ORTHOPAEDIC SURGERY

Limited reliability of grading scapular notching according to Nerot–Sirveaux on anteroposterior radiographs

Marco M. Schneider[1,3] · Felix Toft[1] · Christoph Kolling[1,2] · Barbara Wirth[1] · Robert Vachenauer[1] · Nils Horn[1] · Quinten T. M. Felsch[1] · Laurent Audigé[1,2]

Abstract

Introduction Incidences of scapular notching (SN) in reverse shoulder arthroplasty (RSA) range between 0% and 96%, and controversy remain as to its relevance for patient outcome. We assessed the reliability of scapular notching (SN) grading based on the Sirveaux classification system using anteroposterior radiographs.

Materials and methods 206 RSA procedures with 5-year postoperative anteroposterior shoulder radiographs were classified independently by seven assessors according to Sirveaux (session 1). After a review meeting, three assessors re-classified the radiographs along with quality criteria (session 2). SN grading by the majority of assessors was taken as the reference. Classification interobserver reliability was analyzed using Kappa statistics.

Results The incidence of SN was estimated at 53% and 37% at the first and second sessions, respectively. Interobserver reliability Kappa coefficients resulting from the first and second sessions were 0.27 and 0.43, respectively. Case selection based on radiographic quality criteria did not improve SN grading reliability in the second session.

Conclusion Agreement between individual surgeons was low when grading SN in RSA according to Sirveaux using anteroposterior radiographs. Consensus among several assessors may increase reliability in research settings.

Keywords Scapular notching · Sirveaux classification · Reverse shoulder arthroplasty · Anteroposterior radiographs · Reliability · Agreement

Introduction

Reverse shoulder arthroplasty (RSA) is effective for treating rotator cuff tear arthropathy, fracture sequelae or failed anatomic total shoulder replacement [1–4]. Scapular notching (SN) is a common observation after RSA resulting in erosion of the inferior scapular neck during adduction, and is caused by a mechanical conflict with the humeral RSA component [3, 5–9]. Erosion of the posterior part of the scapula during rotation may also occur [7, 10]. The incidence of SN is between 0% and 96% [11–15]. While the effect of SN on the outcome after RSA remains controversial, SN is believed to influence loosening and therefore long-term prosthesis survival [3, 4, 9, 16–18]. Inferior placement and less medialization of the baseplate is thought to reduce the amount of SN [11, 12].

The Sirveaux classification system comprises four grades describing the extent of SN on anteroposterior and lateral radiographs (Fig. 1) [9, 18]. Its reliability has not been thoroughly evaluated. In a series of 60 shoulders, Sadoghi et al. reported excellent interobserver reliability between two assessors, although quality testing of this system was not their primary focus [19].
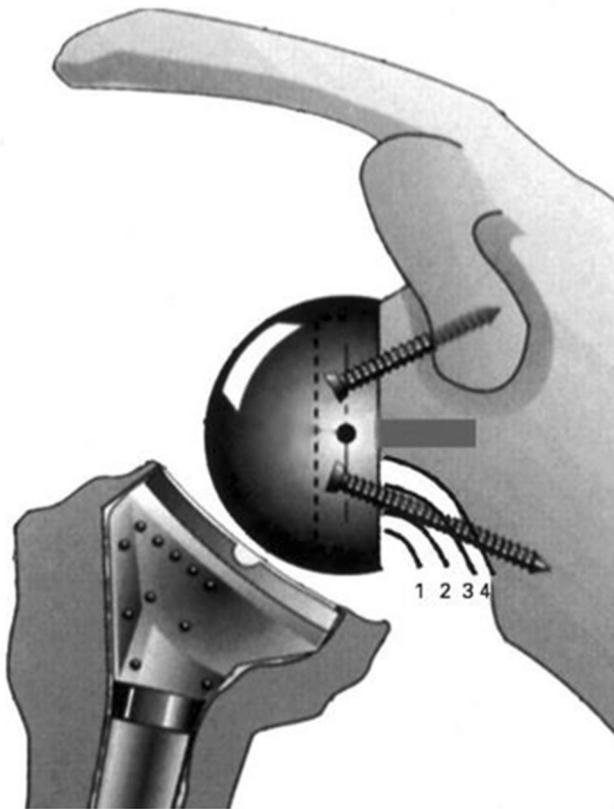
Our primary study aim was to assess the reliability of SN grading based on the Sirveaux classification system using anteroposterior radiographs. In addition, we explored the effect of radiographic quality criteria on SN assessment.

✉ Laurent Audigé
laurent.audige@kws.ch

1   Upper Extremities Department, Schulthess Clinic, Zurich, Switzerland

2   Research and Development Department, Schulthess Clinic, Lengghalde 2, 8008 Zurich, Switzerland

3   University of Witten/Herdecke, Witten, Germany

Materials and methods

We identified a total of 237 consecutive patients who had undergone RSA between July 2006 and May 2010 with a 5-year follow-up clinical examination. The sample size was limited by the size of our register. The use of our register data to assess specific clinically relevant factors such as SN, in conjunction with functional outcome parameters (e.g., Constant-Murley score [20] and patient-reported Shoulder Pain and Disability Index [21]) was approved by the local ethics committee. Anteroposterior radiographs in internal and external rotation of the humerus within the scapular plane made at 5 years or, when missing, at 2 years were selected. Any radiographs with poor visibility of the scapular neck due to an insufficient X-ray beam, superpositions or extracapsular inferior screw placement (Fig. 2) that were clearly insufficient for SN grading were excluded.

First classification session

Available postoperative radiographs [anteroposterior in (a) internal and (b) external rotation] were anonymized, consecutively numbered and given to 7 assessors (i.e., 3 experienced fellowship-trained orthopedic surgeons with shoulder expertise of more than 3 years, 1 orthopedic fellow without special shoulder expertise, 2 fourth-year residents with shoulder experience of more than 12 months, and 1 third-year resident without shoulder expertise) for SN grading according to Sirveaux [9, 18]. All assessors received a description of the classification system; SN grading was independently completed without access to additional information.

Fig. 1 Scapular notching classification according to the Sirveaux classification system (Reproduced with permission and copyright © of the British Editorial Society of Bone and Joint Surgery from Sirveaux F, Favard L, Oudet D, Huquet D, Walch G, Molé D. Grammont inverted total shoulder arthroplasty in the treatment of glenohumeral osteoarthritis with massive rupture of the cuff: results of a multicenter study of 80 shoulders. [9]). A defect contained within the pillar corresponds to a Grade 1 notch. A Grade 2 notch is in contact with the lower screw, Grade 3 notch is localized over the lower screw, and a Grade 4 notch extends under the base plate

Fig. 2 a Five-year postoperative anteroposterior radiograph in internal rotation highlighting extrascapular, inferior screw placement. **b** Five-year postoperative anteroposterior radiograph in internal rotation highlighting the lack of visibility of the scapular neck for the classification of scapular notching
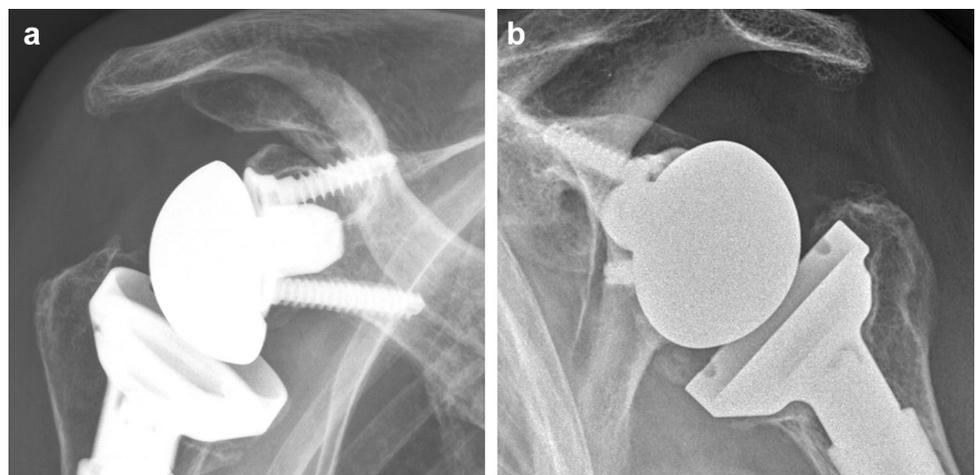
Assessor review meeting

A review meeting attended by all seven assessors was held after the first session. Any grading disagreements and issues were discussed in an attempt to identify the main cause of discrepancy. Radiograph quality was of major concern, and it was suggested that SN grading would show higher reliability when applied on radiographs meeting specific quality criteria. The following quality parameters were defined: (1) optimal tangential view, i.e., when the base plate is seen in profile so that any gap between base plate and glenoid would be clearly visible; (2) optimal horizontal view, i.e., when the projection is horizontal, perpendicular and centered on the central peg so that both superior and inferior base plate screws are seen symmetrical and equidistant to the central screw; and (3) optimal visibility of the scapular neck, i.e., the view onto the inferior scapular neck is not superimposed by ribs, the humeral cup or clavicle (Fig. 3). The additional use of preoperative anteroposterior radiographs in internal and external rotation was suggested among the assessors to more accurately assess individual scapular neck anatomies and ultimately improve SN grading.

Second classification session

A second session was undertaken for the final radiograph set analysis. All images were given to 3 of the original 7
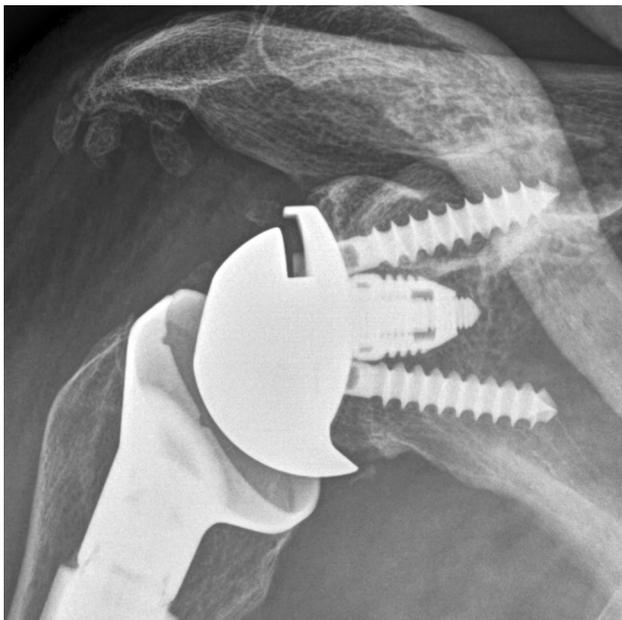
Fig. 3 5-year postoperative anteroposterior radiograph in external rotation fulfilling the three quality criteria for optimal grading of scapular notching according to the Sirveaux classification: (1) projection tangential to the base plate; (2) projection displaying an optimal horizontal view; and (3) a clear view to the inferior scapular neck

assessors (2 experienced fellowship-trained orthopedic surgeons and 1 fourth-year resident) in a new and random order for grading. The assessors were required to initially grade the images only followed by a second assessment with the preoperative (baseline) images. Radiograph quality was also individually evaluated based on the previously defined quality parameters.

Data management and statistical analysis

Data were collected on Excel sheets and combined within Stata 14.0 (StataCorp LP, College Station, TX) for statistical analysis. Data analysis and reporting were made following international guidelines [22]. From the first classification session, a reference grading for each patient was derived either from the grade provided by the majority of assessors or the highest of the two grades when there was an equal number of assessors with the same outcome. This reference grade was described as the most likely SN grade distribution for each patient within the cohort. Interobserver reliability was evaluated by the overall as well as grade-specific Kappa coefficients (κ) and interpreted according to Landis and Koch [23]. In addition, we calculated assessor pairwise κ to evaluate any potential differences in understanding or application of the SN grading system among the assessors. Pairwise κ among assessors with expertise in shoulder surgery were compared with other pairwise κ involving assessors without shoulder expertise using the Wilcoxon rank-sum test. Classification reliability was also assessed for the proposed French Society of Orthopedic Surgery & Traumatology (SOFCOT) classification adjustment (Grades 0–2), where the Sirveaux Grades 2–4 were combined [7].

From the second classification session, the decision of whether a criterion was fulfilled for any radiograph was made by the majority of the three assessors. On this basis, a quality subgroup was formed with radiographs meeting all three quality criteria in one or both views. Combined data were analyzed similar to that handled in the first session. Reliability assessment was performed on all cases without and with additional preoperative radiographs followed by a further assessment only considering the quality subgroup.

Results

From 237 RSAs, 31 were excluded for the first session because the radiographs were clearly insufficient for SN grading due to a lack of scapular neck visibility or extrascapular inferior screw placement (Fig. 4). The overall cohort included 206 RSAs (7.3% had 2-year radiographs) in 200 patients, where 69% were female. Mean patient age at the time of surgery was 71.5 years (SD 7.9). RSAs were performed using either a Promos Reverse® (Smith & Nephew
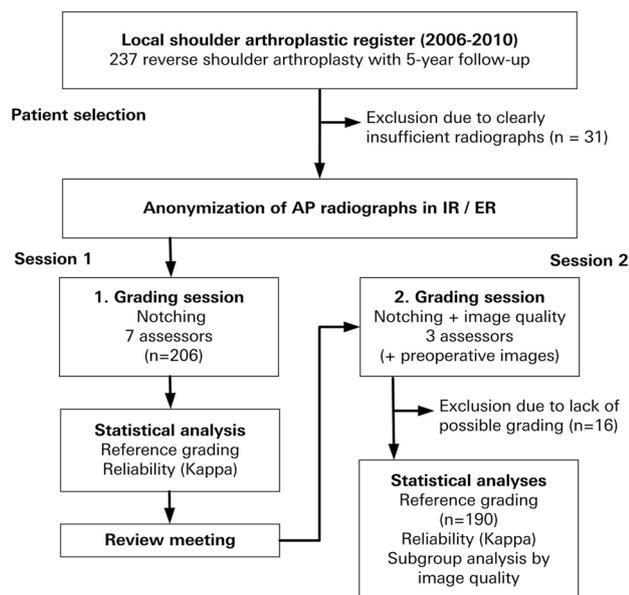
Fig. 4 Flow chart outlining the patient and radiograph selection process for the classification of scapular notching. *AP* anteroposterior, *IR/ER* internal and external rotation

Orthopaedics AG, Aarau, Switzerland) ($n = 137$) or SMR™ Reverse (Lima Switzerland SA, Rotkreuz, Switzerland) prosthesis ($n = 69$).

According to the majority of first session assessors, the SN incidence in our series was estimated at 53%; SN Grades 3 and 4 were noted in 6 and 4 RSAs, respectively (Table 1), and grading could not be achieved using available radiographs in an additional 16 RSAs. Interobserver reliability of the first session showed a κ value of 0.27. Pairwise assessor κ ($n = 21$) ranged from 0.05 to 0.51 with a median of 0.31 (Table 2). In particular, one assessor

(ID 5) with a median pairwise κ of 0.18 showed the most disagreement with the rest of the assessor panel. The experience level of the assessors in shoulder surgery did not influence the reliability of grading SN as measured by pairwise κ ($p = 0.89$).

After the second session, the SN incidence was 37%, and Grades 3 and 4 were noted in 3 and 6 RSAs, respectively (Table 1; Fig. 5). Grading reliability improved to a moderate level with a κ of 0.43 compared to 0.32, which was estimated from the same three assessors in the first classification session. All 3 quality criteria were fulfilled in 98 of 190 cases (52%). Restriction of the analysis to the radiograph quality subgroup with similar κ had no influence on grading reliability. The use of preoperative radiographs had no relevant influence on the assessment of SN grade; 2 of the 3 assessors changed the grade after viewing the preoperative radiograph in only 3 and 15 cases, respectively (i.e., 1 case was upgraded from Grade 0 to 1, 4 cases were downgraded from Grade 2 to 0, and the remaining cases were downgraded from Grade 1 to 0). The third assessor did not change any SN grade at all. Kappa coefficients for classification reliability of the proposed SOFCOT classification adjustment were 0.32 and 0.47 at the first and second sessions, respectively.

Discussion

This study assessed the reliability of SN grading according to Sirveaux in clinical use. The incidence of SN was 37% for the final classification of 190 cases made by 3 assessors with shoulder expertise. Interobserver reliability was low with an overall κ value of 0.43, reaching only 0.46 after considering three radiographic quality criteria.

Table 1 Distribution of cases according to reference scapular notching (SN) grades and Kappa coefficients resulting from two consecutive classification sessions and radiographic quality assessment

SN grade	All cases (Session 1)			All cases (Session 2)			All three quality criteria[a]		
	N	%	Kappa	N	%	Kappa	N	%	Kappa
0 (none)	97	47	0.30	124	65	0.51	63	64	0.51
1	68	33	0.22	45	24	0.41	26	27	0.52
2	31	15	0.28	12	6	0.28	5	5	0.25
3	6	3	0.25	3	2	0.29	2	2	0.34
4	4	2	0.36	6	3	0.43	2	2	0.32
Overall	206		0.27	190		0.43	98		0.46

N = distribution of cases according to a reference grading derived by the majority of seven and three assessors at the first and second grading sessions, respectively. After the second session, cases receiving three different SN grades or at least one Grade 3 or 4 by the assessors were reviewed and graded by consensus

[a]Radiographic quality assessment groups based on the following criteria: (1) sagittal plane view, i.e., when the projection is tangential to the base plate, was considered most essential; (2) transverse plane view, i.e., with both superior and inferior base plate screws symmetrical and equidistant to the central screw and the anterior border aligned with the posterior border of the inferior scapular neck; and (3) optimal visibility of the scapular neck, i.e., the view onto the inferior scapular neck is not superimposed by ribs, the humeral cup or clavicle; all three quality criteria = only cases that met all three quality criteria in one or both views

Table 2 Surgeon pairwise Kappa coefficients of the first classification session

| Assessor ID	Assessor ID					
	2	3	4	5	6	7
1	0.32	0.47	0.34	0.10	0.37	0.34
2		0.37	0.22	0.28	0.23	0.31
3			0.34	0.13	0.51	0.38
4				0.17	0.32	0.30
5					0.05	0.20
6						0.30

Assessors 1, 3 and 4 = trained orthopedic shoulder fellows with expertise in shoulder surgery

Assessors 5 and 7 = 4th-year residents with shoulder expertise; assessor 5 showed overall lower agreement compared with other assessors, which highlights a general misunderstanding of the SN grading process

Assessor 2 = orthopedic fellow without special shoulder expertise

Assessor 6 = 3rd-year resident without shoulder expertise

Fig. 5 a Five-year postoperative anteroposterior radiographs in external rotation showing Grade 3 scapular notching as graded by consensus among three trained experienced assessors according to the Sirveaux classification. **b** Five-year postoperative anteroposterior radiographs in external rotation showing Grade 4 scapular notching as graded by consensus among three trained experienced assessors according to the Sirveaux classification
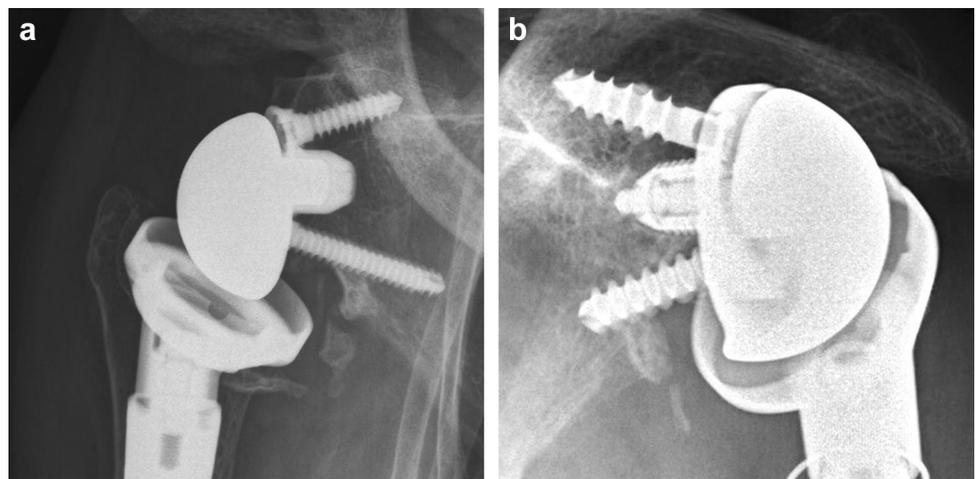
Reported incidences of SN widely vary and may be explained by the use of different prostheses and changes of implant design with inferior or lateral placement of the glenosphere as well as changes in scapular neck angle [24]. Our study did not focus on factors associated with SN, but the reliability of its grading by clinicians independent of the origin of SN. Sirveaux and co-workers initially established this classification to grade scapular neck erosion caused by the reverse shoulder prosthesis humeral component [9, 18]. Levigne et al. described a post-mortem analysis of a shoulder specimen that highlighted the difficulty in assessing SN according to Sirveaux with anteroposterior radiographs [17]. The preserved specimen showed Grade 2 notching, yet the resultant radiographs taken in various projections displayed either Grade 1 or 4 notching depending on the view. Although the assessment of reliability was not a primary aim, Sadoghi et al. reached an "almost perfect" interobserver agreement with a κ of 0.86 in their series of 60 RSA; the classification was done by only two observers over two sessions [19]. However, in view of our results, the controversial influence of SN on clinical outcome in various studies might

be caused by the lack of grading reliability as well as the grading cut-off used [1, 9, 13, 14, 17, 25, 26].

Based on investigations in fracture classification [27], we conducted our study in a systematic way, whereby assessors were required to meet after a classification session to understand the reasons associated with any potential lack of agreement. This approach provides the opportunity to make recommendations on whether a classification system should be revised or further applied under specific circumstances. Any proposed revision in the grading definition or process can be tested in successive sessions. In this study, our assessors failed to understand the circumstances by which SN grading can be performed reliably by single surgeons, despite conducting a review meeting.

Our results suggest that an anteroposterior radiograph with tangential projection of the glenosphere is not an adequate source for reliably grading SN according to Sirveaux or that this classification system has serious limitations. Radiograph quality as well as the availability of preoperative radiographs had no influence on interobserver reliability. In most clinical settings, only one true anteroposterior

radiograph is usually available to determine the presence and degree of SN. Our assessors had access to two anteroposterior radiographs from which the best image could be used for grading. Detailed instructions on how to grade SN, which were fully explained during a review meeting, only coincided with a slight increase in κ of 0.32–0.43. Although the κ values of 0.27, 0.32 and 0.43 would be interpreted as "fair" according to Landis and Koch [23], they remain well below a κ benchmark value of 0.60 that may be acceptable as an indication of clinical usefulness [28].

The Sirveaux classification was initially established for grading SN in patients treated with the Delta shoulder prosthesis [9, 18]. Since then, it has been used for different prostheses and implant designs. In comparison to the Delta shoulder prosthesis, the spaces between the screws and central peg of the Promos Reverse® and SMR™ Reverse prostheses are smaller. This might complicate the ability to differentiate between Grades 3 and 4 SN. Furthermore, with increasingly inferior placement of the base plate and lower screw, the incidence of bony erosion is reduced because the humeral component touches the lower screw before significant abrasion of the inferior scapular neck can occur.

We believe the radiological follow-up assessments in our study adequately reflect standard clinical procedure and 5-year postoperative images represent an acceptable follow-up time point considering the development of SN. Another strength is the reflection of different levels of experience of the observers, since the graduation of SN according to Sirveaux should be reliable with any surgeons. However, this work has limitations. Although image quality did not influence the reliability of SN classification, only 52% of all anteroposterior radiographs fulfilled all 3 quality criteria. The quality evaluation remained subjective and could differ among assessors; therefore, we relied on assessments made by at least two of the three assessors in the second session. Moreover, other factors such as the formation of heterotopic ossifications of the long head of the triceps tendon may impair the radiograph quality and need to be taken into account for correctly grading SN. Additionally, we did not evaluate SN progression since only baseline and 5-year postoperative radiographs were evaluated. By examining the radiographs in sequence over time, the precision of the Sirveaux classification system may increase. Nevertheless, the main issue of the two-dimensional perspective remains, and our inclusion of preoperative radiographs did not improve grading reliability. Another limitation to consider may be that the grading according to Sirveaux was done by clinicians and not radiologists. Yet clinicians tend to rate radiographs themselves and our approach seems to be very relevant for orthopedics in daily practice. The second grading was performed by three of seven assessors because these clinicians were the main investigators most motivated to clarify the role of the Sirveaux classification in assessing

SN. It was planned to involve the other assessors as well to complete the second classification session, however, results were below expectations and we believed the involvement of the remaining four assessors in this second session was no longer justified. The assessment of intraobserver reliability was not relevant in this study.

Conclusion

Anteroposterior radiographs do not provide adequate information for individual surgeons to reliably classify SN in RSA according to Sirveaux. Radiograph quality for SN grading does not seem to influence classification reliability. When SN grading on anteroposterior radiographs is planned, we suggest classifying by consensus among several assessors. In light of our results, we suggest the SN grading system be revisited by an international panel of expert shoulder surgeons. Assessment using alternative imaging such as computer tomography may help to gain a better understanding of this system.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval The study was performed in accordance with the ethical standards of the cantonal ethics committee of Zurich (KEK-ZH) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Approval for analyses of the local clinical register (KEK-ZH Nr. 2014-0483: Clinical and subjective long-term outcome after an implantation of a shoulder arthroplasty) was granted on 23.01.2015.

References

1. Boileau P, Watkinson D, Hatzidakis AM, Hovorka I (2006) Neer Award 2005: the Grammont reverse shoulder prosthesis: results in cuff tear arthritis, fracture sequelae, and revision arthroplasty. J Shoulder Elb Surg 15:527–540. https://doi.org/10.1016/j.jse.2006.01.003
2. Boulahia A, Edwards TB, Walch G, Baratta RV (2002) Early results of a reverse design prosthesis in the treatment of arthritis of the shoulder in elderly patients with a large rotator cuff tear. Orthopedics 25:129–133. https://doi.org/10.3928/0147-7447-20020201-16
3. De Wilde L, Mombert M, Van Petegem P, Verdonk R (2001) Revision of shoulder replacement with a reversed shoulder prosthesis (Delta III): report of five cases. Acta Orthop Belg 67:348–353
4. Guery J, Favard L, Sirveaux F, Oudet D, Molé D, Walch G (2006) Reverse total shoulder arthroplasty. Survivorship analysis of

eighty replacements followed for five to ten years. J Bone Jt Surg Am 88:1742–1747. https://doi.org/10.2106/JBJS.E.00851

5. Kempton LB, Balasubramaniam M, Ankerson E, Wiater JM (2011) A radiographic analysis of the effects of glenosphere position on scapular notching following reverse total shoulder arthroplasty. J Shoulder Elb Surg 20:968–974. https://doi.org/10.1016/j.jse.2010.11.026

6. Kowalsky MS, Galatz LM, Shia DS, Steger-May K, Keener JD (2012) The relationship between scapular notching and reverse shoulder arthroplasty prosthesis design. J Shoulder Elb Surg 21:1430–1441. https://doi.org/10.1016/j.jse.2011.08.051

7. Lévigne C, Garret J, Boileau P, Alami G, Favard L, Walch G (2011) Scapular notching in reverse shoulder arthroplasty: is it important to avoid it and how? Clin Orthop Relat Res 469:2512–2520. https://doi.org/10.1007/s11999-010-1695-8

8. Nicholson GP, Strauss EJ, Sherman SL (2011) Scapular notching: recognition and strategies to minimize clinical impact. Clin Orthop Relat Res 469:2521–2530. https://doi.org/10.1007/s11999-010-1720-y

9. Sirveaux F, Favard L, Oudet D, Huquet D, Walch G, Molé D (2004) Grammont inverted total shoulder arthroplasty in the treatment of glenohumeral osteoarthritis with massive rupture of the cuff. Results of a multicentre study of 80 shoulders. J Bone Jt Surg Br 86:388–395. https://doi.org/10.1302/0301-620X.86B3.14024

10. McFarland EG, Sanguanjit P, Tasaki A, Keyurapan E, Fishman EK, Fayad LM (2006) The reverse shoulder prosthesis: a review of imaging features and complications. Skeletal Radiol 35:488–496. https://doi.org/10.1007/s00256-006-0109-1

11. Holschen M, Franetzki B, Witt KA, Liem D, Steinbeck J (2017) Conversions from anatomic shoulder replacements to reverse total shoulder arthroplasty: do the indications for initial surgery influence the clinical outcome after revision surgery? Arch Orthop Trauma Surg 137:167–172. https://doi.org/10.1007/s00402-016-2595-5

12. Jonušas J, Banytė R, Ryliškis S (2017) Clinical and radiological outcomes after reverse shoulder arthroplasty with less medialized endoprosthesis after mean follow-up time of 45 months. Arch Orthop Trauma Surg 137:1201–1205. https://doi.org/10.1007/s00402-017-2751-6

13. Simovitch RW, Zumstein MA, Lohri E, Helmy N, Gerber C (2007) Predictors of scapular notching in patients managed with the Delta III reverse total shoulder replacement. J Bone Jt Surg Am 89:588–600. https://doi.org/10.2106/JBJS.F.00226

14. Werner CM, Steinmann PA, Gilbart M, Gerber C (2005) Treatment of painful pseudoparesis due to irreparable rotator cuff dysfunction with the Delta III reverse-ball-and-socket total shoulder prosthesis. J Bone Jt Surg Am 87:1476–1486. https://doi.org/10.2106/JBJS.D.02342

15. Wierks C, Skolasky RL, Ji JH, McFarland EG (2009) Reverse total shoulder replacement: intraoperative and early postoperative complications. Clin Orthop Relat Res 467:225–234. https://doi.org/10.1007/s11999-008-0406-1

16. Favard L, Lévigne C, Nerot C, Gerber C, De Wilde L, Molé D (2011) Reverse prostheses in arthropathies with cuff tear: are

survivorship and function maintained over time? Clin Orthop Relat Res 469:2469–2475. https://doi.org/10.1007/s11999-011-1833-y

17. Lévigne C, Boileau P, Favard L, Garaud P, Mole D, Sirveaux F et al (2008) Scapular notching in reverse shoulder arthroplasty. J Shoulder Elb Surg 17:925–935. https://doi.org/10.1016/j.jse.2008.02.010

18. Valenti PH, Boutens D, Nerot C (2001) Delta 3 reversed prosthesis for osteoarthritis with massive rotator cuff tear, long term results (> 5 years). In: Walch G, Boileau P, Molé D (eds) 2000 shoulder prostheses. Two to ten year follow-up. Sauramps Medical, Montpellier, pp 253–259

19. Sadoghi P, Leithner A, Vavken P, Holzer A, Hochreiter J, Weber G et al (2011) Infraglenoidal scapular notching in reverse total shoulder replacement: a prospective series of 60 cases and systematic review of the literature. BMC Musculoskelet Disord 12:101. https://doi.org/10.1186/1471-2474-12-101

20. Constant CR, Gerber C, Emery RJ, Sojbjerg JO, Gohlke F, Boileau P (2008) A review of the Constant score: modifications and guidelines for its use. J Shoulder Elb Surg 17:355–361. https://doi.org/10.1016/j.jse.2007.06.022

21. Angst F, Goldhahn J, Pap G, Mannion AF, Roach KE, Siebertz D et al (2007) Cross-cultural adaptation, reliability and validity of the German Shoulder Pain and Disability Index (SPADI). Rheumatology 46:87–92. https://doi.org/10.1093/rheumatology/kel040

22. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A et al (2011) Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol 64:96–106. https://doi.org/10.1016/j.jclinepi.2010.03.002 (**Blinded for review purposes**)

23. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174

24. Erickson BJ, Frank RM, Harris JD, Mall N, Romeo AA (2015) The influence of humeral head inclination in reverse total shoulder arthroplasty: a systematic review. J Shoulder Elb Surg 24:988–993. https://doi.org/10.1016/j.jse.2015.01.001

25. Wall B, Nove-Josserand L, O'Connor DP, Edwards TB, Walch G (2007) Reverse total shoulder arthroplasty: a review of results according to etiology. J Bone Jt Surg Am 89:1476–1485. https://doi.org/10.2106/JBJS.F.00666

26. Wellmann M, Struck M, Pastor MF, Gettmann A, Windhagen H, Smith T (2013) Short and midterm results of reverse shoulder arthroplasty according to the preoperative etiology. Arch Orthop Trauma Surg 133:463–471. https://doi.org/10.1007/s00402-013-1688-7

27. Audigé L, Bhandari M, Hanson B, Kellam J (2005) A concept for the validation of fracture classifications. J Orthop Trauma 19:404–409. https://doi.org/10.1097/01.bot.0000155310.04886.37 (**Blinded for review purposes**)

28. McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 22:276–282. https://doi.org/10.11613/BM.2012.031