

Clinical Research

# What Are the Interobserver and Intraobserver Variability of Gap and Stepoff Measurements in Acetabular Fractures?

Anne M. L. Meesters MSc, Kaj ten Duis MD, Hester Banierink BSc, Vincent M. A. Stirler MD, PhD, Philip C. R. Wouters MD, Joep Kraeima PhD, Jean-Paul P. M. de Vries MD, PhD, Max J. H. Witjes MD, PhD, Frank F. A. IJpma MD, PhD

Received: 12 February 2020 / Accepted: 15 June 2020 / Published online: 16 July 2020  
Copyright © 2020 by the Association of Bone and Joint Surgeons

## Abstract

**Background** Gap and stepoff values in the treatment of acetabular fractures are correlated with clinical outcomes. However, the interobserver and intraobserver variability of gap and stepoff measurements for all imaging modalities in the preoperative, intraoperative, and postoperative phase of treatment is unknown. Recently, a standardized CT-based measurement method was introduced, which provided the opportunity to assess the level of variability.

**Questions/purposes** (1) In patients with acetabular fractures, what is the interobserver variability in the

measurement of the fracture gaps and articular stepoffs determined by each observer to be the maximum one in the weightbearing dome, as measured on pre- and postoperative pelvic radiographs, intraoperative fluoroscopy, and pre- and postoperative CT scans? (2) What is the intraobserver variability in these measurements?

**Methods** Sixty patients with a complete subset of pre-, intra- and postoperative high-quality images (CT slices of < 2 mm), representing a variety of fracture types with small and large gaps and/or stepoffs, were included. A total of 196 patients with nonoperative treatment (n = 117), inadequate available imaging (n = 60), skeletal immaturity (n = 16), bilateral fractures (n = 2) or a primary THA (n = 1) were excluded. The maximum gap and stepoff values in the weightbearing dome were digitally measured on pelvic radiographs and CT images by five independent observers. Observers were free to decide which gap and/or stepoff they considered the maximum and then measure these before and after surgery. The observers were two trauma surgeons with more than 5 years of experience in pelvic surgery, two trauma surgeons with less than 5 years of experience in pelvic surgery, and one surgical resident. Additionally, the final intraoperative fluoroscopy images were assessed for the presence of a gap or stepoff in the weightbearing dome. All observers used the same standardized measurement technique and each observer measured the first five patients together with the responsible researcher. For 10 randomly selected patients, all measurements were repeated by all observers, at least 2 weeks after the initial measurements. The intraclass correlation coefficient (ICC) for pelvic radiographs and CT images and the kappa value for intraoperative fluoroscopy measurements were calculated to determine the inter- and

---

Each author certifies that neither he or she, nor any member of his or her immediate family, has funding or commercial associations (consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

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. Each author certifies that his or her institution waived approval for the human protocol for this investigation and that all investigations were conducted in conformity with ethical principles of research.

---

A. M. L. Meesters, K. ten Duis, H. Banierink, V. M. A. Stirler, P. C. R. Wouters, J. P. P. M. de Vries, F. F. A. IJpma, Department of Surgery, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

J. Kraeima, M. J. H. Witjes, 3D Lab/Department of Oral and Maxillofacial Surgery, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

F. F. A. IJpma ✉, Department of Trauma Surgery, HPC BA13, Hanzeplein 1, 9713 GZ Groningen, the Netherlands, Email: frankijpma@gmail.com

intraobserver variability. Interobserver variability was defined as the difference in the measurements between observers. Intraobserver variability was defined as the difference in repeated measurements by the same observer. **Results** Preoperatively, the interobserver ICC was 0.4 (gap and stepoff) on radiographs and 0.4 (gap) and 0.3 (stepoff) on CT images. The observers agreed on the indication for surgery in 40% (gap) and 30% (stepoff) on pelvic radiographs. For CT scans the observers agreed in 95% (gap) and 70% (stepoff) of images. Postoperatively, the interobserver ICC was 0.4 (gap) and 0.2 (stepoff) on radiographs. The observers agreed on whether the reduction was acceptable or not in 60% (gap) and 40% (stepoff). On CT images the ICC was 0.3 (gap) and 0.4 (stepoff). The observers agreed on whether the reduction was acceptable in 35% (gap) and 38% (stepoff). The preoperative intraobserver ICC was 0.6 (gap and stepoff) on pelvic radiographs and 0.4 (gap) and 0.6 (stepoff) for CT scans. Postoperatively, the intraobserver ICC was 0.7 (gap) and 0.1 (stepoff) on pelvic radiographs. On CT the intraobserver ICC was 0.5 (gap) and 0.3 (stepoff). There was no agreement between the observers on the presence of a gap or stepoff on intraoperative fluoroscopy images (kappa -0.1 to 0.2).

**Conclusions** We found an insufficient interobserver and intraobserver agreement on measuring gaps and stepoffs for supporting clinical decisions in acetabular fracture surgery. If observers cannot agree on the size of the gap and stepoff, it will be challenging to decide when to perform surgery and study the results of acetabular fracture surgery. **Level of Evidence** Level III, diagnostic study.

## Introduction

Radiographs, fluoroscopic images, and CT images are valuable and essential tools in the care of patients with acetabular fractures. Gap and stepoff measurements derived from these imaging techniques are used as main indications for acetabular fracture surgery, and they are common measurements used for research that deals with prognosis after surgery [2, 8, 24, 25]. If these measurements are unreliable, then the surgical indications are questionable, and the assessments of our work (both clinically and in the clinical research setting) are questionable as well. A misrepresentation of the quality of reduction is especially worrisome when the reduction is correlated to prognosis. Verbeek et al. [24] examined a large series and showed that a postoperative gap of > 5 mm and/or stepoff of > 1 mm are associated with an increased risk of conversion to THA in the long-term. Therefore, it is important that the measurements are reliable and reproducible.

The gap and stepoff measurements consist of some subjective elements, which could be interpreted differently

by different surgeons. The assessment of a fracture relies on where the measurement is performed; that is, which fracture line or CT slice is selected for measurement and how the measurement is performed. Limited evidence is available to guide us on which fracture line to measure for a gap or stepoff. Moreover, there is substantial work that has sought to determine whether surgeons classify fractures reliably, but these papers draw different conclusions. Some conclude that classifying the fractures is reliable [3–5, 7, 16, 19, 21, 27] while others suggest the opposite [9, 17, 18, 20, 26]. Therefore, there is controversy about how well observers agree when looking at the same radiograph or CT image in acetabular fracture treatment. Because gap and stepoff measurements are so important to the indication for surgery and the prognosis, and because it should, in theory, be simpler to measure a gap or a stepoff than it is to apply a complex classifications system, we wanted to see how reliably gaps and stepoffs could be measured.

We therefore asked: (1) In patients with acetabular fractures, what is the interobserver variability in the measurement of the fracture gaps and articular stepoffs determined by each observer to be the maximum one in the weightbearing dome, as measured on pre- and postoperative pelvic radiographs, intraoperative fluoroscopy and pre- and postoperative CT scans? (2) What is the intraobserver variability in these measurements?

## Patients and Methods

### Participants

Between January 1, 2007 and July 1, 2018, we treated 256 patients with acetabular fractures. Of these, 46% (117 out of 256) of the patients were treated nonoperatively. Of the remaining patients, 60 were excluded because imaging was incomplete or of poor quality (a CT scan with a slice thickness of > 2 mm), another 16 were excluded because of skeletal immaturity, two more because of the presence of bilateral fractures, and one was excluded because the patient was primarily treated with total hip arthroplasty. Thus, the patient group for this study included 60 patients with a complete set of high quality pre-, intra-, and postoperative images. All acetabular fractures were classified according to the Letournel and Judet classification system [1, 10] and the AO/OTA classification system [1, 13]. All available fracture types were included to prevent potential bias. Baseline characteristics were retrieved from the patients' electronic records. The median age was 49 years (interquartile range 39 to 62). Fifteen percent (9 of 60) of patients were female. According to the AO/OTA classification, 33% (20 of 60) of patients sustained a Type A fracture, 30% (18 of 60) of patients sustained a Type B fracture, and 37% (22 of 60) of patients sustained a Type C fracture. According to

the Letournel classification, both-column fractures were the most common in our study population (Fig. 1).

This study was reviewed and a waiver was provided by our local medical ethics review committee.

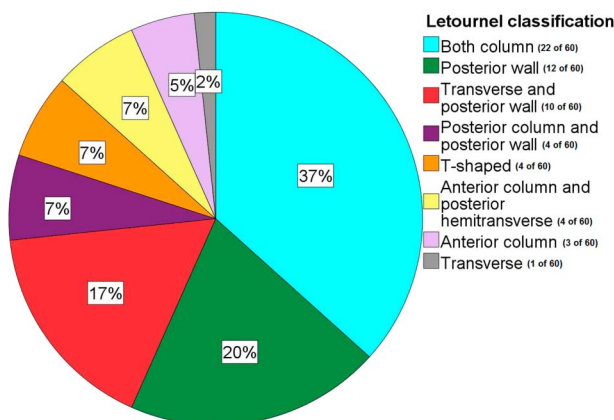
## Measurements

Gap and stepoff measurements were performed on all preoperative and postoperative AP pelvic radiographs and all axial, coronal, and sagittal CT slices. Observers were asked to measure the maximum gap and stepoff in the weightbearing dome, according to Verbeek et al.'s [23] method. The same standardized method was applied to perform the measurement on the pre- and postoperative images. Moreover, each observer was allowed to choose whatever gap and/or stepoff he or she thought was the maximum one before and after surgery. Measurements were performed with a digital tool (accuracy 0.1 mm) in the Carestream Vue Motion imaging system of the patient file and each observer measured the first five patients together with the responsible researcher. The maximum value of the gap and stepoff on the axial, coronal, and sagittal CT slices, as determined by each observer for each patient, was calculated and used to determine the inter- and intraobserver variability. All pelvic radiographs and CT images were automatically calibrated, so it was possible to measure distances in millimeters on the images. All CT scans were performed within 2 weeks postoperatively. Additionally, the definitive intraoperative fluoroscopy images (AP and Judet views after the osteosynthesis has been performed) were assessed for the presence of a gap and/or stepoff. Distance measurements on fluoroscopy images were not possible because the magnification of these images depends on the position of the c-arm. Five observers (FIJ, KtD, VS, RW, HB; two trauma surgeons with more than 5 years of

experience in pelvic surgery, two trauma surgeons with less than 5 years of experience in pelvic surgery, and one surgical resident) performed all measurements. Finally, all five observers repeated all measurements for 10 randomly selected patients at least 2 weeks after the first measurements were performed. For the measurements on radiographs and CT scans it was evaluated whether the observers agreed on the indication for surgery (a preoperative gap or stepoff of  $\geq 2$  mm) and on whether the postoperative reduction was acceptable (defined as a gap  $< 5$  mm or a stepoff  $< 1$  mm [24]).

## Statistical Analysis

Interobserver variability was defined as the difference in the measurements between observers. Intraobserver variability was defined as the difference in repeated measurements by the same observer. Both inter- and intraobserver variability were determined by using SPSS (version 23, IBM, Chicago, IL, USA). We used the intraclass correlation coefficient (ICC), including 95% CIs, in a two-way mixed, single-measurement model with absolute agreement. Additionally, we calculated the median difference between the measurements. This was the absolute median difference between all interobserver measurements. The median difference between intraobserver measurements was the absolute median difference between the first and second measurement by the same observer. The ICC is considered too low to be used with confidence to make clinical decisions when it is  $< 0.7$  [6]. The kappa value was calculated for the intraoperative fluoroscopy assessments using SPSS (version 23, IBM, Chicago, IL, USA). The 95% confidence interval of the kappa was calculated using bootstrapping (1000 samples). The agreement between observers is considered strong if the kappa value is above 0.80, substantial for a kappa value between 0.61 and 0.80, moderate for a kappa value between 0.41 and 0.60, and fair for a kappa value between 0.21 and 0.4, and there is no agreement when the kappa value is 0.20 or lower [11]. The ICC and kappa values were compared (based on whether the 95% CI includes the same range) between the two experienced surgeons and the three less-experienced observers, to see whether experience influences the agreement between observers. Finally, a subgroup analysis of different fracture types, specifically both-column fractures and fractures involving the posterior wall, was performed.



**Fig. 1** This pie chart shows the distribution of fracture types in our study population, according to the Letournel classification. A color image accompanies the online version of this article.

## Results

### Interobserver Agreement

For the interobserver measurements, the preoperative ICC was 0.4 (gap and stepoff) on radiographs and 0.4 (gap) and

0.3 (stepoff) on CT images (Table 1). This is too low to be used with confidence to make clinical decisions. Preoperatively, the observers agreed on whether or not there was an indication for surgery in 40% (24 of 60) of the patients (based on the gap) and 30% (18 of 60) of the patients (based on the stepoff) on pelvic radiographs. For CT scans the observers agreed on whether or not there was an indication for surgery in 95% (57 of 60) of the patients (based on the gap) and 70% (42 of 60) of the patients (based on the stepoff). Intraoperatively, there was no agreement in gap and stepoff detection on intraoperative fluoroscopy images; the kappa values were 0.1 (gap) and 0.2 (stepoff). The kappa values of the two experienced observers were 0.6 for the gap and stepoff, compared with 0.1 for the less-experienced observers. Postoperatively, the interobserver ICC was 0.4 (gap) and 0.2 (stepoff) on radiographs, which is too low to be used with confidence to make clinical decisions. The observers agreed on whether the reduction was acceptable or not in 60% (36 of 60) of patients (based on the gap) and in 40% (24 of 60) of patients (based on the stepoff) on radiographs. On CT scans the ICC was 0.3 (gap) and 0.4 (stepoff), which is also too low for clinical decisions. The observers agreed in 35% (21 of 60) of patients (based on the gap) and 38% (23 of 60) of patients (based on the stepoff). The ICC was only significantly different between experienced observers (0.7) and the three less-experienced observers (0.1) for the postoperative stepoff on radiographs.

**Intraobserver Agreement**

Considering intraobserver measurements, the ICC was 0.6 for the preoperative gap and stepoff on pelvic radiographs

and 0.4 (gap) and 0.6 (stepoff) for CT scans (Table 1), which is considered too low to be used to make clinical decisions. Preoperatively, the observers agreed with themselves on whether or not there was an indication for surgery in 80% (8 of 10) of patients (based on the gap) and 70% (7 of 10) of patients (based on the stepoff) on pelvic radiographs. For CT scans, the observers agreed on whether or not there was an indication for surgery in one 100% (10 of 10) of patients (based on the gap) and 90% (9 of 10) of patients (based on the stepoff). There was no agreement in gap and stepoff detection on intraoperative fluoroscopy images; the kappa values were 0 (gap) and -0.1 (stepoff) for the intraobserver assessments. The kappa value for the intraobserver agreement of the two experienced observers was good enough (0.6) for both the gap and stepoff to know that the fracture is adequately reduced. For the less-experienced observers, there was no agreement (kappa 0.0) in gap and stepoff detection on intraoperative fluoroscopy. Postoperatively, the ICC was 0.7 for the gap on pelvic radiographs, which is suitable for clinical use. It was 0.1 for the stepoff on pelvic radiographs. On CT, the ICC was 0.5 (gap) and 0.3 (stepoff), which is too low to be used to determine whether the fracture reduction is sufficient. The observers agreed with themselves on whether or not the reduction was acceptable in 90% (9 of 10) of patients (based on the gap and stepoff) on radiographs. On CT scans, they agreed in 90% (9 of 10) of patients (based on the gap) and 60% (6 of 10) patients (based on the stepoff). The ICC of all intraobserver measurements was not different based on experience (see Supplemental Digital Content 1, <http://links.lww.com/CORR/A394>). Additionally, the ICC of intraobserver measurements was higher than the ICC of interobserver measurements on both pelvic radiographs and CT images, except for postoperative stepoff (Table 1).

**Table 1.** Inter- and intraobserver measurements

Measurements			Interobserver			Intraobserver		
			ICC (95% CI)	Δ	IQR (mm)	ICC 95% CI	Δ	IQR (mm)
Preoperative	Gap	Radiographs	0.4 (0.3 to 0.6)	4	2 to 7	0.6 (-0.1 to 0.9)	3	1 to 8
		CT	0.4 (0.2 to 0.6)	8	5 to 12	0.4 (-0.1 to 0.9)	6	2 to 8
	Stepoff	Radiographs	0.4 (0.2 to 0.5)	2	0 to 3	0.6 (0.2 to 0.9)	4	1 to 8
		CT	0.3 (0.2 to 0.5)	5	3 to 9	0.6 (-0.5 to 0.6)	3	1 to 5
Postoperative	Gap	Radiographs	0.4 (0.3 to 0.5)	2	0 to 3	0.7 (-0.2 to 0.8)	2	0 to 2
		CT	0.3 (0.2 to 0.5)	4	2 to 6	0.5 (0 to 0.9)	3	1 to 4
	Stepoff	Radiographs	0.2 (0.1 to 0.4)	0	0 to 3	0.1 (0 to 0.9)	0	0 to 1
		CT	0.4 (0.2 to 0.5)	2	1 to 3	0.3 (-0.4 to 0.7)	3	2 to 4
			Kappa (95% CI)			Kappa (95% CI)		
Intraoperative	Gap	Fluoroscopy	0.1 (0 to 0.3)			0 (-0.4 to 0.6)		
	Stepoff		0.2 (-0.1 to 0.4)			-0.1 (-0.3 to 0)		

ICC = intraclass correlation coefficient; Δ = median difference; IQR = interquartile range.

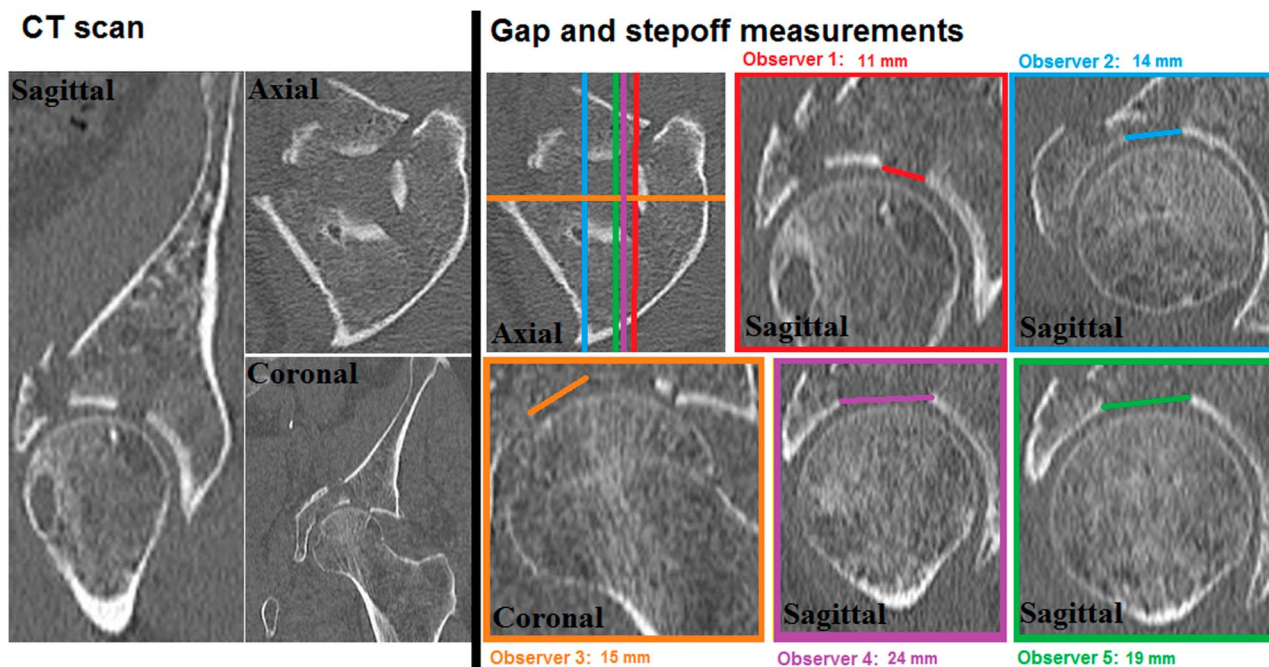
**Other Relevant Findings**

All observers were instructed to measure the maximum gap and stepoff in the weightbearing dome for each patient. In all 60 patients, observers chose different CT slices for performing the measurements, which resulted in differences in measured gaps and stepoff between the observers (Fig. 2). Also, repeated measurements, performed by the same observer, resulted in a selection of different CT slices. Moreover, when measuring on sagittal CT slices, the measurement turned out not to be perpendicular to the fracture line and thus did not represent the true gap in comparison with observing the fracture on an axial CT slice. The subgroup analysis of both-column fractures showed a preoperative interobserver ICC of 0.5 (gap) and 0.4 (stepoff) and a postoperative ICC of 0.4 (gap) and 0.2 (stepoff) on pelvic radiographs (see Supplemental Digital Content 2, <http://links.lww.com/CORR/A395>). On CT scans, the interobserver ICC was 0.5 (gap) and 0.3 (stepoff) preoperatively and 0.3 (gap) and 0.6 (stepoff) postoperatively. The interobserver ICC for the posterior wall was 0.0 to 0.1 on pre- and postoperative pelvic radiographs. On CT scans, the ICC was 0.6 (gap) and 0.0 (stepoff) preoperatively and 0.3 (gap) and 0.2 (stepoff) postoperatively. For combined posterior wall fractures, the

ICC was 0.2 for both the gap and stepoff preoperatively and 0.4 (gap) and 0.3 (stepoff) postoperatively on pelvic radiographs. For CT scans, the ICC was 0.4 (gap) and 0.1 (stepoff) preoperatively; postoperatively it was 0.3 (gap) and 0.2 (stepoff). These ICC values for the different fracture types were too low to be used to make clinical decisions.

**Discussion**

Gap and stepoff measurements are the main indications for acetabular fracture surgery and are commonly associated with prognosis. There is no agreement on how well observers agree when assessing the same radiograph or CT image. Because gap and stepoff measurements are important to support the indication for surgery and predict prognosis, in this study we focused on how reliably gaps and stepoffs could be measured. This study demonstrated a preoperative and postoperative interobserver agreement that was insufficient for making clinical decisions in both pelvic radiographs and CT scans. Furthermore, assessment of intraoperative fluoroscopy images revealed that there was no agreement between the observers on whether gaps and/or stepoffs were still present at the end of surgery.



**Fig. 2** The interobserver variability of gap measurements by the five observers was determined using a preoperative CT image of a patient with a T-shaped fracture. The left side shows sagittal, axial, and coronal CT slices of the fracture. The right side shows the differences in measurements between observers. The colored lines on the axial CT slice represent the slices that were selected by the observers to perform their gap measurements. Overall, every observer chose a different slice, represented in the colored boxes, which resulted in different gap measurements.

More-experienced observers had a higher level of agreement compared with less-experienced observers. Intraobserver reproducibility of measurement was highly variable as well.

### Limitations

The standardized CT-based measurement method was initially intended for analysis of residual displacement on CT scans. In this study, the same method was also applied for preoperative measurements, although it was not specifically designed for this. To our best knowledge, no CT-based preoperative measurement method has been described so far. No clear guideline exists on which CT-slice, plain, direction, fracture line or which part of the fracture should be measured. This caused all observers to measure where they thought the maximum gap or stepoff was situated, leading to different locations being chosen for the measurements that resulted in substantial differences between observers. Despite these limitations, all measurements were performed according to well-described standardized measurement techniques and our current clinical practice. Moreover, our study population consisted of a lot of both-column fractures and posterior wall fractures. These fractures are often hard to measure due to comminution, multiple fracture lines, and often severe displacement. Also, the differences in experience between observers might have contributed to less interobserver agreement. On the other hand, intraobserver agreement was limited as well, and differences in level of experience reflect clinical practice. Moreover, a slight change in gap or a different selection of the CT slices can affect the ICC. This is why we also looked at the agreement between observers in this study. For all patients, the observers reviewed the radiograph first and then they reviewed the CT. This may have caused the CT review to be biased by the radiograph. Yet, this represents the workflow in our current practice. Additionally, implants can obscure residual gaps or stepoffs on postoperative images. Metal artefacts might be mistaken for bone and could influence measurements, although an iterative metal artefact reduction algorithm is usually used for postoperative CT scans. However, in this study all observers have been exposed to these limitations of imaging modalities, and we do not think this will influence our overall conclusion.

Measurements of gaps and stepoffs were barely reproducible between observers in our study. Our results are not consistent with the results of two older studies using a standardized measurement technique that showed excellent interobserver agreement for gap and stepoff detection on pelvic radiographs and CT scans [4, 16]. However, the results are not comparable because former studies focused on the presence instead of the size of gaps and stepoffs. A

more recent study reported a moderate interobserver ICCs using postoperative pelvic radiographs and a fair-to-good ICC (0.7 for gap and 0.5 for stepoff) when using a standardized CT measurement method for assessment of the postoperative residual displacement [23]. Our postoperative ICCs for pelvic radiographs are in line with theirs, but our ICCs for the postoperative CT images are lower (0.3 versus 0.7 for gap and 0.4 versus 0.5 for stepoff) despite using the same standardized measurement method. This can be explained by the differences in sample size (60 versus 40), fracture types (more both-column and posterior wall fractures in our study) and different levels of experience of the observers between studies. Furthermore, one might expect a high agreement on fluoroscopy images, because the operating surgeons agreed with the reduction, though our data showed that there was no agreement between the observers on whether a clinically significant gap or stepoff was present on fluoroscopy images. Yet, this reflects current practice because the postoperative reduction is not always acceptable [8, 25]. Norris et al. [15] conducted one of the few studies that evaluated the use of intraoperative fluoroscopy to assess reduction of acetabular fractures two decades ago. They found that intraoperative fluoroscopy and postoperative radiographs can both be used to evaluate acetabular fracture reductions. However, they did not investigate interobserver and intraobserver variability nor were CT scans used for all patients. Former studies showed that the reliability of other commonly used fracture classifications is low [14, 22, 28], indicating that guidelines are needed for classifications or measurements to be used. Therefore, we also investigated whether the observers agreed on the indication for surgery preoperatively and whether the reduction was acceptable or not postoperatively. Even here, considering a preoperative gap or stepoff of  $\geq 2$  mm as an indication for surgery, the observers agreed on whether there was an indication for surgery in only 30% (stepoff) to 40% (gap) of the patients based on the preoperative radiographs and 70% (stepoff) to 95% (gap) based on CT scans. Yet, our study population only included surgically treated patients with mostly substantial initial fracture displacement. Proceeding to surgery has major implications for the patient. Postoperatively, the observers agreed on whether the reduction was adequate in 40% (stepoff) to 60% (gap) of patients based on radiographs and in 35% (gap) to 38% (stepoff) of patients based on CT scans. This is worrisome because residual displacement is often correlated with prognosis.

Measurements of gaps and stepoffs were barely reproducible between repeated measurements by the same observer, especially the postoperative measurements. Limited data is available on intraobserver reliability in measuring acetabular fractures. A study on intraobserver reliability of Letournel's acetabular fracture classification showed a high agreement, but only for the highly trained

experts in the field [3]. Our results also show that more-experienced surgeons have a higher kappa value.

## Conclusions

We found that the level of agreement on the assessment of acetabular fractures between observers, or between repeated measurements by a single observer, was too low to provide any level of confidence in clinical decision making. This calls into question a substantial proportion of the existing literature, which often attempts to correlate initial or residual displacement with the chance of progressive arthritis or further surgery, such as THA. Considering the known lack of agreement between observers for commonly used classification systems, our findings indicate the need for substantial caution when depending on arbitrary radiographic findings. If this literature will be used for making clinical decisions, surgeons must be clearly informed on how measurements were performed, who performed them, and at what time during the clinical course they were done. Improvements in advanced imaging, such as using an objective three-dimensional fracture assessment tool [12], will be crucial to resolve this problem. Further studies are needed to determine how indications for surgery and prognosis after surgery can be modified so that these are more reliable and independent of the observers.

**Acknowledgments** We thank I.H.F. Reininga PhD, for her contribution to the statistical analysis.

## References

- Alonso JE, Kellam JF, Tile M. Pathoanatomy and classification of acetabular fractures. In: Tile M, Helfet D, Kellam J, Vrahas M, eds. *Fractures of the Pelvis and Acetabulum - Principles and Methods of Management*. Georg Thieme Verlag, Stuttgart, Germany and Thieme New York; 2015:447–470.
- Archdeacon MT, Dailey SK. Efficacy of routine postoperative CT scan after open reduction and internal fixation of the acetabulum. *J Orthop Trauma*. 2015;29:354–358.
- Beaulé PE, Dorey FJ, Matta JM. Letournel Classification for Acetabular Fractures. *J Bone Joint Surg Am*. 2003;85:1704–1709.
- Borrelli J, Goldfarb C, Catalano L, Evanoff BA. Assessment of articular fragment displacement in acetabular fractures: A comparison of computerized tomography and plain radiographs. *J Orthop Trauma*. 2002;16:449–456.
- Borrelli J, Ricci WM, Steger-May K, Totty WG, Goldfarb C. Postoperative radiographic assessment of acetabular fractures: a comparison of plain radiographs and CT scans. *J Orthop Trauma*. 2005;19:299–304.
- Cicchetti D V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1993;6:284–290.
- Garrett J, Halvorson J, Carroll E, Webb LX. Value of 3-D CT in classifying acetabular fractures during orthopedic residency training. *Orthopedics*. 2012;35:e615–e620.
- Jaskolka DN, Di Primio GA, Sheikh AM, Schweitzer ME. CT of preoperative and postoperative acetabular fractures revisited. *J Comput Assist Tomogr*. 2014;38:344–347.
- Jouffroy P, Sebaaly A, Aubert T, Riouallon G. Improved acetabular fracture diagnosis after training in a CT-based method. *Orthop Traumatol Surg Res*. 2017;103:325–329.
- Letournel E, Judet R. Classification. In: *Fractures of the acetabulum*. Springer Verlag Berlin Heidelberg New York; 1993: 63–66.
- Mchugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2019;22:276–282.
- Meesters AML, Kraeima J, Banierink H, Slump CH, de Vries JPPM, ten Duis K, Witjes MJH, IJpma FFA. Introduction of a three-dimensional computed tomography measurement method for acetabular fractures. *PLoS ONE*. 2019:1–11.
- Meinberg EG, Agel J, Roberts CS, Karam MD, Kellam JF. Fracture and dislocation classification compendium - 2018. *J Orthop Trauma*. 2018;32:S1–S170.
- Mellema JJ, Doornberg JN, Molenaars RJ, Ring D, Kloen P, Babis GC, Jeray KJ, Prayson MJ, Pesantez R, Acacio R, et al. Interobserver reliability of the Schatzker and Luo classification systems for tibial plateau fractures. *Injury*. 2016;47:944–949.
- Norris BL, Hahn DH, Bosse MJ, Kellam JF, Sims SH. Intraoperative fluoroscopy to evaluate fracture reduction and hardware placement during acetabular surgery. *J Orthop Trauma*. 1999;13:414–417.
- O’Shea K, Quinlan J, Waheed K, Brady O. The Usefulness of Computed Tomography following Open Reduction and Internal Fixation of Acetabular Fractures. *J Orthop Surg*. 2006;14: 127–132.
- Polesello GC, Nunes MAA, Azuaga TL, De Queiroz MC, Honda EK, Ono NK. Comprehension and reproducibility of the judet and letournel classification. *Acta Ortop Bras*. 2012;20:70–74.
- Prevezas N, Antypas G, Louverdis D, Konstas A, Papatirou A, Sbonias G. Proposed guidelines for increasing the reliability and validity of Letournel classification system. *Injury*. 2009;40: 1098–1103.
- Riouallon G, Sebaaly A, Upex P, Zarea M, Jouffroy P. A New, Easy, Fast, and Reliable Method to Correctly Classify Acetabular Fractures According to the Letournel System. *JBJS Open Access*. 2018:e0032:1–8.
- Sebaaly A, Riouallon G, Zarea M, Upex P, Marteau V, Jouffroy P. Standardized three dimensional computerised tomography scanner reconstructions increase the accuracy of acetabular fracture classification. *Int Orthop*. 2018;42:1957–1965.
- Sinatra PM, Moed BR. CT-generated Radiographs in Obese Patients With Acetabular Fractures: Can They Be Used in Lieu of Plain Radiographs? *Clin Orthop Relat Res*. 2014;472: 3362–3369.
- Teo TL, Schaeffer EK, Habib E, Cherukupalli A, Cooper AP, Aroojis A, Sankar WN, Upasani V V., Carsen S, Mulpuri K, Reilly C. Assessing the reliability of the modified gartland classification system for extension-type supracondylar humerus fractures. *J Child Orthop*. 2019;13:569–574.
- Verbeek DO, van der List JP, Moloney GB, Wellman DS, Helfet DL. Assessing postoperative reduction following acetabular fracture surgery: a standardized digital CT-based method. *J Orthop Trauma*. 2018;32:e284–e288.
- Verbeek DO, van der List JP, Tissue CM, Helfet DL. Predictors for Long-Term Hip Survivorship Following Acetabular Fracture Surgery - Importance of Gap Compared with Step Displacement. *J Bone Joint Surg Am*. 2018;100:922–929.
- Verbeek DO, List JP Van Der, Villa JC, Wellman DS, Helfet DL. Postoperative CT is superior for acetabular fracture reduction

- assessment and reliably predicts hip survivorship. *J Bone Joint Surg Am.* 2017;99:1745–1752.
26. Visutipol B, Chobtangsin P, Ketmalasiri B, Pattarabanjird N, Varodompun N. Evaluation of Letournel and Judet classification of acetabular fracture with plain radiographs and three-dimensional computerized tomographic scan. *Int J Orthop.* 2000;8:33–37.
27. Zhang R, Yin Y, Li A, Wang Z, Hou Z, Zhuang Y, Fan S, Wu Z, Yi C, Lyu G, Ma X, Zhang Y. Three-Column Classification for Acetabular Fractures - Introduction and Reproducibility Assessment. *J Bone Joint Surg Am.* 2019;101:2015–2025.
28. Zhu Y, Hu CF, Yang G, Cheng D, Luo CF. Inter-observer reliability assessment of the Schatzker, AO/OTA and three-column classification of tibial plateau fractures. *J Trauma Manag Outcomes.* 2013;7:1–4.