematic Review and Meta-analysis. The goal of the review is to identify the superior imaging technique when radiographs are unclear. Based on this review, a book Chapter is written to summarize the findings and combines this with a review of cost-effectiveness studies. It is hypothesized that performing immediate additional imaging is a cost-effective approach.

In Chapter 10, a summary and general discussion is given of all studies presented in this thesis. Based on this thesis, an efficient and cost-effective protocol is presented in the final Chapter 11 concerning the full diagnostic work-up for scaphoid injury starting in the ED until an early and definitive diagnosis is established.

REFERENCES

1. Hove LM. Epidemiology of scaphoid fractures in Bergen, Norway. Scand J Plast Reconstr Surg Hand Surg. 1999;33(4):423-6.
2. van Onselen EB, Karim RB, Hage JJ, Ritt MJ. Prevalence and distribution of hand fractures. Journal of hand surgery (Edinburgh, Scotland). 2003;28(5):491-5.
3. van der Molen AB, Groothoff JW, Visser GJ, Robinson PH, Eisma WH. Time off work due to scaphoid fractures and other carpal injuries in The Netherlands in the period 1990 to 1993. Journal of hand surgery (Edinburgh, Scotland). 1999;24(2):193-8.
4. Hey HW, Chong AK, Murphy D. Prevalence of carpal fracture in Singapore. J Hand Surg Am. 2011;36(2):278-83.
5. Gelberman R.H. - Gross MS. The vascularity of the wrist. Identification of arterial patterns at risk. Clinical orthopaedics and related research. 1986(0009-921X (Print)).
6. Breen T, Gelberman RH, Leffert R, Botte M. Massive allograft replacement of hemiarticular traumatic defects of the elbow. J Hand Surg. 1988;13A:900-7.
7. Merrell GA, Wolfe S.W. - Slade JF. Treatment of scaphoid nonunions: quantitative meta-analysis of the literature. Journal of Hand Surgery, American Volume. 2002(0363-5023 (Print)).
8. Dias JJ, Wildin CJ, Bhowal B, Thompson JR. Should acute scaphoid fractures be fixed? A randomized controlled trial. J Bone Joint Surg Am. 2005;87(10):2160-8.
9. Adey L, Souer JS, Lozano-Calderon S, Palmer W, Lee SG, Ring D. Computed tomography of suspected scaphoid fractures. J Hand Surg Am. 2007;32(1):61-6.
10. Lozano-Calderon S, Blazar P, Zurakowski D, Lee SG, Ring D. Diagnosis of scaphoid fracture displacement with radiography and computed tomography. J Bone Joint Surg Am. 2006;88(12):2695-703.
11. Waeckerle JF. A prospective study identifying the sensitivity of radiographic findings and the efficacy of clinical findings in carpal navicular fractures. Ann Emerg Med. 1987;16(7):733-7.
12. Freeland P. Scaphoid tubercle tenderness: a better indicator of scaphoid fractures? Archives of emergency medicine. 1989;6(1):46-50.
13. Brittain HA. Fracture of the Carpal Scaphoid. Br Med J. 1938;2(4055):671-3.
14. Waizenegger M, Barton NJ, Davis TR, Wastie ML. Clinical signs in scaphoid fractures. Journal of hand surgery (Edinburgh, Scotland). 1994;19(6):743-7.
15. Parvizi J, Wayman J, Kelly P, Moran CG. Combining the clinical signs improves diagnosis of scaphoid fractures. A prospective study with follow-up. Journal of hand surgery (Edinburgh, Scotland). 1998;23(3):324-7.
16. Rhemrev SJ, Beeres FJ, van Leerdam RH, Hogervorst M, Ring D. Clinical prediction rule for suspected scaphoid fractures: A prospective cohort study. Injury. 2010;41(10):1026-30.
17. Rhemrev SJ, de Zwart AD, Kingma LM, Meylaerts SA, Arndt JW, Schipper IB, et al. Early computed tomography compared with bone scintigraphy in suspected scaphoid fractures. Clinical nuclear medicine. 2010;35(12):931-4.
18. Ring D, Lozano-Calderon S. Imaging for suspected scaphoid fracture. J Hand Surg Am. 2008;33(6):954-7.
19. Stiell IG, McKnight RD, Greenberg GH, McDowell I, Nair RC, Wells GA, et al. Implementation of the Ottawa ankle rules. JAMA : the journal of the American Medical Association. 1994;271(11):827-32.
20. Yin ZG, Zhang JB-K, S.L. - Wang, X. G. Diagnosing suspected scaphoid fractures: a systematic review and meta-analysis. Clinical orthopaedics and related research. 2010;468(3)(1528-1132 (Electronic)):723-34.
21. Stordahl A, Schjoth A, Woxholt G, Fjermeros H. Bone scanning of fractures of the scaphoid. Journal of hand surgery (Edinburgh, Scotland). 1984;9(2):189-90.
22. 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(1):169-76.
23. Roolker W, Tiel-van Buul MM, Ritt MJ, Verbeeten B, Jr., Griffioen FM, Broekhuizen AH. Experimental evaluation of scaphoid X-series, carpal box radiographs, planar tomography, computed tomography, and magnetic resonance imaging in the diagnosis of scaphoid fracture. J Trauma. 1997;42(2):247-53.

24. 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(8):971-5.
25. 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(1):245-50.
26. Low G, Raby N. Can follow-up radiography for acute scaphoid fracture still be considered a valid investigation? Clinical radiology. 2005;60(10):1106-10.
27. Tiel-van Buul MM, van Beek EJ, Broekhuizen AH, Nooitgedacht EA, Davids PH, Bakker AJ. Diagnosing scaphoid fractures: radiographs cannot be used as a gold standard! Injury. 1992;23(2):77-9.
28. Tiel-van Buul MM, van Beek EJ, Borm JJ, Gubler FM, Broekhuizen AH, van Royen EA. The value of radiographs and bone scintigraphy in suspected scaphoid fracture. A statistical analysis. Journal of hand surgery (Edinburgh, Scotland). 1993;18(3):403-6.
29. Groves AM, Kayani I, Syed R, Hutton BF, Bearcroft PP, Dixon AK, et al. An international survey of hospital practice in the imaging of acute scaphoid trauma. AJR American journal of roentgenology. 2006;187(6):1453-6.
30. Brookes-Fazakerley SD, Kumar AJ-O, J. Survey of the initial management and imaging protocols for occult scaphoid fractures in UK hospitals. Skeletal radiology. 2009;38(11):1045-8.

CHAPTER 2

VARIATIONS IN MANAGEMENT OF SUSPECTED SCAPHOID FRACTURES

Mallee WH
Veltman ES
Doornberg JN
Blankevoort L
Goslings JC
van Dijk CN

Nederlands Tijdschrift voor Geneeskunde. 2012

SAMENVATTING

Doel

Deze studie heeft het doel om het beleid bij patiënten met een vermoeden op een scafoïdfractuur in Nederlandse ziekenhuizen te evalueren en te vergelijken met de aanbevelingen uit de huidige literatuur.

Opzet

Enquête-onderzoek.

Methode

In Nederlandse ziekenhuizen werden artsen die werkzaam waren op de Spoedeisende Hulp gevraagd om een enquête in te vullen van 8 vragen over diagnostische strategieën, het type behandeling en de tijd tussen verschillende stappen van het beleid.

Resultaten

Van de 100 benaderde ziekenhuizen vulden artsen uit 90 ziekenhuizen de enquête in. Van deze 90 ziekenhuizen was bij 71 het beleid in een protocol vastgelegd. Bij de overige 19 was dat afhankelijk van de voorkeur van de behandelend arts. Bij 75 ziekenhuizen werd poliklinische controle binnen 10 dagen afgesproken. In 70 ziekenhuizen werden de röntgenfoto's herhaald vóór aanvullend beeldvormend onderzoek. Als aanvullend onderzoek werd in 35 ziekenhuizen de CT het meest gebruikt, gevolgd door botscintigrafie (12) en MRI (2). In 11 ziekenhuizen werd geen aanvullend onderzoek verricht en werd de behandeling bij niet-afwijkende röntgenfoto's voortgezet op basis van klinische evaluatie. In 72 ziekenhuizen werd de pols geïmmobiliseerd met een onderarmsgips met inclusie van de duim. Een onderarmsgips zonder inclusie van de duim werd door 1 ziekenhuis toegepast.

Conclusie

Er is een grote variatie tussen de Nederlandse ziekenhuizen in de diagnostiek en behandeling van patiënten met het vermoeden op een scafoïdfractuur. Daarnaast is het beleid in de meeste ziekenhuizen niet volgens de recentste aanbevelingen. Er is behoefte aan een evidence-based richtlijn zodat overdiagnostiek en onnodige immobilisatie beperkt kunnen worden.

ABSTRACT

Objective

This study evaluated the daily clinical practice for management of patients with suspected scaphoid fractures in hospitals in the Netherlands and compared it with recommendations from the current literature.

Design

Questionnaire-based investigation.

Method

Doctors working in emergency departments in hospitals in the Netherlands were asked to complete an 8-question survey including questions on diagnostic strategies, treatment type and the time between different steps in the management policy.

Results

Doctors from 90 of the 100 hospitals approached completed the questionnaire. A total of 71 of these 90 hospitals had an established protocol. In the other 19 it depended on the preference of the treating doctor. In 75 hospitals a follow-up outpatient clinic appointment was made for within 10 days. In 70 hospitals X-rays were repeated before additional imaging investigation. CT was the most frequently used additional investigation in 35 hospitals, followed by bone scintigraphy (12) and MRI (2). No additional investigation was carried out in 11 hospitals and when X-rays showed no abnormalities treatment was implemented on the basis of clinical evaluation. In 72 hospitals the wrist was immobilised with a lower-arm plaster cast including the thumb. Lower-arm plaster cast not including the thumb was used in 1 hospital.

Conclusion

There is a great deal of variation in diagnosis and treatment of patients with a suspected scaphoid fracture within hospitals in the Netherlands. Furthermore, management policy in most hospitals is not in keeping with the most recent recommendations. Evidence-based guidelines are required in order to limit over-diagnosis and unnecessary immobilisation.

INTRODUCTIE

Het os scaphoideum speelt een belangrijke rol bij alle bewegingen van de pols. Beperking van de bewegingscapaciteit van het os scaphoideum zal dan ook ernstige gevolgen hebben voor de functie van de hand. Bij trauma van de hand heeft 7% van de patiënten een scafoïdfractuur(1). Er rijst een klinisch vermoeden op een scafoïdfractuur wanneer palpatie van de tabatière anatomique en axiale druk op de 1e straal van de hand pijnlijk zijn(2). Op de eerste röntgenfoto's van de pols kan de fractuur soms nog niet worden gezien. Bij het vermoeden op een scafoïdfractuur met röntgenfoto's waarop geen fractuur te zien is, heeft 16% van de patiënten alsnog een fractuur(3, 4).

Adequate diagnose en behandeling van scafoïdfracturen zijn belangrijk voor een goede genezing en herstel van de polsfunctie en om het risico op 'non-union' te minimaliseren(5-7). Vanwege het risico op complicaties bij onderbehandeling en vanwege de lage sensitiviteit van de röntgenfoto's wordt de hand ook geïmmobiliseerd wanneer er wel een vermoeden op een scafoïdfractuur bestaat, maar op de röntgenfoto's geen fractuur zichtbaar is(8-10). Dit houdt in dat bij 5 van de 6 patiënten de hand onterecht geïmmobiliseerd wordt, met alle gevolgen van dien voor de dagelijkse activiteiten(11, 12).

Diagnostiek

Er is bij dit probleem behoefte aan een structurele aanpak om overbehandeling te minimaliseren. Het gebrek aan consensus over diagnostiek en behandeling van scafoïdfracturen staat de ontwikkeling van een protocol echter in de weg. Een van de diagnostische problemen bij het vermoeden op een scafoïdfractuur is de keuze van het aanvullend beeldvormend onderzoek. In ziekenhuizen varieert de keuze van aanvullend onderzoek van botscintigrafie, MRI en CT tot het herhalen van de röntgenfoto's. Al deze methodes zijn geëvalueerd als diagnostisch hulpmiddel bij het vermoeden op een scafoïdfractuur(13-20). Vanwege de uiteenlopende eigenschappen van deze modaliteiten is het moeilijk een definitieve keuze te maken. Daarnaast speelt de beschikbaarheid van de apparatuur ook een rol.

Behandeling

De behandeling van scafoïdfracturen is ook onderwerp van discussie. Wanneer er het vermoeden bestaat op een scafoïdfractuur, moet het os scaphoideum geïmmobiliseerd worden totdat een definitieve diagnose is gesteld. Immobilisatie is mogelijk met een circulair onderarmsgips met of zonder inclusie van de duim (respectievelijk klassiek scafoïdgips en collesgips) of door middel van een spalk(21, 22). Hoewel 1 niveau 1-studie en een niveau 4-studie suggereren dat een collesgips geschikt is voor de immobilisatie van een scafoïdfractuur, blijkt uit recente studies dat het klassieke gips nog steeds veel toegepast wordt. Dit gips geeft echter meer functiebeperkingen(21-23).

Het aantal röntgenfoto's dat bij eerste presentatie genomen wordt, de timing van aanvullend beeldvormend onderzoek en het moment van poliklinische controle staan ook ter discussie. Het meeste hiervan is nooit onderzocht in studieverband. Gezien het gebrek aan wetenschappelijke onderbouwing en consensus mag verwacht worden dat er aanzienlijke verschillen bestaan tussen ziekenhuizen. Eerdere studies lieten ook al variatie tussen behandelprotocollen zien, maar deze studies concentreerden zich vooral op het beeldvormend onderzoek bij occulte scafoïdfracturen(8-10).

Het doel van deze studie was om de huidige dagelijkse praktijk van de diagnostiek en initiële behandeling van scafoïdfracturen binnen de Nederlandse gezondheidszorg te evalueren. Het huidige beleid in Nederland werd vergeleken met de recentste aanbevelingen uit de literatuur. Onze hypothese was dat er een grote variatie bestaat tussen de Nederlandse ziekenhuizen wat betreft de keuze van beeldvormend onderzoek, manier van immobilisatie en de tijd tussen de eerste presentatie en uiteindelijke diagnose.

METHODE

Om te achterhalen wie verantwoordelijk was voor het diagnostisch en behandelprotocol van scafoïdfracturen, namen wij contact op met het medisch hoofd van de Spoedeisende Hulp (SEH) van alle ziekenhuizen in Nederland. Wanneer het hoofd van de SEH niet bereikbaar was, werd een orthopedisch chirurg, traumachirurg of SEH-arts gevraagd mee te werken aan het onderzoek mits zij voldoende kennis van het protocol hadden. We stelden een enquête samen van 8 vragen. De vragenlijst werd verstuurd via e-mail (tabel 1). Artsen die de vragenlijst niet terugstuurden, werden herhaaldelijk telefonisch benaderd met intervallen van 2 weken om een hoge respons te garanderen. In de periode februari-april 2010 werden 90 academische ziekenhuizen en regioziekenhuizen met een SEH geïncludeerd.

RESULTATEN

Van de 100 benaderde ziekenhuizen werd de enquête door artsen uit 90 ziekenhuizen beantwoord. Traumachirurgen stonden het vaakst aan het hoofd van de SEH (30%) gevolgd door SEH-artsen (27%) en de orthopedisch chirurgen (14%) (tabel 2).

In 71 van de 90 van de ziekenhuizen (79%) was een protocol voor diagnostiek en behandeling van scafoïdfracturen op de SEH aanwezig. In de overige 19 ziekenhuizen (21%) was het beleid afhankelijk van de dienstdoende arts.

Tabel 1. Enquête vragen

1.	In welk ziekenhuis bent u werkzaam?
2.	Wat is uw functie binnen dit ziekenhuis?
3.	Bestaat er een standaard protocol bij een verdenking op een scafoïd fractuur?
4.	In welke richtingen worden de initiële röntgenfoto’s gemaakt?
5.	Als immobilisatie gewenst is, hoe wordt de pols dan gegipst?
6.	Worden de röntgenfoto’s herhaald voordat aanvullend onderzoek wordt verricht? Zo ja, na hoeveel dagen?
7.	Na hoeveel dagen wordt de patiënt voor controle gezien op de poli-kliniek?
8.	Welke aanvullende beeldvorming wordt gebruikt en hoeveel dagen na trauma gebeurt dit?

Het aantal initiële röntgenfoto's varieerde van 2-6. Dat gebeurde het meest in 4 richtingen (42 ziekenhuizen; 47%), en in 3 richtingen (32 ziekenhuizen; 36%) (tabel 3).

Immobilisatie van het os scaphoïdeum gebeurde in 72 ziekenhuizen (80%) met een klassiek scafoïdgips (met inclusie van de duim). Slechts in 1 ziekenhuis (1%) werd gebruik gemaakt van het collesgips (tabel 4). Bij de overige 17 ziekenhuizen werd een spalk aangelegd.

In 75 ziekenhuizen (83%) vond poliklinische controle 7-10 dagen na de SEH-presentatie plaats. Bij 9 ziekenhuizen (10%) gebeurde dit na 10-14 dagen. In 3 ziekenhuizen (3%) werd de patiënt poliklinisch teruggezien binnen 7 dagen na het trauma en in de andere 3 ziekenhuizen werden patiënten pas voor controle gezien nadat het os scaphoideum 6 weken geïmmobiliseerd was geweest.

In 70 ziekenhuizen (78%) werden de röntgenfoto's herhaald voordat een ander soort beeldvormend onderzoek (CT, MRI of botscintigrafie) werd aangevraagd. Bij 65 ziekenhuizen (93%) werden de röntgenfoto's herhaald binnen 2 weken na de eerste presentatie.

Van de verschillende modaliteiten was CT de eerste keuze in 35 ziekenhuizen (39%), gevolgd door botscintigrafie in 12 ziekenhuizen (13%). MRI was de eerste keus in 2 ziekenhuizen (2%). In de overige ziekenhuizen werd de methode van beeldvormend onderzoek bepaald door de beschikbaarheid van de radiologische faciliteiten en de voorkeur van de behandelend arts. De keuze van de behandeld arts was tussen CT en botscintigrafie in 15 ziekenhuizen (17%), tussen MRI en CT in 8 ziekenhuizen (9%), en tussen botscintigrafie en MRI in 3 ziekenhuizen (3%). Er was geen voorkeur voor CT, MRI of botscintigrafie in 4 ziekenhuizen (4%), terwijl in 11 ziekenhuizen (12%) nooit een andere vorm van beeldvormend onderzoek dan röntgenfoto's werd ingezet (tabel 5).

BESCHOUWING

Deze studie laat een grote variatie zien tussen de Nederlandse ziekenhuizen wat betreft diagnostiek en behandeling van een patiënt met het vermoeden op een scafoïdfractuur. Dit is in overeenstemming met vergelijkbare studies die in andere landen zijn uitgevoerd(8-10, 24). In tegenstelling tot deze studies heeft 79% van Nederlandse ziekenhuizen een protocol voor het beleid bij het vermoeden op een scafoïdfractuur. In de overige ziekenhuizen is de keuze voor diagnostiek en behandeling van dit type fractuur afhankelijk van de voorkeur van

Tabel 2. Beroep van de respondenten

Beroep	Aantal respondenten	Percentage
Traumachirurg	27	30%
SEH-arts	24	27%
Orthopaedisch chirurg	13	14%
Arts in opleiding	12	13%
Nurse practitioner	8	9%
Hoofd gipskamer	6	7%

Tabel 3. Aantal initiële röntgenfoto's

Aantal röntgenfoto's	Aantal respondenten	Percentage
2	10	11%
3	32	36%
4	42	47%
5	3	3%
6	2	2%
Incomplete antwoorden	1	1%

Tabel 4. Gipsmethode

Gipsmethode	Aantal respondenten	Percentage
Gips met inclusie van duim	72	80%
Gips zonder inclusie van duim	1	1%
Spalk	17	19%

Tabel 5. Aanvullende diagnostiek

Diagnostische modaliteit	Aantal respondenten	Percentage
CT	35	39%
MRI	2	2%
BS	12	13%
BS of CT	15	17%
MRI of CT	8	9%
MRI of BS	3	3%
MRI of BS of CT	4	4%
Geen	11	12%

de behandelend specialist. Deze voorkeur is onder andere afhankelijk van de beschikbaarheid van beeldvormende modaliteiten op het moment van presentatie van de patiënt. CT is over het algemeen ruimer beschikbaar dan botscintigrafie of MRI.

Beeldvormend onderzoek

De grote diversiteit in het gekozen aanvullend beeldvormend onderzoek bij patiënten bij wie het vermoeden bestaat op een scafoïdfractuur, wordt mogelijk verklaard door het gebrek aan consensus tussen verschillende studies. Er zijn voorstanders van het gebruik van CT (18, 25-28), maar ook mensen die pleiten voor het inzetten van MRI(28-31), en van botscintigrafie (14, 26). Het gebrek aan consensus blijkt ook uit onze resultaten. CT is eerste keus voor aanvullend onderzoek in 39% van de ziekenhuizen. Daarnaast behoort CT in 30%

van de ziekenhuizen tot een van de mogelijke modaliteiten; uiteindelijke keuze is afhankelijk van de behandeld arts.

Uit onze studie werd duidelijk dat de röntgenfoto's in 78% van de ziekenhuizen werden herhaald voordat een ander type beeldvormend onderzoek werd aangevraagd. Eerdere onderzoeken naar kosteneffectiviteit bij dit type fractuur suggereren dat het snel inzetten van aanvullend beeldvormend onderzoek (binnen 5 dagen) niet gepaard gaat met hogere kosten voor de gezondheidszorg. Daarnaast wordt op deze manier onnodige immobilisatie beperkt, waardoor patiënten minder gehinderd worden in hun dagelijkse activiteiten. Zo kunnen uiteindelijk ook de maatschappelijke kosten verminderd worden(12, 32, 33). In geen van de Nederlandse ziekenhuizen wordt dit beleid echter toegepast. Mogelijk is het in de Nederlandse ziekenhuizen op micro-economisch niveau ongunstig om op dezelfde DBC meer aanvullende diagnostiek aan te vragen bij een klinisch vermoeden op een scafoïdfractuur. Macro-economisch blijkt snelle mobilisatie van de patiënt na uitsluiten van een scafoïdfractuur met duurder aanvullend beeldvormend onderzoek voordeliger(12).

Een recente systematische review en meta-analyse heeft laten zien dat MRI en CT bij het vermoeden op een scafoïdfractuur een significant betere specificiteit hebben dan botscintigrafie(28). In 1 van de 2 prospectieve studies waar CT en MRI direct vergeleken waren, werd MRI geadviseerd in verband met een superieure accuratesse(34). Vergelijkbare buitenlandse studies naar de variëteit van beleid en diagnostiek van scafoïdfracturen laten zien dat MRI in 41-58% van de ziekenhuizen wordt toegepast in de Verenigde Staten, Engeland en Australië(8-10).

In onze studie valt op dat MRI slechts in 2% van Nederlandse ziekenhuizen het onderzoek van eerste keus is en in 17% tot een van de opties behoort. Dit verschil kan worden verklaard door de beperktere beschikbaarheid en de hogere kosten van MRI in Nederland en door de voorkeur van Nederlandse artsen. Daarentegen heeft MRI wel het voordeel dat hiermee pathologische afwijkingen van de weke delen beter in beeld gebracht kunnen worden. De klinische relevantie in de acute fase is tot op heden nog onduidelijk.

Immobilisatie

Verschillende studies laten zien dat een collesgips voldoende immobilisatie geeft voor de behandeling van scafoïdfracturen. Dit gips heeft als voordeel dat het de patiënt minder beperkt in zijn handelen gedurende de lange immobilisatie van de pols, doordat de duim niet geïncludeerd wordt(21, 22). Traditionele biomechanische en klinische studies raden echter gipsimmobilisatie aan, waarbij de patiënt meer beperkt wordt(30, 31). In overeenstemming met deze visie werd in nagenoeg alle responderende ziekenhuizen het klassieke gips toegepast om te immobiliseren. Gerandomiseerde gecontroleerde trials kunnen het antwoord geven op de beste methode van immobilisatie bij een fractuur van het os scaphoideum.

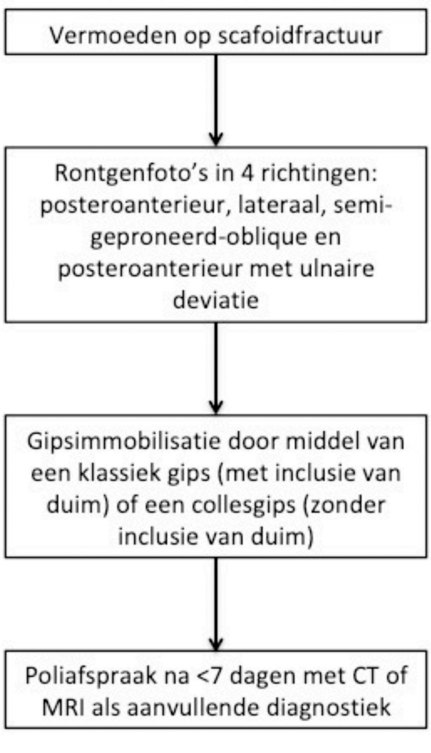
Bias

Lage respons is een bekend nadeel van enquêteonderzoek en kan een aanzienlijke bias veroorzaken. Een van de sterke punten van ons onderzoek is de hoge respons van 90%, waar vergelijkbare studies slechts een respons haalden van 16-73%(8-10). Door de hoge respons geeft deze studie een waarheidsgetrouw overzicht van de variabiliteit van de huidige klinische praktijk in Nederland.

Aan de hand van onze onderzoeksresultaten en de recentste literatuur hebben wij een stroomdiagram gemaakt met een suggestie voor het beleid bij patiënten met het vermoeden op een scafoïdfractuur (figuur 2). Dit zou een opzet kunnen zijn voor het ontwikkelen van een landelijke richtlijn.

CONCLUSIE

Tussen Nederlandse ziekenhuizen bestaat een grote variatie in het beleid bij het vermoeden op scafoïdfractuur. Hoewel de diagnostiek in de Nederlandse praktijk altijd veel wetenschappelijke aandacht heeft genoten(13, 14, 16, 18, 20), ontbreekt tot op heden een landelijke richtlijn. Verder onderzoek moet niet alleen als doel hebben om een evidence-based consensus te bereiken over beeldvormend onderzoek, maar ook over immobilisatie. Het is belangrijk een richtlijn te ontwerpen waarin vroege diagnostiek en adequate behandeling van occulte scafoïdfracturen gebaseerd worden op de recentste literatuur. Dit zal onnodige immobilisatie van patiënten voorkomen en mogelijk de kosten reduceren.



Figuur 1.

LEERPUNTEN

- Bij een klinisch vermoeden op een scafoïdfractuur heeft tot 16% van de patiënten met niet-afwijkende röntgenfoto's toch een fractuur.
- In verband met inadequate röntgenfoto's en het risico op complicaties, wordt de pols ook bij niet-afwijkende röntgenfoto's geïmmobiliseerd met een gips; het is niet bekend welke methode hier het geschiktst voor is.
- De keuze voor beeldvormend onderzoek bij een patiënt met een vermoeden op een scafoïdfractuur verschilt sterk tussen ziekenhuizen.
- Immobilisatie gebeurt meestal met een klassiek scafoïdgips, dit is met inclusie van de duim.
- Het ontbreekt tot op heden aan een richtlijn voor vroege diagnostiek en adequate behandeling bij het vermoeden op een scafoïdfractuur.

REFERENTIES

1. Hove LM. Epidemiology of scaphoid fractures in Bergen, Norway. Scand J Plast Reconstr Surg Hand Surg. 1999;33(4):423-6.
2. Kumar S, O'Connor A, Despois M, Galloway H. Use of early magnetic resonance imaging in the diagnosis of occult scaphoid fractures: the CAST Study (Canberra Area Scaphoid Trial). N Z Med J. 2005;11(118):1209.
3. Hunter JC, Escobedo EM, Wilson AJ, Hanel DP, Zink-Brody GC, Mann FA. MR imaging of clinically suspected scaphoid fractures. AJR American journal of roentgenology. 1997;168(5):1287-93.
4. Jenkins PJ, Slade K, Huntley JS, Robinson CM. A comparative analysis of the accuracy, diagnostic uncertainty and cost of imaging modalities in suspected scaphoid fractures. Injury. 2008;39(7):768-74.
5. Arora R, Gschwentner M, Krappinger D, Lutz M, Blauth M, Gabl M. Fixation of nondisplaced scaphoid fractures: making treatment cost effective. Prospective controlled trial. Arch Orthop Trauma Surg. 2007;127(1):39-46.
6. Dias JJ, Wildin CJ, Bhowal B, Thompson JR. Should acute scaphoid fractures be fixed? A randomized controlled trial. J Bone Joint Surg Am. 2005;87(10):2160-8.
7. Waitayawinyu T, McCallister WV, Katolik LI, Schlenker JD, Trumble TE. Outcome after vascularized bone grafting of scaphoid nonunions with avascular necrosis. J Hand Surg Am. 2009;34(3):387-94.
8. Brookes-Fazakerley SD, Kumar AJ-O, J. Survey of the initial management and imaging protocols for occult scaphoid fractures in UK hospitals. Skeletal radiology. 2009;38(11):1045-8.
9. Groves AM, Kayani I, Syed R, Hutton BF, Bearcroft PP, Dixon AK, et al. An international survey of hospital practice in the imaging of acute scaphoid trauma. AJR American journal of roentgenology. 2006;187(6):1453-6.
10. Kelly AM. Initial management of potential occult scaphoid fracture in Australasia. International journal of emergency medicine. 2010;3(1):45-7.
11. Bretlau T, Christensen OM, Edstrom P, Thomsen HS, Lausten GS. Diagnosis of scaphoid fracture and dedicated extremity MRI. Acta orthopaedica Scandinavica. 1999;70(5):504-8.
12. Dorsay TA, Major NM, Helms CA. Cost-effectiveness of immediate MR imaging versus traditional follow-up for revealing radiographically occult scaphoid fractures. AJR American journal of roentgenology. 2001;177(6):1257-63.
13. 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(8):971-5.
14. Beeres FJ, Hogervorst M, Rhemrev SJ, den Hollander P, Jukema GN. A prospective comparison for suspected scaphoid fractures: bone scintigraphy versus clinical outcome. Injury. 2007;38(7):769-74.
15. Beeres FJ, Rhemrev SJ, den Hollander P, Kingma LM, Meylaerts SA, le Cessie S, et al. Early magnetic resonance imaging compared with bone scintigraphy in suspected scaphoid fractures. J Bone Joint Surg Br. 2008;90(9):1205-9.
16. Rhemrev SJ, de Zwart AD, Kingma LM, Meylaerts SA, Arndt JW, Schipper IB, et al. Early computed tomography compared with bone scintigraphy in suspected scaphoid fractures. Clinical nuclear medicine. 2010;35(12):931-4.
17. Adey L, Souer JS, Lozano-Calderon S, Palmer W, Lee SG, Ring D. Computed tomography of suspected scaphoid fractures. J Hand Surg Am. 2007;32(1):61-6.
18. Mallee W, Doornberg JN, Ring D, van Dijk CN, Maas M, Goslings JC. Comparison of CT and MRI for diagnosis of suspected scaphoid fractures. J Bone Joint Surg Am. 2011;93(1):20-8.
19. Beeres FJ, Hogervorst M, Kingma LM, Le Cessie S, Coerkamp EG, Rhemrev SJ. Observer variation in MRI for suspected scaphoid fractures. The British journal of radiology. 2008;81(972):950-4.
20. Tiel-van Buul MM, Roolker W, Broekhuizen AH, Van Beek EJ. The diagnostic management of suspected scaphoid fracture. Injury. 1997;28(1):1-8.
21. Clay NR, Dias JJ, Costigan PS, Gregg PJ, Barton NJ. Need the thumb be immobilised in scaphoid fractures? A randomised prospective trial. J Bone Joint Surg Br. 1991;73(5):828-32.

22. Schramm JM, Nguyen M - Wongworawat MD, Kjellin I. Does thumb immobilization contribute to scaphoid fracture stability? Hand. 2008;Mar;3(1)(1558-9447 (Print)):41-3.
23. Petheram TG, Garg S, Compson JP. Is the scaphoid cast still alive? A survey of current UK practice in conservative management of scaphoid fractures. The Journal of hand surgery, European volume. 2009;34(2):281-2.
24. Chechik O, Rosenblatt Y. Management of clinically suspected scaphoid fractures: a survey of current practice in Israel. Isr Med Assoc J. 2009;11(4):225-8.
25. Cruickshank J, Meakin A, Breadmore R, Mitchell D, Pincus S, Hughes T, et al. Early computerized tomography accurately determines the presence or absence of scaphoid and other fractures. Emerg Med Australas. 2007;19(3):223-8.
26. Querellou S, Moineau G, Le Duc-Pennec A, Guillo P, Turzo A, Cotonea Y, et al. Detection of occult wrist fractures by quantitative radioscintigraphy: a prospective study on selected patients. Nucl Med Commun. 2009;30(11):862-7.
27. Ty JM, Lozano-Calderon S, Ring D. Computed tomography for triage of suspected scaphoid fractures. Hand (N Y). 2008;3(2):155-8.
28. Yin ZG, Zhang JB-K, S.L. - Wang, X. G. Diagnosing suspected scaphoid fractures: a systematic review and meta-analysis. Clinical orthopaedics and related research. 2010;468(3)(1528-1132 (Electronic)):723-34.
29. Foex B, Speake P, Body R. Best evidence topic report. Magnetic resonance imaging or bone scintigraphy in the diagnosis of plain x ray occult scaphoid fractures. Emerg Med J. 2005;22(6):434-5.
30. Kawamura K, Chung KC. Treatment of scaphoid fractures and nonunions. J Hand Surg Am. 2008;33(6):988-97.
31. Pillai A, Jain M. Management of clinical fractures of the scaphoid: results of an audit and literature review. European journal of emergency medicine : official journal of the European Society for Emergency Medicine. 2005;12(2):47-51.
32. Patel NK, Davies N, Mirza Z, Watson M. Cost and clinical effectiveness of MRI in occult scaphoid fractures: a randomised controlled trial. Emerg Med J. 2013;30(3):202-7.
33. Brooks S, Cicuttini FM, Lim S, Taylor D, Stuckey SL, Wluka AE. Cost effectiveness of adding magnetic resonance imaging to the usual management of suspected scaphoid fractures. British journal of sports medicine. 2005;39(2):75-9.
34. 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(1):169-76.

PART 2

CLINICAL EVALUATION OF SCAPHOID FRACTURES: WHEN IS THERE A CLINICAL SUSPICION?

CHAPTER 3

CLINICAL DIAGNOSTIC EVALUATION FOR SCAPHOID FRACTURES: A SYSTEMATIC REVIEW AND META-ANALYSIS

Mallee WH
Henny EP
van Dijk CN
Kamminga SP
van Enst WA
Kloen P

Journal of Hand Surgery – American Volume. September 2014

ABSTRACT

Purpose

To provide an overview of available clinical evaluation tests for scaphoid fractures and to compare their diagnostic accuracies.

Methods

We performed a systematic review of all studies assessing diagnostic characteristics of clinical evaluation in scaphoid fractures by searching Medline, Embase, Cochrane, and Cinahl databases. Only studies on clinical testing prior to radiographic evaluation and with acceptable reference standard for occult fractures were included. Thirteen relevant articles were analyzed that described a total of 25 tests.

Diagnostic characteristics of the tests were used to construct contingency tables. If possible, data were pooled and summary receiver operating characteristic curves were fitted.

Results

Anatomic snuff-box tenderness (ASB, 8 studies, 1164 patients) and longitudinal thumb compression (LTC, 8 studies, 961 patients) had sufficient data for statistical analyses. Sensitivity for ASB ranged from 0.87 to 1.00; for LTC, this was 0.48 to 1.00. Specificity of ASB ranged from 0.03 to 0.98; for LTC, this was 0.22 to 0.97. Due to considerable heterogeneity, pooled estimate points were not calculated. Other high sensitivity tests were scaphoid tubercle tenderness with sensitivity and specificity ranging from 0.82 to 1.00 and 0.17 to 0.57, respectively, and painful ulnar deviation ranging from 0.67 to 1.00 and 0.17 to 0.60, respectively. Three studies showed that combining tests increased the specificity and post-test fracture probability while maintaining high sensitivity.

Quality assessment showed high or unclear risk of bias and applicability concerns in reference standard and patient selection. Twelve study designs were prospective, one was retrospective.

Discussion

Anatomical snuff box tenderness was the most sensitive clinical test. The low specificity of the clinical tests may result in a considerable number of over treated patients. Combining tests improved the post-test fracture probability. This can be used to limit unnecessary immobilization, number of hospital visits, and use of imaging. The data presented herein may help to develop clinical prediction rules that could increase specificity without reducing sensitivity.

Level of evidence

Diagnostic, level II

INTRODUCTION

Diagnosis of suspected scaphoid fractures faces several challenges. Clinical suspicion is raised with tenderness in the anatomic snuffbox after a fall on the outstretched hand(1, 2). Initial radiographs in multiple views are sometimes ineffective in establishing a definitive diagnosis(3). Sixteen to 27% of patients with normal initial radiographs still has a fracture(4, 5). With the risk of nonunion and/or subsequent degenerative changes in mind(6-8), all suspected fractures are immobilized with a cast until definitive diagnosis is obtained. Approximately 4 out of 5 patients(4, 5) will therefore receive unnecessary immobilization, while a possible wrist sprain could be treated with a soft bandage. In addition, patients must pay additional visit(s) to the hospital to remove the plaster cast and to be further evaluated with radiographs, computed tomography, magnetic resonance imaging, or bone scintigraphy(3). This increases healthcare costs and time expended(9).

The pool of patients that consequently receives unnecessary diagnostic management could be reduced by lowering the number of false positives and raising the post-test fracture probability. This systematic review analyzed all adequately studied clinical tests for suspected scaphoid fractures. The main purpose was to depict the clinical tests with the highest diagnostic accuracy for detecting a scaphoid fracture in patients with wrist trauma.

METHODS

This systematic review is reported according to the 2009 Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist(10).

Literature search and study selection

A systematic literature search of Embase (Online Appendix 1), Medline, CENTRAL, and Cinahl was performed on April 27, 2012. The general search terms were 'scaphoid OR navicular fracture' AND 'clinical evaluation OR physical examination'. Additionally, the reference lists of relevant articles were hand searched, and the related article function in Pubmed was used. No language or quality restrictions were applied. Non-English studies were included if translation of the full article was possible.

Studies were eligible if they included patients presenting to the emergency department or outpatient clinic following wrist trauma, but prior to knowledge of initial radiographic assessment: evaluated one or more clinical tests of the wrist with the presence or absence of a scaphoid fracture as main outcome; used an acceptable reference standard to identify occult fractures as well; and the study provided sufficient data for constructing a 2-by-2 contingency table of the index test(s). Studies were excluded if the index test was performed with knowledge of initial radiographs and if they included a non-consecutive series of patients.

Two authors (WHM & EPH) independently screened the titles and abstracts of all the studies. Full reports were obtained and examined for all citations that were likely to meet inclusion criteria. Any disagreement was solved by a third author (SPK).

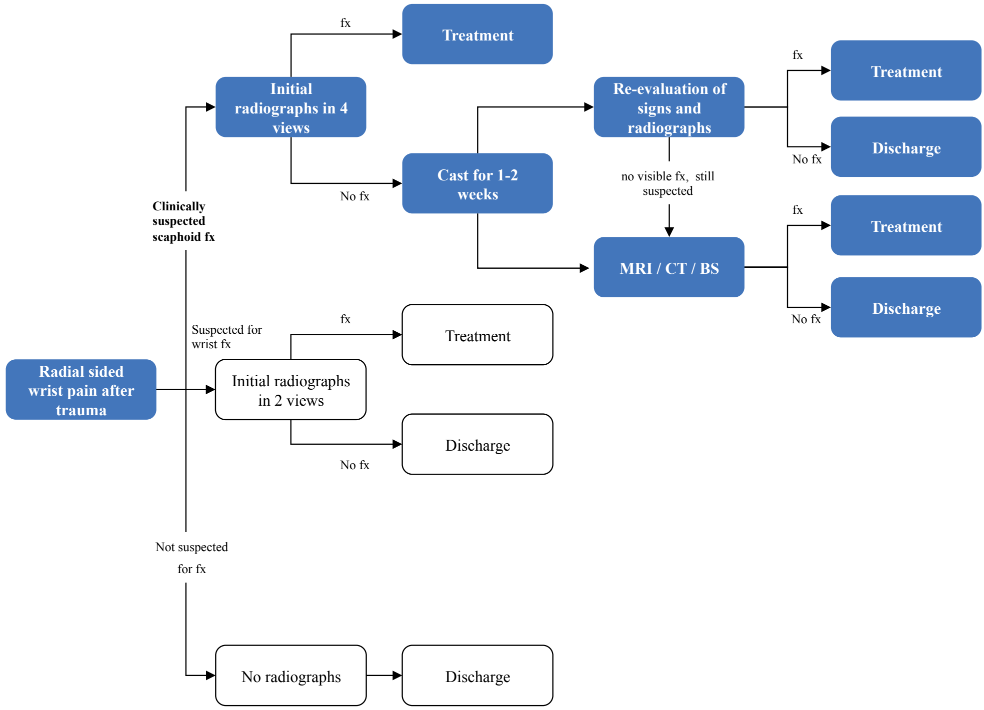


Figure 1. Patient flow emergency department
Fx = fracture

Reference standard

Ongoing discussion on the best available test to establish the presence or absence of a true scaphoid fracture makes the use of multiple reference standards a necessity(3, 11, 12). We accepted the following reference standards to establish if there was a true fracture of the scaphoid (Figure 1, blue section).

Studies only using repeated clinical evaluation were inadequate in the quality assessment and were excluded. Studies that only used initial radiographs and no reference standard for suspected fractures were not included since occult fractures play an important role in diagnostic management.

Data extraction and methodological quality assessment

Two reviewers (WHM & EPH) independently collected the following data for each individual study: year of publication, study design, participants (total, mean age, number of males), mechanism of injury, index test(s), reference test, prevalence of scaphoid fracture within the study population, duration of follow-up (period between test and injury), and data for constructing 2-by-2 tables by using a specially designed extraction form.

The methodological quality of the included studies was assessed by using Quality Assessment of Diagnostic Accuracy Studies criteria 2 (Online Appendix 2)(13). This tool allows for transparent evaluation of the design and conduct of the included studies. The items (risk of bias and applicability concerns) were scored low, high, or unclear. Two authors independently assessed the quality (WHM & EPH). Any disagreement was solved by a third author (SPK). Results were presented in a table showing the individual scores per item and in a summary graph.

Data analysis and synthesis

Diagnostic performance characteristics of all index tests were calculated. The information of the 2-by-2 tables was used to calculate the sensitivity and specificity for each study. All studies describing the same test were included per comparison. The accuracy of every test was calculated regardless of any threshold including their likelihood ratios. If warranted, the likelihood ratio was used to calculate post-test probabilities according to the Centre for Evidence Based Medicine, Oxford, England. A pre-test fracture prevalence, calculated from literature, of 32% was applied (Table 2). Variation in threshold is highly likely if no explicit numerical cut-off point can be defined. In this review, definitions of a positive test were based on judgment rather than measurement (e.g. the amount of pressure given to elicit pain in the anatomical snuffbox). The meta-analyses were executed using the bivariate model. This model uses a random effects approach for both sensitivity and specificity allowing for heterogeneity beyond chance. It also integrates the possibility of differences in precision between how the sensitivity and specificity have been measured in every study by allowing larger studies more weight in the analysis. This method ensures a more equal distribution of study weights. It allows analysis of the diagnostic odds ratio but separately also of the sensitivity and specificity(14). The bivariate model requires 5 parameters to be estimated in the model and therefore requires a minimum of 5 studies for analysis(15). This is different than for a standard random effect model. The parameters estimated from the bivariate model produced a hierarchical summary receiver operating characteristic curve, which represented the change in accuracy according to the change in thresholds.

Heterogeneity is to be expected in any review of diagnostic accuracy. The magnitude of observed heterogeneity was depicted graphically by visual examination of the scatter of points. We addressed patient selection methods as possible source for heterogeneity in a subgroup with a stratified analysis of those solely clinically suspected versus those clinically suspected with normal initial radiographs).

RESULTS

Literature search and study selection

After removal of duplicates, 2072 references were screened, and 147 full articles were assessed for eligibility (Figure 2). The most frequent reason to exclude an article was the lack of any reference standard (48 exclusions) to detect occult fractures. Thirteen studies were included for this review reporting on 25 different index tests (Table 2)(16-28).

Table 1. Study specifications

Study	Year	Number of Patients /(Male)	Study design	Index test(s)	Reference test(s)	Duration of Follow-Up	Mechanisme of Injury	Prevalence No. (%)	Mean age (Range, y)	Mean age of scaphoid fractures
Waeckerle et al.	1987	85/(63)	Prospective	LTC, ASB, PRS	RG	6 months 2 weeks (x-ray)	FOOSH	40 (47%)	26 (NR)	NR
Powell et al.	1988	73/(NR)	Prospective	PUD	RG	3 months	NR	20 (27%)	NR (NR)	NR
Chen et al.	1989	52/(41)	Retrospective	LTC	RG	2 weeks	NR	13 (25%)	NR (11-85)	NR
Freeland et al.	1989	246/(NR)	Prospective	ASB, STT	RG	2 weeks	NR	30 (12%)	NR (10-65)	26
Mehta et al.	1990	90/(NR)	Prospective	ASB, ASB-s	RG	2 weeks	FOOSH Sports	65 (72%)	NR (NR)	NR
Esberger et al.	1994	99/(54)	Prospective	LTC	CE, RG, BS	2 weeks	NR	44 (44%)	NR (10-74	NR
Waizenegger et al.	1994	64/(NR)	Prospective	LTC, ASB, ASB-s, MI, CS, H, PUD, PRD, WT, PRP, PRS	RG, BS	2 weeks 6 weeks	FOOSH	23 (36%)	NR (NR)	NR
Grover et al.	1996	221/(NR)	Prospective	LTC, ASB, STT, S	RG	10 days	N.R	29 (13%)	NR (NR)	28 Male 36 Female
Parvizi et al.	1998	215/(112)	Prospective	LTC, ASB, STT, RMT	RG, BS	2-6 weeks	N.R	56 (26%)	36 (16-79)	28
Mody et al.	2007	58/(NR)	Prospective	ASB, ASB-s, MI, PSF, R	RG	3 weeks	FOOSH	13 (22%)	NR (16-61)	31.5
Unay et al.	2009	41/(29)	Prospective	TIFP, PAbT, PRD, PUD, LTC, PWF, PWE, PG, PPF, PSF	MRI	N.R	FOOSH	12 (29%)	29 (9-50)	NR
Rhemrev et al.	2010	78/(40)	Prospective	ASB-s, MI, H, FEWG, SPSM, GS, PDR	MRI, RG, BS	2 weeks 6 weeks	Fall standing/ height, Sports	13 (17%)	41 (>18)	42

Table 1. (continued)

Study	Year	Number of Patients /(Male)	Study design	Index test(s)	Reference test(s)	Duration of Follow-Up	Mechanisme of Injury	Prevalence No. (%)	Mean age (Range, y)	Mean age of scaphoid fractures
Duckworth et al.	2012	223/(123)	Prospective	LTC, ASB, TIFP, STT, RMT, PUD, PRD	RG, CT, MRI	2 - 6 weeks	FOOSH Sports	62 (28%)	33 (13-95)	27

NR = Not Reported, CE = Clinical Evaluation, RG = Radiographs, BS = Bone Scintigraphy, PSF = Previous Scaphoid Fracture, MI = Mechanisme of Injury, S = swelling, R = local Redness, H = Haematoma snuffbox, ASB-s = Anatomic Snuffbox Swelling, ASB = Anatomic Snuffbox Tenderness, STT = Scaphoid Tubercle Tenderness, LTC = painful Longitudinal Thumb Compression, CS = Clamp Sign, WT = shift (Watson) Test, TIFP = Thumb-Index Finger Pinch, PUD = Painful Ulnar Deviation, PRD = Painful Radial Deviation, PDR = Painful Distal Radius, PPF = Painful Pronation Forearm, PRP = Pain with Resisted Pronation, PSF = Painful Supination Forearm, PRS = Pain with Resisted Supination, PWF = Painful Wrist Flexion, PWE = Painful Wrist Extension, RMT = decreased Range of Motion Thumb, PAbT = Painful Abduction Thumb, FEWG = Flexion and Extension of the Wrist measured using a hand-held Goniometer, SPSM = Supination and Pronation Strength Measured using a custom device, PG = painful Power Grip, GS = Grip Strength

3

Table 2. Diagnostic accuracy per index test

Test	Number of studies	Number of subjects	Sensitivity range	Specificity range	LR+ range	LR- range
Anatomical Snuff Box (ASB) Tenderness	8	1164	0.87 - 1.00	0.03 - 0.98	1.01 - 45.0	0.00 - 0.87
Longitudinal Thumb Compression (LTC)	8	961	0.48 - 1.00	0.22 - 0.97	0.90 - 38.0	0.00 - 1.35
Scaphoid Tubercle Tenderness	4	879	0.82 - 1.00	0.17 - 0.57	1.20 - 2.01	0.00 - 0.46
ASB Swelling	4	276	0.61 - 0.77	0.37 - 0.72	1.15 - 2.64	0.36 - 0.82
Painful Ulnar Deviation	4	394	0.67 - 1.00	0.17 - 0.60	1.02 - 2.52	0.00 -1.01
Painful Radial Deviation	3	316	0.67 - 0.90	0.31 - 0.42	1.01 - 1.55	0.23 -0.98
Decreased Range of Motion Thumb	2	412	0.65 - 0.66	0.38 - 0.59	1.04 - 1.63	0.57 - 0.94
Hematoma	2	130	0.22 - 0.46	0.76 - 0.77	0.90 - 2.00	0.70 - 1.03
Thumb Index Finger Pinch	2	264	0.75 - 0.79	0.44 - 0.76	0.90 - 2.00	0.70 - 1.03
Pain with Resisted Supination	2	137	0.83 - 1.00	0.38 - 0.98	1.33 - 45.0	0.00 - 0.46
Pain with Resisted Pronation	1	52	0.65	0.24	0.86	1.44
Wrist Extension < 50%	1	78	0.85	0.59	2.07	0.25
Grip strength ≤ 25 %	1	78	0.92	0.34	1.39	0.24
Supination strength ≤ 10%	1	78	0.85	0.77	3.70	0.19
Pronation strength ≤ 10%	1	78	0.69	0.65	1.97	0.48
Watson Test	1	52	0.83	0.31	1.20	0.56
Clamp Sign	1	52	0.26	0.79	1.26	0.93
Painful Supination Forearm	1	41	0.76	0.50	1.52	0.48
Painful Pronation Forearm	1	41	0.79	0.58	1.90	0.35
Painful Wrist Extension	1	41	0.72	0.60	1.81	0.46
Painful Wrist Flexion	1	41	0.71	0.50	1.43	0.57
Painful Power Grip	1	41	0.67	0.20	0.83	1.67
Painful Abduction Thumb	1	41	0.73	0.50	1.45	0.55
Local Redness	1	58	0.23	0.67	0.69	1.15
Painful Distal Radius	1	78	0.62	0.57	1.43	0.68

Study characteristics and quality assessment

The percentage of males averaged 58%. The average prevalence of a true scaphoid fracture was 32%. Mean age of patients with a scaphoid fracture was 29 years, but this number was derived from only 5 studies. A fall on outstretched hand, either during sports or while standing, was the only reported trauma mechanism (Table 2).

All studies had some risk of bias or applicability concerns (Table 3). Most risks of bias and applicability concerns were found in the selection of patients. This was mainly caused by

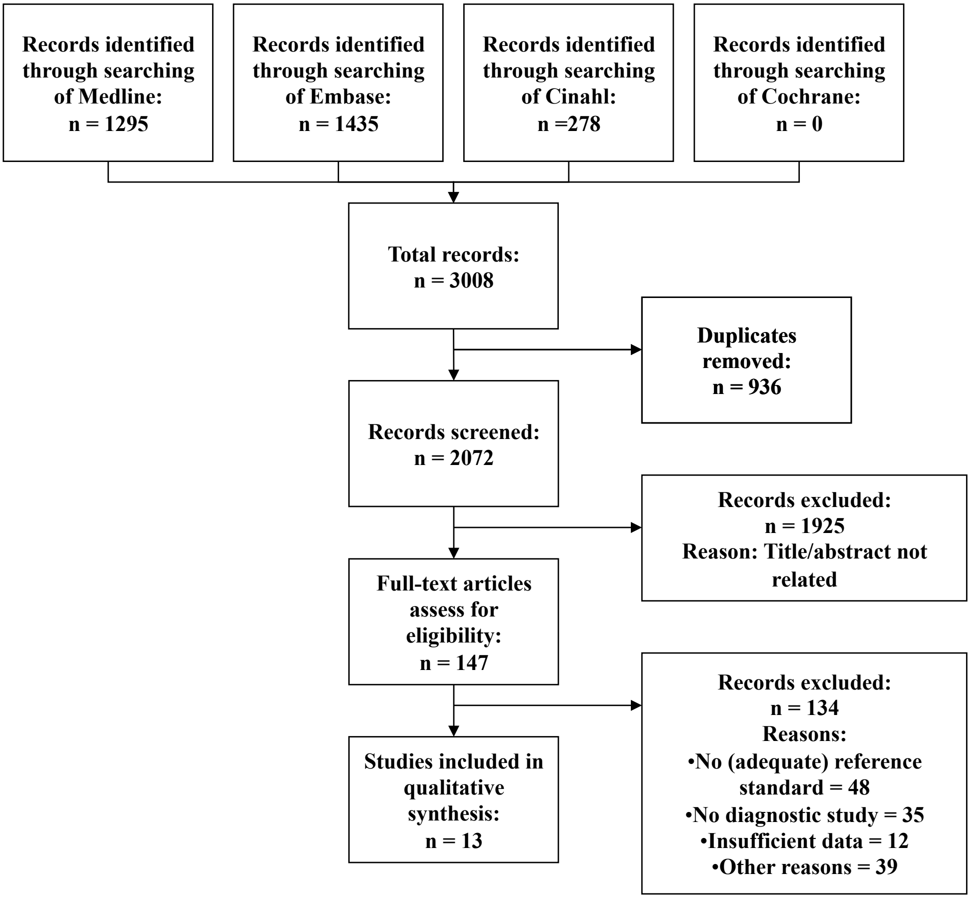


Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.

inclusion of skeletally immature patients. Since several studies used follow-up radiographs at 2 weeks, there were concerns in scoring the reference standard.

Index Tests

The anatomical snuff box (ASB) tenderness and the longitudinal thumb compression (LTC) tests were both evaluated in 8 studies ensuring power to perform statistical analyses. The summary receiver operating characteristic curves showed that sensitivity was higher for anatomical snuff box (ASB) tenderness. For ASB tenderness, 8 papers studying 1164 patients were included. The sensitivity was relatively homogeneous with a range between 0.87 and 1.00. The specificity ranged between 0.03 and 0.98 indicating high heterogeneity. For LTC, 8 papers studying 961 patients were included. The test had heterogeneous results for sensitivity (range between 0.48 and 1.00) and specificity (range between 0.22 and 0.97) (Figure 3). Because of the high heterogeneity we refrained from calculating pooled estimate points. Two studies 25, 26

Table 3. Quality Assessment of Diagnostic Accuracy Studies 2. The following table summarizes the Quality Assessment of Diagnostic Accuracy Studies -2 and lists all signaling, risk of bias and applicability rating questions.

	Risk of bias				Applicability concerns		
Study	Patient selection	Index test	Reference standard	Flow & timing	Patient selection	Index test	Reference standard
Waeckerle et al.	-	-	+	+	-	-	?
Powell et al.	-	-	?	+	?	-	?
Chen et a.l	+	?	+	+	+	-	?
Freeland et al.	-	-	-	+	+	-	?
Metha et al.	-	?	?	+	?	-	?
Esberger et al.	-	-	?	+	+	-	+
Waizenegger et al.	+	?	?	-	?	-	-
Grover et al.	-	-	-	+	+	-	?
Parvizi et al.	-	-	-	-	+	-	?
Mody et al.	-	+	?	+	+	-	-
Unay et al.	+	?	?	-	+	+	-
Rhemrev et al.	+	-	?	-	+	+	-
Duckworth et al.	-	?	?	+	+	-	?

All criteria were scored low (-), high (+) or unclear

investigating different tests included patients with clinical suspicion but who had normal initial radiographs. Power to perform a stratified analysis for the relevant analyses was too low.

The other 23 tests were evaluated in too few studies (1 to 4) to estimate all 5 parameters needed to fit the analysis model. Scaphoid tubercle tenderness (STT) showed reasonably consistent and high sensitivities in 4 papers studying 879 patients, with a range of 0.82 to 1.00. This was also noticed for painful ulnar deviation (PUD) in 4 papers studying 394 patients where sensitivity ranged from 0.67 to 1.00.

Rhemrev et al. studied grip strength and range of motion (supination/pronation and flexion/extension). Both were significantly decreased in patients with a fractured scaphoid compared to those without fractured scaphoids(25).

Three studies combined multiple tests to improve diagnostic accuracy. One study combined ASB tenderness, STT, and LTC to reach a sensitivity of 1.00 and a specificity of 0.74(23). This would result in a post-test fracture probability of 64%. Two studies described diagnostic accuracy for combining tests but also used predicted probabilities. Multivariate logistic regression analysis was performed to calculate the optimal combination of predictors. Predicted probabilities of a scaphoid fracture derived from regression coefficients. Duckworth et al. presented multiple clinical prediction rules with different fracture probabilities and accompanying diagnostic accuracies. The highest fracture probability (91%) was when the patient was male, had a sports injury, PUD within 72 hours after injury, and STT on re-evaluation at 2 weeks. The sensitivity was 0.82 and specificity was 0.89(17). Rhemrev et al. found a fracture probability of 97% when

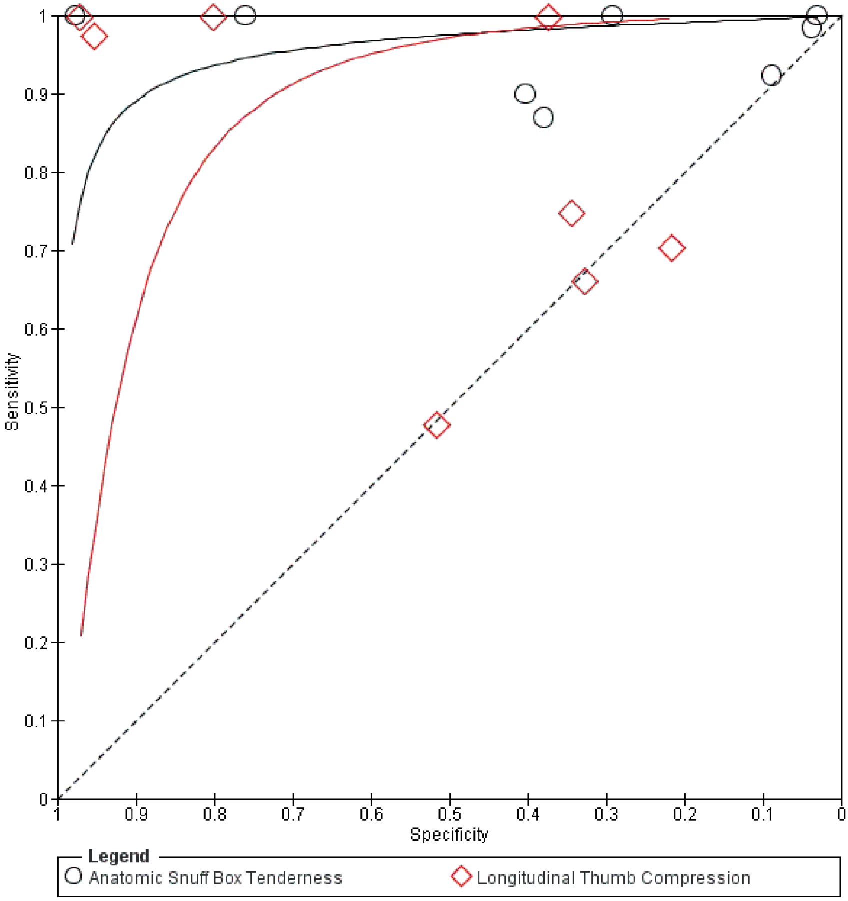


Figure 3. Summary receiver operating characteristic curves of anatomical snuff box tenderness and longitudinal thumb compression.

combining a previous fracture of either one of the wrists, supination strength of ≤ 10%, and wrist extension of <50% compared to the contralateral side.

DISCUSSION

This systematic review and meta-analysis of clinical evaluation tests in scaphoid fractures showed several important findings. With a sensitivity between 0.87 and 1.00, ASB tenderness seems an adequate test to incorporate in clinical evaluation after wrist trauma. However, using only this clinical test, up to 13% of scaphoid fractures can be missed. For LTC, most studies presented high sensitivity rates, however these results were more heterogeneous (range between 0.48 and 1.00). Moreover, for both ASB tenderness and LTC, the specificity was too heterogeneous to pool results. The heterogeneity found within these studies could result from different clinical experience of the clinicians. This was probably caused by implicit threshold variation, which is likely if there is no explicit numerical cut-off point and diagnosis is based on judgment rather than measurement. Unfortunately, specific information to test this hypothesis was not provided in the individual studies. In future studies, clear instructions on how the tests were/should be performed are likely to reduce the heterogeneity.

As it appears that most tests have high sensitivity but lack specificity, a next step can be to combine 2 or more tests. Parvizi studied this technique in scaphoid trauma and described a considerable improvement of specificity (up to 74%) when combining ASB tenderness with STT and LTC, while sensitivity remained 100%(23). This is assuring since fractures will not be missed and it leads to a reasonable post-test fracture probability of 64%. Duckworth et al. combined more predictive signs and developed several clinical prediction rules with fracture probabilities of up to 91% if 3 or 4 factors were positive(17). Their most accurate test was ASB pain with PUD with a sensitivity of 1.00 and a specificity of 0.45. This test might be a promising supplement to daily practice as it seems to be accurate and has less chance of threshold variation. These clinical prediction rules will have a substantial impact on further diagnostic management and result in a more cost-effective process.

The limitations of this review are visualized in the Quality Assessment of Diagnostic Accuracy Studies 2 results (Table 3). The included articles suggest a moderate or low methodological quality. This is mainly based on patient selection and the use of a reference standard. Patient selection was prone to inappropriate exclusions giving rise to applicability concerns. The more recent studies use reference standards that are believed to be more accurate (magnetic resonance imaging , computed tomography, bone scintigraphy, 6 week follow-up radiographs), where older studies use repeated radiographs within 2 weeks. This has already been shown to be less accurate(29).

Each systematic review has the inherent risk of publication bias. As the mechanisms of publication bias are currently not well understood for diagnostic accuracy studies, there are currently no assessment tools available to investigate this risk other than graphical interpretation.

Unfortunately, the number of studies and patients limited analyses for 23 out of 25 index tests. Further research is needed to perform meta-analyses of these tests. However, a more realistic response to this diagnostic challenge might be to combine several tests rather than to find one that is 100% accurate.

In wrist trauma, high sensitivity tests are essential to avoid missing a scaphoid fracture and its complications. Combining several tests by creating a clinical prediction rule, such as the Ottawa Ankle Rules(30), can limit the number of initial radiographs and also the need for additional imaging. These rules are designed to avoid missing fractures (high sensitivity) and raise the post-test probability of the disease.

Our results can be used as the basis to develop 1 clinical prediction rule that is easy to implement. This means that all patients that present with wrist trauma must be included and, prior to performing initial radiographs, the following features are determined: sex, trauma mechanism, presence of swelling in the ASB, tenderness in the ASB, painful LTC, STT and PUD. When initial radiographs appear normal, all patients must receive an adequate reference standard: computed tomography, magnetic resonance imaging, bone scintigraphy, or 6-week follow-up radiographs. Studying these 7 aspects will result in a trial that will have to include at least 70 true scaphoid fractures using both the outcome of initial radiographs and a follow-up reference standard(31).

REFERENCES

1. Duckworth AD, Ring D, McQueen MM. Assessment of the suspected fracture of the scaphoid. J Bone Joint Surg Br. 2011;93(6):713-9.
2. Beeres FJ, Rhemrev SJ, Hogervorst M, den Hollander P, Jukema GN. [Scaphoid fractures: diagnosis and therapy]. Nederlands tijdschrift voor geneeskunde. 2007;151(13):742-7.
3. Yin ZG, Zhang JB, Kan SL, Wang XG. Diagnostic accuracy of imaging modalities for suspected scaphoid fractures: meta-analysis combined with latent class analysis. J Bone Joint Surg Br. 2012;94(8):1077-85.
4. Beeres FJ, Hogervorst M, den Hollander P, Rhemrev S. Outcome of routine bone scintigraphy in suspected scaphoid fractures. Injury. 2005;36(10):1233-6.
5. Jenkins PJ, Slade K, Huntley JS, Robinson CM. A comparative analysis of the accuracy, diagnostic uncertainty and cost of imaging modalities in suspected scaphoid fractures. Injury. 2008;39(7):768-74.
6. Ruby LK, Stinson J, Belsky MR. The natural history of scaphoid non-union. A review of fifty-five cases. J Bone Joint Surg Am. 1985;67(3):428-32.
7. Waitayawinyu T, McCallister WV, Nemechek NM, Trumble TE. Scaphoid nonunion. J Am Acad Orthop Surg. 2007;15(5):308-20.
8. Gelberman R.H. - Gross MS. The vascularity of the wrist. Identification of arterial patterns at risk. Clinical orthopaedics and related research. 1986(0009-921X (Print)).
9. Dorsay TA, Major NM, Helms CA. Cost-effectiveness of immediate MR imaging versus traditional follow-up for revealing radiographically occult scaphoid fractures. AJR American journal of roentgenology. 2001;177(6):1257-63.
10. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.
11. Mallee W, Doornberg JN, Ring D, van Dijk CN, Maas M, Goslings JC. Comparison of CT and MRI for diagnosis of suspected scaphoid fractures. J Bone Joint Surg Am. 2011;93(1):20-8.
12. Rhemrev SJ, de Zwart AD, Kingma LM, Meylaerts SA, Arndt JW, Schipper IB, et al. Early computed tomography compared with bone scintigraphy in suspected scaphoid fractures. Clinical nuclear medicine. 2010;35(12):931-4.
13. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011;155(8):529-36.
14. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58(10):982-90.
15. Macaskill P, Gatsonis C, Deeks J, Harbord M, Takwoingi Y. Analysing and presenting results. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. 2010:46-7.
16. Chen SC. The scaphoid compression test. J Hand Surg [Br]. 1989;14(3):323-5.
17. Duckworth AD, Buijze GA, Moran M, Gray A, Court-Brown CM, Ring D, et al. Predictors of fracture following suspected injury to the scaphoid. J Bone Joint Surg Br. 2012;94(7):961-8.
18. Esberger DA. What value the scaphoid compression test? Journal of hand surgery (Edinburgh, Scotland). 1994;19(6):748-9.
19. Freeland P. Scaphoid tubercle tenderness: a better indicator of scaphoid fractures? Archives of emergency medicine. 1989;6(1):46-50.
20. Grover R. Clinical assessment of scaphoid injuries and the detection of fractures. Journal of hand surgery (Edinburgh, Scotland). 1996;21(3):341-3.
21. Mehta M, Brautigan MW. Fracture of the carpal navicular--efficacy of clinical findings and improved diagnosis with six-view radiography. Ann Emerg Med. 1990;19(3):255-7.
22. Mody RA, Clift B. The "clinical scaphoid fracture." A prospective study of outcome. Journal of Orthopaedics. 2007;4(2):e15.
23. Parvizi J, Wayman J, Kelly P, Moran CG. Combining the clinical signs improves diagnosis of scaphoid fractures. A prospective study with follow-up. Journal of hand surgery (Edinburgh, Scotland). 1998;23(3):324-7.

24. Powell JM, Lloyd GJ, Rintoul RF. New clinical test for fracture of the scaphoid. Canadian journal of surgery Journal canadien de chirurgie. 1988;31(4):237-8.
25. Rhemrev SJ, Beeres FJ, van Leerdam RH, Hogervorst M, Ring D. Clinical prediction rule for suspected scaphoid fractures: A prospective cohort study. Injury. 2010;41(10):1026-30.
26. Unay K, Gokcen B, Ozkan K, Poyanli O, Eceviz E. Examination tests predictive of bone injury in patients with clinically suspected occult scaphoid fracture. Injury. 2009;40(12):1265-8.
27. Waeckerle JF. A prospective study identifying the sensitivity of radiographic findings and the efficacy of clinical findings in carpal navicular fractures. Ann Emerg Med. 1987;16(7):733-7.
28. Waizenegger M, Barton NJ, Davis TR, Wastie ML. Clinical signs in scaphoid fractures. Journal of hand surgery (Edinburgh, Scotland). 1994;19(6):743-7.
29. Tiel-van Buul MM, van Beek EJ, Broekhuizen AH, Nooitgedacht EA, Davids PH, Bakker AJ. Diagnosing scaphoid fractures: radiographs cannot be used as a gold standard! Injury. 1992;23(2):77-9.
30. Stiell IG, McKnight RD, Greenberg GH, McDowell I, Nair RC, Wells GA, et al. Implementation of the Ottawa ankle rules. JAMA : the journal of the American Medical Association. 1994;271(11):827-32.
31. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373-9.

CHAPTER 4

DETECTING SCAPHOID FRACTURES IN WRIST INJURY: A CLINICAL DECISION RULE

Mallee WH
Walenkamp MMJ
Mulders MAM
Goslings JC
Schep NWL

Submitted

ABSTRACT

The purpose of this prospective multicenter study was to develop and validate clinical decision rule for detecting scaphoid fractures that limit the number of unnecessary splint immobilizations and diagnostic follow-up, without increasing the risk of missing fractures.

First, a clinical prediction model for detecting scaphoid fractures in adult patients following wrist trauma was derived. Second, internal validation of the model was performed. Finally a clinical decision rule was formed. The predictors used were previously identified via a systematic review of literature. The outcome measure was the presence of a scaphoid fracture, diagnosed on either initial radiographs or during re-evaluation. After multivariate logistic regression analysis and bootstrapping, the regression coefficient for each significant predictor was calculated. A consecutive series of 893 patients with acute wrist injury was included, encompassing 68 scaphoid fractures. The final prediction rule incorporated sex, swelling of the anatomic snuffbox, tenderness in the anatomic snuffbox, painful ulnar deviation and painful axial thumb compression. Internal validation of the prediction rule showed a sensitivity of 97% and a specificity of 20%. Though external validation is needed, a 15% reduction in unnecessary immobilization and imaging could be achieved with a 50% decreased risk of missing a fracture compared with current clinical practice.

Level II evidence

INTRODUCTION

Scaphoid fractures are difficult to diagnose, especially in the acute setting. Due to the risk of complications, such as non-union and radiocarpal arthritis, patients with clinically suspected scaphoid fractures with normal radiographs are initially immobilized with a splint until further diagnostics are performed. On average, 80% of these patients do not have a scaphoid fracture(1), resulting in substantial overtreatment (e.g. immobilization), unnecessary follow-up imaging and substantial impact on work. Clinical assessment of possible scaphoid fractures in the Emergency Department (ED) is limited due to a lack of evidence supporting adequate tests.(2)

Improving the clinical selection of patients that require imaging and immobilization is warranted and a well-designed clinical decision rule could be the solution. Combining several clinical tests such as tenderness in the anatomic snuffbox and painful longitudinal thumb compression tests has already proven to increase the diagnostic accuracy of clinical assessment of scaphoid injury.(3-5) However, these studies were either underpowered, of uncertain methodology or were impractical for implementation in daily practice.

The aim of this study was to develop and validate an easy to use clinical decision rule, applicable in the ED that limits the number of unnecessary splinting and diagnostic follow-up in suspected scaphoid injury, without increasing the risk of missing a fracture.

METHODS

Study design

This study was part of a comprehensive research project, the Amsterdam Wrist Rules. This study included all wrist injuries to identify predictors for a distal radius or a scaphoid fracture.(6) A prospective multicenter study was performed, consisting of three components: (1) derivation of a clinical prediction model for detecting scaphoid fractures in patients following wrist trauma; (2) internal validation via bootstrapping and (3) design of a clinical decision rule. The study was conducted at the Emergency Departments of five Dutch hospitals from November 11 2010 to June 25 2014. The participating hospitals included one academic hospital and four regional teaching hospitals. Our Institutional Review Board approved this study without the need for informed consent. The trial was registered at the Dutch Trial Registration prior to start of inclusion (NTR 2544, www.trialregister.nl)

The entire dataset of the Amsterdam Wrist Rules comprised of a consecutive series of all adult patients (≥18 years) presenting to the Emergency Department (ED) with acute wrist injury (within three days after the initial trauma). For this trial, all patients that were suspected for a scaphoid fracture according to the treating (ED) physician were included. In addition, the entire Amsterdam Wrist Rules dataset was searched for possible missed scaphoid fractures. Patients were excluded if radiographs were performed prior to clinical evaluation. All patients received radiographs (one postero-anterior, one true lateral and if suspected for a scaphoid fracture: one semipronated oblique and posteroanterior view of the wrist in ulnar deviation); if any previous initial treatment was started in another hospital; if evaluation was performed by a nurse or general practitioner/house doctor; in case of multi-trauma or severe pain preventing examination; and any cognitive disorders limiting accurate response to questions. All patients with a true scaphoid fracture (both initially visible and occult) and all patients without a scaphoid fracture were divided into two groups for comparison of characteristics (Figure 1: Flowchart).

ED-physicians or Residents in (Orthopedic) Trauma clinically evaluated all patients by using a specially designed Case Record Form (CRF) prior to initial radiographs. Since knowledge of anatomy and tests was essential, education was given by presentations and laminated descriptive sheets to the residents and ED-physicians prior to the study initiation (Appendix). Both demographics and clinical tests were implemented in the CRF (Table 1). These variables were based on a recent review(2): Age, sex, presence of swelling in the anatomic snuff box (ASB), tenderness in the ASB, painful longitudinal thumb compression, scaphoid tubercle tenderness and pain over scaphoid with ulnar deviation. Definitive management was not interfered by the outcome of the CRF. An immobilizing cast or splint was given to those patients who were clinically suspected by discretion of the ED physician. The data on the CRFs were extracted by two researchers.

The primary outcome measure was the presence of a scaphoid fracture, diagnosed on initial radiographs, during re-evaluation after one to two weeks with repeated clinical evaluation (painless anatomical snuffbox and no pain with longitudinal thumb compression) and/or repeated radiographs or on additional imaging (MRI or CT). A fissure and an avulsion were classified as a fracture. The attending orthopaedic trauma surgeon and/or a resident in orthopaedic or trauma surgery and a radiologist evaluated the images during the trauma meeting the following day, they received normal clinical information and were blinded to the content of the Case Record Forms.

Sample size & Statistical analysis

A common rule of thumb to determine the sample size of the development of a prediction model is ten events (true scaphoid fractures) per variable.(7) With 7 variables, the inclusion of 70 patients with a true scaphoid fracture was required.

For efficient statistical analysis (8-10), we used imputation techniques to impute the missing values (*aregImpute* function from the *Hmisc* library, R, version 3.0.1.). For each variable

Table 1. Predictors and tests

Predictor	Outcome
Sex	Y / N
Age (per year)	Continuous
Swelling in ASB	Y / N
Tenderness in ASB	Y / N
Painful scaphoid with Longitudinal Thumb Compression	Y / N
Scaphoid tubercle tenderness	Y / N
Pain with ulnar deviation	Y / N

containing missings, the *aregImpute* package draws values from a random sample from the non-missing values with replacement. Using this data, aregImpute fits a flexible model that predicts the missing target variable while finding its optimum transformation. Each missing variable is then imputed with the observed value whose predicted transformed value is closest to the predicted transformed value of the missing variable. We considered an imputation model that included all dichotomous variables. The set of first imputations was used for the analyses.

Descriptive statistics were performed to summarize baseline characteristics. Categorical variables were presented as frequencies and percentages. Normality of the data was assessed by visually inspecting the normality plots. Parametric data was presented as mean and standard deviation (SD) and non-parametric data was presented as median and interquartile range [IQR]. Differences in patient characteristics between the groups with and without a scaphoid fracture were calculated using the Mann-Whitney U test for continuous data and the Chi-square test for categorical data. For each predictor, the sensitivity and specificity were calculated according to standard formulas.

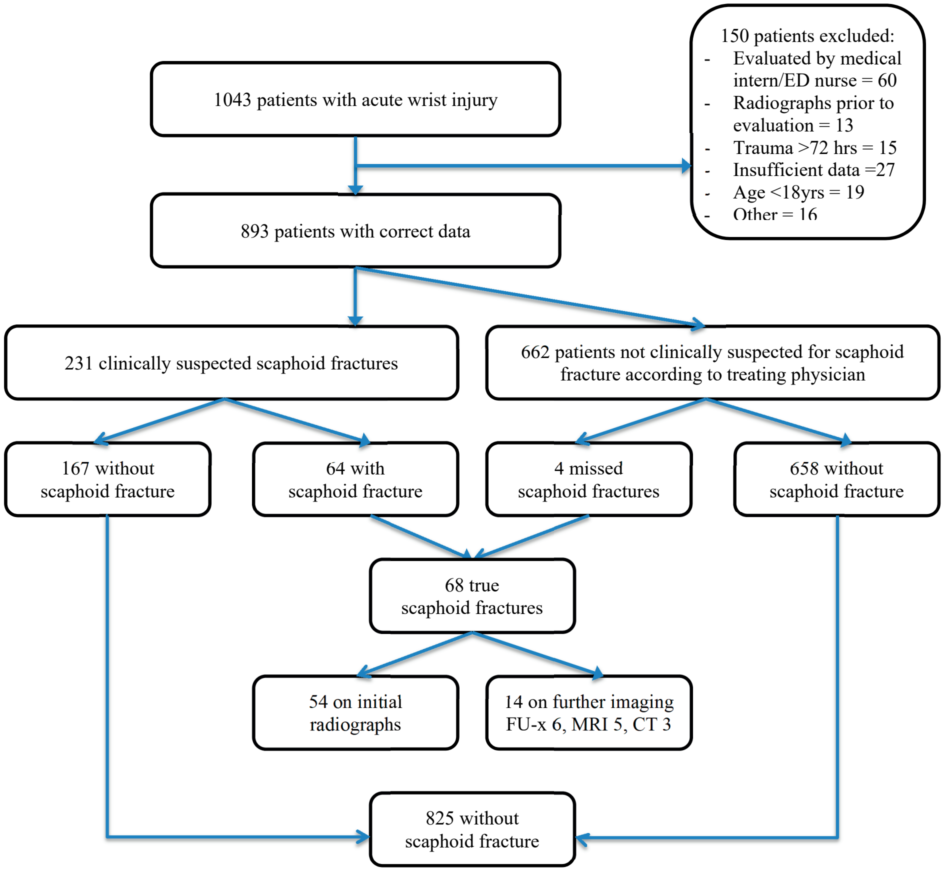


Figure 1. Flowchart of patient in- and exclusion

Model development and internal validation

We derived a clinical prediction model for scaphoid fractures using data on all patients with a clinically suspected fracture including missed fractures, enrolled in the study. A multivariate logistic regression model with all 7 potential predictors was fit. This full model was reduced using a stepwise backward elimination process based on a liberal p-value of 0.15.(11) The coefficient determines the effect of that predictor on the probability of a true scaphoid fracture. The coefficient of each variable represents the amount of change in the probability of a scaphoid fracture. A positive coefficient increases and a negative coefficient decreases the probability of a fracture. To estimate the internal validation of performance we used bootstrapping (500 replications). Bootstrapping was used to quantify the optimism of the prediction model by mimicking the process of sampling from the underlying population: the difference between performance in the bootstrap sample and performance in the original sample. A shrinkage factor, also obtained by bootstrap validation, was used for multiplication of the regression coefficients.(12, 13)

To estimate the ability of the model to discriminate between patients with and without a fracture, we calculated the Areas under the Receiver Operating Characteristics Curve (AUC). The AUC ranges from 0.5 to 1, with a higher score indicating more accurate predictions. The model was also evaluated for their agreement between predicted fractures and observed fractures. This is otherwise known as the model calibration and was assessed by plotting the predicted probability of a fracture and the observed frequency of fractures. The ideal slope of such a plot is 1, indicating perfect agreement between observed and predicted risks.(11)

Clinical decision rule

A clinical prediction *model* provides an estimated risk of a certain outcome. A clinical decision *rule* goes one step further and links a recommendation to the predicted risk. In this study, the recommendation would be to request a radiograph yes or no. If yes, immobilization with a cast/splint is inevitable. A clinical decision rule therefore requires a cut-off value for the predicted probability of a fracture to classify patients as low or high risk (or recommend radiograph yes or no). We decided beforehand to select a cut-off value at which the sensitivity of the rule would not drop below 95% to minimize the risk of missing fractures.

To assess the effect of the rule, number of overtreatment and undertreatment was calculated by applying the rule to the patients that were presumed clinically suspected for a scaphoid fracture by discretion of the ED physician (current clinical practice). The number of missed fractures and number of overtreatment after the rule was applied was compared to current clinical practice.

RESULTS

Patient characteristics

A total of 1043 adult patients with acute wrist injury were included in this study; 893 patients were eventually eligible for analysis. Sixty-eight patients (7.6%) were diagnosed with a scaphoid fracture, 54 patients (79%) during initial presentation and 14 patients (21%) during follow-up (radiographs six, MRI five, CT three) (Figure 1). Patients with a scaphoid fracture were significantly younger (p= 0.001) and males were overrepresented (p< 0.001), compared to patients without a scaphoid fracture. For patients characteristics see Table 2. A fall on the outstretched hand (FOOSH) was the trauma mechanism in 66% of the fractures.

At the discretion of the physician, 231 patients were clinically suspected for a scaphoid fracture, this group included 64 scaphoid fractures. Four scaphoid fractures (5.9%) were missed and did not receive initial immobilization due to lack of clinical signs (e.g. no anatomical snuffbox tenderness and no pain during axial compression of the thumb). These patients were discharged with pressure bandage and presented themselves within one to four weeks after ED presentation with persistent wrist pain. The missed fractures were 3 distal scaphoid fractures (Herbert type A1-2) that received cast immobilization and one complete waist fracture (Herbert type B2) that did not unite and received open reduction and screw fixation after 4 months.

There were no missing values for age, sex and tenderness in the ASB. For swelling of the ASB (4 missing), scaphoid tubercle tenderness (6 missing), painful ulnar deviation (1 missing) and painful longitudinal thumb compression (6 missing), missing data were imputed accordingly.

Results of individual tests

Table 3 shows the diagnostic accuracy, coefficients and odds ratios for each individual predictor. Diagnostic accuracy of tenderness in the ASB showed a sensitivity of 0.71 and a specificity of 0.25. For longitudinal compression of the thumb, sensitivity and specificity were 0.92 and 0.56 respectively. This test had the biggest effect on the clinical prediction rule since its coefficient was the highest (0.8544). Pain with ulnar deviation resulted in a sensitivity of 0.82 and specificity of 0.34.

Model development

The final clinical decision rule included five variables: sex, swelling of the ASB, tenderness of the ASB, painful ulnar deviation and painful longitudinal compression of the thumb (Table 3). The area under the curve (AUC) was 0.72 (95% CI: 0.65 – 0.78), after correcting for optimism by bootstrapping. The calibration of the model was 1 (95% CI: 0.59 – 1.40) indicating perfect agreement between predicted and observed fractures. The final formula for calculating the predicted chance of a true scaphoid fracture is shown in Table 4.

Table 2. Patient characteristics

	All patients N = 893	**Scaphoid fractures N = 68**	**Non scaphoid fractures N = 825**	**p-value fracture vs non fracture**
Median age (IQR)	50 (31-63)	35 (23.1 – 58.5)	50 (32.6 – 63.7)	0.001
Male	40%	62%	38.2%	<0.001
Trauma to dominant side	48%	45,9%	48,2	0.731

IQR: interquartile range

Table 3. Individual test results

Predictors	Sensitivity	Specificity	Coefficients (95% CIs)	P-values	Odds ratios (95% Cis)
Age	n.a.	n.a.	0.0035 (-0.01296 – 0.01996)	0.6793	n.a.
Sex*	n.a.	n.a.	0.8347 (0.2306 – 1.4388)	0.0068	2.3041 (1.26 – 4.22)
Swelling ASB*	0.5224	0.6707	0.6598 (0.0063 – 1.3133)	0.0478	1.9345 (1.01 – 3.72)
Tenderness in ASB*	0.7059	0.2455	-1.0155 (-1.8360 ; -0.1950)	0.0153	0.362 (0.16 – 0.82)
Scaphoid tubercle Tenderness	0.6818	0.5524	0.1333 (-0.6423 – 0.9089)	0.7361	n.a.
Painful ulnar deviation*	0.8235	0.3373	0.7411 (0.0083 – 1.4739)	0.0475	2.0983 (1.01 – 4.37)
Painful longitudinal thumb compression*	0.9242	0.5583	0.8544 (0.0786 – 1.6302)	0.0309	2.35 (1.08 – 5.1)

* = Included in final rule
n.a. = not applicable
ASB = Anatomic snuffbox

Outcome of the decision rule

If a threshold of 15% (the probability of a fracture is ≥ 15%) was applied, 66 scaphoid fractures would have been identified correctly, two scaphoid fractures would have been missed (one minor avulsion of the tubercle (type A1) and one incomplete waist fracture (type A2) in anatomic position) and 36 patients would not have received further imaging and immobilization (Table 5). This results in a 15% reduction of further imaging and a 50% reduction of missed fractures. The sensitivity and specificity were 97% and 20% respectively; positive and negative predictive values were 0.33 and 0.94 respectively; the prevalence of true fractures among clinically suspected fractures increased from 0.27 to 0.33.

LIMITATIONS

First, the derived scaphoid rule is highly sensitive (97%), but the specificity is only 20%. Therefore, overtreatment will still be an issue in clinically suspected scaphoid fractures. The lack of specificity is frequently addressed in scaphoid fracture diagnostics(2, 4), however the results of this study show a reduction in unnecessary diagnostics and treatment of more than 15%.

Second, the AUC of this clinical prediction model after internal validation rule is 0.72. This means that the predicted probability of the rule showed a fair discrimination between patients with and without a scaphoid fracture. The higher the AUC, the better the discriminative value of the rule. Bootstrapping methods provide bias-corrected estimates of the performance of a clinical prediction model and are recommended for internal validation(14). However,

Table 4. Clinical prediction rule for detecting true scaphoid fractures

Linear Predictor	(0.649662618 x if man) + (0.51353467826 x if swelling anatomic snuffbox) + (-0.79038263985 x if painful palpation anatomic snuffbox) + (0.57681198857 x if painful ulnar deviation) + (0.66499549728 x if painful thumb compression) -1.685
Clinical Prediction Formula	1 / (1 + EXP (- Linear Predictor))

Table 5. 2x2 table of the clinical prediction rule

	True fracture	No fracture	Total
Rule +	66	133	199
Rule-	2	34	36
Total	68	167	235
Positive Predictive Value		66 / 199 = 0.33	
Negative Predictive Value		34 / 36 = 0.94	
Sensitivity		66 / 68 = 0.97	
Specificity		34 / 167 = 0.20	
Prevalence without rule: 0.27 (64/235)		Prevalence with rule: 0.33 (66/199)	

external validation was not performed. Validation is most reliable when it is performed in a different patient population, and therefore prior to implementation of this clinical decision rule, it is necessary to externally validate this rule in a different patient population in other hospitals.(14, 15)

Third, 70 scaphoid fractures were required based on the 'rule of thumb' in determining the sample size of a clinical prediction rule development (10 events per variable). This study included 68 patients with a proven scaphoid fracture. However, we believe 2 additional events would not have changed the outcome of the analysis.

A continuing issue in scaphoid research is the lack of an adequate reference standard to detect a true fracture. This study used different standards, MRI/CT/follow-up radiographs, based on the local preferences of the hospital. It is known that none of the modalities are 100% accurate (1, 5) and therefore, misdiagnosis is still possible.

Using subjective measures such as physical examination introduces a possible lack of interobserver agreement. This trial did not incorporate an interobserver reliability study. Prior to the study, informative presentations and laminated descriptive sheets were provided concerning the anatomy and tests in clinical evaluation to limit inadequate application of the rule as much as possible.

DISCUSSION

With this study, we developed a highly sensitive clinical decision rule that is able to select those patients presenting at the ED with wrist trauma, that require further imaging and treatment for a suspected scaphoid fracture. Moreover, when applying the Amsterdam Scaphoid Rule, a reduction of 15% in radiographs and unnecessary immobilization is possible, while reducing the number of missed fractures with 50%.

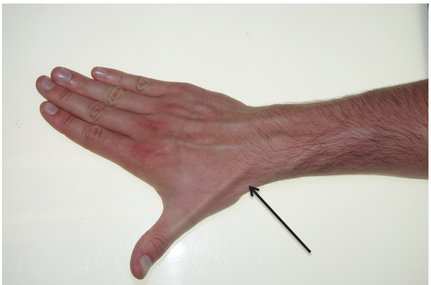
Similar to the Amsterdam Wrist Rules, this scaphoid clinical decision rule will be incorporated in the same smartphone application to simplify its use. The use of a mobile application has recently been studied in applying the Ottawa Ankle Rules.(16) It proved to increase its adherence significantly, which results in improving documentation of key clinical data. Using a mobile application also tackles the known barrier of forgetting details of the decision rule.(17)

In order to link a recommendation to the derived prediction model, it is necessary to set a threshold value. The key is to find a reasonable balance between reduction of unnecessary overtreatment and imaging, and the risk of missing fractures. Several thresholds were used to determine the most suitable in this situation with the goal to keep sensitivity above 95%. With a sensitivity of 97%, this study showed that with a threshold of 15% fracture probability, only 2 of 68 fractures were missed compared to 4 of 68 scaphoid fractures with current clinical practice. Thus this scaphoid rule reduces the number of missed fractures with 50% and reduces unnecessary follow-up with 15%.

This study is a large prospective study on diagnosis of wrist injuries including 68 scaphoid fractures. The variables for clinical assessment were selected by performing a thorough systematic review(2) and analysis encompasses the use of robust statistical and study design methodology.(7, 11, 12) The inclusion of all acute wrist injuries ensured that all scaphoid fractures were included in the analysis in order to get an accurate representation of clinical practice. The statistical models have been thoroughly tested, for this study as well as previous for the Amsterdam Wrist Rules.(6, 18)

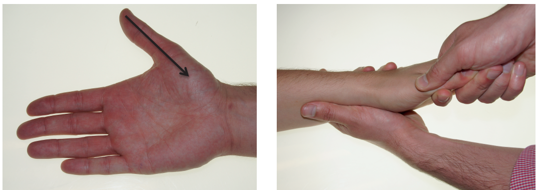
Several findings from previous literature, like the fact that scaphoid fractures mainly occur in young male patients, are emphasized in this study.(4, 19, 20) However, controversial results were also detected, especially on diagnostic performance characteristics of individual tests. A tender anatomic snuffbox and painful longitudinal thumb compression have been described as being highly sensitive for detecting scaphoid fractures .(3, 4) Duckworth et al. showed a 100% sensitivity of ASB tenderness.(4) In contrast, in this study we found a sensitivity of 71% for ASB tenderness. Moreover, ASB tenderness was not present in 20 of 68 (29.4%) scaphoid fractures. In addition, the number of occult fractures was remarkably higher in our data (21% vs 11%). An explanation for the latter is that the current study initially included all wrist injuries; this ensured the inclusion of four scaphoid fractures that were missed with current clinical assessment. Previous studies included patients that were 'clinically suspected for a scaphoid fracture', meaning that there was already a selection prior to inclusion. These selection methods were not described and thus induce selection bias.

Clinical assessment of patients with wrist injury can be protocolized with this decision rule. Moreover, with the Amsterdam Scaphoid Rule the risk of missing a fracture is lower than in current clinical practice. Therefore it has the potential to reduce unnecessary immobilization and diagnostic follow-up without increasing the risk of missed fracture. It can thus provide physicians at the ED an easy and effective tool to select patients with suspected scaphoid fractures for radiography. If the rule does not suggest radiographical evaluation and immobilization, we suggest patients are either treated with a supportive bandage/tubi grip or discharged without treatment and to report back to the outpatient clinic when symptoms persist after 2 weeks. External validation and implementation of this rule will be subject of further research.

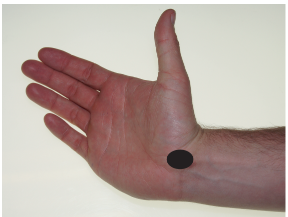


Anatomic Snuffbox (ASB)

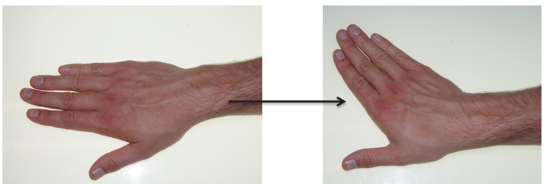
Tenderness in the ASB



Longitudinal Thumb Compression (LTC)



Scaphoid tubercle tenderness,
follow the flexor carpi radialis



Painful ulnar deviation

Appendix. Instructions of physical examination

REFERENCES

1. Mallee WH, Wang J, Poolman RW, Kloen P, Maas M, de Vet HC, et al. Computed tomography versus magnetic resonance imaging versus bone scintigraphy for clinically suspected scaphoid fractures in patients with negative plain radiographs. Cochrane Database Syst Rev. 2015;6:CD010023.
2. Mallee WH, Henny EP, van Dijk CN, Kamminga SP, van Enst WA, Kloen P. Clinical diagnostic evaluation for scaphoid fractures: a systematic review and meta-analysis. J Hand Surg Am. 2014;39(9):1683-91 e2.
3. Parvizi J, Wayman J, Kelly P, Moran CG. Combining the clinical signs improves diagnosis of scaphoid fractures. A prospective study with follow-up. Journal of hand surgery (Edinburgh, Scotland). 1998;23(3):324-7.
4. Duckworth AD, Buijze GA, Moran M, Gray A, Court-Brown CM, Ring D, et al. Predictors of fracture following suspected injury to the scaphoid. J Bone Joint Surg Br. 2012;94(7):961-8.
5. Rhemrev SJ, Beeres FJ, van Leerdam RH, Hogervorst M, Ring D. Clinical prediction rule for suspected scaphoid fractures: A prospective cohort study. (1879-0267 (Electronic)).
6. Walenkamp MM, Bentohami A, Slaar A, Beerekamp MS, Maas M, Jager LC, et al. The Amsterdam wrist rules: the multicenter prospective derivation and external validation of a clinical decision rule for the use of radiography in acute wrist trauma. BMC Musculoskelet Disord. 2015;16(389).
7. Steyerberg E. Study Design for Prediction Models. Clinical Prediction Models: a Practical Approach to Development, Validation, and Updating. New York: Springer; 2009. p. 50-1.
8. Cummings P. Missing data and multiple imputation. JAMA Pediatr. 2013;167(7):656-61.
9. Janssen KJ, Donders AR, Harrell FE, Jr., Vergouwe Y, Chen Q, Grobbee DE, et al. Missing covariate data in medical research: to impute is better than to ignore. J Clin Epidemiol. 2010;63(7):721-7.
10. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393.
11. Steyerberg E. Evaluation of Performance. Anonymous Clinical Prediction Models, a Practical Approach to Development, Validation, and Updating. New York: Springer; 2009. p. 270-9.
12. Steyerberg E. Overfitting and Optimism in Prediction Models. Anonymous Clinical Prediction Models, a Practical Approach to Development, Validation, and Updating. New York: Springer; 2009. p. 94-6.
13. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774-81.
14. Bleeker SE-MHA-S, E.W. - Donders, A.R.T. - Derksen-Lubsen, G. - Grobbee, D.E. - Moons, K.G. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol. 2003;56(0895-4356 (Print)):826-32.
15. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691-8.
16. Silveira PC, Ip IK, Sumption S, Raja AS, Tajmir S, Khorasani R. Impact of a clinical decision support tool on adherence to the Ottawa Ankle Rules. The American journal of emergency medicine. 2016;Mar;34(3)(1532-8171 (Electronic)):412-8.
17. Stiell IG, Bennett C. Implementation of clinical decision rules in the emergency department. Acad Emerg Med. 2007;Nov;14(11)(1553-2712 (Electronic)):955-9.
18. Slaar A, Walenkamp MM, Bentohami A, Maas M, van Rijn RR, Steyerberg EW, et al. A clinical decision rule for the use of plain radiography in children after acute wrist injury: development and external validation of the Amsterdam Pediatric Wrist Rules. Pediatr Radiol. 2016 Jan;46(1):50-60.
19. Duckworth AD, Ring D, McQueen MM. Assessment of the suspected fracture of the scaphoid. J Bone Joint Surg Br. 2011;93(6):713-9.
20. Jenkins PJ, Slade K, Huntley JS, Robinson CM. A comparative analysis of the accuracy, diagnostic uncertainty and cost of imaging modalities in suspected scaphoid fractures. Injury. 2008;39(7):768-74.

PART 3

DIAGNOSTIC IMAGING IN SUSPECTED SCAPHOID FRACTURES: WHAT'S NEXT WHEN RADIOGRAPHS ARE NORMAL?

CHAPTER 5

COMPUTED TOMOGRAPHY FOR SUSPECTED SCAPHOID FRACTURES: COMPARISON OF REFORMATIONS IN PLANE OF THE WRIST VERSUS THE LONG AXIS OF THE SCAPHOID

Mallee WH
Doornberg JN
Ring D
Maas M
Muhl M
van Dijk CN
Goslings JC

Hand (NY). March 2014

ABSTRACT

Background

Definitive diagnosis of occult scaphoid fractures remains difficult. We tested the null hypothesis that, for diagnosis of true fractures among suspected scaphoid fractures, computed tomography (CT) reformations along the long axis of the scaphoid have the same accuracy as reformations made relative to the anatomical planes of the wrist.

Methods

In a prospective trial, thirty-four patients with a suspected scaphoid fracture underwent CT scanning within ten days after trauma. CT reformations along the long axis of the scaphoid (CT-scaphoid) and along planes relative to the wrist (CT-wrist) were made. We used radiographs obtained 6 weeks after injury as the reference standard for a true fracture. A blinded panel including two surgeons and one radiologist came to a consensus diagnosis for each reformation plane.

Results

The reference standard showed six fractures of the scaphoid (prevalence 18%). Using CT–wrist, a scaphoid fracture was diagnosed in five patients (15%), with three false positive, four false negative and two true positive diagnoses. Using CT-scaphoid, a scaphoid fracture was diagnosed in five patients (15%), with one false positive, two false negative and four true positive results. Sensitivity, specificity and accuracy for CT-wrist were 33%, 89% and 79%; and for CT-scaphoid 67%, 96% and 91% respectively. This resulted in PPV's of 36% for CT-wrist and 76% for CT-scaphoid. NPV's were 87% for CT-wrist and 94% for CT-scaphoid. No significant differences were found with the number of patients available.

Conclusions

For diagnosis of true fractures among suspected scaphoid fractures, the diagnostic performance characteristics of CT scans reformatted along the long axis of the scaphoid were better than CT scans in the planes of the wrist, but the differences were not significant.

INTRODUCTION

There is no consensus regarding the best diagnostic strategy for patients with suspected scaphoid fractures, defined as patients with scaphoid tenderness after a fall and normal scaphoid radiographs(1, 2). When Computed Tomography (CT) is used to image the scaphoid for fracture diagnosis or to measure alignment, reformations made along the long axis of the scaphoid are often recommended(3-7). While these reformations have been adopted by experts in the field (5, 8), our sense is that they are not used consistently in all hospitals. We tested the null hypothesis that reformations in planes defined by the long axis of the scaphoid and reformations made in the anatomic planes of the wrist have comparable accuracy for diagnosing true fractures among suspected fractures of the scaphoid.

PATIENTS AND METHODS

The QUADAS-guidelines(9) were used to design the primary study which was approved by our Institutional Review Board.

Patients

This study used a convenience sample of data from a prospective cohort comparing CT and MRI(10). 40 consecutive patients with a clinically suspected scaphoid fracture but normal initial radiographs in four views were enrolled. An attending trauma or orthopaedic surgeon and a musculoskeletal radiologist evaluated the radiographs. Further criteria for inclusion in the study were a minimum age of 18 years, wrist trauma within the previous 24 hours, snuffbox tenderness on palpation and when longitudinally compressing the thumb(11). Exclusion criteria were prior scaphoid fracture, rheumatoid arthritis and dementia. Six patients were lost to follow-up and excluded from this comparison.

Imaging

CT protocol

CT scans were obtained an average of 3.6 days after injury (range 0 to 10 days) using a Brilliance CT-scanner (64 slice, Philips medical System, Eindhoven, The Netherlands) from the distal radioulnar joint to the proximal 1/3 of the metacarpal bones. We used a high-resolution sequence with 0.5 mm slice thickness. Acquired volumetric data were reformatted into axial, coronal and sagittal scans in the anatomical planes of the wrist (CT-wrist) [1] (Figure 1 & 2A). In addition, sagittal and coronal reformations along the central longitudinal axis of the scaphoid (CT scaphoid) were made(3, 4, 6) (Figure 2B & 3).

Reference standard

Six weeks after the initial injury, the same radiographs in 4 standard scaphoid views were repeated. This is one of the most commonly used reference standards in the study of suspected scaphoid fractures(12). An abnormal lucent line within the scaphoid and/or a disruption of the cortex was considered evidence of a fracture(1).

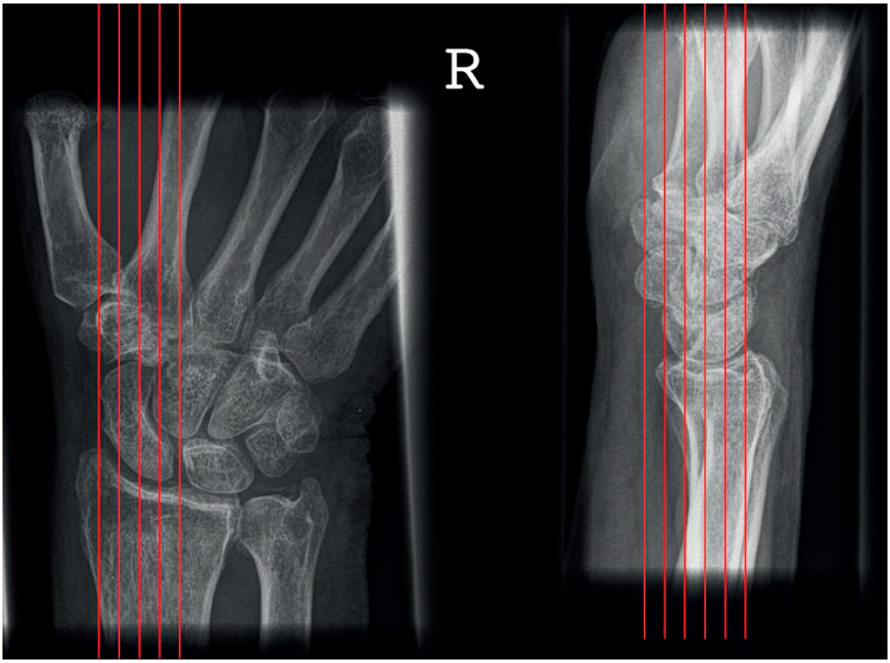


Figure 1. planes of the wrist on PA and lateral radiographs, respectively. The red lines represent the plane of the wrist in which the scans are reformatted

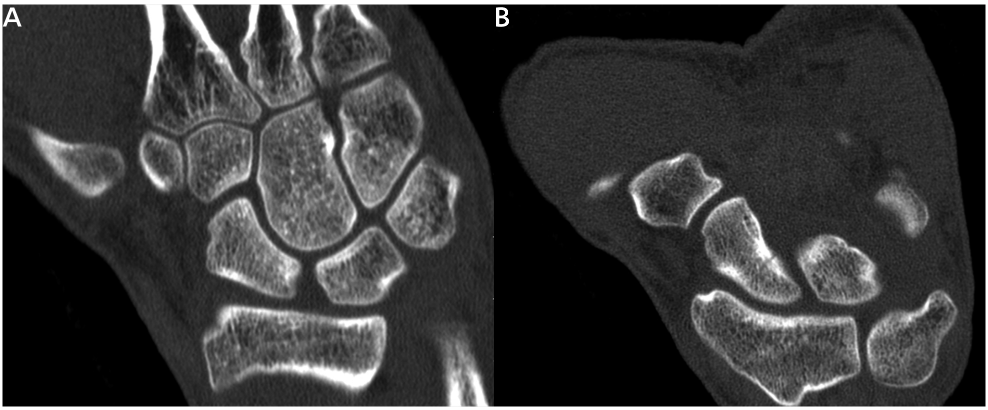


Figure 2. (A) No fracture visible on CT scan reformatted in the plane of the wrist (CT-wrist). (B) The same patient did show a fracture on CT scan reformatted along the longitudinal axis of the scaphoid (CT-scaphoid).

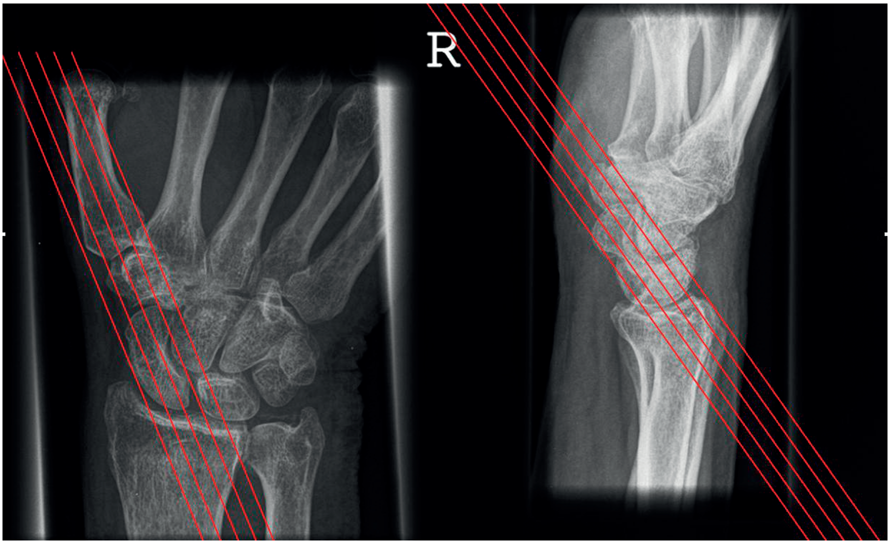


Figure 3. The longitudinal axis of the scaphoid on PA and lateral radiographs, respectively. The red lines represent the long axis of the scaphoid in which the scans are reformatted.

Diagnosis of Fracture

CT-wrist, CT-scaphoid and the reference standard were separated into three rounds and presented to a panel of observers which consisted of an attending musculoskeletal radiologist, and two attending surgeons. The panel evaluated the images for the presence of a scaphoid fracture to reach a consensus opinion. The images were blinded and randomly ordered with a four-week interval between evaluations to limit recall of the images. In the first round, the panel examined the initial radiographs and CT reformation relative to the wrist; in the second round they examined the initial radiographs and the CT with scaphoid axis-reformations; in the third round they evaluated the initial radiographs and the 6-week post-injury radiographs. In each round, the panel was blinded to the results of the other imaging modalities.

Criteria for a fracture on CT images were according to the protocol of Memarsadeghi and colleagues(1): the 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.

Statistical Analysis

Using standard formulas we calculated sensitivity, specificity and accuracy for the detection of a scaphoid fracture with CT-wrist and with CT-scaphoid (Table 1). The positive predictive value (PPV) and negative predictive value (NPV) were determined with use of Bayes' theorem(3, 13). We estimated the prevalence of true scaphoid fractures as 16% in our center for the purpose of these calculations(14). Ninety-five percent confidence intervals were calculated with use of Pratt's normal approximation method for binomial proportions.

The McNemar test for paired binary data was used to test for significant differences between scanning planes(15). For the original study, a sample size of 32 patients provided 80% power (α = 0.05, β = 0.20) to detect significant differences of 20% in diagnostic performance characteristics between MRI and CT. Post-hoc power analysis was performed to calculate if we could reach the same power for this specific study question.

According to post-hoc power analysis 133 patients would be needed to achieve 80% power to show statistical significance of the differences noted.

RESULTS

Reference Standard

The scaphoid specific radiographs obtained six weeks after injury showed a fracture of the wrist in ten patients (29%), of whom six (18%) patients had a fracture of the scaphoid (five waist and one distal pole) and the other fractures were located in the triquetrum (two patients), the capitate (one patient) and the distal radius (one patient). No patients were diagnosed with multiple fractures.

CT-wrist (Table 1)

With CT reformations in planes relative to the wrist, 13 patients were diagnosed with 17 fractures. Fractures to bones other than the scaphoid diagnosed on CT-wrist included a lunate fracture in one, triquetrum fracture in four, trapezium fracture in one, trapezoid fracture in one, capitate fracture in one, hamate fracture in one, distal radius fracture in two and small finger metacarpal fracture in one. Two patients were diagnosed with other fractures along with the scaphoid fracture (distal radius in one and triquetrum in one). In one patient the trapezium, trapezoid, and capitate were fractured. Fracture of the scaphoid was diagnosed in 5 patients (4 waist and 1 proximal pole). There were two true positive, three false positive, and four false negative results according to the reference standard. We calculated a sensitivity of 33% (95% Confidence Interval = 11% to 60%); and a specificity of 89% (95% CI = 84% to 95%) with an accuracy of 79% in depicting scaphoid fractures. The prevalence-adjusted Positive Predictive Value (PPV) was 36% (95% CI = 11% to 69%); prevalence-adjusted Negative Predictive Value (NPV) was 87% (95% CI = 83% to 93%).

CT-scaphoid (Table 1)

With CT-reformations in planes defined by the long axis of the scaphoid, 17 patients were diagnosed with 20 fractures. Fractures to bones other than the scaphoid diagnosed on CT-scaphoid included a lunate fracture in two, triquetrum fracture in four, trapezium fracture in one, capitate fracture in one, hamate fracture in one, distal radius fracture in four and small finger metacarpal fracture in two. One patient was diagnosed with a distal radius fracture along with a fractured scaphoid. One fractured both the small finger metacarpal and triquetrum and one fractured the trapezium and capitate. Fracture of the scaphoid was diagnosed in 5 patients (4 waist and 1 proximal pole). There were four true positive, one false positive and

Table 1. Results of Computed Tomography imaging compared to the 6-week follow-up reference radiographs.

CT-WRIST versus CT-SCAPHOID 4x4									
		CT-wrist				**CT-scaphoid**			
		Scaphoid Fracture	**No scaphoid fracture**	**Total**		**Scaphoid Fracture**	**No scaphoid fracture**	**Total**	
Reference Standard	Scaphoid fracture	2	4	6	Scaphoid fracture	4	2	6	
	No scaphoid fracture	3	25	28	No scaphoid fracture	1	27	28	
	Total	5	29	34	Total	5	29	34	
Sensitivity	2 / 2 + 4 = 33%*				4 / 4 + 2 = 67%*				
Specificity	25 / 3 + 25 = 89%**				27 / 1 + 27 = 96%**				
Accuracy	2 + 25 / 2 + 4 + 3 + 25 = 79%†				4 + 27 / 4 + 2 + 1 + 27 = 91%†				
PPV#	33% x 16% / (33% x 16%) + [(1 – 89) x (1 – 16%)] = 36%				67% x 16% / (67% x 16%) + [(1 – 96) x (1 – 16%)] = 76%				
NPV#	89% x (1 – 16%) / [(1 – 33%) x 16%] + [89% x (1 – 16%] = 87%				96% x (1 – 16%) / [(1 – 67%) x 16%] + [96% x (1 – 16%] = 94%				

* The proportion of patients with a scaphoid fracture according to the reference standard classified as having a positive CT (true positives)
**The proportion of patients with no scaphoid fracture according to the reference standard classified as having a negative CT (true negative)
† The proportion of patients who are correctly classified by CT
# Accounting for prevalence and incidence

two false negative results according to the reference standard. We calculated a sensitivity of 67% (95% Confidence Interval = 36% to 80%) and a specificity of 96% (95% CI = 90% to 99%) with an accuracy of 91% in depicting scaphoid fractures. The prevalence (16%)-adjusted Positive Predictive Value (PPV) was 76% (95% CI = 39% to 94%); prevalence-adjusted Negative Predictive Value (NPV) was 94% (95% CI = 88% to 97%).

Statistical Analysis

According to the McNemar's test of equality of paired proportions(16), we could not detect a significant difference between CT-wrist and CT scaphoid with the number of patients available.

DISCUSSION

Our null-hypothesis that CT-wrist and CT-scaphoid have comparable accuracy (79% vs. 91%, respectively) is confirmed, at least with the number of scans available. With a larger sample, the substantial differences between PPV (CT-wrist: 36% and CT-scaphoid: 76%) and sensitivity (CT-wrist: 33% and CT-scaphoid: 67%) might reach statistical significance.

The study of diagnostic performance characteristics of tests for diagnosing true fractures among scaphoid fractures is hindered by the lack of a consensus reference standard for a true fracture(3, 17). We used one of the more common reference standards (six week post-injury radiographs), but none of the proposed standards (including MRI, which is subject to false positives(10)) are satisfactory. It is conceivable that we will never have a consensus reference standard. We may need to use alternative statistical techniques such as latent class analysis, which relies on associations of specific factors (unobserved or latent classes) rather than a reference standard to estimate diagnostic performance characteristics. Preliminary data from this paper were provided for a study of latent class analysis(18). The numbers changed slightly after all three observers completed the evaluations.

This study should be interpreted in light of several actual or potential shortcomings. First, this was a retrospective analysis of CT scans obtained as part of a prospective study designed to test a different primary study question. This represents a sample of convenience, which offered a prime opportunity to evaluate this issue. Therefore, the comparison of the reformations must be interpreted as a pilot experiment, since the study was not designed or powered to answer this question and showed to be underpowered. This is particularly important given the low prevalence of true fractures among suspected fractures. The differences we observed are substantial and likely would be statistically significant in a larger study.

CONCLUSION

Computed Tomography reformatted along the long axis of the scaphoid is statistically comparable to CT scans in the planes of the wrist for diagnosis of suspected scaphoid fractures although the diagnostic performance characteristics were notably better for CT in the plane of the scaphoid. We encourage further study with larger numbers to test our clinical impression that CT-scaphoid might reach significant superiority over CT-wrist.

ACKNOWLEDGMENTS

We would like to thank drs. S.P. Kamminga for his support to include several patients for this study and Mrs. W.A. van Enst for her dedication in the statistical analysis.

REFERENCES

1. 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(1):169-76.
2. Ring D, Lozano-Calderon S. Imaging for suspected scaphoid fracture. J Hand Surg [Am]. 2008;33(6):954-7.
3. Adey L, Souer JS, Lozano-Calderon S, Palmer W, Lee SG, Ring D. Computed tomography of suspected scaphoid fractures. J Hand Surg Am. 2007;32(1):61-6.
4. Lozano-Calderon S, Blazar P, Zurakowski D, Lee SG, Ring D. Diagnosis of scaphoid fracture displacement with radiography and computed tomography. J Bone Joint Surg Am. 2006;88(12):2695-703.
5. Ring D, Jupiter JB, Herndon JH. Acute fractures of the scaphoid. J Am Acad Orthop Surg. 2000;8(4):225-31.
6. Sanders WE. Evaluation of the humpback scaphoid by computed tomography in the longitudinal axial plane of the scaphoid. J Hand Surg [Am]. 1988;13(2):182-7.
7. Stewart NR, Gilula LA. CT of the wrist: a tailored approach. Radiology. 1992;183(1):13-20.
8. Bain GI, Bennett JD, Richards RS, Slethaug GP, Roth JH. Longitudinal computed tomography of the scaphoid: a new technique. Skeletal radiology. 1995;24(4):271-3.
9. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC medical research methodology. 2003;3:25.
10. Mallee W, Doornberg JN, Ring D, van Dijk CN, Maas M, Goslings JC. Comparison of CT and MRI for diagnosis of suspected scaphoid fractures. J Bone Joint Surg Am. 2011;93(1):20-8.
11. Chen SC. The scaphoid compression test. J Hand Surg [Br]. 1989;14(3):323-5.
12. Ring D, Lozano-Calderon S. Imaging for suspected scaphoid fracture. J Hand Surg Am. 2008;33(6):954-7.
13. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ. 1994;309(6947):102.
14. Jenkins PJ, Slade K, Huntley JS, Robinson CM. A comparative analysis of the accuracy, diagnostic uncertainty and cost of imaging modalities in suspected scaphoid fractures. Injury. 2008;39(7):768-74.
15. Kocher MS, Zurakowski D. Clinical epidemiology and biostatistics: a primer for orthopaedic surgeons. J Bone Joint Surg Am. 2004;86-A(3):607-20.
16. Miettinen OS. The matched pairs design in the case of all-or-none responses. Biometrics. 1968;24(2):339-52.
17. Munk B, Frokjaer J, Larsen CF, Johannsen HG, Rasmussen LL, Edal A, et al. Diagnosis of scaphoid fractures. A prospective multicenter study of 1,052 patients with 160 fractures. Acta orthopaedica Scandinavica. 1995;66(4):359-60.
18. Buijze GA, Mallee WH, Beeres FJ, Hanson TE, Johnson WO, Ring D. Diagnostic performance tests for suspected scaphoid fractures differ with conventional and latent class analysis. Clin Orthop Relat Res. 2011;469(12):3400-7.

CHAPTER 6

COMPARISON OF COMPUTED TOMOGRAPHY AND MAGNETIC RESONANCE IMAGING FOR DIAGNOSIS OF SUSPECTED SCAPHOID FRACTURES

Mallee WH
Doornberg JN
Ring D
van Dijk CN
Maas M
Goslings JC

Journal of Bone and Joint Surgery – American Volume. January 2011

ABSTRACT

Background

The optimal method for diagnosis of true scaphoid fractures among patients with suspected scaphoid fractures is debated. This study tested the null hypothesis that computed tomography (CT) and magnetic resonance imaging (MRI) have the same diagnostic performance characteristics for diagnosis of suspected scaphoid fractures.

Methods

Thirty-four consecutive patients with a suspected scaphoid fracture (tenderness of the scaphoid and normal radiographs after a fall on the outstretched hand) underwent CT and MRI within ten days of wrist injury. The reference standard for a true fracture of the scaphoid was 6-week follow-up radiographs in four views. A panel including surgeons and radiologists came to a consensus diagnosis for each type of imaging considered in a randomly ordered and blinded fashion, independent of the other types of imaging. We calculated sensitivity, specificity and accuracy as well as positive (PPV) and negative predictive values (NPV).

Results

The reference standard revealed six true fractures of the scaphoid (prevalence 18%). CT diagnosed a fracture in five patients (15%), with one false positive, two false negative and four true positive results. MRI diagnosed a fracture in seven patients (21%), with three false positive, two false negative and four true positive results. Sensitivity, specificity and accuracy for CT were 67%, 96% and 91%; and for MRI 67%, 89% and 85% respectively. According to McNemar's test for paired binary data, these differences were not significant. The PPVs using Bayes' formula were 0.76 for CT and 0.54 for MRI. NPVs were 0.94 for CT and 0.93 for MRI.

Conclusions

CT and MRI had comparable diagnostic characteristics. Both are better at excluding than confirming scaphoid fractures and both were subject to false positive and false negative interpretations. The best reference standard is debatable, but for now it is not clear whether bone edema on MRI and small unicortical lines on CT represent a true fracture.

Level of Evidence

Level I-Diagnostic Study-Testing of previously developed diagnostic criteria in series of consecutive patients (with universally applied reference standard)

INTRODUCTION

Displacement and delayed diagnosis are important risk factors for nonunion of scaphoid fractures(1-3). A scaphoid fracture may be suspected following a trauma when there is tenderness of the anatomic snuffbox, even when additional radiographic views are interpreted as normal(4, 5). The prevalence of acute scaphoid fractures is on average 7% among patients with acute wrist injuries(4, 6, 7). In prospective studies where patients had clinical findings of an acute scaphoid fracture but negative plain radiographs –our study population– the reported prevalence in meta-analyses averaged 16%(8, 9). This suggests that on average only one of five patients that present to the emergency room with scaphoid tenderness and normal radiographs actually has a scaphoid fracture(9), and approximately 84% of patients may have unnecessary cast immobilization resulting in a substantial loss of productivity(10, 11).

In 2006, Groves et al. performed a worldwide survey and found substantial variation in imaging and treatment protocols of acute scaphoid injuries(12). Current treatment protocols most commonly include repeat physical examination and radiographs 2 weeks after the initial presentation, or earlier bone scan, CT or MRI imaging(4, 13). Current evidence regarding the optimal protocol for diagnosis of suspected scaphoid fractures lacks methodological quality(14) which may contribute to a lack of consensus(12). It is recommended that MRI is the best radiological test for diagnosis of suspected scaphoid fractures(12), but bone scanning, CT, and ultrasound may also be useful, particularly when MRI is not readily available(4).

Computed tomography (CT) and magnetic resonance imaging (MRI) have both been studied for diagnosis of suspected scaphoid fractures in non-comparative studies(4, 9). MRI has been reported in case series as having a high sensitivity (98-100%) and specificity (100%) (5, 15-20). Disadvantages of MRI include logistic issues (limited availability in some areas—i.e. long wait times), and cost. CT is more readily available and less costly(9). CT imaging is more reliable than radiographs with greater sensitivity (89-100%)(5, 16, 17, 21-23) and specificity (85-100%). All diagnostic radiological studies are better for excluding the diagnosis of a nondisplaced scaphoid fracture than for confirming a fracture, because the prevalence of true fractures among patients with radial sided pain is low, which magnifies the impact of false positive results(21, 23).

We are aware of only one prospective study comparing CT and MRI for confirmation of suspected scaphoid fractures. MRI was both 100% sensitive and specific, while CT was 100% sensitive, but only 73% specific;(5) however, the CT images were made in planes relative to the wrist and forearm as opposed to the recommended reconstructions in planes defined by the long axis of scaphoid(16, 24).

This prospective study evaluates the sensitivity, specificity and accuracy as well as prevalence-adjusted positive predictive values (PPVs) and negative predictive values (NPVs) of CT using reconstructions in the long axis of the scaphoid and MRI in diagnosing suspected scaphoid fractures. We tested the null hypothesis that computed tomography (CT) and magnetic resonance imaging (MRI) have the same performance characteristics for diagnosis of suspected scaphoid fractures.

MATERIALS AND METHODS

This study was designed and reported according to the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) guidelines(14) (Table 1). The study was approved by our Institutional Review Board and all patients gave written informed consent.

Patients

Between April 2008 and October 2008, all patients with clinical symptoms of tenderness in the anatomic snuffbox and normal scaphoid specific radiographs after a fall on an outstretched hand (5, 9, 18, 25, 26), were invited to enroll in a prospective comparison of CT and MRI for diagnosis. The scaphoid specific radiographs consist of four views: 1) a posterior-anterior view in ulnar deviation, 2) a lateral view with fifteen degrees wrist extension, and 3) a lateral view in thirty degrees pronation. 4) and a posterior-anterior view with the x-ray beam directed from distal to proximal in 40 degrees angulation(27). To be included in this study, the patient had to present within 24 hours of injury, have tenderness in the anatomic snuffbox, and normal scaphoid-specific radiographs. Exclusion criteria were patients younger than 18 years of age, any concurrent distal ulna, radius or carpal fracture, previous scaphoid fracture, rheumatoid arthritis, and cognitive dysfunction limiting the physical examination. A radiologist and trauma surgeon evaluated the radiographs.

Forty patients (25 men and 15 women) were enrolled. Twenty-four patients injured their right hand (22 dominant) and sixteen patients injured their left (0 dominant). Both CT and MRI were performed on the same day and within an average of 3.6 days (range 0 to 10 days) after injury. All wrists with suspected scaphoid fractures were immobilized in a thumb spica splint or cast until definitive diagnosis. Five wrists diagnosed with a fracture of the waist of the scaphoid were immobilized for 10 weeks in a below elbow thumb spica cast. One patient diagnosed with a fracture of the distal pole was immobilized in a below elbow thumb spica cast for 6 weeks. Thirty-four patients returned for 6-week follow-up radiographs (average 48 days; range, 35 to 74 days) and five did not. One was a tourist, three patients were lost, and one withdrew. One patient was excluded because of inadequate image quality due to a motion artefact.

Imaging

MRI protocol

MR imaging was performed in all patients with a 1.0 Tesla open MRI –scan (panorama 1.0T, Philips medical System, Eindhoven, The Netherlands). The standard scaphoid protocol (Sense wrist coil) with a slice thickness of 3 mm and 0.6 mm gap, included the following series; localizer, Cor STIR, Cor SE T1 in coronal and sagittal 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 resulting in improved image quality as compared to off centered scanning in the conventional tube MRI(28, 29).

TABLE 1. CT Versus MRI

	6-wk Follow-up Radiographs (Reference Standard)	CT		MRI	
		Scaphoid fracture	No scaphoid fracture	Scaphoid fracture	No scaphoid fracture
Scaphoid fracture	6	4	2	4	2
No scaphoid fracture	28	1	27	3	25
Totals	34	5	29	7	27
Diagnostic Performance Characteristics					
Sensitivity:		4 / 4 + 2 = 67%*		4 / 4 + 2 = 67%*	
Specificity:		27 / 1 + 27 = 96%**		25 / 3 + 25 = 89%**	
Accuracy:		4 + 27 / 4 + 2 + 1 + 27 = 91%***		4 + 25 / 4 + 2 + 3 + 25 = 85%***	
Positive Predictive Value (PPV)		4 / 4 + 1 = 80% #		4 / 4 + 3 = 57% #	
PPV accounting for prevalence and incidence:		76%		54%	
Negative Predictive Value (NPV)		27 / 27 + 2 = 93% §		25 / 25 + 2 = 93% §	
NPV accounting for prevalence and incidence:		94%		93%	

*The proportion of patients with a scaphoid fracture according to the reference standard classified as having a positive MRI (true positives)
**The proportion of patients with no scaphoid fracture according to the reference standard classified as having a negative CT/MRI (true negative)
***The proportion of patients who are correctly classified by CT/MRI
The probability that a patient with a positive CT/MRI has a scaphoid fracture
§ The probability that a patient with a negative CT/MRI does not have a scaphoid fracture

CT protocol

Multidetector high resolution CT-scan was performed in all patients with a Brilliance CT-scan (64 slice, Philips medical System, Eindhoven, The Netherlands) in the following sequence: high-resolution 0.5 mm slices section thickness. The scan covered the wrist from the distal radioulnar joint to the carpometacarpal joints. Patients were positioned –in the superman position– prone with the affected arm above the body with the palm down. Reconstructions in planes defined by the long axis of the scaphoid were made(24). Sagittal plane images of the scaphoid were defined as reconstructions that provided a lateral view of the scaphoid bone as defined by the central longitudinal axis of the scaphoid. Coronal plane images were defined as images that provided a posteroanterior view of the scaphoid in the anatomic plane and in line with the axis of the scaphoid(21, 23).

Reference Standard

Six weeks after the initial injury scaphoid specific radiogrphas were repeated(4). This is the most commonly used reference standard in studies of diagnostic tests for diagnosis of suspected scaphoid fractures and the one used by Memarsadeghi et al in their comparison of multidetector CT and MR imaging(5). An abnormal lucent line within the scaphoid was considered evidence of a fracture.

Study Design

CT, MRI and 6 week follow-up radiographs were separated into three rounds and presented to a panel of three observers: an attending musculoskeletal radiologist, an attending trauma surgeon (trauma surgeons treat fractures in our setting) and an attending orthopaedic surgeon. The panel evaluated the images for the presence of a scaphoid fracture to reach a consensus opinion. In the absence of consensus, the panel openly discussed the case. The images were blinded, randomly ordered according to a computer random number generator, and reviewed in two rounds. In the first round, the panel evaluated the initial radiographs and the CT scan; in the second evaluation they evaluated the initial radiographs and the MRI. ;. The panel was thereby blinded to the results of the other imaging modalities. An interval of 2 weeks between each round of interpretations was used to limit recognition of the radiographs and recall of the CT scan or MRI.

Criteria for a scaphoid fracture on CT images were according to Memarsadghi's study protocol: the 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(5)

Criteria for a bone fracture on MR images included the presence of a cortical fracture line, a trabecular fracture line or combination of both. This is according to the study design used by Memarsadeghi and colleagues. (5, 30). In addition to these criteria, an extensive focal zone of edema without a clear cortical fracture line (comparable to diagnosis of stress fractures(31)) was open for debate to decide if it can be considered a fracture.

Figure 1. Images of the wrist of a sixty-four-year-old patient with a suspected scaphoid fracture after a simple fall on the extended wrist. This patient had evidence of a fracture on CT and MRI, but no fracture was seen with use of the reference standard. Fig. 1-A Normal scaphoid-specific radiographs. Fig. 1-B Evidence of a non-displaced cortical fracture (arrow) on CT (coronal and sagittal planes). Fig. 1-C Evidence of a fracture (arrows) on MRI (a coronal slice of short tau inversion recovery (STIR) sequence and a coronal slice of a spin-echo T1-weighted sequence). Fig. 1-D No evidence of a sustained fracture on the six-week follow-up scaphoid-specific radiographs.

A

B

C

D

Figure 2. A twenty-eight-year-old patient with a suspected scaphoid fracture after a simple fall on the extended wrist. Fig. 2-A Normal scaphoid-specific radiographs. Fig. 2-B Normal initial CT scans (sagittal and coronal planes). Fig. 2-C Fracture (arrows) on MRI (a coronal slice of short tau inversion recovery (STIR) sequence and a coronal slice of a spin-echo T1-weighted sequence). Fig. 2-D No evidence of scaphoid fracture on the six-week follow-up scaphoid-specific radiographs.

Statistical Analysis

The sensitivity, specificity, and accuracy for the detection of a scaphoid fracture with CT and with MRI were calculated according to standard formulas (Table 2) and with 95% confidence intervals constructed with use of Pratt's normal approximation method for binomial proportions. The significance of differences was evaluated with use of the McNemar test for paired binary data(32) for each imaging modality. The positive predictive value (PPV) and negative predictive value were determined with use of Bayes' theorem, which requires an a priori estimate of the prevalence (pretest probability) of the presence of scaphoid fractures. The positive predictive value is the patient's probability of having a scaphoid fracture when the test is positive, and the negative predictive value is the probability of a patient not having a scaphoid fracture when the test is negative. The predictive values of any imaging modality depend critically on the prevalence of the characteristic in the patients being tested; hence the use of the appropriate Bayesian analysis is important. For the determination of positive and negative predictive values, we estimated an average prevalence of scaphoid fractures of 16% on the basis of the best available data(9). The positive predictive value was calculated as: sensitivity x prevalence / (sensitivity x prevalence) + [(1 – specificity) x (1 – prevalence)], and the negative predictive value was calculated as: (specificity) x (1 – prevalence) / [(1 – sensitivity) x prevalence] + [specificity x (1 – prevalence)](4, 33).

Statistical analysis and power analysis were performed to establish the number of patients required for comparing diagnostic performance characteristics (sensitivity, specificity, accuracy, positive predictive value, negative predictive value) between CT and MRI. Using McNemar's test of equality of paired proportions(34), a sample size of 32 patients provided 80% power ($\alpha = 0.05$, $\beta = 0.20$) to detect significant differences in proportions of 20% in each performance characteristic between the two imaging protocols with a 0.05 two-sided level.

Source of Funding

There was no external funding source that played a role in this study.

RESULTS

Reference Standard

Ten patients (29%) had a fracture of the wrist identified on radiographs obtained six weeks after injury, of which 6 (18%) patients had a fracture of the scaphoid (one distal pole fracture, five waist fractures). One evident fracture diagnosed on CT and MRI was not seen on the 6-week scaphoid series (Figure 3). The remaining four fractures were located in the triquetrum (2 patients), the capitate (one patient) and distal radius (one patient). No patients were diagnosed with multiple fractures based on conventional radiographs.

CT Imaging

CT imaging resulted in a diagnosis of 20 fractures in 17 patients. Fractures were located in the scaphoid (five, 4 waist and 1 distal pole), lunate (two), triquetrum (four), trapezium (one), capitate (one), hamate (one), distal radius (four) and the small finger metacarpal (two). Three patients were diagnosed with multiple fractures. One patient had a fracture of the distal radius and the scaphoid bone, one fractured both the small finger metacarpal and triquetrum and one fractured the trapezium and capitate. Computed tomography identified all of the non-scaphoid fractures seen on the six-week post-injury radiographs. Four of six scaphoid fractures as seen on the reference standard were depicted on CT. A total of 5 (15%) scaphoid fractures were found resulting in 67% sensitivity (95% confidence interval 35% to 88%) and 96% specificity (95% confidence interval = 85% to 99%) with an accuracy of 91% in depicting scaphoid fractures. The prevalence (16%(9))-adjusted Positive Predictive Value (PPV) was 0.76 (95% confidence interval 0.43 to 0.95); prevalence-adjusted Negative Predictive Value (NPV) was 0.94 (95% confidence interval 0.81 to 0.98) (Table 2).

MR Imaging

MR imaging identified a total of 19 fractures in 16 patients: 1 less than CT. Fractures were located in the scaphoid (seven, 6 waist and 1 distal pole), lunate (one), triquetrum (three), trapezium (two), capitate (one), hamate (one), distal radius (three) and the small finger metacarpal bone (one). Three patients were diagnosed with multiple fractures. Two patients had a fracture of the distal radius and the scaphoid bone, one fractured both the trapezium and capitate. Four of six scaphoid fractures diagnosed on six-week radiographs were found an MRI. Three additional scaphoid fractures were diagnosed. The panel diagnosed a fracture of the scaphoid based on MRI in two patients with negative CT and negative six-week follow-up radiographs (Figure 2). According to the reference standard the sensitivity of MRI for correct diagnosis of an occult scaphoid fracture was 67% (95% confidence interval = 35% to 88%), specificity 89% (95% confidence interval = 76% to 96%), and accuracy 85%.

The difference between performance characteristics of CT versus MRI were not statistically significant with the numbers available using McNemar's test for paired binary data(32) ($p < 0.05$). The prevalence-adjusted PPV was 0.54 (95% confidence interval 0.29 to 0.81) and the prevalence-adjusted NPV was 0.93 (95% confidence interval 0.80 to 0.98). (Table 2)

DISCUSSION

A systematic review(4) of studies evaluating imaging techniques for the diagnosis of suspected scaphoid fractures found that MRI had an average sensitivity of 98%, specificity of 99%, accuracy of 96%, prevalence-adjusted NPV of 1.00 (meaning that an MRI showing no fracture always corresponded with a true absence of fracture), prevalence-adjusted PPV of 0.88 (meaning that a positive MRI corresponded with a true fracture in 88%). The review showed that CT had an average sensitivity of 94%, specificity of 96%, accuracy of 98%, prevalence-adjusted PPV of 0.75 and NPV of 0.99. In other words, both MRI and CT are better at excluding than confirming scaphoid fractures and MRI has performed slightly better than CT in these non-comparative cohort studies. We could not reproduce these excellent diagnostic performance characteristics in our study. We found a sensitivity of 67% for both MRI and CT, and PPVs of 0.54 and 0.76

respectively. This might be explained by our strict inclusion criteria, as we did not include any fractures that were diagnosed on scaphoid-specific radiographs.

We found that the interpretation of bone marrow edema on MRI is questionable. In this study, our panel decided that a focal zone of bone edema was considered a fracture. In Memarsadeghi's study, evidence of a zone of diffusely increased signal intensity on STIR images was interpreted as bone marrow edema and not a fracture(5). Edema had to be accompanied by a cortical fracture line, trabecular fracture line or a combination, to be compatible with a fracture. If we use these criteria evaluating MRI, this would have resulted in a diagnosed fracture in four patients (seven patients according to our criteria), with one (three in our study) false positive, three (two) false negative and three (four) true positive results. Sensitivity, specificity and accuracy for MRI would have been 50%, 96% and 88%; as compared to the figures in our study being: 67%, 89% and 85% respectively. PPV and NPV according to Bayes' theorem would have been 0.70 and 0.91 compared to 0.54 and 0.93 with our criteria.

One difficulty of the study of suspected scaphoid fractures is the absence of a consensus reference standard for a true fracture(21, 35). While it is accepted that both MRI (possible bone bruise) and CT scanning (possible vascular channels) may have findings that can be misinterpreted as a fracture, it is not clear that a six-week post-injury radiograph can diagnose all fractures, but there is currently no viable alternative. In our study one evident fracture diagnosed with perfect agreement by our observers on CT and MRI was not seen on the 6-week scaphoid series (Figure 3). Therefore we suspect that the reference standard of six-week post-injury radiographs is inadequate. Subsequent to the design and execution of this study we became aware of latent class analysis as a technique for analyzing diagnostic performance characteristics in the absence of a consensus reference standard. Instead of relying on a reference standard to determine diagnostic categories, this statistical technique looks for separate diagnostic groups (classes) in the data. Given that there may never be an adequate reference standard for the diagnosis of true fractures among suspected fractures, latent class analysis may be useful here and will be incorporated in future studies.

The potential weaknesses of this study include the use of a 1 Tesla MRI unit; the use of an open MRI unit; the involvement of trauma surgeons; and the use of a consensus panel to diagnose fracture (which might introduce selection bias). A 1 Tesla open MR was used for practical purposes, but also with the understanding that the use of a dedicated coil and central placement of the wrist within the magnet in a comfortable position that limits motion artefact are more important than the strength of the magnet or the closed vs. open MRI machine. Viewing the involvement of a trauma surgeon as a weakness may be a cultural bias or misunderstanding, because in The Netherlands trauma surgeons level of experience in fracture management at least equal to that of an orthopaedic surgeon.

Considered in the light of these shortcomings, the null hypothesis that computed tomography (CT) and magnetic resonance imaging (MRI) have the same diagnostic performance characteristics for diagnosis of suspected scaphoid fractures was confirmed. CT had accuracy comparable to MRI (91% versus 85%). Both imaging modalities were better for excluding a scaphoid fracture ($NPV_{CT} = 0.94$ versus $NPV_{MRI} = 0.93$) than for confirming a scaphoid fracture

(PPV_{CT} = 0.76 versus PPV_{MRI} = 0.54). While additional study is needed, based on our study and results available in the literature(4), CT with reconstructions made in planes defined by the long axis of the scaphoid is comparable to MRI for diagnosis of suspected scaphoid fractures.

Appendix. This study was reported according to the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) guidelines(14)

QUADAS* guidlines	
Characteristics	**Motivation**
Was the spectrum of patients representative of the patients who will receive the test in practice?	We included patients with a positive clinical examination and negative plain radiographs for suspected scaphoid fracture.
Is the reference standard likely to correctly classify the target condition?	Six week follow-up radiographs in four planes is the best available reference standard according to the current literature(4, 5), however is indeed debatable.
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 two tests?	Not applicable. Six weeks are required for an abnormal lucent line to develop which was considered evidence of a fracture(5).
Did the whole sample or a random selection of the sample receive verification using a reference standard of diagnosis?	The whole sample underwent 6-week follow-up radiographs.
Did patients receive the same reference standard regardless of the index result?	The index result did not influence the study protocol.
Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard?)	Yes, however the best reference standard for a true fracture is debatable. Findings on CT or MRI are sometimes given as part of the reference standard, but we considered this circular given that these diagnostic techniques are under study(23).
Were the index test results interpreted without knowledge of the results of the reference standard?	The panel was blinded to the results of follow-up radiographs.
Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	The initial negative plain radiographs in four planes of the scaphoid, as well as the positive results of clinical examination were available to the panel.
Were uninterpretable/intermediate test results reported?	One patient was excluded because of insufficient image quality of 6-week follow-up radiographs.
Were withdrawals from the study explained?	Five patients were excluded, One patient was a tourist and moved back abroad, 3 were lost and declined to return despite written informed consent. See flow diagram.

*This study was reported according to the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) guidelines(14)

REFERENCES

1. Waitayawinyu T, McCallister WV, Nemechek NM, Trumble TE. Scaphoid nonunion. J Am Acad Orthop Surg. 2007 May;15(5):308-20.
2. Dias JJ, Wildin CJ, Bhowal B, Thompson JR. Should acute scaphoid fractures be fixed? A randomized controlled trial. J Bone Joint Surg Am. 2005 Oct;87(10):2160-8.
3. Arora R, Gschwentner M, Krappinger D, Lutz M, Blauth M, Gabl M. Fixation of nondisplaced scaphoid fractures: making treatment cost effective. Prospective controlled trial. Arch Orthop Trauma Surg. 2007 Jan;127(1):39-46.
4. Ring D, Lozano-Calderon S. Imaging for suspected scaphoid fracture. J Hand Surg [Am]. 2008 Jul-Aug;33(6):954-7.
5. 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 Jul;240(1):169-76.
6. Hove LM. Epidemiology of scaphoid fractures in Bergen, Norway. Scand J Plast Reconstr Surg Hand Surg. 1999 Dec;33(4):423-6.
7. Larsen CF, Lauritsen JM. [Epidemiology of acute wrist injuries]. Ugeskr Laeger. 1994 May 9;156(19):2889-92.
8. 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 May;168(5):1287-93.
9. Jenkins PJ, Slade K, Huntley JS, Robinson CM. A comparative analysis of the accuracy, diagnostic uncertainty and cost of imaging modalities in suspected scaphoid fractures. Injury. 2008 Jul;39(7):768-74.
10. Dorsay TA, Major NM, Helms CA. Cost-effectiveness of immediate MR imaging versus traditional follow-up for revealing radiographically occult scaphoid fractures. AJR Am J Roentgenol. 2001 Dec;177(6):1257-63.
11. Bretlau T, Christensen OM, Edstrom P, Thomsen HS, Lausten GS. Diagnosis of scaphoid fracture and dedicated extremity MRI. Acta Orthop Scand. 1999 Oct;70(5):504-8.
12. Groves AM, Kayani I, Syed R, Hutton BF, Bearcroft PP, Dixon AK, et al. An international survey of hospital practice in the imaging of acute scaphoid trauma. AJR Am J Roentgenol. 2006 Dec;187(6):1453-6.
13. Beeres FJ, Hogervorst M, Rhemrev SJ, den Hollander P, Jukema GN. A prospective comparison for suspected scaphoid fractures: bone scintigraphy versus clinical outcome. Injury. 2007 Jul;38(7):769-74.
14. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003 Nov 10;3:25.
15. Beeres FJ, Rhemrev SJ, Hogervorst M, den Hollander P, Jukema GN. [Scaphoid fractures: diagnosis and therapy]. Ned Tijdschr Geneeskd. 2007 Mar 31;151(13):742-7.
16. Ring D, Jupiter JB, Herndon JH. Acute fractures of the scaphoid. J Am Acad Orthop Surg. 2000 Jul-Aug;8(4):225-31.
17. Bhat M, McCarthy M, Davis TR, Oni JA, Dawson S. MRI and plain radiography in the assessment of displaced fractures of the waist of the carpal scaphoid. J Bone Joint Surg Br. 2004 Jul;86(5):705-13.
18. 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 Apr;203(1):245-50.
19. Nakamura R, Imaeda T, Horii E, Miura T, Hayakawa N. Analysis of scaphoid fracture displacement by three-dimensional computed tomography. J Hand Surg [Am]. 1991 May;16(3):485-92.
20. Plancher KD. Methods of imaging the scaphoid. Hand Clin. 2001 Nov;17(4):703-21.
21. Adey L, Souer JS, Lozano-Calderon S, Palmer W, Lee SG, Ring D. Computed tomography of suspected scaphoid fractures. J Hand Surg [Am]. 2007 Jan;32(1):61-6.
22. Bain GI, Bennett JD, MacDermid JC, Slethaug GP, Richards RS, Roth JH. Measurement of the scaphoid humpback deformity using longitudinal computed tomography: intra- and interobserver variability using various measurement techniques. J Hand Surg [Am]. 1998 Jan;23(1):76-81.

23. Lozano-Calderon S, Blazar P, Zurakowski D, Lee SG, Ring D. Diagnosis of scaphoid fracture displacement with radiography and computed tomography. J Bone Joint Surg Am. 2006 Dec;88(12):2695-703.
24. Sanders WE. Evaluation of the humpback scaphoid by computed tomography in the longitudinal axial plane of the scaphoid. J Hand Surg [Am]. 1988 Mar;13(2):182-7.
25. Phillips TG, Reibach AM, Slomiany WP. Diagnosis and management of scaphoid fractures. Am Fam Physician. 2004 Sep 1;70(5):879-84.
26. Beeres FJ, Rhemrev SJ, den Hollander P, Kingma LM, Meylaerts SA, le Cessie S, et al. Early magnetic resonance imaging compared with bone scintigraphy in suspected scaphoid fractures. J Bone Joint Surg Br. 2008 Sep;90(9):1205-9.
27. Toth F, Sebestyen A, Balint L, Mester S, Szabo G, Nyarady J, et al. Positioning of the wrist for scaphoid radiography. Eur J Radiol. 2007 Oct;64(1):126-32.
28. Nikken JJ, Oei EH, Ginai AZ, Krestin GP, Verhaar JA, van Vugt AB, et al. Acute peripheral joint injury: cost and effectiveness of low-field-strength MR imaging--results of randomized controlled trial. Radiology. 2005 Sep;236(3):958-67.
29. Nikken JJ, Oei EH, Ginai AZ, Krestin GP, Verhaar JA, van Vugt AB, et al. Acute wrist trauma: value of a short dedicated extremity MR imaging examination in prediction of need for treatment. Radiology. 2005 Jan;234(1):116-24.
30. Andrews CL. From the RSNA Refresher Courses. Radiological Society of North America. Evaluation of the marrow space in the adult hip. Radiographics. 2000 Oct;20 Spec No:S27-42.
31. Lee JK, Yao L. Stress fractures: MR imaging. Radiology. 1988 Oct;169(1):217-20.
32. Kocher MS, Zurakowski D. Clinical epidemiology and biostatistics: a primer for orthopaedic surgeons. J Bone Joint Surg Am. 2004 Mar;86-A(3):607-20.
33. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ. 1994 Jul 9;309(6947):102.
34. Miettinen OS. The matched pairs design in the case of all-or-none responses. Biometrics. 1968 Jun;24(2):339-52.
35. Munk B, Frokjaer J, Larsen CF, Johannsen HG, Rasmussen LL, Edal A, et al. Diagnosis of scaphoid fractures. A prospective multicenter study of 1,052 patients with 160 fractures. Acta Orthop Scand. 1995 Aug;66(4):359-60.

CHAPTER 7

6-WEEK RADIOGRAPHS UNSUITABLE FOR DIAGNOSIS OF SUSPECTED SCAPHOID FRACTURES

Mallee WH
Mellema JJ
Guitton TG
Goslings JC
Ring D
Doornberg JN

In collaboration with the Science of Variation Group (83 Collaborators).

Archives of Orthopaedic and Trauma Surgery. June 2016

ABSTRACT

Introduction

Six week follow-up radiographs are a common reference standard for the diagnosis of suspected scaphoid fractures. The main purpose of this study was to evaluate the interobserver reliability and diagnostic performance characteristics of 6-weeks radiographs for the detection of scaphoid fractures. In addition, two online techniques for evaluating radiographs were compared.

Materials & Methods

A total of 81 orthopedic surgeons affiliated with the Science of Variation Group assessed initial and 6-week scaphoid-specific radiographs of a consecutive series of 34 patients with suspected scaphoid fractures. They were randomized in two groups for evaluation, one used a standard website showing JPEG files and one a more sophisticated image viewer (DICOM). The goal was to identify the presence or absence of a (consolidated) scaphoid fracture. Interobserver reliability was calculated using the multirater kappa measure. Diagnostic performance characteristics were calculated according to standard formulas with CT and MRI upon presentation in the emergency department as reference standards.

Results

The interobserver agreement of 6-week radiographs for the diagnosis of scaphoid fractures was slight for both JPEG and DICOM (k=0.15 and k=0.14, respectively). The sensitivity (range, 42-79%) and negative predictive value (range, 79-94%) were significantly higher using a DICOM viewer compared to JPEG images. There were no differences in specificity (range, 53-59%), accuracy (range, 53-58%), and positive predictive value (range, 14-26%) between the groups.

Conclusions

Due to low agreement between observers for the recognition of scaphoid fractures and poor diagnostic performance, 6-week radiographs are not adequate for evaluating suspected scaphoid fractures. The online evaluation of radiographs using a DICOM viewer seem to improve diagnostic performance characteristics compared to static JPEG images and future reliability and diagnostic studies should account for variation due to the method of delivering medical images.

Level of Evidence

Diagnostic level II

INTRODUCTION

In management of suspected scaphoid fractures, overtreatment (i.e. immobilization and restrictions of activities) must be balanced against the risks of nonunion associated with undertreatment(1). Overtreatment can be limited by establishing early definitive diagnosis using bone scintigraphy(2-4), magnetic resonance imaging (MRI)(5-8) and computed tomography (CT)(5, 8, 9). However, there is no consensus on scaphoid imaging protocols due to limited evidence regarding diagnostic performance of these advanced imaging techniques(10).

The absence of a consensus reference standard for the diagnosis of scaphoid fractures makes the interpretation of diagnostic performance characteristics and improvement of diagnostic imaging tests difficult(11). Latent class analysis can be used to estimate diagnostic test accuracy without using a reference standard(1, 12), but this approach has considerable limitations and must be viewed with skepticism(13). The most commonly used reference standard in studies that evaluated diagnostic tests for scaphoid fractures are scaphoid-specific radiographs made 6 weeks after initial injury(5, 8, 9, 11, 14-18), while some authors question the use of follow-up radiographs as reference standard(19-21).

The Science of Variation Group, a collaborative effort to improve the study of variation in interpretation and classification of injuries, performed numerous studies by evaluating images using JPEG format(22-24). Since this could limit diagnostic performance due to lack of several functions (window level, zoom, lower quality image), a new online tool was created using an embedded DICOM viewer. This tool mimics clinical practice, however, larger data files and use of multiple functions increases duration of assessment. It is unknown if this tool could be of true value.

As the reliability and accuracy of 6-week radiographs for suspected scaphoid fractures remain subject of discussion and important for the interpretation of diagnostic accuracy of alternative imaging modalities, CT and MRI in particular, there is a need to assess its reliability as well as diagnostic performance characteristics. Therefore, the purpose of this study was to evaluate the interobserver reliability and diagnostic performance characteristics of 6-week radiographs for the recognition of scaphoid fractures in patients with suspected scaphoid fractures. In addition, this study compared the online evaluation of radiographs in JPEG and DICOM format.

METHODS

Study design

Orthopaedic surgeons affiliated with the Science of Variation Group were asked to log on to http://www.scienceofvariationgroup.org or http://www.traumaplatform.org for an online evaluation of suspected scaphoid fractures. In an invitation email observers were informed that participation would be credited on the study by acknowledgement or group authorship (25, 26) and links were provided that directed to the respective web-based study platforms. Our Institutional Review Board approved this study.

Subjects

The initial and 6-week radiographs were used from our previous study(5) of a consecutive series of 34 patients aged 18 years or greater with a suspected scaphoid fracture (tenderness of the scaphoid and normal radiographic findings after a fall on the outstretched hand). All patients presented within twenty-four hours after injury and underwent CT and MRI within ten days after wrist injury between April 2008 and October 2008 in a level I trauma center.

The number of subjects in reliability studies is determined based on an appropriate balance between the number of observers evaluating each subject and the number of subjects(27). Our web-based study platforms (i.e. Science of Variation Group and Traumaplatform) aim to increase the number of observers in interobserver reliability studies for maximizing power and generalizability and to allow comparison between and within subgroups. For this reason, we prefer to select a limited number of subjects to limit burden on observers and increase participation rate (i.e. number of observers).

Observers

Orthopedic surgeons trained in hand surgery and listed in the Science of Variation Group as active members were randomized (1:1) by computer-generated random numbers (Microsoft Excel, Redmond, WA, USA) to assess the selected radiographs online in JPEG or DICOM format.

Online evaluation

Scaphoid-specific radiographs at baseline and 6 weeks after initial trauma were presented and consisted of four views: (1) a posteroanterior view with the wrist in ulnar deviation, (2) a lateral view with the wrist in 15° extension, (3) a lateral view with the wrist in 30° of pronation, and (4) a posteroanterior view with the x-ray beam directed from distal to proximal and with the wrist positioned in 40° of angulation. Observers were asked to answer 1 question for each of the 34 cases: Is there a (consolidated) scaphoid fracture?

Before starting the online evaluation and upon log on to the website, observers received a short description of the study procedure. Observers assigned to the JPEG group evaluated radiographs that were converted to images in JPEG format (www.scienceofvariationgroup.org) and observers assigned to the DICOM group evaluated radiographs provided by an online DICOM viewer (www.traumaplatform.org). Both groups evaluated the same initial and 6-week radiographs, however, the JPEG group was not able to use the window level, scroll, and zoom options available in the online DICOM viewer software.

Statistical analysis

A post-hoc power analysis was performed using the method as described by Guitton and Ring(23). It was calculated that 81 observers provided 5.8% power to detect a 0.003 difference in kappa value (i.e. interobserver reliability) between the JPEG and DICOM group using a two-sample z-test (alfa = 0.05). However, 81 observers provided 100% power to detect a clinically relevant difference in kappa value, defined as a difference of one category as describe by Landis and Koch(28) (Δ_{kappa}=0.20), between the groups with alfa = 0.05.

Interobserver reliability was calculated using the multirater kappa as described by Siegel and Castellan(29). The kappa statistic is a frequently used measure of chance-corrected agreement between observers and interpreted according to the guidelines of Landis and Koch(28): a value of 0.01 to 0.20 indicates slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 to 0.99, almost perfect agreement. A two-sample z-test was used to compare kappa values and P values of <0.05 were considered significant. For a better understanding of the underlying data, the proportion of agreement was calculated for each case (in absolute percentages, %) and defined as the proportion of observers agreeing with the most provided answer.

Diagnostic performance characteristics (sensitivity, specificity, accuracy, positive predictive value, and negative predictive values) of 6-week radiographs for the recognition of (consolidated) scaphoid fractures were calculated according to standard formulas. The reference standard for the diagnosis of scaphoid fractures was CT and MRI. A panel of three observers, an attending musculoskeletal radiologist, an attending trauma surgeon who treats fractures, and an attending orthopaedic surgeon, evaluated the images for the presence of a scaphoid fracture until a consensus opinion was reached(5). The 95% confidence intervals (95% CIs) were calculated by using the formula for the standard error of proportion, based on normal approximation method for binomial proportions, and differences were considered significant when the 95% CIs did not overlap(30).

RESULTS

Observer Characteristics

A total of 288 invitation emails were sent, of which 143 went to the JPEG group and 145 to the DICOM group. Fifty-seven respondents started with the evaluation in the JPEG group, of which 53 (93%) completed the online evaluation, and 45 respondents started in the DICOM group, of which 28 (62%) completed the online evaluation. After incomplete responses were excluded, 53 (65%) observers were left in the JPEG group and 28 (35%) in the DICOM group. Observers were predominately male (95%), from the U.S. (78%), hand and wrist surgeons (96%), and in independent practice for more than 5 years (68%) (Table 1).

Reliability of 6-Week Radiographs for Scaphoid Fractures

The interobserver reliability of 6-week radiographs for the diagnosis of scaphoid fractures was the same for the JPEG and DICOM viewer group and slight in both groups (k=0.15 and k=0.14, respectively; P = 0.75). In addition, subgroup analysis showed that interobserver agreement ranged from slight to fair and no significant differences in kappa value between subgroups were detected (Table 2). The average proportion of agreement was 68% in the JPEG group and 68% in the DICOM group (Table 3).

Table 1. Observer Characteristics

	JPEG (n=53)		DICOM Viewer (n=28)	
	n	%	n	%
Sex				
Men	50	94	27	96
Women	3	5,7	1	3,6
Area				
United States	41	77	22	79
Europe	7	13	3	11
Other	5	9,4	3	11
Specialization				
Hand and wrist	53	100	25	89
Schoulder and elbow	-	-	2	7,1
Trauma	-	-	1	3,6
Years in independent practice				
0-5	18	34	8	29
6-10	9	17	4	14
11-20	16	30	10	36
21-30	10	19	6	21
Fractures per year				
0-10	12	23	5	18
11-20	33	62	7	25
more than 20	8	15	16	57

Diagnostic Performance Characteristics of 6-Week Radiographs for Scaphoid Fractures

The sensitivity of 6-week radiographs for the diagnosis of scaphoid fractures ranged from 42% to 79% and was significantly higher in the DICOM group compared to the JPEG group with MRI, CT, and MRI with CT combined as reference standard. Specificity ranged from 53% to 59%, accuracy ranged from 53% to 58%, and positive predictive value ranged from 14% to 26% and were not significantly different between the DICOM and JPEG group with MRI, CT and MRI with CT combined as reference standard. The negative predictive value ranged from 79% to 94% and was significantly higher using the DICOM viewer compared to JPEG images with MRI, CT, and MRI with CT combined as reference standard (Table 4).

DISCUSSION

Scaphoid-specific radiographs at 6 weeks follow-up are most commonly used as reference standard for scaphoid fractures despite its alternatives, such as latent class analysis and MRI, but its use remains subject of discussion(1, 5, 7-9, 12, 14-18). This study was designed to evaluate the interobserver reliability and diagnostic performance characteristics of 6-week radiographs for the recognition of scaphoid fractures in patients with suspected scaphoid fractures and to

Table 2. Interobserver Agreement for the Recognition of (Consolidated) Scaphoid Fractures Based on 6-Week Radiographs (JPEG versus DICOM Viewer)

	JPEG (n=53)			DICOM Viewer (n=28)			
	Kappa	**Agreement**	**95% CI**	**Kappa**	**Agreement**	**95% CI**	**P Value**
Overall	0,15	Slight	0.13-0.16	0,14	Slight	0.12-0.16	0,75
Area							
United States	0,14	Slight	0.12-0.16	0,16	Slight	0.14-0.18	0,24
Europe	0,18	Slight	0.09-0.26	0,28	Fair	0.06-0.50	0,40
Other	0,04	Slight	-0.06-0.15	0,18	Slight	-0.04-0.41	0,28
Years in independent practice							
0-5	0,12	Slight	0.09-0.15	0,06	Slight	0.00-0.13	0,094
More than 5 years	0,16	Slight	0.13-0.18	0,16	Slight	0.13-0.18	0,94
Fractures per year							
0-10	0,17	Slight	0.11-0.23	0,22	Fair	0.09-0.35	0,47
11-20	0,14	Slight	0.12-0.15	0,21	Fair	0.12-0.29	0,12
more than 20	0,12	Slight	0.03-0.22	0,13	Slight	0.10-0.16	0,94

7

compare the online evaluation of radiographs using images in JPEG and DICOM format. We found that the interobserver reliability for 6-week radiographs was slight in both the JPEG and DICOM group. The diagnostic performance characteristics of 6-week radiographs were poor as well, but significantly better when radiographs were evaluated using a DICOM viewer compared to JPEG images.

The strengths of our study include the large number of observers, which allowed a more complex study design with randomization and subgroup analysis, the use of prospectively collected data from our previous study(5) that evaluated a consecutive series of 34 patients with a suspected scaphoid fracture that returned for follow-up after 6 weeks and underwent CT and MRI scans, and the use of DICOM viewers for the online evaluation of radiographs that resembles evaluation in clinical practice. The limitations include the heterogeneous group of surgeons that evaluated the radiographs, which were from multiple countries and different levels of experience and therefore more likely to disagree compared to observers from a single institute with the same level of experience. A possible limitation was the use of a reference standard for the diagnosis of scaphoid fractures that was based on CT and MRI findings and the consensus agreement of three senior authors.

In this study, the interobserver reliability for the recognition of scaphoid fractures based on 6-week radiographs was low in the JPEG and DICOM group and comparable with agreement reported in previous studies(19-21). Tiel-van Buul et al.(19) selected follow-up radiographs (2 and 6 weeks after injury) of a consecutive series of 60 patients with suspected scaphoid fractures that were rated by 4 observers and found slight to fair interobserver agreement (range, k=0.20 to k=0.39). A similar study by Tiel-van Buul et al.(20) reported slight to moderate agreement (range, k=0.19 to k=0.50) among 3 observers that evaluated 6-week radiographs of a consecutive series of 78 patients with clinically suspected scaphoid fractures. Low et al.(21) found fair agreement (range, k=0.30 to k=0.40) for scaphoid-specific follow-up radiographs between 4 observers that rated 50 patients with a suspected scaphoid fracture.

We found that the diagnostic performance characteristics of 6-week radiographs for scaphoid fractures were poor with MRI, CT, and MRI with CT combined as reference standard using radiographs in JPEG and DICOM format. Six-week radiographs seem better at excluding scaphoid fractures (negative predictive value ranged from 79% to 94%) than recognizing a scaphoid fracture (positive predictive value ranged from 14% to 26%). Moreover, our data suggest that almost 50% of the ratings were inaccurate (accuracy ranged from 53% to 58%). Low et al.(21) reported low negative predictive value (range, 30% to 40%) and high positive predictive value (range, 75% to 88%) of follow-up radiographs in patients with suspected scaphoid fractures with MRI as reference standard, which were not consistent with our findings. These differences can be explained as the prevalence influences the negative predictive value and positive predictive value (9, 31). Our study evaluated the radiographs of a consecutive series of patients (prevalence 18%) and Low et al. selected patients retrospectively if they had both follow-up radiographs and MRI after injury (prevalence 75%).

Our results show that the method of presenting radiographs may affect their evaluation by surgeon observers. We found that the interobserver reliability was the same in the JPEG and

DICOM group, but the diagnostic performance was better when radiographs were evaluated using a DICOM viewer compared to static JPEG images. The ability to window level, scroll, and zoom using a DICOM viewer improved the diagnosis of scaphoid fractures, in terms of sensitivity and negative predictive value, significantly. Since the format of medical images could

Table 3. Proportion of Agreement for the Recognition of (Consolidated) Scaphoid Fractures Based on 6-Week Radiographs (JPEG and DICOM Viewer)

	JPEG (n=53)		DICOM Viewer (n=28)	
Case No.	**Most provided answer**	**PA***	**Most provided answer**	**PA***
1	Present	79	Absent	57
2	Absent	64	Present	86
3	Absent	66	Absent	57
4	Present	57	Absent	75
5	Absent	62	Absent	61
6	Absent	70	Absent	75
7	Absent	57	Present	82
8	Present	51	Absent	75
9	Present	85	Present	57
10	Present	68	Absent	57
11	Present	74	Absent	54
12	Present	72	Present	93
13	Absent	60	Absent	64
14	Absent	83	Absent	71
15	Absent	85	Absent	79
16	Present	62	Present	68
17	Absent	70	Present	86
18	Absent	77	Present	79
19	Absent	55	Present/Absent	50
20	Present	53	Absent	61
21	Absent	77	Absent	64
22	Absent	87	Present	61
23	Absent	66	Present	57
24	Present	55	Present	61
25	Absent	74	Present	75
26	Present	70	Absent	64
27	Absent	87	Absent	57
28	Present	62	Absent	61
29	Absent	66	Absent	79
30	Absent	57	Present	75
31	Present	60	Absent	79
32	Absent	87	Absent	71
33	Absent	62	Absent	61
34	Absent	60	Present	61

* Proportion of agreement: the proportion of observers agreeing with the most provided answer.

7

be a source of variation between surgeons, it should be accounted for in future reliability and diagnostic studies.

Given the low agreement and poor diagnostic accuracy of 6-week radiographs for the recognition of scaphoid fractures in this study, surgeons and patients must accept that they are dealing with probabilities rather than certainties in the management of scaphoid fractures. For example, we cannot reduce the probability of missing a fracture to 0% with a negative predictive value of less than 100%. Using 6-week radiographs as reference standard for studying suspected scaphoid fractures is not advised for future studies. To date, observer experience, training, image presentation, training, and simplification of classifications are shown to have a limited effect on the reliability and accuracy of diagnosis and classification of fractures. At this time it remains unclear what interventions will improve reliability and accuracy, but our collaborative plans to continue studying variation between surgeons to attempt to reduce it.

Table 4. Diagnostic Performance of 6-Week Radiographs for the Recognition of (Consolidated) Scaphoid Fractures (JPEG versus DICOM viewer)

	JPEG (n=53)		**DICOM Viewer (n=28)**	
	%	**95%CI**	**%**	**95%CI**
Reference standard: MRI				
Sensitivity	42	37-47	64	57-71
Specificity	56	54-59	53	50-57
Accuracy	53	51-56	56	52-59
Positive predictive value	20	17-23	26	22-30
Negative predictive value	79	76-81	85	82-88
Reference standard: CT				
Sensitivity	56	50-62	79	72-85
Specificity	59	56-61	55	51-58
Accuracy	58	56-61	58	55-61
Positive predictive value	19	16-22	23	19-27
Negative predictive value	89	87-90	94	91-96
Reference standard: MRI + CT				
Sensitivity	52	45-59	75	67-83
Specificity	58	55-60	53	50-56
Accuracy	57	55-59	56	52-59
Positive predictive value	14	12-17	18	14-21
Negative predictive value	90	88-92	94	92-96

REFERENCES

1. Duckworth AD, Ring D, McQueen MM. Assessment of the suspected fracture of the scaphoid. J Bone Joint Surg Br. 2011;93(6):713-9.
2. Rhemrev SJ, de Zwart AD, Kingma LM, Meylaerts SA, Arndt JW, Schipper IB, et al. Early computed tomography compared with bone scintigraphy in suspected scaphoid fractures. Clinical nuclear medicine. 2010;35(12):931-4.
3. Akdemir UO, Atasever T, Sipahioglu S, Turkolmez S, Kazimoglu C, Sener E. Value of bone scintigraphy in patients with carpal trauma. Annals of nuclear medicine. 2004;18(6):495-9.
4. Beeres FJ, Rhemrev SJ, den Hollander P, Kingma LM, Meylaerts SA, le Cessie S, et al. Early magnetic resonance imaging compared with bone scintigraphy in suspected scaphoid fractures. J Bone Joint Surg Br. 2008;90(9):1205-9.
5. Mallee W, Doornberg JN, Ring D, van Dijk CN, Maas M, Goslings JC. Comparison of CT and MRI for diagnosis of suspected scaphoid fractures. J Bone Joint Surg Am. 2011;93(1):20-8.
6. De Zwart AD, Beeres FJ, Ring D, Kingma LM, Coerkamp EG, Meylaerts SA, et al. MRI as a reference standard for suspected scaphoid fractures. The British journal of radiology. 2012;85(1016):1098-101.
7. Tibrewal S, Jayakumar P, Vaidya S, Ang SC. Role of MRI in the diagnosis and management of patients with clinical scaphoid fracture. International orthopaedics. 2012;36(1):107-10.
8. 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(1):169-76.
9. Adey L, Souer JS, Lozano-Calderon S, Palmer W, Lee SG, Ring D. Computed tomography of suspected scaphoid fractures. J Hand Surg Am. 2007;32(1):61-6.
10. Groves AM, Kayani I, Syed R, Hutton BF, Bearcroft PP, Dixon AK, et al. An international survey of hospital practice in the imaging of acute scaphoid trauma. AJR American journal of roentgenology. 2006;187(6):1453-6.
11. Ring D, Lozano-Calderon S. Imaging for suspected scaphoid fracture. J Hand Surg Am. 2008;33(6):954-7.
12. Buijze GA, Mallee WH, Beeres FJP, Hanson TE, Johnson WO, Ring D. Diagnostic Performance Tests for Suspected Scaphoid Fractures Differ with Conventional and Latent Class Analysis. Clinical orthopaedics and related research. 2011;469(12):3400-7.
13. Pepe MS, Janes H. Insights into latent class analysis of diagnostic test performance. Biostatistics. 2007;8(2):474-84.
14. Gabler C, Kukla C, Breitenseher MJ, Trattnig S, Vecsei V. Diagnosis of occult scaphoid fractures and other wrist injuries. Are repeated clinical examinations and plain radiographs still state of the art? Langenbeck's archives of surgery / Deutsche Gesellschaft fur Chirurgie. 2001;386(2):150-4.
15. Munk B, Frokjaer J, Larsen CF, Johannsen HG, Rasmussen LL, Edal A, et al. Diagnosis of scaphoid fractures. A prospective multicenter study of 1,052 patients with 160 fractures. Acta orthopaedica Scandinavica. 1995;66(4):359-60.
16. 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(1):245-50.
17. Groves AM, Cheow HK, Balan KK, Courtney HM, Bearcroft PW, Dixon AK. Case report: False negative 16 detector multislice CT for scaphoid fracture. The British journal of radiology. 2005;78(925):57-9.
18. Mallee WH, Doornberg JN, Ring D, Maas M, Muhl M, van Dijk CN, et al. Computed tomography for suspected scaphoid fractures: comparison of reformations in the plane of the wrist versus the long axis of the scaphoid. Hand (N Y). 2014;9(1):117-21.
19. Tiel-van Buul MM, van Beek EJ, Broekhuizen AH, Nooitgedacht EA, Davids PH, Bakker AJ. Diagnosing scaphoid fractures: radiographs cannot be used as a gold standard! Injury. 1992;23(2):77-9.
20. Tiel-van Buul MM, van Beek EJ, Borm JJ, Gubler FM, Broekhuizen AH, van Royen EA. The value of radiographs and bone scintigraphy in suspected scaphoid fracture. A statistical analysis. Journal of hand surgery (Edinburgh, Scotland). 1993;18(3):403-6.
21. Low G, Raby N. Can follow-up radiography for acute scaphoid fracture still be considered a valid investigation? Clinical radiology. 2005;60(10):1106-10.

22. Doornberg J, Lindenhovius A, Kloen P, van Dijk CN, Zurakowski D, Ring D. Two and three-dimensional computed tomography for the classification and management of distal humeral fractures. Evaluation of reliability and diagnostic accuracy. J Bone Joint Surg Am. 2006;88(8):1795-801.
23. Guitton TG, Ring D. Interobserver reliability of radial head fracture classification: two-dimensional compared with three-dimensional CT. J Bone Joint Surg Am. 2011;93(21):2015-21.
24. Neuhaus V, Bot AG, Guitton TG, Ring DC, Abdel-Ghany MI, Abrams J, et al. Scapula fractures: interobserver reliability of classification and treatment. J Orthop Trauma. 2014;28(3):124-9.
25. Mellema JJ, Doornberg JN, Guitton TG, Ring D. Biomechanical studies: science (f)or common sense? Injury. 2014;45(12):2035-9.
26. Doornberg JN, Guitton TG, Ring D. Diagnosis of elbow fracture patterns on radiographs: interobserver reliability and diagnostic accuracy. Clin Orthop Relat Res. 2013;471(4):1373-8.
27. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Statistics in medicine. 1998;17(1):101-10.
28. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics. 1977;33(2):363-74.
29. Siegel S, Castellan JN. Nonparametric Statistics for the Behavioral Sciences. Siegel S, Castellan JN, editors. New York: McGraw-Hill; 1988.
30. Harper R, Reeves B. Reporting of precision of estimates for diagnostic accuracy: a review. BMJ. 1999;318(7194):1322-3.
31. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ. 1994;309(6947):102.

CHAPTER 8

DIAGNOSTIC PERFORMANCE TESTS FOR SUSPECTED SCAPHOID FRACTURES DIFFER WITH CONVENTIONAL AND LATENT CLASS ANALYSIS

Buijze GA
Mallee WH
Beeres FJP
Hanson TE
Johnson WO
Ring D

Clinical Orthopaedics and Related Research. September 2011

ABSTRACT

Background

Evaluation of the diagnostic performance characteristics of radiographic tests for diagnosing a true fracture among suspected scaphoid fractures is hindered by the lack of a consensus reference standard. Latent class analysis is a statistical method that takes advantage of unobserved, or latent, classes in the data that can be used to determine diagnostic performance characteristics when there is no consensus reference (gold) standard.

Purposes

We therefore compared the diagnostic performance characteristics of MRI, CT, bone scintigraphy, and physical examination to identify true fractures among suspected scaphoid fractures.

Patients and Methods

We used data from two studies, one that prospectively studied 34 patients who had MRI and CT of the wrist, and a second that studied 78 patients who had MRI, bone scintigraphy, and structured physical examination. We compared the diagnostic performance characteristics calculated by latent class analysis with those calculated using formulas based on a reference standard.

Results

In the first cohort, the calculated sensitivity and specificity with latent class analysis were different than those with traditional reference standard-based calculations for the CT in the scaphoid planes (sensitivity, 0.78 versus 0.67; specificity, 1.0 versus 0.96) and the MRI (sensitivity, 0.80 versus 0.67; specificity, 0.93 versus 0.89). In the second cohort, the greatest differences were in the sensitivity of MRI (0.84 versus 0.75) and the sensitivities of physical examination maneuvers (range, 0.63–0.73 versus 1.0).

Conclusions

The diagnostic performance characteristics calculated using latent class analysis may differ from those calculated according to formulas based on a reference standard. We believe latent class analysis merits further study as an option for assessing diagnostic performance characteristics for orthopaedic conditions when there is no consensus reference standard.

Level of Evidence

Level II, prognostic study.

INTRODUCTION

Investigations to evaluate the diagnostic performance characteristics of tests used to diagnose true fractures among suspected scaphoid fractures are hindered by the lack of a consensus reference standard. Reference standards for a true fracture in various studies have included followup radiography and/or clinical signs between 10 days and 12 months after injury (2, 33), followup MRI (17), and standards based on a combination of test results (5). A recent systematic review of diagnostic tests for suspected scaphoid fractures documents substantial variation in diagnostic performance characteristics, ie, sensitivity (Se) and specificity (Sp) for MRI (Se, 0.80–1.0; Sp, 0.95–1.0), CT (Se, 0.73–1.0; Sp, 1.0), bone scintigraphy (Se, 0.78–1.0; Sp, 0.52–1.0), and ultrasound (Se, 0.78–1.0; Sp, 0.89–0.98) (38). Inconsistency in imaging protocols and reference standards might account for much of this variation.

Latent class analysis is a statistical method that identifies unobserved or latent classes (factors associating with one another) in data. Latent class analysis has proved helpful for the evaluation of diagnostic tests when no reference standard is available (18). An example of a disease for which there is no accepted reference (gold) standard for diagnosis is compartment syndrome. Latent class analysis takes advantages of known but unobserved groupings of patients based on disease status. Although there can be more than two groups, only two are considered here, namely 'diseased' or 'not diseased'. A statistical analysis of these two groups leads to calculations of estimated probabilities of disease, without knowing which patients have the disease and which do not.

Latent class analysis relies on the results of multiple data points or test results in a population of patients. The estimation of test accuracies and prevalence are performed using either maximum likelihood (ML) (which is a standard method of statistical inference (9, 14) that obtains parameter estimates that maximize the probability of observing the actual data), or the Bayesian method (4, 10) (which incorporates scientific knowledge into the data analysis that is independent of the currently sampled data, and which does so by simply obeying known probability laws), or both. The quality of inferences based on ML estimation depend on having a reasonable model for the data, on having large sample sizes, and on not having estimates that are too close to one or zero (for example, they will not work well if one of the tests is nearly perfect). Bayesian methods also rely on having a reasonable model for the data, but they do not rely on having large sample sizes or on having estimates that are not near zero or one. The downside of Bayesian methods is that they rely on expert estimations of the actual situation, which if accurate will improve inferences, but if not will hinder them unless sample sizes are reasonably large (in which case these estimations play a lesser role in the final inference).

Other diseases lacking a consensus reference standard have been studied using latent class analysis, such as peripheral joint psoriatic arthritis (32), carpal tunnel syndrome (22), and various infectious diseases (3, 8, 11, 34). Its use in some of these studies confirm that the diagnostic performance values of various tests are similar to those found with traditional analysis based on a reference standard, which supports the accuracy of the reference standard.

In a previous publication we explored the application of latent class analysis in orthopaedic diagnostic studies and provided a brief description of the current study and its conclusions

(7). This publication is intended to provide a complete description of that study to assess the diagnostic performance characteristics of true fractures among suspected scaphoid fractures using latent class analysis and using standard formulas based on a reference standard.

PATIENTS AND METHODS

We applied latent class analysis to data from two prospective cohort studies: in one we compared MRI with CT, and in the other we compared MRI with bone scintigraphy and clinical tests. Both trials were approved by a Medical Ethical Committee and all patients gave written informed consent for participation.

The first cohort (MRI versus CT) included 34 patients diagnosed with a suspected scaphoid fracture in the Emergency Department between April and October 2008 (25). We included adult patients presenting within 24 hours of injury and having tenderness of the scaphoid in the anatomic snuffbox and normal scaphoid-specific radiographs with a minimum of three views. We excluded patients with any concurrent distal ulna, radius, or carpal fracture, previous scaphoid fracture, rheumatoid arthritis, and cognitive dysfunction limiting clinical evaluation. At the time of treatment all radiographs were independently evaluated by the treating radiologist and the treating trauma surgeon. All patients underwent MRI and CT. We performed both examinations on the same day, at an average of 3.6 days (range, 0–10 days) after initial trauma. All MRI studies were performed with an open 1.0 Tesla MR scanner (Panorama 1.0 T, Philips Medical Systems, Eindhoven, The Netherlands). The standard scaphoid protocol (Sense wrist coil), with a slice thickness of 3 mm and 0.6-mm gap, included the following series: localizer, Cor STIR, and Cor SE T1. The patient was positioned supine with the forearm and wrist alongside the body. The open MR scanner allowed for central placement of the hand relative to the magnetic field, resulting in improved image quality when compared with off-centered scanning in a conventional tube. Multidetector, high-resolution CT was performed in all patients using a 64-slice CT-scan (Brilliance, Philips Medical Systems, Eindhoven, The Netherlands) in the following sequence: high-resolution 0.5-mm slices section thickness. The scan covered the wrist from the distal radioulnar joint to the carpometacarpal joints. Patients were positioned in the "superman" position, prone with the affected arm above the body and the palm facing down. We made reconstructions in planes, defined by the long axis of the scaphoid (30). Sagittal plane images of the scaphoid were defined as reconstructions that provided a lateral view of the scaphoid bone, as defined by the central longitudinal axis of the scaphoid. Coronal plane images were those that provided a posteroanterior view of the scaphoid in the anatomic plane and in line with the axis of the scaphoid (1, 24). Criteria for a scaphoid fracture on CT images were the presence of a sharp lucent line within the trabecular bone pattern, break in the continuity of the cortex, sharp step in the cortex, or dislocation of bone fragments. Criteria for a fracture on MRI included the presence of a cortical fracture line, trabecular fracture line, or combination of both. In addition to these criteria, any extensive focal zone of edema without a clear cortical fracture line, comparable with that seen with a stress fracture, was discussed to decide if the findings represented a fracture. Three of us (JCG, MM, and CNvD) formed the panel that evaluated MR images, CT images, and all radiographs at the nominal 6-week followup (average, 48 days; range, 35–74 days postinjury) until a consensus opinion was reached. Interobserver reliability, measured with the multirater kappa measure described by Siegel and Castellan (31), was κ = 0.62, which reflects overall substantial agreement. The reference standard for a true scaphoid fracture was an abnormal lucent line in the scaphoid (26).

The second cohort (MRI versus bone scintigraphy and clinical tests) included 78 patients who visited one emergency department for a suspected scaphoid fracture between April 2004 and January 2007 (28). We included adult patients with a suspected scaphoid fracture (tender anatomic snuffbox and pain in the snuffbox when applying axial pressure on the first or second digit), recent trauma (within 48 hours), and no evidence of a fracture on scaphoid-specific radiographs. We excluded patients with polytraumatic injuries and patients with bilateral suspected scaphoid fractures. Clinical tests were performed at initial presentation, MRI within 24 hours, and bone scintigraphy between 3 and 5 days after trauma. Experienced physicians performed all clinical tests, according to a predefined and standardized method on the suspected and contralateral sides, consisting of (1) inspection of the anatomic snuffbox for the presence of a hematoma and/or swelling in comparison to the contralateral side, (2) measurements of range of wrist flexion and extension, (3) measurements of supination and pronation strength using a custom-made hydraulic dynamometer (LUMC, Leiden, Netherlands), and (4) measurements of grip strength using a hydraulic hand dynamometer (Saehan Corporation, Masan, Korea). All measurements were expressed as a percentage of the uninjured side. Motion and strength tests were considered positive if there was a loss of 25% or greater compared with the uninjured side. MRI studies were performed with a 1.5 Tesla MR scanner (Siemens Medical Solutions, Erlangen, Germany). The patient lay prone on the scanner couch with the hand suspected of a scaphoid fracture extended forward, palm down, over his or her head. The flexible surface coil then was wrapped around the wrist. The MRI protocol included coronal T1-weighted turbo spin-echo images with a TR of 450 ms, TE of 13 ms, field view of 180 mm, base resolution of 512, two averages, slice thickness of 3 mm with a distance factor of 10%, and scan time of 2.17 minutes. The parameters for the coronal fat-suppressed T2-weighted fast spin-echo images were 5220/73 ms (TR/TE), field of view of 220 mm, base resolution of 448, three averages, slice thickness of 3 mm with a distance factor of 10%, and scan time of 4.33 minutes. All MRI scans were independently rated by two radiologists (EGC and LMK). Bone scintigraphy was performed using a standard protocol of images of the early static phase, on a SKYlight gamma camera (Philips Medical Systems, Eindhoven, The Netherlands). Palmar and dorsal images of both wrists were obtained between 2.5 and 4 hours after injection of 500 MBq of technetium-99 m diphosphonate (Tc-99 m-HDP) to observe the osteoblast activity. Observations were performed by an experienced clinical nuclear physician (JWA). The reference standard for a true scaphoid fracture was a combination of MRI, bone scintigraphy, and clinical examination results. Where there was a discrepancy between MRI and bone scintigraphy (ie, only one tested positive), a true fracture was defined as an abnormal lucent line in the scaphoid observed on radiographs at the 6-week followup or as scaphoid tenderness more than 2 weeks after injury.

Latent class analysis looks for groups of test results (or latent classes) that represent levels of disease probability. The latent classes cannot be observed directly (eg, a fracture), but the resultant (eg, a sharp lucent line in the trabecular bone pattern on CT) from which these latent characteristics are inferred can be observed. Depending on whether the results of the tests are related, two methods can be used. These methods have been described previously in more detail with examples (7).

The ML-based method, developed by Hui and Walter (18), assumes conditional independence of the tests, meaning that presence or absence of one symptom, sign, or test result is unrelated to the presence or absence of all others, conditional on true disease status. Walter designed the program LATENT1 (Latent1 Software, Version 3, McMaster University, Hamilton, Ontario, Canada), which calculates the ML estimates and gives confidence intervals for test accuracies and prevalence. In addition to the basic parameters, LATENT1 provides positive predictive values for each pattern of test results or latent class. Because we assumed that the four diagnostic tests of the first cohort (MRI, CT in the planes of the scaphoid or wrist, and 6-week radiography) met the conditional independence criteria, we used LATENT1 software for analysis. This assumption was based on the fact that the interpretation of each these tests was blinded to the result of the other tests.

In the second cohort (MRI versus bone scan), we did not expect the seven clinical test results to be unrelated to the others because the examiner knew the result of each test. Therefore, the data violate the conditional independence assumption of standard latent class analysis, and require the use of a recently developed latent class analysis model based on Bayesian methods that allow for conditional dependence among multiple test results by relying on surgeon estimation of plausible dependencies between test results (21). Johnson et al. (20) provided Bayesian methods for the Hui-Walter model and Dendukuri and Joseph (12) extended the model to incorporate two additional dependence parameters (one for each latent class), in the case of two diagnostic tests. We considered all clinical tests to be conditionally dependent, and we considered MRI to be independent of all tests other than the reference standard and bone scintigraphy.

We based our surgeon estimates on the lowest thresholds of each parameter's range that was reported in a review of the literature (38). Specifically we selected the following values: 0.78 (6) for sensitivity and 0.52 (27) for specificity of bone scintigraphy; 0.8 (5) for sensitivity and 0.95 (19) for specificity of MRI; and 0.05 for prevalence of true fractures among suspected scaphoid fractures (37) (Appendix 1).

RESULTS

The diagnostic performance characteristics calculated using latent class analysis differed from those calculated using the traditional methods based on a reference standard in both cohorts. In the first cohort, both methods showed CT in the scaphoid planes had the highest diagnostic performance values, and CT in the axial and sagittal planes had the lowest (Table 1). For the latter, the diagnostic performance values were similar in both methods (Se, 0.16 versus 0.17; Sp, 0.89 versus 0.89). However, the sensitivity and specificity of CT in the scaphoid planes

(Se, 0.78 versus 0.67; Sp, 1.0 versus 0.96) and MRI (Se, 0.80 versus 0.67; Sp, 0.93 versus 0.89) were slightly higher in the latent class analysis than the reference standard calculations and the prevalence was slightly greater (18.9% versus 17.6%). The positive predictive value in latent class analysis for a true scaphoid fracture in case of positive CT and MRI was 1.0, regardless of a negative 6-week radiography result; and in case of negative CT and MRI studies and positive 6-week radiography, the positive predictive value was 0.21 (Table 2).

Table 1. Latent class analysis versus reference standard calculations for Cohort 1 (MRI versus CT)

	Latent class analysis		**Calculations using reference standard**	
Diagnostic test	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
CT (scaphoid plane)	0.78 (0.36–1.0)	1.0 (0.96–1.0)	0.67 (0.36–0.80)	0.96 (0.90–0.99)
CT (wrist)	0.16 (0–0.45)	0.89 (0.77–1.0)	0.17 (0.03–0.44)	0.89 (0.86–0.95)
MRI	0.80 (0.41–1.0)	0.93 (0.83–1.0)	0.67 (0.34–0.89)	0.89 (0.82–0.94)
Radiographs (6 weeks)	0.80 (0.40–1.0)	0.97 (0.89–1.00)		
Prevalence (%)	18.9		17.6	

CI = Confidence Interval

Table 2. Positive predictive values for fracture given a particular pattern of test outcomes using data from Cohort 1

CT scaphoid plane	**CT wrist**	**MRI**	**Radiographs (6 weeks)**	**Probability of scaphoid fracture**
-	-	-	-	0.0021
+	-	-	-	0.9986
-	+	-	-	0.032
+	+	-	-	0.9991
-	-	+	-	0.1055
+	-	+	-	1
-	+	+	-	0.1535
+	+	+	-	1
-	-	-	+	0.2105
+	-	-	+	1
-	+	-	+	0.2906
+	+	-	+	1
-	-	+	+	0.9377
+	-	+	+	1
-	+	+	+	0.9586
+	+	+	+	1

+ = positive test result
- = negative test result

When compared with the calculations based on a reference standard in the second cohort, the latent class analysis sensitivity was slightly lower for bone scintigraphy (Se, 0.94 versus 1.0) and the specificity was equal, whereas for the MRI, the sensitivity was substantially higher and the specificity was slightly lower (Se, 0.75 versus 0.84; Sp, 1.0 versus 0.99) (Table 3). Motion and strength test sensitivities of five tests ranged between 0.63 and 0.73 in the latent class analysis versus a sensitivity of 1.0 with the reference standard. With the exception of the test for loss greater than 25% of wrist flexion, the specificities of the other four tests on motion and strength were slightly higher in the latent class analysis (range, 0.14–0.27 versus range, 0.6–0.23). The latent class analysis estimates of the presence of snuffbox swelling showed a higher sensitivity (0.63 versus 0.42) and a lower specificity (0.41 versus 0.76), whereas for the presence of a hematoma, the estimates showed a lower sensitivity (0.36 versus 0.92) and a higher specificity (0.71 versus 0.32).

DISCUSSION

There is no consensus reference standard for a true scaphoid fracture. All previous studies have calculated diagnostic performance characteristics based on debatable reference standards such as radiographs of the scaphoid obtained 2 or 6 weeks after fracture. Our analysis shows that diagnostic performance characteristics calculated with latent class analysis are notably different from those calculated using traditional methods based on a reference standard. It

Table 3. Latent class analysis versus reference standard calculations for Cohort 2 (MRI versus bone scintigraphy versus clinical tests)

	Latent class analysis		Calculations using reference standard	
Diagnostic test	**Sensitivity (95% PI)**	**Specificity (95% PI)**	**Sensitivity (95% CI)**	**Specificity (95% CI)**
Snuffbox swelling	0.63 (0.45–0.8)	0.41 (0.31–0.54)	0.42 (0.20–0.66)	0.76 (0.72–0.80)
Hematoma	0.36 (0.18–0.54)	0.71 (0.61–0.80)	0.92 (0.68–0.99)	0.32 (0.27–0.33)
Flexion loss less than 25%	0.63 (0.46–0.82)	0.33 (0.22–0.43)	1.00 (0.78–1.0)	0.29 (0.25–0.29)
Extension loss less than 25%	0.72 (0.56–0.87)	0.27 (0.20–0.38)	1.00 (0.79–1.0)	0.23 (0.19–0.23)
Grip strength loss less than 25%	0.73 (0.52–0.89)	0.14 (0.07–0.22)	1.00 (0.83–1.0)	0.08 (0.05–0.08)
Pronation strength loss less than 25%	0.70 (0.53–0.87)	0.15 (0.10–0.25)	1.00 (0.82–1.0)	0.09 (0.06–0.09)
Supination strength loss less than 25%	0.65 (0.38–0.81)	0.15 (0.09–0.23)	1.00 (0.85–1.0)	0.06 (0.03–0.06)
MRI	0.84 (0.65–0.96)	0.99 (0.96–1.0)	0.75 (0.57–0.75)	1.00 (0.97–1.0)
Bone Scan	0.94 (0.80–0.99)	0.89 (0.79–0.95)	1.00 (0.80–1.0)	0.89 (0.86–0.89)
Reference standard	0.86 (0.56–0.97)	0.97 (0.91–0.99)		
Prevalence (%)	15.8		15.4	

PI = Probability Interval ; CI = Confidence Interval is not possible to state which numbers are more accurate, but the differences in the numbers emphasize that we are dealing with probabilities rather than certainties of fracture and that our choice of the reference standard can affect those probabilities. It is possible that latent class analysis will provide more accurate and meaningful probabilities, but this would need to be tested prospectively, using meaningful outcomes such as union, disability, time away from work and sport, and costs.

Our study had some limitations. First, our analysis is based on data made available to us and is subject to all its weaknesses enumerated in the previous publications, but primarily relate to small sample size for our purposes. Although the first cohort had a small sample size, the ML was applicable as all diagnostic tests met the conditional independence criteria and not only sample size but the ratio of the number of tests to sample size is important for reliability of the method. The ratio in Cohort 1 was deemed large enough for this method to be reliable, and additionally we presented the estimated 95% bootstrap confidence intervals, which are more appropriate than ML-based intervals when sample sizes are small to moderate, as is the case here. Second, there is the possibility that the estimations used in the Bayesian analysis of the second cohort are inaccurate. Third, the assumption of conditional independence of some of the tests could be incorrect in the first cohort (which introduces large biases if there is more than slight dependence (16), which we think is unlikely). Fourth, the model could not be validated as cross validation has not been used in latent class analyses and bootstrap is used to cope with small sample ML problems, among others.

Latent class analysis is increasingly used to study diagnostic tests for diseases lacking a consensus reference standard (3, 8, 11, 22, 32, 34), particularly in the field of psychiatry (13). In a study similar to ours, Faraone and Tsuang (13) analyzed prior data for the diagnosis of major depressive disorder (29) with traditional reference standard-based calculations and latent class analysis and found consistency between the statistical methods, suggesting that psychiatric diagnoses may be highly accurate.

Meta-analyses of diagnostic tests also can account for the lack of a reference standard by calculating adjusted summary receiver operating characteristic (SROC) curves using pooled diagnostic performance characteristics, allowing for the possibility of errors in the reference standard, through use of a latent class model (36). The model presumes the true disease status of each subject is unknown, or latent, and uses parameter estimates to calculate a set of fitted frequencies for the numbers of true (but unobserved) cases and noncases, adjusted for the misclassification in the reference standard.

Given the imperfect reference standards for diagnosis of a true fracture among suspected scaphoid fractures, it is not surprising that there were notable differences in the diagnostic performance characteristics calculated using traditional and latent class analyses in our two cohorts. In the first cohort, it is notable that the sensitivity and specificity of CT and MRI calculated by latent class analysis were in the range of those in a previous study (38), whereas those calculated using traditional analysis were not. The sensitivity and specificity of MRI calculated using analysis based on a reference standard were lower than the lowest previously reported (5, 19). All diagnostic performance parameters calculated by latent class analysis were closer to the average diagnostic parameters based on pooled data in a meta-analysis (38).

In the second cohort, the sensitivity of MRI calculated by latent class analysis also was closer to the average sensitivity based on pooled data in a meta-analysis (38). Physical tests of strength and motion were 100% sensitive according to calculations using a reference standard and only 63% to 73% sensitive when using latent class analysis, indicating their utility for triage of suspected scaphoid fractures is questionable. These results were comparable to those of Unay et al. (35), who evaluated the diagnostic performance characteristics of 10 physical examination maneuvers for the triage of suspected scaphoid fractures using MRI as the reference standard. In their study, sensitivities ranged between 67% and 79% and the specificities ranged between 20% and 75%. The reason that the traditional analysis overestimates the sensitivity of physical examination maneuvers in the second cohort is probably because physical examination was part of the reference standard for defining a true fracture. Latent class analysis can help determine shortcomings of reference standards.

According to latent class analysis, the reference standard used in the first cohort (radiographs taken 6 weeks after injury) is only 80% sensitive and 97% specific for a true fracture, and the reference standard used in the second cohort (a combination of radiographic and physical examination test results) from MRI and bone scintigraphy is only 86% sensitive and 97% specific. The most commonly used reference standard in the evaluation of diagnostic tests for triage of suspected scaphoid fractures is the absence of radiographic evidence of a scaphoid fracture on scaphoid-specific radiographs obtained a minimum of 6 weeks after injury (38). This reference standard is controversial (15, 23). Low and Raby (23) reported poor accuracy and reliability for followup radiography as a diagnostic test for scaphoid fractures with normal initial radiographs. Nondisplaced scaphoid fractures can be subtle, such that we cannot agree on a reliable reference standard. Furthermore, some nondisplaced fractures are not visible at the bone or articular surface because the cartilage is not disrupted, making even arthroscopy imperfect as a reference standard. It is conceivable that there will never be a consensus reference standard for the diagnosis of true fractures among suspected scaphoid fractures.

Given that the diagnostic performance characteristics of tests used for the diagnosis of true fractures among suspected scaphoid fractures are notably different depending on whether traditional or latent class analysis is used, additional research is needed to determine which method leads to better patient care. An imperfect or debated reference standard is commonplace in orthopaedic surgery and latent class analysis might merit wider utilization if it provides more accurate information that leads to better patient care. Given the inherent uncertainty in many diagnostic methods it might be appropriate, for many if not most illnesses, that patients and doctors base decisions on probabilities of disease rather than the traditional dichotomous, all or none, concept of disease.

ACKNOWLEDGEMENTS

We thank J.W. Arndt, MD, E.G. Coerkamp, MD, L.M. Kingma, MD, PhD, C.N. van Dijk, MD, PhD, J.C. Goslings, MD, PhD, and M. Maas, MD, PhD, for their contributions to this study.

APPENDIX 1. MODEL PARAMETERIZATION

The following information was incorporated into the model. This required that each parameter would be assumed to be larger than each of the 'lowest threshold' values with high certainty. Specifically, the prior probability of each parameter being larger than the lowest threshold was set to 95%. The beta (a,b) distribution describes a figure starting at 0 and ending at 1 that is entirely above the horizontal axis and which has a total area of 1, as areas underneath it correspond to modeled probabilities. The curves we used have 95% of the area above the lower threshold value and simply increase from that point on, indicating a lack of specific knowledge about particular values above them. Values of a and b were selected to have these characteristics. The beta (12.06, 1) (lower threshold is 0.78) and beta (4.58, 1) (lower threshold is 0.51) distributions were selected for the sensitivity and specificity of bone scintigraphy; and beta (13.43, 1) (lower threshold is 0.8) and beta (50.40, 1) (lower threshold is 0.94) for the sensitivity and specificity of MRI. For prevalence we selected beta (2.73, 9.0) which has a lower threshold of 0.07 and a most likely value of 0.18.

REFERENCES

1. Adey L, Souer JS, Lozano-Calderon S, Palmer W, Lee SG, Ring D. Computed tomography of suspected scaphoid fractures. J Hand Surg Am. 2007;32:61–66.
2. Akdemir UO, Atasever T, Sipahioglu S, Turkolmez S, Kazimoglu C, Sener E. Value of bone scintigraphy in patients with carpal trauma. Ann Nucl Med. 2004;18:495–499.
3. Baughman AL, Bisgard KM, Cortese MM, Thompson WW, Sanden GN, Strebel PM. Utility of composite reference standards and latent class analysis in evaluating the clinical accuracy of diagnostic tests for pertussis. Clin Vaccine Immunol. 2008;15:106–114.
4. Bayes T. An essay towards solving a problem in the doctrine of chances. Phil Trans R Soc London. 1763;53:370–418.
5. Beeres FJ, Rhemrev SJ, den Hollander P, Kingma LM, Meylaerts SA, le Cessie S, Bartlema KA, Hamming JF, Hogervorst M. Early magnetic resonance imaging compared with bone scintigraphy in suspected scaphoid fractures. J Bone Joint Surg Br. 2008;90:1205–1209.
6. Breederveld RS, Tuinebreijer WE. Investigation of computed tomographic scan concurrent criterion validity in doubtful scaphoid fracture of the wrist. J Trauma. 2004;57:851–854.
7. Buijze GA, Hanson TE, Johnson W, Ring D. Latent class analysis to determine the accuracy of diagnostic tests in orthopaedics. Orthop J Harvard Med School. 2010;12:106–108.
8. Butler JC, Bosshardt SC, Phelan M, Moroney SM, Tondella ML, Farley MM, Schuchat A, Fields BS. Classical and latent class analysis evaluation of sputum polymerase chain reaction and urine antigen testing for diagnosis of pneumococcal pneumonia in adults. J Infect Dis. 2003;187:1416–1423.
9. Casella G, Berger RL. Statistical Inference. Ed 2. Pacific Grove, CA: Duxbury Press; 2001.
10. Christensen R, Johnson WO, Branscum A, Hanson TF. Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians. Ed 3. Boca Raton, FL: CRC Press; 2010.
11. De La Rosa GD, Valencia ML, Arango CM, Gomez CI, Garcia A, Ospina S, Osorno S, Henao A, Jaimes FA. Toward an operative diagnosis in sepsis: a latent class approach. BMC Infect Dis. 2008;8:18.
12. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics. 2001;57:158–167.
13. Faraone SV, Tsuang MT. Measuring diagnostic accuracy in the absence of a "gold standard". Am J Psychiatry. 1994;151:650–657.
14. Fisher RA. On the mathematical foundations of theoretical statistics. Phil Trans R Soc A.1922;222:309–368.
15. Gabler C, Kukla C, Breitenseher MJ, Trattnig S, Vecsei V. Diagnosis of occult scaphoid fractures and other wrist injuries: are repeated clinical examinations and plain radiographs still state of the art? Langenbecks Arch Surg. 2001;386:150–154.
16. Georgiadis MP, Johnson WO, Singh R, Gardner IA. Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. J Royal Stat Soc C. 2003;52:63–76.
17. Groves AM, Kayani I, Syed R, Hutton BF, Bearcroft PP, Dixon AK, Ell PJ. An international survey of hospital practice in the imaging of acute scaphoid trauma. AJR Am J Roentgenol. 2006;187:1453–1456.
18. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. Biometrics.1980;36:167–171.
19. 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;168:1287–1293.
20. Johnson WO, Gastwirth JL, Pearson LM. Screening without a "gold standard": the Hui-Walter paradigm revisited. Am J Epidemiol. 2001;153:921–924.
21. Jones G, Johnson WO, Hanson TE, Christensen R. Identifiability of models for multiple diagnostic testing in the absence of a gold standard. Biometrics. 2010;66:855–863.
22. LaJoie AS, McCabe SJ, Thomas B, Edgell SE. Determining the sensitivity and specificity of common diagnostic tests for carpal tunnel syndrome using latent class analysis. Plast Reconstr Surg. 2005;116:502–507.
23. Low G, Raby N. Can follow-up radiography for acute scaphoid fracture still be considered a valid investigation? Clin Radiol. 2005;60:1106–1110.

24. Lozano-Calderon S, Blazar P, Zurakowski D, Lee SG, Ring D. Diagnosis of scaphoid fracture displacement with radiography and computed tomography. J Bone Joint Surg Am. 2006;88:2695–2703.
25. Mallee W, Doornberg JN, Ring D, van Dijk CN, Maas M, Goslings JC. Comparison of CT and MRI for diagnosis of suspected scaphoid fractures. J Bone Joint Surg Am.2011;93:20–28.
26. Memarsadeghi M, Breitenseher MJ, Schaefer-Prokop C, Weber M, Aldrian S, Gabler C, Prokop M. Occult scaphoid fractures: comparison of multidetector CT and MR imaging. Initial experience. Radiology. 2006;240:169–176.
27. Nielsen PT, Hedeboe J, Thommesen P. Bone scintigraphy in the evaluation of fracture of the carpal scaphoid bone. Acta Orthop Scand. 1983;54:303–306.
28. Rhemrev SJ, Beeres FJ, van Leerdam RH, Hogervorst M, Ring D. Clinical prediction rule for suspected scaphoid fractures: a prospective cohort study. Injury. 2010;41:1026–1030.
29. Rice JP, Endicott J, Knesevich MA, Rochberg N. The estimation of diagnostic sensitivity using stability data: an application to major depressive disorder. J Psychiatr Res.1987;21:337–345.
30. Sanders WE. Evaluation of the humpback scaphoid by computed tomography in the longitudinal axial plane of the scaphoid. J Hand Surg Am. 1988;13:182–187.
31. Siegel S, Castellan JN. Nonparametric Statistics for the Behavioral Sciences. New York, NY: McGraw-Hill; 1988.
32. Symmons DP, Lunt M, Watkins G, Helliwell P, Jones S, McHugh N, Veale D. Developing classification criteria for peripheral joint psoriatic arthritis: Step I. Establishing whether the rheumatologist's opinion on the diagnosis can be used as the "gold standard". J Rheumatol. 2006;33:552–557.
33. Thorpe AP, Murray AD, Smith FW, Ferguson J. Clinically suspected scaphoid fracture: a comparison of magnetic resonance imaging and bone scintigraphy. Br J Radiol.1996;69:109–113.
34. Tuyisenge L, Ndimubanzi CP, Ndayisaba G, Muganga N, Menten J, Boelaert M, Van den Ende J. Evaluation of latent class analysis and decision thresholds to guide the diagnosis of pediatric tuberculosis in a Rwandan reference hospital. Pediatr Infect Dis J.2010;29:e11–e18.
35. Unay K, Gokcen B, Ozkan K, Poyanli O, Eceviz E. Examination tests predictive of bone injury in patients with clinically suspected occult scaphoid fracture. Injury.2009;40:1265–1268.
36. Walter SD, Irwig L, Glasziou PP. Meta-analysis of diagnostic tests with imperfect reference standards. J Clin Epidemiol. 1999;52:943–951.
37. Wilson AW, Kurer MH, Peggington JL, Grant DS, Kirk CC. Bone scintigraphy in the management of X-ray-negative potential scaphoid fractures. Arch Emerg Med.1986;3:235–242.
38. Yin ZG, Zhang JB, Kan SL, Wang XG. Diagnosing suspected scaphoid fractures: a systematic review and meta-analysis. Clin Orthop Relat Res. 2010;468:723–734.

CHAPTER 9a

COMPUTED TOMOGRAPHY VERSUS MAGNETIC RESONANCE IMAGING VERSUS BONE SCINTIGRAPHY FOR CLINICALLY SUSPECTED SCAPHOID FRACTURES IN PATIENTS WITH NEGATIVE PLAIN RADIOGRAPHS

Mallee WH
Wang J
Poolman RW
Kloen P
Maas M
de Vet HCW
Doornberg JN

Cochrane Database of Systematic Reviews. June 2015

** The tables: 'Characteristics of included studies' and 'Characteristics of excluded studies' can be found in the online application only*

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.

PLAIN LANGUAGE SUMMARY

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.

BACKGROUND

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(1). The scaphoid is the most commonly fractured carpal bone(1-4). 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(5, 6). If untreated, this can lead to non-union, with or without avascular necrosis, and finally carpal collapse and disability(5, 7). Early detection and adequate treatment can provide predictable and satisfactory rates of healing(8). In contrast, delay of diagnosis and failure to recognise displacement are important risk factors for non-union of scaphoid wrist fractures(9, 10).

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(11-13). 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(14). Most scaphoid fractures will be identified with this imaging technique, but up to 16% are missed on initial radiographs(15, 16). 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(6, 7). 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(16, 17).

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)(18-21). 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(22, 23). 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 revealed equivalent imaging strategies for suspected scaphoid fractures in only 6.7% of the, mainly university, hospitals(23). 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(24).

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(16). 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(24). 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(25). 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(8, 26, 27). To avoid this burden, operative fixation with a headless compression screw can be performed(28). Surgical treatment is favourable in terms of time off work and functional outcome, but can lead to more (minor) complications(29).

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(16, 30, 31). 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(16). No clear criteria for a bruise or a fracture are established. When bone bruising is detected, the possibility of fracture development must be remembered (32); 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(33), 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(17). 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(34);
- in the lead author's hospital, the costs of BS are comparable with those for MRI.

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(35). In addition, an international survey of imaging strategies among hospitals revealed no use of US for these injuries(23). This review therefore does not consider US.

Another test, six-week follow-up radiographs, is extensively used in literature as a reference standard(16, 36); but its accuracy is being questioned(16). 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.

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(37). 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(23). 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 (14, 35). 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 (14). 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.

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.

Secondary objectives

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

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).

METHODS

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 (online only).

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.

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.

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 retrospective and consecutive versus non-consecutive; randomised);
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(38) (Whiting 2003). 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 1).

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 where 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 x 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(39)), we treated these as negative findings. If the data presented in trials had been uninterpretable in that 2 x 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), 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.

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 x 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.

Table 1. QUADAS checklist and assessment criteria

	Item question
1.	Was the spectrum of patients representative of the patients who will receive the test in practice? (representative spectrum)
2.	Is the reference standard likely to classify the target condition correctly? (acceptable reference standard)
3.	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)
4.	Did the whole sample or a random selection of the sample receive verification using the intended reference standard? (partial verification avoided)
5.	Did patients receive the same reference standard irrespective of the index test result? (differential verification avoided)
6.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)? (incorporation avoided)
7.	Were the reference standard results interpreted without knowledge of the results of the index test? (index test results blinded)
8.	Were the index test results interpreted without knowledge of the results of the reference standard? (reference standard results blinded)
9.	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice? (relevant clinical information)
10.	Were uninterpretable/intermediate test results reported? (uninterpretable results reported?)
11.	Were withdrawals from the study explained? (withdrawals explained)
12.	Did the study provide a clear definition of what was considered to be a ‘positive’ result?

Item answer
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
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
Yes: 1) if reference standard is 6-week follow-up radiographs (this is the most commonly used reference standard); 2) if 2 index tests report the same outcome; or 3) if 1 index test is used as a reference standard combined with clinical evaluation
Unclear: suboptimal would be if only 1 index test is used
No: 1) if only clinical evaluation after 6 weeks is considered to be the reference standard; or 2) if only clinical evaluation or radiographs, or both, after 2 weeks is considered to be the reference standard; 3) if insufficient information is given
Yes: if average interval between trauma and follow-up radiographs was 6 to 14 weeks. We will allow follow-up radiographs taken at least 2 weeks after trauma although this is considered to be a suboptimal reference standard
No: if interval was not clearly reported or before 2 weeks or greater than 14 weeks after trauma
Yes: if all patients received both index test and reference standard. We will allow for a random selection
Unclear: if insufficient information was available to judge this
No: if some of the patients who received the index test did not receive verification of their true disease state, and the selection of patients to receive the reference standard was not random
Yes: if all patients received the same reference standard, irrespective of the index test result
Unclear: if it is unclear whether different reference standards were used
No: if the outcome of the index test influenced the choice of reference standard
Yes: if index test was not part of the reference standard
Unclear: unclear
No: if index test was part of the reference standard
Yes: if the evaluation was blinded from the index test results
Unclear: if insufficient information was given on the blinded evaluation of the reference standard
No: if the index test results were present during evaluation of the reference standard
Yes: if the evaluation of the index test results was blinded from the results of the reference standard
Unclear: if insufficient information was given on the blinded evaluation of the index test
No: if the results of the reference standard were present during evaluation of the index test
Yes: if available clinical data during evaluation of the test are the same as in daily practice
Unclear: if insufficient information is given on the available clinical data during evaluation of the test
No: if the usual clinical data were not available during evaluation of the test
Yes: if the number of uninterpretable/intermediate test results is stated or if results match the number of initially included patients
Unclear: if insufficient information to permit judgement
No: if uninterpretable/intermediate test results are reported, without amount, or were excluded
Yes: if any withdrawals are stated and explained
Unclear: if insufficient information to permit judgement
No: if withdrawals are not mentioned or explained
Yes: if fracture criteria are well defined, even though they can differ between studies
Unclear: if insufficient information but evaluation was performed by at least 2 observers
No: if no fracture criteria are defined

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.

RESULTS

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(16, 18, 21, 36, 39-45); and 45 studies were excluded, one of which was published in two reports(46). 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 1.

Included studies

The characteristics of the individual studies are reported in the Characteristics of included studies.

Four studies evaluated CT (16, 36, 41, 42); five studies evaluated MRI(16, 18, 36, 40, 44); and six studies evaluated BS(21, 39-41, 43, 45). Of these studies, two compared CT with MRI (16, 36); one study compared BS with CT (41); and one compared MRI with BS(40).

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 (18); and the other four of which evaluated BS (21, 39, 43, 45). The gender distribution was available for 10 studies, in which the proportion of men ranged from 49.7% (41), to 100% (42).

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 (18, 39, 43, 45). In seven studies, the index test was performed within 10 days of trauma (16, 18, 36, 39-42).

Tenderness in the anatomical snuffbox was clearly incorporated in clinical evaluation in six studies (16, 18, 21, 40-42). One study reported 'pain over the scaphoid' as being clinically suspected (36). Four studies did not define the content of clinical evaluation (39, 43-45). Images of BS were evaluated by a consultant clinical nuclear physician in four studies (21, 40, 41, 44); 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) (39, 43, 45). 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 (online only). 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.

Methodological quality of included studies

The included studies were diverse but all were of moderate to good quality (Figure 2 and Figure 3). Five studies were considered 'good quality' (16, 18, 21, 36, 39); and six studies were considered 'moderate quality'. Of these, two studies had three items scored as low quality (40, 41) and one study had five items scored as unclear and one item scored as low quality (43).

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 was the spectrum of patients not clear(45); 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.

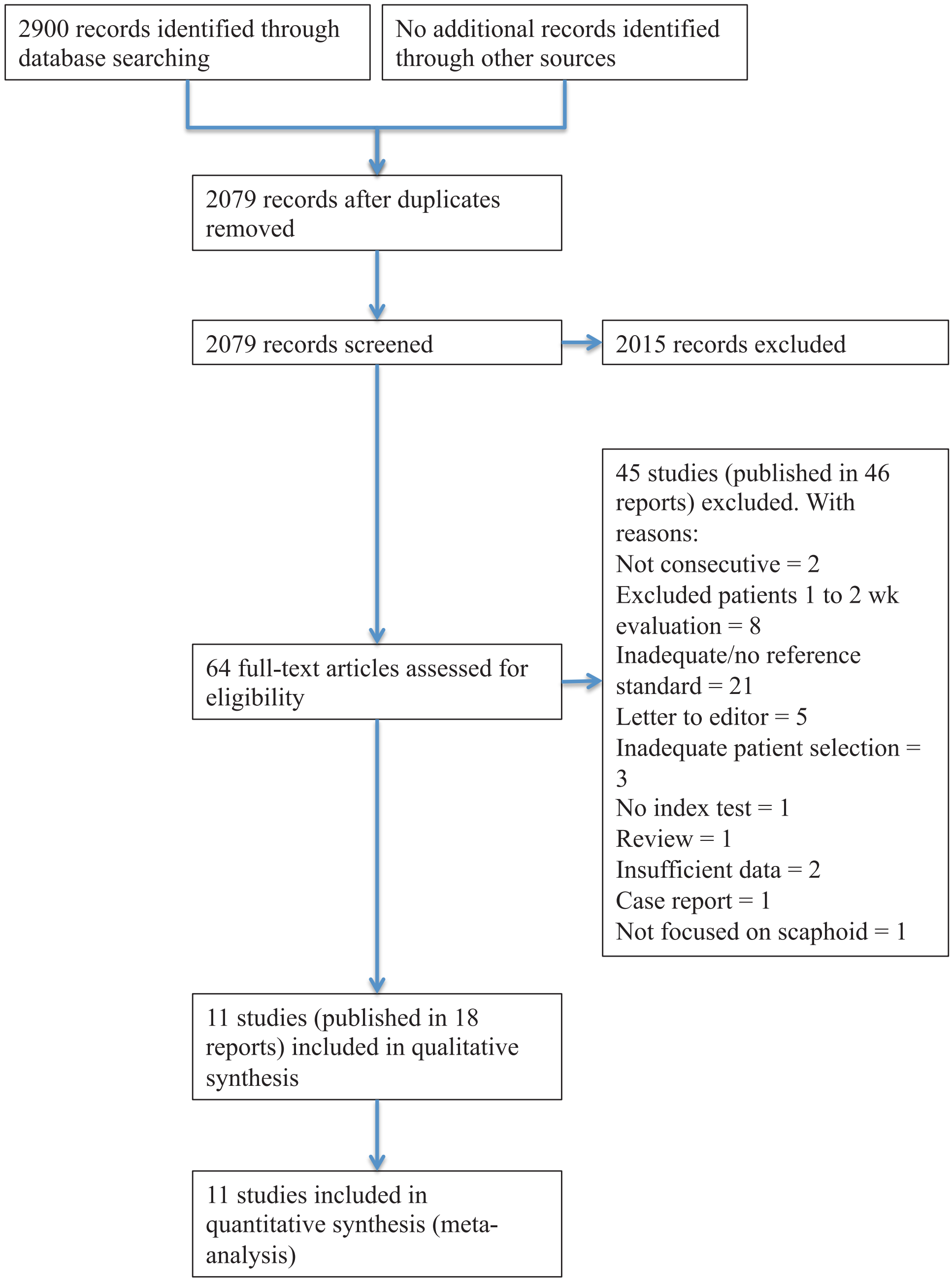


Figure 1. Study Flow Diagram

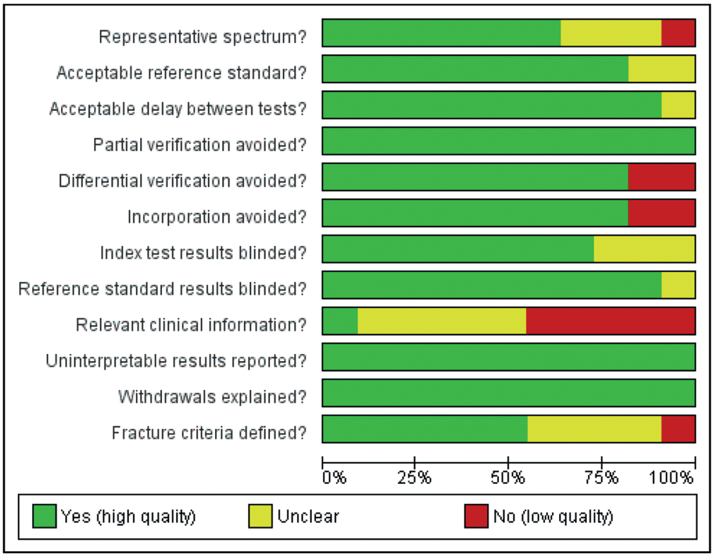


Figure 2. Methodological Quality Graph: review authors' judgements about each methodological quality item presented as percentages across all included studies

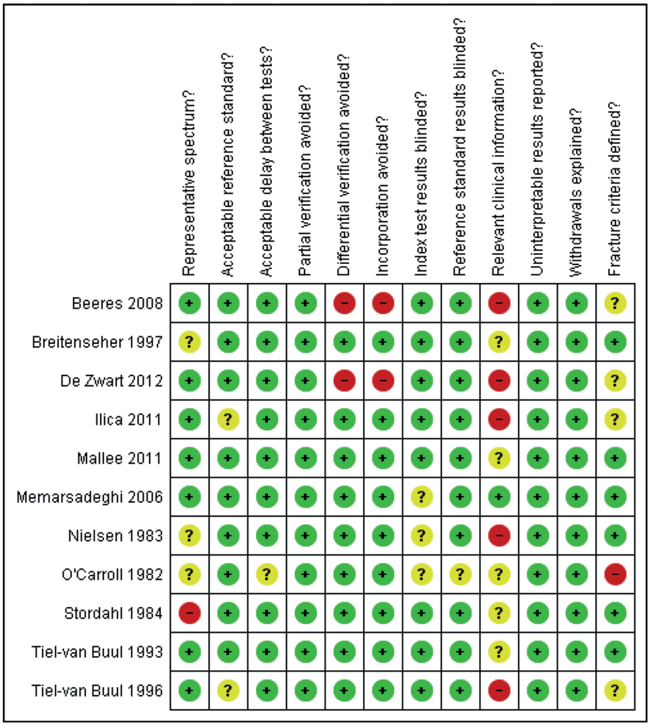


Figure 3. Methodological quality summary: review authors' judgements about each methodological quality item for each included study

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 (16, 18, 21, 36, 39, 43, 45); and two studies used a mixed reference standard (same outcome in two index tests or six-week follow-up radiographs) (40, 41). Two studies were judged at lower quality as they used suboptimal reference standards: one used MRI (42); and the other used BS (44). 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 (40, 41). Only one study reported the use of clinically relevant information during evaluation of the images (36); five studies excluded this information intentionally (39-42, 44). The criteria for diagnosing a fracture was not defined in O'Carroll 1982 for BS; for CT in De Zwart 2012 and Ilica 2011; and for MRI in Beeres 2008 and Tiel-van Buul 1996. 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.

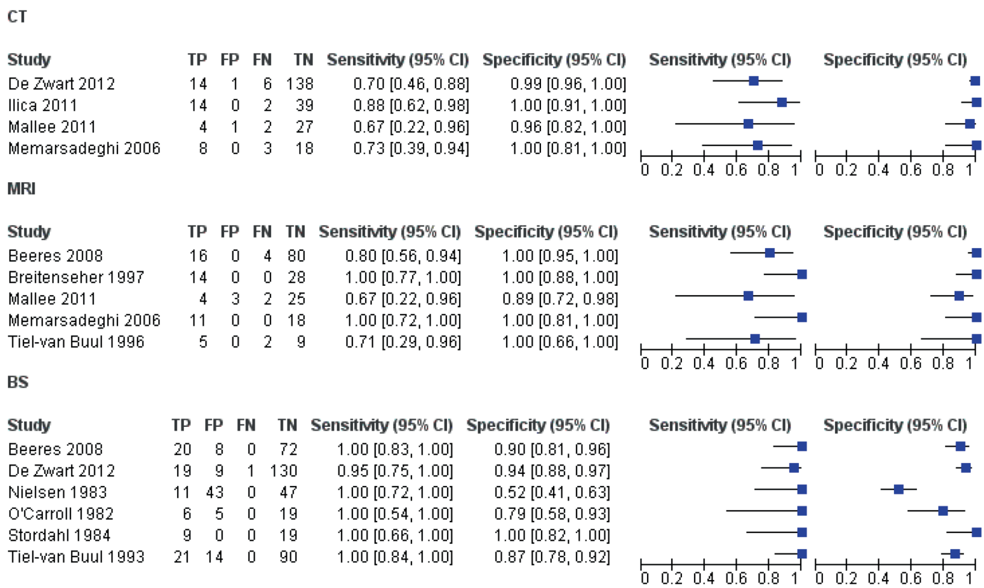
FINDINGS

Indirect comparisons

The forest plots of the diagnostic performance characteristics of CT, MRI and BS are presented in Figure 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 0.22 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 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. 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' ($Chi^2 = 50.3$, $df = 1$, P value < 0.01) and 'MRI versus BS' ($Chi^2 = 29.7$, $df = 1$, 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 ($Chi^2 = 1.9$, $df = 1$, P value $= 0.17$). These results



CT

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
De Zwart 2012	14	1	6	138	0.70 [0.46, 0.88]	0.99 [0.96, 1.00]
Ilica 2011	14	0	2	39	0.88 [0.62, 0.98]	1.00 [0.91, 1.00]
Mallee 2011	4	1	2	27	0.67 [0.22, 0.96]	0.96 [0.82, 1.00]
Memarsadeghi 2006	8	0	3	18	0.73 [0.39, 0.94]	1.00 [0.81, 1.00]

MRI

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Beeres 2008	16	0	4	80	0.80 [0.56, 0.94]	1.00 [0.95, 1.00]
Breitenseher 1997	14	0	0	28	1.00 [0.77, 1.00]	1.00 [0.88, 1.00]
Mallee 2011	4	3	2	25	0.67 [0.22, 0.96]	0.89 [0.72, 0.98]
Memarsadeghi 2006	11	0	0	18	1.00 [0.72, 1.00]	1.00 [0.81, 1.00]
Tiel-van Buul 1996	5	0	2	9	0.71 [0.29, 0.96]	1.00 [0.66, 1.00]

BS

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Beeres 2008	20	8	0	72	1.00 [0.83, 1.00]	0.90 [0.81, 0.96]
De Zwart 2012	19	9	1	130	0.95 [0.75, 1.00]	0.94 [0.88, 0.97]
Nielsen 1983	11	43	0	47	1.00 [0.72, 1.00]	0.52 [0.41, 0.63]
O'Carroll 1982	6	5	0	19	1.00 [0.54, 1.00]	0.79 [0.58, 0.93]
Stordahl 1984	9	0	0	19	1.00 [0.66, 1.00]	1.00 [0.82, 1.00]
Tiel-van Buul 1993	21	14	0	90	1.00 [0.84, 1.00]	0.87 [0.78, 0.92]

Figure 4. Forest plot of tests: 1 CT, 2 MRI, 3 BS

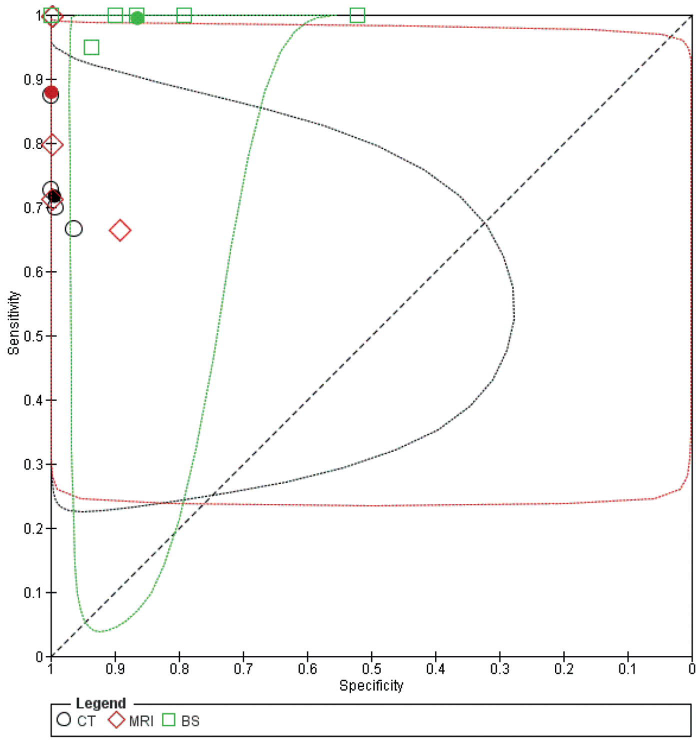


Figure 5. Study specific and pooled estimates of test performance for CT, MRI and BS with 95% confidence regions

Table 2. -2 Log Likelihood of models in each pairwise comparison

	-2 Log Likelihood		
	Model (a)	**Model (b)**	**Model (c)**
CT vs MRI	48.0	48.2	50.1
CT vs BS	63.8	65.0	115.3
MRI vs BS	72.3	72.6	102.3

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

may be explained by the findings from the summary estimates (see summary of findings Table): 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 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 and specificities for the two tests (Mallee 2011; Memarsadeghi 2006), with neither trial finding statistically significant differences between tests (reported P values > 0.05). The study directly comparing CT with BS (De Zwart 2012) 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 (Beeres 2008), 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).

DISCUSSION

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

Table 3. Direct comparisons (CT versus MRI; CT versus BS; MRI versus BS) for detection of scaphoid fractures

Study	Cases	Non-cases	CT Sensitivity (95% CI)	CT Specificity (95% CI)	MRI Sensitivity (95% CI)	MRI Specificity (95% CI)	Difference in sensitivity (95% CI) P value	Difference in specificity (95% CI) P value
Mallee 2011	6	28	0.67 (0.22 to 0.96)	0.96 (0.82 to 1.00)	0.67 (0.22 to 0.96)	0.89 (0.72 to 0.98)	0.00 (-0.53 to 0.53) P = 1.0	0.07 (-0.06 to 0.21) P = 0.611
Memarsadeghi 2006	11	11	0.73 (0.39 to 0.94)	1.00 (0.81 to 1.00)	1.00 (0.72 to 1.00)	1.00 (0.81 to 1.00)	-0.27 (-0.54 to -0.01) P = 0.214	0.00 (Standard error is zero; P undefined)
			CT		**BS**			
De Zwart 2012	20	139	0.70 (0.46 to 0.88)	0.99 (0.96 to 1.00)	0.95 (0.75 to 1.00)	0.94 (0.88 to 0.97)	-0.25 (-0.47 to -0.03) P = 0.092	0.05 (0.01 to 0.10) P = 0.019
			MRI		**BS**			
Beeres 2008	20	80	0.80 (0.56 to 0.94)	1.00 (0.95 to 1.00)	1.00 (0.83 to 1.00)	0.90 (0.81 to 0.96)	-0.20 (-0.38 to -0.03) P = 0.106	0.10 (0.03 to 0.17) P = 0.007

Sensitivity, specificity and their CIs are recalculated with Review Manager
P values were based on the 2-sided Fisher's exact test

9a

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 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 Table .

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.

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; Figure 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.

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 Table). 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.

AUTHORS' CONCLUSIONS

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.

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.

Summary of findings. Summary of findings Summary of findings: Diagnostic accuracy data

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	
Patient population	Patients with a clinically suspected scaphoid fracture but normal radiographs after trauma of the wrist
Prior testing	Clinical evaluation
Setting	Emergency departments
Index tests	Computed tomography (CT), magnetic resonance imaging (MRI) and bone scintigraphy (BS)
Reference standard	Most studies used radiographs obtained after 6 weeks. Otherwise, 1 index test or 2 index tests with the same diagnosis (fracture or no fracture) were used
Target condition	Scaphoid fractures
Importance	Early definitive diagnosis of a scaphoid fracture ensures adequate treatment, prevents unnecessary immobilisation and minimises the risk of long-term complications (e.g. non-union)
Included studies	4 studies for CT; 5 studies for MRI; 6 studies for BS 2 studies compared CT and MRI; 1 study directly compared CT with BS; and 1 compared MRI with BS
Number of suspected fractures (patients) studied	277 (276 patients) for CT; 221 (221 patients) for MRI; 543 (542 patients) for BS
Quality concerns	Overall quality of the included studies was moderate to good. Of most concern was the lack of availability of relevant clinical information during evaluation of the images as this does not mimic daily practice. Five studies did not clearly describe fracture criteria for a positive test
Limitations	No study compared all three tests (CT, MRI and BS) in the same population Only four comparison studies were included. Current comparisons are based on indirect evidence with possible variations in confounding factors like patient population and study characteristics Some studies were performed with only small cohorts The confidence intervals for summary estimates are wide for all three tests

Test	Number of studies	Number of suspected fractures	Summary sensitivity (95% CI)	Summary specificity (95%)	Summary LR+1 (95% CI)	Summary LR-2 (95% CI)	Consequences in a cohort of 10003: Missed fractures	Consequences in a cohort of 10003: Overtreated
CT	4	277	0.72 (0.36 to 0.92)	0.99 (0.71 to 1.00)	119.98 (1.49 to 9655.66)	0.28 (0.10 to 0.85)	56	8

9a

Summary of findings. (continued)

MRI	5	221	0.88 (0.64 to 0.97)	1.00 (0.38 to 1.00)	826.64 (0.51 to 1334596)	0.12 (0.03 to 0.42)	24	0
BS	6	543	0.99 (0.69 to 1.00)	0.86 (0.73 to 0.94)	7.35 (3.51 to 15.37)	0.01 (0.00 to 0.49)	2	112

Comparisons of the imaging tests

Comparison	**Findings**
CT, MRI and BS for diagnosis of clinically suspected scaphoid fractures	Overall diagnostic accuracy (DOR) of BS was significantly higher compared with CT (Chi2 = 50.3, df = 1, P < 0.01) and MRI (Chi2 = 29.7, df = 1, P < 0.01) CT and MRI were comparable in diagnostic accuracy (Chi2 = 1.9, df = 1, P = 0.17) The direct comparisons had similar patterns of differences in sensitivity and specificity as for the indirect comparisons. Given a median prevalence of 20%, 200 out of 1000 patients will have a scaphoid fracture. Of 200 cases, 56 will be missed if diagnosed using CT, 24 will be missed if diagnosed using MRI and 2 will be missed if diagnosed using BS. Of 800 patients without a scaphoid fracture, 8 will receive unnecessary treatment when CT is used for diagnosis, 0 when MRI is used for diagnosis and 112 if BS is used for diagnosis

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. Conversely, specificities of CT and MRI are both higher than for BS. CT and MRI have comparable diagnostic accuracy. Direct comparisons 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 the number of over-treated patients with BS is considerable, as well as the number of missed fractures on CT.

Quality of included studies was moderate to good, but there were only four direct comparison studies.

Well-designed studies directly comparing CT, MRI and BS could give valuable additional information.

1. LR+ Positive likelihood ratio

2. LR- Negative likelihood ratio

3. The median prevalence was 20%, calculated by using all studies. Missed fractures and over-treated patients were calculated using the median prevalence

REFERENCES

1. Hove LM. Epidemiology of scaphoid fractures in Bergen, Norway. Scand J Plast Reconstr Surg Hand Surg. 1999;33(4):423-6.
2. Hey HW, Chong AK, Murphy D. Prevalence of carpal fracture in Singapore. J Hand Surg Am. 2011;36(2):278-83.
3. van der Molen AB, Groothoff JW, Visser GJ, Robinson PH, Eisma WH. Time off work due to scaphoid fractures and other carpal injuries in The Netherlands in the period 1990 to 1993. Journal of hand surgery (Edinburgh, Scotland). 1999;24(2):193-8.
4. van Onselen EB, Karim RB, Hage JJ, Ritt MJ. Prevalence and distribution of hand fractures. Journal of hand surgery (Edinburgh, Scotland). 2003;28(5):491-5.
5. Gelberman R.H. - Gross MS. The vascularity of the wrist. Identification of arterial patterns at risk. Clinical orthopaedics and related research. 1986(0009-921X (Print)).
6. Rhemrev SJ, Ootes D, Beeres FJ, Meylaerts SA, Schipper IB. Current methods of diagnosis and treatment of scaphoid fractures. International journal of emergency medicine. 2011;4:4.
7. Merrell GA, Wolfe S.W. - Slade JF. Treatment of scaphoid nonunions: quantitative meta-analysis of the literature. Journal of Hand Surgery, American Volume. 2002(0363-5023 (Print)).
8. Dias JJ, Wildin CJ, Bhowal B, Thompson JR. Should acute scaphoid fractures be fixed? A randomized controlled trial. J Bone Joint Surg Am. 2005;87(10):2160-8.
9. Adey L, Souer JS, Lozano-Calderon S, Palmer W, Lee SG, Ring D. Computed tomography of suspected scaphoid fractures. J Hand Surg Am. 2007;32(1):61-6.
10. Lozano-Calderon S, Blazar P, Zurakowski D, Lee SG, Ring D. Diagnosis of scaphoid fracture displacement with radiography and computed tomography. J Bone Joint Surg Am. 2006;88(12):2695-703.
11. Pillai A, Jain M. Management of clinical fractures of the scaphoid: results of an audit and literature review. European journal of emergency medicine : official journal of the European Society for Emergency Medicine. 2005;12(2):47-51.
12. Rhemrev SJ, Beeres FJ, van Leerdam RH, Hogervorst M, Ring D. Clinical prediction rule for suspected scaphoid fractures: A prospective cohort study. Injury. 2010;41(10):1026-30.
13. Unay K, Gokcen B, Ozkan K, Poyanli O, Eceviz E. Examination tests predictive of bone injury in patients with clinically suspected occult scaphoid fracture. Injury. 2009;40(12):1265-8.
14. Yin ZG, Zhang JB-K, S.L. - Wang, X. G. Diagnosing suspected scaphoid fractures: a systematic review and meta-analysis. Clinical orthopaedics and related research. 2010;468(3)(1528-1132 (Electronic)):723-34.
15. Jenkins PJ, Slade K, Huntley JS, Robinson CM. A comparative analysis of the accuracy, diagnostic uncertainty and cost of imaging modalities in suspected scaphoid fractures. Injury. 2008;39(7):768-74.
16. Mallee W, Doornberg JN, Ring D, van Dijk CN, Maas M, Goslings JC. Comparison of CT and MRI for diagnosis of suspected scaphoid fractures. J Bone Joint Surg Am. 2011;93(1):20-8.
17. Rhemrev SJ, de Zwart AD, Kingma LM, Meylaerts SA, Arndt JW, Schipper IB, et al. Early computed tomography compared with bone scintigraphy in suspected scaphoid fractures. Clinical nuclear medicine. 2010;35(12):931-4.
18. 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(1):245-50.
19. Roolker W, Tiel-van Buul MM, Ritt MJ, Verbeeten B, Jr., Griffioen FM, Broekhuizen AH. Experimental evaluation of scaphoid X-series, carpal box radiographs, planar tomography, computed tomography, and magnetic resonance imaging in the diagnosis of scaphoid fracture. J Trauma. 1997;42(2):247-53.
20. Senall JA, Failla JM, Bouffard JA, van Holsbeeck M. Ultrasound for the early diagnosis of clinically suspected scaphoid fracture. J Hand Surg Am. 2004;29(3):400-5.
21. Tiel-van Buul MM, van Beek EJ, Broekhuizen AH, Bakker AJ, Bos KE, van Royen EA. Radiography and scintigraphy of suspected scaphoid fracture. A long-term study in 160 patients. J Bone Joint Surg Br. 1993;75(1):61-5.

22. Brookes-Fazakerley SD, Kumar AJ-O, J. Survey of the initial management and imaging protocols for occult scaphoid fractures in UK hospitals. Skeletal radiology. 2009;38(11):1045-8.
23. Groves AM, Kayani I, Syed R, Hutton BF, Bearcroft PP, Dixon AK, et al. An international survey of hospital practice in the imaging of acute scaphoid trauma. AJR American journal of roentgenology. 2006;187(6):1453-6.
24. Geissler WB, Adams JE, Bindra RR, Lanzinger WD, Slutsky DJ. Scaphoid fractures: what's hot, what's not. Instr Course Lect. 2012;61:71-84.
25. Rettig ME, Raskin KB. Retrograde compression screw fixation of acute proximal pole scaphoid fractures. J Hand Surg Am. 1999;24(6):1206-10.
26. Bond CD, Shin AY, McBride MT, Dao KD. Percutaneous screw fixation or cast immobilization for nondisplaced scaphoid fractures. J Bone Joint Surg Am. 2001;83-A(4):483-8.
27. Vinnars B, Pietreanu M, Bodestedt A, Ekenstam F, Gerdin B. Nonoperative compared with operative treatment of acute scaphoid fractures. A randomized clinical trial. J Bone Joint Surg Am. 2008;90(6):1176-85.
28. Fowler JR, Ilyas AM. Headless compression screw fixation of scaphoid fractures. Hand Clin. 2010;26(3):351-61, vi.
29. Buijze GA, Doornberg JN, Ham JS, Ring D, Bhandari M, Poolman RW. Surgical compared with conservative treatment for acute nondisplaced or minimally displaced scaphoid fractures: a systematic review and meta-analysis of randomized controlled trials. J Bone Joint Surg Am. 2010;92(6):1534-44.
30. Sanders WE. Evaluation of the humpback scaphoid by computed tomography in the longitudinal axial plane of the scaphoid. J Hand Surg [Am]. 1988;13(2):182-7.
31. Ty JM, Lozano-Calderon S, Ring D. Computed tomography for triage of suspected scaphoid fractures. Hand (N Y). 2008;3(2):155-8.
32. Thavarajah D, Syed T, Shah Y, Wetherill M. Does scaphoid bone bruising lead to occult fracture? A prospective study of 50 patients. Injury. 2011;42(11):1303-6.
33. Biswas D, Bible JE, Bohan M, Simpson AK, Whang PG, Grauer JN. Radiation exposure from musculoskeletal computerized tomographic scans. J Bone Joint Surg Am. 2009;91(8):1882-9.
34. McDougall IR, Rieser RP. Scintigraphic techniques in musculoskeletal trauma. Radiologic clinics of North America. 1989;27(5):1003-11.
35. Ring D, Lozano-Calderon S. Imaging for suspected scaphoid fracture. J Hand Surg Am. 2008;33(6):954-7.
36. 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(1):169-76.
37. Dorsay TA, Major NM, Helms CA. Cost-effectiveness of immediate MR imaging versus traditional follow-up for revealing radiographically occult scaphoid fractures. AJR American journal of roentgenology. 2001;177(6):1257-63.
38. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC medical research methodology. 2003;3:25.
39. Nielsen PT, Hedeboe J, Thommesen P. Bone scintigraphy in the evaluation of fracture of the carpal scaphoid bone. Acta orthopaedica Scandinavica. 1983;54(2):303-6.
40. Beeres FJ, Rhemrev SJ, den Hollander P, Kingma LM, Meylaerts SA, le Cessie S, et al. Early magnetic resonance imaging compared with bone scintigraphy in suspected scaphoid fractures. J Bone Joint Surg Br. 2008;90(9):1205-9.
41. de Zwart A - Rhemrev SJ-K, Lucas M. - Meylaerts, Sven A. G. - Arndt, Jan-Willem - Schipper, Inger B. - Beeres, Frank J. Early CT compared with bone scintigraphy in suspected schapoid fractures. Clinical nuclear medicine. 2012;Oct;37(10)(1536-0229 (Electronic)).
42. Ilica AT, Ozyurek S, Kose O, Durusu M. Diagnostic accuracy of multidetector computed tomography for patients with suspected scaphoid fractures and negative radiographic examinations. Jpn J Radiol. 2011;29(2):98-103.

43. O'Carroll PF, Doyle J, Duffy G. Radiography and scintigraphy in the diagnosis of carpal scaphoid fractures. Ir J Med Sci. 1982;151(7):211-3.
44. 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(8):971-5.
45. Stordahl A, Schjoth A, Woxholt G, Fjermeros H. Bone scanning of fractures of the scaphoid. Journal of hand surgery (Edinburgh, Scotland). 1984;9(2):189-90.
46. Lepage D, Obert L, Garbuio P, Tropet Y, Paratte B, Runge M, et al. [Contribution of quantitative radio-scintigraphy to diagnosis of wrist injuries undetected on plain films: a prospective study of 154 cases]. Revue de chirurgie orthopedique et reparatrice de l'appareil moteur. 2004;90(6):542-9.

CHAPTER 9b

DIAGNOSTIC WORK-UP FOR SUSPECTED SCAPHOID FRACTURES

Mallee WH
de Vos MJ

Book chapter in: Scaphoid Fractures: Evidence-Based Management. 2018

KEY POINTS

Fear of missing scaphoid fractures results in unnecessary overtreatment

Magnetic Resonance Imaging or Computed Tomography are able to establish a definitive diagnosis

Magnetic Resonance Imaging and Computed Tomography are more expensive diagnostic tools than radiography but with the right timing, it could lead to lower costs

Societal costs benefit from immediate MRI or CT

> A 27 year-old male visited the emergency department with complaints of his right wrist after a fall on the outstretched hand during a tennis match. He has tenderness of the anatomic snuffbox and pain around the scaphoid with ulnar deviation. Radiographs in 4 views show no fracture (figure 1). He is temporarily immobilized with a cast for 14 days and an appointment for repeat radiographs is made.

Panel 1. Case Scenario

INTRODUCTION

Delayed diagnosis of a true scaphoid fracture in clinically suspected scaphoid fractures increases the risk of persistent wrist problems and non-union(2, 3). Its blood supply is fragile and can be interrupted when fractured. Early treatment results in satisfactory results(4).

Trauma mechanism and physical examination in the emergency department (ED) can point towards a suspected scaphoid fracture. If radiographs of the wrist and scaphoid appear normal, 20% of the patients will still have a scaphoid fracture(5). With the risks of delayed treatment and occult fracture rate in mind, defensive management is usually started(6). Usually, temporary cast immobilization is prescribed until further examination or imaging is performed. The fear of undertreatment therefore results in overtreating 80% of patients with normal radiographs.

Numerous studies have been done to determine the diagnostic accuracy of imaging modalities such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and Bone Scintigraphy (BS) in detecting occult scaphoid fractures(7-13). The variety in techniques consequently results in a variation in diagnostic work-up as showed in national and international surveys(14-16). This is due to the availability of the scanning tools per hospital, differences in costs per technique and the lack of a consensus reference standard.

At follow-up after 1 to 2 weeks, repeat radiographs can detect the occult fracture, however the diagnostic accuracy is questionable. Additional imaging is often needed for definitive diagnosis.

Cost effectiveness

Advanced imaging techniques used to be costly diagnostic tools, however in present days these prices have dropped significantly in most countries. Most imaging techniques, especially of CT, are becoming more available in the Emergency Departments, which provides the physician with better options for immediate diagnosis and thereby adequate treatment. In suspected scaphoid fractures, unnecessary cast treatment is one of the main issues due to its effect on loss of productivity of the patient. By improving the diagnostic work-up, both healthcare and societal costs could drop significantly.

MAIN QUESTIONS

What is the best and most cost-effective diagnostic work-up for clinically suspected scaphoid fractures with normal radiographs?

CURRENT OPINION

There is a wide variety in diagnostic management for suspected scaphoid fractures. Repeating radiographs is often performed after 1 to 2 weeks of cast immobilization.

Figure 1. Initial radiographs in four views

FINDING THE EVIDENCE

The literature search was done in the following libraries:

- Cochrane Database of Systematic Reviews
- Cochrane Register of Diagnostic Test Accuracy Studies
- NHS Economic Evaluation Database
- Pubmed

Diagnostic accuracy studies:

- Search extracted studies till 2012
- The Cochrane libraries were searched using the term: Scaphoid
- The NHS database was searched using the term: Scaphoid
- Pubmed (Medline) was searched using: “Scaphoid Bone” [Mesh] OR scaphoid fracture*[tiab] OR scaphoid bone Fracture*[tiab] OR scaphoid[tiab]) AND “diagnos* OR Computed Tomography OR Magnetic Resonance Imaging OR Bone Scintigraphy OR Diagnos*
- Only systematic reviews and prospective cohort studies were included. This chapter is based on a recent Cochrane Review for Diagnostic Test Accuracy(5)

Cost-effectiveness studies:

- Search extracted studies from 2000-2016
- Additional search terms used: “Cost-effectiveness” OR “costs”

QUALITY OF EVIDENCE

Diagnostic accuracy studies

1a. Systematic Review of Prospective cohort studies / direct comparison studies
1b. Prospective cohort studies / direct comparison studies
2a. Systematic Review Cohort Studies
2b. Cohort Study/Low Quality RCT

Economic and decision analysis

1. Reasonable costs and alternatives used in study with values obtained from many studies, study used multi-way sensitivity analysis
2. Reasonable costs and alternatives used in study with values obtained from limited studies, study used multi-way sensitivity analysis
3. Analysis based on a limited section of alternatives and costs, or poor estimates of costs
4. No sensitivity analysis performed
5. Expert opinion

Quality of Evidence for Diagnostic accuracy studies

Level 1A:

- Systematic Review and Meta-Analyses: 1(5)

Quality of Evidence for Cost-effectiveness studies

- Level 1B (17-19)
- Level 2B (20)
- Economic and decision analysis level II (21)
- Economic and decision analysis level IV (22-24)

FINDINGS

Diagnostic accuracy

11 prospective studies were selected(7-13, 25-28). They reported diagnostic accuracy data for CT (4 studies; 277)), MRI (5 studies; 221 patients) and BS (6 studies; 543 patients). The papers were both comparison and non-comparison studies. Figure 2 presents the diagnostic performance characteristics of the imaging modalities. For overall diagnostic accuracy, CT showed a summary sensitivity and specificity of 0,72 (0,36 – 0,92) and 0,99 (0,71 – 1.00) respectively. For MRI, summary sensitivity and specificity was 0,88 (0,64 – 0,97) and 1,00 (0,38 – 1,00) respectively. For BS, this was 0,99 (0,69 – 1,00) and 0,86 (0,73 – 0,94). BS showed a significant higher diagnostic accuracy compared with CT and MRI (P<0,01), this is based on a higher sensitivity rate. CT and MRI showed comparable diagnostic accuracies (P=0,17).

Only one study directly compared CT and BS and did not report a significant higher sensitivity, however, CT showed a significant higher specificity over BS(11). One study directly compared MRI and BS and did not report a significant difference in correct predictions(8).

The median prevalence of missed fractures was 20%. Given this prevalence, it is interesting to evaluate the effect of these diagnostic accuracies in a large cohort of 1000 patients (200 occult scaphoid fractures). These data are presented in table 1.

Cost-effectiveness

The literature review found eight eligible studies evaluating the costs applicable to incorporating advanced imaging techniques in suspected scaphoid fractures(17-24). Most studies compared

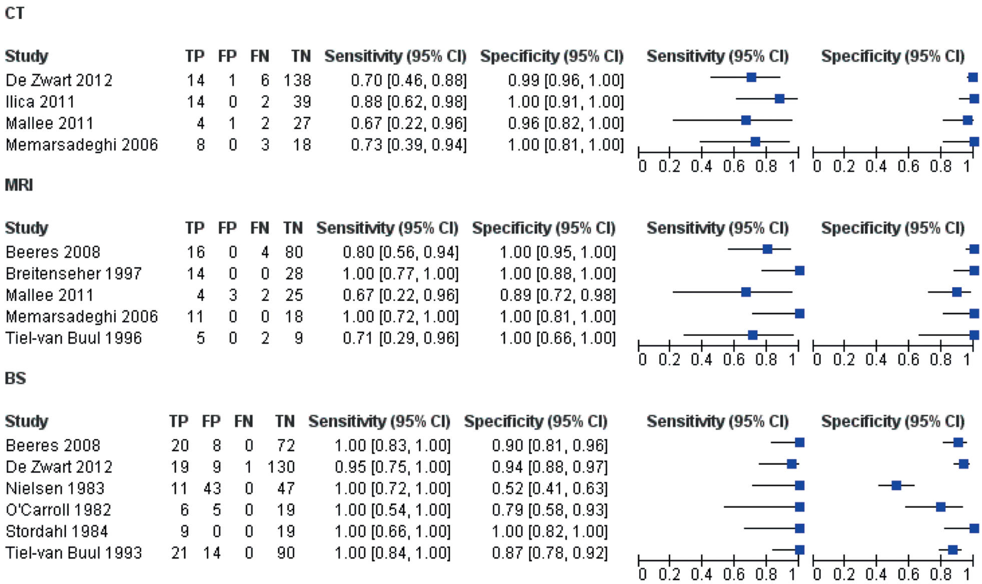


Figure 2. Forrest plots for diagnostic accuracies of computed tomography (CT), magnetic resonance imaging (MRI) and bone scintigraphy (BS). *TP* True Positives; *FP* False Positives; *TN* True Negatives; *FN* False Negatives. (From Mallee WH, Wang J, Poolman RW, et al. Computed tomography versus magnetic resonance imaging versus bone scintigraphy for clinically suspected scaphoid fractures in patients with negative plain radiographs. Cochrane Database 2015, with permission.)

Table 1. Effect of diagnostic accuracy on a cohort of 1000 patients with 200 scaphoid fractures

	Number of missed fractures	Number of overtreated patients
CT	56	8
MRI	24	0
BS	2	112

MRI with conventional protocols that mainly consist of two weeks of cast immobilization and repeat radiography. Since data was scarce and methodology too diverse, pooling of data could not be performed.

Using MRI in the diagnostic work-up was the subject of all included studies. CT was subject in two studies(21, 22). Most studies looked at immediate scanning or within 3 days after injury compared with follow-up radiography and cast immobilization for 10-14 days. Immediate MRI or CT was the most cost-effective approach in three studies(18, 21, 23). Costs were comparable in four studies (19, 20, 22, 24) and only one study reported slightly higher costs for MRI, however this was due to known high costs per MRI in Australia(17).

All studies stress the effect of early diagnosis by immediate MRI and CT on societal costs. It is also emphasized that CT and MRI have the ability to detect other injuries such as carpal or distal radius fractures and ligamentous injury. Benefit is thereby expected in treatment of other occult injuries as well.

RECOMMENDATIONS

- In patients with a clinically suspected scaphoid fracture and normal initial radiographs:
- Comparable results are presented for MRI, CT and BS. Direct comparisons showed no differences, pooled estimates showed a higher sensitivity for BS over CT and MRI [Overall quality: high]
- In large cohorts, BS still results in a considerable amount of overtreatment [Overall quality: high]
- Immediate CT or MRI is comparable or beneficial over 10-14 days of cast immobilization and follow-up in terms of direct hospital costs [Overall quality: moderate]
- Immediate CT or MRI is of great benefit for societal costs [Overall quality: moderate]
- Incorporating CT or MRI in the diagnostic work-up results in detection of other occult injuries besides scaphoid fractures. [Overall quality: low]

CONCLUSIONS

In patients with a clinically suspected scaphoid fracture and normal radiographs, the best and most cost-effective diagnostic work-up is to perform immediate MRI or CT. This management will not increase direct hospital costs, can decrease the societal costs and assures early diagnosis and adequate treatment.

Figure 3. CT and MRI scans show an occult scaphoid waist fracture

In this case, we prefer to perform MRI in the acute setting when a patient presents with a clinically suspected scaphoid fracture and normal radiographs. A comparable alternative is CT (Fig 7.3). A scaphoid fracture was detected on both scans. This will establish an early and definitive diagnosis and thereby adequate treatment without the risk for unnecessary immobilization and loss in productivity. If scans are negative, pressure bandage without further follow-up is sufficient.

Panel 2. Author's Preferred Technique

Pearls:

- With CT scanning in the acute setting, other fractures in the carpus and/or distal radius can be frequently found. CT scanning in the longitudinal axis of the scaphoid might result in an improved visualization of the scaphoid to distinguish vascular channels from minor fractures(1)
- With MRI in the acute setting, other fractures and ligamentous injuries in the wrist can be frequently found.
- Immediate scanning protocols have been shown to be cost-effective

Pitfalls:

- Performing Bone Scintigraphy is both a time consuming and an invasive diagnostic tool and therefore not recommended.

Panel 3. Pearls and Pitfalls

REFERENCES

1. Mallee WH, Doornberg JN, Ring D, Maas M, Muhl M, van Dijk CN, et al. Computed tomography for suspected scaphoid fractures: comparison of reformations in the plane of the wrist versus the long axis of the scaphoid. Hand (N Y). 2014;9(1):117-21.
2. Merrell GA, Wolfe S.W. - Slade JF. Treatment of scaphoid nonunions: quantitative meta-analysis of the literature. Journal of Hand Surgery, American Volume. 2002(0363-5023 (Print)).
3. Gelberman R.H. - Gross MS. The vascularity of the wrist. Identification of arterial patterns at risk. Clinical orthopaedics and related research. 1986(0009-921X (Print)).
4. Dias JJ, Wildin CJ, Bhowal B, Thompson JR. Should acute scaphoid fractures be fixed? A randomized controlled trial. J Bone Joint Surg Am. 2005;87(10):2160-8.
5. Mallee WH, Wang J, Poolman RW, Kloen P, Maas M, de Vet HC, et al. Computed tomography versus magnetic resonance imaging versus bone scintigraphy for clinically suspected scaphoid fractures in patients with negative plain radiographs. Cochrane Database Syst Rev. 2015(6):CD010023.
6. Reigstad O, Grimsgaard C, Thorkildsen R, Reigstad A, Rokkum M. Scaphoid non-unions, where do they come from? The epidemiology and initial presentation of 270 scaphoid non-unions. Hand Surg. 2012;17(3):331-5.
7. Mallee W, Doornberg JN, Ring D, van Dijk CN, Maas M, Goslings JC. Comparison of CT and MRI for diagnosis of suspected scaphoid fractures. J Bone Joint Surg Am. 2011;93(1):20-8.
8. Beeres FJ, Rhemrev SJ, den Hollander P, Kingma LM, Meylaerts SA, le Cessie S, et al. Early magnetic resonance imaging compared with bone scintigraphy in suspected scaphoid fractures. J Bone Joint Surg Br. 2008;90(9):1205-9.
9. 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(1):245-50.
10. 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(1):169-76.
11. de Zwart A - Rhemrev SJ-K, Lucas M. - Meylaerts, Sven A. G. - Arndt, Jan-Willem - Schipper, Inger B. - Beeres, Frank J. Early CT compared with bone scintigraphy in suspected schapoid fractures. Clinical nuclear medicine. 2012;Oct;37(10)(1536-0229 (Electronic)).
12. Ilica AT, Ozyurek S, Kose O, Durusu M. Diagnostic accuracy of multidetector computed tomography for patients with suspected scaphoid fractures and negative radiographic examinations. Jpn J Radiol. 2011;29(2):98-103.
13. 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(8):971-5.
14. Mallee WH, Veltman ES, Doornberg JN, Blankevoort L, van Dijk CN, Goslings JC. [Variations in management of suspected scaphoid fractures]. Nederlands tijdschrift voor geneeskunde. 2012;156(28):A4514.
15. Groves AM, Kayani I, Syed R, Hutton BF, Bearcroft PP, Dixon AK, et al. An international survey of hospital practice in the imaging of acute scaphoid trauma. AJR American journal of roentgenology. 2006;187(6):1453-6.
16. Brookes-Fazakerley SD, Kumar AJ-O, J. Survey of the initial management and imaging protocols for occult scaphoid fractures in UK hospitals. Skeletal radiology. 2009;38(11):1045-8.
17. Brooks S, Cicuttini FM, Lim S, Taylor D, Stuckey SL, Wluka AE. Cost effectiveness of adding magnetic resonance imaging to the usual management of suspected scaphoid fractures. British journal of sports medicine. 2005;39(2):75-9.
18. Patel NK, Davies N, Mirza Z, Watson M. Cost and clinical effectiveness of MRI in occult scaphoid fractures: a randomised controlled trial. Emerg Med J. 2013;30(3):202-7.
19. Gooding A, Coates M Fau - Rothwell A, Rothwell A. Cost analysis of traditional follow-up protocol versus MRI for radiographically occult scaphoid fractures: a pilot study for the Accident Compensation Corporation. (1175-8716 (Electronic)).

20. Kelson T, Davidson R, Baker T. Early MRI versus conventional management in the detection of occult scaphoid fractures: what does it really cost? A rural pilot study. (2051-3909 (Electronic)).
21. Yin ZG, Zhang JB, Gong KT. Cost-Effectiveness of Diagnostic Strategies for Suspected Scaphoid Fractures. J Orthop Trauma. 2015;29(8):e245-52.
22. Karl JW, Swart E, Strauch RJ. Diagnosis of Occult Scaphoid Fractures: A Cost-Effectiveness Analysis. (1535-1386 (Electronic)).
23. Saxena P, McDonald R, Gull S, Hyder N. Diagnostic scanning for suspected scaphoid fractures: an economic evaluation based on cost-minimisation models. Injury. 2003;34(7):503-11.
24. Dorsay TA, Major NM, Helms CA. Cost-effectiveness of immediate MR imaging versus traditional follow-up for revealing radiographically occult scaphoid fractures. AJR American journal of roentgenology. 2001;177(6):1257-63.
25. Nielsen PT, Hedeboe J, Thommesen P. Bone scintigraphy in the evaluation of fracture of the carpal scaphoid bone. Acta orthopaedica Scandinavica. 1983;54(2):303-6.
26. O'Carroll PF, Doyle J, Duffy G. Radiography and scintigraphy in the diagnosis of carpal scaphoid fractures. Ir J Med Sci. 1982;151(7):211-3.
27. Stordahl A, Schjoth A, Woxholt G, Fjermeros H. Bone scanning of fractures of the scaphoid. Journal of hand surgery (Edinburgh, Scotland). 1984;9(2):189-90.
28. Tiel-van Buul MM, van Beek EJ, Broekhuizen AH, Bakker AJ, Bos KE, van Royen EA. Radiography and scintigraphy of suspected scaphoid fracture. A long-term study in 160 patients. J Bone Joint Surg Br. 1993;75(1):61-5.

PART 4

CONCLUSION & DISCUSSION: IMMEDIATE AND ACCURATE DIAGNOSIS IS POSSIBLE!

CHAPTER 10

SUMMARY & GENERAL DISCUSSION

The diagnostic work-up of patients with (suspected) scaphoid fractures is flawed by "under-diagnosis" and "over-treatment".

We are unable to identify all scaphoid fractures during patients' initial visit in Emergency Department, if we continue to rely on painful anatomic snuffbox on clinical examination and plain radiographs as the mainstay of diagnostic imaging. To assure that we do not miss –"under-diagnose"- any scaphoid fractures and do not unnecessarily "over-"treat patients, we have to improve our diagnostic protocol from both clinical- as well as imaging perspectives.

The overall aim of this PhD Thesis was to assess current management and to improve both clinical and diagnostic strategies, in order to create a new and more efficient protocol that leads to earlier and more accurate diagnosis, reduces overtreatment, follow-up imaging and outpatient clinic visits without an increased risk of missing a fracture.

PART 1

Current Hospital Management

Chapter 2 focused on current management in Dutch Hospitals from initial diagnosis to treatment. The aim was to evaluate if there were differences between hospitals and if management was comparable with current literature. With a 90% response rate on our survey, we presented an accurate representation of national differences. Most hospitals have a protocol in place for scaphoid fracture management (79%). In clinically suspected scaphoid fractures and normal radiographs, patients were treated with a lower arm cast including the thumb and reviewed in outpatient clinics 7-14 days later in more than 90% of hospitals. Usually radiographs were repeated prior to obtaining advanced imaging. In 35 of the 90 hospitals Computed Tomography was used to establish a definitive diagnosis when radiographs could not provide, 12 used Bone Scintigraphy and only 2 used Magnetic Resonance Imaging; all other hospitals based further evaluation upon availability of advanced imaging, follow-up radiographs or clinical evaluation. The study showed that there was a large variety between hospitals and that they do not incorporate recommendations from recent literature.

PART 2

Clinical Evaluation of Scaphoid Fractures

The second part focused on improving the selection of patients that are clinically suspected for a scaphoid fracture, without increasing the risk of missing a fracture.

Chapter 3 presents a systematic review and meta-analysis that identified predictors of scaphoid fractures that are used in clinical evaluation in the acute setting. Thirteen included studies described 25 different clinical tests. Nearly all studies were prospective cohorts. The most studied and valuable tests were a painful anatomic snuffbox (ASB) and pain with longitudinal thumb compression (LTC). For ASB, sensitivity ranged from 0.87 to 1.00 and specificity 0.03 to 0.98. For LTC, sensitivity ranged from 0.48 to 1.00 and specificity 0.22 to 0.97. There were clear differences in sensitivity and specificity for both ASB and LTC. Other potentially valuable tests were scaphoid tubercle tenderness (STT) and painful ulnar deviation (PUD). Combining several tests seems to improve diagnostic accuracy, however data on this specific matter was scarce.

The overall methodological quality of the included studies was low to moderate highlighting the lack of quality evidence. One of the recurrent issues in diagnostic test accuracy studies for scaphoid fractures is the absence of a consensus reference standard. Some older studies used 2-week follow-up radiographs, which are known to be inaccurate. The more recent papers included a variety of reference standards such as 6-week follow-up radiographs and CT. The data of this review was used to identify predictors that could be used to develop a clinical decision rule; these predictors were: sex, trauma mechanism, presence of swelling in the ASB, tenderness in the ASB, painful LTC, Scaphoid Tubercle Tenderness and Painful Ulnar Deviation.

In **Chapter 4** these predictors were studied in a large, prospective trial. The study was part of a multicenter project that studied both distal radius and scaphoid fractures: The Amsterdam Wrist Trial. The aim of the study was to develop a clinical decision rule that could aid physicians in the acute setting when diagnosing patients with a clinically suspected scaphoid fracture. The ultimate goal was to prove that a clinical prediction rule could result in a reduction of follow-up CT/MRI/Bone Scintigraphy and lower the number of patients that are initially treated with a cast even though radiographs appear normal. One of the risks that had to be minimized simultaneously was missing a scaphoid fracture. All patients that were seen with wrist injury in the ED were evaluated prior to radiographs using a Case Report Form that included the predictors identified in **Chapter 3**. In a consecutive series of 893 patients, 68 scaphoid fractures were diagnosed on radiographs of which 21% was occult and visualized on follow-up imaging such as repeat radiographs or CT. The final clinical prediction rule combines sex, swelling of the ASB, tenderness in the ASB, PUD and painful LTC and advised to obtain wrist and scaphoid radiographs when the probability of a fracture was at least 15%. This rule showed a 0.97 sensitivity and 0.20 specificity. When compared with current clinical practice, it can result in a reduction of overtreatment of 15%, including a 50% lower risk of missing a scaphoid fracture. The rule can be easily used in the ED with clear benefits.

PART 3

Imaging in Suspected Scaphoid Fractures

If a patient is clinically suspected for a scaphoid fracture and radiographs show no fracture, it is well known that many patients can suffer from an 'occult' fracture. This group of patients will receive cast treatment until definitive diagnosis is established. The most common applied imaging modalities are CT, MRI and BS.

In **Chapter 5**, two different reformations of CT scans were compared to identify the most accurate reformation regarding scaphoid fracture diagnosis in clinically suspected scaphoid fractures with normal radiographs. A prospective consecutive series of 34 patients was included. Standard wrist reformations (CT-wrist) and reformations in the longitudinal axial plane of the scaphoid (CT-scaphoid) were evaluated by a panel of 3 experts, 2 surgeons and one musculoskeletal radiologist, to come to a consensus diagnosis. The reference standard was

6-week follow-up radiographs. The latter identified 6 occult scaphoid fractures (18%). Sensitivity and specificity for both reformations were calculated; for CT-scaphoid this was 0.67 and 0.96 respectively, CT-wrist was lower with 0.33 and 0.89 respectively. Positive and negative predictive values for CT-scaphoid were 0.76 and 0.94, for CT-wrist this was 0.36 and 0.87. Differences in diagnostic accuracy were better for CT-scaphoid, but not significant.

In **Chapter 6**, the diagnostic accuracy of CT and MRI was compared in the same prospective consecutive series of 34 patients. Both CT and MRI were performed within 10 days after trauma and were ultimately compared with 6-week follow-up radiographs (reference standard) and were all evaluated by the same panel of experts. Both CT and MRI were comparable in diagnostic accuracy and showed notably higher specificity (0.89 and 0.96 respectively) than sensitivity (both 0.67). The positive predictive values for CT and MRI were 0.76 and 0.54 respectively; the negative predictive values were 0.94 and 0.93 respectively. Bayes theorem was applied for calculations of diagnostic accuracy. Both techniques were thereby better in excluding fractures then confirming fractures. A considerable amount of fractures other than the scaphoid were identified as well.

In **Chapter 7**, the value of 6-week follow-up radiographs as a reference standard was explored. In the first worldwide study by the Traumaplatform Study Group, an online survey was rolled out in which 81 Orthopedic and Trauma surgeons evaluated the follow-up radiographs of 34 consecutive patients. One group received images in JPEG format and one group evaluated the original DICOM files on a dedicated viewer (the latter mimics daily practice). All observers received both initial as well as 6-week follow-up radiographs. There was slight interobserver agreement for both JPEGs and DICOM images with a kappa of 0.14 and 0.15 respectively. MRI and CT were used as reference standards in calculating diagnostic accuracy. When DICOM images were evaluated, both sensitivity and negative predictive value was significantly higher compared to the JPEGs, still diagnostic accuracy was low. No differences were detected in specificity and positive predictive value. The low interobserver agreement and diagnostic accuracy of follow-up radiographs showed that this is an inadequate reference standard in clinically suspected scaphoid fractures

In diagnostic test accuracy studies without a reference standard, latent class analysis could be a solution. With latent class analysis, unobserved or latent classes in data can be identified and have shown to be helpful for evaluating diagnostic test accuracy in the absence of reference tests(1). In other words, this method provides a different way to look at this type of data if diagnostic accuracy is calculated. In diagnostic test accuracy studies of scaphoid fractures, several reference standards are used due to the lack of consensus (e.g. 6-week follow-up radiographs, repeated physical examination, MRI or combining several tests). In **Chapter 8**, two prospective cohorts were analyzed using latent class methodology, one cohort compared MRI and CT; one cohort compared MRI with BS and clinical tests. Analysis of the first cohort was done by the maximum likelihood-based method, the second cohort by using a Bayesian based latent class analysis model. In the first cohort, latent class analysis showed quite substantial increase in sensitivity and specificity for both CT (0.78 and 1.0 vs 0.67 and 0.89) and MRI (0.8 and 0.93 vs 0.67 and 0.96) respectively. For the second cohort, sensitivity of MRI increased

10

(0.84 vs 0.75) and specificity remained nearly equal (0.99 vs 1.0). For BS, sensitivity decreased (0.94 vs 1.0) and specificity remained equal (both 0.89).

The answer to the question what imaging technique should be used in a clinically suspected scaphoid fracture with normal radiographs including when to perform it in terms of timing, cost-effectiveness and patient benefits is given in **Chapter 9a** and **b**. A Cochrane Review for Diagnostic Test Accuracy of CT, MRI and BS and an overview of available cost-effectiveness studies was performed. The Cochrane Review included 11 moderate to good prospective studies and provided pooled estimates based on indirect comparisons. With a summary sensitivity and specificity of 0.72 and 0.99 for CT, 0.88 and 1.00 for MRI and 0.99 and 0.86 for BS, it is again showed that CT and MRI have better specificity than sensitivity, while the opposite accounts for BS. The data showed that BS was statistically favorable over CT and MRI; the latter were comparable in diagnostic accuracy. However, since BS is an invasive modality and can only be executed after a 3-day delay, it is advised to perform CT or MRI in a clinically suspected scaphoid fracture with normal initial radiographs.

Furthermore, it is showed in **Chapter 9b** that it is cost-effective in terms of direct hospital costs as well as beneficial from a patient perspective to perform CT or MRI immediately. This reduces overtreatment and follow-up visits to the outpatient clinic substantially without increasing costs. Another advantage of immediate CT or MRI is the detection of fractures other than scaphoid fractures resulting in more adequate treatment.

GENERAL DISCUSSION

From a patients' perspective, one could argue that it is hard to understand that their diagnostic work-up and choice of installed treatment in this type of wrist injury depends on the hospitals' favors and capabilities rather than science. Even though most hospitals have a protocolled work-up for scaphoid fractures that should be based on evidence from available literature, there is a wide variation in both the former- and thus the latter.

Physicians use predictors to select or triage patients that need further diagnostic work-up to reliably rule-in or rule-out possible disease. Well-known examples of an evidence-based set of predictors that selects a certain group of patients are the Ottawa Ankle Rules. Applying these tests during initial presentation in the ED results in an advice to obtain radiographs of the ankle-foot or if imaging is unnecessary since a fracture is not suspected. Several implementation studies proved its efficacy and efficiency in reducing the amount of radiographs and waiting time in the ED without increasing the risk of missing foot-ankle fractures. In other words, they are designed to be highly sensitive and raise the pre-test probability of the disease for the following step in diagnostic work-up. The same applies for scaphoid fractures. It is emphasized in this thesis that clinical evaluation cannot rely on a single predictor. With a sensitivity of 0.87 to 1.00 found in previous studies, this test will not identify all patients with a scaphoid fracture. The specificity of this test, though heterogeneous results were found, is insufficient as well. A surprising find in the development of the scaphoid clinical prediction rule was the lack of a painful ASB in 30% of all confirmed scaphoid fractures. Its sensitivity was considerably lower than reported in previous literature. The risk of missing scaphoid fractures in the first step of diagnostic management could be substantial if this was the key test for clinical suspicion of a fractured scaphoid.

Therefore the goal of the scaphoid clinical prediction rule was to minimize this risk. With 97% sensitivity, this was achieved within the studied cohort. The second goal was to reduce the number of patients that warrants further diagnostic imaging. Applying the rule can reduce this with 15% compared with current clinical practice, however the low specificity remains an issue. If standard radiographic imaging appears normal, advanced imaging will still play an important role.

Bone Scintigraphy showed a statistically significant better sensitivity for definitive diagnosis in a clinically suspected scaphoid fracture with normal radiographs. However, this is an invasive technique with the highest radiation exposure and can only be performed 3 days after injury; it is therefor not a preferred technique. Immediate CT or MRI is found to be the designated imaging strategy. These strategies have both proven to be highly accurate and cost-effective. A surprising high number of other occult fractures of the wrist region can be detected as well. Accurate diagnosis during the initial ED presentation ensures adequate treatment and avoids unnecessary treatment and follow-up. Impact on production loss will be limited substantially. However, not all hospitals have these techniques available in the acute setting.

Still, both CT and MRI present their own interpretation difficulties. Small unicortical lines on CT may present vascular channels rather than a fracture. It is argumented that CT scans must be made with reformations in the long axis of the scaphoid to create the most adequate visualization(2). CT is an important imaging tool in scaphoid injury, not only for detection of fractures. Identifying several aspects such as location, plane of the fracture and displacement are essential in predicting union and length of immobilization or surgical treatment options; in addition CT can be used to evaluate union progress throughout treatment since radiographs are inaccurate(3-7). One could argue that every scaphoid fracture requires a CT scan to assess these aspects and present the patient with more accurate prognosis and that CT is the ultimate tool throughout the scaphoid fracture imaging strategy. And with the improvements of CT with Dual Energy, the disadvantages compared with MRI in visualizing soft tissues could be overcome(8).

When MRI is obtained, the presence of bone marrow edema without a cortical fracture line presents a diagnostic debate: Are we looking at a fracture or only bone bruising? Our panel of experts discussed that an extensive focal zone of edema (as seen in stress fractures) is considered a true fracture and does not have to be accompanied with a clear trabecular or cortical fracture lines. This is a somewhat different opinion than other studies suggested(9). An advantage of using MRI is the ability to detect ligamentous injury besides bony injury(10, 11). With the same trauma mechanism as a scaphoid fracture (FOOSH), ligaments such as the scapholunate ligament can rupture. This type of wrist injury is challenging to detect with either radiographs or CT.

It is highlighted that there is a lack of a consensus reference standard for definitive diagnosis of suspected scaphoid fractures. The preferred standard in recent literature was 6-week follow-up radiographs, however with the results found with this thesis it is discouraged to use this ‘reference standard’ in following studies. Latent class analysis could be the solution for future diagnostic accuracy studies in scaphoid fracture research. It is not stated that these calculations are more accurate compared to the more regular analysis using a reference standard. One must keep in mind that the choice for a certain reference standard influences the outcomes of the tested modalities.

MRI and CT have shown to be highly accurate tools and should be used as the new gold standard instead of 6-week follow-up radiographs.

REFERENCES

1. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. Biometrics. 1980;36(1):167-71.
2. Mallee WH, Doornberg JN, Ring D, Maas M, Muhl M, van Dijk CN, et al. Computed tomography for suspected scaphoid fractures: comparison of reformations in the plane of the wrist versus the long axis of the scaphoid. Hand (N Y). 2014;9(1):117-21.
3. Grewal R, Frakash U, Osman S, McMurtry RY. A quantitative definition of scaphoid union: determining the inter-rater reliability of two techniques. Journal of orthopaedic surgery and research. 2013;8:28.
4. Grewal R, Suh N, Macdermid JC. Use of computed tomography to predict union and time to union in acute scaphoid fractures treated nonoperatively. J Hand Surg Am. 2013;38(5):872-7.
5. Grewal R, Lutz K, MacDermid JC, Suh N. Proximal Pole Scaphoid Fractures: A Computed Tomographic Assessment of Outcomes. J Hand Surg Am. 2016;41(1):54-8.
6. Gilley E, Puri SK, Hearns KA, Weiland AJ, Carlson MG. Importance of Computed Tomography in Determining Displacement in Scaphoid Fractures. J Wrist Surg. 2018;Feb;7(1)(2163-3916 (Print)):38-42.
7. Buijze GA, Wijffels MM, Guitton TG, Grewal R, van Dijk CN, Ring D. Interobserver reliability of computed tomography to diagnose scaphoid waist fracture union. J Hand Surg Am. 2012;37(2):250-4.
8. Wong WD, Shah S, Murray N, Walstra F, Khosa F, Nicolaou S. Advanced Musculoskeletal Applications of Dual-Energy Computed Tomography. Radiologic clinics of North America. 2018;Jul;56(4)(1557-8275 (Electronic)):587-600.
9. 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(1):169-76.
10. Meister DW, Hearns KA, Carlson MG. Dorsal Scaphoid Subluxation on Sagittal Magnetic Resonance Imaging as a Marker for Scapholunate Ligament Tear. J Hand Surg Am. 2017;Sep;42(9)(1531-6564 (Electronic)):717-21.
11. Hafezi-Nejad N, Carrino JA, Eng J, Blackmore C, Shores J, Lifchez SD, et al. Scapholunate Interosseous Ligament Tears: Diagnostic Performance of 1.5 T, 3 T MRI, and MR Arthrography-A Systematic Review and Meta-analysis. Acad Radiol. 2016;Sep;23(9)(1878-4046 (Electronic)):1091-103.

CHAPTER 11

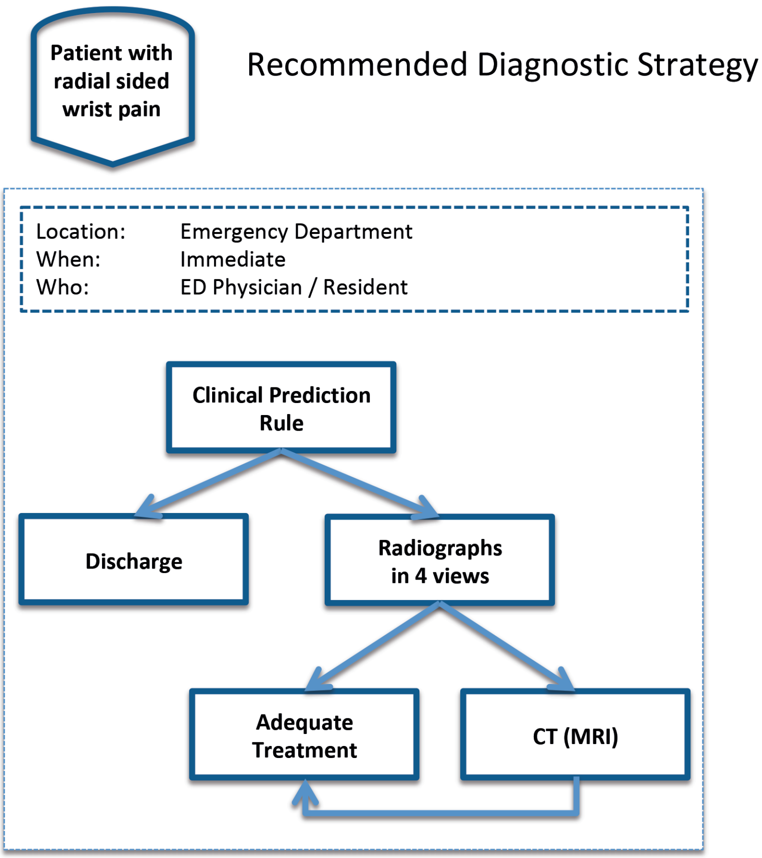
CONCLUSIONS & FUTURE RESEARCH

This thesis provides the answers to the previously set study questions concerning issues in the diagnostic strategies of suspected scaphoid fractures. In 2010 there was a large variety in the diagnostic work-up of scaphoid fractures, both nationally and internationally. A guideline was needed. Since then, a considerable amount of research has been performed that created the scientific foundation to propose a new, quicker and more efficient strategy that benefits both patient and healthcare system. We found an extensive amount of clinical tests that can be used to select the patients that are clinically suspected for a scaphoid fracture, however their diagnostic accuracies are limited. The five most accurate tests (sex, swelling of the anatomic snuffbox, tenderness in the anatomic snuffbox, painful ulnar deviation and painful axial thumb compression) were used to create a clinical prediction rule which showed that it is possible to achieve a 15% reduction of unnecessary immobilization and use of imaging while decreasing the risk of missing a scaphoid fracture with 50%. If a patient is clinically suspected for having a scaphoid fracture, CT and MRI are the best imaging tools to establish a definitive diagnosis. We prefer CT (with reformations along the longitudinal axis of the scaphoid) since this imaging modality can also be used in the follow-up to evaluate union progression. Obtaining these images in the acute setting is very beneficial for both patient and healthcare provider since it could reduce overtreatment to almost zero and it has shown to be a cost-effective approach. In addition, it detects other wrist/carpal fractures besides the scaphoid. We provided evidence that discourages the use of 6-week follow-up radiographs as a referenced standard; future studies should use MRI or CT instead.

Based on this thesis, a new diagnostic work-up for scaphoid fractures is proposed in Flowchart 2: Recommended Diagnostic Strategy

FUTURE RESEARCH SHOULD FOCUS ON:

- Performing external validation and implementation studies of the clinical prediction rule in different cohorts. If its ability to lower the number of clinically suspected scaphoid fractures still remains without increasing the number of missed fractures, it could and should be implemented in the ED.
- Evaluation of the quality of physical examination. Knowledge of both the anatomy of the wrist and how to perform certain tests is crucial in its diagnostic value. At this point, it is unknown if this is at a desired level in both ED physicians as well as residents in (Orthopedic) trauma.
- The role of initial radiographs in clinically suspected scaphoid fractures: if radiographs are normal, CT is warranted for definitive diagnosis; if radiographs show a scaphoid fracture, CT is warranted for determining fracture characteristics (e.g. displacement) and future follow-up. Can we abandon radiographs?
- Reducing the duration of cast treatment for occult fractures since severity of fracture is mostly limited. CT could be used to evaluate fracture healing.
- Evaluation of the implementation of a new national guideline with a similar survey.
- Artificial Intelligence algorithms based on machine learning. Can they replace the evaluation of radiographs and/or advanced imaging performed by the radiologist in the acute setting?



Flowchart 2. Clinical prediction rule: 5 tests → sex, swelling of the anatomic snuffbox, tenderness in the anatomic snuffbox, painful ulnar deviation and painful axial thumb compression. If the outcome of the model shows a fracture chance of 15% or higher, it is advised to obtain radiographs CT: with reformations in the longitudinal axis of the scaphoid. This is also used for Follow-up CT Discharge: e.g. pressure bandage. Adequate Treatment: depends on fracture type/location and patient characteristics

APPENDIX &

NEDERLANDSE SAMENVATTING

DANKWOORD / ACKNOWLEDGEMENTS

PORTFOLIO

CURRICULUM VITAE

NEDERLANDSE SAMENVATTING

Het diagnostische proces bij verdenking op een scafoïd fractuur moet verbeterd. We zijn nog niet in staat om alle scafoïd fracturen te detecteren tijdens een eerste bezoek aan de Spoedeisende Hulp (SEH) als we alleen een pijnlijke anatomische snuifdoos en röntgenfoto's als steunpilaren gebruiken. Geavanceerde beeldvormende technieken blijven een belangrijke rol spelen. Om er zeker van te zijn dat we geen scafoïd fracturen missen en tegelijkertijd voorkomen dat het onnodige behandelingen starten moeten we het diagnostische proces verbeteren op zowel het vlak van klinische evaluatie als beeldvorming.

Dit proefschrift had als doel om het diagnostische proces van scafoïd fracturen grondig te evalueren en de klinische evaluatie en beeldvorming te verbeteren zodat een efficiënter protocol ontstaat welke leidt tot snellere en accuratere diagnostiek, overbehandeling verminderd en verdere diagnostiek en polikliniek bezoeken onnodig maakt, zonder dat er meer fracturen worden gemist.

DEEL 1

Het Huidige Ziekenhuis Beleid

Hoofdstuk 2 focust zich op het huidige beleid binnen de Nederlandse ziekenhuizen, vanaf SEH presentatie tot uiteindelijke behandeling. Het doel was om de verschillen tussen de ziekenhuizen te evalueren en dit te vergelijken met de literatuur. Met een response-rate van 90% op de enquête, kon een realistische weergave van de werkelijkheid worden gegeven. Hieruit werd duidelijk dat veel ziekenhuizen een protocol hebben opgesteld voor diagnostisch management van scafoïd fracturen (79%). Bij een klinische verdenking op een scafoïd fractuur en normale röntgenfoto's werden in 90% van de ziekenhuizen patiënten behandeld met een onderarmsgips of –spalk met immobilisatie van de duim en werden ze opnieuw gezien op de polikliniek na 7-14 dagen. Van de 90 ziekenhuizen gebruikten 35 Computed Tomography (CT) om een definitieve diagnose te stellen, 12 ziekenhuizen gebruikten Bot Scintigrafie (BS) en 2 Magnetic Resonance Imaging (MRI). Alle andere ziekenhuizen gebruikte alleen klinische evaluatie en herhaalden röntgenfoto's. De studie liet zien dat er een enorm verschil was in management tussen de Nederlandse ziekenhuizen en dat de laatste aanbevelingen uit de literatuur niet in de protocollen werden geïncorporeerd.

DEEL 2

Klinische Evaluatie van Scafoïd Fracturen

Het tweede onderdeel concentreert zich op het verbeteren van de selectie van patiënten die klinisch verdacht worden van een scafoïd fractuur zonder hierbij het risico op het missen van een fractuur te vergroten.

Hoofdstuk 3 betreft een systematische review en meta-analyse waarbij we alle mogelijke klinische testen identificeren die gebruikt worden in de acute klinische evaluatie van een scafoïd fractuur. Dertien studies beschreven 25 verschillende testen. Bijna alle studies waren prospectieve cohorten. De meest bestudeerde en accurate testen waren een pijnlijke anatomische snuifdoos

(ASB) en pijn bij longitudinale compressie van de duim (LTC). Voor ASB was de sensitiviteit tussen de 0.87 en 1.00, de specificiteit was 0.03 tot 0.98. Voor LTC was de sensitiviteit 0.48 tot 1.00 en de specificiteit 0.22 tot 0.97. Er zijn duidelijke verschillen in sensitiviteit en specificiteit voor beide testen. Andere potentieel waardevolle testen waren een pijnlijk tuberculum van het scafoïd (STT) en pijn bij ulnair deviatie (PUD). Het combineren van verschillende testen lijkt de diagnostische accuratesse te verbeteren, helaas was er op dit specifieke gebied onvoldoende data beschikbaar.

De methodologische kwaliteit van de geïncludeerde studies was laag tot matig; dit onderstreept het gebrek aan kwalitatieve wetenschappelijke onderbouwing. Een van de terugkerende knelpunten in diagnostische accuratesse studies van scafoïd fracturen is het gebrek aan een referentie standaard. Enkele oudere studies gebruikten röntgenfoto's na 2 weken, welke bekend staan om hun gebrek aan accuratesse. Meer recente studies gebruikten een diversiteit aan referentie standaarden waaronder het herhalen van röntgenfoto's na 6 weken en CT. De data van deze systematische review is gebruikt om voorspellende testen van scafoïd fracturen te vinden die gebruikt kunnen worden om een klinische beslisregel te ontwikkelen; deze testen waren: geslacht, trauma mechanisme, zwelling in de anatomische snuifdoos, drukpijn van de anatomische snuifdoos, pijnlijke longitudinale compressie van de duim, pijnlijk tuberculum scafoïdeum en pijnlijke ulnair deviatie.

In **Hoofdstuk 4** werden deze voorspellers onderzocht in een grote prospectieve trial. De studie was onderdeel van een groot multicenter project waarin gekeken werd naar distale radius en scafoïd fracturen: The Amsterdam Wrist Rules. Het doel van de studie was het ontwikkelen van een klinische beslisregel die in de acute setting gebruikt worden voor detectie van scafoïd fracturen. Deze beslisregel moest resulteren in een reductie van overdiagnostiek (CT/MRI/Bot scintigrafie) en –behandeling en tegelijkertijd de kans op het missen van een scafoïd fractuur zo laag mogelijk te maken. Alle patiënten die op de spoedeisende hulp gezien werden na polstrauma, werden geëvalueerd aan de hand van een Case Report Form waarin onder andere bovenstaande voorspellers opgenomen waren, hierna werd röntgendiagnostiek verkregen.

In een opeenvolgende serie van 893 patiënten werden 68 scafoïd fracturen gediagnosticeerd, hiervan was 21% van de fracturen niet zichtbaar op de initiële röntgenfoto's maar pas bij follow-up beeldvorming (w.o. röntgenfoto's, CT). De uiteindelijke klinische beslisregel combineert geslacht, zwelling van de anatomische snuifdoos, drukpijn van de anatomische snuifdoos, pijnlijke ulnair deviatie en pijnlijke longitudinale compressie van de duim. Bij een fractuur kans van 15% of hoger is er sprake van een klinische verdenking op een scafoïd fractuur en wordt geadviseerd om pols en specifieke scafoïd röntgenfoto's te verkrijgen.

De beslisregel liet een sensitiviteit van 0.97 zien en een specificiteit van 0.20. Als we de uitkomsten van de regel vergelijken met de dagelijkse praktijk waarin de spoedeisende hulp arts zelf de beslissing mag maken, laat de beslisregel een 15% reductie zien van overbehandeling en overdiagnostiek en tegelijkertijd een 50% lager risico op het missen van een scafoïd fractuur. De regel kan gemakkelijk toe worden gepast op de spoedeisende hulp.

DEEL 3
Beeldvorming bij Verdenking op een Scafoïd Fractuur

Als een patiënt een klinische verdenking op een scafoïd fractuur heeft en de gemaakte röntgenfoto's in 4 richtingen laten geen fractuur zien, is het welbekend dat een aanzienlijk deel een 'occulte' fractuur kan hebben. Deze groep patiënten krijgt gipsimmobilisatie totdat er een definitieve diagnose is vastgesteld. Meestal betekent dit een tot twee weken gips en dan worden röntgenfoto's herhaald of gebruikt men CT, MRI of BS.

In **Hoofdstuk 5** worden 2 verschillende CT reconstructies vergeleken om te kijken welke manier van reconstrueren het meest accuraat is in het diagnosticeren van een scafoïd fractuur bij een klinische verdenking en normale röntgenfoto's. Een prospectieve serie van 34 patiënten werd geïncludeerd. Standaard pols reconstructies (CT-pols) en reconstructies in het longitudinale vlak van het scafoïd (CT-scafoïd) werden geëvalueerd door een panel bestaande uit 3 experts (1 Orthopedisch chirurg, 1 Traumachirurg, 1 Musculoskeletaal radioloog) om zo tot een consensus diagnose te komen. De referentie standaard was het herhalen van röntgenfoto's na 6 weken. Laatstgenoemde identificeerde 6 occulte scafoïd fracturen (18%). Sensitiviteit en specificiteit voor beide CT reconstructies werden berekend; voor CT-scafoïd was dit respectievelijk 0.67 en 0.96, voor CT-pols was dit lager respectievelijk 0.33 en 0.89. De positief en negatief voorspellende waardes voor CT-scafoïd waren 0.76 en 0.94, voor CT-pols was dit 0.36 en 0.87. De diagnostische accuratesse van CT-scafoïd is beter, maar niet statistisch significant met deze getallen.

In **Hoofdstuk 6** werden de diagnostische karakteristieken van CT en MRI vergeleken in dezelfde prospectieve groep van 34 patiënten. CT en MRI werden binnen 10 dagen na trauma vervaardigd en wederom vergelijken met de 6-weekse röntgenfoto's (referentie standaard). Evaluatie van de beelden werd gedaan door hetzelfde expertpanel. CT en MRI waren vergelijkbaar in diagnostische accuratesse en lieten beiden een aanzienlijk hogere specificiteit (respectievelijk 0.89 en 0.96) dan sensitiviteit (beiden 0.67) zien.

&

De positief voorspellende waardes voor CT en MRI waren respectievelijk 0.76 en 0.54, de negatief voorspellende waardes waren 0.94 en 0.93. Bayes' theorie werd toegepast voor het berekenen van de diagnostische accuratesse. Beide modaliteiten bleken beter in het excluderen van een fractuur dan in het confirmeren van een fractuur. Er werd tevens een aanzienlijke hoeveelheid fracturen anders dan het scafoïd geïdentificeerd.

In **Hoofdstuk 7** werd de diagnostische waarde van de 6-weekse röntgenfoto's als referentie standaard bestudeerd. Dit was de eerste wereldwijde studie van de Traumaplatform Studie Groep. Een online survey werd verstuurd waarbij 81 Trauma- en Orthopedische chirurgen de 6-weekse röntgenfoto's evalueerden van 34 patiënten met een klinische verdenking op een scafoïd fractuur en normale initiële röntgenfoto's. Een groep evalueerde de beelden in JPEG format, de andere groep evalueerde de beelden in DICOM format in een daarvoor bestemde viewer (dagelijkse praktijk). Alle observers beoordeelden zowel de initiële als de 6-weekse röntgenfoto's. Er was een geringe interobserver overeenstemming binnen zowel de JPEG als de DICOM groep met een kappa van respectievelijk 0.14 en 0.15. MRI en CT werden gebruikt als referentie standaarden om zo de diagnostische karakteristieken van de 6-weekse röntgenfoto's te berekenen. De DICOM beelden lieten een significant hogere sensitiviteit en negatief voorspellende waarde zien in vergelijking met de JPEG beelden, desondanks was de diagnostische accuratesse laag. Er waren geen significante verschillen in specificiteit en positief voorspellende waarde. De lage interobserver overeenstemming in combinatie met de lage diagnostische accuratesse laat duidelijk zien dat 6-weekse röntgenfoto's niet adequaat zijn om te gebruiken als referentie standaard bij klinische verdenking op een scafoïd fractuur met normale initiële röntgenfoto's.

Bij diagnostische accuratesse studies waarbij een referentie standaard ontbreekt is latent class analysis. Hiermee kunnen kansberekeningen worden gemaakt die gebaseerd zijn op meerdere diagnostische testen zonder de standaard formules toe te passen waar een referentie standaard voor nodig is. Met andere woorden, deze statistische methode is een andere manier om naar data te kijken als er diagnostische accuratesse van tests berekent moeten worden

Bij berekenen van diagnostische accuratesse van een test voor scafoïd fracturen wordt vaak gebruikt gemaakt van verschillende referentie standaarden door gebrek aan consensus (w.o. 6-weekse röntgenfoto's, herhalen van klinische evaluatie, MRI, combineren van verschillende tests. In **Hoofdstuk 8** worden 2 prospectieve cohorten geanalyseerd middels latent class methodologie; een cohort vergelijkt MRI en CT; een cohort vergelijkt MRI met BS en klinische evaluatie. Analyse van het eerste cohort is uitgevoerd door middel van de maximum likelihood-based methode; analyse van de tweede groep door middel van een Bayesian gebaseerd latent class analysis model. In het eerste cohort liet de latent class analysis een aanzienlijke verbetering van sensitiviteit en specificiteit zien voor zowel CT (respectievelijk 0.78 en 1.00) als MRI (respectievelijk 0.8 en 0.93). Bij analyse van het tweede cohort, steeg de sensitiviteit van MRI (0.84 vs 0.75) en bleef de specificiteit nagenoeg gelijk (0.99 vs 1.00). De sensitiviteit van BS verminderde (0.94 vs 1.0) en de specificiteit bleef gelijk (beiden 0.89).

Het antwoord op de vraag welke beeldvormende techniek gebruikt moet worden bij een klinische verdenking op een scafoïd fractuur met normale röntgenfoto's, evenals de timing van het verkrijgen van deze beeldvorming, de kosteneffectiviteit en de belangrijkste voordelen voor arts en patiënt wordt gegeven in **Hoofdstukken 9a** en **9b**. Een Cochrane Review voor Diagnostische Test Accuratesse van CT, MRI en BS wordt beschreven in **Hoofdstuk 9a**. De review includeerde 11 prospectieve studies van gemiddelde tot goede kwaliteit en gepoolde gemiddelde waardes gebaseerd op indirecte vergelijkende studies konden worden berekend. De gemiddelde sensitiviteit en specificiteit voor CT was 0.72 en 0.99, 0.88 en 1.00 voor MRI en 0.99 en 0.86 voor BS. Wederom wordt aangetoond dat dat CT en MRI een betere specificiteit dan sensitiviteit, terwijl het tegenovergestelde geldt voor BS. Statistisch gezien lijkt BS de beste modaliteit is en dat CT en MRI vergelijkbaar zijn. Doordat BS een invasieve methode is en diagnostiek pas na 3 dagen plaats kan vinden, wordt het geadviseerd eerder voor CT of MRI te kiezen.

Hoofdstuk 9b gaat verder in op de kosteneffectiviteit en timing van deze modaliteiten. Middels een literatuur review wordt duidelijk dat CT of MRI in de acute setting (dus tijdens Spoedeisende hulp presentatie) een kosteneffectieve strategie is en zeer bevorderlijk voor de patiënt. Er wordt direct een accurate diagnose verkregen (waarbij ook veel fracturen anders dan het scafoïd worden gevonden) en daaropvolgend adequate behandeling ingezet kan worden. Overbehandeling en onnodige polikliniek bezoeken worden hiermee geminimaliseerd.

ALGEMENE DISCUSSIE

Vanuit het perspectief van de patiënt is het lastig voor te stellen dat het diagnostische proces en soort behandeling bij dergelijke polsletsels afhangt van de voorkeuren en mogelijkheden van het ziekenhuis en niet gebaseerd is op wetenschap. Ook al heeft bijna elk ziekenhuis een protocol voor scafoïd fracturen dat met wetenschap onderbouwd moet zijn, toch is er een grote variatie tussen ziekenhuizen in zowel diagnostiek als behandeling.

Om patiënten te selecteren die verdere diagnostiek nodig hebben zijn klinische tests nodig zoals lichamelijk onderzoek. Een bekend voorbeeld van een evidence-based set van voorspellende factoren om een dergelijke groep patiënten te selecteren zijn de Ottawa Ankle Rules en worden gebruikt in voet-enkel trauma. Deze set voorpellers die toe te passen is op de SEH resulteren in het advies om wel of geen röntgenfoto's te verkrijgen. De regels zorgen voor het verminderen van het aantal röntgenfoto's en wachttijd op de SEH zonder het risico op het missen van fracturen te verhogen. Anders gezegd zijn deze regels zeer sensitief en verhogen ze de pre-test kans op een fractuur bij het maken van röntgenfoto's. Hetzelfde geldt voor scafoïd fracturen. Dit proefschrift onderschrijft het belang dat we niet alleen op een pijnlijke anatomische snuifdoos kunnen vertrouwen bij klinische evaluatie van scafoïd fracturen. Met een gemiddelde sensitiviteit van 0.87 tot 1.0 uit eerdere studies, worden nog steeds fracturen gemist. Ook de specificiteit van deze test is insufficiënt. In de ontwikkeling van onze klinische beslisregel bleek zelfs 30% van de bewezen scafoïd fracturen geen drukpijn in de anatomische snuifdoos te hebben en werd een aanzienlijk lagere sensitiviteit berekend dan tot dusver werd aangenomen. Fracturen missen is dus aannemelijk als dit de belangrijkste eerste stap van het diagnostische proces zou zijn.

Het doel van de klinische beslisregel voor scafoïd fracturen was om dit risico te minimaliseren. Met een sensitiviteit van 0.97 binnen het studiecohort is dit doel behaald. Het tweede doel was om een strenge selectie te maken van patiënten die verdere diagnostiek nodig hebben. De regel kan dit aantal met 15% reduceren als we het vergelijken met de huidige praktijk, helaas blijft de specificiteit achter. Als röntgenfoto's normaal blijken, zal verdere diagnostiek nog steeds belangrijk blijven.

Bot scintigrafie lijkt bij patiënten met een klinische verdenking op een scafoid fractuur en normale röntgenfoto's statistisch beter dan MRI en CT. Echter, BS is een invasieve techniek die pas 3 dagen na trauma uitgevoerd kan worden en geniet daardoor niet de voorkeur. Direct MRI of CT is de beste methode. Beiden zijn accuraat en kosteneffectief als ze op de SEH worden uitgevoerd. Ook veel andere pols en carpale fracturen kunnen hiermee gevonden worden. Door accurate diagnose tijdens het eerste SEH bezoek wordt direct adequaat behandeld en wordt onnodige behandeling en follow-up voorkomen. De impact op verlies van functie daarmee ook. Helaas zijn deze modaliteiten niet overal direct beschikbaar.

Maar ook CT en MRI zijn soms lastig te beoordelen. Minimale unicorticale lijnen op de CT kunnen haarvaten zijn in plaats van fractuurlijntjes. Dit proefschrift benadrukt het belang om de CT reconstructies in het longitudinale vlak van het scafoïd te maken voor optimale visualisatie(1). CT is niet alleen cruciaal voor het detecteren van fracturen, maar ook voor het classificeren van de breuk. Verplaatsing, comminutie en locatie zijn belangrijke factoren in het voorspellen van fractuurheling en bepalen van immobilisatie duur en/of operatieve opties. Daarbij kan CT ook gebruikt worden in de evaluatie van progressie van heling gedurende de behandeling. Röntgenfoto's zijn hierbij niet accuraat(2-6). Het is zelfs te beargumenteren dat elke scafoïd fractuur geëvalueerd moet worden middels CT om bovenstaande aspecten te kunnen beoordelen en dat CT dus de ultieme beeldvormende techniek. Daarbij komt dat de innovatieve verbeteringen van CT met onder andere Dual Energy ook de voordelen van MRI (wekedelen letsels) overgenomen kunnen worden(7).

Bij het beoordelen van MRI is bone bruising of beenmerg oedeem een uitdaging: Is dit een fractuur of niet? Het expertpanel in dit proefschrift beschreef een uitgebreide focale zone van oedeem (zoals in stress fracturen) zonder trabeculaire of corticale onderbreking als fractuur. Een voordeel van MRI is het detecteren van ligamentaire letsel(8, 9). Een FOOSH (val op de uitgestrekte hand) is een traumamechanisme waarbij bijvoorbeeld ook letsel van het scafolunaire ligament kan optreden. Dergelijke wekedelen letsels zijn moeilijker te detecteren met röntgenfoto's of CT.

Er is een gebrek aan een referentie standaard voor definitieve diagnose van scafoïd fracturen. Waar recente studies vaak gebruik maken van 6-weekse röntgenfoto's, laat dit proefschrift zien dat deze standaard inaccuraat is en niet praktisch in onderzoeksverband. Verder gebruik wordt daarom afgeraden. Latent class analysis kan een oplossing bieden in dergelijke gevallen, echter is het nog onduidelijk of dit een verbetering is ten opzichte van de reguliere statistische analyses. MRI of CT is door de praktische toepassing en zijn diagnostische accuratesse een betere referentie standaard.

REFERENTIES

1. Mallee WH, Doornberg JN, Ring D, Maas M, Muhl M, van Dijk CN, et al. Computed tomography for suspected scaphoid fractures: comparison of reformations in the plane of the wrist versus the long axis of the scaphoid. Hand (N Y). 2014;9(1):117-21.
2. Buijze GA, Wijffels MM, Guitton TG, Grewal R, van Dijk CN, Ring D. Interobserver reliability of computed tomography to diagnose scaphoid waist fracture union. J Hand Surg Am. 2012;37(2):250-4.
3. Gilley E, Puri SK, Hearns KA, Weiland AJ, Carlson MG. Importance of Computed Tomography in Determining Displacement in Scaphoid Fractures. J Wrist Surg. 2018;Feb;7(1)(2163-3916 (Print)):38-42.
4. Grewal R, Frakash U, Osman S, McMurtry RY. A quantitative definition of scaphoid union: determining the inter-rater reliability of two techniques. Journal of orthopaedic surgery and research. 2013;8:28.
5. Grewal R, Lutz K, MacDermid JC, Suh N. Proximal Pole Scaphoid Fractures: A Computed Tomographic Assessment of Outcomes. J Hand Surg Am. 2016;41(1):54-8.
6. Grewal R, Suh N, Macdermid JC. Use of computed tomography to predict union and time to union in acute scaphoid fractures treated nonoperatively. J Hand Surg Am. 2013;38(5):872-7.
7. Wong WD, Shah S, Murray N, Walstra F, Khosa F, Nicolaou S. Advanced Musculoskeletal Applications of Dual-Energy Computed Tomography. Radiologic clinics of North America. 2018; Jul;56(4)(1557-8275 (Electronic)):587-600.
8. Hafezi-Nejad N, Carrino JA, Eng J, Blackmore C, Shores J, Lifchez SD, et al. Scapholunate Interosseous Ligament Tears: Diagnostic Performance of 1.5 T, 3 T MRI, and MR Arthrography-A Systematic Review and Meta-analysis. Acad Radiol. 2016;Sep;23(9)(1878-4046 (Electronic)):1091-103.
9. Meister DW, Hearns KA, Carlson MG. Dorsal Scaphoid Subluxation on Sagittal Magnetic Resonance Imaging as a Marker for Scapholunate Ligament Tear. J Hand Surg Am. 2017;Sep;42(9)(1531-6564 (Electronic)):717-21.

DANKWOORD / ACKNOWLEDGEMENTS

Allereerst wil ik mijn onnavolgbare promotieteam bedanken.
Mijn Promotores:
Niek van Dijk, veel dank voor de vrijheid die je me gaf om niet alleen dit proefschrift af te ronden, maar ook om alle sidesteps te kunnen nemen die in m'n onderzoekstijd bij me op kwamen. Die maakte de taaie wetenschap behapbaar.
&
Carel Goslings, door jou kwam ik altijd vaak en graag aan de overkant. Je energie, enthousiasme en gedrevenheid zijn bewonderenswaardig en aanstekelijk. Ook buiten de muren van het ziekenhuis gaat dat door. Veel dank voor al je steun, sturing en waardevolle input.

Mijn Co-promotores:
Mario Maas, voor alle musculoskeletale radiologie bel ik je maar al te graag. Je bent laagdrempelig te benaderen voor overleg, zowel op klinisch als wetenschappelijk gebied. Open-minded en oprecht geïnteresseerd. Dank voor je continue steun tijdens mijn carrière.
&
Job Doornberg, hoe kan ik je bedanken. De ontmoeting bij BIS was de kick-off van een vruchtbare samenwerking die inmiddels wat verder gaat dan alleen dit proefschrift. Traumaplatform, de Artificial Intelligence projecten en ontelbare andere zaken. Je bent een machine op de werkvloer. Dank voor de 'spark'. Maar vooral ook dank voor de mooie vriendschap. Je bent ongelofelijk gul en open. Deze productieve symbiose hoeft wat mij betreft nooit te stoppen.

Gino Kerkhoffs, je bent een schitterend voorbeeld van hoe het kan en moet. Tomeloze energie, kennis, kunde en humor. Op het veld, maar zeker ook daarbuiten. Grazie mille voor het vertrouwen in deze jonge hond, voor het sparren, voor het pushen, voor het mogelijk maken van mooie stappen en niet te vergeten voor het beoordelen van mijn proefschrift!

Aan alle overige leden van de promotiecommissie, Professor Nollet, Professor Oostra, Professor Bennink, Professor Coert, dr. Rood en dr. Rhemrev, zeer veel dank voor het beoordelen van mijn proefschrift.

Alle co-auteurs en mede-onderzoekers van de afgelopen jaren, waaronder Peter Kloen, Marjolein Mulders, Annefloor van Enst, Geert Buijze, Niels Schep, Monique Walenkamp, David Ring, Maaike Muhl, Wout Veltman, Leendert Blankevoort, Erik Henny, Sjoerd Kamminga, Thierry Guitton, Jos Mellema, Frank Beeres, Riekie de Vet, Junfeng Wang, Rudolph Poolman en Maarten de Vos, ongelofelijk veel dank voor de support. Samenwerken in teamverband is noodzakelijk om elk onderzoek van de grond te krijgen en te publiceren, no matter what the outcome. Zonder jullie was dit proefschrift er niet geweest.

Alle collega's van het AMC, Tergooi ziekenhuis, MC Slotervaart, Flevoziekenhuis, de kweekvijver, TWG, SWOAHS en HealthPlus. Het is heerlijk werken in al deze teams. Dank dat ik door jullie altijd fluitend naar mijn werk ga.

De staf van het AMC, Tergooi Ziekenhuis, MC Slotervaart en Flevoziekenhuis, dank voor jullie inzet, energie en vertrouwen om van deze Rotterdammer een Orthopedisch chirurg te maken. Het is een eer.

Het secretariaat van het AMC, Veerle, Tineke, Ellen, Marga en Rosalie, een speciale dank voor jullie, ik hoefde maar te roepen en jullie hielpen me bij alles! En dat was altijd erg hard nodig!

Stuf, Died, Jetje, Nine, Lucas, Jan Maerten, Wyb, Pieter, Geert en iedereen die Traumaplatform een warm hart toedraagt: dank voor jullie inzet, weergaloze en krankzinnige ideeën en eindeloze energie die Traumaplatform heeft gebracht waar het nu is.

Lieve Nien, dank voor alle jaren die jij mij hebt geholpen om door de promotie heen te ploegen!

Familie Backes, hoor ik er nu eindelijk echt bij? Wat een fijne familie zijn jullie waar ik me al vanaf het begin thuis voel, dank voor het warme welkom van deze laaggeschoolde...

Baar, Bauk, Bik, Faus, Jil, Pietje, Sytz, Mike, Boer en alle Roffa-boys van de legendarische kerstdiners, Ronnie en de maandagavond gasten, Nessie, Seppie, Bernie, Wim, Shack, Stock, B, Art, G, JuJu: One Love

M’n paranimfen: Koen en Maai. Twee van mijn oudste en beste maten waar ik eigenlijk al vanaf hun geboorte mee opgroei en nu nog steeds graag naast me heb staan. Dat zegt genoeg.

Rudi en Karlijn, jullie houden mijn broertje en zus in het gareel. Daar ben ik jullie meer dan dankbaar voor. Nu weet ik dat het wel goed komt met ze!

Uiteraard veel dank aan mijn kleine grote zus. Tien, de liefste zus, je blijft ons achter de broek aan zitten zodat we elkaar regelmatig zien. Jij houdt ons hecht.

Pap en mam, waar moet ik beginnen. Dank voor letterlijk alles; de kansen, de vrijheid en de mogelijkheden, de immer aanwezige trots ook al is die nergens op gebaseerd, het corrigeren en het niet-corrigeren, de open deur en het altijd fijne thuiskomen.

Lieve Nouk, wij maken er in korte tijd zo’n mooi leven van samen. Dit geluk is niet te omschrijven. En binnenkort komt er ook nog een kleine baas bij! Het is ongelofelijk hoe heerlijk alles is met jou.

PORTFOLIO

Courses

		Year
Knee multi-ligament course, Arthrex	Munich, Germany	2018
Knee arthroscopy course	Utrecht	2016
AO fracture management, basic principles	Atlanta, USA	2014
Advanced Trauma Life Support	New York, USA	2013
Upper extremity trauma, Acumed	Lissabon, Portugal	2012
Legislation and Organization for Clinical Researchers	Amsterdam	2011
Upper extremity trauma, Acumed	Lissabon, Portugal	2011

Seminars, workshops and master classes

Medical Business Masterclass	Amsterdam	2017
Entrepreneurship in Health and Life Sciences	Amsterdam	2012

Review activities

Abstract reviewer SEOHS 2014 - present

Oral Presentations

When is surgery indicated for treatment of acute scaphoid fractures? *Dutch Orthopaedic Association (NOV), 2018, Rotterdam, The Netherlands*

When is surgery indicated for treatment of delayed or non-union of scaphoid fractures? *Dutch Orthopaedic Association (NOV), 2018, Rotterdam, The Netherlands*

Clinical Decision Rule for Detection of Scaphoid Fractures after Wrist Injury. *Annual Meeting Dutch Trauma Association, 2016, Amsterdam, The Netherlands*

Staples or Sutures for Wound Closure after Total Hip Arthroplasty, a Multicenter Randomized Controlled Trial. *European Hip Society 2016, Munich, Germany*

Battle: Computed Tomography vs Bone Scintigraphy in Suspected Scaphoid fractures. *Annual Meeting, Dutch Trauma Association, 1-2 November 2012, Amsterdam, The Netherlands*

Comparison of Magnetic Resonance Imaging and Computed Tomography in Diagnosis of Suspected Scaphoid Fractures. *11th EFORT Congress, June 2-5 2010, Madrid, Spain*

Clinically Suspected for a Scaphoid Fracture, Then What? *Symposium Verstoord Bewegen AMC Amsterdam, 2011*

Comparison of Magnetic Resonance Imaging and Computed Tomography in Diagnosis of Suspected Scaphoid Fractures. *Annual Meeting, Dutch Trauma Association, 5-6 November 2009, Amsterdam, The Netherlands*

Comparison of Magnetic Resonance Imaging and Computed Tomography in Diagnosis of Suspected Scaphoid Fractures. *64th Annual Meeting of the American Society for Surgery of the Hand, September 2009, San Francisco,USA*

Poster Presentations

Staples or Sutures for Wound Closure after Total Hip Arthroplasty. *Annual meeting American Academy of Orthopaedic Surgeons, 2017, San Diego, USA*

Lack of Knowledge in Anatomy and Inconsistent Clinical Evaluation of the Wrist in Scaphoid Injury. *Combined Meeting of Orthopaedic Research Societies 13-16 Oktober 2013, Venetië, Italië*

Management of Suspected Scaphoid Fractures in The Netherlands. *Annual Meeting, Dutch Trauma Association, 11-12 November 2010, Amsterdam, the Netherlands*

Teaching	Year
Lectures	
Ankle Fractures, common trunk residency program	2017
The new hip fracture guidelines, AMC residency program	2016
Stress fractures of the lower leg, AMC residency program	2015
Research supervising	
E.P. Henny, medical student	2014
E.S. Veltman, medical student	2012
Parameters of esteem	
H2020 SME phase 1 for Artificial Intelligence based Algorythm 3dCT segmentation	2018
TKI-LSH research and development fund for Patient Reported Outcome Platform	2018
Marti Keunig Eckhardt Foundation: Start-up fund for interobserver websites	2013
Zon MW – Doematigheid subsidie, t.b.v. Amsterdam Wrist Rules	2013
Zon MW – Network Fund for SEOHS 2011	2011
AMC PhD – Scholarship: Diagnostic work-up of scaphoid fractures	2010
Stichting Professor Boerema Reisfonds	2009
Spinoza fonds	2009

Publications

Articles

Mallee WH, Doornberg JN, Ring D, van Dijk CN, Maas M and Goslings JC
Comparison of Magnetic Resonance Imaging and Computed Tomography in Diagnosis of Suspected Scaphoid Fractures.
J Bone Joint Surg Am. 2011 Jan 5;93(1):20-8

Buijze GA, **Mallee WH**, Beeres FJ, Hanson TE, Johnson WO, Ring D
Diagnostic Performance Tests for Suspected Scaphoid Fractures differ with Conventional and Latent Class Analysis.
Clin Orthop Relat Res. 2011 Dec;469(12):3400-7

Mallee WH, Veltman ES, Doornberg JN, Blankevoort L, Goslings JC, Van Dijk CN
Variatie in Management bij Verdenking op Scafoïdfracturen.
Ned Tijdschr Geneeskd. 2012;156(28):A4514

Mallee WH, Doornberg JN, Ring D, Maas M, Muhl M, van Dijk CN, Goslings JC
Computed Tomography for Suspected Scaphoid Fractures: Comparison of Reformations in Plane of the Wrist Versus the Long Axis of the Scaphoid.
Hand (N Y). 2014 Mar;9(1):117-21

Mallee WH, Henny EP, Van Dijk CN, Kamminga SP, Van Enst WA, Kloen P
Clinical evaluation of Suspected Scaphoid Fractures: a Systematic Review and Meta-analysis.
J Hand Surg Am. 2014 Sep;39(9):1683-1691

Kerkhoffs GMMJ, Hendrickx RPM, **Mallee WH**, De Leeuw PAJ, Van Dijk CN
Arthroscopic Ankle Fusion: The Posterior Approach.
Orthopaedic Knowledge Online Journal 2014

Mallee WH, Weel H, van Dijk CN, van Tulder MW, Kerkhoffs GM, Lin CW
Surgical versus conservative treatment for high-risk stress fractures of the lower leg (anterior tibial cortex, navicular and fifth metatarsal base): a systematic review.
Br J Sports Med. 2015 Mar;49(6):370-6

Mallee WH, Wang J, Poolman RW, Kloen P, Maas M, de Vet HC, Doornberg JN
Computed tomography versus magnetic resonance imaging versus bone scintigraphy for clinically suspected scaphoid fractures in patients with negative plain radiographs.
Cochrane Database Syst Rev. 2015 Jun 5;(6):CD010023

Weel H, **Mallee WH**, van Dijk CN, Blankevoort L, Goedegebuure S, Goslings JC, Kennedy JG, Kerkhoffs GM
The effect of concentrated bone marrow aspirate in operative treatment of fifth metatarsal stress fractures; a double-blind randomized controlled trial.
BMC Musculoskelet Disord. 2015 Aug 20;16:211

Meijer DT, Doornberg JN, **Mallee WH**, van Dijk CN, Kerkhoffs GM, Stufkens SA
Guesstimation of posterior malleolar fractures on lateral plain radiographs.
Injury. 2015 Oct;46(10):2024-9

Mallee WH, Mellema JJ, Guitton TG, Goslings JC, Ring D, Doornberg JN; Science of Variation Group.
6-Week Radiographs Usuitable for Diagnosis of Suspected Scaphoid Fractures.
Arch Orthop Trauma Surg. 2016 Jun;136(6):771-8

Claessen FM, Meijer DT, van den Bekerom MP, Gevers Deynoot BD, **Mallee WH**, Doornberg JN, van Dijk CN.
Reliability of Classification for Post-traumatic Ankle Osteoarthritis.
Knee Surg Sports Traumatol Arthrosc. 2016 Apr;24(4):1332-7

Mellema JJ, **Mallee WH**, Guitton TG, van Dijk CN, Ring D, Doornberg JN, SOVG and Traumaplatform Study Collaborative
Online Studies on Variation in Orthopedic Surgery: Computed Tomography in MPEG4 versus Dicom Format.
J Digit Imaging. 2017 Oct;30(5):547-554

Zwiers R, Weel H, **Mallee WH**, Kerkhoffs GMMJ, van Dijk CN; Ankle Platform Study Collaborative – Science of Variation Group.
Large Variation in Use of Patient-reported Outcome Measures: a Survey of 188 Foot and Ankle Surgeons.
Foot Ankle Surg. 2018 Jun;24(3):246-251

Hodel S, Link BC, Babst R, **Mallee WH**, Posso P, Beeres FJP, Traumaplatform Study Collaborative
Perioperative Management of External Fixation in Staged Protocols: an International Survey.
Eur J Orthop Surg Traumatol. 2018 May;28(4):565-572

Chapters

Mallee WH, de Vos MJ
Diagnostic Work-up for Suspected Scaphoid Fractures
Scaphoid Fractures: Evidence-Based Management 2018

De Boer BNP, Doornberg JN, **Mallee WH**, Buijze GA
Surgical versus Conservative Treatment for Nondisplaced Scaphoid Waist Fractures
Scaphoid Fractures: Evidence-Based Management 2018

CURRICULUM VITAE

Wouter Mallee was born in Rotterdam on the 4th of July 1984. In 2003 he graduated from the Montessori Lyceum and travelled for the following year. In 2004 he started medical school at the Academic Medical Center of Amsterdam (UvA). At the end of his medical study he started working at BISLife in the human donor explantation team, where he met Job Doornberg; with him the scientific endeavors began in the field of Orthopaedic Trauma. Together with his brilliant team (Prof. Dr. van Dijk, Prof. Dr. Goslings, Prof. Dr. Maas and Dr. Doornberg) he received a full AMC PhD Scholarship for this PhD project in 2010. In 2011, he stepped outside the hospital environment to participate in the National Think-tank to focus on 'the Future of Work'. Prof. Dr. Kerkhoffs excited his interest in Orthopaedic Sports Medicine and soon this enthusiasm resulted in a research fellowship at the George Institute of Global Health in Sydney in 2013. In 2014, he founded the Traumaplatform Foundation together with Dr. Doornberg and Dr. Stufkens with the goals to stimulate collaboration between Trauma-, Orthopaedic and Plastic/Reconstructive surgery and to facilitate research in the field of (Orthopaedic) Trauma; a successful initiative so far. 2014 was also the year that his Orthopaedic Surgery training started. He worked as a resident general surgery in the Flevohospital and as a resident orthopaedic surgery at the Academic Medical Center in Amsterdam, Tergooi Hospital in Hilversum and the Slotervaart Hospital. His final year of training will be in the Flevohospital, where he aims to further develop his knowledge and skills in Orthopaedic Trauma and Sports Injuries.