



## UvA-DARE (Digital Academic Repository)

### Scaphoid fractures: anatomy, diagnosis and treatment

Buijze, G.A.

**Publication date**  
2012

[Link to publication](#)

#### **Citation for published version (APA):**

Buijze, G. A. (2012). *Scaphoid fractures: anatomy, diagnosis and treatment*. [Thesis, fully internal, Universiteit van Amsterdam].

#### **General rights**

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

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Chapter

# 4

Diagnostic performance tests for  
suspected scaphoid fractures  
differ with conventional and latent class  
analysis

Buijze GA, Mallee WH, Beeres FJ, Hanson TE, Johnson WO,  
Ring D



## Abstract

**Background** Evaluation of the diagnostic performance characteristics of radiological tests for diagnosis of a true fracture among suspected scaphoid fractures are 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 and can be used to determine diagnostic performance characteristics when there is no consensus (gold) reference standard.

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

**Patients and Methods** Using data from two studies, one that prospectively studied 34 patients who had both 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 to those calculated using formulas based on a reference standard.

**Results** In the first cohort, the calculated sensitivity and specificity with latent class analysis was different than it was with traditional reference standard based calculations for both the CT in the scaphoid planes (sensitivity: 0.78 vs. 0.67; specificity: 1.0 vs. 0.96) and the MRI (sensitivity: 0.80 vs. 0.67; specificity: 0.93 vs. 0.89). In the second cohort, the biggest differences were in the sensitivity of MRI (0.84 vs. 0.75) and the sensitivities of physical examination maneuvers (range, 0.63-0.73 vs. 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. 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<sup>1,2</sup>, followup MRI<sup>3</sup>, and standards based on a combination of test results<sup>4</sup>. A recent systematic review of diagnostic tests for suspected scaphoid fractures documents substantial variation in diagnostic performance characteristics – i.e., 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)<sup>5</sup>. Inconsistency in imaging protocols and reference standards may 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<sup>6</sup>. An example of a disease for which there is no accepted reference standard is compartment syndrome. Since we cannot measure the diagnostic performance of various tests for compartment syndrome against a reference standard, the usual formulas for sensitivity and specificity do not apply. Latent class analysis takes advantages of known but unobserved groupings of patients based on disease status. While there can be more than two groups, only two are considered here, namely “diseased” or “not diseased.” A statistical analysis of these 2 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 with use of either maximum likelihood (which is a standard method of statistical inference<sup>7,8</sup> that obtains parameter estimates that maximize the probability of observing the actual data), or the Bayesian method<sup>9,10</sup> (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 maximum likelihood 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 won’t 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).



Many other diseases lacking a consensus reference standard have been studied using latent class analysis, such as peripheral joint psoriatic arthritis<sup>11</sup>, carpal tunnel syndrome<sup>12</sup>, and various infectious diseases<sup>13-16</sup>. Its use in some of these studies confirmed that the diagnostic performance values of various tests were similar to those found with traditional analysis based on a reference standard, which supports the accuracy of the reference standard. The possibilities for application of latent class analysis in orthopedic diagnostic studies have been previously reviewed by us<sup>17</sup>.

The aim of this study was to test the null hypothesis that the diagnostic performance characteristics calculated using latent class analysis and those calculated using standard formulas based on a reference standard are the same for the diagnosis of true fractures among suspected scaphoid fractures.

## Patients and Methods

We applied latent class analysis to data from two prospective cohort studies: in one we compared MRI to CT, and in the other we compared MRI to 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 vs. CT) included 34 patients diagnosed with a suspected scaphoid fracture in the Emergency Department between April and October 2008<sup>18</sup>. We included adult patients presenting within 24 hours of injury and having tenderness of the scaphoid in the anatomical 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. All radiographs were independently evaluated by the treating radiologist and the treating trauma surgeon. All patients underwent an MRI and CT scan. The authors performed both examinations on the same day, at an average of 3.6 days (range, 0 to 10 days) after initial trauma. All MRIs were performed with an open 1.0 Tesla MR scanner (Panorama 1.0T, Philips Medical Systems, Eindhoven, 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. They positioned the patient 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 to off-centered scanning in a conventional tube. They performed multidetector, high-resolution CT 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 CMC-joints. They positioned patients in the "superman" position, prone

with the affected arm above the body and the palm facing down. The authors made reconstructions in planes, defined by the long axis of the scaphoid<sup>19</sup>. They defined sagittal plane images of the scaphoid as reconstructions that provided a lateral view of the scaphoid bone, as defined by the central longitudinal axis of the scaphoid. They defined coronal plane images those that provided a posteroanterior view of the scaphoid in the anatomic plane and in line with the axis of the scaphoid<sup>20,21</sup>. Criteria for a scaphoid fracture on CT images were 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. Criteria for a fracture on MRI included the presence of a cortical fracture line, a trabecular fracture line, or a 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 or not. Three of us (JCG, MM, and CNvD) formed the panel that evaluated MRI 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<sup>22</sup> – was  $\kappa=0.62$ , which reflects overall substantial agreement. The reference standard for a true scaphoid fracture was an abnormal lucent line in the scaphoid<sup>23</sup>.

The second cohort (MRI vs. bone scintigraphy and clinical tests) included 78 patients who visited a single emergency department for a suspected scaphoid fracture between April 2004 and January 2007<sup>24</sup>. All 78 patients were previously reported in that study including all data from the MRIs and bone scans. We included adult patients with a suspected scaphoid fracture (tender anatomical snuffbox and pain in the snuffbox when applying axial pressure on the first or second digit), a 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. The authors performed clinical tests 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 both the suspected and contralateral side, and consisted of (1) inspection of the anatomical snuffbox for the presence of a hematoma and/or swelling in comparison with the contra-lateral side, (2) measurements of range of wrist flexion and extension, (3) measurements 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. The authors considered the motion and strength tests positive if there was a loss of 25% or more compared to the uninjured side. MRIs 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 the patient's head. The authors then wrapped a flexible surface coil around the wrist. The MRI protocol included coronal T1-weighted turbo spin-echo images with a TR of 450 ms, a TE of 13 ms, a field view of 180 mm, a base resolution of 512, two averages, a slice thickness of 3 mm with a distance factor of 10%, and a scan time of 2.17 minutes. The parameters for the coronal fat-suppressed T2-weighted fast spin-echo images were 5220/73 ms (TR/TE), a field of view of 220 mm, a base resolution of 448, three averages, a slice thickness of 3 mm with a distance factor of 10%, and a 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 Sky-Light gamma camera (Philips Medical Systems, Eindhoven, Netherlands). They performed palmar and dorsal images of both wrists between 2.5 and 4 hours after the injection of 500 MBq of technetium-99m diphosphonate (Tc-99m-HDP) in order to visualize the osteoblast activity. Observations were performed by an experienced clinical nuclear physician (JWA). The reference standard for a true scaphoid fracture was a combination MRI, bone scintigraphy, and clinical examination results. Where there was a discrepancy between MRI and bone scintigraphy (ie., only one tested positive), they defined a true fracture as an abnormal lucent line within the scaphoid on radiographs at the 6 week follow-up 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 previously described by us in more detail with examples<sup>17</sup>.

The maximum likelihood-based (ML) method, developed by Hui and Walter<sup>6</sup>, assumes conditional independence of the tests, meaning that presence or absence of one symptom sign or test results 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 maximum likelihood 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. Since 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 vs. 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 allows for conditional dependence among multiple test results by relying on surgeon estimation of plausible dependencies between test results<sup>25</sup>. Johnson, Gastwirth and Pearson<sup>26</sup> provided Bayesian methods for the Hui-Walter model and Dendukuri and Joseph<sup>27</sup> 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 has been previously reported in a review of the literature<sup>5</sup>. Specifically we selected the following values: 0.78<sup>28</sup> for the sensitivity and 0.52<sup>29</sup> for the specificity of bone scintigraphy; 0.8<sup>4</sup> for the sensitivity and 0.95<sup>30</sup> for the specificity of MRI; and 0.05 for prevalence of true fractures among suspected scaphoid fractures<sup>31</sup>.

## 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 the CT in the scaphoid planes had the highest diagnostic performance values, and the CT in the axial and sagittal plane had the lowest (Table 1). For the latter, the diagnostic performance values were nearly identical in both methods (Se: 0.16 vs. 0.17; Sp: 0.89 vs. 0.89). However, the sensitivity and specificity of both the CT in the scaphoid planes (Se: 0.78 vs. 0.67; Sp: 1.0 vs. 0.96) and the MRI (Se: 0.80 vs. 0.67; Sp: 0.93 vs. 0.89) were

| Diagnostic test       | Latent class analysis |                      | Calculations using reference standard |                      |
|-----------------------|-----------------------|----------------------|---------------------------------------|----------------------|
|                       | 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 (0.89-1.0)      |                                       |                      |
| 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                           |
| -                 | +        | -   | -                     | 0.0032                           |
| +                 | +        | -   | -                     | 0.9991                           |
| -                 | -        | +   | -                     | 0.1055                           |
| +                 | -        | +   | -                     | 1                                |
| -                 | +        | +   | -                     | 0.1535                           |
| +                 | +        | +   | -                     | 1                                |
| -                 | -        | -   | +                     | 0.2105                           |
| +                 | -        | -   | +                     | 1                                |
| -                 | +        | -   | +                     | 0.2906                           |
| +                 | +        | -   | +                     | 1                                |
| -                 | -        | +   | +                     | 0.9377                           |
| +                 | -        | +   | +                     | 1                                |
| -                 | +        | +   | +                     | 0.9586                           |
| +                 | +        | +   | +                     | 1                                |

+ Postive test result; - Negative test result

slightly higher in the latent class analysis than the reference standard calculations and the prevalence was slightly higher (18.9% vs. 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 and positive 6 week radiography, the positive predictive value was 0.21 (Table 2).

When compared to 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 vs. 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 vs. 0.84; Sp: 1.0 vs. 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 of over 25% of wrist flexion, the specificities of the other four tests on motion and strength were

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

| Diagnostic test                        | Latent class analysis   |                         | Calculations using reference standard |                         |
|--|-------------------------|-------------------------|---------------------------------------|-------------------------|
|  | Sensitivity<br>(95% PI) | Specificity<br>(95% PI) | Sensitivity<br>(95% CI)               | Specificity<br>(95% CI) |
| Snuffbox swelling                      | 0.63 (0.45-0.80)        | 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

slightly higher in the latent class analysis (range 0.14-0.27 vs. range 0.6-0.23). The latent class analysis estimates of the presence of snuffbox swelling showed a higher sensitivity (0.63 vs. 0.42) and a lower specificity (0.41 vs. 0.76), whereas for the presence of a hematoma, the estimates showed a lower sensitivity (0.36 vs. 0.92) and a higher specificity (0.71 vs. 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 two or six 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 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.

Readers should be aware of limitations to our study. First, is that our analysis is based upon data made available to us and is subject to all its weaknesses, which were enumerated in the prior publications of those works, but relate primarily to small sample size for our purposes. Although the first cohort had a small sample size, the maximum likelihood 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 the 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 maximum likelihood based intervals when sample sizes are small to moderate, as is the case here. Second, 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<sup>32</sup>, which we feel 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 maximum likelihood problems, among others.

Latent class analysis is increasingly used to study diagnostic tests for diseases lacking a consensus reference standard<sup>11-16</sup>, particularly in the field of psychiatry<sup>33</sup>. In a study similar to ours, Faraone and Tsuang<sup>33</sup> analyzed prior data on the diagnosis of major depressive disorder<sup>34</sup> 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 can also 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<sup>35</sup>. 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 both the CT and MRI calculated by latent class analysis were in the range of previously reported studies<sup>5</sup>, while 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<sup>4,30</sup>. All diagnostic performance parameters calculated by latent class analysis were closer to the average diagnostic parameters based on pooled data in a meta-analysis<sup>5</sup>.

In the second cohort, the sensitivity of MRI calculated by latent class analysis was also closer to the average sensitivity based on pooled data in a meta-analysis<sup>5</sup>. 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 with those of Unay et al., who evaluated the diagnostic performance characteristics of 10 physical examination maneuvers for the triage of suspected scaphoid fractures using MRI as the reference standard<sup>36</sup>. 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 that 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<sup>5</sup>. This reference standard is controversial<sup>37,38</sup>. Low and Raby<sup>38</sup> showed poor accuracy and reliability for follow-up radiography as a diagnostic test for scaphoid fracture with normal initial radiographs. Non-displaced scaphoid fractures can be subtle, such that we cannot agree on a reliable reference standard. Furthermore, some non-displaced 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 upon whether traditional or latent class analysis is used, additional research is needed to determine which method leads to better patient care. The situation of an imperfect or debated reference standard is commonplace in orthopaedic surgery and latent class analysis may merit wider utilization if in fact it provides more accurate information that leads to better patient care. In any case, given the inherent uncertainty in many diagnostic methods it may 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

The authors 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.

## References

1. 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(818):109-13.
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(6):495-9.
3. 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;187(6):1453-6.
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. Yin ZG, Zhang JB, Kan SL, Wang XG. Diagnosing suspected scaphoid fractures: a systematic review and meta-analysis. *Clin Orthop Relat Res* 2010;468(3):723-34.
6. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. *Biometrics* 1980;36(1):167-71.
7. Casella G, Berger RL. *Statistical Inference*. Ed 2. Pacific Grove, CA: Duxbury Press; 2001.
8. Fisher RA. On the mathematical foundations of theoretical statistics. *Phil Trans R Soc A* 1922;222:309-368.
9. Bayes T. An essay towards solving a problem in the doctrine of chances. *Phil Trans R Soc London* 1763;53:370-418.
10. Christensen R, Johnson W, Branscum A, Hanson T. *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*. Boca Raton, FL: CRC Press; 2010.
11. Symmons DP, Lunt M, Watkins G, Helliwell P, Jones S, McHugh N, et al. 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(3):552-7.
12. 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(2):502-7.
13. 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(1):106-14.
14. De La Rosa GD, Valencia ML, Arango CM, Gomez CI, Garcia A, Ospina S, et al. Toward an operative diagnosis in sepsis: a latent class approach. *BMC Infect Dis* 2008;8:18.
15. Butler JC, Bosshardt SC, Phelan M, Moroney SM, Tondella ML, Farley MM, et al. 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(9):1416-23.
16. Tuyisenge L, Ndimubanzi CP, Ndayisaba G, Muganga N, Menten J, Boelaert M, et al. 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(2):e11-8.
17. 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.
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. 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.
20. 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.
21. 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.
22. Siegel S, Castellan JN. *Nonparametric Statistics for the Behavioral Sciences*. New York, NY: McGraw-Hill; 1988.
23. 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.
24. 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.
25. 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(3):855-63.
26. Johnson WO, Gastwirth JL, Pearson LM. Screening without a “gold standard”: the Hui-Walter paradigm revisited. *Am J Epidemiol* 2001;153(9):921-4.
27. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001;57(1):158-67.
28. Breederveld RS, Tuinebreijer WE. Investigation of computed tomographic scan concurrent criterion validity in doubtful scaphoid fracture of the wrist. *J Trauma* 2004;57(4):851-4.
29. Nielsen PT, Hedeboe J, Thommesen P. Bone scintigraphy in the evaluation of fracture of the carpal scaphoid bone. *Acta Orthop Scand* 1983;54(2):303-6.
30. 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(5):1287-93.
31. 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(4):235-42.
32. 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.
33. Faraone SV, Tsuang MT. Measuring diagnostic accuracy in the absence of a “gold standard”. *Am J Psychiatry* 1994;151(5):650-7.
34. 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(4):337-45.
35. Walter SD, Irwig L, Glasziou PP. Meta-analysis of diagnostic tests with imperfect reference standards. *J Clin Epidemiol* 1999;52(10):943-51.
36. 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.
37. Gabler C, Kukla C, Breitenseher MJ, Trattng 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(2):150-4.
38. Low G, Raby N. Can follow-up radiography for acute scaphoid fracture still be considered a valid investigation? *Clin Radiol* 2005;60(10):1106-10.