

What Middle Phalanx Base Fracture Characteristics are Most Reliable and Useful for Surgical Decision-making?

Stein J. Janssen MD, Jeroen Molleman BSc, Thierry G. Guitton MD, PhD,
David Ring MD, PhD, on behalf of the Science Of Variation Group

Received: 19 February 2015 / Accepted: 1 June 2015 / Published online: 18 June 2015
© The Association of Bone and Joint Surgeons® 2015

Abstract

Background Fracture-dislocations of the proximal interphalangeal joint are vexing because subluxation and articular damage can lead to arthrosis and the treatments are imperfect. Ideally, a surgeon could advise a patient, based on radiographs, when the risk of problems merits operative intervention, but it is unclear if middle phalanx base fracture characteristics are sufficiently reliable to be useful for surgical decision making.

Questions/purposes We evaluated (1) the degree of interobserver agreement as a function of fracture

characteristics, (2) the differences in interobserver agreement between experienced and less-experienced hand surgeons, and (3) what fracture characteristics and surgeon characteristics were associated with the decision for operative treatment.

Methods Ninety-nine (33%) of 296 hand surgeons evaluated 21 intraarticular middle phalanx base fractures on lateral radiographs. Eighty-one surgeons (82%) were in academic practice and 57 (58%) had less than 10 years experience. Participants assessed six fracture characteristics and recommended treatment (nonoperative or operative;

Science Of Variation Group: J. Adams; A. Ilyas; A. Marcus; A. Schefer; A. L. Terrono; W. Baer; F. L. Walter; M. Boyer; B. P. D. Wills; R. P. Calfee; C. H. Fernandes; C. Swigart; C. Cassidy; C. D. Miranda; C. L. Moreno-Serrano; C. M. Jones; D. M. McKee; D. M. Kalainov; D. Polatsch; C. J. R. Barreto; R. Nyszkiewicz; S. Sodha; D. Ruchelsman; E. Bonatz; E. T. Walbeehm; E. D. Schumer; F. Suarez; G. DeSilva; R. G. Gaston; G. R. Hernandez; M. Nancollas; E. Hofmeister; J. C. Fanuele; J. G. Boretto; J. Choueka; J. I. Huang; J. M. Wolf; J. M. Erickson; J. Howlett; J. Taras; J. Wint; S. Kakar; K. Chivers; K. J. Prommersberger; K. Lee; K. M. Rumball; G. A. Kraan; L. Weiss; L. Lattanza; L. A. de Mendonça Jr; L. Paz; L. Catalano III; M. Calcagni; J. McAuliffe; M. Darowish; R. M. Costanzo; M. M. Wood; M. Baskies; M. Behrman; M. W. Kessler; M. J. Palmer; M. W. Grafe; D. L. Nelson; N. M. Akabudike; J. T. Capo; O. M. Semenkin; F. E. Jaime; B. A. Palmer; P. A. Martineau; P. Blazar; P. Hahn; P. Muhl; P. Jebson; P. W. Owens; R. de Bedout; R. Shatford; R. M. Fricker; R. R. L. Gray; R. Papandrea; R. S. Gilbert; M. Rizzo; R. L. Hutchison; C. Rodner; R. Klinefelter; S. Spruijt; S. A. Kennedy; S. Beldner; S. Mitchell; S. Dodds; S. Hilliard; S. M. Jacoby; S. Kaplan; S. Moghtaderi; T. Baxamusa; F. T. D. Kaplan; T. Siff; T. Apar; T. F. Varecka; W. C. Hammert.

One of the authors (DR) certifies that he, or a member of his immediate family, has or may receive payments or benefits, during the study period, an amount less than USD 10,000 from Wright Medical (Memphis, TN, USA); an amount less than USD 10,000 from Skeletal Dynamics (Miami, FL, USA); an amount less than USD 10,000 from Biomet (Warsaw, IN, USA); an amount less than USD 10,000 from AO North America (Paoli, PA, USA); and an amount

less than USD 10,000 from AO International (Dubendorf, Switzerland).

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

Clinical Orthopaedics and Related Research® neither advocates nor endorses the use of any treatment, drug, or device. Readers are encouraged to always seek additional information, including FDA-approval status, of any drug or device prior to clinical use.

Each author certifies that his or her institution approved the human protocol for this investigation, that all investigations were conducted in conformity with ethical principles of research, and that informed consent for participation in the study was obtained.

Electronic supplementary material The online version of this article (doi:10.1007/s11999-015-4394-7) contains supplementary material, which is available to authorized users.

S. J. Janssen, J. Molleman, D. Ring (✉)
Hand and Upper Extremity Service, Department of Orthopaedic Surgery, Yawkey Center, Massachusetts General Hospital, Suite 2100, 55 Fruit Street, Boston 02114, MA, USA
e-mail: dring@partners.org

T. G. Guitton
Department of Orthopaedic Surgery, Academic Medical Center
Amsterdam, Amsterdam, The Netherlands

extension block pinning, external fixation, open reduction and internal fixation, volar plate arthroplasty, or hemihamate autograft arthroplasty) for all cases.

Results With all surgeons pooled together, the interobserver agreement for fracture characteristics was substantial for assessment of a 2-mm articular step or gap (κ , 0.73; 95% CI, 0.60–0.86; $p < 0.001$), subluxation or dislocation (κ , 0.72; 95% CI, 0.58–0.86; $p < 0.001$), and percentage of articular surface involved (intraclass correlation coefficient [ICC], 0.67; 95% CI, 0.54–0.81; $p < 0.001$); moderate for comminution (κ , 0.55; 95% CI, 0.39–0.70; $p < 0.001$) and stability (κ , 0.54; 95% CI, 0.39–0.69; $p < 0.001$); and fair for the number of fracture fragments (ICC, 0.39; 95% CI, 0.27–0.57; $p < 0.001$). When recommending treatment, interobserver agreement was substantial (κ , 0.69; 95% CI, 0.50–0.88; $p < 0.001$) for the recommendation to operate or not to operate, but only fair (κ , 0.34; 95% CI, 0.21–0.47; $p < 0.001$) for the specific type of treatment, indicating variation in operative techniques. There were no differences in agreement for any of the fracture characteristics or treatment preference between less-experienced and more-experienced surgeons, although statistical power on this comparison was low. None of the surgeon characteristics was associated with the decision for operative treatment, whereas all fracture characteristics were, except for stable and uncertain joint stability. Articular step or gap (β , 0.90; R-squared, 0.89; 95% CI, 0.75–1.05; $p < 0.001$), likelihood of subluxation or dislocation (β , 0.80; R-squared, 0.76; 95% CI, 0.59–1.02; $p < 0.001$), and unstable fractures (β , 0.88; R-squared, 0.81; 95% CI, 0.67–1.1; $p < 0.001$), are most strongly associated with the decision for operative treatment.

Conclusions We found that assessment of a step or gap and likelihood of subluxation were most reliable and are strongly associated with the decision for operative treatment. Surgeons largely agree on which fractures might benefit from surgery, and the variation seems to be with the operative technique. Efforts at improving the care of these fractures should focus on the comparative effectiveness of the various operative treatment options.

Level of Evidence Level III, diagnostic study.

Introduction

Fracture-dislocations of the proximal interphalangeal joint can be problematic as they can result in stiffness, pain, and secondary osteoarthritis [7, 18, 21]. Characteristics of the articular fracture of the middle phalanx that are thought to indicate benefit of operative treatment are subluxation of the joint, greater than 2-mm articular stepoff, and more than 40% of articular surface involvement [13, 18, 21]. These fracture characteristics also are used in selecting a specific

operative treatment with options extending from extension block pinning to hemihamate autograft arthroplasty [13, 17, 18, 21]. However, it is unclear if middle phalanx base fracture characteristics can be assessed with sufficient reliability to be useful for surgical decision making.

Validation of a specific criterion for recommending operative treatment includes evidence that: (1) patients do better with surgery when that criterion is met; (2) the criterion accurately measures pathophysiologic features; and (3) measures of the criterion are reliable. Putting aside whether the criteria for recommending operative treatment of a proximal interphalangeal joint fracture-dislocation are useful and accurate thresholds, if observers cannot agree on the criteria (unreliable) they may not be useful. Studies of fractures at other anatomic sites (such as the humerus, clavicle, and olecranon) showed that some fracture characteristics are assessed more reliably than others [1, 2, 5, 8, 10]. In addition to measuring the reliability of radiographic criteria for recommending operative treatment, better understanding of the criteria that have the greatest influence on recommendation for treatment of middle phalanx base fractures would help direct the development of guidelines and measurement techniques for those criteria.

We assessed interobserver agreement of assessment of middle phalanx base fracture characteristics and resulting treatment recommendations by hand surgeons. Specifically, we evaluated: (1) the degree of interobserver agreement as a function of fracture characteristics, (2) differences in interobserver agreement between experienced and less-experienced hand surgeons, and (3) what fracture characteristics and surgeon characteristics were associated with the decision for operative treatment.

Materials and Methods

This study was approved by our institutional review board and a waiver of informed consent was obtained. With the use of International Classification of Diseases, 9th Revision codes (code 816.01: closed middle or proximal phalanx fracture), patients who received a diagnosis of a phalangeal fracture at a tertiary hand clinic between July 2011 and August 2013 were identified ($n = 315$). There were 21 (6.7%) acute intraarticular middle phalanx base fractures assessed with a true lateral radiograph (assessed by looking at projection of the proximal phalanx condyles) (Appendix 1. Supplemental material is available with the online version of CORR®).

We used SurveyMonkey (Palo Alto, CA, USA), an online tool for development of a survey including the 21 cases. For all cases, the following seven questions were asked in the following order: (1) What percentage of the articular surface is fractured? (2) Is there an articular

step or gap greater than 2 mm? (3) Is there comminution or fragmentation of the articular surface? (4) How many fracture fragments? (5) Is the proximal interphalangeal joint subluxated or dislocated? (6) How would you classify this injury based on imaging: stable, tenuous, or unstable? (7) What is your proposed treatment: nonoperative, extension block pinning, external fixation, open reduction and internal fixation, volar plate arthroplasty, or hemihamate autograft arthroplasty? The answer to Question 7 then was dichotomized into nonoperative and operative treatment and was analyzed separately.

An invitation to participate was sent to a list of surgeons built as part of the Science Of Variation Group (SOVG) [11, 12]. The SOVG consists of 691 orthopaedic, trauma, and plastic surgeons, all with an interest in treating upper extremity conditions. We approached only hand surgeons ($n = 296$); 108 (36%) responded and 99 completed (33%) all seven questions for all 21 cases. The SOVG aims to study variation in the definition and treatment of human illness without financial incentive.

Surgeon Characteristics

Of these 99 hand surgeons, there were 90 men (91%). The majority of hand surgeons were from the United States (80%). There was substantial variability in the number of years the surgeons were in practice; 57 (58%) had less than 10 years in practice. Eighty-one (82%) hand surgeons were involved in supervising trainees (Table 1).

Statistical Analysis

We used Fleiss' kappa analysis to assess the interobserver agreement for the questions with categorical answers (Questions 2, 3, 5, 6, and 7). Bootstrapping (number of

Table 1. Surgeon characteristics ($n = 99$)

Characteristic	Number (%)
Male sex	90 (91)
Location of practice	
United States and Canada	79 (80)
Europe	12 (12)
Other	8 (8.1)
Years in practice	
0–5	36 (36)
6–10	21 (21)
11–20	30 (30)
20–30	12 (12)
Supervising trainees	81 (82)

resamples = 1000) was used to calculate a standard error, z statistic, p value, and 95% CI for the kappa values [16]. Kappa is a quantitative measure of agreement among observers and takes into account that observers sometimes will choose the same answer to a question by chance [23]. A perfect agreement among observers would be reflected as a kappa of 1, whereas agreement totally based on chance would equate to a kappa of 0. Interpretation of kappa often is done by a classification proposed by Landis and Koch [15] in which a kappa between 0.01 and 0.20 is considered to reflect slight agreement, a value between 0.21 and 0.40 as fair agreement, between 0.41 to 0.60 as moderate agreement, 0.61 and 0.80 as substantial agreement, and greater than 0.81 as almost perfect agreement.

Intraclass correlation coefficient (ICC) with 95% CI was calculated through a two-way mixed-effects model with absolute agreement and shows interobserver agreement for questions with a continuous answer (Questions 1 and 4). To calculate an ICC, the variability of ratings per subject is compared with the total possible variability in ratings. Absolute agreement in an ICC assesses how much each measurement performed per observer differs from the other observers. As with kappa, a score of 1 reflects perfect agreement in ICC, whereas 0 reflects no agreement. Fisher's z transformation was used to calculate p values comparing ICCs in the experienced and less-experienced groups [3, 4].

We compared the kappa values and ICCs of experienced (≥ 10 years of practice) and less-experienced (< 10 years of experience) hand surgeons for every question.

We used multivariable linear regression analysis to assess if surgeon characteristics (sex, location of practice, years in practice, or supervising trainees) were independently associated with the likelihood of choosing surgery per surgeon (99 surgeons). We calculated an overall surgery score per surgeon by dividing the amount of cases they would operate on by 21 (the total number of cases). The score ranges from 0 to 1 with a higher score indicating a higher likelihood of choosing surgery. Linear regression analysis provides a β regression coefficient that indicates the difference in surgery score (likelihood for surgery) in one group compared with another.

Simple linear regression was used to assess which fracture characteristics influenced the recommendation for operative treatment per case (21 cases). We calculated an overall surgery score per case by dividing the number of surgeons who would operate by 99 (the total number of surgeons) and assessed the association of fracture characteristics with the surgery score. We used simple linear regression because fracture characteristics were highly collinear resulting in an unstable multivariable model. The R-squared indicates how much of the

variation in surgery score per case is explained by the specific fracture characteristic.

A two-tailed *p* value less than 0.05 was considered significant; all statistical analyses were performed using Stata® 14 (StataCorp LP, College Station, TX, USA).

Results

Interobserver Reliability for Fracture Characteristics

When all surgeons are pooled together, the interobserver agreement for fracture characteristics was greatest for assessment of articular stepoff or gap of 2 mm or greater (kappa, 0.73; 95% CI, 0.60–0.86, *p* < 0.001) and lowest for the number of fracture fragments (ICC, 0.39; 95% CI, 0.27–0.57, *p* < 0.001) (Fig. 1). There was substantial agreement on the percentage of articular surface area that was involved in the fracture (ICC, 0.67; 95% CI, 0.54–0.81, *p* < 0.001) (Table 2).

Interobserver Reliability Based on Surgeon Experience

We found no difference, with the numbers available, in interobserver agreement for any of the fracture characteristics between less-experienced and more-experienced hand surgeons (Table 3).

Factors Associated with Recommendation of Surgery

We found that all fracture characteristics, except for stable and uncertain joint stability, were associated with a recommendation for operative treatment. Articular stepoff or gap greater than 2 mm was most strongly associated with a recommendation for operative treatment (β , 0.90; R-squared, 0.89; 95% CI, 0.75–1.05; *p* < 0.001) (Table 4).

After controlling for relevant confounding variables, we found that male sex (β regression coefficient, -0.019; 95% CI, -0.082 to 0.044; *p* = 0.55), more experience (β , 0.0035; 95% CI, -0.033 to 0.040; *p* = 0.85), and supervising trainees (β , -0.028; 95% CI, -0.075 to 0.020; *p* = 0.26) were not associated with the recommendation for operative treatment (Table 5). There also was no difference in recommendation for surgery among practice locations: Europe (β , -0.018; 95% CI, -0.074 to 0.038; *p* = 0.53) and Other (β , 0.046; 95% CI, -0.021 to 0.11; *p* = 0.18) compared with the United States and Canada (Table 5).

Interobserver agreement was fair (kappa, 0.34; 95% CI, 0.21–0.47, *p* < 0.001) for the type of treatment recommended but substantial (kappa, 0.69; 95% CI, 0.50–0.88, *p* < 0.001) for the recommendation to operate or not (Table 2). The fair agreement for type of treatment indicates considerable variation in decision for specific treatment techniques among included

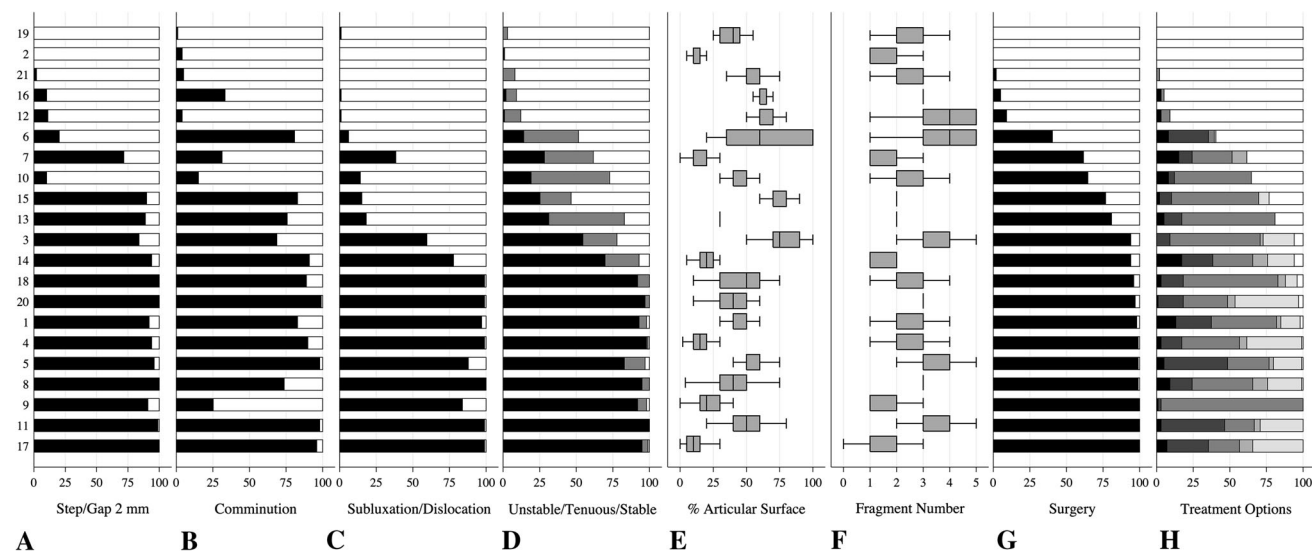


Fig. 1A–H The bar graphs show the percentage of surgeons indicating (A) 2-mm step or gap (black bar); (B) comminution (black bar); (C) subluxation or dislocation (black bar); (D) Unstable (black bar), tenuous (grey bar) or stable (white bar) joint per case (case number on y-axis). The box plots show the median percentage of (E) fractured articular surface and (F) number of fracture fragments, with the interquartile range and range per case. (G) The bar graph shows the percentage of surgeons recommending operative treatment (black bar)

per case. Cases (Appendix 1) are ordered from no surgeons recommending surgery (Cases 19 and 2) to all surgeons recommending surgery (Cases 9, 11, and 17). (H) The graph shows the variation in recommended treatment options: nonoperative (white bar), hemi-hamate autograft arthroplasty (light gray bar), volar plate arthroplasty (medium light gray bar), open reduction and internal fixation (medium dark gray bar), external fixation (dark gray bar), and extension block pinning (black bar).

Table 2. Interobserver agreement among all hand surgeons

Fracture characteristics and recommendation for treatment	Kappa (95% CI)	p value
Is there an articular step or gap greater than 2 mm: yes or no?	0.73 (0.60–0.86)	< 0.001
Is there comminution/fragmentation of the articular surface: yes or no?	0.55 (0.39–0.70)	< 0.001
Is the proximal interphalangeal joint subluxated or dislocated: yes or no?	0.72 (0.58–0.86)	< 0.001
How would you classify this injury based on imaging: stable, tenuous, or unstable?	0.54 (0.39–0.69)	< 0.001
What is your proposed treatment: nonoperative, extension block pinning, external fixation, open reduction and internal fixation, volar plate arthroplasty, or hemihamate autograft arthroplasty?	0.34 (0.21–0.47)	< 0.001
What is your proposed treatment: operative or nonoperative?	0.69 (0.50–0.88)	< 0.001
	Intraclass correlation coefficient* (95% CI)	
What percentage of the articular surface is involved in the fracture?	0.67 (0.54–0.81)	< 0.001
How many fracture fragments are there?	0.39 (0.27–0.57)	< 0.001

* Intraclass correlation coefficient (two-way mixed-effects model with absolute agreement).

Table 3. Comparison of interobserver agreement between less experienced and experienced hand surgeons

Fracture characteristics and recommendation for treatment	Less experienced surgeons (< 10 years experience) kappa (95% CI)	Experienced surgeons (> 10 years experience) kappa (95% CI)	p value
Is there an articular step or gap greater than 2 mm: yes or no?	0.73 (0.60–0.86)	0.72 (0.59–0.85)	0.88
Is there comminution/fragmentation of the articular surface: yes or no?	0.52 (0.37–0.67)	0.58 (0.41–0.76)	0.60
Is the proximal interphalangeal joint subluxated or dislocated: yes or no?	0.72 (0.58–0.86)	0.72 (0.57–0.87)	0.99
How would you classify this injury based on imaging: stable, tenuous, or unstable?	0.55 (0.40–0.70)	0.52 (0.36–0.69)	0.80
What is your proposed treatment: nonoperative, extension block pinning, external fixation, open reduction and internal fixation, volar plate arthroplasty, or hemihamate autograft arthroplasty?	0.35 (0.22–0.48)	0.33 (0.19–0.46)	0.79
What is your proposed treatment: operative or nonoperative?	0.70 (0.51–0.90)	0.67 (0.48–0.86)	0.80
	Intraclass correlation coefficient* (95% CI)	Intraclass correlation coefficient* (95% CI)	
What percentage of the articular surface is fractured?	0.67 (0.54–0.81)	0.66 (0.53–0.81)	0.93
How many fracture fragments?	0.37 (0.25–0.55)	0.43 (0.30–0.61)	0.73

* Intraclass correlation coefficient (two-way mixed effects model with absolute agreement).

surgeons (Fig. 1). We found no difference in interobserver agreement when deciding for specific strategies ($p = 0.79$) or the decision to operate ($p = 0.80$) between the less-experienced and experienced hand surgeons with the numbers evaluated (Table 3).

Discussion

Several middle phalanx base fracture characteristics are commonly used in selecting treatment options. However, it

is unclear if these characteristics can be assessed with sufficient reliability to be useful for surgical decision-making and what factors are most strongly associated with the decision for surgery. We assessed interobserver agreement among 99 hand surgeons in assessing morphologic features of fractures and decision for treatment on lateral radiographs of 21 intraarticular middle phalanx base fractures. We found that: (1) interobserver agreement was greatest for assessment of articular step or gap, likelihood of subluxation or dislocation, and percentage articular surface involvement; (2) there were no differences in

Table 4. Simple linear regression of fracture characteristics associated with decision for surgery per case

Fracture characteristics	β regression coefficient (95% CI)	Standard error	R-squared	p value
Percentage of articular surface involvement	0.015 (0.0090–0.021)	0.0030	0.58	< 0.001
Articular step or gap greater than 2 mm	0.90 (0.75–1.05)	0.073	0.89	< 0.001
Is there comminution or fragmentation	0.86 (0.56–1.2)	0.15	0.65	< 0.001
Number of fracture fragments	0.33 (0.14–0.52)	0.092	0.41	0.002
Proximal interphalangeal joint subluxated or dislocated	0.80 (0.59–1.02)	0.10	0.76	< 0.001
Fracture classification				< 0.001
Stable	Reference group	Reference group		Reference group
Tenuous	0.072 (–1.1 to 1.2)	0.56	0.001	0.90
Unstable	0.88 (0.67–1.1)	0.097	0.81	< 0.001

Table 5. Multivariable linear regression of surgeon characteristics associated with decision for surgery

Surgeon characteristics	β regression coefficient (95% CI)	Standard error	p value
Male sex (reference group: female sex)	–0.019 (–0.082 to 0.044)	0.032	0.55
Location of practice			
United States and Canada	Reference group	Reference group	Reference group
Europe	–0.018 (–0.074 to 0.038)	0.028	0.53
Other	0.046 (–0.021 to 0.11)	0.034	0.18
Experienced surgeons (reference group: less experienced surgeons)	0.0035 (–0.033 to 0.040)	0.018	0.85
Supervising trainees (reference group: not supervising trainees)	–0.028 (–0.075 to 0.020)	0.024	0.26

agreement between experienced and less-experienced surgeons; (3) articular step or gap, likelihood of subluxation or dislocation, and unstable fractures are most strongly associated with the decision for operative treatment; and (4) there was substantial agreement when deciding for operative versus nonoperative treatment, whereas there was only fair agreement when deciding for specific strategies.

This study has several limitations. First, with our sample size we found no difference in agreement between experienced and less-experienced hand surgeons; however, a larger sample size might have resulted in a significant difference. We performed a post hoc power analysis assessing effect size, achieved power ($1 - \beta$), and required sample size—assuming a similar effect—for assessing comminution, as the difference between experienced (kappa, 0.58) and less-experienced (kappa, 0.52) hand surgeons seems to be largest for this characteristic. This is a very small effect size (0.11), which we had only 8.1% power to detect. This small difference is probably clinically unimportant, because to achieve a power of 0.80 to be more assured that this difference is not spurious, we would need 1180 experienced and 1602 less-experienced hand surgeons [22]. Second, only $\frac{1}{3}$ of the hand surgeons in the SOVG completed the questionnaire. Because most of the people with email

addresses in the SOVG database are not active participants (ie, their email addresses are not regularly updated and we do not weed out the nonresponders), the rate of participation is not a true response rate. However, there is no difference between responding and nonresponding SOVG hand surgeons regarding sex ($p = 0.56$, by Fisher's exact test), years in practice ($p = 0.081$, by Fisher's exact test), and location of practice ($p = 0.47$, by Fisher's exact test). Participants (81 of 99; 82%) more often were involved in academic medicine (ie, supervising trainees) compared with the nonresponding hand surgeons (122 of 197; 62%; $p < 0.001$).

Furthermore, surgeons in the SOVG probably are more involved in academic medicine compared with the larger community of hand surgeons. However, we do not believe this difference compromises the generalizability of our results, as phalangeal fractures are common and typically assessed with radiographs. We found no trend toward a difference in agreement based on experience, nor was any of the included factors (ie, sex, practice location, years of experience, supervising trainees) associated with the decision for operative treatment. We therefore believe that our results are generalizable to the hand surgeon community as a whole. Third, we used only lateral radiographs for assessment of morphologic features of fractures. Use of

other imaging modalities such as CT, and physical examination, might influence interobserver agreement. We see this as a minor limitation because lateral radiographs are used most commonly and have a strong influence on the treatment of these fractures. Fourth, experience varies among participants; however, as experience is continuous and cumulative with time we decided to dichotomize level of experience based on practice years. We chose 10 years of experience as a cutoff as this resulted in more or less equally sized groups increasing statistical power. Fifth, we did not assess classification systems as these often group fractures based on multiple characteristics (for example, the AO classification uses fragment size, impaction, instability, and dislocation). We chose to study individual characteristics instead of a classification system to allow for more-detailed exploration of which of these specific factors are associated with decision for treatment.

There was moderate to substantial agreement among hand surgeons for assessment of all fracture characteristics, except for the number of fracture fragments. Assessing 2-mm articular step or gap, likelihood of subluxation or dislocation, and percentage articular surface involvement on a lateral radiograph are most reliable. This means that these three factors are most reliable to base the decision for treatment and for communication between surgeons. Classifications and guidelines should incorporate these factors, and future studies need to assess how these fracture characteristics relate to functional outcome, stiffness, and secondary osteoarthritis. Agreement on the number of fracture fragments on lateral radiographs is poor, rendering this characteristic less useful for decision-making. Assessment of fracture stability could be more reliable in clinical practice as it might improve with examination and stress radiographs [7]. Therefore it is difficult to draw firm conclusions on the usefulness of this parameter. Overall interobserver agreement for assessment of morphologic features of a middle phalanx base fracture was much greater than other anatomic areas (humerus, elbow, clavicle, and olecranon) [1, 2, 5, 8, 10]. We found only fair agreement, indicating variation, regarding specific proposed treatments, which might be a reflection of the lack of high-level evidence and relatively small case series on which surgeons base their decisions [7]. Variation in reported indications for surgical strategies also might influence decision making among hand surgeons when deciding for specific surgical strategies [7, 13, 18, 21]. Creating guidelines by reaching consensus among hand surgeons might reduce variation and improve quality of care [9, 14].

Our study showed that, with the numbers evaluated, experienced and less-experienced hand surgeons are equally good at assessing morphologic features of fractures. Hand surgeons seem to reach proficiency assessing middle phalanx base fractures on radiographs early during their experience. A previous study assessing the learning curve of pediatric

residents in diagnosing ankle fractures on radiographs showed proficiency after evaluation of 50 cases, after which learning slowed but did not stop until all 234 cases were reviewed [19]. A study of distal radius fracture classification on radiographs in children also showed that experienced surgeons were more reliable than junior registrars [20]. However, additional training, even of already fully trained surgeons, might further improve the assessment of morphologic features of fractures, as interobserver reliability improved after training of 64 fully trained orthopaedic and trauma surgeons in diagnosing scaphoid fracture displacement [6].

None of the included surgeon characteristics was associated with the decision for operative treatment, whereas all fracture characteristics were, except for uncertain fracture stability. This means that hand surgeons are consistent and likeminded regardless of sex, experience, academic involvement, and practice location when recommending operative treatment, bearing in mind the limited sample size. Our findings that articular step or gap greater than 2 mm, injuries judged unstable, and proximal interphalangeal joint subluxation or dislocation are associated most strongly with the decision for operative treatment, are in line with the indications usually cited [7, 13, 18, 21]. R-squared values (0.76 to 0.89) and β regression coefficients (0.80 to 0.90) of these three characteristics are high and vary slightly, meaning that these factors drive the decision for operative treatment to a similar degree and fashion. Future studies need to assess how these factors relate to functional outcome, stiffness, and secondary osteoarthritis.

Surgeons mostly agree on which fractures might benefit from surgery and the radiographic criteria they use most for this determination (articular step or gap and likelihood of subluxation or dislocation). The effectiveness of these thresholds for recommending surgery would be supported by additional research comparing initial operative and nonoperative approaches for these injuries, but the ethics of such research are a consideration given the current high level of agreement of hand surgeons based largely on wisdom and experience. Efforts at improving the care of proximal interphalangeal joint fracture-dislocations can focus primarily on the comparative effectiveness of the various operative treatment options.

References

1. Benetton CA, Cesa G, El-Kouba Junior G, Ferreira AP, Vissoci JR, Pietrobon R. Agreement of olecranon fractures before and after the exposure to four classification systems. *J Shoulder Elbow Surg.* 2015;24:358–363.
2. Bishop JY, Jones GL, Lewis B, Pedroza A. Intra- and interobserver agreement in the classification and treatment of distal third clavicle fractures. *Am J Sports Med.* 2015;43:979–984.
3. Bond CF Jr, Richardson K. Seeing the FisherZ-transformation. *Psychometrika.* 2004;69:291–303.

4. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Effect Sizes Based on Correlations. *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons, Ltd; 2009:41–43.
5. Bruinsma WE, Guitton TG, Warner JJ, Ring D; Science of Variation Group. Interobserver reliability of classification and characterization of proximal humeral fractures: a comparison of two and three-dimensional CT. *J Bone Joint Surg Am*. 2013;95:1600–1604.
6. Buijze GA, Guitton TG, van Dijk CN, Ring D; Science of Variation Group. Training improves interobserver reliability for the diagnosis of scaphoid fracture displacement. *Clin Orthop Relat Res*. 2012;470:2029–2034.
7. Calfee RP, Sommerkamp TG. Fracture-dislocation about the finger joints. *J Hand Surg Am*. 2009;34:1140–1147.
8. Doornberg JN, Guitton TG, Ring D; Science of Variation Group. Diagnosis of elbow fracture patterns on radiographs: interobserver reliability and diagnostic accuracy. *Clin Orthop Relat Res*. 2013;471:1373–1378.
9. Every NR, Hochman J, Becker R, Kopecky S, Cannon CP. Critical pathways : a review. Committee on Acute Cardiac Care, Council on Clinical Cardiology, American Heart Association. *Circulation*. 2000;101:461–465.
10. Gradl G, Neuhaus V, Fuchsberger T, Guitton TG, Prommersberger KJ, Ring D; Science of Variation Group. Radiographic diagnosis of scapholunate dissociation among intra-articular fractures of the distal radius: interobserver reliability. *J Hand Surg Am*. 2013;38:1685–1690.
11. Guitton TG, Ring D; Science of Variation Group. Interobserver reliability of radial head fracture classification: two-dimensional compared with three-dimensional CT. *J Bone Joint Surg Am*. 2011;93:2015–2021.
12. Hageman MG, Guitton TG, Ring D; Science of Variation Group. How surgeons make decisions when the evidence is inconclusive. *J Hand Surg Am*. 2013;38:1202–1208.
13. Jones NF, Jupiter JB, Lalonde DH. Common fractures and dislocations of the hand. *Plast Reconstr Surg*. 2012;130:722e–736e.
14. Kirkpatrick DH, Burkman RT. Does standardization of care through clinical guidelines improve outcomes and reduce medical liability? *Obstet Gynecol*. 2010;116:1022–1026.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
16. Lee J, Fung KP. Confidence interval of the kappa coefficient by bootstrap resampling. *Psychiatry Res*. 1993;49:97–98.
17. McElfresh EC, Dobyns JH, O'Brien ET. Management of fracture-dislocation of the proximal interphalangeal joints by extension-block splinting. *J Bone Joint Surg Am*. 1972;54:1705–1711.
18. Ng CY, Oliver CW. Fractures of the proximal interphalangeal joints of the fingers. *J Bone Joint Surg Br*. 2009;91:705–712.
19. Pusic M, Pecaric M, Boutis K. How much practice is enough? Using learning curves to assess the deliberate practice of radiograph interpretation. *Acad Med*. 2011;86:731–736.
20. Randsborg PH, Sivertsen EA. Classification of distal radius fractures in children: good inter- and intraobserver reliability, which improves with clinical experience. *BMC Musculoskelet Disord*. 2012;13:6.
21. Seno N, Hashizume H, Inoue H, Imatani J, Morito Y. Fractures of the base of the middle phalanx of the finger: classification, management and long-term results. *J Bone Joint Surg Br*. 1997;79:758–763.
22. Texeira A, Rosa A, Calapez T. Statistical power analysis with Microsoft Excel: normal tests for one or two means as a prelude to using non-central distributions to calculate power. *J Stat Educ*. 2009;17:1–21.
23. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005;37:360–363.