



Contents lists available at ScienceDirect

JSES International

journal homepage: www.jseinternational.org

Interobserver and intraobserver agreement of three-dimensionally printed models for the classification of proximal humeral fractures

Hannah Bougher, MBBS ^a, Petra Buttner, BSc, MSc, PhD ^b, Jonathon Smith, BMBS ^c, Jennifer Banks, MBS, PhD ^a, Hyun Su Na, MD ^c, David Forrestal, BEng ^d, Clare Heal, MBChB, PhD ^{a,*}

^a James Cook University, Mackay Clinical School, Mackay, QLD, Australia

^b James Cook University, Smithfield, QLD, Australia

^c Mackay Base Hospital, Mackay, QLD, Australia

^d Queensland University of Technology, Brisbane City, QLD, Australia

ARTICLE INFO

Keywords:

Proximal humeral fracture
interobserver agreement
intraobserver agreement
3D modeling
fracture classification
Neer system

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

Hypothesis: This study aimed to examine whether three-dimensionally printed models (3D models) could improve interobserver and intraobserver agreement when classifying proximal humeral fractures (PHFs) using the Neer system. We hypothesized that 3D models would improve interobserver and intraobserver agreement compared with x-ray, two-dimensional (2D) and three-dimensional (3D) computed tomography (CT) and that agreement using 3D models would be higher for registrars than for consultants.

Methods: Thirty consecutive PHF images were selected from a state-wide database and classified by fourteen observers. Each imaging modality (x-ray, 2D CT, 3D CT, 3D models) was grouped and presented in a randomly allocated sequence on two separate occasions. Interobserver and intraobserver agreements were quantified with kappa values (κ), percentage agreement, and 95% confidence intervals (CIs).

Results: Seven orthopedic registrars and seven orthopedic consultants classified 30 fractures on one occasion (interobserver). Four registrars and three consultants additionally completed classification on a second occasion (intraobserver). Interobserver agreement was greater with 3D models than with x-ray ($\kappa = 0.47$, CI: 0.44–0.50, 66.5%, CI: 64.6–68.4% and $\kappa = 0.29$, CI: 0.26–0.31, 57.2%, CI: 55.1–59.3%, respectively), 2D CT ($\kappa = 0.30$, CI: 0.27–0.33, 57.8%, CI: 55.5–60.2%), and 3D CT ($\kappa = 0.35$, CI: 0.33–0.38, 58.8%, CI: 56.7–60.9%). Intraobserver agreement appeared higher for 3D models than for other modalities; however, results were not significant. There were no differences in interobserver or intraobserver agreement between registrars and consultants.

Conclusion: Three-dimensionally printed models improved interobserver agreement in the classification of PHFs using the Neer system. This has potential implications for using 3D models for surgical planning and teaching.

© 2020 The Author(s). Published by Elsevier Inc. on behalf of American Shoulder and Elbow Surgeons. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Three-dimensional (3D) printing is an emerging technology in orthopedics, with its uses ranging from the development of customized implants, surgical templates, and bioprinted bone to the use of models for teaching and surgical planning.^{6,13,19,37}

This research was supported by a \$1000 honors research study grant from James Cook University.

This study was approved by the James Cook University Human Research Ethics Committee [H7392].

* Corresponding author: Clare Heal, MBChB, PhD, James Cook University, Mackay Clinical School, Mackay, QLD 4740, Australia.

E-mail address: clare.heal@jcu.edu.au (C. Heal).

<https://doi.org/10.1016/j.jseint.2020.10.019>

2666-6383/© 2020 The Author(s). Published by Elsevier Inc. on behalf of American Shoulder and Elbow Surgeons. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Proximal humeral fractures (PHFs) are the fourth most common osteoporotic fracture, affecting Australian men and women at rates of 40.6 and 73.2 per 100,000 person-years, respectively.^{14,16} Currently 43% of patients are hospitalized, 21% receive surgery, and 15% die within 3 months,³⁰ and with an aging population, the incidence and burden of disease will likely increase. Accurate fracture classification has important implications for diagnosis, surgical management, and planning, as well as estimation of patient prognosis.²²

PHFs are most commonly classified using the Neer system according to the number of parts displaced by greater than 1 cm or a 45° angle and grades complexity increasing from one- to four-part fractures.^{18,27} Without a gold standard for the Neer system,

interobserver and intraobserver agreement can be used as surrogates for validity and reliability. Like other PHF classification systems, the Neer system has shown limited interobserver and intraobserver agreement.^{7,8,11,28} A recent systematic literature review found that levels of interobserver and intraobserver agreement for PHF classification was lowest for x-ray, increased with two-dimensional (2D) computed tomography (CT), and highest with 3D CT. The same study suggested CT may increase interobserver agreement to a greater extent for less experienced observers.⁴

Conventional imaging modalities may limit the ability of surgeons to interpret *in vivo* anatomy.^{11,27} Three-dimensionally printed models (herein referred to as 3D models) have theoretical advantages over x-rays and CT as they allow tactile examination of anatomy,^{15,19} and unlimited 360° visualization of the fracture,^{35–37} thereby avoiding the need to interpret 3D patho-anatomy from a 2D screen.^{11,15} The Neer system was designed to be applied after examining intraoperative anatomy.²⁶ By replicating the fracture and simulating the intraoperative findings, 3D models allow classification to be applied similarly to the original design of the Neer system.

The aim of our study was to investigate orthopedic surgeons' interobserver and intraobserver agreement with 3D models using the Neer system. The primary hypothesis was that 3D models would improve interobserver agreement by a kappa value of 0.1 in comparison with x-ray. The secondary hypotheses were that 1) 3D models would improve agreement compared with 2D and 3D CT and 2) agreement using 3D models would be higher (kappa 0.15) for registrars than for consultants.

Materials and methods

Setting

The study was conducted from March to July 2019, at an Australian regional general hospital.

Participants

Fourteen observers (seven orthopedic registrars and seven orthopedic consultants, who comprised all relevant staff members at the participating regional hospital) were invited to participate. Registrars (the Australian term for the equivalent of the US resident) were principal house officers or held an Australian Orthopaedic Association Surgical Education and Training Program position. Consultants were general orthopedic surgeons employed as specialists in the private or public system.

The head of the department retrospectively selected thirty eligible PHFs from a state-wide database from December 18, 2018, until equal numbers of consecutive two-, three-, and four-part fractures were available (age range: 49–96 years; median age: 73 years; 28 female). To be eligible for inclusion, x-rays and 2D and 3D CT scans must have been available. For the purpose of fracture selection only, fracture severity was determined by the head of the department with all available imaging according to the Neer system.²⁷ One-part fractures were excluded as many of these patients do not receive a CT and the lack of displacement would have made fracture lines difficult to visualize with 3D models.¹⁵ All imaging had been used clinically and captured before callus formation. Two-dimensional CT had an axial primary image plane and slice thickness of 1 mm or less. Two-dimensional CT DICOM (Digital Imaging and Communications in Medicine) files were converted to STL (Standard Tessellation Language) files using Slicer, version 4.10.1, and Blender, version 2.79, before being printed with a 3D printer.

Models were printed from polylactic acid thermoplastic material using an Ultimaker 2+ (Ultimaker B.V, Utrecht, Netherlands) 3D printer using the fused filament fabrication method. STL files containing the 3D model data were imported into Ultimaker Cura (4.2.1) software and converted to GCODE files containing the 3D printing machine instructions for manufacturing the models. Models were printed with a 0.4-mm nozzle, 0.15-mm layer height, and 20% grid infill density, and sacrificial support material was added on regions of the model with 50° or greater overhang angle. To minimize the amount of required support material, models were aligned with the humeral head located on the build plate and the shaft extending vertically upward. After printing, the support material was removed before use in the study.

Procedure

Before classification session one, observers watched a 5-minute prerecorded PowerPoint presentation defining the Neer system.²⁷ Classification was recorded on fixed response surveys (Appendix 1). Images were deidentified and presented in a randomly allocated sequence for each grouped image modality, to prevent observers correlating images across modalities. Observers classified fractures individually without time restriction (representative images for each modality are provided in Figure 1) for representation. No clinical details were provided. Anterior-posterior and lateral x-rays were displayed as JPEG files. Coronal, sagittal, and longitudinal 2D CT scans, as well as axially rotating 3D CT scans, were displayed on interactive software (IntelViewer, Intelrad, Montreal, Canada). Observers could manipulate imaging and handle 3D models.

Observers completed a second identical classification session three to eight weeks later. The same images were viewed in a different randomly allocated sequence. No feedback was provided at any point in the study, and images were not available between sessions.

Sample size

It was calculated that a sample size of 30 fractures (ten each of two-, three-, and four-part fractures) would result in 95% confidence intervals (95% CIs) for kappa smaller than 0.1 in width when the assessments of fourteen observers are combined and 95% CIs for kappa of 0.15 in width for subgroups with seven registrars and seven consultants separately. This sample size was calculated to confirm or reject the primary hypothesis, comparing x-rays with 3D models. It assumed that the study would recruit 14 observers (seven registrars and seven consultants) and that two-, three-, and four-part fractures (three categories) would be differentiated, each occurring with similar frequency. The sample size estimation was adjusted for multiple testing ($k = 6$), allowing the assessment of the interobserver agreement of all fourteen observers together, of the subgroup of registrars and the subgroup of consultants, separately for x-rays and 3D models.¹⁰

Statistical analysis

Stata IC 13 (StataCorp, College Station, TX, USA) was used to calculate interobserver and intraobserver agreement kappa (κ) values (nonunique raters, no weighting) and percentage agreement with 95% CIs. Interobserver and intraobserver percentage agreements were calculated as the mean values of overall agreement for all possible combinations of assessors within each imaging modality. Kappa for interobserver agreement was based on all assessors in the respective analysis. Intraobserver agreement κ values were calculated for each of the seven observers who repeated the

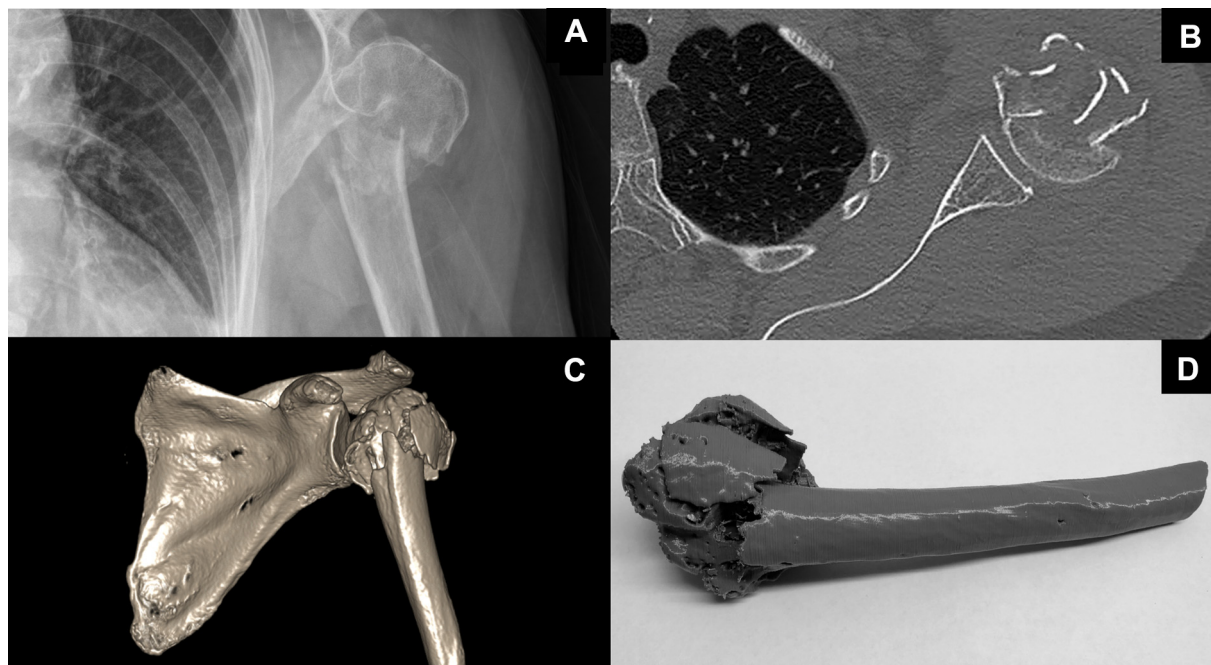


Figure 1 Representative images of proximal humeral fractures that observers were asked to classify using the Neer system: (A) X-ray, (B) 2D CT, (C) 3D CT, (D) 3D printed model. 2D, two-dimensional; 3D, three-dimensional; CT, computed tomography.

classification session and then averaged. Kappa values were interpreted using Landis and Koch criteria (Table 1).²⁰ The difference between two κ values was considered statistically significant if 95% CIs did not overlap ($P < .05$). Furthermore, P values were calculated for each κ value to test for statistically significant difference from “0.”

Results

Fourteen observers (seven orthopedic registrars and seven orthopedic consultants) participated in the initial interobserver agreement study. Of these, seven (three consultants and four registrars) completed the intraobserver component.

Interobserver agreement

Three-dimensionally printed models significantly improved overall interobserver agreement compared with x-ray ($\kappa = 0.47$, CI: 0.44–0.50, 66.5%, CI: 64.6–68.4% and $\kappa = 0.29$, CI: 0.26–0.31, 57.2%, CI: 55.1–59.3%, respectively, Table II, Figure 2). They also produced significantly better agreement than 2D CT ($\kappa = 0.30$, CI: 0.27–0.33, 57.8%, CI: 55.5–60.2%) and 3D CT ($\kappa = 0.35$, CI: 0.33–0.38, 58.8%, CI: 56.7–60.9%, Table II, Figure 2). Three-dimensionally printed models achieved moderate interobserver agreement with the Landis and Koch criteria compared with fair agreement for x-rays and 2D and 3D CT (Table II, Figure 2).²⁰

There was no significant difference in the level of interobserver agreement produced by 3D models in registrars compared with consultants ($\kappa = 0.51$, CI: 0.44–0.57, 68.3%, CI: 64.4–72.3% and $\kappa = 0.48$, CI: 0.42–0.55, 66.8%, CI: 62.3–71.3%, respectively, Table II, Figure 2).

Intraobserver agreement

Three-dimensionally printed models produced better intraobserver agreement than x-rays (mean $\kappa = 0.60$, CI: 0.46–0.73,

Table 1
Landis and Koch criteria²⁰

Kappa value	Agreement
Less than 0.00	Poor agreement
0.00–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–1.00	Almost perfect agreement

mean agreement: 75.0%, CI: 66.4–83.6% and mean $\kappa = 0.45$, CI: 0.36–0.54, mean agreement: 70.5%, CI: 62.8–78.1%, respectively, Table III, Figure 3), but results were not statistically significant. Three-dimensionally printed models also provided higher intraobserver agreement than 2D CT and 3D CT (Table III, Figure 3), although results were also not statistically significant. Consultants achieved higher intraobserver agreement than registrars with 3D models (mean $\kappa = 0.69$ and mean agreement: 82.0% for consultants versus mean $\kappa = 0.52$ and mean agreement: 69.8% for registrars, Table III, Figure 3). We did not analyze these results further because of the small sample size.

Discussion

For the interobserver agreement component of the study, the primary hypothesis was confirmed, with 3D models significantly increasing interobserver agreement compared with x-ray. Three-dimensionally printed models also significantly improved interobserver agreement in comparison with 2D CT and 3D CT. There was no significant difference in agreement using 3D models between consultants and registrars. Under the Landis and Koch criteria, interobserver agreement was classified as moderate with 3D models, while fair agreement was achieved with x-rays, 2D CT, and 3D CT. Only seven of the original fourteen observers completed the intraobserver component of the study. While 3D models produced higher intraobserver agreement than x-ray, 2D CT, and 3D CT,

Table II
Interobserver agreement

	X-ray	2D CT	3D CT	3D models
Overall (consultants and registrars, n = 14)				
kappa				
κ	0.29	0.30	0.35	0.47
95% CI	0.26–0.31	0.27–0.33	0.33–0.38	0.44–0.50
Agreement*	Fair	Fair	Fair	Moderate
P value†	<.001	<.001	<.001	<.001
% agreement				
%	57.2	57.8	58.8	66.5
95% CI	55.1–59.3	55.5–60.2	56.7–60.9	64.6–68.4
Number of images	30	29	28	26
Consultants (n = 7)				
kappa				
κ	0.26	0.37	0.30	0.48
95% CI	0.20–0.32	0.31–0.43	0.24–0.36	0.42–0.55
Agreement	Fair	Fair	Fair	Moderate
P value	<.001	<.001	<.001	<.001
% agreement				
%	56.0	62.9	57.1	66.8
95% CI	50.2–61.9	56.3–69.4	52.0–62.3	62.3–71.3
Number of images	30	29	30	27
Registrars (n = 7)				
kappa				
κ	0.35	0.25	0.39	0.51
95% CI	0.29–0.40	0.19–0.31	0.34–0.45	0.44–0.57
Agreement	Fair	Fair	Fair	Moderate
P value	<.001	<.001	<.001	<.001
% agreement				
%	60.3	54.1	60.1	68.3
95% CI	57.3–63.4	49.0–59.3	55.1–65.0	64.4–72.3
Number of images	30	30	28	29

2D, two-dimensional; 3D, three-dimensional; CI, confidence interval; CT, computed tomography.

* Agreement has been defined using the Landis and Koch criteria (Table I).²⁰

† P value less than .05 shows that kappa was statistically significantly different from "0".

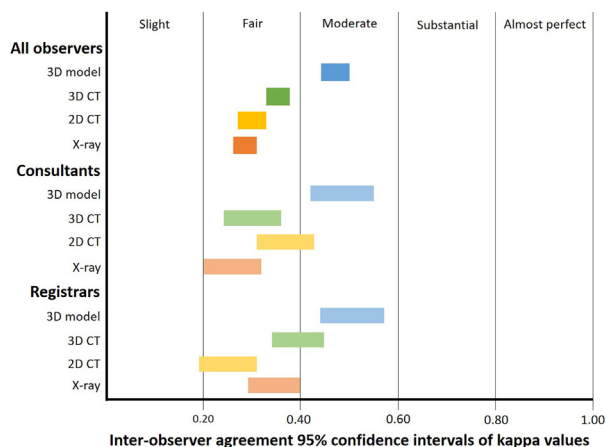


Figure 2 Interobserver agreement using Landis and Koch criteria for each modality and group – all observers (top), consultants (middle), and registrars (bottom).

this did not reach statistical significance. Three-dimensionally printed models achieved substantial intraobserver agreement with the Landis and Koch criteria, while all other imaging modalities achieved moderate agreement.

This study had a number of strengths. First, the consecutive selection of fractures from a state-wide database resulted in clinically realistic images. Second, for the interobserver agreement component of the study, the predetermined sample size was achieved, and the study was adequately powered for the primary hypothesis, resulting in narrow CIs and significant results. This study included more observers than many previous studies.^{2,7,11,17,29}

Finally, the use of a standard presentation before rating ensured a standard definition of the Neer system was applied by raters.

This study also had a number of limitations. In the absence of a gold standard, the original selection of 10 of each of 2-, 3-, and 4-part fracture for the purpose of the study was conducted by the head of the department; however, equality of the number of each type of fracture cannot be assured. With the absence of a gold standard for classification, and also with low baseline levels of agreement, the allocation of fractures to a level of complexity to conduct a subanalysis by complexity would have been arbitrary; therefore, this subanalysis was not performed.² The intraobserver part of the study was only conducted with seven observers, leading to reduced statistical power for the analysis. As a consequence, comparisons between observer groups were not conducted.

In clinical practice, different imaging modalities are used simultaneously to assess a fracture. However, our study required grouping by image modality to allow agreement to be attributed to a specific imaging modality rather than cumulative familiarity with the fracture.

The time between repeat fracture classifications varied for observers between three to eight weeks. Although identical timing would have been ideal, this was not feasible owing to doctor availability. Images were not available between sessions, and feedback was not given after the first session. We believe it is unlikely that observers could recall previous classification and consequently unlikely that the difference in timing impacted intraobserver results.

Kappa values were used to allow comparison to previous studies. Although kappa statistics correct for agreement occurring by chance, they have limitations. Prevalence and bias effect, resulting from marginal proportions inherent in calculating kappa, means kappa can be low when percentage agreement is high.³² Because the Landis and Koch criteria are arbitrary, the effects of prevalence and bias effects on kappa should still be considered.³² Supplementing kappa with percentage agreement adds statistical credibility as it is not affected by prevalence and bias effects. Although the Neer classification system is known to have limited reliability and validity, it is the most widely used system and therefore the most appropriate for our study.¹⁸

There has been very limited research addressing 3D models as an imaging modality for PHF classification, and to the authors' knowledge, this is the first study specifically measuring agreement.^{9,37} However, there has been more extensive research for other types of fractures. Three-dimensionally printed models have been found to increase interobserver agreement in the classification of acetabular, calcaneal, and coronoid and distal humeral fractures in comparison with 2D CT.^{6,12,15,24}

A previous study had hypothesized that 3D models might be more helpful for less experienced surgeons diagnosing PHFs.³⁷ However, in the classification of acetabular and calcaneal fractures, 3D models did not produce greater agreement in less experienced observers.^{15,24} Similarly, registrars did not appear to benefit to a greater extent from the use of 3D models than consultants in our study.

A key factor contributing to limited agreement with the Neer system is a lack of clarity regarding the threshold for displacement.^{3,4,34} Historically, this has varied leading to some uncertainty regarding its definition.²⁵ Displacement threshold of 1 cm likely appears more substantial if the humeral head diameter is 3.5 cm rather than 5 cm, which could contribute to disagreement. Fractures with displacement closer to the 1-cm threshold are likely to be more contentious.^{4,27,31} Muscles attached to fragments may cause displacement to change between different images.^{26,31} Owing to the round structure, it is challenging to appreciate angulation or displacement of the humeral head.²⁷ There is likely more disagreement about complex fractures,^{4,5,23,34} where it

Table III
Intraobserver agreement

	X-ray	2D CT	3D CT	3D models
Consultants and registrars (n = 7)				
kappa				
Mean κ	0.45	0.41	0.43	0.60
Range of κ	0.32-0.63	0.10-0.64	0.36-0.53	0.45-0.89
95% CI	0.36-0.54	0.25-0.57	0.37-0.48	0.46-0.73
Agreement*	Moderate	Moderate	Moderate	Moderate
Range of P values [†]	P = .016 to P < .0001	P = .201 to P < .0001	P = .004 to P < .0001	P = .0005 to P < .0001
Mean % agreement				
%	70.5	65.6	65.5	75.0
95% CI	62.8-78.1	53.9-77.3	61.2-69.9	66.4-83.6
Number of images	30	30 [‡]	30 [‡]	30 [§]
Consultants (n = 3)				
kappa				
Mean κ	0.46	0.45	0.39	0.69
Range of κ	0.42-0.48	0.33-0.56	0.36-0.44	0.51-0.89
Agreement	Moderate	Moderate	Moderate	Substantial
Mean % agreement	74.4	70.8	65.6	82.0
Registrars (n = 4)				
kappa				
Mean κ	0.44	0.37	0.45	0.52
Range of κ	0.32-0.63	0.10-0.64	0.40-0.53	0.45-0.60
Agreement	Moderate	Fair	Moderate	Moderate
Mean % agreement	67.5	61.7	65.5	69.8

2D, two-dimensional; 3D, three-dimensional; CI, confidence interval; CT, computed tomography.

* Agreement has been defined using the Landis and Koch criteria (Table 1).²⁰

[†] P value less than .05 shows that kappa was statistically significantly different from "0". Range of P values assessing kappa for each observer.

[‡] 29 images for one observer

[§] 29 images for two observers.

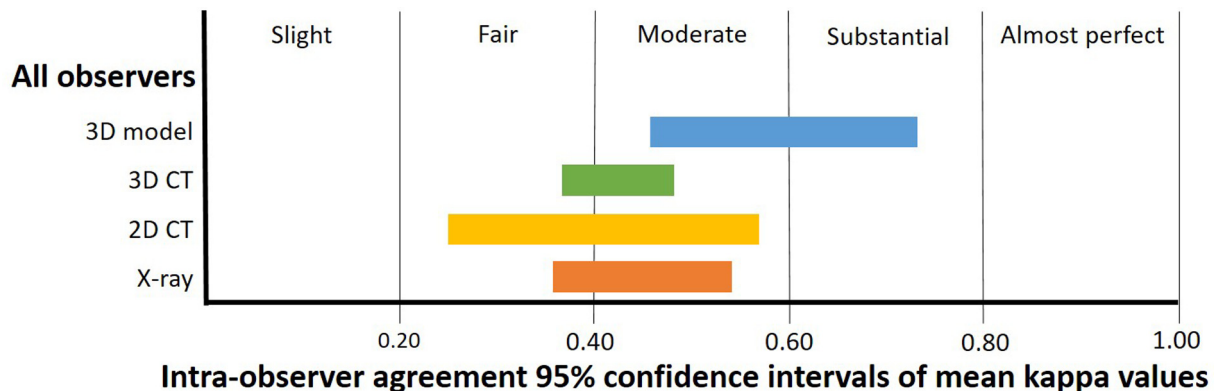


Figure 3 Intraobserver agreement using Landis and Koch criteria for each modality for all observers.

becomes increasingly difficult to assess displacement.⁴ However, the amount of displacement helps determine management and is integral to PHF classification systems.²⁵⁻²⁷

Fracture classification is important for reporting of injury severity, surgical management, and surgical planning and estimating the likely prognosis, but there is currently no evidence that fracture classifications improve patient outcomes. However, the ability to visualize individual patient anatomy using 3D models for PHF surgery has been shown to decrease operating time, blood loss, and radiation exposure owing to reduced imaging and improve functional outcomes (including shoulder range of motion and Short Form-36 physical component summary scores) compared with the conventional 2D and 3D imaging.⁹ The improvement in agreement using the Neer system in our study is in essence a surrogate marker for the increase in ability for surgeons to visualize and understand fractures using 3D models compared with conventional imaging.

Clinical implications

We suggest that x-rays should still remain first line in view of low cost and limited exposure to radiation.¹ The improved inter-observer agreement in our study with the use of CT supports the current practice of adding CT when further information is required. The addition of 3D models does not require additional radiation or patient discomfort.^{21,26} With newer technology, 3D models will be quicker to produce, cheaper, and more widely available and therefore more feasible to use as an additional modality. Surgical planning with PHF 3D models has been shown to improve patient outcomes and reduce operative time, blood loss, and the use of intraoperative x-rays,^{9,37} thereby reducing costs and potentially offsetting the cost of 3D printing. Use of 3D models for informed consent may also improve patient understanding and satisfaction and reduce potential litigation.^{15,35-37}

Research implications

Limited agreement was found with the Neer system, suggesting new classification systems should be investigated, including the HGLS system, which has been shown to have superior interobserver and intraobserver reliability compared with both the Neer and AO systems.³³ Three-dimensionally printed models are likely to increase agreement for other classification systems. Future, adequately powered research should determine if 3D models are more useful for complex fractures. The use of an expert panel as a gold standard to define fracture complexity could be useful in this setting. Prospective research is required to confirm if improved agreement about PHF classification with 3D models improves treatment consistency and patient outcomes. A formal cost-effectiveness study could assess if improved outcomes justify the cost of 3D printing. The ability of PHF 3D models to improve patient understanding and expectations, and aid in gaining informed consent, should also be formally assessed.

Conclusions

The use of 3D models significantly improved interobserver agreement about the Neer classification system in comparison with x-rays. Three-dimensionally printed models also significantly improved interobserver agreement over 2D and 3D CT. There was also increased intraobserver agreement with 3D models, although not statistically significant owing to lack of statistical power. Three-dimensionally printed models did not significantly benefit agreement of registrars compared with consultants. Future research into the use of 3D models in PHF management should further investigate the ability of this technology to improve patient outcomes.

Acknowledgments

The authors thank the orthopedic consultants and registrars who participated in this study.

Disclaimer

No financial remunerations were received by the authors or their family members related to the subject of this article. The authors have no conflicts of interest to declare.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jseint.2020.10.019>.

References

1. Australian GDoh. MBS Online. <http://www9.health.gov.au/mbs/search.cfm?rpp=10&q=57506%2D61110&qt=&sopt=S&st=y&start=1>. Accessed November 6, 2019.
2. Berkes MB, Dines JS, Little MT, Garner MR, Shifflett GD, Lazaro LE, et al. The impact of three-dimensional CT imaging on intraobserver and interobserver reliability of proximal humeral fracture classifications and treatment recommendations. *J Bone Joint Surg Am* 2014;96:1281-6. <https://doi.org/10.2106/JBJS.M.00199>.
3. 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.
4. Bougher H, Nagendiram A, Banks J, Hall LM, Heal C. Imaging to improve agreement for proximal humeral fracture classification in adult patient: A systematic review of quantitative studies. *J Clin Orthop Trauma* 2020;11(Suppl 1):S16-24. <https://doi.org/10.1016/j.jcot.2019.06.019>.
5. Brorson S. Fractures of the proximal humerus. *Acta Orthop Suppl* 2013;84(suppl351):1-32. <https://doi.org/10.3109/17453674.2013.826083>.
6. Brouwer KM, Lindenhovius AL, Dyer GS, Zurakowski D, Mudgal CS, Ring D. Diagnostic accuracy of 2- and 3-dimensional imaging and modeling of distal

7. humerus fractures. *J Shoulder Elbow Surg* 2012;21:772-6. <https://doi.org/10.1016/j.jse.2012.01.009>.
7. Brunner A, Honigsmann P, Treumann T, Babst R. The impact of stereo-visualisation of three-dimensional CT datasets on the inter- and intra-observer reliability of the AO/OTA and Neer classifications in the assessment of fractures of the proximal humerus. *J Bone Joint Surg Br* 2009;91:766-71. <https://doi.org/10.1302/0301-620X.91B6.22109>.
8. Carrerra EdF, Wajnsztein A, Lenza M, Archetti Netto N. Reproducibility of three classifications of proximal humeral fractures. *Einstein (São Paulo)* 2012;10:473-9. <https://doi.org/10.1590/S1679-45082012000400014>.
9. Chen Y, Jia X, Qiang M, Zhang K, Chen S. Computer-assisted virtual surgical technology versus three-dimensional printing technology in preoperative planning for displaced three and four-part fractures of the proximal end of the humerus. *J Bone Joint Surg Am* 2018;100:1960-8. <https://doi.org/10.2106/JBJS.18.00477>.
10. Fleiss JL. *Statistical methods for raters and proportions*. 2nd ed. New York, NY: John Wiley & Sons; 1981.
11. Foroohar A, Tosti R, Richmond JM, Gaughan JP, Ilyas AM. Classification and treatment of proximal humerus fractures: inter-observer reliability and agreement across imaging modalities and experience. *J Orthop Surg Res* 2011;6:38. <https://doi.org/10.1186/1749-799X-6-38>.
12. Guitton T, Kinaci A, Ring D. Diagnostic accuracy of 2- and 3-dimensional computed tomography and solid modeling of coronoid fractures. *J Shoulder Elbow Surg* 2013;22:782-6. <https://doi.org/10.1016/j.jse.2013.02.009>.
13. Hoekstra H, Rosseels W, Sermon A, Nijs S. Corrective limb osteotomy using patient specific 3D-printed guides: a technical note. *Injury* 2016;47:2375-80. <https://doi.org/10.1016/j.injury.2016.07.021>.
14. Holloway KL, Bucki-Smith G, Morse AG, Brennan-Olsen SL, Kotowicz MA, Moloney DJ, et al. Humeral fractures in south-eastern Australia: epidemiology and risk factors. *Calcif Tissue Int* 2015;97:453-65. <https://doi.org/10.1007/s00223-015-0039-9>.
15. Hurson C, Tansey A, O'Donnchadha B, Nicholson P, Rice J, McElwain J. Rapid prototyping in the assessment, classification and preoperative planning of acetabular fractures. *Injury* 2007;38:1158-62. <https://doi.org/10.1016/j.injury.2007.05.020>.
16. Imai N, Endo N, Shobugawa Y, Oinuma T, Takahashi Y, Suzuki K, et al. Incidence of four major types of osteoporotic fragility fractures among elderly individuals in Sado, Japan, in 2015. *J Bone Miner Metab* 2019;37:484-90. <https://doi.org/10.1007/s00774-018-0937-9>.
17. Iordens GI, Mahabier KC, Buisman FE, Schep NW, Muradin GS, Beenen LF, et al. The reliability and reproducibility of the Hertel classification for comminuted proximal humeral fractures compared with the Neer classification. *J Orthop Sci* 2016;21:596-602. <https://doi.org/10.1016/j.jos.2016.05.011>.
18. Khmelnitskaya E, Lamont L, Taylor S, Lorich D, Dines D, Dines J. Evaluation and management of proximal humerus fractures. *Adv Orthop* 2012;2012:861598. <https://doi.org/10.1155/2012/861598>.
19. Lal H, Patralekh MK. 3D printing and its applications in orthopaedic trauma: a technological marvel. *J Clin Orthop Trauma* 2018;9:260-8. <https://doi.org/10.1016/j.jcot.2018.07.022>.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
21. Mahadeva D, Mackay DC, Turner SM, Drew S, Costa ML. Reliability of the Neer classification system in proximal humeral fractures: a systematic review of the literature. *Eur J Orthop Surg Traumatol* 2008;18:415-24. <https://doi.org/10.1007/s00590-008-0325-6>.
22. Martin JS, Marsh JL. Current classification of fractures. Rationale and utility. *Radiol Clin North Am* 1997;35:491-506.
23. Matsushige T, Franco VP, Pierami R, Tamaoki MJS, Netto NA, Matsumoto MH. Do computed tomography and its 3D reconstruction increase the reproducibility of classifications of fractures of the proximal extremity of the humerus? *Rev* 2014;49:174-80. <https://doi.org/10.1016/j.rboe.2014.03.016>.
24. Misselyn D, Nijs S, Fieuws S, Shaheen E, Schepers T. Improved Interobserver Reliability of the Sanders Classification in Calcaneal Fractures Using Segmented Three-Dimensional Prints. *J Foot Ankle Surg* 2018;57:440-4. <https://doi.org/10.1053/j.jfas.2017.10.014>.
25. Mora Guix JM, Gonzalez AS, Bruggalla JV, Carril EC, Banos FG. Proposed protocol for reading images of humeral head fractures. *Clin Orthop* 2006;448:225-33. <https://doi.org/10.1097/01.blo.0000205899.28856.98>.
26. Neer CS. Displaced proximal humeral fractures: part I. classification and evaluation. *J Bone Joint Surg Am* 1970;52:1077-89.
27. Neer CS. Four-segment classification of proximal humeral fractures: purpose and reliable use. *J Shoulder Elbow Surg* 2002;11:389-400. <https://doi.org/10.1067/mse.2002.124346>.
28. Papakonstantinou 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: Interobserver agreement of Neer and AO classifications. *ANZ J Surg* 2016;86:280-4. <https://doi.org/10.1111/ans.13451>.
29. Ramappa AJ, Patel V, Goswami K, Zurakowski D, Yablon C, Rodriguez EK, et al. Using computed tomography to assess proximal humerus fractures. *Am J Orthop (Belle Mead NJ)* 2014;43:E43-7.
30. Roux A, Decroocq L, El Batti S, Bonnevalle N, Moineau G, Trojani C, et al. Epidemiology of proximal humerus fractures managed in a trauma center. *Orthop Traumatol Surg Res* 2012;98:715-9. <https://doi.org/10.1016/j.otsr.2012.05.013>.

31. Shrader MW, Sanchez-Sotelo J, Sperling JW, Rowland CM, Cofield RH. Understanding proximal humerus fractures: image analysis, classification, and treatment. *J Shoulder Elbow Surg* 2005;14:497-505. <https://doi.org/10.1016/j.jse.2005.02.014>.
32. Sim J, Wright C. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257-68. <https://doi.org/10.1093/ptj/85.3.257>.
33. Sukthankar AV, Leonello DT, Hertel RW, Ding GS, Sandow MJ. A comprehensive classification of proximal humeral fractures: HGLS system. *J Shoulder Elbow Surg* 2013;22:e1-6. <https://doi.org/10.1016/j.jse.2012.09.018>.
34. Sumrein BO, Mattila VM, Lepola V, Laitinen MK, Launonen AP, Group N. Intraobserver and interobserver reliability of recategorized Neer classification in differentiating 2-part surgical neck fractures from multi-fragmented proximal humeral fractures in 116 patients. *J Shoulder Elbow Surg* 2018;27:1756-61. <https://doi.org/10.1016/j.jse.2018.03.024>.
35. Yang L, Grottkau B, He Z, Ye C. Three dimensional printing technology and materials for treatment of elbow fractures. *Int Orthop* 2017;41:2381-7. <https://doi.org/10.1007/s00264-017-3627-7>.
36. Yang L, Shang X-W, Fan J-N, He Z-X, Wang J-J, Liu M, et al. Application of 3D printing in the surgical planning of trimalleolar fracture and doctor-patient communication. *BioMed Res Int* 2016;2016:2482086. <https://doi.org/10.1155/2016/2482086>.
37. 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 (Neer 3-part and 4-part) in old people. *Orthop Traumatol Surg Res* 2016;102(7):897-903. <https://doi.org/10.1016/j.otsr.2016.06.009>.