



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

8

Training improves interobserver reliability
for the diagnosis of scaphoid fracture
displacement

Buijze GA, Guitton TG, Van Dijk CN, Ring D,
The Science of Variation Group*



Abstract

Background The diagnosis of displacement in scaphoid fractures is notorious for poor interobserver reliability.

Questions/purposes We tested whether training can improve interobserver reliability and sensitivity, specificity, and accuracy for the diagnosis of scaphoid fracture displacement on radiographs and CT scans.

Methods Sixty-four orthopaedic surgeons rated a set of radiographs and CT scans of 10 displaced and 10 nondisplaced scaphoid fractures for the presence of displacement, using a web-based rating application. Before rating, observers were randomized to a training group (34

observers) and a nontraining group (30 observers). The training group received an online training module before the rating session, and the nontraining group did not. Interobserver reliability for training and nontraining was assessed by Siegel's multirater kappa and the Z-test was used to test for significance.

Results There was a small, but significant difference in the interobserver reliability for displacement ratings in favor of the training group compared with the nontraining group. Ratings of radiographs and CT scans combined resulted in moderate agreement for both groups. The average sensitivity, specificity, and accuracy of diagnosing displacement of scaphoid fractures were, respectively, 83%, 85%, and 84% for the nontraining group and 87%, 86%, and 87% for the training group. Assuming a 5% prevalence of fracture displacement, the positive predictive value was 0.23 in the nontraining group and 0.25 in the training group. The negative predictive value was 0.99 in both groups.

Conclusions Our results suggest training can improve interobserver reliability and sensitivity, specificity and accuracy for the diagnosis of scaphoid fracture displacement, but the improvements are slight. These findings are encouraging for future research regarding interobserver variation and how to reduce it further.

*The Science of Variation Group consists of the following members: Osterman AL, Wahegaonkar AL, Ladd A, Barquet A, van Vugt AB, Shyam AK, Swigart C, Coles CP, Zalavras C, Goldfarb CA, Cassidy C, Allan C, Beingessner D, Kalainov DM, Eygendaal D, Sancheti P, Feibel RJ, Rocha S, Grosso E, Frihagen F, Dyer GSM, Athwal GS, Goslings JC, Della Rocca GJ, Harris I, Fanuele JC, Lawton J, Jiuliano J, McAuliffe J, Capo JT, Conflitti JM, Segalman K, Egol K, Ponsen KJ, Jeray K, Lattanza L, Catalano L 3rd, Swiontkowski M, Boyer M, Richardson M, Soong M, Baskies M, Prayson M, Mckee M, Chen NC, van Eerten PV, Brink PRG, Evans PJ, Jebson P, Kloen P, Blazar P, Rhemrev SJ, Page RS, Papandrea R, Nelissen RG, Zura RD, Pesantez R, Poolman RW, Sodha S, Duncan S, Wolfe S, Gosens T, Wright T, and Davis TR.

Introduction

Diagnosis of displacement is a critical factor of the management of scaphoid fractures. Displacement is one of the factors most strongly associated with a greater risk of nonunion in scaphoid waist fractures¹³. Although scaphoid waist fractures have a 90% to 95% union rate overall, fractures with greater than 1 mm displacement are associated with high rates of nonunion, as much as 55%^{10,5}. Open reduction and internal fixation therefore should be considered for these fractures.

Three previous methodologic studies showed poor to moderate levels of interobserver reliability for displacement of scaphoid fractures^{5,12,15}. Combined radiographs and CT had a higher level of agreement than each modality alone¹⁵; however, the interobserver reliability was only moderate ($\kappa = 0.48$; 95% confidence interval, $\kappa = 0.37-0.57$). The fact that observers disagree on the diagnosis of displacement even with sophisticated imaging suggests observers interpret images based on factors other than the quality of the anatomic depiction. Interobserver variability is a clinically important issue as it represents the human aspect or art of medicine, the unwarranted variation of clinical observations. It should be possible to limit interobserver variability caused by observer error (eg, lack of discernment) and/or observer bias (eg, different/false preconceptions) by improved definitions, diagnostic thresholds, training, and experience. Prior studies in various medical fields have shown improved interobserver agreement with formal training of observers^{3,4,7,16-18,21}.

The results from a pilot study among eight orthopaedic surgeons and residents at our institution suggested training could improve interrater agreement for scaphoid fracture classification, but the study was underpowered to show significant differences between pretraining and posttraining. We therefore tested whether training can improve interobserver reliability for the diagnosis of scaphoid fracture displacement on radiographs and CT scans in a larger group of orthopaedic surgeons. Second, we tested whether the sensitivity, specificity, and accuracy of these diagnostic tests could be improved by training.

Materials and Methods

We invited independent observers (all orthopaedic and trauma surgeons) from several countries to evaluate 20 cases of scaphoid fractures in an online survey. We randomly assigned them on a 1:1 basis to a short training session or no training session before their ratings. The study was performed under a protocol approved by the Institutional Research Board at the principal investigator's (DR) hospital.

This study was part of a nascent collaborative called the Science of Variation Group (SOVG). The objectives of the collaborative are to study variation in the definition,



interpretation, and classification of injury and disease. The SOVG has created a web-based platform (www.scienceofvariationgroup.org, Amsterdam, The Netherlands) that facilitates large international interobserver studies. With multiple fully trained surgeons from diverse countries and institutions participating in studies, this approach should provide a powerful forum for studying, understanding, and ultimately reducing interobserver variation in aspects of patient care.

The training group received an online training module before the rating session, and the nontraining group did not. The module instructed observers in definitions and rating techniques. The training group was instructed that for purposes of this study no measurements should be taken on any of the radiographic images. Displacement was defined as any gapping, angulation, or translation of the fracture (regardless of comminution)—anything more than a crack. Instead of measurements, the module provided instructions for a definition of displacement based on CT scans. For each type of fracture displacement, image examples were provided to illustrate a displaced fracture and a nondisplaced fracture (Fig. 1).

For the rating session, a set of combined radiographs and CT scans of 10 displaced and 10 apparently nondisplaced scaphoid fractures was selected. Radiographs consisted of a set of four views: posteroanterior with the wrist in neutral position and ulnar deviation, lateral, and oblique. The CT scans consisted of reconstructions, presented as movies, in the coronal and sagittal longitudinal planes of the scaphoid, as described by Sanders²².

For purposes of this study, observers were asked to give their overall judgment regarding fracture displacement based on the combined radiographs and CT scans of each patient. The classification was simplified to displaced or nondisplaced, thus excluding the possibility of minimal displacement.

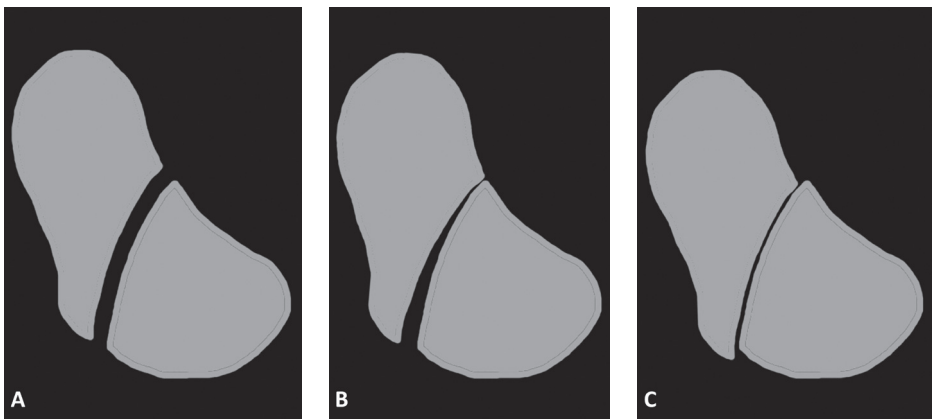


Figure 1A–C Diagrams from the training module illustrate (A) gapping, (B) angulation, and (C) translation in a displaced fracture.

For the selection of the 10 displaced and 10 apparently nondisplaced fractures, we based the reference standard of displacement on a consensus of interpretations between the treating surgeon, a musculoskeletal radiologist, and an independent physician. All displaced fractures were internally fixed with a screw, and displacement was confirmed intraoperatively by arthroscopy or a miniincision. Fractures diagnosed by consensus as nondisplaced were not confirmed intraoperatively, therefore, we speak of apparently nondisplaced fractures. These fractures were treated by either cast immobilization or screw fixation.

The multirater kappa (κ) measure was used to estimate agreement among training and nontraining surgeons. It is a commonly used statistic to describe chance-corrected agreement in various intraobserver and interobserver studies^{8,14,19}. Agreement among observers was calculated with use of the multirater κ measure described by Siegel and Castellan²³. The κ values were interpreted using the guidelines proposed by Landis and Koch¹⁴: values of 0.01 to 0.20 indicate slight agreement; 0.21 to 0.40 fair agreement; 0.41 to 0.60 moderate agreement; 0.61 to 0.80 substantial agreement; and greater than 0.81 almost perfect agreement. Zero indicates no agreement beyond that expected attributable to chance alone, -1.00 indicates total disagreement, and $+1.00$ indicates perfect agreement^{14,19}. The agreement among the training group was compared with the nontraining group using Z-tests, which assume the two samples are independent. As the samples compared in this study were not independent (the same set of patients were rated by both groups), this method produced conservative estimates of the p values. The study had 80% power to detect a significant difference in a multirater κ of 0.1 between observers randomized to training or no training.

For each group, sensitivity, specificity, and accuracy were calculated using standard formulas with 95% CI. The positive and negative predictive values were determined with use of Bayes' theorem, with an a priori estimate of the prevalence (pretest probability) of displacement of scaphoid fractures set at 5%.

Results

Sixty-four fully trained orthopaedic and trauma surgeons who practice in various parts of the world participated in this observer study through a recently developed online forum for the Science of Variation (www.scienceofvariationgroup.org). The total group of observers consisted of 57 male and seven female attending orthopaedic and trauma surgeons from multiple countries, with the majority practicing in the United States (58%). The group consisted for the most part of surgeons with at least 6 years of independent practice (79%), who supervise trainees (95%), and treat at least 10 scaphoid fractures on a yearly basis (63%). Most surgeons specialized in either hand surgery (38%) or orthopaedic traumatology (36%). Before rating, the observers were



randomized to a training group (34 observers) and a nontraining group (30 observers) (Table 1).

The κ values representing the interobserver reliability for scaphoid fracture displacement were 0.60 (95% CI, 0.58–0.61) for the training group and 0.52 (95% CI, 0.50–0.54)

Table 1. Demographics of the two rating groups.						
	Non-Training Group (n=30)	%	Training Group (n=34)	%	Total (n=64)	%
Gender						
Male	27	90	30	88	57	89
Female	3	10	4	12	7	11
Country of practice						
Asia	2	7	2	6	4	6
Australia	2	7	1	3	3	5
Canada	2	7	2	6	4	6
Europe	6	20	7	21	13	20
United Kingdom	1	3	0	0	1	2
Unites States	17	56	20	59	37	58
Other	0	0	2	6	2	3
Years of independent practice						
0-5	8	27	5	15	13	21
6-10	8	27	10	29	18	28
11-20	11	36	16	47	27	42
21-30	3	10	3	9	6	9
Trainee supervision						
Yes	30	100	31	91	61	95
No	0	0	3	9	3	5
Nr of treated scaphoid fractures per year						
0-5	9	30	7	20	16	25
6-10	4	13	4	12	8	12
11-20	9	30	18	53	27	43
>20	8	27	5	15	13	20
Specialization						
General orthopaedics	0	0	1	3	1	2
Orthopaedic traumatology	12	40	11	32	23	36
Shoulder and elbow	6	20	6	18	12	18
Hand and wrist	10	33	14	41	24	38
Other	2	7	2	6	4	6

for the nontraining group, a difference ($p < 0.001$) in favor of the training group. According to the benchmark scores of Landis and Koch¹⁴, rating of displacement on radiographs and CT scans combined resulted in moderate agreement for both groups. The average sensitivity of diagnosing displacement of scaphoid fractures was 83% (95% CI, 52%–100%) for the nontraining group and 87% (95% CI, 60%–100%) for the training group. The average specificity was 85% (95% CI, 51%–100%) for the nontraining group and 86% (95% CI, 64%–100%) for the training group. The average accuracy was 84% (95% CI, 65%–100%) for the nontraining group and 87% (95% CI, 72%–100%) for the training group. According to Bayes' theorem, the positive predictive values were 0.23 in the nontraining group and 0.25 in the training group. The negative predictive value was 0.99 in both groups.

Discussion

The diagnosis of displacement in scaphoid fractures is notorious for poor interobserver reliability. We therefore tested whether training can improve interobserver reliability for the diagnosis of scaphoid fracture displacement on radiographs and CT scans.

Our study had several limitations. First, the apparently nondisplaced fractures were not arthroscopically confirmed, thus leaving the possibility that some fractures were misdiagnosed as being nondisplaced, although this is unlikely given the relative infrequency of displaced fractures. Second, we included an even amount of displaced and nondisplaced fractures, which may introduce context or spectrum bias because the incidence of displaced fractures is somewhat lower in real practice, as much as 30% of fractures¹⁰. No normal or nonfractured scaphoid cases were included and observers were informed that all cases were fractures. Finally, the test conditions may create less realistic circumstances such that participants may be more likely to err on the side of caution.

This experiment determined a training session before rating can improve interobserver reliability for the diagnosis of scaphoid fracture displacement using radiographs and CT scans; however, the degree of agreement was categorized as moderate in both groups, and the improvement was slight. All diagnostic performance characteristics (eg, accuracy) were greater in the training group, but the difference was not substantial and we do not have a method to test for statistical significance in the differences. Because of the low prevalence of displacement, the positive predictive value is relatively low and the negative predictive value very high. Thus, regardless of training, these two modalities combined are good for ruling out displacement but not for ruling it in.

In 1980, Cooney et al.⁹ defined displacement (or instability) of a scaphoid fracture as (1) a fracture gap larger than 1 mm on any radiographic projection, (2) a scapholunate



angle larger than 60°, or (3) a radiolunate angle larger than 15°. These criteria have been widely applied since then. Amadio et al.¹ added the criterion that the intrascaphoid angle should not exceed 35°. Previous work has shown the radiolunate¹⁵ and intrascaphoid^{2,20} angles to be unreliable.

Interobserver variability is affected by multiple factors, including errors of discernment, classification, recording, and observer bias. Although variation and errors in discernment might be addressed with more sophisticated imaging, studies consistently show more sophisticated imaging improved intraobserver variation more than interobserver variation^{6,15,24}. The other types of observer error and bias may be reduced by clear consensus definitions, measurement techniques, and classification/diagnostic criteria, and experience and training.

In both of our groups, there were comparable numbers of observers who tended to overdiagnose displacement (increasing sensitivity and decreasing specificity) or underdiagnose displacement (decreasing sensitivity and increasing specificity), which balanced out into an accuracy that almost equals the exact average between sensitivity and specificity. Given the potential consequences of inadequate treatment of a displaced fracture, underdiagnosing is more of a concern than overdiagnosing. Therefore, we attempted to simplify the definition in the training group by defining displacement as anything more than a crack.

It is hoped training can improve interobserver agreement by reducing observer error and bias. It is of interest to understand why observers make ratings in variable ways and how training can reduce this. Training could be considered a way to counteract the tendency of the mind to rate everything intuitively without consistently following a guideline, which is a potential form of observer bias. This study shows us, even though observers are encouraged and trained to classify fractures using clear and simplified guidelines, many observers still see things that others do not see, regardless of their experience. Another possible method to address this is by a joint discussion session to discuss cases and achieve consensus where possible. Such efforts have shown great improvement in interrater agreement in histopathologic grading among experienced pathologists¹¹.

Our results suggest training can improve interobserver agreement and sensitivity, specificity, and accuracy for the diagnosis of scaphoid fracture displacement, but the improvements are slight. These findings are encouraging for future research, which is needed to improve our understanding of interobserver variation and how to reduce it further.

References

1. Amadio PC, Berquist TH, Smith DK, Ilstrup DM, Cooney WP III, Linscheid RL. Scaphoid malunion. *J Hand Surg Am.* 1989; 14:679–687.
2. 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;23:76–81.
3. Bankier AA, Fleischmann D, De Maertelaer V, Kontrus M, Zontsich T, Hittmair K, Mallek R. Subjective differentiation of normal and pathological bronchi on thin-section CT: impact of observer training. *Eur Respir J.* 1999;13:781–786.
4. Berg WA, D’Orsi CJ, Jackson VP, Bassett LW, Beam CA, Lewis RS, Crewson PE. Does training in the Breast Imaging Reporting and Data System (BI-RADS) improve biopsy recommendations or feature analysis agreement with experienced breast imagers at mammography? *Radiology.* 2002;224:871–880.
5. Bernard SA, Murray PM, Heckman MG. Validity of conventional radiography in determining scaphoid waist fracture displacement. *J Orthop Trauma.* 2010;24:448–451.
6. Bernstein J, Adler LM, Blank JE, Dalsey RM, Williams GR, Iannotti JP. Evaluation of the Neer system of classification of proximal humeral fractures with computerized tomographic scans and plain radiographs. *J Bone Joint Surg Am.* 1996;78: 1371–1375.
7. Brorson S, Bagger J, Sylvest A, Hrobjartsson A. Improved interobserver variation after training of doctors in the Neer system: a randomised trial. *J Bone Joint Surg Br.* 2002;84:950–954.
8. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20:37–46.
9. Cooney WP, Dobyns JH, Linscheid RL. Fractures of the scaphoid: a rational approach to management. *Clin Orthop Relat Res.* 1980;149:90–97.
10. Dabezies EJ, Mathews R, Faust DC. Injuries to the carpus: fractures of the scaphoid. *Orthopedics.* 1982;5:1510–1515.
11. de Vet HC, Koudstaal J, Kwee WS, Willebrand D, Arends JW. Efforts to improve interobserver agreement in histopathological grading. *J Clin Epidemiol.* 1995;48:869–873.
12. Desai VV, Davis TR, Barton NJ. The prognostic value and reproducibility of the radiological features of the fractured scaphoid. *J Hand Surg Br.* 1999;24:586–590.
13. Eddeland A, Eiken O, Hellgren E, Ohlsson NM. Fractures of the scaphoid. *Scand J Plast Reconstr Surg.* 1975;9:234–239.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174.
15. 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.
16. Lujan ME, Chizen DR, Peppin AK, Kriegler S, Leswick DA, Bloski TG, Pierson RA. Improving interobserver variability in the evaluation of ultrasonographic features of polycystic ovaries. *Reprod Biol Endocrinol.* 2008;6:30.
17. Magnan MA, Maklebust J. The effect of Web-based Braden Scale training on the reliability of Braden subscale ratings. *J Wound Ostomy Continence Nurs.* 2009;36:51–59.
18. Patel AB, Amin A, Sortey SZ, Athawale A, Kulkarni H. Impact of training on observer variation in chest radiographs of children with severe pneumonia. *Indian Pediatr.* 2007;44:675–681.
19. Posner KL, Sampson PD, Caplan RA, Ward RJ, Cheney FW. Measuring interrater reliability among



- multiple raters: an example of methods for nominal data. *Stat Med.* 1990;9:1103–1115.
20. Ring D, Patterson JD, Levitz S, Wang C, Jupiter JB. Both scanning plane and observer affect measurements of scaphoid deformity. *J Hand Surg Am.* 2005;30:696–701.
 21. Ripsweden J, Mir-Akbari H, Brolin EB, Brismar T, Nilsson T, Rasmussen E, Ruck A, Svensson A, Werner C, Winter R, Cederlund K. Is training essential for interpreting cardiac computed tomography? *Acta Radiol.* 2009;50:194–200.
 22. 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.
 23. Siegel S, Castellan NJ. *Nonparametric Statistics for the Behavioral Sciences.* 2nd ed. New York, NY: McGraw-Hill; 1988.
 24. Stieber J, Quirno M, Cunningham M, Errico TJ, Bendo JA. The reliability of computed tomography and magnetic resonance imaging grading of lumbar facet arthropathy in total disc replacement patients. *Spine (Phila Pa 1976).* 2009;34:E833–840.
 25. Szabo RM, Manske D. Displaced fractures of the scaphoid. *Clin Orthop Relat Res.* 1988;230:30–38.