



Interobserver reliability of radiographic assessment after radial head arthroplasty

Rens Bexkens^{1,2,3}, Femke M. A. P. Claessen¹,
Izaäk F. Kodde^{1,2}, Luke S. Oh³, Denise Eygendaal^{1,2}
and Michel P. J. van den Bekerom⁴; on behalf of the
Shoulder and Elbow Platform*

Abstract

Background: Radiographs are part of routine clinical care after radial head arthroplasty (RHA). Therefore, the aim of this diagnostic study was to assess the interobserver reliability of radiographic assessment following RHA.

Methods: Anteroposterior (AP) and lateral radiographs of 24 consecutive patients who underwent press-fit bipolar RHA were evaluated with respect to 14 parameters by 14 orthopaedic surgeons specializing in the elbow: shaft loosening (AP, lateral), subcollar bone resorption, nonbridging heterotopic ossification, capitellar erosion, capitellar osteopenia, implant size, ulnohumeral joint gapping, ulnohumeral joint degeneration, proximal radio-ulnar joint congruency, stem size, stem positioning (AP, lateral) and component dissociation or polyethylene wear of the head with increased angulation. Observer agreement was evaluated using the multirater kappa (κ) measure.

Results: Nine of 14 parameters had poor interobserver agreement [$\kappa = 0.0$ to 0.20 , confidence interval (CI) = 0.0 to 0.31]. Four parameters had fair agreement: subcollar bone resorption ($\kappa = 0.27$, CI = 0.12 to 0.40), capitellar erosion ($\kappa = 0.30$, CI = 0.20 to 0.40), ulnohumeral joint degeneration ($\kappa = 0.35$, CI = 0.22 to 0.51) and stem positioning in AP view ($\kappa = 0.24$, CI = 0.14 to 0.36). One parameter had moderate agreement: nonbridging heterotopic ossification ($\kappa = 0.47$, CI = 0.31 to 0.64).

Conclusions: The overall interobserver reliability of radiographic assessment following press-fit bipolar RHA was poor among experienced elbow surgeons. Therefore, radiographic evaluation after RHA should be interpreted with caution when making treatment decisions.

Keywords

arthroplasty, evaluation, interobserver, radial head, radiographic assessment, reliability

Date received: 24th January 2017; accepted: 7th June 2017

Introduction

Radial head arthroplasty (RHA) is indicated for selected displaced or comminuted radial head fractures, especially in unstable elbows or when reduction cannot be accomplished with open reduction and internal fixation.^{1–5} RHA is also performed in patients with persistent post-traumatic elbow symptoms, including nonunion and malunion of the radial head, and symptomatic elbow instability following previous excision of the radial head.^{3,4,6–8}

Clinical and radiographic follow-up is important after RHA because loosening of the prosthesis, osteolysis, erosion of the capitellum, and implant failure can affect long-term outcomes.^{1,9,10} Conventional radiographs have been widely used in orthopaedic practice to monitor

¹Department of Orthopaedic Surgery, Academic Medical Center, Amsterdam, The Netherlands

²Department of Orthopaedic Surgery, Amphia Hospital, Breda, The Netherlands

³Department of Orthopaedic Surgery, Sports Medicine Service, Massachusetts General Hospital, Boston, MA, USA

⁴Department of Orthopaedic Surgery, Shoulder and Elbow Unit, Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands

*A list of the collaborating authors is provided at the end of the article.

Corresponding author:

Rens Bexkens, Department of Orthopaedic Surgery, Academic Medical Center, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands.
Email: rensbexkens@gmail.com

adverse events after RHA.^{1,9,10} Several studies have demonstrated that radiographic abnormalities (e.g. heterotopic ossification, subcollar bone resorption, and signs of implant loosening) result in poor clinical outcome in RHA.^{9–11} Ha et al.⁹ reported an association between radiographic abnormalities and the presence of symptoms (e.g. pain, range of motion and instability). Additionally, a significant association between heterotopic ossification and restricted range of motion was found by Rotini et al.¹⁰ at 2-year follow-up.

Ideally, radiographic assessment would be reliable and consistent among surgeons because it is widely used for follow-up of RHA and clinical research purposes. However, interpretation of radiographs may vary from surgeon to surgeon,^{12–14} and thus it would be useful to determine whether radiographic assessment for signs of malpositioning and loosening in RHA is reliable among experienced elbow surgeons. To our knowledge, this has not yet been investigated.

The purpose of the present study was to evaluate the interobserver reliability of radiographic assessment following press-fit bipolar radial head arthroplasty. We hypothesized that there would be moderate interobserver agreement among experienced elbow surgeons for the evaluation of 14 radiographic parameters in press-fit bipolar RHA.

Materials and methods

Orthopaedic surgeons specializing in elbow surgery from several countries were invited to participate in this interobserver study. They were asked to evaluate postoperative radiographs of 24 consecutive patients who had press-fit bipolar radial head arthroplasty (RHS[®]; Tornier, Montbonnot-Saint-Martin, France). Indications for RHA were post-traumatic symptoms following a radial head fracture, such as persistent pain, restricted range of motion and instability. The mean (SD) radiographic follow-up was 27 (10) months. The right elbow was treated in 13 patients and the left elbow in 11 patients.

Plain anteroposterior (AP) and lateral radiographs of the elbow were available for each patient. A research fellow not involved in patient care removed all identifying information from the radiographs and uploaded the Digital and Communications in Medicine (DICOM) files to a web-based study platform (www.shoulderelbowplatform.com). Observers evaluated radiographs using a built-in DICOM viewer and were able to adjust brightness, contrast, window leveling and zoom in. All questions related to one case had to be completed to proceed to the next case. The observers completed the study at their own pace and in their own time on various computers, if necessary.

A total of 20 orthopaedic surgeons with elbow experience, members of the Shoulder & Elbow Platform, were

invited to participate in this interobserver study via an invitation e-mail that included a short study description. Upon website log-in, observers were asked about demographics and professional information: sex (male or female), location of practice (Europe, USA & Canada, or other), years in practice (1 to 5, 6 to 10, 11 to 15 or ≥ 16), involvement in training residents or fellows (yes or no) and number of RHAs performed per year (1 to 5, 6 to 10, 11 to 15 or > 15). Subsequently, observers were asked to evaluate AP and lateral radiographs of 24 patients on 14 parameters,^{9,10,15–19,25} which were related to implant size, positioning of the stem, signs of loosening, radiocapitellar degeneration, nonbridging heterotopic ossification, capitellar erosion and radioulnar congruency, amongst others. An overview of all radiographic parameters is shown in Table 1.

This retrospective study was approved by the Medical Ethics Committee of our institution. Data were collected as part of routine clinical care and each patient consented that their patient data could be used for scientific purposes.

Statistical analysis

Agreement among observers was calculated by using a multirater kappa (k), as described by previously.^{20–22} Point estimates and two-sided 95% confidence intervals (CIs) were calculated for each radiographic parameter. The multi-rater kappa is a commonly used statistic to describe chance-corrected agreement in various interobserver studies.^{20–22} A value of 0 indicates no agreement beyond chance alone. A value of -1.00 indicates total disagreement and $+1.00$ represents total agreement.^{21,22} A value of 0.01 to 0.20 is defined as poor agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and more than 0.81, almost perfect agreement.^{21,22} Subgroups (number of years in practice and number of RHAs performed per year) were compared using the Z-test.^{23,24} Multi-rater kappas were calculated with use of Stata, version 12.0 (StataCorp LP, College Station, TX, USA).

Results

Participants

Fourteen of the 20 invited members (70%) of the Shoulder & Elbow Platform completed the online interobserver study. Most observers in this interobserver study were from Europe (71%), followed by Asia (21%). Twelve of 14 observers (86%) were in practice for ≥ 6 years. Seven surgeons (50%) performed more than five RHAs per year; the remaining group performed between one and five RHAs annually. Detailed demographics are reported in Table 2.

Table 1. Radiographic assessment after radial head arthroplasty ($n = 14$).

Parameter	Response
1. Any signs of loosening of the shaft (AP)?	No, zone 1, 2, 3, 4, 5, 6, or 7 ²⁵
Sub-analysis: any signs of loosening?	Yes or no
2. Any signs of loosening of the shaft (lateral)?	No, zone 1, 2, 3, 4, 5, 6, or 7 ²⁵
Sub-analysis: any signs of loosening?	Yes or no
3. Any signs of subcollar bone resorption?	Yes or no
4. Any signs of nonbridging heterotopic ossification?	Yes or no
5. Any signs of erosion of capitellum?	Yes or no
6. Any signs of osteopenia of capitellum?	Yes or no
7. Is the radial head implant size correct?	Too large, too small, or just right
8. Is there gapping of the ulnohumeral joint?	Yes or no
9. Any signs of degeneration of ulnohumeral joint?	Yes or no
10. Is the proximal radio-ulnar joint congruent?	Yes or no
11. Is the stem size correct?	Too large, too small, or just right
12. Is the stem positioning correct (AP)?	Malpositioned, heading into the radial tuberosity, or right positioned
13. Is the stem positioning correct (lateral)?	Malpositioned, heading into the radial tuberosity, or right positioned
14. Any component dissociation or polyethylene wear of the head with increased angulation of the head in relation to the shaft?	Yes or no

AP, anteroposterior.

Interobserver agreement

Observers had poor interobserver agreement in nine of 14 evaluated radiographic parameters ($\kappa = 0.0$ to 0.20 , $CI = 0.0$ to 0.31). Two examples of parameters with poor agreement (i.e. implant size and loosening of the shaft) are shown in Figs 1 and 2. There was fair agreement in four parameters: signs of subcollar bone resorption ($\kappa = 0.27$, $CI = 0.12$ to 0.40), signs of erosion of the capitellum ($\kappa = 0.30$, $CI = 0.20$ to 0.40), signs of degeneration of the ulnohumeral joint ($\kappa = 0.35$, $CI = 0.22$ to 0.51) and stem positioning in AP view ($\kappa = 0.24$, $CI = 0.14$ to 0.36). There was moderate agreement for assessment of nonbridging heterotopic ossification ($\kappa = 0.47$, $CI = 0.31$ to 0.64) (Fig. 3). Additional analysis performed by dichotomizing the responses of shaft loosening (parameters 1 and 2) into ‘loosening present’ or ‘loosening not present’ did not improve interobserver agreement. A detailed summary for each parameter is shown in Table 3.

For each of the 14 parameters, subgroup analysis showed that there was no difference in interobserver agreement between surgeons who performed more than five RHAs per year and surgeons who performed between one and five RHAs annually ($p > 0.05$). For parameter 5 (presence of capitellar erosion) and parameter 14 (component dissociation or polyethylene wear of the head with increased angulation of the head in relation to the shaft), there was greater agreement among surgeons who were in practice for ≤ 10 years compared to surgeons who were in practice for > 10 years ($p < 0.001$). No difference in agreement was found for the remaining 12 parameters ($p > 0.05$).

Discussion

Plain radiographs are routinely performed in follow-up of radial head arthroplasty. It is important to determine whether radiographic assessment is reliable among

Table 2. Observer demographics ($n = 14$).

	<i>n</i> (%)
Sex	
Male	11 (79)
Female	3 (21)
Location of practice	
Europe	11 (71)
Asia	3 (21)
USA & Canada	1 (8)
Years in practice	
1 to 5	2 (14)
6 to 10	7 (50)
11 to 15	2 (14)
> 16	3 (22)
Supervises resident or fellow	12 (86)
Radial head arthroplasty cases per year	
1 to 5	7 (50)
6 to 10	3 (21)
11 to 15	3 (21)
> 16	1 (8)

surgeons. The present study shows that the interobserver agreement for radiographic evaluation after placement of a press-fit bipolar radial head prosthesis is poor, even among experienced elbow surgeons. Only one of 14 parameters (i.e. nonbridging heterotopic ossification) showed moderate interobserver agreement. Fair agreement was found in four parameters: signs of degeneration of ulnohumeral joint, signs of capitellar erosion, signs of subcollar bone resorption and stem positioning in AP view. Interobserver agreement was poor for the remaining parameters.

No direct comparison can be made because RHA is a fairly new procedure and there are no studies that have investigated the reliability of radiographic assessment following such procedures. Nevertheless, similar values were reported in previous studies that evaluated the reliability of radiographic assessment in total hip arthroplasty.^{12–14} Smith et al.¹⁴ found limited interobserver agreement ($k = 0.26$) in an assessment of radiolucency and loosening in the Gruen zones; a system of dividing the interface between the femoral component and femur in seven areas.²⁵ Muir and colleagues described comparable kappa values ($k = 0.24 - 0.41$) in an evaluation of the Engh Grading Scale; a scale that is widely used in the follow-up of uncemented total hip arthroplasty.¹³ No difference in interobserver variability was seen between different specialties: two arthroplasty surgeons, a senior orthopaedic resident, and a radiologist.¹³ In our study, for two parameters (component dissociation or polyethylene wear of the head with increased angulation of the head in relation to the shaft), there was greater interobserver agreement among surgeons who were in practice for ≤ 10 years in comparison with surgeons in

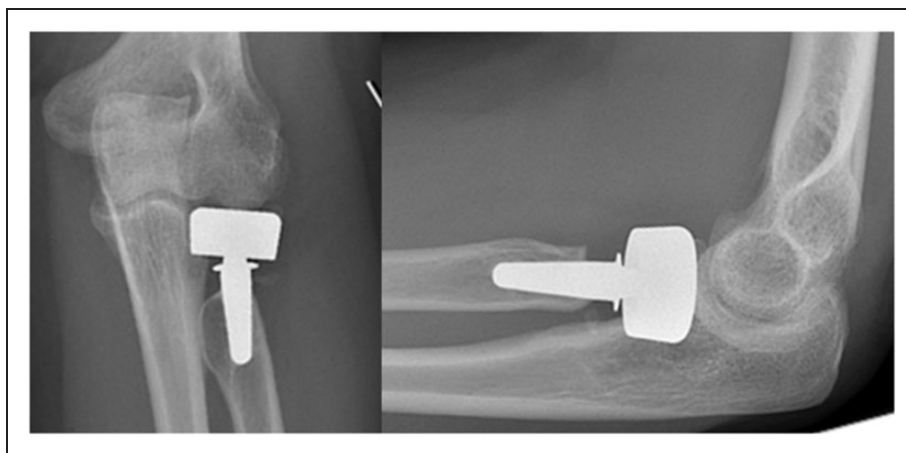


Figure 1. Radiographs of a 55-year-old female who underwent radial head arthroplasty after a comminuted radial head fracture and dislocation. Observers disagreed on radial head implant size (diameter); eight observers assessed implant size as just right and six observers assessed implant size as too large.

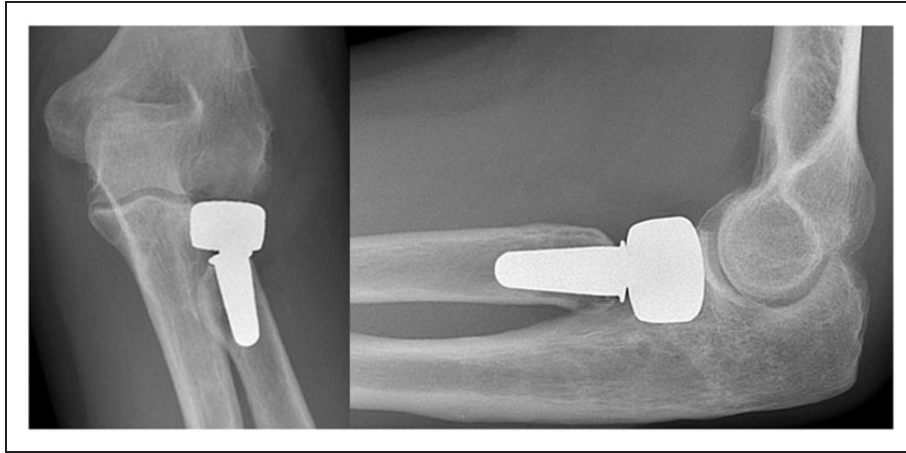


Figure 2. Radiographs of a 48-year-old female who underwent radial head arthroplasty for post-traumatic symptoms following radial head excision and a fracture of the proximal ulna that was previously treated operatively. Observers disagreed on signs of loosening of the shaft; six out of 14 observers assessed shaft loosening in anteroposterior view and 10 out of 14 assessed shaft loosening in lateral view.

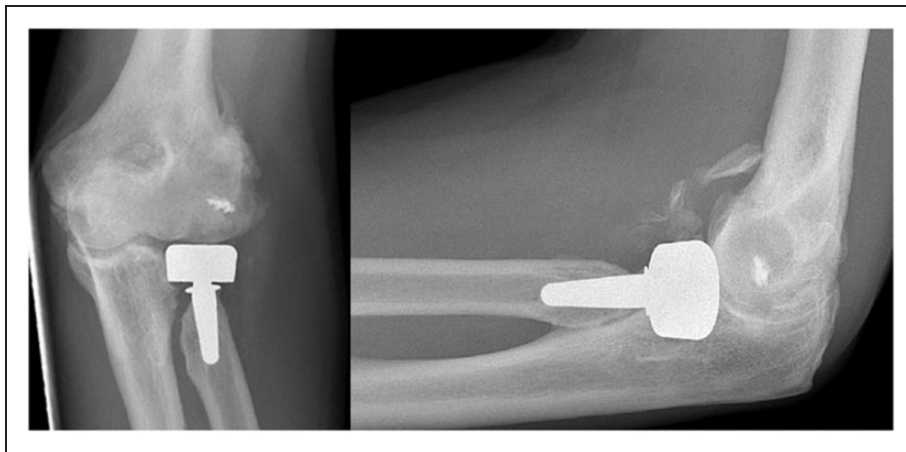


Figure 3. Radiographs of a 56-year-old man who underwent radial head arthroplasty and lateral collateral ligament repair after a malunion following a comminuted radial head fracture. All observers agreed on presence of nonbridging heterotopic ossification.

practice > 10 years. Because this was only the case in two parameters and the fact that the interobserver agreement was still only fair at best, we consider that no hard conclusions can be drawn from the subgroup analysis performed in the present study.

Limited interobserver reliability of radiographic assessment may be attributable to its subjective nature and the fact that most parameters require a categorical response. For example, in our study, to indicate whether the stem size is correct, observers had to fill in whether the stem was 'too large', 'too small' or 'just right'. Similarly, to assess gapping of the ulnohumeral joint, observers responded with either 'yes' or 'no'. We hypothesize that replacing categorical parameters (qualitative) with continuous parameters (quantitative) may improve interobserver reliability; however,

Al-Ahaideb et al.¹² found only slightly greater reliability after adding a quantitative component in radiographic evaluation of total hip arthroplasty. Although the overall interobserver agreement was poor in our study, it is important to note that agreement for nonbridging heterotopic ossification was moderate ($\kappa = 0.47$).

The findings of the present study have implications for radiographic assessment in the clinical setting and its use in research. Radiographic evaluation after RHA is a routine part of follow-up surveillance in orthopaedic practice. However, based on the inconsistency among observers in the evaluation of 24 consecutive cases, one should be cautious when interpreting post-operative radiographs and one should not solely rely on them. Also, one should take into consideration the overall poor interobserver agreement when

Table 3. Interobserver agreement of radiographic assessment after radial head arthroplasty ($n = 14$).

Parameter	Category	Kappa	95% CI
1. Any signs of loosening of the shaft (AP)?	Poor	0.0051	0.016 to 0.035
Sub-analysis: any signs of loosening?	Poor	0.0094	0.025 to 0.057
2. Any signs of loosening of the shaft (lateral)?	Poor	0.043	0.018 to 0.16
Sub-analysis: any signs of loosening?	Poor	0.032	0.043 to 0.14
3. Any signs of subcollar bone resorption?	Fair	0.27	0.12 to 0.40
4. Any signs of non-bridging heterotopic ossification?	Moderate	0.47	0.31 to 0.64
5. Any signs of erosion of capitellum?	Fair	0.30	0.20 to 0.40
6. Any signs of osteopenia of capitellum?	Poor	0.17	0.098 to 0.26
7. Is the radial head implant size correct?	Poor	0.10	0.033 to 0.20
8. Is there gapping of the ulnohumeral joint?	Poor	0.16	0.06 to 0.31
9. Any signs of degeneration of ulnohumeral joint?	Fair	0.35	0.22 to 0.51
10. Is the proximal radio-ulnar joint congruent?	Poor	0.11	0.013 to 0.25
11. Is the stem size correct?	Poor	0.14	0.0070 to 0.34
12. Is the stem positioning correct (AP)?	Fair	0.24	0.14 to 0.36
13. Is the stem positioning correct (lateral)?	Poor	0.18	0.056 to 0.31
14. Any component dissociation or polyethylene wear of the head with increased angulation of the head in relation to the shaft?	Poor	0.049	0.009 to 0.14

AP, anteroposterior.

using multiple observers for radiographic assessment in longitudinal follow-up studies of RHA. Future studies should focus on standardizing radiographic assessment following RHA. Precise and universal definitions and terminology for radiographic parameters need to be developed to improve interobserver agreement among surgeons. Furthermore, training in radiographic assessment is needed because RHA is a relatively new procedure, and not intimately familiar to a lot of surgeons. Future studies should also investigate the observer agreement for different designs of radial head prostheses, with and without the use of cement.

Some limitations of the present study should be taken into consideration. First, all patients in the present study had a press-fit bipolar radial head prosthesis. Therefore, our findings may not be applicable to other types of ingrowth radial head prostheses or cases in which cement was utilized. Second, none of the observers were specifically trained prior to participation in the present study. Interobserver agreement may have been greater if the observers were educated how to

evaluate plain radiographs. Third, to date, none of the 24 consecutive patients needed subsequent surgery at mean follow-up of 27 months. This may be the reason that only a few obvious radiographic abnormalities could be observed.

Conclusions

The present study demonstrates overall poor interobserver reliability of radiographic assessment in press-fit bipolar radial head arthroplasty among surgeons with elbow expertise. In a clinical or research setting, caution is warranted when interpreting postoperative radiographs in RHA.

Acknowledgments

Collaborating authors: A. L. Boerboom, D. Eygendaal, D. F. van Deurzen, R. J. Blokzijl, A. Heijink, P. Mansat, L. S. Oh, M. P. J. van den Bekerom, A. R. Kachooei, M. Nishiwaki, K. Ochi, E. E. J. Raven, I. B. Schipper, M. Ostendorf and M. Clerici. Website: www.shoulderelbowplatform.com

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical Review and Patient Consent

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. For this type of study, formal consent is not required. Ethics Committee Review (Amphia Hospital, Breda, The Netherlands) was not required for the present study according to Dutch law because it was a retrospective review that used anonymous data.

References

1. Popovic N, Lemaire R, Georis P and Gillet P. Midterm results with a bipolar radial head prosthesis: radiographic evidence of loosening at the bone-cement interface. *J Bone Joint Surg Am* 2007; 89: 2469–2476.
2. Flinkkila T, Kaisto T, Sirnio K, Hyvonen P and Leppilahti J. Short- to mid-term results of metallic press-fit radial head arthroplasty in unstable injuries of the elbow. *J Bone Joint Surg Br* 2012; 94: 805–810.
3. Chapman CB, Su BW, Sinicropi SM, Bruno R, Strauch RJ and Rosenwasser MP. Vitallium radial head prosthesis for acute and chronic elbow fractures and fracture-dislocations involving the radial head. *J Shoulder Elbow Surg* 2006; 15: 463–473.
4. Kodde IF, Heijink A, Kaas L, Mulder PG, van Dijk CN and Eygendaal D. Press-fit bipolar radial head arthroplasty, midterm results. *J Shoulder Elbow Surg* 2016; 25: 1235–1242.
5. Heijink A, Kodde IF, Mulder PG, et al. Radial head arthroplasty: a systematic review. *JBJS Rev* 2016; 4:pil: 01874474-201610000-00001.
6. Shore BJ, Mozzon JB, MacDermid JC, Faber KJ and King GJ. Chronic posttraumatic elbow disorders treated with metallic radial head arthroplasty. *J Bone Joint Surg Am* 2008; 90: 271–280.
7. Berschback JC, Lynch TS, Kalainov DM, Wysocki RW, Merk BR and Cohen MS. Clinical and radiographic comparisons of two different radial head implant designs. *J Shoulder Elbow Surg* 2013; 22: 1108–1120.
8. Brinkman JM, Rahusen FT, de Vos MJ and Eygendaal D. Treatment of sequelae of radial head fractures with a bipolar radial head prosthesis: good outcome after 1-4 years follow-up in 11 patients. *Acta Orthop* 2005; 76: 867–872.
9. Ha AS, Petscavage JM and Chew FS. Radial head arthroplasty: a radiologic outcome study. *Am J Roentgenology* 2012; 199: 1078–1082.
10. Rotini R, Marinelli A, Guerra E, Bettelli G and Cavaciocchi M. Radial head replacement with unipolar and bipolar SBi system: a clinical and radiographic analysis after a 2-year mean follow-up. *Musculoskeletal Surg* 2012; 96(Suppl 1): S69–S79.
11. Bowman SH, Barfield WR, Slone HS, Shealy GJ and Walton ZJ. The clinical implications of heterotopic ossification in patients treated with radial head replacement for trauma: a case series and review of the literature. *J Orthopaedics* 2016; 13: 272–277.
12. Al-Ahaideb A, Muir SW, Huckell J, et al. Interobserver reliability of the radiographic assessment of cement fixation in total hip arthroplasty. *Eur J Orthop Surg Trauma* 2013; 23: 889–894.
13. Muir SW, Al-Ahaideb A, Huckell J, Johnson MA, Johnston DB and Beaupre LA. Radiographic assessment of uncemented total hip arthroplasty: reliability of the Engh Grading Scale. *Canadian J Surg* 2011; 54: 185–188.
14. Smith TO, Williams TH, Samuel A, Ogonda L and Wimhurst JA. Reliability of the radiological assessments of radiolucency and loosening in total hip arthroplasty using PACS. *Hip Int* 2011; 21: 577–582.
15. Birkedal JP, Deal DN and Ruch DS. Loss of flexion after radial head replacement. *J Shoulder Elbow Surg* 2004; 13: 208–213.
16. Van Glabbeek F, Van Riet RP, Baumfeld JA, et al. Detrimental effects of overstuffing or understuffing with a radial head replacement in the medial collateral-ligament deficient elbow. *J Bone Joint Surg Am* 2004; 86A: 2629–2635.
17. Van Glabbeek F, van Riet RP, Baumfeld JA, et al. The kinematic importance of radial neck length in radial head replacement. *Med Engineering Physics* 2005; 27: 336–342.
18. Van Riet RP, Van Glabbeek F, Verborgt O and Gielen J. Capitellar erosion caused by a metal radial head prosthesis. A case report. *J Bone Joint Surg Am* 2004; 86A: 1061–1064.
19. Kim HM, Roush EP and Kiser C. Intraoperative fluoroscopic assessment of proper prosthetic radial head height. *J Shoulder Elbow Surg* 2016; 25: 1874–1881.
20. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; 70: 213–220.
21. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.
22. Posner KL, Sampson PD, Caplan RA, Ward RJ and Cheney FW. Measuring interrater reliability among multiple raters: an example of methods for nominal data. *Stat Med* 1990; 9: 1103–1115.
23. Claessen FM, van den Ende KI, Doornberg JN, Guitton TG, Eygendaal D and van den Bekerom MP. Osteochondritis dissecans of the humeral capitellum: reliability of four classification systems using radiographs and computed tomography. *J Shoulder Elbow Surg* 2015; 24: 1613–1618.
24. Guitton TG and Ring D. Interobserver reliability of radial head fracture classification: two-dimensional compared with three-dimensional CT. *J Bone Joint Surg Am* 2011; 93: 2015–2021.
25. Gruen TA, McNeice GM and Amstutz HC. ‘Modes of failure’ of cemented stem-type femoral components: a radiographic analysis of loosening. *Clin Orthop Rel Res* 1979; 17–27.