



Advanced CT visualization improves the accuracy of orthopaedic trauma surgeons and residents in classifying proximal humeral fractures: a feasibility study

Jan Dauwe^{1,2} · Karen Mys¹ · Guy Putzeys³ · Jana F. Schader¹ · R. Geoff Richards¹ · Boyko Gueorguiev¹ · Peter Varga¹ · Stefaan Nijs⁴

Received: 6 June 2020 / Accepted: 29 July 2020 / Published online: 6 August 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Purpose Osteosynthesis of proximal humeral fractures remains challenging with high reported failure rates. Understanding the fracture type is mandatory in surgical treatment to achieve an optimal anatomical reduction. Therefore, a better classification ability resulting in improved understanding of the fracture pattern is important for preoperative planning. The purpose was to investigate the feasibility and added value of advanced visualization of segmented 3D computed tomography (CT) images in fracture classification.

Methods Seventeen patients treated with either plate-screw-osteosynthesis or shoulder hemi-prosthesis between 2015 and 2019 were included. All preoperative CT scans were segmented to indicate every fracture fragment in a different color. Classification ability was tested in 21 orthopaedic residents and 12 shoulder surgeons. Both groups were asked to classify fractures using three different modalities (standard CT scan, 3D reconstruction model, and 3D segmented model) into three different classification systems (Neer, AO/OTA and LEGO).

Results All participants were able to classify the fractures more accurately into all three classification systems after evaluating the segmented three-dimensional (3D) models compared to both 2D slice-wise evaluation and 3D reconstruction model. This finding was significant ($p < 0.005$) with an average success rate of 94%. The participants experienced significantly more difficulties classifying fractures according to the LEGO system than the other two classifications.

Conclusion Segmentation of CT scans added value to the proximal humeral fracture classification, since orthopaedic surgeons were able to classify fractures significantly better into the AO/OTA, Neer, and LEGO classification systems compared to both standard 2D slice-wise evaluation and 3D reconstruction model.

Keywords Proximal humeral fracture · Fracture classification · Computed tomography · Semi-automatic segmentation · Computer-assisted segmentation

Introduction

Proximal humeral fractures (PHF) are common injuries and comprise approximately 5% of all fractures in adults [1], with women two-to-three times more frequently affected than male patients [2]. Most of the PHF can be managed conservatively; however, surgery is indicated in unstable and displaced fractures. Goal of the surgical treatment is to reduce the fracture and stabilize the fragments, thereby alleviating pain during the healing process and enabling immediate mobilization. Angular stable plate osteosynthesis has become a common procedure for surgical management of PHF [3], but even with these type of implants, it has remained a challenging procedure with a high failure

✉ Jan Dauwe
dauwejan@gmail.com

¹ AO Research Institute Davos, Clavadelerstrasse 8,
7270 Davos, Switzerland
² Department of Orthopaedic Surgery, University Hospitals
Leuven, Leuven, Belgium
³ Department of Orthopaedic and Trauma Surgery, AZ
Groeninge, Kortrijk, Belgium
⁴ Department of Trauma Surgery, University Hospitals Leuven,
Leuven, Belgium

rate up to 35% reported in the literature [4–7]. This might be due to the complexity of these fractures and the variety of influencing factors which must be considered for successful treatment. One important aspect is the fracture type that has to be correctly assessed as a prerequisite for the selection of the appropriate treatment. Nevertheless, fracture classification is a challenging task for complex PHF, partially due to difficulties in recognizing the entire three-dimensional (3D) extent of the fracture, including the number and dislocation of fragments.

Several classification systems have been developed for PHF, including the Neer [8], AO/OTA (Arbeitsgemeinschaft für Osteosynthesefragen/Orthopaedic Trauma Association) [9] and LEGO [10] systems. The AO/OTA classification is widely accepted and frequently used to differentiate A- and B-type fractures (extra-articular) from C-type fractures (intra-articular). These groups are usually divided into subgroups based on the degree of displacement and dislocation of the fracture elements. According to the Neer classification system, the proximal humerus is divided into four anatomical fragments (humeral head, greater tuberosity, lesser tuberosity, and humeral shaft) as Codman suggested in 1934 [11]. Fracture displacement is defined as a 1 cm distance and/or 45 degree angle between elements [12]. A newer classification system, known as the LEGO system described by Hertel et al. [10], classifies proximal humeral fractures into 14 (12 plus 2) possible fracture patterns and is based on the location of fracture planes rather than fracture elements. With a classification based on fracture planes, Hertel was more able to assess humeral head vascularity.

Overall, classification systems aim to help the surgeon in making the correct treatment decision, but should also allow easier intra- and interdisciplinary communication [13, 14]. A better classification ability is expected to result in an improved understanding of the fracture pattern, which is important for the preoperative planning of fracture reduction and fixation. For a given classification system, independent observers should ideally agree on the same fracture type category regardless the level of experience. However, the amount and quality of information available to the observer may affect the accuracy of classification. Three-dimensional imaging of complex fractures using preoperative computed tomography (CT) is state of the art in most trauma centers. Nevertheless, the clinical interpretation of the CT image is often performed by scrolling through the 2D slices, not utilizing the entire potential of the data. Three-dimensional reconstructions are (occasionally) requested and used; however, the understanding of the number and the spatial arrangement of fragments remains challenging and may not aid the classification and preoperative planning. This may be overcome by enhanced visualization of the 3D CT images, in particular, by clearly demarcating the fragments. Computer-assisted image segmentation techniques can be used for this

purpose [15–19]. Nevertheless, it remains unknown if the 3D CT reconstructions can improve classification accuracy versus the 2D slice-wise analysis and if segmented 3D images provide any added value. Therefore, the aim of this study was to investigate the feasibility and added value of advanced visualization of segmented 3D CT images in proximal humeral fracture classification, compared to the conventional 3D CT reconstruction and the standard 2D slice-wise analysis.

Patients and methods

Data collection

Data of 17 patients with a preoperative CT scan were collected retrospectively from the database of the AZ Groeninge (Kortrijk, Belgium) hospital from 2015 until 2019, with treatment codes being either plate-screw-osteosynthesis or shoulder hemi-prosthesis. The mean age was 74 years (standard deviation: 10.6 years, range 58–87 years) and 14 patients were females. The left side was involved in 12 cases. Detailed patient characteristics are presented in Table 1. The study was approved by the ethical committee of the hospital under number AZGS2019072.

Image processing

All CT scans were anonymized using Synedra View Personal Version 19.0.0.1 (Synedra information technologies GmbH, Austria). The CT images were then segmented by means of a custom-developed script in MATLAB software package (Version R2019A, MathWorks, USA). Hereby, all fracture fragments were indicated with a different color (Fig. 1). The segmentation process was semi-automatic. First, the software pre-segmented the image automatically by separating clearly isolated fragments and presented these to the user in different colors in a 3D view and as 2D slices in three orthogonal views. The user could then refine the segmentation by indicating fragments that were not recognized in the first step by selecting one point on each fragment on the 3D view or on the 2D slices. This procedure was performed by a third-year orthopaedic resident (JD) selecting bone parts that were displaced more than 1 cm and/or angulated more than 45°, following the criteria of the Neer classification [12]. After this manual step, the software automatically identified the fracture line separating the two parts.

Fracture classification study

The ground truth of fracture classification was established for each case as a consensus between a senior experienced shoulder surgeon (GP) and a third-year orthopaedic trauma

Table 1 Patient characteristics

	Sex	Age at operation date (years)	Fractured side	Neer classification	AO/OTA classification	LEGO classification
Case 1	F	73	Left	3	11.B1	7
Case 2	F	75	Left	4	11.C3	12
Case 3	F	71	Right	3	11.B1	7
Case 4	M	58	Left	4	11.C3	12
Case 5	F	59	Left	4	11.C3	12
Case 6	F	82	Right	4	11.C3	12
Case 7	F	79	Right	Head split	11.C3	13
Case 8	F	83	Left	Head split	11.C3	14
Case 9	F	78	Right	Head split	11.C3	13
Case 10	F	70	Left	3	11.B1	7
Case 11	F	86	Left	2	11.A2	1
Case 12	F	74	Left	2	11.A3	1
Case 13	F	84	Left	3	11.B1	7
Case 14	F	87	Right	4	11.C3	12
Case 15	M	48	Left	Head split	11.C3	13
Case 16	F	76	Left	4	11.C3	12
Case 17	M	70	Left	3	11.B1	7

AO/OTA (Arbeitsgemeinschaft für Osteosynthesefragen/Orthopaedic Trauma Association)

resident (JD) based on careful inspection of all image data of a given patient. Fracture classification ability was then evaluated for two groups of surgeons, based on three modalities, into three classification systems and in two study designs. The first participant group consisted of orthopaedic residents in the first to fourth year of their training, whereas the second group consisted of experienced shoulder surgeons. The task was explained just before starting to each participant individually. They were asked to give their final answer after seeing every single modality. All individuals did the evaluations independently from each other in a quiet environment and were able to use a given paper sheet with the three printed classification systems (Neer [8], AO/OTA [9] and LEGO [10]) as a helping tool during the task. The time to complete the test was not recorded; however, all participants fulfilled the whole task at once. The three modalities were 2D CT slices, 3D CT reconstructions, and 3D segmentations with colored fragments, and were shown on a computer screen. For the 2D view, the participants were given the possibility to scroll through the standard CT slices of each proximal humeral fracture in the coronal, axial, and sagittal planes similar to the clinical situation. The 3D reconstruction, created by the radiology department of the hospital, could be rotated spatially to observe the fracture at different angles. The segmented 3D visualization was presented also with the possibility to turn the view three-dimensionally, and the body side was told to the participant, since the scapula was removed from the segmentation (Fig. 1). The participants were asked to classify the fracture into three

different classification systems: Neer classification [8], AO/OTA classification [9], and the LEGO system [10].

In the first study design, the tests were conducted in the hospital during the breaks of the daily practice and the three modalities of the same fracture were shown consecutively. The participants were twelve residents and five shoulder surgeons. In the second study design, nine residents and seven expert participants performed the tests in their free or research time outside the hospital and the fractures were shown in a random order and modality.

Statistical analysis

The effect of the image modality and the classification system was first evaluated with the Friedman test followed by the ad hoc Wilcoxon Signed-Rank test, separately for both study designs. A Bonferroni correction for multiple comparisons was applied when necessary. Within each study design, the results of surgeons and residents were pooled to improve the power of the test, since no significant differences between the participant groups were found. It was not the purpose of the study to compare results between the participants (interobserver variation). All statistical analyses were performed using Matlab R2019a (Mathworks, USA).

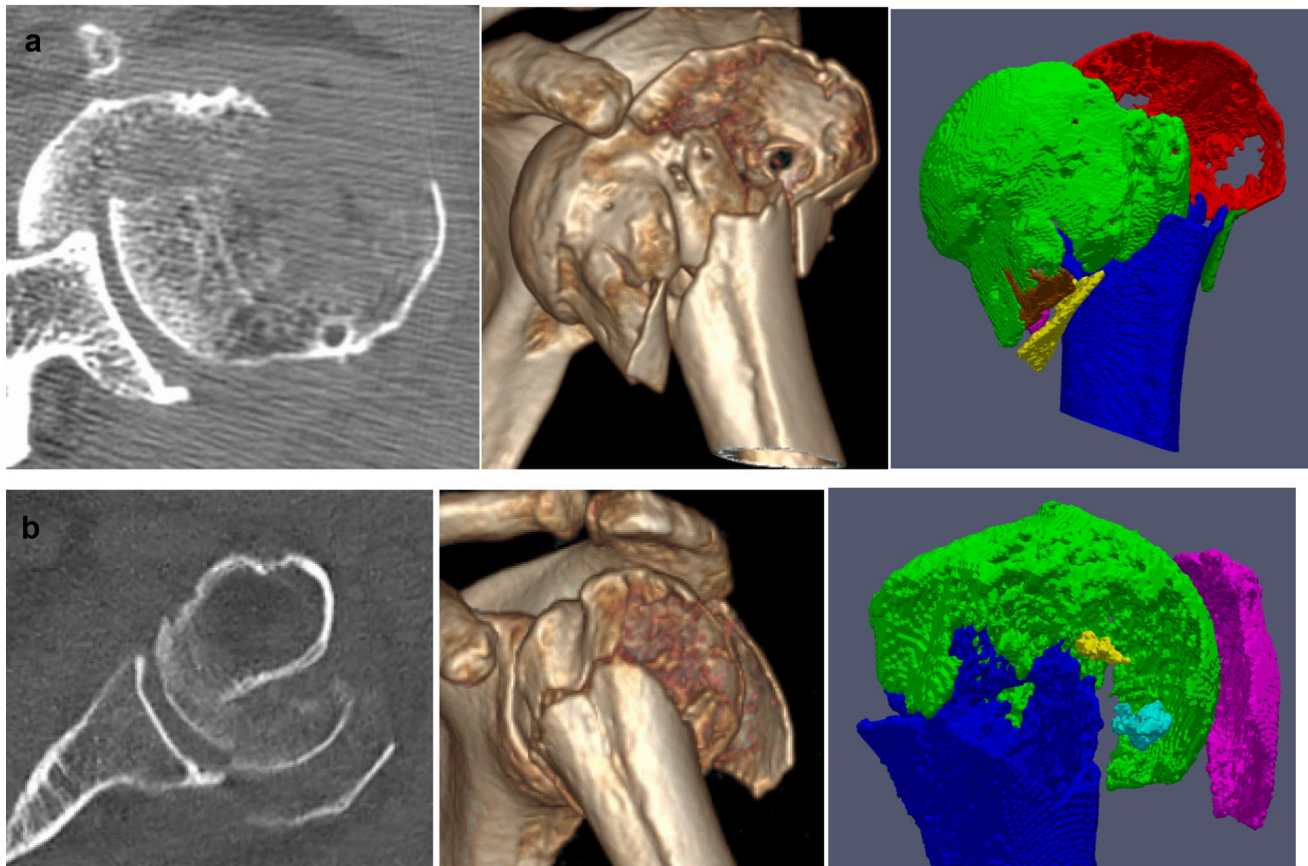


Fig. 1 **a** Head split proximal humeral fracture (case number 8 in Table 1). **b** Three-part proximal humeral fracture (case number 13 in Table 1). Respectively from left to right: a standard spiral CT scan

(axial view), the 3D model (rotating), and the 3D segmented model (rotating) with colored fracture fragments

Results

Regarding fracture classification of our included 17 patients, most were four-part (or LEGO type 12) fractures accounting for a total of six cases (35%). Five patients (29%) were diagnosed with three-part (or LEGO type 7) fractures and two patients with two-part (or LEGO type 1) fractures (12%). A total of four cases (24%) with head-split fractures (or LEGO type 13 and 14) were included in the study. A clear overview is presented in Table 1.

All participants in both study designs were able to classify proximal humeral fractures significantly more accurately ($p < 0.005$) into all three classification systems based on the 3D segmentations compared to the 2D slices or the 3D reconstructions (Fig. 2). Using the 3D segmentations, an average accuracy of $94\% \pm 5\%$ (mean \pm standard deviation (SD)) was achieved and the results were closely the same in both study designs and all three classification systems (range of mean values: 86–98%). Moreover, not only the average results were better with the 3D segmentation, but the results of the different participants were more

consistent, as well. These were indicated by the smaller SD values of the 3D segmentations (5% on average) compared to the 3D reconstructions (11%) or 2D slices (22%). An overview of the statistical results including p values is presented in Table 2.

The accuracies obtained with the 2D slice-wise analysis in the first study design ($23\% \pm 11\%$) showed a significantly inferior performance compared to the 3D reconstructions ($53\% \pm 10\%$) in all three classification systems ($p < 0.005$). Nevertheless, there were no significant differences between the two modalities in the second study design.

The participants experienced more difficulties with classifying proximal humeral fractures according to the LEGO system compared to the two other systems. The LEGO classification showed significant ($p < 0.05$) differences compared to the Neer and AO/OTA classification in all three modalities, except in comparison to the Neer classification in the 2D CT. Furthermore, no significant differences were found between the Neer and AO/OTA classification.

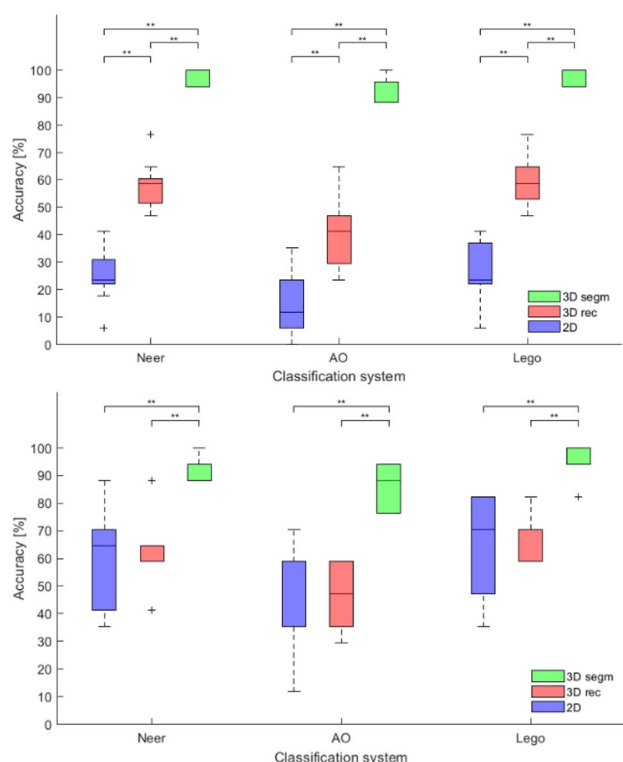


Fig. 2 Overview of the classification accuracies in study design 1 (top) and 2 (down) concerning the three different modalities and classification systems. **Statistically significant, 3D rec 3D reconstruction, 3D segm 3D segmentation

Discussion

The aim of the current study was to investigate the feasibility and added value of advanced visualization of segmented 3D CT images in proximal humeral fracture classification, compared to the conventional 3D CT reconstruction and the standard 2D slice-wise analysis. The 3D segmentation model, indicating every fracture fragment with another color, was the most successful modality. These models significantly improved the ability of all participants in classifying proximal humeral fractures into all three classification systems. The achieved accuracy was consistently the highest with the smallest scatter among the participants, indicating a robust and reproducible outcome using this modality. Our study utilized two different designs. In the first study design, the fractures were shown per patient in a fixed modality order, starting with the standard 2D CT slices, followed by the 3D reconstruction and finally the segmented 3D model. This situation is similar to the clinical practice where CT images can be viewed by the treating surgeon in as much detail as desired. Nevertheless, due to the pre-defined sequence of the three modalities, a learning effect or bias could be assumed in favor of the segmented 3D model; however, the correct classification of the fracture was never

provided to the participants. To avoid bias due to a possible learning effect, all image modalities and fractures were shown randomly in the second study design. Using the 3D segmentations, very similar classification accuracies were achieved in both study designs, indicating a high reproducibility, consistency, and clear superiority independent of the potential learning effect. These results were reproducible in the two observer groups, i.e., surgeons versus residents.

After discussing our results regarding the first modality, we focus on the conventional 3D CT reconstruction and the standard 2D slice-wise analysis. The poorest results were obtained in general with the 2D slice-wise evaluation. However, the study design influenced these results, showing much poorer outcomes in the first study design. The 2D slice-wise analysis was inferior compared to the 3D reconstruction in the first study design only. In the second study design, these two modalities showed similar level of benefit. These results were consistent for all three classification systems. The findings of the first study design question the relevance of using the 2D slice-wise analysis as the first step of the fracture interpretation using computed tomography in the clinical setting.

When comparing the classification systems, Neer and AO/OTA showed similar results for all modalities and both study designs. The LEGO classification appeared to be the most difficult for all participants compared to the other systems. A plausible explanation could lie in the fact that this is the youngest classification system and, therefore, not yet commonly used in every clinical practice. However, this classification system is not inferior compared to the two other systems when using the 3D segmentation model according to our results.

In addition to the above discussed results, some extra (qualitative) information was noticed during the experiment; the speed of the classification process was not precisely recorded, but all participants were noteworthy faster in classifying fractures using the 3D segmented model compared to the other two modalities (especially than the 2D slice-wise evaluation).

The present study found improved classification accuracy with segmented CT scans, which, in the future, could be incorporated into clinical practice. Better understanding of the fracture type may help orthopaedic surgeons to reconstruct proximal humeral fractures more reliably and accurately. Extensive preoperative planning might help to obtain a better and faster reduction in the operation room [20]. Optimal fracture reduction is one of the most important prerequisites to avoid fracture failure [4]. Therefore, enhanced understanding of the 3D extent and location of fracture fragments enabled by the advanced visualization may help surgeons in performing the required reduction maneuvers during surgery. However, there are several limitations of this study. First, semi-automatic segmentation depended partially

Table 2 Results of the statistical analysis with Bonferroni correction for multiple comparisons when necessary

	Test	Significancy between	Taken into account for analyses	Study design 1	Study design 2
Modalities	Friedman	2D—3D rec—3D segm	Neer	< 0.001	< 0.001
		2D—3D rec—3D segm	LEGO	< 0.001	< 0.001
		2D—3D rec—3D segm	AO	< 0.001	< 0.001
	Wilcoxon signed-rank	2D—3D rec	Neer	< 0.001	0.873
		3D rec—3D segm	Neer	< 0.001	< 0.001
		2D—3D segm	Neer	< 0.001	< 0.001
		2D—3D rec	LEGO	< 0.001	0.582
		3D rec—3D segm	LEGO	< 0.001	< 0.001
		2D—3D segm	LEGO	< 0.001	< 0.001
		2D—3D rec	AO	< 0.001	0.992
		3D rec—3D segm	AO	< 0.001	< 0.001
		2D—3D segm	AO	< 0.001	< 0.001
		LEGO-Neer-AO	2D	< 0.001	< 0.001
		LEGO-Neer-AO	3D rec	< 0.001	< 0.001
		LEGO-Neer-AO	3D segm	< 0.001	< 0.001
Classifications	Friedman	LEGO-Neer-AO	2D	< 0.001	< 0.001
		LEGO-Neer-AO	3D rec	< 0.001	< 0.001
		LEGO-Neer-AO	3D segm	< 0.001	< 0.001
	Wilcoxon signed-rank	Neer-LEGO	2D	0.052	0.067
		LEGO-AO	2D	0.018	0.014
		Neer-AO	2D	0.613	0.638
		Neer-LEGO	3D rec	< 0.001	0.003
		LEGO-AO	3D rec	< 0.001	< 0.001
		Neer-AO	3D rec	0.657	0.981
		Neer-LEGO	3D segm	< 0.001	0.010
		LEGO-AO	3D segm	< 0.001	0.003
		Neer-AO	3D segm	0.986	0.341

AO Arbeitsgemeinschaft für Osteosynthesefragen, 3D rec 3D reconstruction, 3D segm 3D segmentation

Statistical significance was defined as $p < 0.05$ in bold

on the observer selecting fracture fragments based on the program's initial suggestion. While this was done using the Neer criteria requiring a fracture gap of 1 cm and/or an angulation of 45°, the quality of the clinical CT scan may have influenced this by not allowing an appropriate segmentation of the cortical surfaces. The procedure was not fully automated as there was some user interaction required to supervise the segmentation process. However, this study did not aim to develop an autonomous procedure, yet the input of the operator was kept to a minimal. Second, in clinical practice, the first image modality requested to diagnose a fracture is usually X-rays in pre-defined views; afterwards, the CT-based analysis is performed only in complex cases. In this study, the X-rays were not shown before evaluating the CT scan in any modality. However, the CT scanning includes more detailed 3D information than the projected X-ray views and, therefore, this limitation is not expected to influence the findings. This study is considered a feasibility study because of the rather small sample size used for the number of analyses performed. Furthermore, implementing a semi-automatic segmentation process into clinical practice

would require further development and validation, resulting in a certified medical device. This was clearly beyond the scope of the current study and is a second reason that this work is considered as a feasibility study. Nevertheless, we could demonstrate the added value of this technique and, thus, believe that a fully automatic segmentation program in the future could be of great value in preoperative planning and, therefore, also help during the operation.

Conclusion

Semi-automated segmentation of CT scans provided a clear added value in proximal humeral fracture classification accuracy. Using this modality, orthopaedic shoulder surgeons and residents were able to classify fractures into the AO/OTA, Neer, and LEGO classification systems significantly better compared to the standard 2D CT slice-wise evaluation or the 3D reconstructions. The results of this feasibility study suggest that segmented CT scans are useful towards enhanced understanding the fracture, which could help to

achieve better fracture reduction and may decrease the risk of fixation failures.

Acknowledgements The authors are not compensated, and there are no other institutional subsidies, corporate affiliations, or funding sources supporting this work unless clearly documented and disclosed. This investigation was performed with the assistance of the AO Foundation. The authors would like to thank Carolin Danker for her contributions to writing of the semi-automatic segmentation program.

Funding None.

Compliance with ethical standards

Conflict of interest Stefaan Nijs reports grants from DePuy Synthes, as well as personal fees from DePuy Synthes, Zimmer Biomet, and Mathys Medical that are outside the submitted work. The other authors declare that they have no conflicts of interest to disclose.

Ethical approval This retrospective study was approved by the ethical committee of the hospital AZ Groeninge, Kortrijk. Study number AZGS2019072.

References

1. Court-Brown CM, Garg A, McQueen MM. The epidemiology of proximal humeral fractures. *Acta Orthop*. 2001;72:365–71.
2. Burkhart KJ, Dietz SO, Bastian L, Thelen U, Hoffmann R, Müller LP. Behandlung der proximalen Humerusfraktur des Erwachsenen. *Dtsch Arztebl Int*. 2013;110:591–7.
3. Hardeman F, Bollars P, Donnelly M, Bellemans J, Nijs S. Predictive factors for functional outcome and failure in angular stable osteosynthesis of the proximal humerus. *Injury*. 2012;43:153–8.
4. Krappinger D, Bizzotto N, Riedmann S, Kammerlander C, Hengg C, Kralinger FS. Predicting failure after surgical fixation of proximal humerus fractures. *Injury*. 2011;42:1283–8. <https://doi.org/10.1016/j.injury.2011.01.017> (Elsevier Ltd).
5. Meier RA, Messmer P, Regazzoni P, Rothfischer W, Gross T. Unexpected high complication rate following internal fixation of unstable proximal humerus fractures with an angled blade plate. *J Orthop Trauma*. 2006;20:253–60.
6. Agudelo J, Schurmann M, Stahel P, Helwig P, Morgan SJ, Zechel W, et al. Analysis of efficacy and failure in proximal humerus fractures treated with locking plates. *J Orthop Trauma*. 2007;21:676–81.
7. Kralinger F, Blauth M, Goldhahn J, Käch K, Voigt C, Platz A, et al. The influence of local bone density on the outcome of one hundred and fifty proximal humeral fractures treated with a locking plate. *J Bone Joint Surg Am*. 2014;96:1026–32.
8. Neer CS. 2nd. Displaced proximal humeral fractures: part I. Classification and evaluation. 1970. *Clin Orthop Relat Res*. 2006;442:77–82.
9. Meinberg E, Agel J, Roberts C. Fracture and dislocation classification compendium—2018. *J Orthop Trauma*. 2018;32:s1–s10.
10. Hertel R, Hempfing A, Stiehler M, Leunig M. Predictors of humeral head ischemia after intracapsular fracture of the proximal humerus. *J Shoulder Elb Surg*. 2004;13:427–33.
11. Codman E. Fractures in relation to the subacromial bursa. The shoulder. Boston: Thomas Todd; 1934.
12. Neer CS. 2nd. Displaced proximal humeral fractures. Part I. Classification and evaluation By Charles S. Neer, I 1970. *Clin Orthop Relat Res*. 1987;223:3–10.
13. Papakostantinou MK, Hart MJ, Farrugia R, Gabbe BJ, Kamali Moaveni A, van Bavel D, et al. Interobserver agreement of Neer and AO classifications for proximal humeral fractures. *ANZ J Surg*. 2016;86:280–4. <https://doi.org/10.1111/ans.13451>.
14. Burstein AH. Fracture classification systems: do they work and are they useful? *J Bone Joint Surg Am*. 1993;75:1743–4.
15. Fürnstahl P, Székely G, Gerber C, Hodler J, Snedeker JG, Harders M. Computer assisted reconstruction of complex proximal humerus fractures for preoperative planning. *Med Image Anal*. 2012;16:704–20.
16. Khanna K, Brabston EW, Qayyum U, Gardner TR, Levine WN, Jobin CM, et al. Proximal Humerus Fracture 3-D Modeling. *Am J Orthop (Belle Mead NJ)*. 2018;47:4.
17. Xia S, Zhang Y, Wang X, Wang Z, Wang W, Ma X, et al. Computerized virtual surgery planning for ORIF of proximal humeral fractures. *Orthopedics*. 2015;38:e428–e433433.
18. Chen Y, Zhang K, Qiang M, Li H, Dai H. Computer-assisted preoperative planning for proximal humeral fractures by minimally invasive plate osteosynthesis. *Chin Med J (Engl)*. 2014;127:3278–85.
19. You W, Liu LJ, Chen HX, Xiong JY, Wang DM, Huang JH, et al. Application of 3D printing technology on the treatment of complex proximal humeral fractures (Neer3-part and 4-part) in old people. *Orthop Traumatol Surg Res*. 2016;102:897–903.
20. Wang D, Zhang K, Qiang M, Jia X, Chen Y. Computer-assisted preoperative planning improves the learning curve of PFNA-II in the treatment of intertrochanteric femoral fractures. *BMC Musculoskelet Disord*. 2020;21:34. <https://doi.org/10.1186/s12891-020-3048-4>.