



Intraobserver and interobserver reliability of recategorized Neer classification in differentiating 2-part surgical neck fractures from multi-fragmented proximal humeral fractures in 116 patients

Bakir O. Sumrein, MD^{a,*}, Ville M. Mattila, PhD^{a,b}, Vesa Lepola, PhD^a,
Minna K. Laitinen, PhD^a, Antti P. Launonen, PhD^a, NITEP Group[†]

^aDivision of Orthopaedics and Traumatology, Department of Trauma, Musculoskeletal Surgery and Rehabilitation, Tampere University Hospital, Tampere, Finland

^bSchool of Medicine, University of Tampere, Tampere, Finland

[†]Juha Paloneva, PhD^{1,2}, Kenneth Jonsson, PhD³, Olof Wolf, PhD³, Peter Ström, MD³,
Hans Berg, PhD^{4,5}, Li Felländer-Tsai, PhD⁴, Inger Mechlenburg, PhD^{6,7},
Kaj Døssing, MD⁸, Helle Østergaard, PT, MSc in Health Science⁸, Timo Rahnel, MD⁹,
Aare Märtson, PhD¹⁰

¹Department of Orthopedics and Traumatology, Central Finland Hospital, Jyväskylä, Finland

²Department of Orthopedics and Traumatology, University of Eastern Finland, Kuopio, Finland

³Department of Orthopaedics, Institute of Surgical Sciences, Uppsala University Hospital, Uppsala, Sweden

⁴Department of Clinical Science, Intervention and Technology, Division of Orthopedics and Biotechnology, Karolinska Institutet, Stockholm, Sweden

⁵Division of Orthopedics, Karolinska University Hospital Huddinge, Stockholm, Sweden

⁶Department of Orthopedic Surgery, Aarhus University Hospital, Aarhus, Denmark

⁷Centre of Research in Rehabilitation (CORIR), Department of Clinical Medicine, Aarhus University Hospital and Aarhus University, Denmark

⁸Orthopedics Department, Viborg Regional Hospital, Viborg, Denmark

⁹Tallinn Surgery Clinic, Orthopedics Centre, North Estonia Medical Centre Foundation, Tallinn, Estonia

¹⁰Department of Traumatology and Orthopedics, University of Tartu, Tartu, Estonia

The proximal humeral fracture randomized controlled trial protocol has been approved by the Regional Ethics Committee of Tampere University Hospital (approval No. R10127).

[†]Contributing authors to NITEP (Nordic Innovative Trial to Evaluate Osteoporotic Fractures) group: Juha Paloneva (Central Finland Hospital), Kenneth Jonsson (Uppsala University Hospital), Olof Wolf (Uppsala University Hospital), Peter Ström (Uppsala University Hospital), Hans Berg (Karolinska Institutet and Karolinska University Hospital Huddinge), Li Felländer-Tsai (Karolinska Institutet), Inger Mechlenburg

(Aarhus University Hospital and Aarhus University), Kaj Døssing (Viborg Regional Hospital), Helle Østergaard (Viborg Regional Hospital), Timo Rahnel (North Estonia Medical Centre Foundation), Aare Märtson (University of Tartu).

*Reprint requests: Bakir O. Sumrein, MD, Division of Orthopaedics and Traumatology, Department of Trauma, Musculoskeletal Surgery and Rehabilitation, Tampere University Hospital, Teiskontie 35, 116 Bakir Omar Sumrein, 33520, Tampere, Finland.

E-mail address: dr.bakir.sumrein@gmail.com (B.O. Sumrein).

Background: Optimal fracture classification should be simple and reproducible and should guide treatment. For proximal humeral fractures, the Neer classification is commonly used. However, intraobserver and interobserver reliability of the Neer classification has been shown to be poor. In clinical practice, it is essential to differentiate 2-part surgical neck fractures from multi-fragmented fractures. Thus, the aim of this study was to evaluate whether surgeons can differentiate 2-part surgical neck fractures from multi-fragmented fractures using plain radiographs and/or computed tomography (CT).

Methods: Three experienced upper limb specialists and trauma surgeons (B.O.S., A.P.L., and V.L.) independently reviewed and classified blinded plain radiographs and CT scans of 116 patients as showing 2-part surgical neck fractures or multi-fragmented fractures. Each imaging modality was reviewed and classified separately by each surgeon, after which each surgeon reviewed both modalities at the same time. This process was repeated by all surgeons after 24 weeks. Intraobserver and interobserver analyses were conducted using Cohen and Fleiss κ values, respectively.

Results: The κ coefficient for interobserver reliability showed substantial correlation (0.61-0.73) and was as follows: 0.73 for radiographs alone, 0.61 for CT scans alone, and 0.72 for radiographs and CT scans viewed together. After 24 weeks, the process was repeated and intraobserver reliability was calculated. The κ coefficient for intraobserver reliability showed substantial correlation (0.62-0.75) and was as follows: 0.62 for radiographs alone, 0.64 for CT scans alone, and 0.75 for radiographs and CT scans viewed together.

Conclusion: Clinicians were able to differentiate 2-part surgical neck fractures from multi-fragmented fractures based on plain radiographs reliably.

Level of evidence: Basic Science Study; Validation or Development of Classification System

© 2018 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Proximal humeral fracture; Neer classification; recategorized; interobserver; intraobserver; reliability

Proximal humeral fracture (PHF) is one of the most common fractures among elderly persons.¹⁸ A Swedish population-based study was recently published that reported a national PHF incidence of 122 per 100,000 person-years in 2012, with the highest fracture incidence and surgical treatment rate observed in individuals aged 60 years or older.²⁶

The optimal treatment for PHF has been controversial. This is especially true for the treatment of PHFs in patients older than 60 years. A recent Cochrane review, as well as a high-quality review from Finland, reported on current randomized controlled trials (RCTs) that have compared different treatment options for elderly patients with 3- or 4-part PHFs.^{9,15} Both reviews suggested that the functional outcome after operative treatment is not superior to that after nonoperative treatment. In recent years, reverse total shoulder arthroplasty (RTSA) has gained popularity in the treatment of multi-fragmented PHFs, and 1 RCT stated its superiority over hemiarthroplasty.²⁴ It is interesting that no published RCT has compared the commonly used treatment options of plating and nonoperative treatment in 2-part surgical neck fractures that constitute the majority of displaced PHFs in the elderly population,⁵ although 1 trial protocol has been published.¹⁶ Case series have shown promising results after surgery with locking plates,¹⁰ and the incidence of plating has increased significantly in many countries.^{12,26} When the scientific evidence on these varying treatment options is taken into account, it seems that it is essential to differentiate 2-part surgical neck fractures from multi-fragmented fractures.

The Neer classification (NC)²⁰ is probably one of the most popularized and most used classification systems. However,

the 4-segment classification system defines PHFs by the number of displaced segments (humeral head and shaft, greater and lesser tuberosity), with additional categories for articular fractures and fracture-dislocations, making 16 different categories in total (Fig. 1). According to the original publication, a fracture is defined as displaced if there is more than 1 cm of distance between segments or 45° of angulation.²⁰ A limitation of the NC is the arbitrary definition of “displacement,”²¹ the detrimental effect of which is amplified in 3- and 4-part fractures as to whether all fractured segments should be displaced according to the NC definition.² Intraobserver and interobserver studies of the NC are abundant, many of which have concluded fair to moderate agreement.^{3,19}

Fracture classification systems should be easy to implement, and they should guide the decision-making process to select an adequate method of treatment based on high-quality evidence. On the basis of the current evidence, treatment recommendations for 2-part surgical neck and multi-fragmented fractures may vary, and thus it is essential to differentiate these categories. The NC with 16 categories seems too complicated in clinical practice, and according to a study by Court-Brown et al,⁵ two-thirds of the displaced fractures fall into 3 categories: surgical neck (2-part) fracture and 3- and 4-part fractures.

With the limitations of the NC and recent scientific evidence on the treatment of PHFs being taken into account, the aim of this study was to assess the intraobserver and interobserver reliability of a simplified and recategorized NC in which we recategorized 3- and 4-part fractures into a single category of multi-fragmented fractures while otherwise retaining the original NC and its criteria. Radiographs and

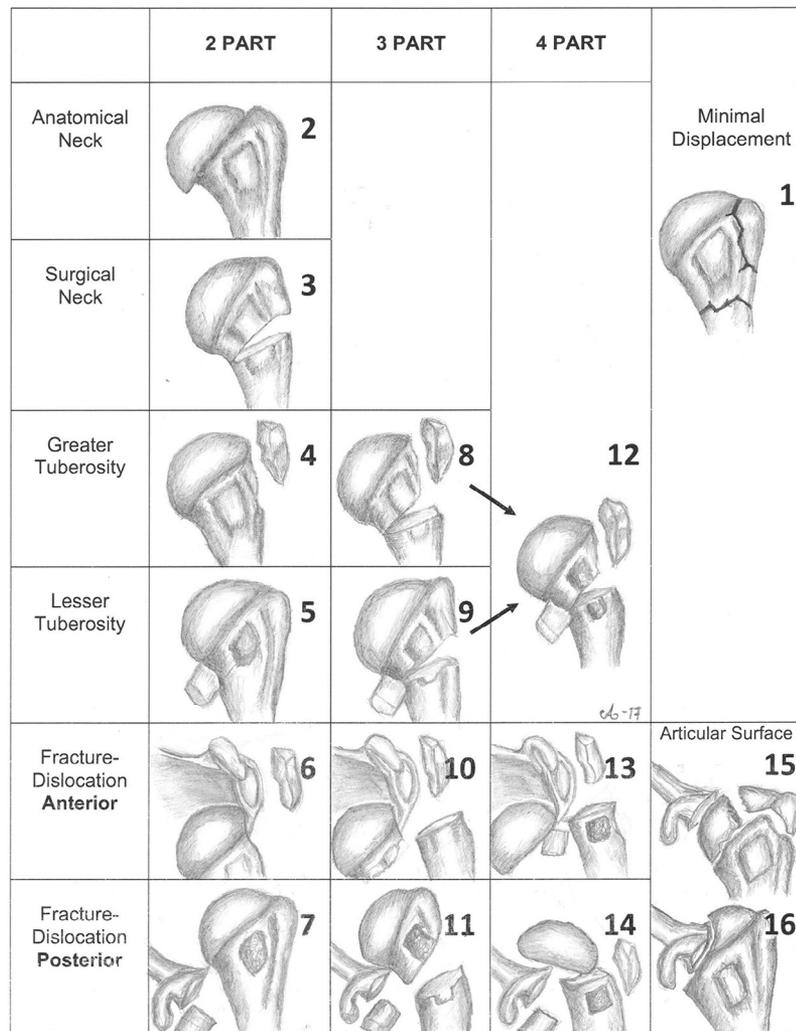


Figure 1 Neer classification for proximal humeral fractures. One of the authors (A.P.L.) created this illustration using the original Neer classification²⁰ as the data source. The proposed multi-fragmented fracture category includes original Neer classification categories 8, 9, and 12; otherwise, the original Neer classification was retained.

computed tomography (CT) scans were used to differentiate 2-part surgical neck and multi-fragmented PHFs. We hypothesized that trauma surgeons could differentiate 2-part surgical neck fractures from multi-fragmented fractures based on plain radiographs with substantial intraobserver and interobserver reliability using the recategorized NC.

Materials and methods

This prospective study sample included patients enrolled in the ongoing Nordic Innovative Trial to Evaluate Osteoporotic Fractures (NITEP) international multicenter RCT (n = 116). All patients recruited at Tampere University Hospital between February 1, 2011, and March 1, 2016, were included. As such, patient radiographs were readily available and easy to access for research purposes. The NITEP trial on PHFs compares nonsurgical and surgical treatment in the population aged 60 years or older, and more specific details on this trial have been published previously.¹⁶ In accordance with the RCT protocol, all 116 PHFs were diagnosed using plain radiographs (anteroposterior and lateral views) taken on average 1 day (range,

0-3 days) after trauma, followed by a routine CT scan using the GE Lightspeed RT16 scanner (GE Healthcare, Buckinghamshire, UK), Philips Brilliance 64 scanner (Philips Medical Systems, Andover, MA, USA), or GE Revolution GSI scanner (GE Healthcare). The CT scan included the entire scapula and the upper third of the fractured humerus, with a slice thickness of 0.6 mm. Coronal, sagittal, and axial images were obtained, and 3-dimensional volume reformatting was performed.¹⁶ The mean period between plain radiographs and CT scans was 1 day (range, 0-3 days). The Carestream Vue PACS (picture archiving and communication system) workstation (version 11.4.0.1253; Carestream Health, Rochester, NY, USA) was used to evaluate the radiographs and CT images; the raters were able to adjust the contrast and brightness and to zoom in and out on the images. According to the RCT recruitment consensus classification, this study included 53 multi-fragmented and 63 two-part surgical neck fractures. All radiographs included in this study will be available on the NITEP homepage (NITEP.eu) after June 2018.

The patients' radiographs and CT scans were rendered anonymous by removing names, identity numbers, and dates, as well as any other references. Three experienced upper extremity specialists and trauma surgeons (B.O.S., A.P.L., and V.L.), all of whom worked

at Tampere University Hospital, were selected as raters. They are perceived as experts in PHF management with a minimum of 5 years (range, 5-10 years) of experience in upper limb trauma and elective surgery including total shoulder arthroplasty and RTSA. Each rater independently reviewed and classified the plain radiographs and CT scans individually using our proposed simplified and recategorized NC in which we combined 3- and 4-part fractures into a single category of multi-fragmented fractures while otherwise retaining the original NC's displacement criteria (a fracture is defined as displaced if there is >1 cm of distance between segments or 45° of angulation). In the first stage, each rater independently reviewed the set of plain radiographic images and classified the fractures into 2-part surgical neck fractures and multi-fragmented fractures. In the second stage, the process was repeated using only CT scans. In the third stage, the raters reviewed and classified the fractures using all available imaging studies as in normal clinical practice (radiographic images and CT scans). On average, each set was reviewed within 3 days, and there was a 4-week delay before the next set was distributed for review. Before each viewing session, the radiographs were re-randomized and the raters re-blinded to their previous responses. The process was repeated by all 3 surgeons after a period of no less than 24 weeks to allow intraobserver reliability score calculation.

The reviewing process was conducted in the clinical setting; thus, the images were not calibrated, no time limit was set for viewing, and we did not hold a teaching session on the NC nor did we distribute a chart showing the NC. No additional instruments were used (eg, goniometer or ruler). NC fracture displacement criteria (1 cm or angulated at least 45°) were observed and judged pragmatically as in a normal clinical setting.

Statistics

Interobserver and intraobserver reliabilities were calculated using Fleiss and Cohen κ statistics, respectively. Results were interpreted according to the Landis and Koch criteria¹⁴ (0.00-0.20, slight agreement; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; and 0.81-1.00, almost perfect), and 95% confidence intervals were calculated. All analyses were completed using a web-based intercoder reliability calculator (<http://dfreelon.org/utlils/recalfront/recal>).

Results

By use of the recategorized NC, the κ coefficient for interobserver reliability showed substantial correlation (0.61-0.73) and was as follows: 0.73 for radiographs alone, 0.61 for CT scans alone, and 0.72 for radiographs and CT scans viewed together. After 24 weeks, the process was repeated and intraobserver reliability was calculated. The κ coefficient for intraobserver reliability showed substantial correlation (0.62-0.75) and was as follows: 0.62 for radiographs alone, 0.64 for CT scans alone, and 0.75 for radiographs and CT scans viewed together.

Discussion

The principal finding of this study was that experienced upper extremity orthopedic and trauma surgeons were able to differentiate 2-part surgical neck fractures from multi-fragmented

fractures based on plain radiographs with substantial intraobserver and interobserver reliability using the recategorized NC. In addition, CT scans did not markedly improve differentiation. The interobserver and intraobserver reliability of the NC in PHFs using radiographs has been shown in the literature to have great variation and is mostly graded as poor.^{7,19} Therefore, the purpose of this study was not to validate the entire NC but was to show that the reliability of the recategorized NC that specifically focuses on differentiating 2-part surgical neck fractures from multi-fragmented fractures in elderly patients is substantial.

The justification for recategorizing the NC into 2-part surgical neck and multi-fragmented fractures is based on the literature and treatment recommendations. In 2-part surgical neck fractures, surgical treatment with locking plates and nonsurgical treatment have been commonly used, while in 3- and 4-part fractures, the treatment options suggested by the current evidence are RTSA and nonsurgical treatment. Indeed, most surgeons would not consider arthroplasty for the treatment of 2-part surgical neck fractures. Moreover, it has been stated that the most important fracture-related factor predicting increased surgical treatment of PHFs in elderly patients is the severity, that is, the fracture pattern.²³ It has also been previously shown that the poor intraobserver and interobserver reliability of the NC mainly arises from differentiating between multi-fragmented fractures. Majed et al¹⁹ found that the poorest κ coefficient was recorded for 3-part fractures. Handoll et al⁸ defined the fracture population of the Proximal Fracture of the Humerus: Evaluation by Randomization study using the NC. They noted an increase in interobserver agreement after lowering the criteria for assessing displacement to include "displaced but unclear if Neer displacement criteria met."

In concordance studies, the κ coefficient is used as an index of reliability. In this study, we used the categorization suggested by the Landis and Koch criteria¹⁴ (0.00-0.20, slight agreement; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; and 0.81-1.00, almost perfect). We acknowledge that these values are not a gold-standard reference.¹³ Moreover, the κ coefficient is difficult to interpret unless the prevalence of positive and negative cases is taken into account. Therefore, the best way for the investigator to avoid paradoxical behavior of the κ coefficient is to design a study with approximately equal numbers of positive and negative cases.^{4,6,11}

In the previous publication by our research group, we discovered that 68% of upper extremity surgeons in Nordic countries preferred CT scans for diagnostic purposes and 86% used them for preoperative planning.¹⁷ In the present study, we found that CT scans did not improve the level of intraobserver or interobserver reliability in differentiating between the recategorized NC categories. These findings are in accordance with the previous literature.^{1,25} CT scans have been shown to improve intrarater and inter-rater reproducibility in analyzing complex multi-fragmented fractures,²² whereby CT obviously has its place in preoperative planning as it reveals the morphology of the fracture, guiding the surgeon during the operation.

We acknowledge that our study has limitations. One of the strengths of the study was that we used a defined prospectively collected cohort of PHFs in the population aged 60 years or older, where the method of collection resulted in a consistent homogeneous group. However, the predefined nature of the cohort may have biased the study results; the same surgeons (B.O.S., A.P.L., and V.L.) recruited the original 116 patients to the ongoing NITEP study on PHFs, as described in the “Materials and methods” section. The inclusion period was rather long, spanning a period of more than 5 years. However, the long inclusion period may have lessened the effect of our first limitation because the details of individual patients will have been forgotten over time.

Conclusion

We introduced a recategorized NC by which experienced upper extremity specialists were able to differentiate 2-part surgical neck fractures from multi-fragmented fractures based on plain radiographs and/or CT scans reliably and in a reproducible manner. An interesting finding was that CT scans did not increase interobserver or intraobserver reliability. With the newly introduced recategorized NC, we expect to better guide PHF treatment policies and make them easier to implement and generalize into the clinical setting.

Acknowledgment

The authors of the NITEP Group would like to thank the Academy of Finland.

Disclaimer

This research was supported by a grant from the Academy of Finland (grant 275481). The Academy of Finland is a governmental funding body for scientific research in Finland.

The authors, their immediate families, and any research foundations with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

References

- Bernstein J, Adler LM, Blank JE, Dalsey RM, Williams GR, Iannotti JP. Evaluation of the Neer system of classification of proximal humeral fractures with computerized tomographic scans and plain radiographs. *J Bone Joint Surg Am* 1996;78:1371-5.
- Brorson S, Olsen BS, Frich LH, Jensen SL, Sørensen AK, Krosgaard M, et al. Surgeons agree more on treatment recommendations than on classification of proximal humeral fractures. *BMC Musculoskelet Disord* 2012;13:114. <http://dx.doi.org/10.1186/1471-2474-13-114>
- Carofino BC, Leopold SS. Classifications in brief: the Neer classification for proximal humerus fractures. *Clin Orthop Relat Res* 2013;471:39-43. <http://dx.doi.org/10.1007/s11999-012-2454-9>
- Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990;43:551-8.
- Court-Brown CM, Garg A, McQueen MM. The epidemiology of proximal humeral fractures. *Acta Orthop Scand* 2001;72:365-71.
- Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43:543-9.
- Gracitelli MEC, Dotta TAG, Assunção JH, Malavolta EA, Andrade-Silva FB, Kojima KE, et al. Intraobserver and interobserver agreement in the classification and treatment of proximal humeral fractures. *J Shoulder Elbow Surg* 2017;26:1097-102. <http://dx.doi.org/10.1016/j.jse.2016.11.047>
- Handoll HH, Brealey SD, Jefferson L, Keding A, Brooksbank AJ, Johnstone AJ, et al. Defining the fracture population in a pragmatic multicentre randomised controlled trial: PROFHER and the Neer classification of proximal humeral fractures. *Bone Joint Res* 2016;5:481-9. <http://dx.doi.org/10.1302/2046-3758.510.bjr-2016-0132.r1>
- Handoll HH, Brorson S. Interventions for treating proximal humeral fractures in adults. *Cochrane Database Syst Rev* 2015;(11):CD000434. <http://dx.doi.org/10.1002/14651858.CD000434.pub4>
- Hauschild O, Konrad G, Audige L, de Boer P, Lambert SM, Hertel R, et al. Operative versus non-operative treatment for two-part surgical neck fractures of the proximal humerus. *Arch Orthop Trauma Surg* 2013;133:1385-93. <http://dx.doi.org/10.1007/s00402-013-1798-2>
- Hoehler FK. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *J Clin Epidemiol* 2000;53:499-503.
- Huttunen TT, Launonen AP, Pihlajamäki H, Kannus P, Mattila VM. Trends in the surgical treatment of proximal humeral fractures—a nationwide 23-year study in Finland. *BMC Musculoskelet Disord* 2012;13:261. <http://dx.doi.org/10.1186/1471-2474-13-261>
- Krippendorff K. *Content analysis: an introduction to its methodology*. Thousand Oaks, CA: SAGE; 2018. ISBN: 978-1506395661
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- Launonen AP, Lepola V, Flinkkilä T, Laitinen M, Paavola M, Malmivaara A. Treatment of proximal humerus fractures in the elderly: a systematic review of 409 patients. *Acta Orthop* 2015;86:280-5. <http://dx.doi.org/10.3109/17453674.2014.999299>
- Launonen AP, Lepola V, Flinkkilä T, Strandberg N, Ojanperä J, Rissanen P, et al. Conservative treatment, plate fixation, or prosthesis for proximal humeral fracture. A prospective randomized study. *BMC Musculoskelet Disord* 2012;13:167. <http://dx.doi.org/10.1186/1471-2474-13-167>
- Launonen AP, Lepola V, Laitinen M, Mattila VM. Do treatment policies for proximal humerus fractures differ among three Nordic countries and Estonia? Results of a survey study. *Scand J* 2016;105:186-90. <http://dx.doi.org/10.1177/1457496915623149>
- Lauritzen JB, Schwarz P, Lund B, McNair P, Transbol I. Changing incidence and residual lifetime risk of common osteoporosis-related fractures. *Osteoporos Int* 1993;3:127-32.
- Majed A, Macleod I, Bull AM, Zyto K, Resch H, Hertel R, et al. Proximal humeral fracture classification systems revisited. *J Shoulder Elbow Surg* 2011;20:1125-32. <http://dx.doi.org/10.1016/j.jse.2011.01.020>
- Neer CS II. Four-segment classification of proximal humeral fractures: purpose and reliable use. *J Shoulder Elbow Surg* 2002;11:389-400. <http://dx.doi.org/10.1067/mse.2002.124346>
- Neer CS II. Displaced proximal humeral fractures. I. Classification and evaluation. *J Bone Joint Surg Am* 1970;52:1077-89.
- Ohl X, Mangin P, Barbe C, Brun V, Nerot C, Sirveaux F. Analysis of four-fragment fractures of the proximal humerus: the interest of 2D and 3D imagery and inter- and intra-observer reproducibility. *Eur J Orthop*

- Surg Traumatol 2017;27:295-9. <http://dx.doi.org/10.1007/s00590-017-1911-2>
23. Okike K, Lee OC, Makanji H, Harris MB, Vrahas MS. Factors associated with the decision for operative versus non-operative treatment of displaced proximal humerus fractures in the elderly. *Injury* 2013;44:448-55. <http://dx.doi.org/10.1016/j.injury.2012.09.002>
 24. Sebastián-Forcada E, Cebrián-Gómez R, Lizaur-Utrilla A, Gil-Guillén V. Reverse shoulder arthroplasty versus hemiarthroplasty for acute proximal humeral fractures. A blinded, randomized, controlled, prospective study. *J Shoulder Elbow Surg* 2014;23:1419-26. <http://dx.doi.org/10.1016/j.jse.2014.06.035>
 25. Sjoden GO, Movin T, Aspelin P, Guntner P, Shalabi A. 3D-radiographic analysis does not improve the Neer and AO classifications of proximal humeral fractures. *Acta Orthop Scand* 1999;70:325-8.
 26. Sumrein BO, Huttunen TT, Launonen AP, Berg HE, Felländer-Tsai L, Mattila VM. Proximal humeral fractures in Sweden-a registry-based study. *Osteoporos Int* 2016;28:901-7. <http://dx.doi.org/10.1007/s00198-016-3808-z>