

# Online Studies on Variation in Orthopedic Surgery: Computed Tomography in MPEG4 Versus DICOM Format

Jos J. Mellema<sup>1</sup> · Wouter H. Mallee<sup>2</sup> · Thierry G. Guitton<sup>2</sup> · C. Niek van Dijk<sup>2</sup> · David Ring<sup>3</sup> · Job N. Doornberg<sup>2</sup> · Science of Variation Group & Traumatplatform Study Collaborative

Published online: 24 January 2017

© Society for Imaging Informatics in Medicine 2017

**Abstract** The purpose of this study was to compare the observer participation and satisfaction as well as interobserver reliability between two online platforms, Science of Variation Group (SOVG) and Traumatplatform Study Collaborative, for the evaluation of complex tibial plateau fractures using computed tomography in MPEG4 and DICOM format. A total of 143 observers started with the online evaluation of 15 complex tibial plateau fractures via either the SOVG or Traumatplatform Study Collaborative websites using MPEG4 videos or a DICOM viewer, respectively. Observers were asked to indicate the absence or presence of four tibial plateau fracture characteristics and to rate their satisfaction with the evaluation as provided by the respective online platforms. The observer participation rate was significantly higher in the SOVG (MPEG4 video) group compared to that in the Traumatplatform Study Collaborative (DICOM viewer) group (75 and 43%, respectively;  $P < 0.001$ ). The median observer satisfaction with the online evaluation was seven (range, 0–10) using MPEG4 video compared to six (range, 1–9) using DICOM viewer ( $P = 0.11$ ). The interobserver reliability for

recognition of fracture characteristics in complex tibial plateau fractures was higher for the evaluation using MPEG4 video. In conclusion, observer participation and interobserver reliability for the characterization of tibial plateau fractures was greater with MPEG4 videos than with a standard DICOM viewer, while there was no difference in observer satisfaction. Future reliability studies should account for the method of delivering images.

**Keywords** Variation · Orthopedic surgery · Tibial plateau fractures · Computed tomography · MPEG · DICOM

## Introduction

The Science of Variation Group (SOVG) is a collaborative effort to improve the study of variation in interpretation and classification of injuries. Through <http://www.scienceofvariationgroup.org>, many “interobserver reliability” [1–13] studies were completed with large numbers of fully trained, practicing, and

✉ Jos J. Mellema  
josjmellema@gmail.com

Wouter H. Mallee  
w.h.mallee@amc.uva.nl

Thierry G. Guitton  
guitton@gmail.com

C. Niek van Dijk  
c.n.vandijk@amc.uva.nl

David Ring  
david.ring@austin.utexas.edu

Job N. Doornberg  
doornberg@traumatplatform.org

Science of Variation Group & Traumatplatform Study Collaborative

<sup>1</sup> Hand and Upper Extremity Service, Department of Orthopedic Surgery, Massachusetts General Hospital, Harvard Medical School, Yawkey Center, Suite 2100, 55 Fruit Street, Boston, MA 02114, USA

<sup>2</sup> Department of Orthopedic Surgery, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1100 DD, Amsterdam, The Netherlands

<sup>3</sup> Department of Surgery and Perioperative Care, Dell Medical School, The University of Texas at Austin, 1400 Barbara Jordan Blvd. Suite 2.834; MC: R1800, Austin, TX 78723, USA

experienced surgeons from all over the world predominantly the USA and Europe. Interobserver reliability studies require an appropriate balance between the number of observers evaluating each subject and the number of subjects [14]. Specifically, it is the number of observations rather than the number of observers or subjects that creates the power in this type of study. By having a large number of observers that make a small number of observations, we can increase participation. Increased participation (i.e., large number of observers) leads to greater generalizability of study findings and allows study of many more aspects of variation.

Previous studies on factors of variation among surgeons revealed that (1) training improves interobserver reliability [3, 15], (2) simple classification systems lead to better—but still not perfect—agreement [2], (3) objective-quantified measurements improve agreement on surgical decision-making [7], and that (4) advanced imaging modalities such as two-dimensional (2D) multiplanar reformatting (MPR) versus three-dimensional (3D) volume-rendering technique (VRT) have little, but statistically significant influence on interobserver variation [2, 16–20]. However, the effect of these factors on interobserver reliability seems to differ between fracture classifications and anatomical sites.

A proposed methodological limitation of interobserver reliability studies performed via the SOVG has been that of a standard Digital Imaging and Communications in Medicine (DICOM) viewer, as used in daily clinical practice, could not be used for the evaluation of injuries due to technical limitations of online DICOM viewers. Consequently, observers did not have the ability to adjust contrast, brightness, and window level and to zoom in or out. The incorporation of DICOM viewers into the online study of variation should more closely resemble a day-to-day practice. It is therefore plausible that the formatting of medical images (i.e., MPEG4 video versus DICOM) might affect interobserver variability. On the other hand, using the DICOM viewer might be more cumbersome for busy clinicians that are willing to participate in brief (less than 30 min) surveys, but not much longer.

The purpose of this study was to compare the observer participation and satisfaction as well as interobserver reliability between two online platforms: Science of Variation Group (MPEG4 video) and Traumaplatfrom Study Collaborative (DICOM viewer). We tested the null hypotheses (1) that there is no difference in observer participation between observers who started with the online evaluation of computed tomography (CT) scans of tibial plateau fractures using MPEG4 videos provided by the SOVG and observers who started with the same evaluation using a DICOM viewer provided by the Traumaplatfrom and (2) that there is no difference in observer satisfaction and interobserver reliability between observers who complete the online evaluation via the SOVG (using MPEG4 video) and Traumaplatfrom (using DICOM viewer).

## Materials and Methods

### Study Design

This study is based on secondary use of data from an online evaluation in which 143 observers started evaluating tibial plateau fracture characteristics of 15 subjects with complex tibial plateau fractures. The online evaluation was designed to study two primary endpoints: (1) interobserver reliability and accuracy for CT-based evaluation of tibial plateau fracture characteristics and (2) interobserver reliability of the Schatzker and Luo tibial plateau fracture classification. Observers were randomized (1:1) within the online platform groups (i.e., SOVG and Traumaplatfrom Study Collaborative) by computer-generated random numbers to assess CT scans provided in MPR with or without additional 3D volume-rendered reconstructions.

For these studies, our Institutional Review Board approved a web-based evaluation using an anonymized CT scan of complex tibial plateau fractures. Musculoskeletal trauma surgeons and residents affiliated with the SOVG and Traumaplatfrom Study Collaborative, as well as authors who have published on tibial plateau fractures in the last decade (between 2004 and 2014 in orthopedic peer-reviewed literature) were invited to participate. All corresponding authors and members of our collaborative were approached via email, and the only incentive to participate was to be credited on the study by acknowledgement.

### Observers

Five hundred and twenty-two observers were invited to participate via the SOVG (<http://www.scienceofvariationgroup.org>) or Traumaplatfrom Study Collaborative (<http://www.traumaplatfrom.org>) websites. A total of 143 respondents started with the evaluation: 63 (44%) via the SOVG (MPEG4 videos) and 80 (56%) via the Traumaplatfrom (DICOM viewer). Among the SOVG respondents, there was a greater percentage of observers from Europe and of observers 0–5 years in practice, although these differences were not significant (Table 1).

### Online Evaluation

The SOVG provided a link to [www.surveymonkey.com](http://www.surveymonkey.com) (SurveyMonkey, Palo Alto, CA, USA) for online evaluation of complex tibial plateau fractures. Computed tomography scans of the selected tibial plateau fractures were converted into videos (MPEG4 format) using DICOM viewer software and uploaded into SurveyMonkey. On SurveyMonkey, CT scans provided in MPR with or without additional 3D volume-rendered reconstructions were displayed with the use of a syntax that allowed proper relative position of videos. Observers were able to play and scroll

**Table 1** Observer characteristics<sup>a</sup>

	MPEG4 video (n = 63)		DICOM viewer (n = 80)		Total (n = 143)
	n	%	n	%	n
Sex					
Male	61	97	72	90	133
Female	2	3	8	10	10
Area					
USA	8	13	18	23	26
Europe	38	60	32	40	70
Asia	4	6	8	10	12
Canada	3	5	5	6	8
U.K.	3	5	8	10	11
Australia	3	3	0	0	3
Other	4	6	9	11	13
Years in independent practice					
0–5	22	35	19	24	41
6–10	15	24	26	33	41
11–20	17	27	22	28	39
21–30	9	14	13	16	22
Specialization					
General orthopedics	12	19	11	14	23
Orthopedic traumatology	44	70	58	73	102
Shoulder and elbow	1	2	0	0	1
Hand and wrist	0	0	2	3	2
Others	6	10	9	11	15
Supervision of trainees					
Yes	55	87	72	90	127
No	8	13	8	10	16

<sup>a</sup>Demographics of observers that started with the evaluation

through the MPEG4 videos of transverse, sagittal, and coronal planes and 3D reconstructions of the selected CT scans. This online viewer did not facilitate adjustments of contrast, brightness, window leveling, nor any other zoom options.

The Traumaplatfom Study Collaborative provided a link to [www.traumaplatfom.org/onlinestudies](http://www.traumaplatfom.org/onlinestudies). Digital Imaging and Communications in Medicine files of CT scans of the selected tibial plateau fractures were uploaded to the website and displayed in built-in DICOM viewers. To be able to use the DICOM viewer, observers were asked to install the most recent Adobe Flash Player and to use a web browser with JavaScript enabled prior to their login. In general, the online DICOM viewer is limited by computer properties and internet speed. Upon login, observers received CT scans provided in MPR with or without additional 3D volume-rendered reconstructions. Observers were able to window level, scale, position, rotate, and scroll back and forth through transverse,

sagittal, and coronal planes and 3D reconstruction images of selected CT scans (Fig. 1).

**Outcome Measures**

The primary outcome measure was observer participation. The observer participation was defined as the proportion of observers who completed the evaluation among the observers who started with the evaluation. Observers who started with the evaluation were included if they clicked on the link and answered at least one question.

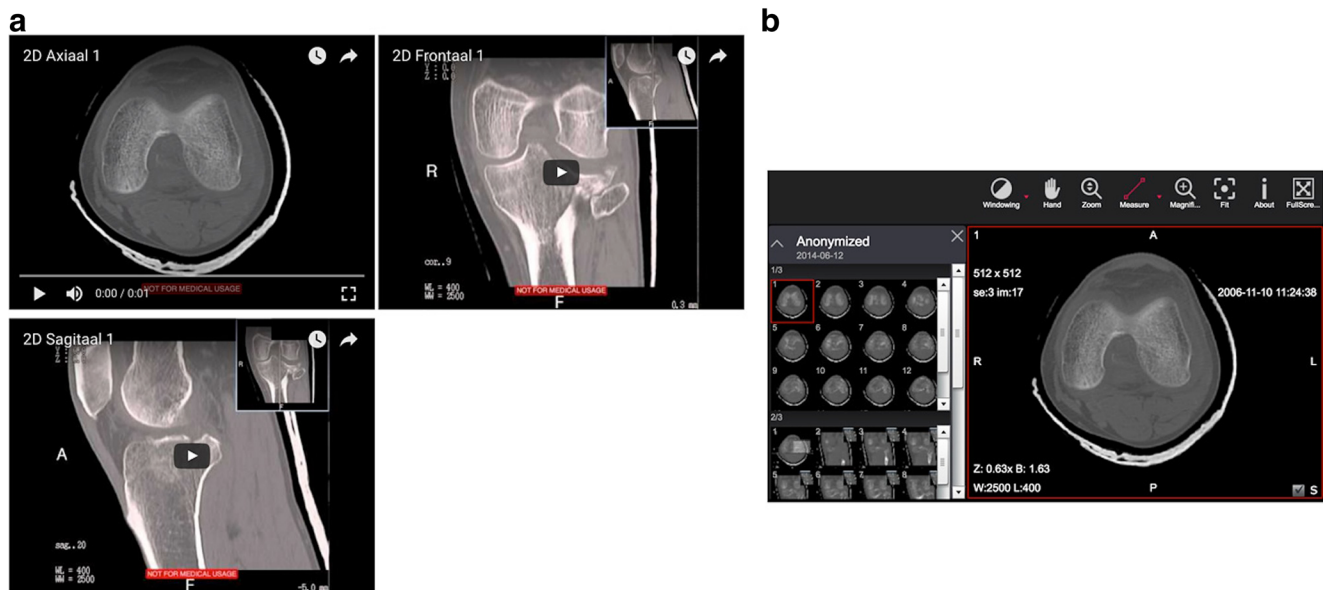
The secondary outcome variables were observer satisfaction and interobserver reliability. The observer satisfaction was obtained directly after the online evaluation. Observers were asked to rate their satisfaction with the online evaluation of complex tibial plateau fractures provided by SOVG or Traumaplatfom on an 11-point ordinal scale. The interobserver reliability of CT-based characterization of tibial plateau fractures was calculated for the MPEG4 video group and the online DICOM viewer group. The majority of observers (62%) participated previously in online SOVG studies evaluating videos in MPEG4 format via SurveyMonkey, and a smaller proportion of observers (17%) were familiar with online DICOM viewer technique through studies via [www.ankleplatfom.com](http://www.ankleplatfom.com) as part of our collaborative.

**Statistical Analysis**

Post hoc, it was determined that for the primary hypothesis, based on the chi-square test, 143 observers provided 99% power to detect a 0.32 difference (effect size  $w = 0.35$ ) in proportion of observers between the MPEG4 and DICOM viewer group ( $\alpha = 0.05$ ).

Observer characteristics were summarized with frequencies and percentages. The chi-squared or Fisher exact test was used to determine difference in observer participation between the SOVG and Traumaplatfom and within subgroups. The Wilcoxon rank-sum and Kruskal-Wallis test was used to determine difference in observer satisfaction as the assumption of normality that was not met. Subgroup categories were grouped in case of low numbers to ensure equal to or greater than five responses in each group.

Interobserver reliability was determined with the use of Siegel and Castellan’s multirater kappa [21], which is a frequently used measure of chance-corrected agreement between multiple observers. The kappa values were interpreted according to the guidelines of Landis and Koch: a value of 0.01 to 0.20 indicates slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 to 0.99, almost perfect agreement [22]. Kappa values were compared using the two-sample z test. P values of <0.05 were considered significant.



**Fig. 1** The online evaluation of computed tomography scans of complex tibial plateau fractures. **a** Science of Variation Group provided CT scans converted into videos (MPEG4 format). Observers were able to play and scroll through the MPEG4 videos of transverse, sagittal, and coronal

planes. **b** The Traumaplatform Study Collaborative provided CT scans displayed in built-in DICOM viewers. Observers were able to window level, scale, position, rotate, and scroll back and forth through transverse, sagittal, and coronal planes

## Results

### Observer Participation of MPEG4 Video Versus DICOM Viewer

There was a significant difference in observer participation between the MPEG4 video and DICOM viewer group ( $P < 0.001$ ); in the MPEG4 video group, 75% of the observers who started the evaluation completed it, compared to 43% in the DICOM viewer group (Table 2).

In the MPEG4 video group, the observer participation was greatest among women and observers from Canada (100%) and lowest in observers from Asia (50%). For the DICOM viewer group, the observer participation was greatest among observers from Canada (80%) and lowest among observers from areas other than the USA, Europe, Canada, and Asia (24%). Observers in the USA were more likely to complete

the study in the MPEG4 video group (75% versus 33%). No observer characteristics were significantly associated with observer participation in both groups (Table 3).

### Observer Satisfaction of MPEG4 Video Versus DICOM Viewer

The median observer satisfaction with the online evaluation of observers who completed the evaluation was seven (range, 0–10) using MPEG4 video compared to six (range, 1–9) using DICOM viewer ( $P = 0.11$ ). The median ranged from eight (range, 3–10) among observers 6–10 years in practice to six (range, 3–9) among observers from the USA in the MPEG4 video group. In the DICOM viewer group, the median ranged from seven (range, 5–9) among observers 21–30 years in practice to six (range, 1–8) among observers from the USA.

**Table 2** Observer participation and satisfaction by web-based study platform (MPEG4 video versus DICOM viewer)

	MPEG4 video		DICOM viewer		P value
	n	%	n	%	
Observer participation <sup>a</sup>					<0.001
Number of observers that completed the evaluation	47	75	34	43	
Number of observers that not completed the evaluation	16	25	46	58	
Observer satisfaction <sup>b</sup>	7	0–10	6	1–9	0.11

<sup>a</sup> Participation of observers that started with the evaluation ( $n = 143$ )

<sup>b</sup> Satisfaction of observers that completed the evaluation ( $n = 81$ )

**Table 3** Observer participation<sup>a</sup> by observer characteristics and web-based study platform (MPEG4 video and DICOM viewer)

	MPEG4 video ( <i>n</i> = 63)				<i>P</i> value	DICOM viewer ( <i>n</i> = 80)				<i>P</i> value
	Observers that not completed the evaluation		Observers that completed the evaluation			Observers that not completed the evaluation		Observers that completed the evaluation		
	<i>n</i>	%	<i>n</i>	%		<i>n</i>	%	<i>n</i>	%	
Sex					0.99					0.46
Male	16	100	45	96		40	87	32	94	
Female	0	0	2	4		6	13	2	6	
Area					0.43					0.14
USA	2	13	6	13		12	26	6	18	
Europe	8	50	30	64		16	35	16	47	
Asia	2	13	2	4		4	9	4	12	
Canada	0	0	3	6		1	2	4	12	
Other	4	25	6	13		13	28	4	12	
Years in independent practice					0.51					0.45
0–5	4	25	18	38		12	26	7	21	
6–10	4	25	11	23		12	26	14	41	
11–20	4	25	13	28		15	33	7	21	
21–30	4	25	5	11		7	15	6	18	
Specialization					0.66					0.94
General orthopedics	4	25	8	17		6	13	5	15	
Orthopedic traumatology	10	63	34	72		33	72	25	74	
Other	2	13	5	11		7	15	4	12	
Supervision of trainees					0.41					0.99
Yes	13	81	42	89		41	89	31	91	
No	3	19	5	11		5	11	3	9	

<sup>a</sup> Participation of observers that started with the evaluation

No factors were significantly associated with observer satisfaction in either group (Tables 2 and 4).

**Interobserver Reliability of MPEG4 Video versus DICOM Viewer**

Interobserver reliability was significantly better for observers who completed the online evaluation of CT scans of tibial plateau fractures using MPEG4 videos. More specifically, the interobserver reliability of posteromedial component ( $k_{\text{MPEG4}} = 0.54$  and  $k_{\text{DICOM}} = 0.26$ ,  $P < 0.001$ ), lateral component ( $k_{\text{MPEG4}} = 0.56$  and  $k_{\text{DICOM}} = 0.48$ ,  $P = 0.038$ ), and tibial spine component ( $k_{\text{MPEG4}} = 0.49$  and  $k_{\text{DICOM}} = 0.37$ ,  $P < 0.001$ ) were higher for the evaluation using MPEG4 video compared to that using DICOM viewer (Table 5).

**Discussion**

Online study platforms that study variation in orthopedic surgery allow more complex study design and improve power

and generalizability. Using prepared videos and photos that cannot be adjusted makes it easier for the observers to get through a survey efficiently. But some express concern that the lack of a DICOM viewer similar to that used in daily practice might limit reliability and accuracy. We compared two online platforms: the Science of Variation Group (using MPEG4 videos) and Traumaplatform (using a built-in DICOM viewer) for the online evaluation of CT scans. We found that the observer participation and interobserver reliability was higher for observers who evaluated CT scans converted into MPEG4 videos provided by the SOVG compared to observers who did the same evaluation using an online DICOM viewer provided by the Traumaplatform.

This study has several limitations. First, all outcome variables but the primary outcome variable (i.e., observer satisfaction and interobserver reliability) could only be evaluated for observers who completed the online evaluation. The observers who did not complete the evaluation (25% in the MPEG4 video group and 57% in the DICOM viewer group) could have had different satisfaction scores and agreement. Second, a standardized environment with the same state-of-the-art computer using a uniform



**Table 4** Observer satisfaction<sup>a</sup> by observer characteristics and web-based study platform (MPEG4 video and DICOM viewer)

	MPEG4 video (n = 47)		P value	DICOM viewer (n = 34)		P value
	Observer satisfaction			Observer satisfaction		
	Median	Range		Median	Range	
Area			0.68			0.43
USA	6	3–9		6	1–8	
Europe	7	0–10		7	4–9	
Other	8	2–10		6	1–9	
Years in independent practice			0.27			0.80
0–5	7	0–10		6	5–7	
6–10	8	3–10		7	1–9	
11–20	7	1–9		7	2–9	
21–30	8	2–10		7	5–9	
Specialization			0.55			0.17
Orthopedic traumatology	8	0–10		6	1–9	
Others	7	1–10		7	5–9	

<sup>a</sup> Satisfaction of observers that completed the evaluation

web browser with the required JavaScript enabled could have prevented technical difficulty related to the online DICOM viewer. Third, we did not control for difference in imaging modality between the groups; however, randomization to assess CT scans provided in MPR with or without additional 3D volume-rendered reconstructions within the video and DICOM group guaranteed well-balanced groups. Fourth, we did not measure the time of image reading while this may be an important determinant of reading incompleteness and observer satisfaction. Finally, the evaluation of radiographs was not included in this study. Therefore, our findings apply best to CT images.

Observer participation is higher for prepared MPEG4 videos than that for a DICOM viewer, most likely because an observer can more efficiently get through the cases and faces fewer potential technical problems. A high participation rate and a good experience that make observers willing to participate in future online evaluations are both critically important and seem better achieved with prepared videos.

There was a trend that observers of videos were more satisfied than those who used the DICOM viewer, but it was not significant with the numbers available. It seems safe to assume that observers who did not complete the study were less satisfied.

Extensive effort has been made to improve reliability, including training [3, 15], different imaging modalities [2, 16–20, 23], and simplifying classification systems [2, 24]. Our findings show that the interobserver reliability for the recognition of tibial plateau fracture characteristics on CT scans is higher on MPEG4 videos compared to that on DICOM viewer. Critics of the use of videos thought that being able to adjust contrast, brightness, and window leveling on a DICOM viewer would reduce variation, but it turns out that standardization of these aspects of the image might help decrease variability.

In conclusion, our results showed that online CT-based evaluation of tibial plateau fracture characteristics using MPEG videos resulted in higher observer participation and

**Table 5** Overall Interobserver Agreement of Fracture Characteristics by Web-Based Study Platform (MPEG4 video versus DICOM viewer)

	MPEG4 video (n = 47)			DICOM viewer (n = 34)			P value
	Kappa <sup>a</sup>	Agreement	95% CI	Kappa	Agreement	95% CI	
Fracture characteristics							
Posteromedial component	0.54	Moderate	0.47–0.61	0.26	Fair	0.16–0.36	<0.001
Lateral component	0.56	Moderate	0.51–0.61	0.48	Moderate	0.43–0.54	0.038
Tibial tubercle component	0.27	Fair	0.24–0.30	0.33	Fair	0.27–0.38	0.073
Tibial spine component	0.49	Moderate	0.47–0.50	0.37	Fair	0.35–0.40	<0.001

<sup>a</sup> Interobserver agreement of fracture characteristics among observers that completed the evaluation

interobserver reliability compared to the use of an online DICOM viewer for the same evaluation, while we found no difference in observer satisfaction between the groups. However, the use of DICOM viewers for the online evaluation of radiographs has not been demonstrated and might be different as the evaluation of radiographs in a DICOM viewer may be less technically demanding compared to its use for CT scans. Moreover, web-based technologies will keep improving and it is most likely that better online DICOM viewers will be available in the years to come allowing online CT-based evaluation with comparable or better performance than the current SOVG methodology.

**Acknowledgement** The authors acknowledge the Science of Variation Group and Traumaplatform Study Collaborative:

Babis, G.C.; Jeray, K.J.; Prayson, M.J.; Pesantez, R.; Acacio, R.; Verbeek, D.O.; Melvanki, P.; Kreis, B.E.; Mehta, S.; Meylaerts, S.; Wojtek, S.; Yeap, E.J.; Haapasalo, H.; Kristan, A.; Coles, C.; Marsh, J.L.; Mormino, M.; Menon, M.; Tyllianakis, M.; Schandelmaier, P.; Jenkinson, R.J.; Neuhaus, V.; Shahriar, C.M.H.; Belangero, W.D.; Kannan, S.G.; Leonidovich, G.M.; Davenport, J.H.; Kabir, K.; Althausen, P.L.; Weil, Y.; Toom, A.; Sa da Costa, D.; Lijoi, F.; Koukoulas, N.E.; Manidakis, N.; Van den Bogaert, M.; Patczai, B.; Grauls, A.; Kurup, H.; van den Bekerom, M.P.; Lansdaal, J.R.; Vale, M.; Ousema, P.; Barquet, A.; Cross, B.J.; Broekhuysse, H.; Haverkamp, D.; Merchant, M.; Harvey, E.; Stojkowska Pemovska, E.; Frihagen, F.; Seibert, F.J.; Garnavos, C.; van der Heide, H.; Villamizar, H.A.; Harris, I.; Borris, L.C.; Brink, O.; Brink, P.R.G.; Choudhari, P.; Swiontkowski, M.; Mittlmeier, T.; Tosounidis, T.; van Rensen, I.; Martinelli, N.; Park, D.H.; Lasanianos, N.; Vide, J.; Engvall, A.; Zura, R.D.; Jubel, A.; Kawaguchi, A.; Goost, H.; Bishop, J.; Mica, L.; Pirpiris, M.; van Helden, S.H.; Bouaicha, S.; Schepers, T.; Havlicek, T.; and Giordano, V.

### Compliance with Ethical Standards

**Conflict of Interest** Each author certifies that he or she has no commercial associations (e.g., consultancies, stock ownership, equity interest, patent/licensing arrangements) that might pose a conflict of interest in connection with the submitted article.

**Ethical Review Committee** The Institutional Review Board of our institution approved this study under protocol #2009P001019/MGH.

**The Location Where the Work Was Performed** The work was performed at the Hand and Upper Extremity Service, Department of Orthopedic Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA and Department of Orthopedic Surgery, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

### References

- Bruinsma WE, Guitton T, Ring D: Radiographic loss of contact between radial head fracture fragments is moderately reliable. *Clinical orthopaedics and related research*, 2014
- Bruinsma WE, Guitton TG, Warner JJ, Ring D: Interobserver reliability of classification and characterization of proximal humeral fractures: a comparison of two and three-dimensional CT. *The Journal of bone and joint surgery American volume* 95:1600–1604, 2013
- Buijze GA, Guitton TG, van Dijk CN, Ring D: Training improves interobserver reliability for the diagnosis of scaphoid fracture displacement. *Clinical orthopaedics and related research* 470:2029–2034, 2012
- Buijze GA, Wijffels MM, Guitton TG, Grewal R, van Dijk CN, Ring D: Interobserver reliability of computed tomography to diagnose scaphoid waist fracture union. *The Journal of hand surgery* 37: 250–254, 2012
- Doornberg JN, Guitton TG, Ring D: Diagnosis of elbow fracture patterns on radiographs: interobserver reliability and diagnostic accuracy. *Clinical orthopaedics and related research* 471:1373–1378, 2013
- Gratl G, Neuhaus V, Fuchsberger T, Guitton TG, Prommersberger KJ, Ring D: Radiographic diagnosis of scapholunate dissociation among intra-articular fractures of the distal radius: interobserver reliability. *The Journal of hand surgery*, 2013
- Tosti R, Ilyas AM, Mellema JJ, Guitton TG, Ring D: Interobserver variability in the treatment of little finger metacarpal neck fractures. *The Journal of hand surgery* 39:1722–1727, 2014
- Wijffels MM, Guitton TG, Ring D: Inter-observer variation in the diagnosis of coronal articular fracture lines in the lunate facet of the distal radius. *Hand (New York, NY)* 7:271–275, 2012
- Neuhaus V, et al.: Scapula fractures: interobserver reliability of classification and treatment. *Journal of orthopaedic trauma* 28: 124–129, 2014
- van Kollenburg JA, Vrahas MS, Smith RM, Guitton TG, Ring D: Diagnosis of union of distal tibia fractures: accuracy and interobserver reliability. *Injury* 44:1073–1075, 2013
- Mellema JJ, Doornberg JN, Molenaars RJ, Ring D, Kloen P: Tibial plateau fracture characteristics: reliability and diagnostic accuracy. *Journal of orthopaedic trauma* 30:e144–151, 2016
- Mallee WH, Mellema JJ, Guitton TG, Goslings JC, Ring D, Doornberg JN: 6-week radiographs unsuitable for diagnosis of suspected scaphoid fractures. *Archives of orthopaedic and trauma surgery* 136:771–778, 2016
- Mellema JJ, Doornberg JN, Molenaars RJ, Ring D, Kloen P: Interobserver reliability of the Schatzker and Luo classification systems for tibial plateau fractures. *Injury* 47:944–949, 2016
- Walter SD, Eliasziw M, Donner A: Sample size and optimal designs for reliability studies. *Statistics in medicine* 17:101–110, 1998
- Brorson S, Bagger J, Sylvest A, Hrobjartsson A: Improved interobserver variation after training of doctors in the Neer system. A randomised trial. *The Journal of bone and joint surgery British volume* 84:950–954, 2002
- Brunner A, Heeren N, Albrecht F, Hahn M, Ulmar B, Babst R: Effect of three-dimensional computed tomography reconstructions on reliability. *Foot & ankle international* 33:727–733, 2012
- Bishop JY, Jones GL, Rerko MA, Donaldson C: 3-D CT is the most reliable imaging modality when quantifying glenoid bone loss. *Clinical orthopaedics and related research* 471:1251–1256, 2013
- Doornberg J, Lindenhovius A, Kloen P, van Dijk CN, Zurakowski D, Ring D: Two and three-dimensional computed tomography for the classification and management of distal humeral fractures. Evaluation of reliability and diagnostic accuracy. *The Journal of bone and joint surgery American volume* 88:1795–1801, 2006
- Lindenhovius A, Karanicolas PJ, Bhandari M, van Dijk N, Ring D: Interobserver reliability of coronoid fracture classification: two-dimensional versus three-dimensional computed tomography. *The Journal of hand surgery* 34:1640–1646, 2009
- Guitton TG, Ring D: Interobserver reliability of radial head fracture classification: two-dimensional compared with three-dimensional CT. *The Journal of bone and joint surgery American volume* 93: 2015–2021, 2011

21. Siegel S, Castellan JN: Nonparametric statistics for the behavioral sciences, New York: McGraw-Hill, 1988
22. Landis JR, Koch GG: An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 33:363–374, 1977
23. Berkes MB, et al.: The impact of three-dimensional CT imaging on intraobserver and interobserver reliability of proximal humeral fracture classifications and treatment recommendations. *The Journal of bone and joint surgery American volume* 96:1281–1286, 2014
24. Laoutliv B, Havsteen I, Bech BH, Narvestad E, Christensen H, Christensen A: Interobserver agreement in fusion status assessment after instrumental desis of the lower lumbar spine using 64-slice multidetector computed tomography: impact of observer experience. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 21:2085–2090, 2012