Navigating new waters in proximal humerus fractures

How to optimize fracture assessment, clinical decision-making, and pre-operative planning?

Reinier W.A. Spek

NAVIGATING NEW WATERS IN PROXIMAL HUMERUS FRACTURES

How to optimize fracture assessment,
clinical decision-making, and pre-operative planning?

Reinier W.A. Spek

Navigating new waters in proximal humerus fractures
Cotutelle PhD Flinders University and University of Groningen

This work was performed at:
Flinders Medical Centre, Adelaide
University Medical Centre Groningen, Groningen
OLVG, Amsterdam
Financial support for printing was kindly received from:
Flinders University, the University of Groningen, ChipSoft, and the Nederlandse Orthopaedische Vereniging

ISBN: 978-94-6473-942-8

Printed by Ipskamp Printing | proefschriften.net
Cover design and illustrations: Siebren Posthuma
Photography: Puck Veen
Layout and design: Indah Hijmans, persoonlijkproefschrift.nl

Navigating new waters in proximal humerus fractures

How to optimize fracture assessment, clinical decision-making, and pre-operative planning?

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. J.M.A. Scherpen
and in accordance with
the decision by the College of Deans

and

to obtain the degree of PhD at
Flinders University
on the authority of the
Dean of Graduate Research Prof. T. Cavagnaro
and in accordance with
the requirements of Higher Degrees by Research

This thesis will be defended in public on
Monday 17 November 2025 at 9:00 hours

by

Reinier Willem Alfred Spek

born on 29 April 1994

Supervisors
Prof. P.C. Jutte
Prof. R.L. Jaarsma
Prof. J.N. Doornberg
Prof. M.P.J. van den Bekerom

Assessment committee
Prof. P.M.A. van Ooijen
Prof. D. Eygendaal
Dr. T.C. Kwee
Dr. P.J. Smitham

Flinders University

Navigating new waters in proximal humerus fractures

How to optimize fracture assessment, clinical decision-making, and pre-operative planning?

By

Reinier W.A. Spek
MD, MSc

Thesis
Submitted to Flinders University
for the degree of

Doctor of philosophy
College of medicine and public health

Date of approval: 2 May 2025

I certify that this thesis: does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and the research within will not be submitted for any other future degree or diploma without the permission of Flinders University; and to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Signed by Reinier Spek, 15 May 2024

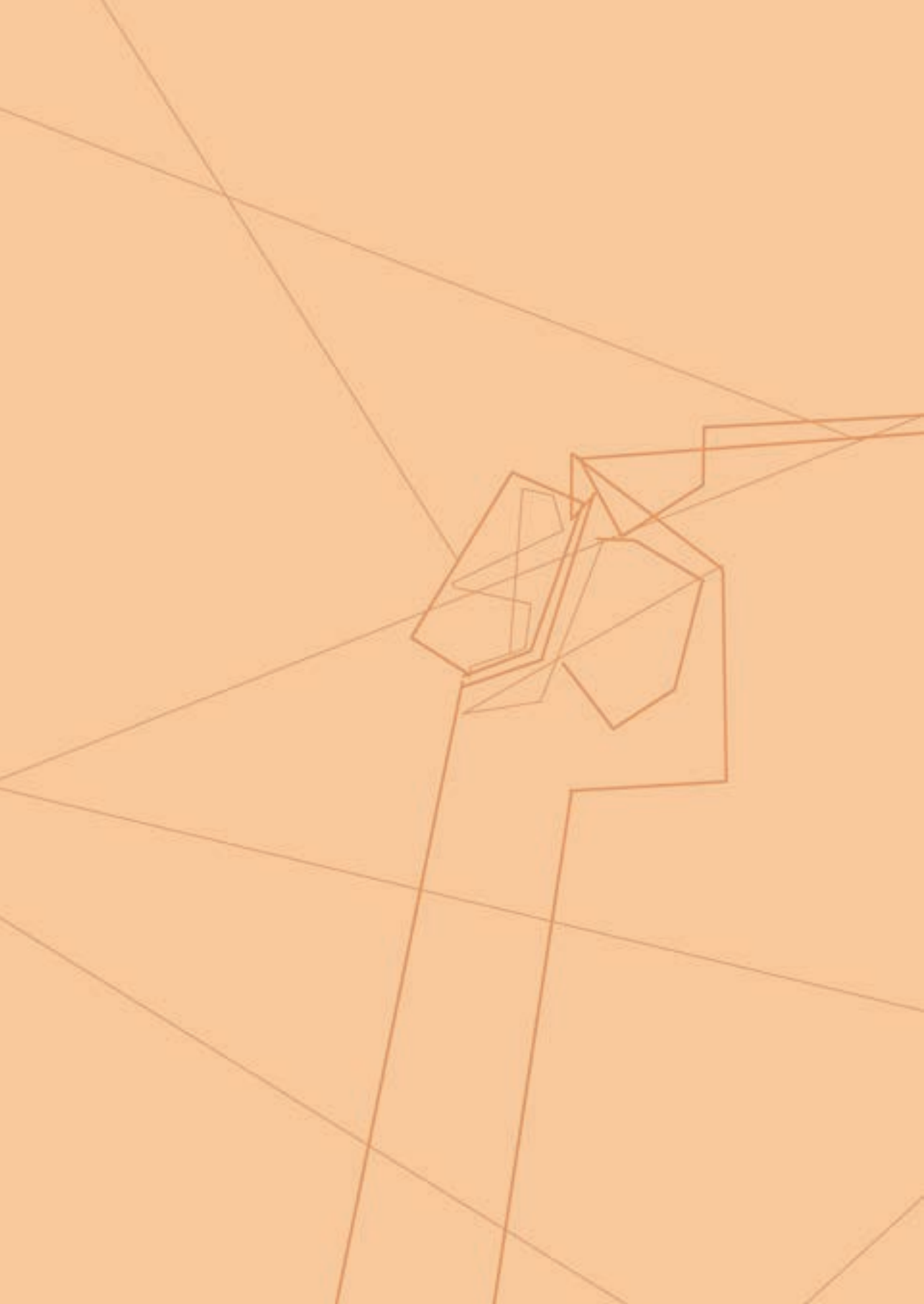
“Nooit meer slapen” - Willem Frederik Hermans

Aan mijn ouders
Aan mijn broertje en zus
Aan mijn naaste vrienden

TABLE OF CONTENTS

GENERAL DISCUSSION AND SUMMARY

12 SUPPLEMENTS



1

Health impact, background, clinical case, and research objectives

PATIENT IMPACT AND HEALTH BURDEN

Proximal humerus fractures are the fourth most common osteoporotic fracture, following vertebral, distal radius, and hip fractures. These fractures typically result from a simple, low-energy fall, with females being four times more likely to be affected than males. The lifetime risk of sustaining a proximal humerus fracture is estimated at approximately 13% for females over 50 years of age, with the highest age-specific incidence observed between 80 and 89 years [1] .

The impact of these fractures on geriatric patients is often underestimated. The one-year mortality rate is 15%, which is three times higher than in patients without such a fracture [2]. It also results in considerable morbidity, including impaired mobility, loss of functional independence, and psychological distress [3]. Furthermore, patients are at an increased risk of sustaining a subsequent fracture (within one year, nearly one in six patients experiences a hip fracture) [4].

The resulting loss of mobility often leads to greater dependence on others, which may necessitate long-term care, rehabilitation, or placement in a nursing facility. When combined with the costs of surgery, hospital admissions, and the aging population, these factors exacerbate the pressure on healthcare systems and contribute to rising healthcare expenditures [5,6].

BACKGROUND AND CLINICAL CASE

Case 1. A 26-year-old male presents at the emergency department following a heavy lateral contact injury to the right shoulder while playing Australian football. He is in severe pain and is unable to lift his arm (Fig. 1). After careful assessment, radiographic assessment shows a proximal humerus fracture.

Case 2. A 72-year-old female presents at the emergency department with acute shoulder pain following a low-energy trauma while walking the dog with her spouse. Despite her age, she has otherwise no relevant co-morbidities, does not require home care, and goes for a long walk or swim almost every day (Fig. 1). The subsequently performed radiograph reveals multiple fracture lines through the proximal humerus.

Figure 1. A young Australian football player *versus* an active older aged woman

Leave it, fix it, or replace it? Clinicians are repeatedly confronted with this question when encountering a proximal humerus fracture [7–10]. Although the question is simple, the answer is not. Proximal humerus fractures are complex injuries, and many variables should be considered for adequate surgical decision-making. Largely, this can be broken down into three mainstays: patient variables, surgical skills, and fracture pattern.

1. Patient variables. Managing patients' expectations and understanding patient's needs may be one of the most vital aspects for definitive decision-making [11–13]. Different paths in the treatment algorithm can be taken, depending on several factors such as life expectancy, co-existing comorbidities, bone quality, psychological health, and functional demand. For instance, suppose that the patients described in case 1 and 2, would have sustained the same valgus impacted proximal humerus fracture with >1cm displacement of both the tuberosities (Fig. 2), how would you treat them? One could argue that the young football player has a high functional demand and should be considered as a good surgical candidate for open reduction and internal fixation. Low-demanding older adults are better off with non-operative treatment while a reverse shoulder arthroplasty would be an adequate treatment option for the active old woman [14–17].

Figure 2. Valgus impacted proximal humerus fracture with displacement of the greater and lesser tuberosity.

2. Surgical skills. It is hard for surgeons to maintain their operative skills as the volumes of fractures considered for surgical intervention is low (~87% of proximal humerus fractures can be treated non-operatively) [18–20]. Moreover, there is a myriad of promising new surgical implants available, but their outcomes have not yet been studied in large clinical trials to prove superiority. A few examples are the interlocking reversed total shoulder arthroplasties, intramedullary nails (technically extremely demanding) and locking plates with intramedullary caging [21–23] .

3. Fracture patterns. A high risk for vascular compromise may alter surgical decision-making but it remains challenging to predict this adequately [24,25]. Moreover, proximal humerus fractures may present with varying levels of displacement and can follow a wide spectrum of fracture patterns (e.g., presence of metaphyseal extension, intra-articular involvement or medial calcar comminution) [26]. The ongoing challenge is to understand this wide spectrum of different fractures and to identify the prognostically poor fracture hallmarks

amongst them (as they may require surgical repair). But *how* can we do that? *How* can we do that if the interobserver reliability of fracture classifications and characteristics is inadequate? In other words, *how* can fracture characteristics drive surgical decision-making when such biases exist? The overall aim of this thesis is to quantify this problem and to provide solutions with emerging innovations in orthopaedic trauma.

RESEARCH OBJECTIVES

In the *first* part, fracture assessment. the interobserver agreement of fractures patterns is quantified: firstly, we assess a new –and simple– classification system on radiographs in a big panel of clinicians ranging from young residents to orthopaedic surgeons (are radiographs sufficiently informative to accurately classify angulated and separated surgical neck fractures?) (chapter 2). Secondly, we incorporate two classification systems and several fracture characteristics into our variables of interest and ask observers to assess them with and without 3D-printed fracture models (does the interobserver reliability improve if you have such a hand-held model?) (chapter 3). Finally, the ongoing debate on the low interobserver agreement is tackled from a new point of view: should clinicians outsource fracture assessment tasks to machine learning algorithms? Two convolutional neural networks are evaluated: one on fracture detection and classification (chapter 4), and one on assessment of ≥1cm greater tuberosity displacement, neck-shaft angle ≤100˚, shaft translation, and articular fracture involvement on plain radiographs (chapter 5).

The *second* part, patient counselling and decision-making, addresses the results of lesser tuberosity fractures in current clinical practice to inform patient on expected outcomes after this uncommon fracture (chapter 6). Next, we focus on isolated surgical neck fractures to evaluate if hanging down the arm in collar and cuff improves fracture re-alignment in the first weeks after the injury or whether this should be considered a true dogma (chapter 7).

In the *third* part, pre-operative planning, we outline the pros and cons of virtual 3D planning software in the work-up of proximal humerus locking plate fixation: does it lead to alterations in fracture reduction, plate position, calcar screw positioning, and screw lengths (chapter 8)?

REFERENCES

1. Court-Brown CM, Garg A, McQueen MM. The epidemiology of proximal humeral fractures. *Acta Orthop Scand*. 2001;72(4):365-371.
2. Walter N, Szymski D, Kurtz SM, et al. Proximal humerus fractures - epidemiology, comparison of mortality rates after surgical versus non-surgical treatment, and analysis of risk factors based on Medicare registry data. *Bone Joint Res*. 2023;12(2):103-112.
3. Iking J, Fischhuber K, Stolberg-Stolberg J, Raschke MJ, Katthagen JC, Köppe J. Quality of Life and Pain after Proximal Humeral Fractures in the Elderly: A Systematic Review. *Medicina (Kaunas)*. 2023;59(10).
4. Clinton J, Franta A, Polissar NL, et al. Proximal humeral fracture as a risk factor for subsequent hip fractures. *J Bone Joint Surg Am*. 2009;91(3):503-511.
5. Thorsness R, Shields E, Iannuzzi JC, Zhang L, Noyes K, Voloshin I. Cost Drivers After Surgical Management of Proximal Humerus Fractures in Medicare Patients. *J Orthop Trauma*. 2016;30(5):262-268.
6. Fjalestad T, Hole M, Jørgensen JJ, Strømsøe K, Kristiansen IS. Health and cost consequences of surgical versus conservative treatment for a comminuted proximal humeral fracture in elderly patients. *Injury*. 2010;41(6):599-605.
7. Rangan A, Handoll H, Brealey S, et al. Surgical vs nonsurgical treatment of adults with displaced fractures of the proximal humerus: the PROFHER randomized clinical trial. *JAMA*. 2015;313(10):1037-1047.
8. Fraser AN, Bjørdal J, Wagle TM, et al. Reverse Shoulder Arthroplasty Is Superior to Plate Fixation at 2 Years for Displaced Proximal Humeral Fractures in the Elderly: A Multicenter Randomized Controlled Trial. *J Bone Joint Surg Am*. 2020;102(6):477-485.
9. Beks RB, Ochen Y, Frima H, et al. Operative versus nonoperative treatment of proximal humeral fractures: a systematic review, meta-analysis, and comparison of observational studies and randomized controlled trials. *J Shoulder Elbow Surg*. 2018;27(8):1526-1534.
10. Orman S, Mohamadi A, Serino J, et al. Comparison of surgical and non-surgical treatments for 3- and 4-part proximal humerus fractures: A network meta-analysis. *Shoulder Elbow*. 2020;12(2):99-108.
11. Mancuso CA, Altchek DW, Craig E V., et al. Patients' expectations of shoulder surgery. *J Shoulder Elbow Surg*. 2002;11(6):541-549.
12. Iles RA, Davidson M, Taylor NF, O'Halloran P. Systematic review of the ability of recovery expectations to predict outcomes in non-chronic non-specific low back pain. *J Occup Rehabil*. 2009;19(1):25-40.
13. Waljee J, McGlinn EP, Sears ED, Chung KC. Patient expectations and patient-reported outcomes in surgery: a systematic review. *Surgery*. 2014;155(5):799-808.
14. Spross C, Meester J, Mazzucchelli RA, Puskás GJ, Zdravkovic V, Jost B. Evidence-based algorithm to treat patients with proximal humerus fractures—a prospective study with early clinical and overall performance results. *J Shoulder Elbow Surg*. 2019;28(6):1022-1032.
15. Belayneh R, Haglin J, Lott A, Kugelman D, Konda S, Egol KA. Underlying Mental Illness and Psychosocial Factors Are Predictors of Poor Outcomes After Proximal Humerus Repair. *J Orthop Trauma*. 2019;33(9):E339-E344.
16. Neuhaus V, Swellengrebel CHJ, Bossen JKJ, Ring D. What are the factors influencing outcome among patients admitted to a hospital with a proximal humeral fracture? General. *Clin Orthop Relat Res*. 2013;471(5):1698-1706.

17. Seebeck J, Goldhahn J, Städele H, Messmer P, Morlock MM, Schneider E. Effect of cortical thickness and cancellous bone density on the holding strength of internal fixator screws. *J Orthop Res*. 2004;22(6):1237-1242.
18. Hammond JW, Queale WS, Kim TK, McFarland EG. Surgeon experience and clinical and economic outcomes for shoulder arthroplasty. *J Bone Joint Surg Am*. 2003;85(12):2318-2324.
19. Jain NB, Kuye I, Higgins LD, Warner JJP. Surgeon volume is associated with cost and variation in surgical treatment of proximal humeral fractures shoulder. *Clin Orthop Relat Res*. 2013;471(2):655-664.
20. Brorson S, Viberg B, Gundtoft P, Jalal B, Ohrt-Nissen S. Epidemiology and trends in management of acute proximal humeral fractures in adults: an observational study of 137,436 cases from the Danish National Patient Register, 1996-2018. *Acta Orthop*. 2022;93:750-755.
21. Nourissat G, Corsia S, Srikumaran U, et al. Use of a locking stem for reverse shoulder arthroplasty is a rare but reliable option. *Int Orthop*. 2022;46(9):2097-2104.
22. Hudgens JL, Jang J, Aziz K, Best MJ, Srikumaran U. Three- and 4-part proximal humeral fracture fixation with an intramedullary cage: 1-year clinical and radiographic outcomes. *J shoulder Elb Surg*. 2019;28(6S):S131-S137.
23. Boileau P, d'Ollonne T, Bessière C, et al. Displaced humeral surgical neck fractures: classification and results of third-generation percutaneous intramedullary nailing. *J Shoulder Elbow Surg*. 2019;28(2):276-287.
24. Campochiaro G, Rebuzzi M, Baudi P, Catani F. Complex proximal humerus fractures: Hertel's criteria reliability to predict head necrosis. *Musculoskelet Surg*. 2015;99 Suppl 1:9-15.
25. Hertel R, Hempfing A, Stiehler M, Leunig M. Predictors of humeral head ischemia after intracapsular fracture of the proximal humerus. *J Shoulder Elbow Surg*. 2004;13(4):427-433.
26. Hasan AP, Phadnis J, Jaarsma RL, Bain GI. Fracture line morphology of complex proximal humeral fractures. *J Shoulder Elbow Surg*. 2017;26(10):e300-e308.


AMSTERDAM
ADELAIDE

2

What is the interobserver agreement of displaced humeral surgical neck fracture patterns?

Reinier W.A. Spek
Laura J. Kim

The Traumaplatform 3D Consortium
Clinics in Shoulder and Elbow. 2022 Dec;25(4):304-310

ABSTRACT

Aims

The Boileau classification distinguishes three surgical neck fracture patterns: type A, B and C. However, the reproducibility of this classification on plain radiographs is unclear. Therefore, we questioned: what is the interobserver agreement and accuracy of displaced surgical neck fracture patterns categorized according to the modified Boileau classification? And does the reliability to recognize these fracture patterns differ between orthopaedic residents and attending surgeons?

Methods

This interobserver study consisted of a randomly retrieved series of 30 plain radiographs representing clinical practice in a Level 1 and a Level 2 trauma centre. Radiographs were included from patients (≥18 years) who sustained an isolated displaced surgical neck fracture taken ≤1 week after initial injury. A ground truth was established by consensus among three senior orthopaedic surgeons. All images were assessed by 17 orthopaedic residents and 17 attending orthopaedic trauma surgeons.

Results

Agreement for the modified Boileau classification was fair (κ = 0.37; 95% confidence interval [CI], 0.36 - 0.38) with an accuracy of 62% (95% CI, 57% - 66%). Comparison of interobserver variability between residents and attending surgeons revealed a significant but clinically irrelevant difference in favour of attending surgeons (0.34 *versus* 0.39, respectively, Δ κ = 0.05, 95% CI, 0.02 - 0.07).

Conclusion

The modified Boileau classification yields a low interobserver agreement with an unsatisfactory accuracy in a panel of orthopaedic residents and attending surgeons. This supports the hypothesis that surgical neck fractures are challenging to categorize, and that this classification should not be used to determine prognosis if only plain radiographs are available.

INTRODUCTION

Two-part surgical neck fractures of the humerus entail 28% of proximal humerus fractures and can be treated non-operatively or by several surgical modalities (e.g., plate fixation and intramedullary nailing) [1–3]. However, substantial treatment variability is observed between clinicians, hospitals, and even among countries [4]. Among other things, classification of the fracture is important for determining the optimal treatment [5]. Ideally, classification should guide the surgeons' decision-making and be taken into account to determine the optimal treatment for proximal humerus fractures.

Currently available classification systems for surgical neck fractures are the fracture patterns according to Neer [6] and Arbeitsgemeinschaft für Osteosynthesefragen (AO) [7]. Neer created three subgroups (impacted angulated, separated, and comminuted two-part surgical neck fractures), while the AO created two subgroups (impacted and non-impacted two-part surgical neck fractures). Nevertheless, clinical implications of these distinct fracture patterns are unclear.

To determine the optimal entry point for intramedullary nailing, Boileau et al. [8] developed a new classification system which categorized displaced surgical neck fractures into three types: type A, partial medial shaft translation with valgus humeral head angulation; type B, entire medial shaft translation without humeral head tilt or angulation; and type C, lateral shaft translation with varus humeral head angulation. Although numerous studies have investigated the agreement on the full array of two-, three-, and four-part proximal humerus fractures, no interobserver study has been carried out regarding surgical neck fracture patterns in particular [9,10]. A reproducible fracture classification is a prerequisite to comparing patient outcomes of different clinical trials [5]. Moreover, if a high level of agreement can be reached, fracture patterns could potentially influence surgical decision-making and might predict prognosis.

The Boileau classification was originally based on radiographs and computed tomography (CT) scans, but as CT scans are not routinely available for every patient, this study aimed to assess its reproducibility on plain radiographs. The following research questions were asked: what is the interobserver agreement and accuracy of displaced surgical neck fracture patterns categorized according to the modified

Boileau criteria? And does the reliability to recognize these fracture patterns differ between orthopaedic residents and attending surgeons?

PATIENTS AND METHODS

Ethical approval was received from OLVG (Amsterdam, The Netherlands, No. 19.135) and Flinders Medical Centre (Adelaide, Australia, No. 234.19). Informed consent from patients was waived.

Setting and study design

This is an interobserver study in which 30 radiographs were assessed and categorized according to the modified Boileau's classification of displaced surgical neck fractures [8]. The study was carried out in March and April 2021, and an observer panel was created with participants from the orthopaedic and trauma units of four different teaching hospitals. The panel consisted of 17 orthopaedic residents and 17 attending orthopaedic trauma surgeons with different levels of experience and subspecialties.

Images

Anteroposterior (true or standard) and lateral radiographic views were included from patients (≥18 years) who sustained an isolated displaced surgical neck fracture which could be classified according to the Boileau classification. Patients were deemed eligible irrespective of the treatment provided; thus, trauma radiographs of both non-operatively treated patients and surgically treated patients were included. Patients were excluded if they presented to the emergency department more than 1 week after the initial injury or had a concomitant fracture (Hill-Sachs lesion, proximal humerus, humeral shaft, or pathologic fracture).

Classification

Boileau et al. [8] developed this classification system to categorize displaced surgical neck fractures into three types: type A, partial medial shaft displacement with valgus humeral head angulation; type B, entire medial shaft translation without humeral head tilt; and type C, lateral shaft displacement with varus humeral head angulation. A fracture was considered displaced if it was translated >25% of the humeral midshaft width. Displacement was measured from the outer cortex of the

most proximal part of the humeral shaft fragment to the outer cortex of the most distal humeral head fragment. To cover all displaced surgical neck fractures, an additional category was incorporated in this study: "non-classifiable". This meant that the head angulation and humeral shaft translation did not match Boileau's criteria (e.g., partial lateral humeral shaft translation without head angulation). Therefore, four categories could be chosen by the observers: type A, type B, type C, or non-classifiable (Fig. 1).

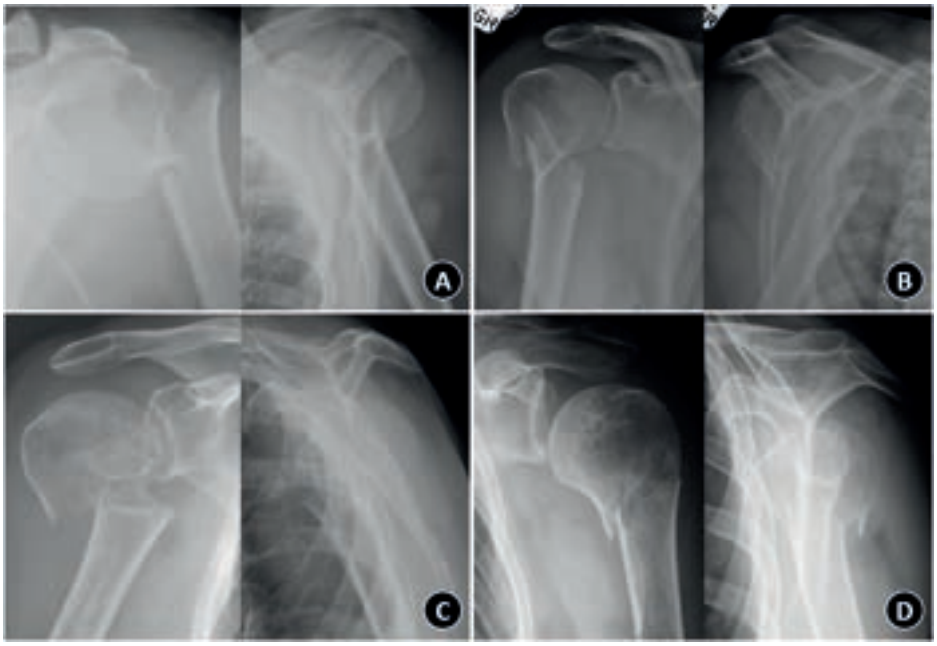


Figure. 1. The modified Boileau classification covers four options: type A, type B, type C and non-classifiable displaced surgical neck fractures. (A) Type C: lateral shaft displacement with varus angulation of the head. (B) Type B: entire medial (or ventral) shaft translation without humeral head tilt. (C) Type A: medial shaft translation with valgus humeral head tilt. (D) Non-classifiable: shaft translation and/or head angulation do not match with Boileau's classification. In this example, there is no varus angulation of the head meaning it could not be classified according to Boileau. Type A and C were used for training; Type B and the non-classifiable radiograph were used for the actual assessments

Selection of radiographs

Radiographs of eligible patients were collected from a Level 1 trauma centre in Australia (March 1, 2016, to July 31, 2020) and a Level 2 trauma centre in the Netherlands (January 1, 2004, to June 30, 2018). A total of 614 surgical neck

fractures were identified of which 236 patients had a displaced fracture. Among these displaced fractures, 121 patients could be classified according to Boileau's classification (type A, n = 41; type B, n = 20; type C, n = 60). While maintaining this mutual distribution between the three Boileau types, we randomly selected 9 type A fractures, 5 type B, 11 type C, and 5 non-classifiable fractures. The number selected for the non-classifiable category was equal to that of the group with the lowest number (i.e., type B fractures). Randomization was carried out in Microsoft Excel version 2102 (Microsoft Corp., Redmond, WA, USA) by assigning a randomization number which was sorted from low to high. Cases with the lowest randomization number were selected until the predefined sample size (n = 30) was reached. The mean age (range) of included patients was 72.4 years (29 - 96 years), and the majority were females (80%).

Ground truth

A ground truth was generated by consensus among three senior orthopaedic attending surgeons (two with >20 years of experience and one with >15 years of experience after finishing their training). Each of these orthopaedic surgeons completed the study prior to the consensus meeting, so they classified all fractures independently before answers were compared. The meeting was led by the first author (R.S.), and discrepancies were resolved by discussion.

Observer panel

The observer panel consisted of 34 participants: 17 orthopaedic residents and 17 attending orthopaedic surgeons. Six attending orthopaedic surgeons had <5 years of experience. All other attending surgeons had >5 years of experience: five were seniors (>20 years of experience), three were shoulder specialists (they completed fellowship training on the upper extremity), two were dedicated attending trauma surgeons, and one was an orthopaedic oncologist. All attending surgeons had substantial experience in treating trauma, and years of experience was defined as years in clinical practice after finishing the training program.

Training and assessment

Prior to assessment, each observer received training in recognizing the fracture patterns according to Boileau's classification. The first part of the training consisted of an explanation of the fracture patterns and the following rules: (1) dorsal head angulation is not considered (e.g., medial translation with valgus head angulation

and dorsal head angulation should be classified as a type A fracture) and (2) type B fractures require entire medial or entire ventral humeral shaft translation. It was also emphasized that both head angulation and shaft displacement had to match Boileau's criteria (e.g., medial humeral shaft translation with varus angulation should be categorized as non-classifiable). Following this, four training cases were provided (one case covering each category) (Fig. 2). At the discretion of observers, training was provided either face-to-face (by R.S. or L.K.) or as self-study via REDCap [11,12]. Face-to-face training was provided to 73.5% of observers, and 26.5% followed the self-study on REDCap. Each observer classified 30 displaced surgical neck fractures with both anteroposterior and lateral views. Questions and radiographs were both presented on-screen. Illustration sheets depicting the classification system were displayed during the observation. There was no time limit on assessment, and radiographs were presented in identical order for each observer. Observers could not use radiographic measurement tools. However, they could go back if needed and adjust their answer for each radiograph.

Statistical analysis

IBM SPSS software version 27 (IBM Corp., Armonk, N.Y., USA) was used for statistical analysis. To determine interobserver variability, the multi-rater Fleiss' kappa (κ) was calculated. Values were interpreted according to Landis and Koch: $\kappa < 0.00$ (poor), $\kappa = 0.00 - 0.20$ (slight), $\kappa = 0.21 - 0.40$ (fair), $\kappa = 0.41 - 0.60$ (moderate), $\kappa = 0.61 - 0.80$ (substantial), and $\kappa = 0.81 - 1.00$ (almost perfect) [13]. Accuracy was defined as the degree to which each given answer corresponded with the ground truth and expressed as a percentage from 0 to 100. If the accuracy was 0%, no cases were classified the same as the ground truth. If the accuracy was 100%, all cases were classified the same as the ground truth. To calculate accuracy, the accuracy per observer was determined and subsequently averaged across all participants. To compare residents *versus* attending surgeons, delta (Δ) κ was computed and depicted with a two-tailed *p*-value. Accuracy among residents and attending surgeons was compared with an independent samples t-test. Multi-rater Fleiss' κ as well as accuracy was displayed with a 95% confidence interval (CI).

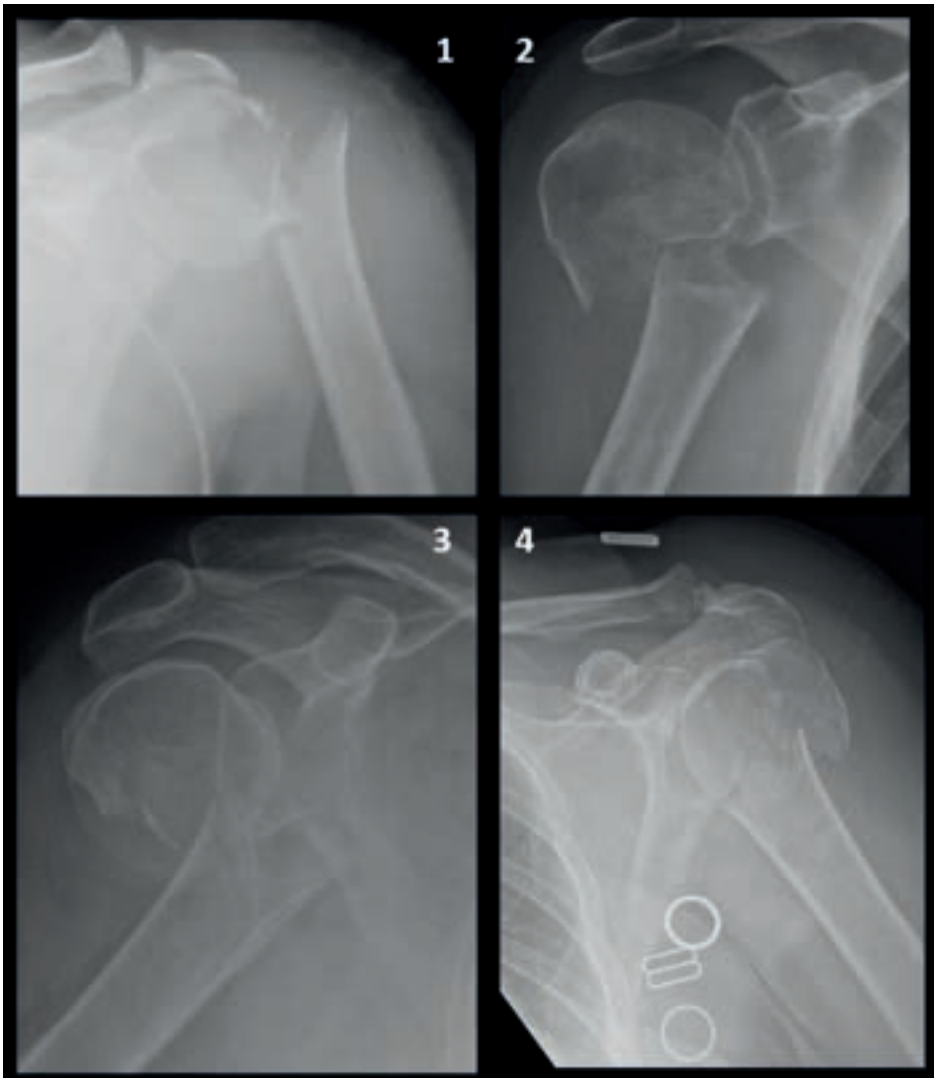


Figure 2. Radiographs used for training, shown in the order from 1 to 4, with 1 = type C, 2 = type A, 3 = type B, and 4 = non-classifiable. Although present on image 3 and 4, fracture dislocations and concomitant greater tuberosity fractures were not included in the actual assessment. This was explained to the observers accordingly.

RESULTS

Interobserver variability and accuracy

Interobserver agreement to classify fractures according to the modified Boileau criteria among all observers was fair (κ = 0.37; 95% CI, 0.36 - 0.38) (Fig. 3). In type A and C fractures, concordance was moderate (κ = 0.42; 95% CI, 0.40 - 0.44 and κ = 0.58; 95% CI, 0.57 - 0.59, respectively). Observers disagreed the most on type B (κ = 0.23; 95% CI, 0.21 - 0.25) and non-classifiable fractures (κ = 0.18; 95% CI, 0.16 - 0.20). Accuracy amongst all participants was 62% (95% CI, 57% - 66%) and the highest for type C fractures: 79% (95% CI, 74% - 85%) (Table 1).

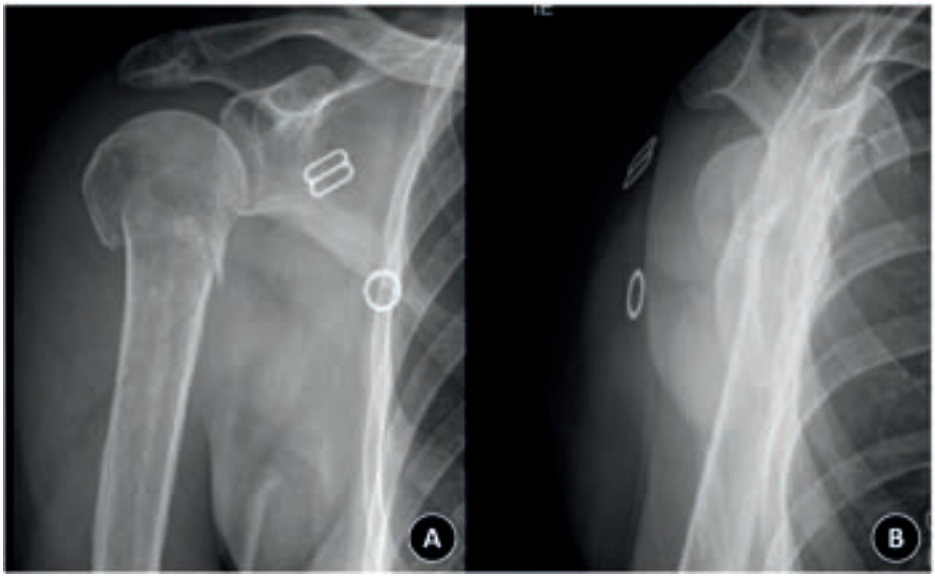


Figure 3. Assessment of a radiograph with substantial variability amongst the observers: 53% classified this as type A (18 observers), 3% as type B (1 observer), 3% as type C (1 observer), and 41% as "non-classifiable" (14 observers). (A) Standard anterior-posterior view. (B) Lateral view.

Table 1. Agreement and accuracy among all observers

Category	Kappa (95% CI)		Agreement	Accuracy (95% CI), %	
Overall	0.37	(0.36 - 0.38)	Fair	62	(57 - 66)
Type A	0.42	(0.40 - 0.44)	Moderate	64	(57 - 71)
Type B	0.23	(0.21 - 0.25)	Fair	69	(59 - 79)
Type C	0.58	(0.57 - 0.59)	Moderate	79	(74 - 85)
Non-classifiable	0.18	(0.16 - 0.20)	Slight	57	(49 - 65)

Abbreviations: CI, confidence interval.

Residents versus attending surgeons

Comparison of interobserver variability between residents and attending surgeons revealed a significant but intuitively clinically irrelevant difference in favour of attending surgeons (fair *versus* fair, Δ κ = 0.05; 95% CI, 0.02 - 0.07). Residents showed an accuracy of 60% (95% CI, 55% - 65%) in correctly classifying the fractures, whereas attending surgeons revealed an accuracy of 63% (95% CI, 55% - 72%). No statistically significant difference was found between both groups (Δ κ = 0.03; 95% CI, -0.06 to 0.12) (Table 2).

Table 2. Agreement and accuracy compared between 17 residents and 17 attending surgeons

Parameter	Kappa (95% CI)		Agreement	Accuracy (95% CI), %	
Residents	0.34	(0.33 - 0.36)	Fair	60	(55 - 65)
Surgeons	0.39	(0.37 - 0.41)	Fair	63	(55 - 72)
Delta	0.05	(0.02 - 0.07)		3	(-6 to 12)
p-value	<0.001			0.47	

Abbreviations: CI, confidence interval.

DISCUSSION

Boileau's classification is a recently introduced classification to enhance the humeral nail entry point in treatment for displaced surgical neck fractures. Its inter-surgeon reliability on plain radiographs is unclear, hence our aim was to assess the interobserver variability and accuracy. This study revealed an overall kappa of 0.37 with 62% accuracy for the modified Boileau classification on radiographs. The interobserver variability is a measure that represents the extent of variation between observers for the same radiographs expressed as the kappa coefficient and should be considered together with accuracy. A kappa value of 0.37 is relatively low and implies strong variability in the classification which can lead to misdiagnosis and a potential delay in best treatment. In other words, our study demonstrated that 62% of radiographs were classified correctly, but there was substantial disagreement in the misclassified radiographs.

The interobserver reliability of the general AO and full Neer classification systems has been studied intensively. However, many of these studies had a limited number

of observers, which could result in overestimation of agreement, and the question remained unanswered as to the interobserver agreement was for the subgroups of surgical neck fractures (Neer included three subgroups, and AO included two subgroups) [14,15]. Regarding the AO classification, the largest study included 46 observers and found a kappa of 0.18 [10]. Another study included 18 observers and investigated the agreement on two-, three-, and four-part fractures according to Neer. They revealed a kappa ranging from 0.03 to 0.07 for classifying two-part fractures [9]. Additionally, kappa values do not improve when fractures are assessed with CT scans [8,9,14,16]. Our study therefore demonstrated a better kappa (0.38); however, this is still inadequate for clinical use. Furthermore, the low interobserver agreement of Boileau's classification has implications for surgical decision-making in clinical practice: it is unlikely that surgeons can solely rely on radiographs for surgical planning of humeral nailing.

Assessment of three- and four-part proximal humerus fractures is thought to be better among shoulder specialists compared to general orthopaedic surgeons [9]. Additionally, some studies advocate that attending surgeons outperform residents [16]. In this study, we did not find a clinically relevant difference between assessments by residents compared to attending surgeons. As opposed to three- and four-part fractures, this study therefore suggests that two-part displaced surgical neck fractures do not require a certain level of expertise, potentially due to their less complex nature or due to the matter that nobody had any experience with this classification.

It has yet to be established whether or not Boileau's classification has clinical implications aside from humeral nailing, and if it can determine prognosis. Nevertheless, one could argue that this classification may be useful for decision-making. For instance, in type B fractures, the entire shaft is translated which, in our experience, may require surgical intervention. Moreover, type C fractures are likely to respond well to non-operative treatment due to traction of the pectoralis major muscle while wearing a collar and cuff. Decision-making in type A fractures could depend on the degree of valgus angulation, as patients with ≥160° may be better off with surgical fixation [17].

This work reconfirms the challenges clinicians are facing to improve interobserver agreement for proximal humerus fracture patterns. As the era of artificial intelligence is approaching, it is speculated that we should make a transition to data-

driven care: potentially, an algorithm trained on fracture classification by the input of senior surgeons could neutralize current misconceptions and observation bias [17].

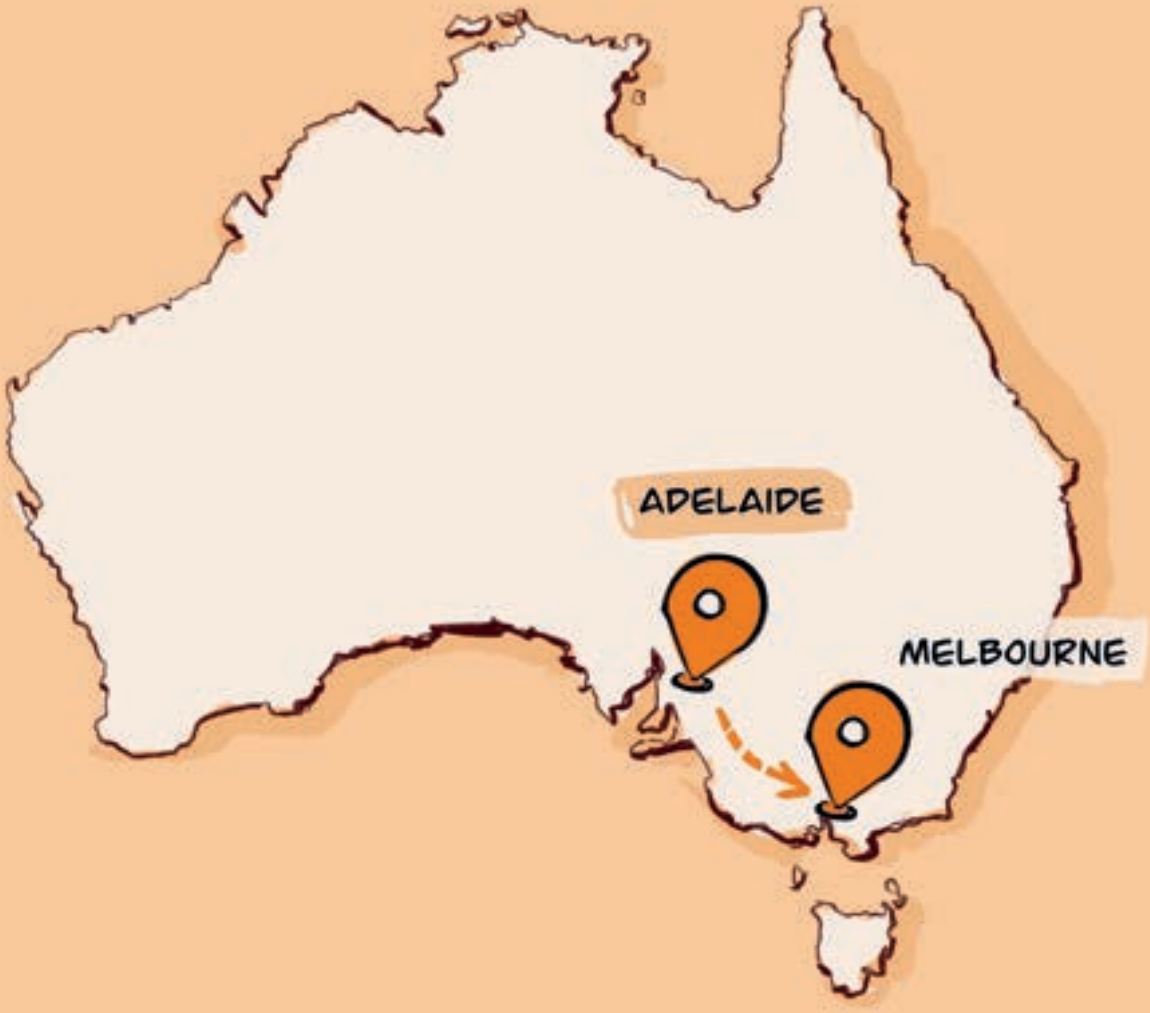
Several shortcomings should be considered: firstly, the quality of radiographs varied as not all radiographs were taken with similar radiographic imaging settings. In some patients, the true anteroposterior radiographic views were not obtained, which may have changed the perception of humeral shaft translation as well as head angulation. Additionally, internal humeral head rotation makes it difficult to assess head deformity as the greater tuberosity is not well profiled. However, our aim was to evaluate the classification on radiographs which would reflect the hospital setting well: in clinical practice, it is well known that radiographic quality can be low, and that patients retain their shoulders in internal rotation due to pain. As opposed to the original classification, CT scans were not used for this study. The rationale for assessing this classification was to assess whether it could be applied to all patients presenting at the emergency department, and as CTs are not routinely performed for these patients, this was not feasible. Hence, we coined it the modified Boileau classification: a fourth category (non-classifiable) was added to cover all displaced surgical neck fractures. One could argue that by mitigating these factors, interobserver variability could improve. Secondly, in clinical practice, radiographs are usually discussed between colleagues (e.g., between orthopaedic residents and attending surgeons). This is a limitation for interobserver studies in general so it would be interesting to assess its impact on agreement. For instance, during the consensus meeting there was hardly any significant dispute on radiographs even though the attending surgeons classified 12 radiographs differently during initial assessment. This underscores the suggestion that group discussion might improve agreement. Thirdly, the intra-observer agreement was not evaluated. One of the study strengths was the representativeness of the observer panel, which was a good reflection of potential users of this classification.

Conclusion

Displaced surgical neck fractures are hard to classify on plain radiographs: the modified Boileau classification yields a poor interobserver agreement with an accuracy of 62% in a panel of orthopaedic residents and attending surgeons with different levels of experience. This suggests that two-part displaced surgical neck fractures do not require a certain level of expertise, and that surgeons cannot rely solely on radiographs for surgical planning of humeral nailing.

REFERENCES

1. Yoon RS, Dziadosz D, Porter DA, Frank MA, Smith WR, Liporace FA. A comprehensive update on current fixation options for two-part proximal humerus fractures: a biomechanical investigation. *Injury*. 2014;45(3):510-514.
2. Setaro N, Rotini M, Luciani P, Facco G, Gigante A. Surgical management of 2- or 3-part proximal humeral fractures: comparison of plate, nail and K-wires. *Musculoskelet Surg*. 2022;106:163-167.
3. Court-Brown CM, Garg A, McQueen MM. The epidemiology of proximal humeral fractures. *Acta Orthop Scand*. 2001;72(4):365-371.
4. Launonen AP, Sumrein BO, Reito A, et al. Operative versus non-operative treatment for 2-part proximal humerus fracture: a multicenter randomized controlled trial. Handoll H, ed. *PLoS Med*. 2019;16(7):e1002855.
5. Handoll HH, Brorson S. Interventions for treating proximal humeral fractures in adults. *Cochrane database Syst Rev*. 2012;12:CD000434.
6. Neer CS 2nd. Displaced proximal humeral fractures. I. Classification and evaluation. *J Bone Joint Surg Am*. 1970;52(6):1077-1089.
7. Meinberg EG, Agel J, Roberts CS, Karam MD, Kellam JF. Fracture and Dislocation Classification Compendium-2018. *J Orthop Trauma*. 2018;32 Suppl 1:S1-S170.
8. Boileau P, d'Ollonne T, Bessière C, et al. Displaced humeral surgical neck fractures: classification and results of third-generation percutaneous intramedullary nailing. *J Shoulder Elbow Surg*. 2019;28(2):276-287.
9. 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.
10. Bruinsma WE, Guitton TG, Warner JJP, Ring D; Science of Variation Group. Interobserver reliability of classification and characterization of proximal humeral fractures: a comparison of two and three-dimensional CT. *J Bone Joint Surg Am*. 2013;95(17):1600-1604.
11. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208.
12. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap): a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
14. Iordens GIT, Mahabier KC, Buisman FE, 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(5):596-602.
15. Marongiu G, Leinardi L, Congia S, Frigau L, Mola F, Capone A. Reliability and reproducibility of the new AO/OTA 2018 classification system for proximal humeral fractures: a comparison of three different classification systems. *J Orthop Traumatol*. 2020;21:4.
16. 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(9):1371-1375.
17. Chung SW, Han SS, Lee JW, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop*. 2018;89(4):468-473.


ADELAIDE
MELBOURNE

3

3D-printed handheld models do not improve recognition of specific characteristics and patterns of three-part and four-part proximal humerus fractures

Reinier W.A. Spek
Bram J.A. Schoolmeesters
Jacobien H. F. Oosterhoff
Job N. Doornberg
Michel P.J. van den Bekerom
Ruurd L. Jaarsma
Denise Eygendaal
Frank F.A. IJpma

The Traumaplatform 3D Consortium
Clinical Orthopaedics and Related Research®. 2022 Jan;480(1):150-159

ABSTRACT

Aims

Reliably recognizing the overall pattern and specific characteristics of proximal humerus fractures may aid in surgical decision-making. With conventional onscreen imaging modalities, there is considerable and undesired interobserver variability, even when observers receive training in the application of the classification systems used. It is unclear whether three-dimensional (3D) models, which now can be fabricated with desktop printers at relatively little cost, can decrease interobserver variability in fracture classification. Do 3D-printed handheld models of proximal humerus fractures improve agreement among residents and attending surgeons regarding (1) specific fracture characteristics and (2) patterns according to the Neer and Hertel classification systems?

Methods

Plain radiographs, as well as two-dimensional (2D) and 3D CT images were collected from 20 patients (aged 18 years or older) who sustained a three-part or four-part proximal humerus fracture treated at a Level 1 trauma centre between 2015 and 2019. The included images were chosen to comprise images from patients whose fractures were considered as difficult-to-classify, displaced fractures. Consequently, the images were assessed for eight fracture characteristics and categorized according to the Neer and Hertel classifications by four orthopaedic residents and four attending orthopaedic surgeons during two separate sessions. In the first session, the assessment was performed with conventional onscreen imaging (radiographs and 2D and 3D CT images). In the second session, 3D-printed handheld models were used for assessment, while onscreen imaging was also available. Although proximal humerus classifications such as the Neer classification have, in the past, been shown to have low interobserver reliability, we theorized that by receiving direct tactile and visual feedback from 3D-printed handheld fracture models, clinicians would be able to recognize the complex 3D aspects of classification systems reliably. Interobserver agreement was determined with the multi-rater Fleiss kappa and scored according to the categorical rating by Landis and Koch. To determine whether there was a difference between the two sessions, we calculated the delta (difference in the) kappa value with 95% confidence intervals and a two-tailed *p*-value. Post hoc power analysis revealed that with the current sample size, a delta kappa value of 0.40 could be detected with 80% power at alpha = 0.05.

Results

Using 3D-printed models in addition to conventional imaging did not improve interobserver agreement of the following fracture characteristics: more than 2 mm medial hinge displacement, more than 8 mm metaphyseal extension, surgical neck fracture, anatomic neck fracture, displacement of the humeral head, more than 10 mm lesser tuberosity displacement, and more than 10 mm greater tuberosity displacement. Agreement regarding the presence of a humeral head-splitting fracture was improved but only to a level that was insufficient for clinical or scientific use (fair to substantial, delta kappa = 0.33 [95% CI 0.02 to 0.64]). Assessing 3D-printed handheld models in adjunct to onscreen conventional imaging did not improve the interobserver agreement for pattern recognition according to Neer (delta kappa = 0.02 [95% CI -0.11 to 0.07]) and Hertel (delta kappa = 0.01 [95% CI -0.11 to 0.08]). There were no differences between residents and attending surgeons in terms of whether 3D models helped them classify the fractures, but there were few differences to identify fracture characteristics. However, none of the identified differences improved to almost perfect agreement (kappa value above 0.80), so even those few differences are unlikely to be clinically useful.

Conclusion

Using 3D-printed handheld fracture models in addition to conventional onscreen imaging of three-part and four-part proximal humerus fractures does not improve agreement among residents and attending surgeons on specific fracture characteristics and patterns. Therefore, we do not recommend that clinicians expend the time and costs needed to create these models if the goal is to classify or describe patients' fracture characteristics or pattern, since doing so is unlikely to improve clinicians' abilities to select treatment or estimate prognosis.

INTRODUCTION

Recognizing the overall pattern and specific characteristics of proximal humerus fractures may aid in decision-making and determining prognosis. However, there is considerable and undesired interobserver variability, even when observers receive training in the application of the classification systems used [1]. Because the relationship between fracture lines and displacement can be difficult to assess on plain radiographs [2], two-dimensional (2D) and three-dimensional (3D) CT images are part of the routine diagnostic workup in many institutions. 2D and 3D CT images result in better inter-surgeon reliability than radiographs and are particularly valuable for assessing more severe fracture configurations (such as head-splitting fractures and three-part and four-part fractures) [3,4]. Despite the improvements seen with the use of 2D and 3D CT onscreen imaging, overall agreement on fracture patterns between attending surgeons remains low (slight to fair concordance) [4]. Another contentious issue is the value of 3D CT images for attending surgeons with different levels of experience; although one study concluded that residents benefit the most from using 3D CT images [5], other studies found improvement among specialists only [1,3].

Printing of 3D models for diagnostic assessment and surgical planning of fractures is now widely available using freely available software and relatively inexpensive desktop 3D printers, without the need to rely on commercial vendors [6]. In distal humerus fractures, 3D-printed models have been demonstrated to improve inter-surgeon agreement in determining fracture characteristics [7]. However, the clinical value of 3D printing for diagnostic workup of proximal humerus fractures, as well as the potential value in aiding residents to recognize patterns, has yet to be determined. To date, one study found that agreement improved regarding the choice of treatment (non-operative *versus* osteosynthesis *versus* arthroplasty) when proximal humerus fractures were assessed with 3D-printed models [8]. Nonetheless, two studies showed that 3D-printed models improved agreement for the Neer and AO classification systems among both residents and attending surgeons, but did not reveal a difference between both groups [9,10]. Although they conducted valuable work, they did not account for characterization and other fracture classification systems. Therefore, it remains unclear whether 3D-printed models can decrease interobserver variability in fracture assessment.

To fill this knowledge gap, we asked: Do 3D-printed handheld models of proximal humerus fractures improve agreement among residents and attending surgeons regarding (1) specific fracture characteristics and (2) patterns according to the Neer and Hertel classification systems?

PATIENTS AND METHODS

Setting and study design

This diagnostic study was performed between August 2019 and June 2020 in a Level 1 trauma centre in Australia and a Level 2 trauma centre in the Netherlands. During this period, four orthopaedic residents and four attending orthopaedic surgeons (DE, three Traumaplatform 3D Consortium members) assessed 20 proximal humerus fractures for eight specific fracture characteristics and the full Neer [11] and Hertel [12] fracture patterns during two separate observation sessions with a minimum interval of 1 month between reads. As all participants were involved in the treatment of hundreds of trauma patients monthly, it was assumed that a 1-month interval would be sufficient to minimize information bias. The Neer classification categorizes proximal humerus fractures into four groups (minimally displaced, two-part, three-part, and four-part fractures) while distinguishing four anatomic segments (the shaft, articular segment, lesser tuberosity, and greater tuberosity). The segments are considered as a separate part if they are displaced more than 1 centimetre or angulated more than 45°. If not, the fracture part is considered minimally displaced. The classification also accounts for the presence of dislocation and head-splitting fractures. Altogether, 16 different categories can be chosen [13]. The Hertel classification consists of 12 different fracture patterns, which are determined by identifying the fracture planes between the greater tuberosity, humeral head, lesser tuberosity, and the shaft. Unlike the Neer classification, this system does not consider displacement or angulation between any of the segments. This classification is illustrated by LEGO bricks, and can be found in the original study by Hertel et al. [12]. Despite the poor inter-surgeon agreement of the Neer classification (kappa = 0.07; 18 observers; used modality = 3D CT reconstruction images) [3] and the relatively low agreement of the Hertel classification (kappa = 0.44; four observers; used modality = rapid sequence prototype models) [14], this study incorporated both fracture patterns in the assessment.

Both classification systems have limited value for clinical decision-making, but they are still widely used to report outcomes of proximal humerus fractures in conjunction with specific fracture characteristics. For this reason, we wanted to establish how reliably these injuries could be assessed: If clinicians cannot agree on fracture characteristics and classification, it will be challenging to study results of proximal humerus fractures. The 3D-printed fracture models were designed to be held in the hand and freely rotated in space in every direction. We theorized that observers could move one step closer to reality by handling the models, allowing them to better determine angulation, displacement, and recognize the anatomic parts (such as the lesser tuberosity). Although proximal humerus classifications such as the Neer classification have, in the past, been shown to have low interobserver reliability, and in particular, the Neer classification is a complex classification system that requires 3D understanding of the fracture morphology, we wondered whether 3D-printed models could decrease its high interobserver variability.

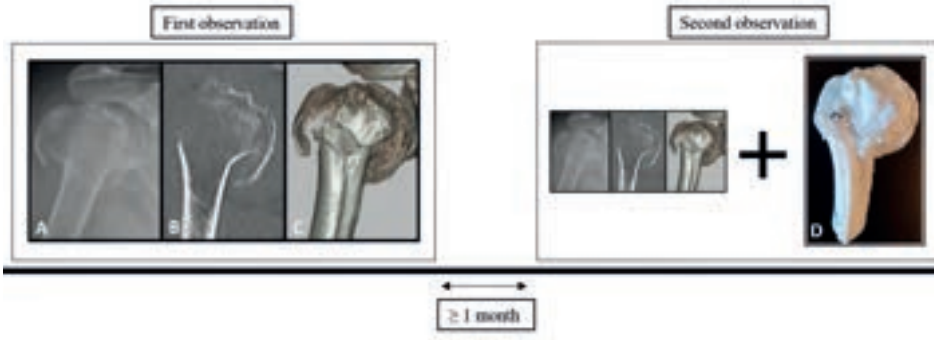


Figure 1. During the first observation, proximal humerus fractures were assessed with conventional onscreen imaging. During the second observation, 3D-printed models were added. The image labelled with the letter A represents the trauma radiograph, B the 2D CT image (coronal plane), C the 3D CT image (anterolateral aspect), and D the 3D-printed handheld model.

In the first session, assessment was completed with conventional imaging, which comprised standard trauma radiographs (AP and Y-view) and 2D and 3D CT images. During the second session, the same proximal humerus fractures were evaluated, but now a 3D-printed handheld model was used in adjunct to conventional imaging (Fig. 1). Conventional imaging was presented in RadiAnt DICOM viewer (Medixant, version 2020.1). With this software, participants could toggle through the radiographs, scroll through the various 2D CT slices, and rotate the 3D CT reconstructions over the x- and y-axes. Tools to perform measurements, and the

option to adjust contrast and brightness, were also available. Participants were not allowed to discuss cases; in both sessions they completed the assessment on their own.

Study patients

We considered patients potentially eligible to have their images included if they were aged 18 years or older, sustained a three-part or four-part proximal humerus fracture between 2015 and 2019 that was treated at the Level 1 trauma centre, and if they received a series of plain radiographs (AP and Y-view) and a 2D CT scan with 3D CT reconstruction images. All images, as well as the 3D CT reconstructions, were obtained as part of routine patient care and retrospectively collected from the medical imaging system Carestream Vue PACS (Carestream Health). In our Level 2 trauma centre, it was standard practice to perform a CT scan in patients with a displaced three- or four-part proximal humerus fracture. All 3D-printed fractures models were fabricated specifically for this study. Images were collected by the second author (B.S.) who reviewed all CT shoulder scans between January 1, 2015, and July 1, 2019. Within this period there were 77 three- and four-part proximal humerus fractures. Of those fractures, the second author (B.S.) included conventional imaging from 20 patients who were considered as having especially difficult-to-classify fractures. Availability of 2D CT image data in DICOM format was a prerequisite to create 3D-printed handheld models; thus, patients without CT images or those with poor-quality CT images were excluded. For each patient, trauma radiographs and CT images were downloaded and saved as a DICOM file and subsequently anonymized with a DICOM Cleaner (PixelMed Publishing, LLC). No attending surgeons who were part of the original care of these patients were involved in the study.

Description of 3D printing

All CT images were uploaded into a 3D slicer (The Slicer Community, version 4.10.2) for preprocessing of the 3D-printed model. To develop these skills, we followed online tutorials on the 3D slicer website [15]. Because CT images included the entire shoulder complex and part of the thorax, the proximal part of the humerus had to be cropped and contoured. This was done with a volume rendering tool by indicating the region of interest. After this, the shoulder was further segmented using thresholding. Thresholding is a semiautomatic segmentation process that selects areas based on signal intensity. The threshold used in this study to select the

proximal humerus while minimizing adjacent tissues or structures was between 250 pixels and 300 pixels for the lowest volume intensity and 2002 pixels for the highest volume intensity. All surrounding bone structures were removed with the island feature. Successively, the 3D surface model was built and exported to Ultimaker Cura (version 4.6, Ultimaker B.V.) as an OBJ file (file format for 3D images containing all necessary object data and coordinates). In this software, the models were sliced and subsequently printed with a standard nozzle (diameter of 0.4 mm) on a 1:1 scale with the layer height at 0.15 mm, infill density at 20%, and printing speed at 100 mm per second. All models were printed with support material that was manually removed after the printing was finished. The prints were made with an Ultimaker 2D + 3D printer (Ultimaker BV). The costs of 3D printing depend on the preprocessing time, type of printer, and printing material [16]. Preprocessing of the models required 45 minutes, the actual printing process required approximately 6 to 8 hours, and removal of support material required less than 15 minutes. In this study, we used a printer valued at USD 2650 with polylactic acid as the printing material. Polylactic acid costs approximately USD 30 per kg and labour of a resident at the start of his/her training approximately USD 26.75 per hour; thus, considering an average of 35 g needed per 3D-printed model, the cost of one model was USD 160 (printer = USD 132.50 [2650 ÷ 20], material = USD 1.05 [0.035 x 30], labour = USD 26.75).

Variables and outcome measures

The primary outcome was interobserver reliability and observers with different levels of experience were included in this study to represent a group of clinicians working within an orthopaedic department. The participants were two orthopaedic residents who just started their training, two residents who were halfway through their training, three attending orthopaedic surgeons: two with 11 to 20 years of experience, one who was within 5 years of finishing orthopaedic training (these residents and attending orthopaedic surgeons were members of the Traumaplatform 3D Consortium) and one attending orthopaedic surgeon with fellowship training in upper extremity surgery (D.E., >21 years of experience). All participants assessed the presence or absence of the following fracture characteristics on two occasions with an interval of at least 1 month: humeral head split, more than 2 mm medial hinge displacement, more than 8 mm metaphyseal extension, surgical neck fracture, anatomic neck fracture, displacement of the humeral head (varus, valgus, or no displacement), more than 10 mm lesser tuberosity displacement, and more than 10 mm greater tuberosity displacement. Observers also classified the fractures

according to the full Neer classification (16 options) and the Hertel binary LEGO description system (12 options). Answers to each question were provided on questionnaires and captured via REDCap (Vanderbilt University Medical Center) [17,18]. Participants were able to spend as much time on the assessment as they wished. Before each session, every participant was trained by two study authors (R.S., B.S.) using a sheet of paper with figures depicting all respective fracture classifications. Observers were allowed to keep these sheets during the assessments.

Ethical approval

This study was approved by the institutional review board at Flinders Medical Centre, Adelaide, Australia (reference number 50.19).

Statistical analysis

The statistical analysis was performed with Stata Statistical Software (Release 16, StataCorp LLC). The interobserver variability was determined with a multi-rater Fleiss kappa using bootstrapping with 1000 iterations and scored according to the Landis and Koch rating with the following categories: poor (kappa <0.00), slight (kappa 0.00-0.20), fair (kappa 0.21-0.40), moderate (kappa 0.41-0.60), substantial (kappa 0.61-0.80), and almost perfect (kappa 0.81-1.00) [19]. If values were missing, all ratings within a participant were excluded. The multi-rater Fleiss kappa values are provided with corresponding 95% confidence intervals. To determine whether there was a difference between the sessions, delta (difference in the) kappa was calculated with a 95% CI and a two-tailed *p*-value. A *p*-value less than 0.05 was considered statistically significant. A post hoc power analysis was conducted in PASS (version 21.0.2, NCSS LLC) by comparison of two independent proportions. Although this test could not control for the number of observers, it revealed that with 20 images a delta (difference in the) kappa of 0.40 could be established between the group with and without the 3D-printed handheld model at 80% power with alpha = 0.05.

We note that eight fracture-assessment questions were not completed. As the multi-rater Fleiss kappa analysis cannot handle missing data, these 8 of 400 fracture assessments were excluded through listwise deletion (20 patients assessed with conventional imaging, 20 patients with 3D-printed models, eight specific fractures characteristics, and two fracture patterns; [20 + 20] * [8 +2] = 400). The missing values were only present among residents.

Finally, a kappa value less than 0 indicates poor agreement. If the group with and without the 3D-printed handheld model are compared, a delta kappa value less than 0 indicates that agreement for conventional imaging with 3D models is lower than for conventional imaging only. If attending surgeons are compared with residents, it means that agreement for residents is lower than for attending surgeons.

RESULTS

Agreement on fracture characteristics and classification

Among the eight observers (four orthopaedic residents and four attending surgeons), assessment by 3D-printed handheld models together with onscreen imaging did not improve agreement regarding the following fracture characteristics: more than 2 mm of medial hinge displacement, more than 8 mm of metaphyseal extension, surgical neck fracture, anatomic neck fracture, displacement of the humeral head, more than 10 mm of lesser tuberosity displacement, and more than 10 mm of greater tuberosity displacement. Interobserver agreement for the presence of a humeral head-splitting fracture improved to a level that was still inadequate for clinical use (fair to substantial, delta kappa = 0.33 [95% CI 0.02 to 0.64]). The interobserver agreement for the Neer fracture patterns using conventional imaging was 0.13 (95% CI 0.06 to 0.20) and did not improve when assessed with 3D-printed models (delta kappa = 0.02 [95% CI -0.11 to 0.07]). Similarly, the agreement on the Hertel fracture patterns using conventional imaging was 0.14 (95% CI 0.06 to 0.22), and additional 3D-printed models did not result in improvement (delta kappa = 0.01 [95% CI -0.11 to 0.08]) (Table 1).

Table 1. Agreement for conventional imaging and 3D-printed models among all eight observers

Parameter	Conventional			Conventional + 3D			Δ kappa		*p*-value
	Kappa		Agreement	Kappa		Agreement			
Characteristics									
Humeral head split	0.39	(0.17-0.60)	Fair	0.72	(0.50-0.94)	Substantial	0.33	(0.02-0.64)	0.04
Medial hinge displacement >2 mm	0.19	(0.00-0.38)	Slight	0.35	(0.06-0.64)	Fair	0.16	(-0.19 to 0.51)	0.36
Metaphyseal extension >8 mm	0.14	(0.00-0.31)	Slight	0.28	(0.12-0.44)	Fair	0.14	(-0.09 to 0.36)	0.24
Surgical neck fracture	0.10	(0.00-0.26)	Slight	0.27	(0.05-0.50)	Fair	0.17	(-0.10 to 0.45)	0.21
Anatomic neck fracture	0.16	(0.01-0.31)	Slight	0.29	(0.10-0.47)	Fair	0.13	(-0.11 to 0.37)	0.29
Displacement of humeral head	0.44	(0.22-0.65)	Moderate	0.35	(0.16-0.53)	Fair	0.09	(-0.37 to 0.19)	0.55
LT displacement >10 mm	0.03	(-0.05 to 0.12)	Slight	0.16	(0.02-0.30)	Slight	0.13	(-0.04 to 0.29)	0.13
GT displacement >10 mm	0.13	(0.01-0.25)	Slight	0.16	(0.00-0.36)	Slight	0.03	(-0.21 to 0.26)	0.81
Fracture patterns									
Neer classification	0.13	(0.06-0.20)	Slight	0.11	(0.05-0.17)	Slight	0.02	(-0.11 to 0.07)	0.67
Hertel classification	0.14	(0.06-0.22)	Slight	0.13	(0.07-0.19)	Slight	0.01	(-0.11 to 0.08)	0.81

A kappa value less than 0 indicates poor agreement; a Δ kappa value less than 0 indicates that agreement for conventional imaging with 3D models is lower than conventional imaging only. Abbreviations: LT, lesser tuberosity; GT, greater tuberosity.

Agreement among residents and attending surgeons

Among residents, additional 3D-printed handheld models did not improve agreement regarding fracture characteristics and patterns (Table 2). Among attending surgeons, only agreement on lesser tuberosity displacement more than 10 mm improved from poor to slight (delta kappa = 0.22 [95% CI 0.01 to 0.42]), which was still insufficient for clinical use. Thus, adding 3D-printed handheld models to the diagnostic process likewise did not improve concurrence among attending surgeons (Table 3). There were no differences between residents and attending surgeons in terms of whether 3D models helped them to classify the fractures, and there were few differences in terms of whether the 3D models helped them to identify fracture characteristics. However, none of the identified differences improved to almost perfect agreement (kappa value above 0.80), so we do not see even those few differences as likely to be clinically useful (Table 4).

Table 2. Agreement for conventional imaging and 3D-printed models among four residents

Parameter	Conventional			Conventional + 3D					
	Kappa		Agreement	Kappa		Agreement	Δ kappa		*p*-value
Characteristics									
Humeral head split	0.48	(0.21-0.74)	Moderate	0.66	(0.37-0.95)	Substantial	0.18	(-0.21 to 0.58)	0.18
Medial hinge displacement >2 mm	0.02	(-0.15 to 0.20)	Slight	0.07	(-0.18 to 0.33)	Slight	0.05	(-0.25 to 0.35)	0.74
Metaphyseal extension >8 mm	0.27	(0.02-0.52)	Fair	0.28	(0.08-0.48)	Fair	0.01	(-0.32 to 0.33)	0.48
Surgical neck fracture	0.02	(-0.20 to 0.24)	Slight	0.13	(-0.15 to 0.40)	Slight	0.11	(-0.25 to 0.46)	0.28
Anatomic neck fracture	0.17	(-0.03 to 0.38)	Slight	0.28	(0.00-0.55)	Fair	0.10	(-0.24 to 0.45)	0.57
Displacement of the humeral head	0.29	(0.07 to 0.52)	Fair	0.1	(-0.07 to 0.26)	Slight	0.19	(-0.79 to 0.40)	0.52
LT displacement >10 mm	0.18	(-0.06 - 0.41)	Slight	0.17	(-0.06 to 0.40)	Slight	0.01	(-0.33 to 0.32)	0.97
GT displacement >10 mm	0.16	(-0.06 to 0.38)	Slight	0.13	(-0.13 to 0.39)	Slight	0.03	(-0.37 to 0.31)	0.43
Fracture patterns									
Neer classification	0.14	(0.03-0.25)	Slight	0.04	(-0.05 to 0.13)	Slight	0.10	(-0.25 to 0.04)	0.16
Hertel classification	0.14	(0.00-0.27)	Slight	0.13	(0.03-0.23)	Slight	0.01	(-0.17 to 0.15)	0.46

A kappa value less than 0 indicates poor agreement; a Δ kappa value less than 0 indicates that agreement for conventional imaging with 3D models is lower than for conventional imaging only. Abbreviations: LT, lesser tuberosity; GT, greater tuberosity.

Table 3. Agreement for conventional imaging and 3D-printed models among four attending surgeons

Parameter	Conventional			Conventional + 3D			Δ kappa		*p*-value
	Kappa		Agreement	Kappa		Agreement			
Characteristics									
Humeral head split	0.37	(0.05-0.69)	Fair	0.75	(0.48-1.00)	Substantial	0.38	(-0.04 to 0.79)	0.08
Medial hinge displacement >2 mm	0.38	(0.05-0.70)	Fair	0.62	(0.24-1.00)	Substantial	0.24	(-0.26 to 0.74)	0.34
Metaphyseal extension >8 mm	0.00	(-0.18 to 0.17)	Slight	0.16	(-0.02 to 0.33)	Slight	0.16	(-0.09 to 0.40)	0.20
Surgical neck fracture	0.22	(-0.16 to 0.59)	Fair	0.54	(0.10-0.98)	Moderate	0.33	(-0.25 to 0.90)	0.27
Anatomic neck fracture	0.13	(-0.09 to 0.34)	Slight	0.42	(0.17-0.66)	Moderate	0.29	(-0.03 to 0.61)	0.08
Displacement of the humeral head	0.58	(0.36-0.81)	Moderate	0.69	(0.41-0.97)	Substantial	0.11	(-0.25 to 0.47)	0.55
LT displacement >10 mm	-0.11	(-0.23 to 0.02)	Poor	0.11	(-0.05 to 0.28)	Slight	0.22	(0.01-0.42)	0.04
GT displacement >10 mm	-0.01	(-0.15 to 0.13)	Poor	0.13	(-0.14 to 0.40)	Slight	0.14	(-0.16 to 0.45)	0.36
Fracture patterns									
Neer classification	0.13	(-0.01 to 0.27)	Slight	0.05	(-0.04 to 0.13)	Slight	0.08	(-0.25 to 0.08)	0.32
Hertel classification	0.09	(-0.02 to 0.20)	Slight	0.06	(-0.01 to 0.19)	Slight	0.03	(-0.18 to 0.12)	0.68

A kappa value less than 0 indicates poor agreement; a Δ kappa value less than 0 indicates that agreement for conventional imaging with 3D models is lower than for conventional imaging only. Abbreviations: LT, lesser tuberosity; GT, greater tuberosity.

Table 4. Comparison of agreement between four residents and four attending surgeons using 3D-printed models

Parameter	Residents (conventional + 3D)			Attending surgeons (conventional + 3D)			Δ kappa		*p*-value
	Kappa		Agreement	Kappa		Agreement			
Characteristics									
Humeral head split	0.66	(0.37-0.95)	Substantial	0.75	(0.48-1.00)	Substantial	0.09	(-0.31 to 0.49)	0.65
Medial hinge displacement >2 mm	0.07	(-0.18 to 0.33)	Slight	0.62	(0.24-1.00)	Substantial	0.54	(0.09-0.99)	0.02
Metaphyseal extension >8 mm	0.28	(0.08-0.48)	Fair	0.16	(-0.02 to 0.33)	Slight	0.12	(-0.39 to 0.14)	0.36
Surgical neck fracture	0.13	(-0.15 to 0.40)	Slight	0.54	(0.10-0.98)	Moderate	0.42	(-0.10 to 0.94)	0.11
Anatomic neck fracture	0.28	(0.00-0.55)	Fair	0.42	(0.17-0.66)	Moderate	0.14	(-0.23 to 0.51)	0.45
Displacement of the humeral head	0.10	(-0.07 to 0.26)	Slight	0.69	(0.41-0.97)	Substantial	0.60	(0.27-0.92)	<0.001
LT displacement >10 mm	0.17	(-0.06 to 0.40)	Slight	0.11	(-0.05 to 0.28)	Slight	0.06	(-0.34 to 0.22)	0.68
GT displacement >10 mm	0.13	(-0.13 to 0.39)	Slight	0.13	(-0.14 to 0.40)	Slight	0.00	(-0.38 to 0.38)	>0.99
Fracture patterns									
Neer classification	0.04	(-0.05 to 0.13)	Slight	0.05	(-0.04 to 0.13)	Slight	0.01	(-0.12 to 0.13)	0.90
Hertel classification	0.13	(0.03-0.23)	Slight	0.06	(-0.01 to 0.19)	Slight	0.07	(-0.22 to 0.07)	0.30

A kappa value less than 0 indicates poor agreement; a Δ kappa value less than 0 indicates that agreement for residents is lower than for attending surgeons.
Abbreviations: LT, lesser tuberosity; GT, greater tuberosity.

DISCUSSION

Recognizing the overall pattern and specific characteristics of proximal humerus fractures may aid in decision-making and determining prognosis. However, there is considerable and undesired interobserver variability, even when observers receive training in the application of the classification systems used. Both the Neer and Hertel classifications are routinely reported in research studies, so to enhance our knowledge, we wanted to evaluate how reliably these injuries can be assessed with the assistance of 3D models. We, therefore, sought to determine whether cutting-edge technology (3D-printed fracture models), which now can be fabricated with desktop printers at relatively little cost, could deliver its promise and reduce the great undesired interobserver variability in fracture classification and characterization. If clinicians cannot agree, it will be challenging to evaluate results of proximal humeral fractures based on these classification schemes. In summary, we found that using 3D-printed handheld models with conventional imaging to assess three-part and four-part proximal humerus fractures did not improve agreement for fracture characteristics to a level that was adequate for clinical or scientific use. No improvement in agreement on fracture pattern recognition according to Neer and Hertel was established by using 3D-printed models together with onscreen conventional imaging. Residents did not seem to benefit more from 3D-printed handheld models than attending surgeons did. Hence, we do not recommend using these models in clinical practice if the goal is to improve classification reliability or to describe patients' fracture patterns or characteristics.

Limitations

Several limitations must be considered when interpreting our findings. First, we included only eight observers, which resulted in wide 95% CIs. However, as 3D-printed models are still relatively expensive and time-consuming, they must show strong value to be incorporated in clinical practice. Therefore, even a small study like ours should have been able to demonstrate the added value of 3D-printed models. For this reason, it was powered to detect profound differences between conventional imaging and 3D-printed models and not to detect subtle changes (post hoc power analysis revealed that with 20 images a delta kappa of 0.40 could be detected). Second, results should be inferred considering differences in experience among residents and attending surgeons. One may argue that residents and attending surgeons do not have the same knowledge level compared with an upper

extremity expert. This could have decreased the agreement; however, our goal was to include an observer panel that would represent typical orthopaedic practice in public hospitals. Third, in our observer cohort, there were missing values. To address this, proximal humerus fractures were listwise excluded from the analysis. This was 2% of the total number of proximal humerus fractures and could therefore have influenced our 95% confidence intervals. Notably, missing values only occurred among residents and mainly in the 3D group. In addition, we did not analyse intra-observer reliability. Classifying proximal humerus fractures is challenging; so much so that even advanced technology such as the 3D models used in our study could not improve agreement. We therefore argue that the classification is the main flaw and must be revised. We also note that diagnostic parameters, such as accuracy, were not included in this study. Ideally, this would be established intra-operatively, but because not all fractures were treated surgically, this was not feasible. A potential limitation here was that a cost analysis showing at what price such models would become cost-effective was not reported in this work. Something that is not effective cannot be cost effective; we therefore decided that a cost analysis should not be performed, given that effectiveness (in other words, interobserver reliability) was not established. Lastly, the Hertel fracture patterns were classified according to the original binary description system comprising 12 different categories [20] and thus without the two humeral-head split fracture types [12].

Agreement on fracture characteristics and classification

This study revealed that using 3D-printed handheld models in adjunct to onscreen imaging did not improve agreement regarding fracture characteristics. Based on these results, we cannot recommend using these models in the diagnostic workup of patients with proximal humerus fractures, especially because these models require time, materials, and money to produce. Only one study reported on the use of 3D-printed models to assess the characteristics of proximal humerus fractures [21]. They retrospectively compared pre-operative planning with conventional imaging, virtual planning, and 3D-printed models in patients undergoing internal fixation with locking plates and assessed clinical outcomes and the accuracy of fracture characteristics. Their results were based on intra-operative findings as the reference standard, and they therefore determined diagnostic parameters and not interobserver agreement. However, consistent with our study, they did not reveal any differences between 3D-printed models and conventional imaging.

The kappa value for fracture patterns according to the Neer classification in prior studies ranges from 0.07 to 0.14 (16 observers) [3], and from 0.39 to 0.60 (four observers) for the Hertel classification [22] with the availability of radiographs and 2D and 3D CT images. In our study, fracture pattern recognition according to the Neer and Hertel classifications had low interobserver agreement in all imaging modalities despite 3D-printed modelling (Neer: kappa = 0.11, Hertel: kappa = 0.13). One study demonstrated fair-to-moderate agreement for the simplified Neer classification (three categories: two-part, three-part, and four-part fractures) among 20 residents (kappa = 0.40) and 20 attending surgeons (kappa = 0.50) when using 3D-printed models only (without additional imaging). This supports that 3D models are not clinically useful for classifying proximal humerus fractures but the question remained unanswered if other classifications, such as the Hertel LEGO description system, or specific fracture characteristics would improve with 3D modelling [10]. Another study found moderate agreement (kappa = 0.47) among 14 assessors, but they also simplified the Neer classification to three categories: two-part, three-part, and four-part fractures, and assessed the 3D-printed models without additional radiographs or CT images [9]. Again, the question whether 3D fracture models would be useful for characterization and assessment of other fracture classification systems was left open. Combining these studies with our data, it seems justifiable to say that the utility of 3D models in determining fracture assessment of proximal humerus fractures is negligible. Nevertheless, 3D models may help create surgical strategies and approaches, such as guides to place K-wires and screws. They may also be valuable for educational purposes (such as teaching medical students or explaining surgical plans pre-operatively), but well-designed follow-up studies are needed to identify any potential benefits.

Agreement among residents and attending surgeons

There were no important differences between residents and attending surgeons in whether 3D models helped them to classify or describe the fractures, and the few observed differences were not sufficiently large to be clinically useful (Table 4). These findings were in line with another study that did not find any differences in agreement between residents and attending surgeons [10]. It is likely that because of the complexity of three-part and four-part proximal humerus fractures, assessment is difficult and debatable for both residents and attending surgeons. It also confirms that the hallmarks of proximal humerus fractures are seen differently and subjectively by observers, and that they are difficult to categorize in any

classification scheme. Considering this, we do not recommend using the currently available classification systems for supporting clinical decisions or to report on patient outcomes. Time-consuming interventions like the 3D-printed models used in this study did not overcome the shortcomings of difficult-to-use classifications; keeping those classifications as simple as possible therefore seems important.

Conclusion

Using 3D-printed handheld models with onscreen conventional imaging (radiographs and 2D and 3D CT images) to assess three-part and four-part proximal humerus fractures did not improve agreement regarding fracture characteristics and patterns. Therefore, we cannot recommend that clinicians expend the time and costs needed to create these models if the goal is to classify or describe patients' fracture characteristics. Future studies are needed to establish the value of 3D modelling in practicing fracture fixation and templating a pre-operative plan.

REFERENCES

1. Bruinsma WE, Guitton TG, Warner JJP, Ring D. Interobserver reliability of classification and characterization of proximal humeral fractures. *J Bone Joint Surg Am*. 2013;95(17):1600-1604.
2. Janssen SJ, Hermanussen HH, Guitton TG, van den Bekerom MPJ, van Deurzen DFP, Ring D. Greater tuberosity fractures: does fracture assessment and treatment recommendation vary based on imaging modality? *Clin Orthop Relat Res*. 2016;474(5):1257-1265.
3. 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.
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-S24.
5. Berkes MB, Dines JS, Little MTM, 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(15):1281-1286.
6. Mitsouras D, Liacouras P, Imanzadeh A, et al. Medical 3D printing for the radiologist. *Radiographics*. 2015;35(7):1965-1988.
7. Brouwer KM, Lindenhovius AL, Dyer GS, Zurakowski D, Mudgal CS, Ring D. Diagnostic accuracy of 2- and 3-dimensional imaging and modeling of distal humerus fractures. *J Shoulder Elbow Surg*. 2012;21(6):772-776.
8. Cocco LF, Aihara AY, Franciozi C, Dos Reis FB, Luzo MVM. Three-dimensional models increase the interobserver agreement for the treatment of proximal humerus fractures. *Patient Saf Surg*. 2020;14:33.
9. Bougher H, Buttner P, Smith J, et al. Interobserver and intraobserver agreement of three-dimensionally printed models for the classification of proximal humeral fractures. *JSES Int*. 2021;5(2):198-204.
10. Cocco LF, Yazzigi JA, Kawakami EFKI, Alvachian HJF, Dos Reis FB, Luzo MVM. Inter-observer reliability of alternative diagnostic methods for proximal humerus fractures: A comparison between attending surgeons and orthopedic residents in training. *Patient Saf Surg*. 2019;13(1):1-13.
11. Neer CS 2nd. Displaced proximal humeral fractures. I. Classification and evaluation. *J Bone Joint Surg Am*. 1970;52(6):1077-1089.
12. Hertel R, Hempfing A, Stiehler M, Leunig M. Predictors of humeral head ischemia after intracapsular fracture of the proximal humerus. *J Shoulder Elbow Surg*. 2004;13(4):427-433.
13. Sumrein BO, Mattila VM, Lepola V, et al. 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(10):1756-1761.
14. Majed A, Macleod I, Bull AMJ, et al. Proximal humeral fracture classification systems revisited. *J Shoulder Elbow Surg*. 2011;20(7):1125-1132.
15. 3D Slicer. Documentation/4.10/Training - Slicer Wiki. Available at: https://www.slicer.org/wiki/Documentation/4.10/Training#Introduction:_Slicer_4.10_Tutorials.
16. Morgan C, Khatri C, Hanna SA, Ashrafian H, Sarraf KM. Use of three-dimensional printing in preoperative planning in orthopaedic trauma surgery: A systematic review and meta-analysis. *World J Orthop*. 2020;11(1):57-67.
17. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208.

18. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
20. Hertel R, Mees C, Scholl E, Ballmer FT, Siebenrock K. Morphologic classification of fractures of the proximal humerus. A validated, teachable and practicable alternative. Presented at the 8th International Conference on Shoulder Surgery (ICSS). 2001:23-26.
21. 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(22):1960-1968.
22. Iordens GIT, Mahabier KC, Buisman FE, 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(5):596-602.


CAIRNS
MELBOURNE

4

Convolutional neural networks can accurately detect but not classify proximal humerus fractures

Reinier W.A. Spek
Marat Sverdlov
William J. Smith
Yang Zhao
Zhibin Liao
Johan W. Verjans
Jasper Prijs
Sebastiaan Broos
Minh-Son To
Henrik Åberg
Wael Chiri
Bhavin Jadav
John White
Gregory I. Bain
Frank F.A. IJpma
Paul C. Jutte
Michel P.J. van den Bekerom
Ruurd L. Jaarsma
Job N. Doornberg

The Machine Learning Consortium
The Bone & Joint Journal. 2024 Nov;106-B(11):1348-1360

ABSTRACT

Aims

To develop a convolutional neural network (CNN) and evaluate diagnostic performance characteristics for (1) fracture detection; and (2) classification for proximal humerus fractures (PHFs) on plain radiographs.

Methods

A CNN was designed and tested on radiographs sourced from eleven public hospitals in Australia. This was subsequently externally validated using radiographs from two Dutch hospitals. The study included radiographic images from patients with a PHF as well as those with healthy shoulders. A prerequisite for inclusion was the availability of corresponding CT scans, which established the reference standard based on multi-rater consensus. Each radiograph underwent evaluation on the presence of a fracture. If a fracture was detected, it was further categorized into one of four classifications, using both plain radiographs and two-dimensional and three-dimensional (2D and 3D) CT scans: 1) non- to minimally displaced; 2) two-part; 3) multi-part; and 4) glenohumeral dislocation.

Results

The algorithm was trained on 1,709 radiographs (n = 803, mean age: 59.5±20.7; 452 healthy shoulders, 351 fractures), tested on 567 radiographs (n = 244), and subsequently externally validated on 535 radiographs (n = 227). The overall accuracy for fracture detection was 94% (AUC = 0.98) with excellent diagnostic performance on external data as well: 92% (AUC = 0.96). However, the overall accuracy for classification was 78% while non- to minimally displaced was 36% (AUC = 0.68), two-part 53% (AUC = 0.80), multi-part 80% (AUC = 0.93), and dislocated fractures 46% (AUC = 0.73). External validation yielded a similar overall accuracy of 75% for fracture classification.

Conclusion

CNNs proficiently rule out PHFs on plain radiographs. Despite rigorous methodology with multi-rater consensus on advanced CT imaging as the reference standard, diagnostic parameters for AI driven classification are insufficient for clinical implementation. One could argue that identification of pathoanatomy on plain radiographs is on par with human observation for simple tasks such as PHF recognition, but diagnostic performance decreases significantly when tasks become increasingly complex.

INTRODUCTION

Although shared decision-making has improved due to better understanding of patient demographics, fracture related factors and surgical challenges that are all associated with outcomes in proximal humerus fractures (PHFs), debate is ongoing on which patients to select for surgical treatment [1]. One of the challenges related to this issue is the low interobserver reliability and lack of adequate –and reproducible– definitions for classifications leading to potentially biased interpretation of clinical studies [2,3]. In an attempt to reduce poor interobserver reliability for fracture patterns, various (new) technologies –like three-dimensional (3D) Computed Tomography (CT) scans, and 3D-printed hand held fracture models– have been introduced but did not provide sufficient improvement in human interobserver reliability to date (kappa values did not exceed 0.47) [2,4].

Deep learning algorithms are rapidly gaining attraction for their unparalleled speed in data processing and their unbiased capacity to tackle intricate challenges in image analysis [5]. Consequently, they are increasingly employed in contemporary medicine and fracture care [6–8]. To date, only one study has explored deep learning for PHFs in patients over 18 years old [9]. The authors developed a convolutional neural network (CNN) to detect and classify PHFs on conventional radiographs with an accuracy of 96% (CI: 94 - 97%) for fracture detection and between 65% and 86% for fracture classification. While the authors achieved satisfactory results and showed performance on par to human observers (fracture detection = 93%, classification = 65 - 93%), the algorithm was trained without using CT scans for ground truth (i.e. unreliable plain radiographic classification of PHF), and lacked external validation. The current presented study was conducted to fill the knowledge gaps essential for adoption of such an algorithm into clinical practice.

The purpose was to develop a CNN and evaluate its performance for (1) fracture detection; and (2) classification of PHFs on conventional radiographs. Multi-rater consensus agreement based on advanced CT imaging was used as the reference standard for training. We asked: what are the diagnostic performance characteristics for: a) AI driven “simple” fracture detection; and b) “more complex” PHF classification?

PATIENTS AND METHODS

Setting and study design

This retrospective diagnostic imaging study was primarily conducted at a Level 1 trauma centre and a specialized affiliated machine learning institute (both situated in Adelaide, South Australia). External validation was performed on radiographs from two Dutch hospitals: one Level 1 trauma centre and one Level 2 trauma centre. This manuscript was written in concordance with the CONSORT-AI checklist [10]. Ethics approval was obtained from each participating centre.

Population: fracture dataset

Inclusion criteria were availability of a CT scan (within four weeks after initial presentation at the Emergency Department) and a non-pathological PHF with at least one glenohumeral (x-ray beam perpendicular to the glenohumeral joint) or one patient-oriented (x-ray beam oriented in the sagittal plane of the patient) anteroposterior (AP) view. Radiographs were excluded if patients wore a brace, quality was poor (e.g., over- an underexposed images, substantial pixel loss), fracture extended beyond the metaphysis into the diaphysis, or if the amount of radiographic displacement did not resemble the displacement on the CT scan (e.g., if the radiograph and CT scan were taken at different time point, the shaft –or other fragments– could have been substantially more displaced on the radiograph or CT scan).

Population: healthy shoulder dataset

Radiographs of patients were included if they showed a non-fractured, non-dislocated shoulder without old- or acute fractures at other sites (e.g., humeral shaft, or scapula fractures). In both the fracture and healthy shoulder dataset, a minimum age of 18 years or older was required and patients were excluded if their radiograph(s) revealed rotator cuff pathology (Hamada stage 2,3,4 or 5 [11]), severe glenohumeral osteoarthritis (stage 3 and 4 according to Samilson and Prieto [12]), an open epiphyseal plate, or if there was inconclusive evidence regarding the presence or absence of a fracture.

Data collection: fracture dataset

The CNN was developed and validated on anonymized radiographs collected from eleven hospitals within South Australia. All humerus- and shoulder CT scans with radiographs and radiology reports were downloaded from the South Australian

medical imaging servers (inclusion period: 2007 - April 2021). After assessment of 2541 CT scans, 468 patients were included (Supplement 1). External validation was performed on data from two Dutch hospitals: one Level 1 trauma centre and one Level 2 trauma centre. In the Level 1 trauma centre, period of inclusion was January 2016 - June 2021. In the Level 2 trauma centre, this was between March 2014 and September 2020. Eventually, 395 patients were screened of which 126 were included. The radiograph closest to the date of the CT scan was retrieved. All available radiographic views were collected, and if certain views were taken more than once, the best quality image was selected.

Data collection: healthy shoulder dataset

All shoulder and humerus radiographs obtained between November 2015 and August 2020 were downloaded from a Level 1 South Australian hospital (n = 10,563). Eligible radiographs were identified by filtering on keywords in the radiology reports and radiographic assessment. This led to a database of 1,292 healthy shoulders with ≥2 different views per patient and allowed us to randomly select as many shoulders as needed. For external validation radiographs were collected from a Level 2 Dutch hospital: 254 patients were screened (2018 - November 2021) of which 125 patients were included. Assessment against the inclusion and exclusion criteria across all datasets was done by two researchers independently.

Figure 1. Fractured *versus* non fractured shoulder.

Reference standard: fracture dataset

Each radiograph underwent evaluation on the presence of a fracture. If present, the following labels were allocated with the aid of 2D and 3D CT scan images: 1) non- to minimally displaced; 2) two-part; 3) multi-part or 4) glenohumeral dislocation (Fig. 1, 2). RadiAnt DICOM viewer (Medixant, Poznan, Poland) was used to assess the CT scans and/or radiographs and allowed for multiplanar reconstruction and creation of virtual 3D models aside from the standard assessment tools [13]. Determining the presence of a fracture and its classification was always performed by two or more independent observers with consensus obtained during in-person meetings. Multiple surgeons were involved and assessed 44% of the data, the following levels of consensus were distinguished (Table 1):

a) *Medical researcher assessment*: independent assessment by two trained medical researchers (R.S. *versus* M.S., J.S., or S.M.) with doubts to be reviewed by a trauma or upper limb surgeon.
b) *Single surgeon assessment*: assessment by one upper limb or trauma surgeon. Their chosen classification was compared to the answers from the medical researchers with discrepancies to be resolved by discussion. Three surgeons were involved: one upper limb fellow (W.C.) who was within one year of finishing surgical training, one trauma surgeon (H.A.) who practiced independently for five years after finishing training, and one academically oriented upper limb surgeon with more than 10 years of clinical experience (M.vdB.).
c) *Dual surgeon assessment:* assessment by two academically oriented upper limb, or trauma surgeons. After they agreed on the correct label, their answer was compared to the medical researchers' consensus. If needed, discrepancies were resolved by discussion. Four expert orthopaedic surgeons were involved, each with ≥15 years of clinical experience after finishing surgical training (G.B., J.W., R.J., and B.J.).

For external validation, assessment by the surgeon was completed based on key images. At this stage, medical researchers' assessment skills were close to dual surgeon judgement and fractures were already extensively studied, discussed, and (re-)defined. Hence, consensus with two surgeons was deemed unnecessary: for fracture detection, there was 100% agreement, the kappa for classification was almost perfect (kappa = 0.86, 95% CI: 0.78 - 0.94) based on assessment of 117 CT scans with an absolute agreement of 90.6% [14].

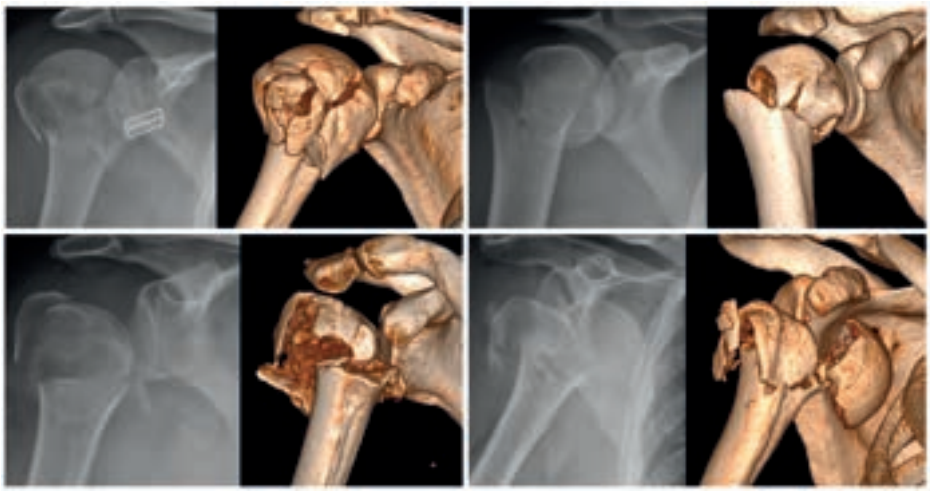


Figure 2. Left upper image = non- to minimally displaced proximal humerus fracture; right upper image = two-part proximal humerus fracture; left bottom image = multi-part proximal humerus fracture; right bottom image = proximal humerus fracture with glenohumeral dislocation.

Reference standard: healthy shoulder dataset

Each shoulder radiograph was judged without CT scans: 433 radiographs were scored by medical researchers and 249 by an upper limb or trauma surgeon (W.C., H.A., J.D., or M.vdB.) (Table 1). Radiographs were shown to the surgeons on Labelbox (Labelbox, San Francisco, United States of America) which allowed them to pan, zoom, and adjust contrast.

Definitions

Classification was done based on the fundamentals of Neer's system [15]. Some adjustments were made however, to improve interobserver reliability. The following rules were applied: 1) four anatomic segments were distinguished: humeral shaft, greater tuberosity, lesser tuberosity, and the articular segment. If any of these fragments were displaced ≥1 cm or angulated ≥45º they were considered a separate part, 2) three- and four-part fractures were grouped and named multi-part fractures [16], 3) greater tuberosity displacement was judged in relation the articular segment and/or the lesser tuberosity [17], 4) if there was disagreement between observers, the maximum displacement was measured by consensus on CT. Each anatomic fragment was defined according to Hasan et al (Figure 2 in their paper [18]). Displacement of the greater tuberosity was calculated with the following formula: (combined anterior/posterior and medial/lateral displacement)^2 + (superior

or inferior displacement)^2 = (total greater tuberosity displacement)^2. Within the defined volume of the greater tuberosity, we chose to measure the biggest amount of displacement. Multiplanar reconstruction was used to adequately align the CT planes with the greater tuberosity (Supplement 2 - 5). Glenohumeral dislocations were defined as the absence of contact between the humeral head and glenoid (head located anterior or posterior to the glenoid) or full posterior rotation of the humeral head (subluxations or extreme varus tilted fractures were *not* included) (Supplement 6 - 8).

Annotations

All PHFs and healthy shoulders were annotated in Labelbox to guide the CNN to the area of interest. As this software did not support the use of DICOM files (which was the default format of our imaging data), the different views from the anonymized DICOMs were first converted to a PNG file using ImageMagick (ImageMagick Studio LLC, 2023) and then augmented to one image with a Matlab script (Matlab version 9.12, MathWorks, Natick, United States of America). Each humerus was annotated with a bounding box, and if present also the fracture. All fractures lines, (displaced) fragments, and gaps were delineated within each annotation. (Fig. 3 - 5). Annotations were completed by the first author. The training and external validation images were reviewed by another medical researcher (J.S., M.S., or S.M.), the test set images by an attending surgeon (R.J., M.vdB., or J.D.). Healthy shoulders were annotated by a medical researcher (J.S., M.S., or C.L.) and reviewed by the first author.

Algorithm development

Detectron2 implementation of Mask R-CNN was used [19]. The backbone of the Faster R-CNN model was set to the Microsoft Research Asia version of the ResNet-50 CNN model, pretrained on ImageNet [20,21]. To specify the training strategy, the Faster R-CNN was trained with stochastic gradient descent for 6250 iterations with the initial learning rate of 0.02 and a mini batch of 8 images per iteration. The learning rate was reduced by a factor of 10 at iteration 3750 and 5625, respectively. Weight decay and momentum were set at $1x10^{-4}$ and 0.9. Our backbone networks were initialized with the weights pre-trained on ImageNet. All experiments were performed with PyTorch deep learning framework on an Nvidia RTX 3090 GPU (Nvidia, United States).

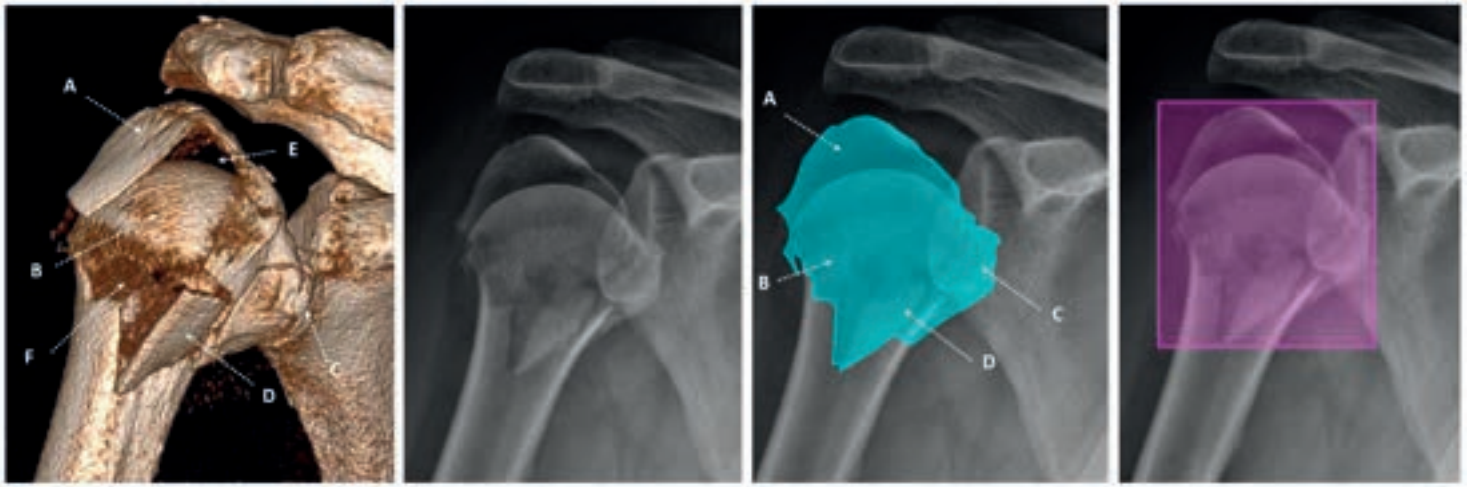


Figure 3. Annotations in the fracture group: this proximal humerus fracture involves the anatomic neck, surgical neck and both the tuberosities. Hence each of these segments were annotated. A = Greater tuberosity, B = articular segment, C = lesser tuberosity, D = surgical neck. Fracture gaps (= E) and areas of cancellous bone (= F) were also annotated. The final annotation with a bounding box is shown on the most rightward image.

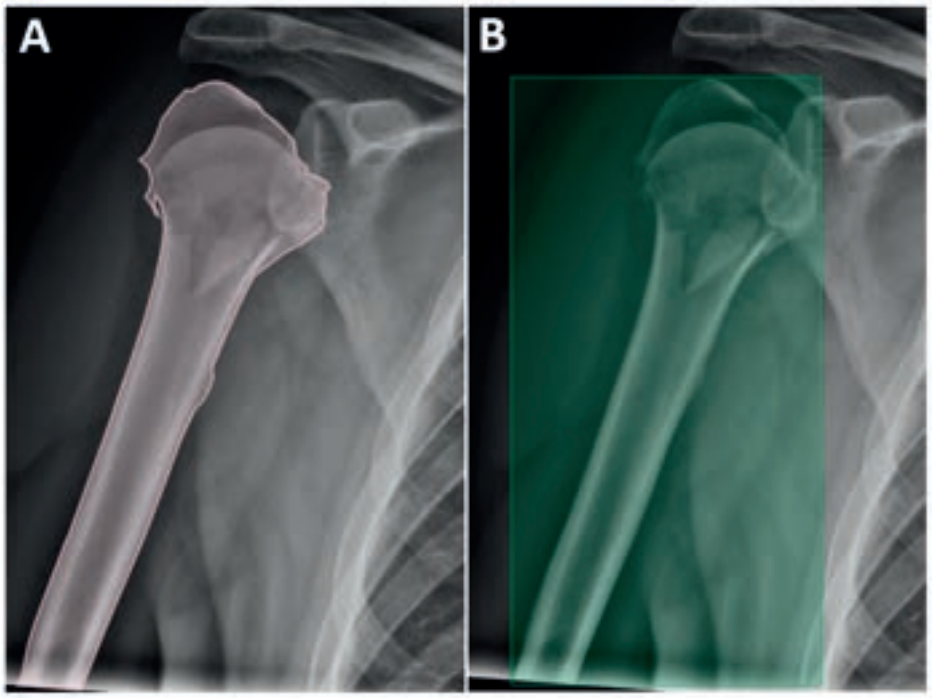


Figure 4. Annotations in the fracture group: each humerus (A) was annotated with a bounding box (B).

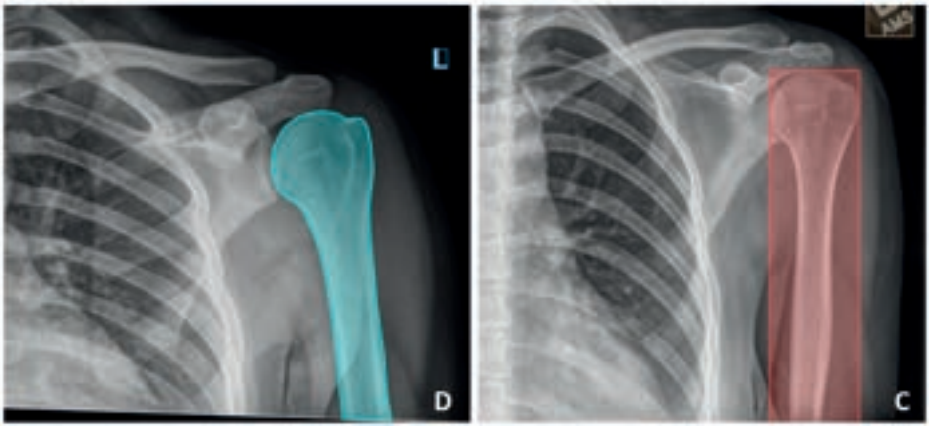


Figure 5. Healthy shoulder annotations: each humerus (A) was annotated with a bounding box (B).

Baseline demographics

The algorithm was trained on 1,709 radiographs from 803 patients with a mean age of 59.5 ± 20.7 (452 healthy shoulders, 351 fractures). In the fracture group, 11% sustained a non- to minimally displaced fracture, 34% a two-part fracture, 45% a multi-part fracture, and 11% a dislocation fracture (based on CT scans). After finishing the training process, the performance was internally validated 567 radiographs (n = 244) (Table 1, 2).

Table 2. Distribution of fracture patterns in each dataset

	Training (n = 351)	Internal validation (n = 117)	External validation (n = 124)
Classification			
Non- to minimally displaced	37 (11%)	14 (12%)	9 (7%)
Two-part	118 (34%)	45 (39%)	46 (37%)
Multi-part	157 (45%)	45 (39%)	57 (46%)
Dislocation	39 (11%)	13 (11%)	12 (10%)

Outcomes measures

To evaluate the performance of the model the following diagnostic performance metrics were calculated with 95% confidence interval: area under the receiving operating characteristic (AUC) curve, top-1 accuracy, sensitivity, specificity and the Youden index.

Statistical analysis

IBM SPSS software version 27 (IBM Corp., Armonk, N.Y., USA) was used for calculating the baseline demographics. Cohen's kappa values were calculated to determine the agreement on fracture assessments between the medical researchers and dual surgeon assessment.

RESULTS

Internal validation

To answer the first research question, what are the diagnostic performance characteristics for AI driven "simple" fracture detection, we found an overall accuracy for fracture detection of 0.94. For the second question, what is the diagnostic performance for "more complex" PHF classification, the accuracy for the classification task was 78% and showed the best performance for multi-part fractures (80%, AUC = 0.93, sensitivity = 0.80, specificity = 0.91) and the worst performance for non- to minimally displaced fractures (36%, AUC = 0.68, sensitivity = 0.36, specificity = 0.97) (Table 3).

External validation

External validity was evaluated on 535 radiographs (n = 227). Fractures were detected with 92% accuracy. Compared to the internal validation group, classification on external data yielded a similar overall accuracy of 75% but lower accuracies for the individual classifications. Best CNN performance was observed for two-part fractures with an accuracy of 65% (AUC = 0.82, sensitivity = 0.65, specificity = 0.86). Non- to minimally displaced fractures were the most challenging to detect (44% accuracy) (Table 3).

Classification pitfalls and strengths

Notably, the CNN found it challenging to distinguish two-part fractures from dislocations on the internal validation dataset. Moreover, it struggled to differentiate the multi-part fractures from the two-part fractures (external validation) but if the CNN was shown a multi-part fracture it could almost flawlessly separate it from a healthy shoulder or non- to minimally displaced fracture (Table 4, Fig. 6 - 9).

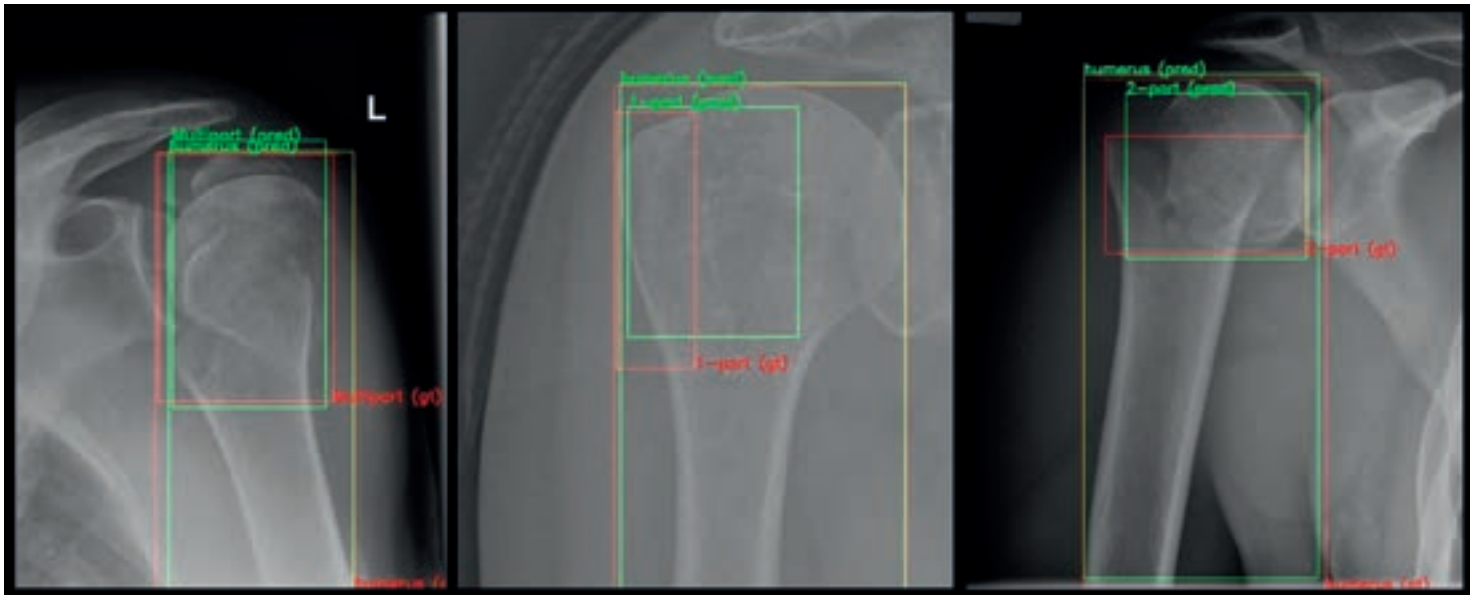


Figure 6. Example of three adequate classifications. Of note, the area of interest from the algorithm did not correspond completely with our annotation of interest: in the two images on the right, the algorithm's attention was drawn to an area wider than the actual fracture. Abbreviations: GT, ground truth, pred, prediction.

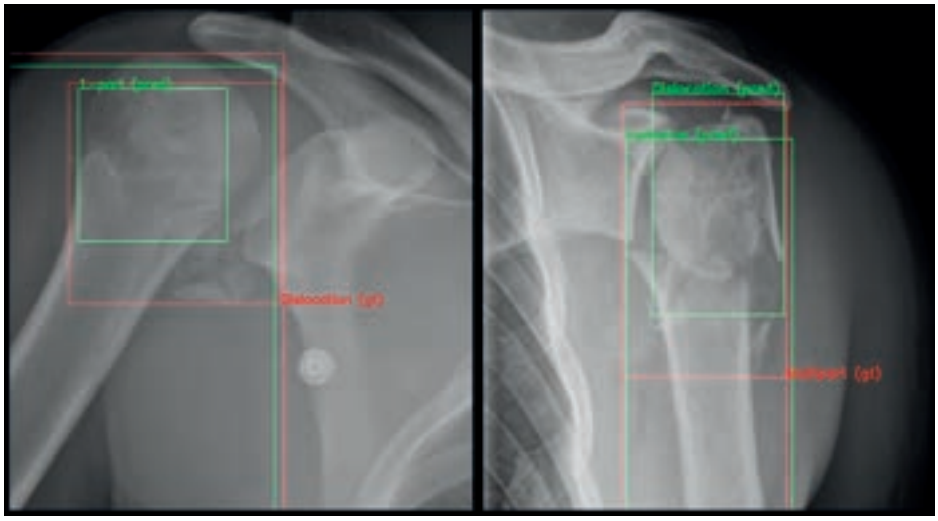


Figure 7. Example of two misclassifications. The left radiograph showing a posterior dislocation was misclassified as a one-part fracture. The multi-part fracture on the right was misclassified as a glenohumeral dislocation. Abbreviations: GT, ground truth; pred, prediction.

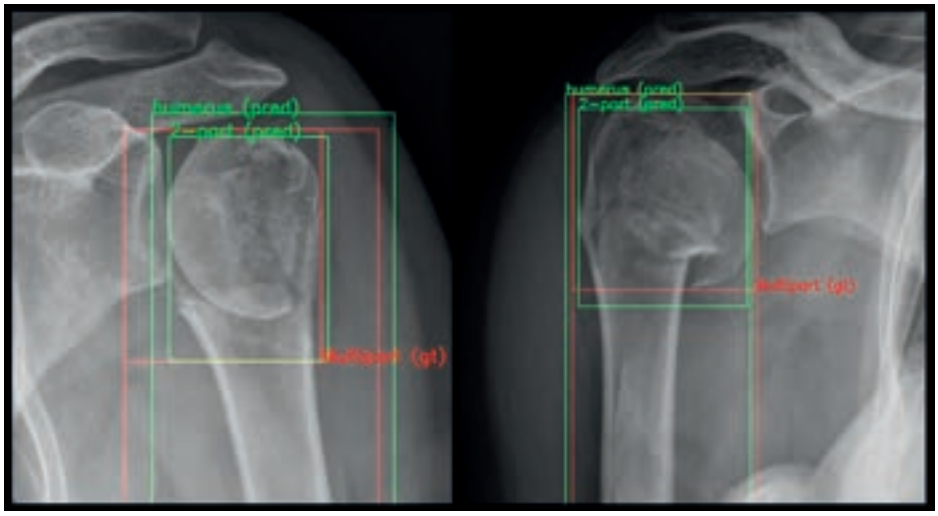


Figure 8. Example of a common mistake. Both multi-part fractures were misclassified as a two-part fracture. Abbreviations: GT, ground truth, pred, prediction.

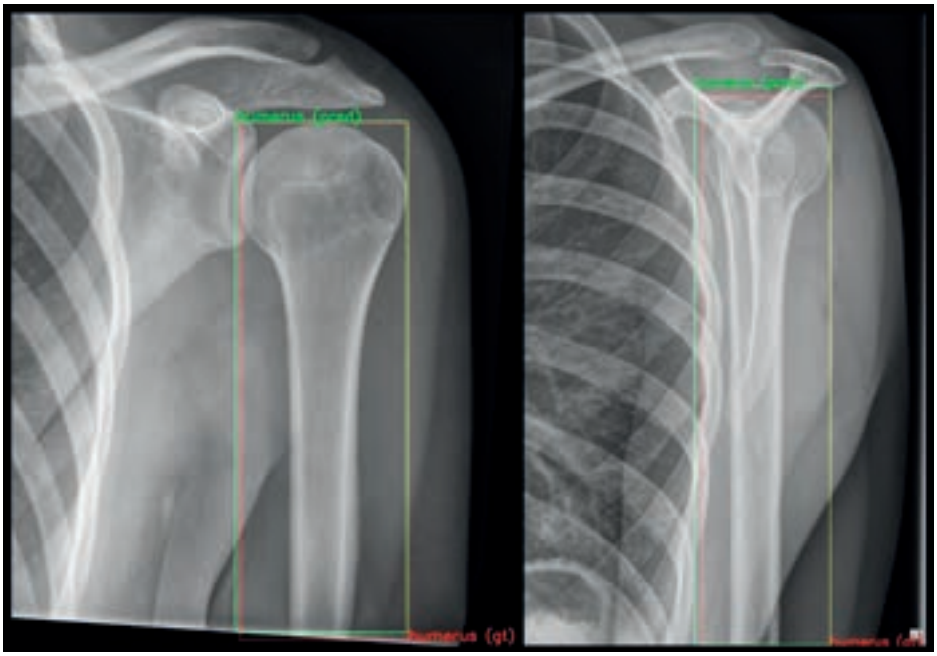


Figure 9. Left image: example of an almost-perfect prediction of the humeral bone in a healthy shoulder (the boxes are close to 100% overlap). Right image: localisation of the humerus bone (green) was relatively close to the true location of the humerus bone (red). Abbreviations: GT, ground truth; pred, prediction.

DISCUSSION

The classification of PHFs is often inconsistent and is challenged by poor interobserver reliability, since surgeons tend to align with their own interpretations rather than others due to inherent human biases [2,22–24]. This inconsistency in clinical trials introduces bias and reduces comparability between studies [1,25]. A CNN could theoretically overcome human biases and increase reliability of fracture classification. Our algorithm detected fractures with an accuracy of 94% and classified them with 78% accuracy according to the ground truth established through multi-rater consensus agreement. External performance evaluation yielded comparable results.

This diagnostic imaging study should be interpreted in the light of strengths and weaknesses: The core strengths of this project are (1) adherence to CONSORT-AI checklist [10]; (2) external validation to achieve more generalisability since an algorithm may perform differently on different datasets from other hospitals worldwide [26];

(3) a robust ground truth established through collaborative evaluations by multiple observers, guided by various attending surgeons, with CT scans and interobserver statistics; and (4) use of state-of-the-art machine learning technology, created by experienced developers in an academic setting [27].

Several limitations must be considered. First, there was an imbalance across the datasets with respect to non- to minimally displaced fracture and glenohumeral dislocations. This may have resulted in limited performance as the CNN had limited exposure to these entities during training. Collecting cases with these fracture patterns may improve training and accuracy and is advised to be done in a follow-up project, The current distribution however resembled clinical practice and the distribution of these patterns in our consecutive CT population. Second, machine specifications were not uniform across hospitals and considerable variability in patient positioning was observed. This could have compromised the accuracy of classification. For instance, some images were captured with too much inferior tilt, some views had their center point between a standard and glenohumeral AP view, and for most patients only patient-oriented standard AP and lateral views were obtained (the ideal composite would include an axillary and glenohumeral AP as well). One could argue however that this resembles clinical practice and therefore enhances the generalizability of our CNN. Third, nuances of each definition should be carefully read to adequately understand the algorithms output. For example, humeral head subluxations are relatively common (especially inferior subluxations). As these patterns still have some degree of contact with the glenoid cartilage, these were not classified as glenohumeral dislocation.

Our results are similar to the performance of Chung's model: their algorithm returned a 96% accuracy with an AUC of 1.0 (515 healthy shoulders, 1376 fractured shoulders) [9]. Noteworthy though is that our study merely included PHFs confirmed on CT scans. As such, Chung's study included more patients with simple fractures which could have resulted in their slightly better AUC. Our results are in line with algorithms for other common fractures trained solely on plain radiographs, such as studies on hip (AUC = 0.91, n = 1554 non-fractures, n = 1472 fractures), and distal radius fractures (AUC = 0.96, n = 849 non-fractures, n = 1491 fractures) [28,29]. Additionally, Zech et al. developed an algorithm to identify upper extremity fractures (including the proximal humerus) in paediatric patients and revealed 89.7 % accuracy with an AUC of 0.96. [30].

Current CNN correctly classified PHFs with an accuracy of 78%. Chung's algorithm reached better results with an accuracy of 86% to detect greater tuberosity fractures, 80% for surgical neck fractures, 65% for three-part fractures and 75% for four-part fractures [9]. This study however, lacked external validation. Performance of our CNN on external validation was unsatisfactory (49%) and therefore too low to use in clinical practice. CT scans are known to be much more precise in determining fracture patterns but were in Chung's paper only used in difficult cases. As such, they may have introduced bias towards simpler fractures. Although the authors were the first to develop an algorithm on PHFs which has showed a great step forward in computerised possibilities in health care, this is an important drawback and the main reason for us to use CT scans as reference standard.

As even orthopaedic surgeons merely reach 93% accuracy for fracture detection on radiographs, one could argue that current CNN should already be tested in clinical practice with prospective studies for ongoing validation and audit [9]. For the most optimal classification performance, we recommend training the CNN on CT scans first. Due to the endless opportunities and applications, we strongly argue for further algorithm development on clinically useful tasks [31]. It should be stressed that algorithms have abilities which cannot be mastered by humans to the same extent. CNNs outpace humans with regards to data volume, speed, pixel analysis, and (most importantly) objective decision-making. In clinical practice, this can be translated to a more efficient health care system where doctors closely collaborate with AI: human talents such as social intelligence and creativity are assisted on-demand by the perks computers have to offer. Other future benefits may range from providing help in developing countries to standardizing surgical indications and combining it with machine learning prediction tools to reduce bias in clinical decision-making. CNNs are currently best used as a tool in adjunct to human judgement.

Conclusion

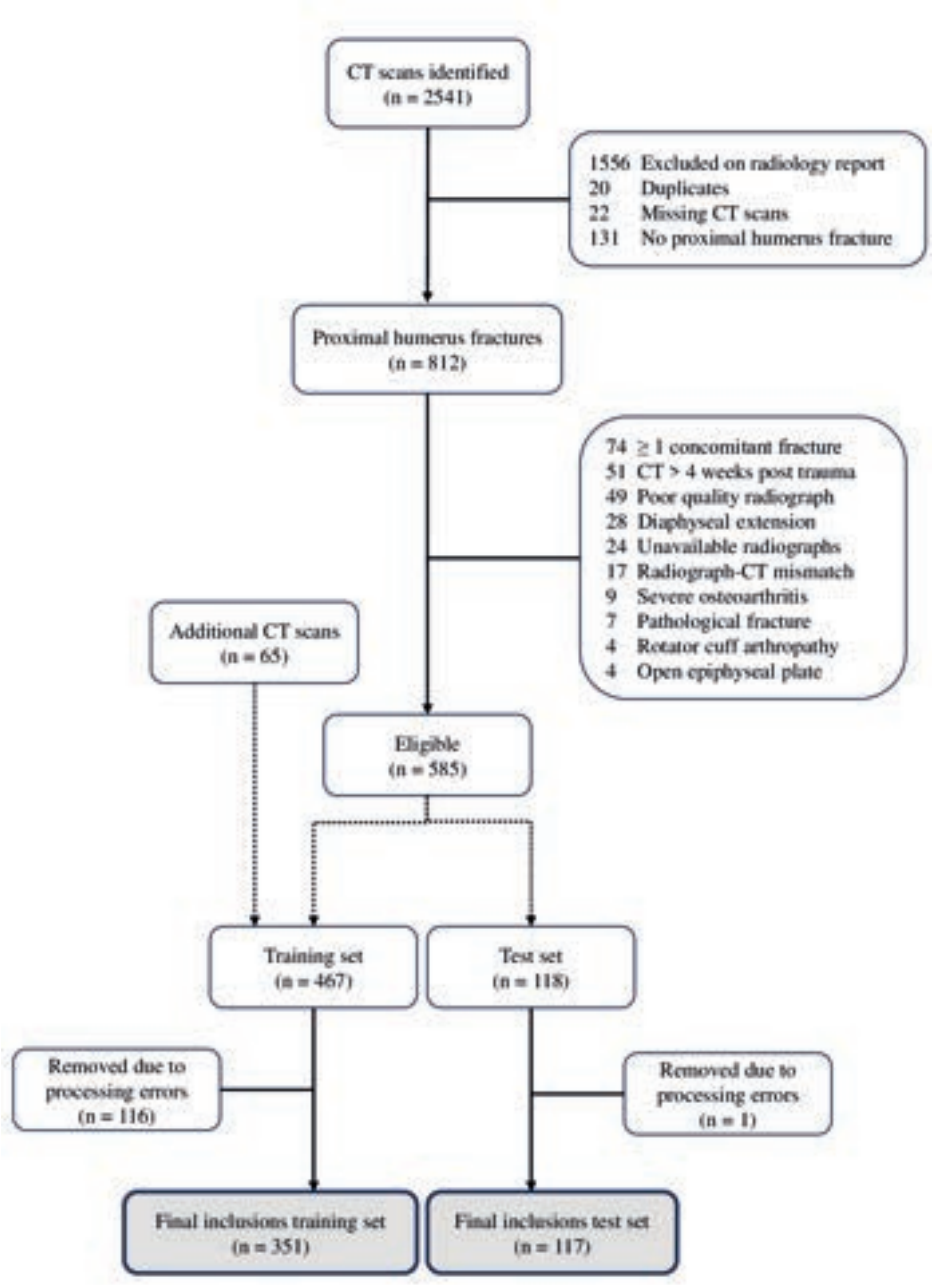
CNNs can proficiently rule out PHFs on radiographs. While it can adequately distinguish multi-part fractures from healthy or minimally displaced fractures, the classification metrics are not yet ready for clinical implementation despite rigorous methodology with multi-rater consensus on advanced CT imaging as the reference standard. One could argue that AI driven identification of pathoanatomy on plain radiographs is on par with human observation for simple tasks such as PHF recognition, but diagnostic performance of machine learning decreases significantly when tasks become increasingly complex.

REFERENCES

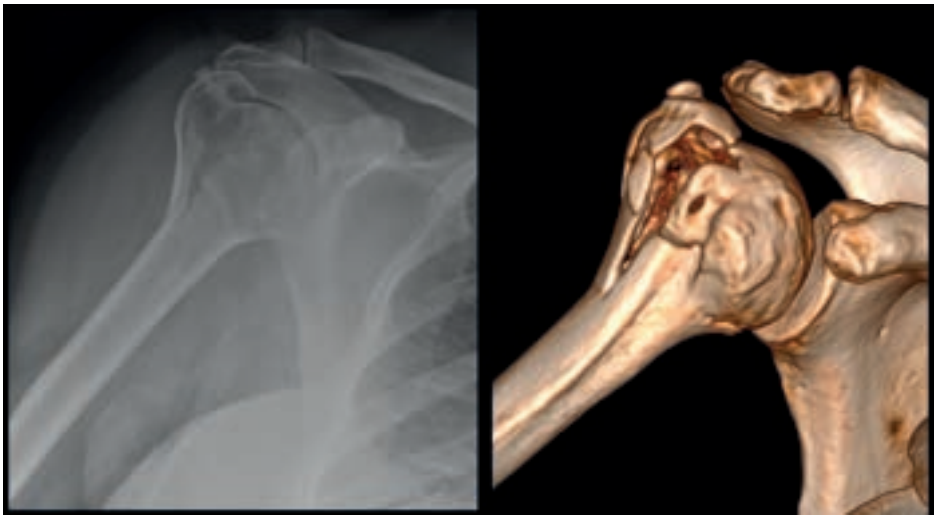
1. Rangan A, Handoll H, Brealey S, et al. Surgical vs nonsurgical treatment of adults with displaced fractures of the proximal humerus: the PROFHER randomized clinical trial. *JAMA*. 2015;313(10):1037-1047.
2. Bruinsma WE, Guitton TG, Warner JJP, Ring D. Interobserver reliability of classification and characterization of proximal humeral fractures. *J Bone Joint Surg Am*. 2013;95(17):1600-1604.
3. 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-S24.
4. Spek RWA, Schoolmeesters BJA, Oosterhoff JHF, et al. 3D-printed Handheld Models Do Not Improve Recognition of Specific Characteristics and Patterns of Three-part and Four-part Proximal Humerus Fractures. *Clin Orthop Relat Res*. 2022 1;480(1):150-159.
5. Langerhuizen DWG, Janssen SJ, Mallee WH, et al. What Are the Applications and Limitations of Artificial Intelligence for Fracture Detection and Classification in Orthopaedic Trauma Imaging? A Systematic Review. *Clin Orthop Relat Res*. 2019;477(11):2482-2491.
6. Langerhuizen DWG, Bulstra AEJ, Janssen SJ, et al. Is Deep Learning On Par with Human Observers for Detection of Radiographically Visible and Occult Fractures of the Scaphoid? *Clin Orthop Relat Res*. 2020;478(11):2653-2659.
7. Prijs J, Liao Z, To M-S, et al. Development and external validation of automated detection, classification, and localization of ankle fractures: inside the black box of a convolutional neural network (CNN). *Eur J trauma Emerg Surg*. 2023;49(2):1057-1069.
8. Groot OQ, Bongers MER, Ogink PT, et al. Does Artificial Intelligence Outperform Natural Intelligence in Interpreting Musculoskeletal Radiological Studies? A Systematic Review. *Clin Orthop Relat Res*. 2020;478(12):2751-2764.
9. Chung SW, Han SS, Lee JW, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop*. 2018;89(4):468-473.
10. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26(9):1364-1374.
11. Hamada K, Fukuda H, Mikasa M, Kobayashi Y. Roentgenographic findings in massive rotator cuff tears. A long-term observation. *Clin Orthop Relat Res*. 1990:(254):92-6.
12. Samilson RL, Prieto V. Dislocation arthropathy of the shoulder. *J Bone Joint Surg Am*. 1983;65(4):456-460.
13. Medixant. RadiAnt DICOM Viewer. Available at: https://www.radiantviewer.com.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
15. Neer CS 2nd. Displaced proximal humeral fractures: part I. Classification and evaluation. 1970. *Clin Orthop Relat Res*. 2006;442:77-82.
16. Sumrein BO, Mattila VM, Lepola V, et al. 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(10):1756-1761.
17. Matsumura N, Furuhata R, Seto T, et al. Reproducibility of the modified Neer classification defining displacement with respect to the humeral head fragment for proximal humeral fractures. *J Orthop Surg Res*. 2020;15(1).

18. Hasan AP, Phadnis J, Jaarsma RL, Bain GI. Fracture line morphology of complex proximal humeral fractures. *J Shoulder Elbow Surg*. 2017;26(10):e300-e308.
19. Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y. & Girshick R. Detectron2. Available at: https://github.com/facebookresearch/detectron2.
20. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition.* 2016:770-778.
21. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*. 2010:248-255.
22. Meijer DT, de Muinck Keizer R-JO, Doornberg JN, et al. Diagnostic Accuracy of 2-Dimensional Computed Tomography for Articular Involvement and Fracture Pattern of Posterior Malleolar Fractures. *Foot ankle Int*. 2016;37(1):75-82.
23. Guitton TG, Ring D. Interobserver reliability of radial head fracture classification: two-dimensional compared with three-dimensional CT. *J Bone Joint Surg Am*. 2011;93(21):2015-2021.
24. Doornberg JN, Rademakers MV, van den Bekerom MPJ, et al. Two-dimensional and three-dimensional computed tomography for the classification and characterisation of tibial plateau fractures. *Injury*. 2011;42(12):1416-1425.
25. Lopiz Y, Alcobía-Díaz B, Galán-Olleros M, García-Fernández C, Picado AL, Marco F. Reverse shoulder arthroplasty versus nonoperative treatment for 3- or 4-part proximal humeral fractures in elderly patients: a prospective randomized controlled trial. *J Shoulder Elbow Surg*. 2019;28(12):2259-2271.
26. Oliveira E Carmo L, van den Merkhof A, Olczak J, et al. An increasing number of convolutional neural networks for fracture recognition and classification in orthopaedics : are these externally validated and ready for clinical application? *Bone Jt Open*. 2021;2(10):879-885.
27. Health TLD. A digital (r)evolution: introducing The Lancet Digital Health. *Lancet Digit Health*. 2019;1(1):e1.
28. Krogue JD, Cheng K V., Hwang KM, et al. Automatic hip fracture identification and functional subclassification with deep learning. *Radiol Artif Intell*. 2020;2(2):e190023.
29. Gan K, Xu D, Lin Y, et al. Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. *Acta Orthop*. 2019;90(4):394-400.
30. Zech JR, Jaramillo D, Altosaar J, Popkin CA, Wong TT. Artificial intelligence to identify fractures on pediatric and young adult upper extremity radiographs. *Pediatr Radiol*. 2023;53(12):2386-2397.
31. Bossuyt PMM, Reitsma JB, Linnet K, Moons KGM. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clin Chem*. 2012;58(12):1636-1643.

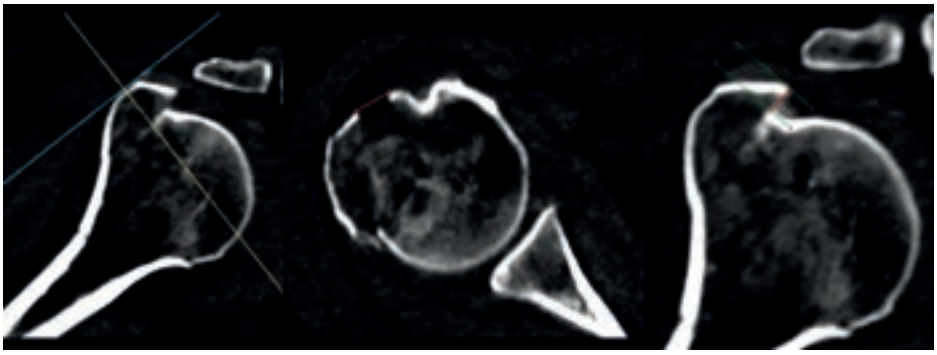
SUPPLEMENTARY MATERIAL



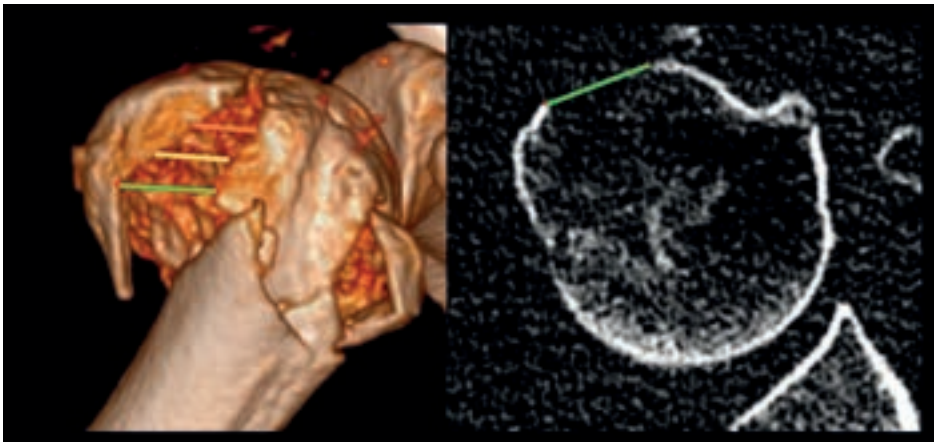
Supplement 1. Breakdown of patients selected for algorithm training and internal validation.



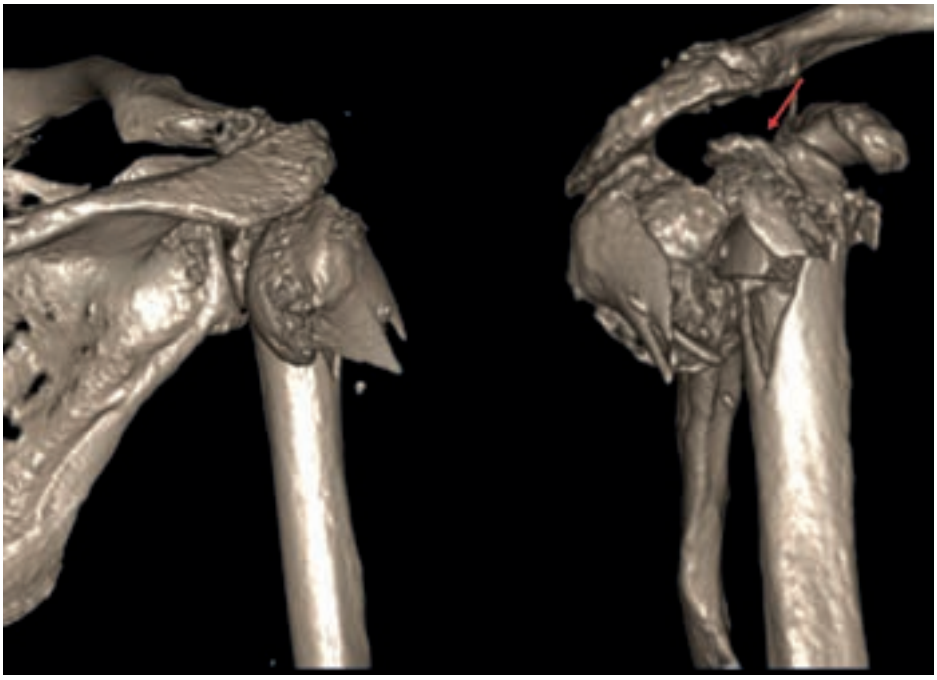
Supplement 2. Radiograph of a greater tuberosity fracture. Displacement was defined as displacement more than 10 mm in relation to the lesser tuberosity or the humeral head. The maximum displacement was measured.



Supplement 3. Left image: multiplanar reconstruction was used to align with the greater tuberosity. Middle image: on axial plane the maximum posterolateral displacement was 8.8 mm (red line). Right image: superior displacement was determined by measuring the distance between the two perpendicular lines (red line here). This was 5.1 mm. The total displacement equalled: (8.8)2 + (5.1)2 = (103.5)2 = 10.2 mm. Hence this fracture was classified as a two-part fracture.

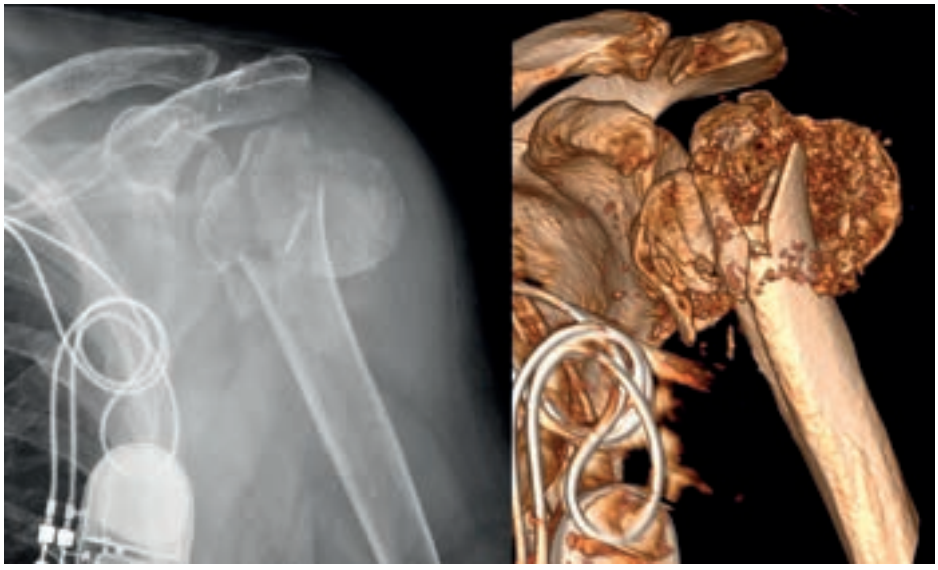


Supplement 4. This image highlights that greater tuberosity fractures can be measured on different levels which all result in different measurements. We chose to measure the maximum displacement on the greater tuberosity (green line).

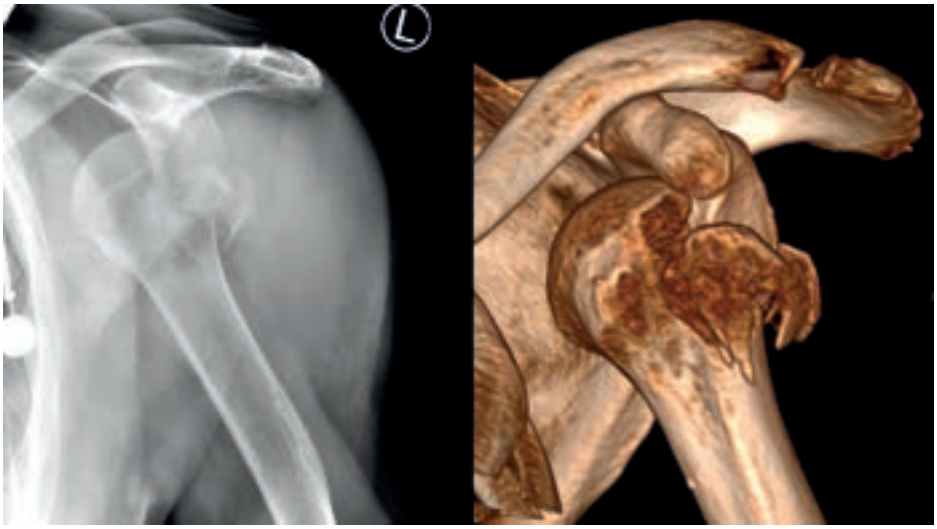


Supplement 5. This image highlights that greater tuberosity fractures were judged anatomically. The anterior facet (red arrow) is separated from the posterior and middle facet. As these parts were more than 1 cm displaced in relation to each other, this tuberosity was classified as being displaced.

Supplement 6. A severely valgus impacted fracture. These types were not classified as a glenohumeral dislocation.



Supplement 7. Dislocated fracture with severe posterior humeral rotation. This fracture was classified as a glenohumeral dislocation.



Supplement 8. Image of a “typical” glenohumeral (anterior) dislocation.


DARWIN
CAIRNS

5

Identification of proximal humerus fracture characteristics on plain radiographs: do convolutional neural networks still outperform humans when the task becomes increasingly more complex?

Reinier W.A. Spek
William J. Smith
Marat Sverdlov
Sebastiaan Broos
Yang Zhao
Zhibin Liao
Johan W. Verjans
Jasper Prijs
Minh-Son To
Henrik Åberg
Frank F.A. IJpma
Wael Chiri
Bhavin Jadav
John White
Gregory I. Bain
Paul C. Jutte
Michel P.J. van den Bekerom
Ruurd L. Jaarsma
Job N. Doornberg

The Machine Learning Consortium
The Bone & Joint Journal. 2024 Nov;106-B(11):1348-1360

ABSTRACT

Aims

Artificial intelligence driven computer vision models to characterise proximal humerus fracture (PHFs) may be more reproducible than human fracture evaluation. The purpose of this study was to develop a convolutional neural network (CNN) to identify; 1) greater tuberosity displacement ≥1cm; 2) neck-shaft angle ≤100˚; 3) shaft translation; and 4) articular fracture involvement; on plain radiographs.

Methods

The CNN was trained using radiographs from eleven hospitals in Australia and externally validated on radiographs from the Netherlands. Each radiograph was paired with corresponding computed tomography (CT) scans to serve as the reference standard based on multi-rater consensus (dual independent evaluation by trained researchers and attending orthopaedic surgeons). The following four fracture characteristics were determined on two-dimensional (2D) and three-dimensional (3D) CT scans and subsequently allocated to each series of radiographs: 1) greater tuberosity displacement ≥1cm; 2) neck-shaft angle ≤100˚; 3) shaft translation (0% to <75%, 75% - 95%, >95%); and 4) the extent of articular involvement (0% to <15%, 15% - 35%, or >35%).

Results

The dataset comprised 562 radiographs for training, 235 for internal validation, and 223 for external validation -including 1020 corresponding 2D and 3D CTs characterized with multi-rater consensus- representing 281, 103, and 105 patients, respectively. Accuracy to detect greater tuberosity fracture displacement ≥1cm was 35.0% (AUC = 0.57). The CNN did not recognize neck-shaft angles ≤100˚ (AUC = 0.42), nor fractures with ≥75% shaft translation (AUC = 0.51 - 0.53), or with ≥15% articular involvement (AUC = 0.48 - 0.49). The model's performance on the external dataset showed similar poor accuracy levels.

Conclusion

Despite rigorous training methodology based on CT imaging with multiple-observer consensus agreement to serve as the reference standard; our developed CNN exhibited poor diagnostic ability to detect greater tuberosity displacement ≥1cm and failed to identify neck-shaft angles ≤100˚, shaft translations or articular fractures on plain radiographs.

INTRODUCTION

In addition to patient- and surgeon factors, specific fracture characteristics may guide the need for surgical repair and type of surgery (e.g., intramedullary nail *versus* plate *versus* joint replacement) in proximal humerus fractures (PHFs) [1,2]. Despite decades of research on commonly used classification systems of Neer [3], Hertel [4], and AO [5], their clinical relevance has been extensively debated due to their low reproducibility and surgeons' experiences that classification does not guide surgical decision-making [6–8]. Some argue therefore, to use these classification systems for documentation and research purposes only, and move forward with specific characteristics to describe fracture patterns that have more clinical relevance in decision-making [9].

Deep learning in healthcare gained significant interest, as it holds the promise to improve productivity, augment surgical decision-making (data-driven evidence based) and the ability predict patient-specific outcomes [10–12]. Also in orthopaedic trauma, deep learning algorithms yield promising results [13,14]. Not merely on their ability to process more complex imaging modalities (CT or MRI) but also on model integration (multiple algorithms are combined into one neural network to identify different fractures at one body site) and their ability to reduce human interobserver bias for fracture detection, classification, and characterisation tasks [15–17]. For instance, human accuracy in detecting hip fractures ranges from 78 - 94%, but if they are aided by deep learning algorithms this improves to 91 - 97% [18,19]. To date, no machine learning algorithms have been developed to recognize characteristics specific to PHFs. Hence, we identified the four most important fracture hallmarks for surgical decision-making, (re-)defined them and undertook this study.

The purpose was to develop a convolutional neural network (CNN) that can assess on plain radiographs; greater tuberosity displacement ≥1cm, neck-shaft angle ≤100˚, shaft translation, and the extent of articular involvement. Multi-rater consensus agreement based on advanced Computer Tomography (CT) imaging was used as the reference standard for CNN training. We asked: Do CNNs still outperform humans when the task becomes increasingly more complex?

PATIENTS AND METHODS

Our Institutional Review Board waived requirement for approval of this diagnostic imaging study, in accordance to the Declaration of Helsinki.

Setting and study design

This study was carried out in Adelaide, Australia at a Level 1 trauma centre in collaboration with a dedicated machine learning institute (Australian Institute for Machine Learning) (Fig. 1). This work was written according to CONSORT-AI and CAIR checklists [20,21].

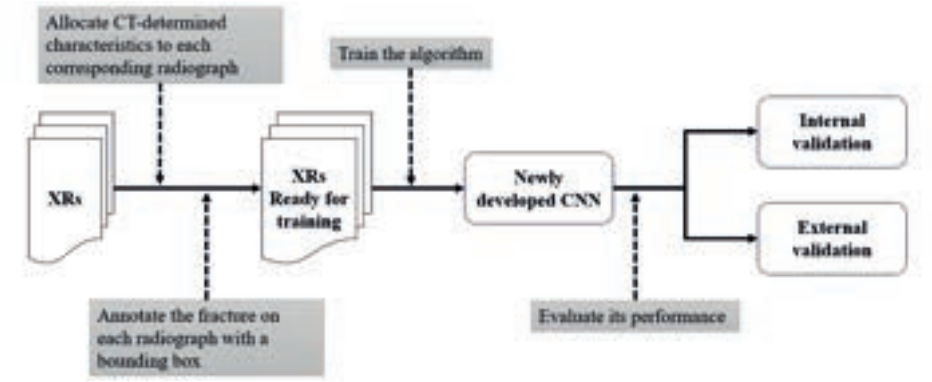


Figure 1. Flow diagram which summarizes the steps undertaken for the development of our CNN.

Population

The following inclusion criteria were used: (1) CT scans must be available for each patient who sustained a fracture (obtained <4 weeks after initial presentation at the emergency department), (2) radiographs had to reveal a PHF from patients aged 18 years or older, (3) availability of at least one standard or glenohumeral anteroposterior (AP) view. Radiographs with one of the following criteria were excluded: glenohumeral dislocation, metadiaphyseal fracture extension, humeri with orthotics, glenohumeral joint osteoarthritis (stage 3 and 4 according to Samilson and Prieto [22]), open epiphyseal plates, or inconclusive evidence of a fracture or pathology. Poor quality radiographs were also excluded as well as those with discrepancies between the amount of radiographic and CT displacement (e.g., if the shaft was substantially more displaced on the radiograph than on the CT scan, or vice versa. Although rare, this may have occurred if both imaging modalities were taken at different time points).

Data collection and screening

For the training and internal validation dataset, the South Australian Medical Imaging database was used which contains radiographic imaging modalities from eleven hospitals across South Australia. Shoulder CT scans together with the corresponding radiographs were collected. If the radiology report revealed a PHF, the CT scan was assessed against the exclusion criteria by two researchers [23]. The radiographs performed closest to the date of the CT scan were used. Out of the 2541 CT scans, 384 consecutive patients were selected (Supplement 1). For external validation, a Dutch database was utilized of which 105 patients from two hospitals (one Level 1 and one Level 2 trauma centre) were included. Periods of inclusion were January 2016 to June 2021 for the Level 1 trauma centre and March 2014 to September 2020 for the Level 2 trauma centre.

Definitions of fracture characteristics

The following four fracture characteristics were allocated to each radiograph (Fig. 2 - 6, Supplement 2) [1,24,25].

1. Greater tuberosity displacement ≥1cm. The greater tuberosity was anatomically defined as the bony prominence posterior to the bicipital groove, with the base demarcated by the line running from the medial hinge to the most lateral prominence of the humerus (Figure 2 in Hasan et al. [26]). The following rules were applied for assessment: (a) displacement was judged in relation to the articular fragment and/or the lesser tuberosity so *not* in relation to the humeral shaft displacement [27], (b) fractures were considered from an anatomical point of view (Supplement 3), (c) axial and coronal images representing the largest displacement were used for measurements (Supplement 4), (d) displacement was calculated by the following formula: (maximum displacement axial plane)^2 + (maximum displacement coronal plan)^2 = (total greater tuberosity displacement)^2 (Fig. 7) [28].

2. Neck-shaft angle (NSA) ≤100˚. Using 3D CT virtual models, we measured the angle between the line parallel to the humeral diaphysis and the line perpendicular to the anatomic neck (Fig. 8). NSAs were measured by two researchers. Mean NSAs were used for algorithm training- and validation. NSA reproducibility was determined on 40 randomly selected cases: the ICC was 0.83 (95% CI: 0.58 - 0.93) with a mean difference between observers of 5.8˚ (range: 0.7˚ - 19.3˚).

3. Shaft translation. Shaft translation was defined according to the remaining contact between the surgical neck of the humerus (at the level of the fracture) in relation to the articular fragment including the tuberosities (independent of displacement) [24]. Less bony contact between the head and shaft, has a lower percentage of contact, and a higher degree of translation. The following categories were distinguished: 0% to <75%, 75% - 95%, or >95%.

4. Extent of articular involvement. Articular involvement was defined as the percentage of the fracture extending across the articular cartilage of the humeral head and was subdivided into three categories: 0% to <15%, 15% - 35%, or >35%. Two assessors measured articular involvement and their mean determined which category the fracture was grouped in (Fig. 9). Intraclass correlation was determined on 20 randomly selected CT scans with two evaluators. The reproducibility was good (ICC = 0.83, 95%: 0.62 - 0.92) with a mean difference of 6.2% (range: 0 - 18%).

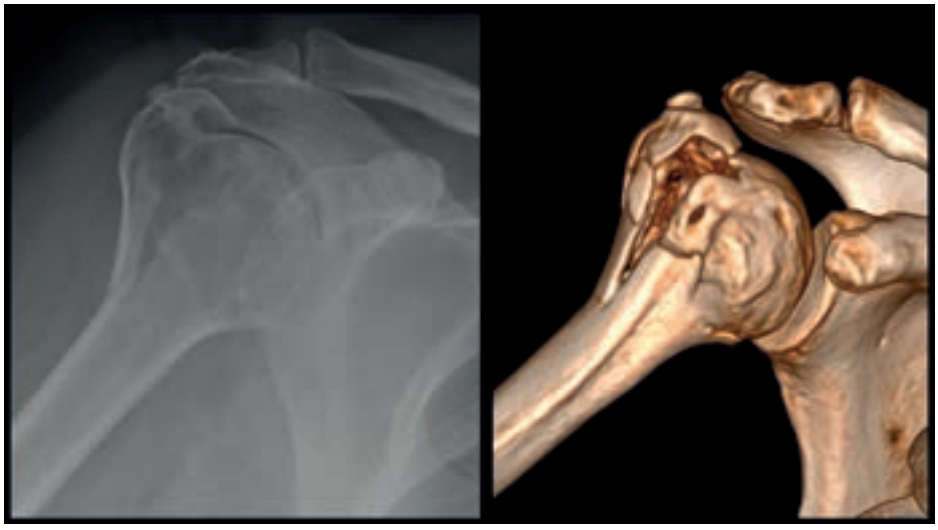


Figure 2. Greater tuberosity displacement ≥1cm.

Figure 3. Neck-shaft angle ≤100˚. The angle, as determined by two assessors, mean 89.5˚.

Figure 4. Three subcategories of shaft translation. A = 0 to <75%; B = 75% - 95%; C = >95% displacement.

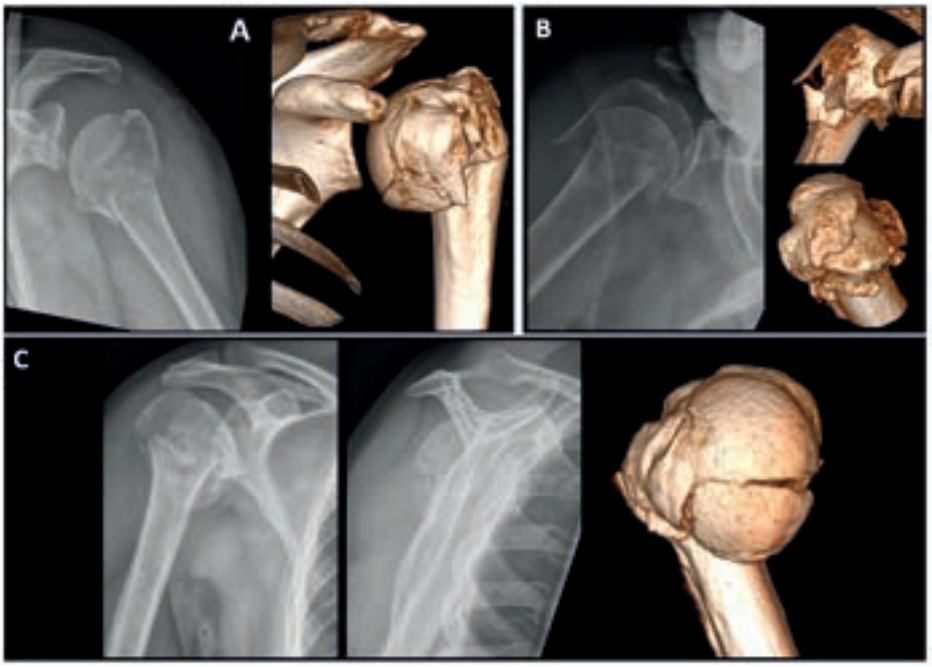


Figure 5. Three subcategories of articular involvement: A = 0% to <15% (the percentage, as determined by two evaluators, yielded an average of 14.5%, B = 15 - 35% (averaged at 23.4%), C = >35% (mean: 44.2%).

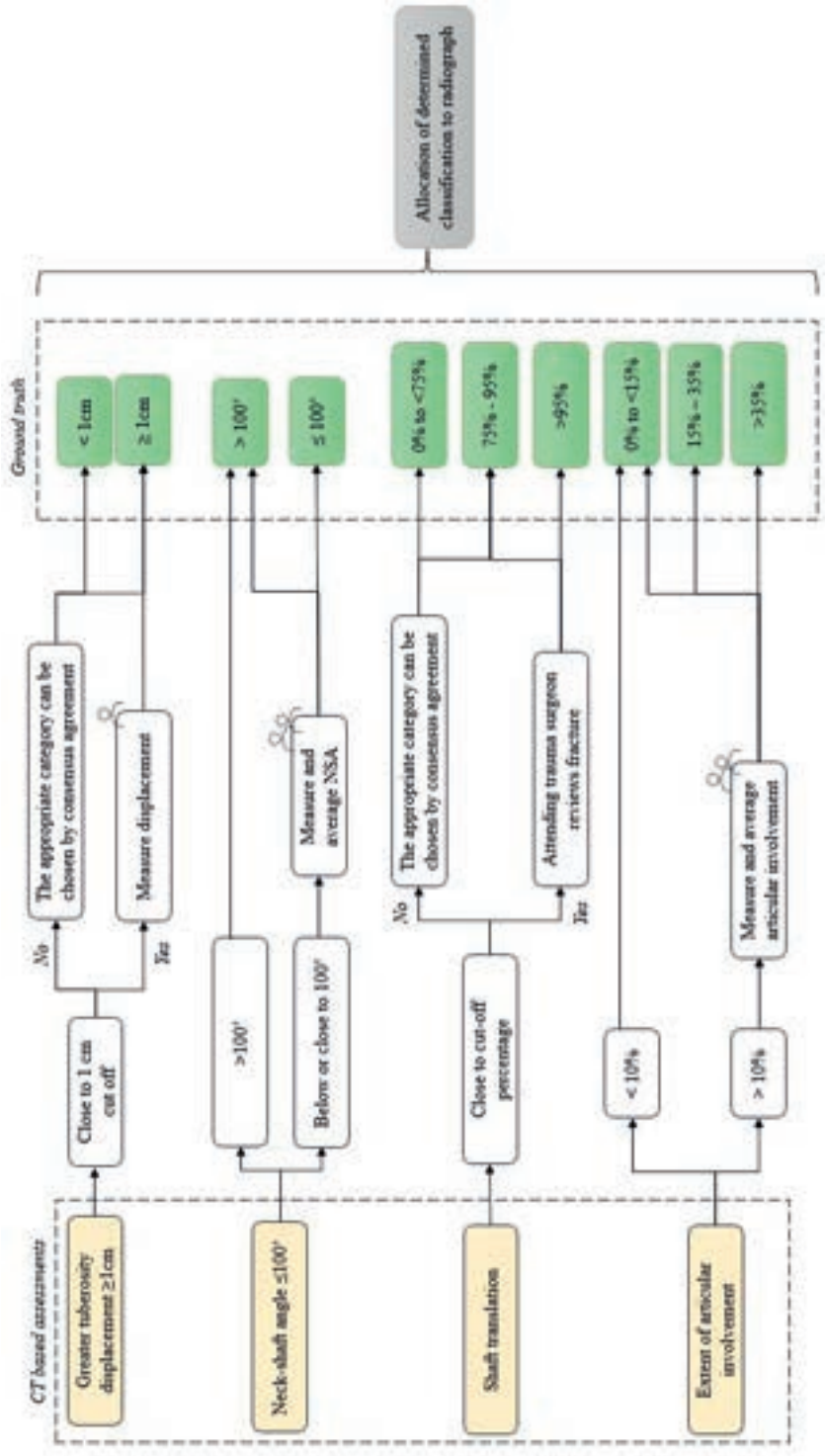


Figure 6. Data labelling: establishment of the ground truth. The number of participants measuring the fracture is represented with the person icon.

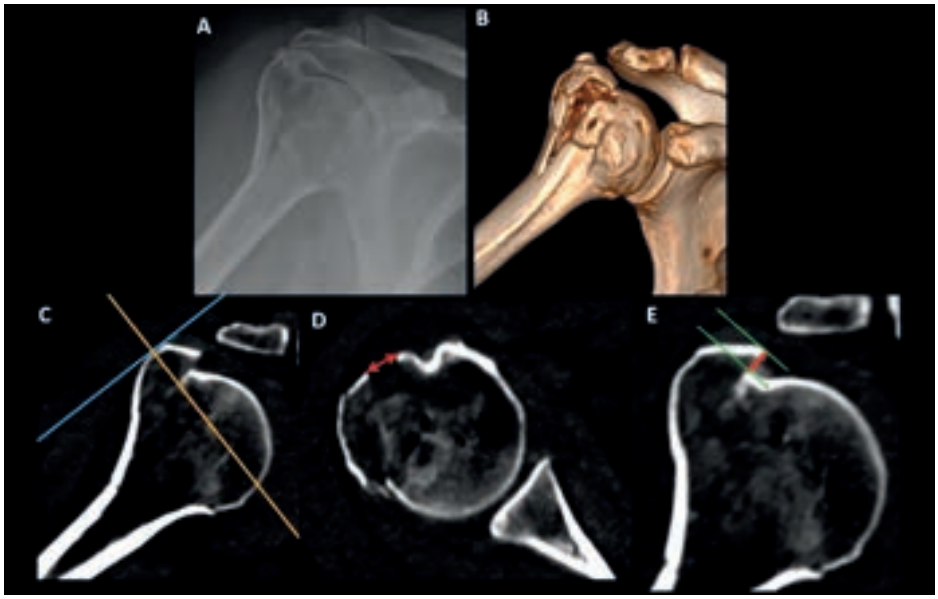


Figure 7. Measurement of displacement of the greater tuberosity (A and B). First, multiplanar reconstruction was used to align the CT scan with the greater tuberosity so that the coronal plane was parallel to the lateral border of the greater tuberosity (C). Second, the maximum posterolateral displacement on the axial plane was measured: 8.8 mm (red line, D). Third, superior displacement was measured on the coronal view: 5.1 mm (distance between the two perpendicular lines, E). The total displacement equalled: $(8.8)^2 + (5.1)^2 = (10.2)^2$. Therefore, the total displacement was 10.2mm, so this patient was scored as a fracture with ≥1cm greater tuberosity displacement.

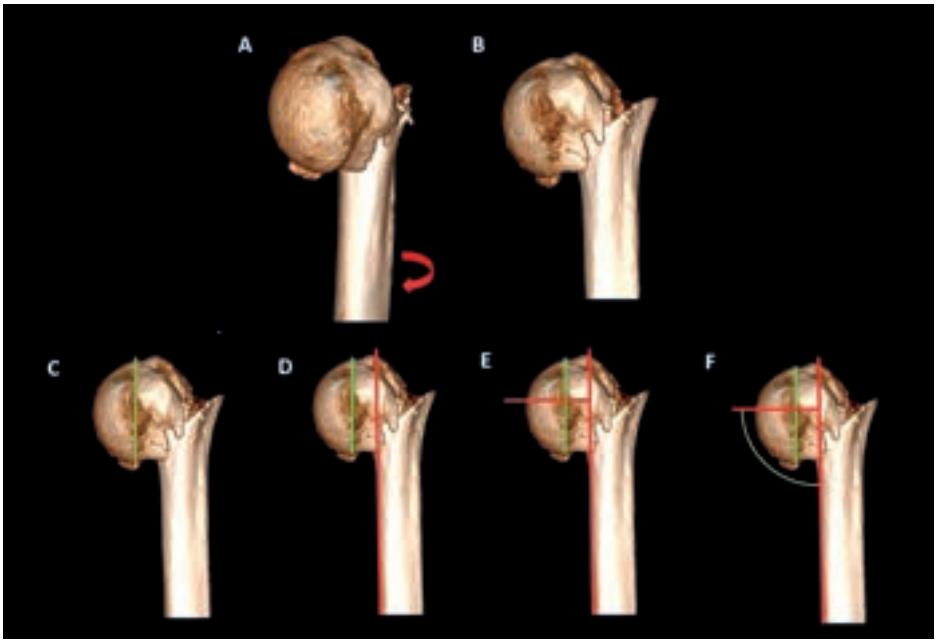


Figure 8. Measurement of neck-shaft angle. The 3D reconstructed models, with all structures other than the proximal humerus removed, were freely rotated in space until approximately 50% of the articular segment was visible (A and B). To determine the neck-shaft angle a line was drawn along the anatomic neck (C), humeral diaphysis (D) and a line perpendicular to the anatomic neck (E). The neck-shaft angle (F) was then calculated by measuring the angle between line.

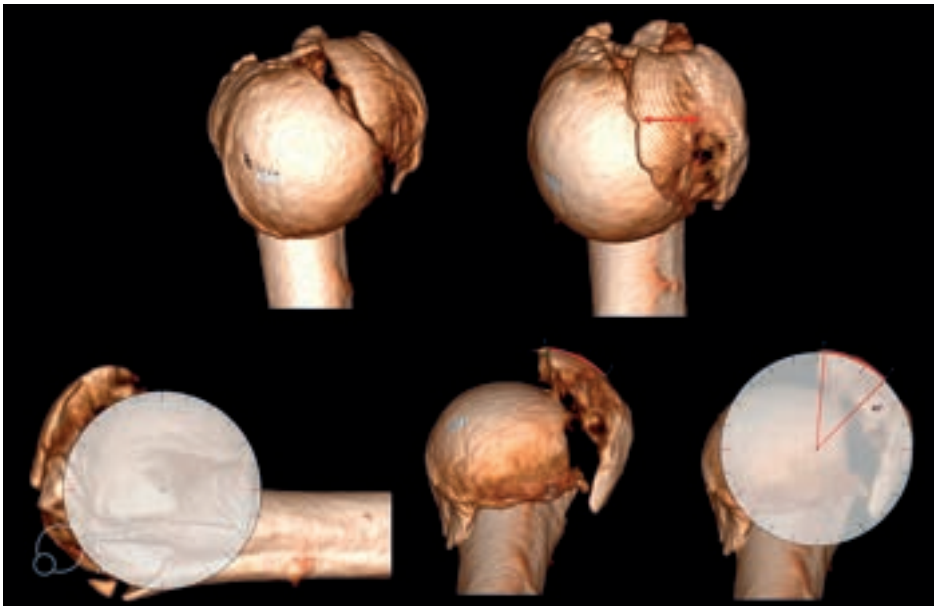


Figure 9. Measurement of articular fractures: A = identify the articular fracture; B = determine the location with the largest fracture extension across the head; C = position a circle that best fits and overlaps with the humeral head. ; D = identify the fracture plane perpendicular to the view in step B by rotating the 3D model in space; E = Copy the circle created in step C onto the fracture to measure its angle. The measured fracture angle was then divided by 145° (population mean of the articular arc derived from a cohort study with non-pathological shoulders) and multiplied with 100%. In this example, the fracture angle was 40°. Hence, the percentage of articular extension across the head equalled 27.6% ((40°/ 145°) *100%).

Ground truth

RadiAnt DICOM viewer (Medixant, Poznan, Poland) was used for fracture assessments and was performed by two or more assessors. Radiographs were used for algorithm training and evaluation, but the characteristics were always derived from the corresponding CT scans [23]. Consensus was obtained during in-person meetings. As many as 51%; 194 out of 384) was assessed by a trauma or upper limb surgeon. The following levels of assessment were utilized to determine the characteristics (years of experience were counted after completion of surgical training) (Table 1):

a) Medical researcher assessment: independent assessment was performed by two trained researchers (from R.S., J.S., M.S., or S.M.).
b) Single surgeon assessment: independent assessment by an upper limb fellow (W.C.), an attending orthopaedic surgeon (>5 years of experience) (H.A.), or an academically oriented upper limb surgeon (>10 years of experience) (M.vdB.).
c) Dual surgeon assessment: assessment by two attending trauma or upper limb surgeons (≥15 years of experience) (G.B./ J.W., or R.J. / B.J.).

For all assessments results were discussed between reviewers, to form a consensus. Any continuing discrepancies were be resolved by discussion with the first author (R.S.) and attending clinician (R.J.). For external validation, PHFs were reviewed by a an academically oriented upper limb surgeon on key CT images only. Answers were compared to the medical researcher's assessment: absolute agreement on greater tuberosity displacement ≥1cm was 100%, categories of shaft translation 96.0% (κ = 0.87, 95% CI: 0.63 - 1.12) (n = 23). No articular fractures or fractures with NSAs ≤100˚ were missed.

Table 1. Demographics of the three datasets

	Training (n = 281)	Internal validation (n = 103)	External validation (n = 105)
Age (years)	65.4 ± 13.9	64.9 ± 15.9	59.1 ± 14.9
Gender			
Male	70 (24.9%)	30 (29.1%)	35 (33.3%)
Female	210 (74.7%)	73 (70.9%)	70 (66.7%)
Side			
Left	154 (54.8%)	45 (43.7%)	57 (54.3%)
Right	127 (45.2%)	58 (56.3%)	48 (45.7%)
Greater tuberosity displacement ≥1cm			
Yes	164 (58.4%)	59 (57.3%)	58 (55.2%)
No	117 (41.6%)	44 (42.7%)	47 (44.8%)
Neck-shaft angle ≤100˚			
Yes	89 (31.7%)	32 (31.1%)	23 (21.9%)
No	192 (68.3%)	71 (68.9%)	82 (78.1%)

Table 1. Demographics of the three datasets (continued)

	Training (n = 281)	Internal validation (n = 103)	External validation (n = 105)
Shaft translation			
0% to <75%	235 (83.6%)	87 (84.5%)	83 (79%)
75% - 95%	17 (6%)	5 (4.9%)	9 (8.6%)
>95%	29 (10.3%)	11 (10.7%)	13 (12.4%)
Extent of articular involvement			
0% to <15%	230 (81.9%)	87 (84.5%)	96 (91.4%)
15% - 35%	45 (16%)	13 (12.6%)	8 (7.6%)
>35%	6 (2.1%)	3 (2.9%)	1 (1%)
Level of assessment			
Medical researcher	190 (67.6%)	-	82 (78.1%)
Single surgeon*	91 (32.4%)	-	23 (21.9%)
Dual surgeon	-	103 (100%)	-

* In the training set, 21 fractures were assessed for greater tuberosity displacement only.

Annotations

All included radiographs were annotated with a bounding box to guide the algorithm to the fracture. During this process, the corresponding CT scan was reviewed to ensure the fracture boundaries were adequately delineated. Annotations included all fracture segments (irrespective of displacement), lines and comminution, and were performed in Labelbox (Labelbox, San Francisco, United States of America) (Fig. 10) (Supplement 5) [29]. Annotations were completed by the first author (R.S.). The training- and external validation images were reviewed by a second researcher (J.S., M.S., or S.M.) and the internal validation images by an attending trauma surgeon (R.J., M.vdB., or J.D.).

Algorithm development

A pretrained version of the ResNet-152 model was used and adjusted to accommodate for two binary and two multiclass tasks, and was trained on the datasets in the cloud environment of Google's Colab Pro [30–32]. To avoid overloading the GPU, each image was pre-processed by resizing it to 512x512 pixels. The aspect ratio was preserved by adding the required padding to the images before resizing.

During training, a variety of image augmentations (i.e. shifting, scaling and random rotation) were used and the model was trained for 100 epochs [31]. To account for any variations in the outcomes due to entrapment in local minima, the full training procedure was performed three times, and the best AUC was chosen as the result.

Outcome measures

The following diagnostic outcome metrics were evaluated: accuracy, area under the receiving operating characteristic curve (AUC), sensitivity, and specificity.

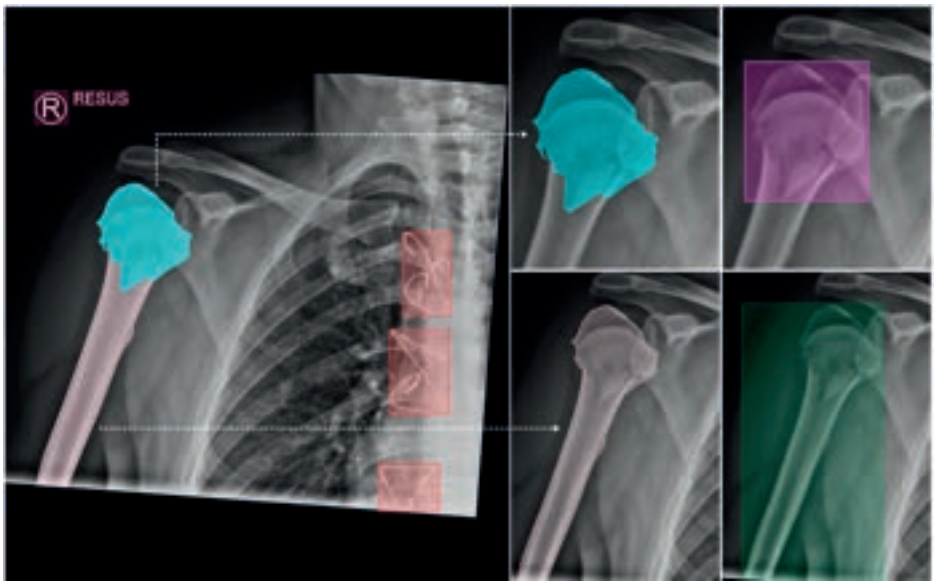


Figure 10. Fracture annotation. A bounding box was drawn around the fracture as well as the humerus. Throughout the training and evaluation phase, the CNN can utilize this box to focus on the area of interest to optimize its performance.

Statistical analysis

Microsoft Excel version 2021 (Microsoft, Redmond, United States of America) was used for calculating baseline demographics. Intraclass correlation coefficients (ICC) were calculated with a reliability analysis performed in a two-way mixed model with absolute agreement. Cohen's kappa values with corresponding 95% confidence intervals were calculated to determine the agreement on fracture characteristics.

RESULTS

Baseline demographics training set

The algorithm was trained with 562 radiographic views from 281 patients (mean age: 65.4 ± 13.9, 74.7% females), with 562 corresponding 2D and 3D CT scans with multi-rater consensus agreement as the reference standard for CNN training. This group entailed 68.3% ≥1cm displaced greater tuberosity fractures, 31.7% with NSAs ≤100˚, 6% with 75% - 95% shaft translation, 10.3% with >95% shaft translation, 16% articular fractures with 15% - 35% involvement, and 2.1% with >35% articular involvement (Table 1).

Internal validation

Internal validation comprised 235 radiographs (n = 103) (mean age: 64.9 ± 15.9, 70.9% females) (Table 1). Accuracy to detect a displaced greater tuberosity fracture ≥1cm was 35.0% (AUC = 0.57). For neck-shaft angles ≤100˚ this was 68.9% (AUC = 0.42), 84.5% for distinguishing the three grades of shaft translation (AUC 0% to <75% = 0.54, AUC 75% - 95% = 0.53, AUC >95% = 0.51), and 84.5% for categorizing fractures into the percentage of articular involvement (AUC 0% to <15% = 0.51, AUC 15% - 35% = 0.49, AUC >35% = 0.48) (Table 2). Only 8 out of 59 ≥1cm displaced greater tuberosity fractures were classified correctly as a displaced fracture and the CNN was unable to detect a neck-shaft angle ≤100˚, shaft translations with ≥75% displacement, or fractures with ≥15% articular involvement. In fact, the CNN did not classify a single fracture into any of these groups (Table 3).

External validation

External validation was conducted on 223 Dutch radiographs (n = 105) (mean age: 59.1 ± 14.9, 66.7% females) (Table 1). Detection of ≥1cm displaced greater tuberosity fractures revealed 48.6% accuracy, but a lower AUC (0.45) compared to internal validation (Table 2). The ability for the CNN to recognize greater tuberosity displacement was again poor: merely 5 out of 58 fractures were adequately categorized as displaced ≥1cm. The CNN could not identify the other three fracture characteristics (Table 3).

DISCUSSION

Identifying specific fracture hallmarks may be helpful in clinical practice for surgical decision-making, but assessment is hampered by observer variability [24,33]. Therefore, this study aimed to provide an objective AI-driven tool to aid in uniform accurate assessment for all patients presenting at the emergency department with a PHF [9,34]. Such an algorithm could be used in the emergency department to aid in selecting patients for surgical treatment and to ascertain that these hallmarks are recognized by all doctors, regardless of seniority or specialty. The CNN however, showed insufficient accuracy to detect ≥1cm displaced greater tuberosity fractures, and was unable to recognize any of the other fracture characteristics.

This study should be interpreted in the light of strengths and weaknesses. This study was designed according to CONSORT-AI and CAIR checklists [20,21]. Additional merits are threefold. Firstly, its clinical significance is underscored by the selection of fracture characteristics, each with the potential to influence decision-making processes. Secondly, the ground truth is ensured by the collaborative efforts of an international expert panel who contributed to the fracture definitions, as well as their hands-on involvement in the assessment process, supported by CT scans. Lastly, the deployment of what is currently considered a cutting-edge algorithm, conceived and refined by an institution experienced in this domain. We expect that with technological advances, innovations, and extended training, future computer algorithms will be able to identify these radiological variables.

For adequate interpretation of this work, including nuances in definitions should be well understood: e.g., greater tuberosity fractures were judged in relation to the head and/or lesser tuberosity and *not* in relation to the humeral shaft. Another example is that NSAs were determined on 3D CT reconstructions at the plane where 50% of the head was visible. Moreover, the computer was trained and validated on plain radiographs, with variables that were defined by the 3D CT scan images to serve as the ground truth in the annotation process.

This study also requires acknowledgement of various limitations. Firstly, not every patient had all four views radiographic views (standard AP, glenohumeral AP, lateral, and axillary). This may have caused issues for the algorithm as certain fracture characteristics are better visible on certain radiographic views (e.g., glenohumeral

AP views may be more illustrative than standard AP views). Secondly, in clinical practice some fracture hallmarks (like fractures >35% articular involvement) are uncommon and were thus scarcely present in our dataset. This could have resulted in training difficulties, as in machine learning equal distribution of categories is known to be beneficial for optimal performance [35].

To date, many different algorithms have been developed but most focus on fracture recognition and/or classification but not on specific hallmarks which would alter surgical decision-making [36–39]. One could argue the clinical benefit of automated fracture detection, as human performance is already excellent, with only a very small "accuracy gap" between human and machine with very little clinical relevance [40]. The CNN from Chung et al., was trained to detect greater tuberosity fractures and reached an accuracy of 86% (AUC 0.98) with radiographs from 1376 patients but did not incorporate a displacement threshold like we did [31]. Another research group trained an algorithm to detect all separate AO/OTA subclassifications using 574 patients. The algorithm was able to identify A1.1 (two-part greater tuberosity fractures) with an AUC of 0.91 [41]. Noteworthy however, is that both studies did not incorporate CT scans as ground truth and that our algorithm was trained to identify greater tuberosity displacement in each fracture type, which added another layer of complexity. A 2013-study revealed poor interobserver agreement on greater tuberosity fractures and humeral head-splits of 0.30 (kappa value) (sample size: 15 CT scans, 61 observers) [8]. Although accuracies were not reported in this study, our algorithm is not superior to human observers and therefore –not yet– ready for clinical implementation. As the chosen fracture characteristics are also challenging for humans to characterize on plain radiographs, we advise to repeat this study with CT scans as the input for the algorithm instead of radiographs [17,42].

The body of evidence on artificial intelligence (AI) to assist clinicians in patient care is growing rapidly [10,36]. This surge is driven by its potential to reduce human decision-making biases, relieve workload, and decrease costs[43]. Over the years, we have started to better understand the limits and benefits of AI. Our paper adds to the scarce number of studies, to disclose limitations of such AI driven applications. It underscores that humans cannot teach everything to algorithms and emphasizes that complex tasks on simple imaging (plain radiographs instead of CT scans) is not feasible (yet). Given the aforementioned benefits of deep learning, we encourage

researchers to continue developing algorithms and train clinically relevant fundamentals aside from fracture recognition and classifications.

Conclusion

The CNN trained and evaluated in this study showed poor diagnostic accuracy to detect greater tuberosity displacement ≥1cm, neck-shaft angles ≤100˚, shaft translations or articular fractures on plain radiographs. This outcome highlights the intrinsic challenge of automating complex diagnostic tasks and its generalisability—those that are demanding for human experts are often even more so for CNNs. Moving forward, it would be prudent to conduct a follow-up study where the diagnostic capabilities are exclusively assessed on CT scans.

Clinical relevance

CNNs have been successfully developed for –debatably– easy tasks such as PHF detection and classification. However, one could argue that AI still underperforms for complex and clinically more relevant tasks for surgical decision-making, such as PHF characterisation that may be too complex to identify on plain radiographs for both humans and machines. Despite the promise of an unparalleled potential, our study illustrates that applications of AI are not unlimited. At this point, AI driven fracture characterisation is not reliable on plain radiographs and requires advanced CT imaging.

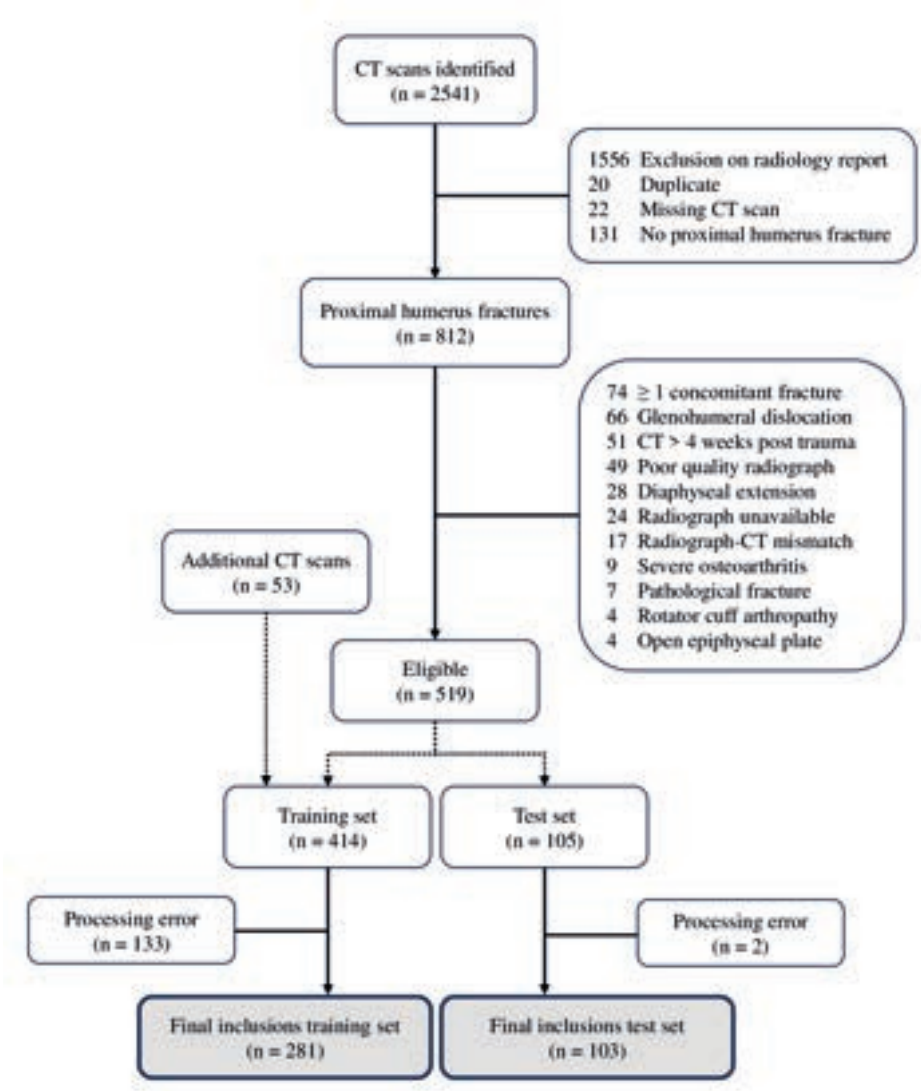
REFERENCES

1. Murray IR, Amin AK, White TO, Robinson CM. Proximal humeral fractures: Current concepts in classification, treatment and outcomes. *J Bone Joint Surg Br*. 2011;93 B(1):1-11.
2. Boileau P, d'Ollonne T, Bessière C, et al. Displaced humeral surgical neck fractures: classification and results of third-generation percutaneous intramedullary nailing. *J Shoulder Elbow Surg*. 2019;28(2):276-287.
3. Neer CS 2nd. Displaced proximal humeral fractures: part I. Classification and evaluation. 1970. *Clin Orthop Relat Res*. 2006;442:77-82.
4. Hertel R, Hempfing A, Stiehler M, Leunig M. Predictors of humeral head ischemia after intracapsular fracture of the proximal humerus. *J Shoulder Elbow Surg*. 2004;13(4):427-433.
5. Meinberg EG, Agel J, Roberts CS, Karam MD, Kellam JF. Fracture and Dislocation Classification Compendium-2018. *J Orthop Trauma*. 2018;32(1):S1-S170.
6. Chelli M, Gasbarro G, Lavoué V, et al. The reliability of the Neer classification for proximal humerus fractures: a survey of orthopedic shoulder surgeons. *JSES Int*. 2022;6(3):331-337.
7. Iordens GIT, Mahabier KC, Buisman FE, 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(5):596-602.
8. Bruinsma WE, Guitton TG, Warner JJP, Ring D. Interobserver reliability of classification and characterization of proximal humeral fractures. *J Bone Joint Surg Am*. 2013;95(17):1600-1604.
9. Spek RWA, Schoolmeesters BJA, Oosterhoff JHF, et al. 3D-printed Handheld Models Do Not Improve Recognition of Specific Characteristics and Patterns of Three-part and Four-part Proximal Humerus Fractures. *Clin Orthop Relat Res*. 2022;480(1):150-159.
10. Lång K, Josefsson V, Larsson A-M, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening ac. *Lancet Oncol*. 2023;24(8):936-944.
11. Yang S, Zhu F, Ling X, Liu Q, Zhao P. Intelligent Health Care: Applications of Deep Learning in Computational Medicine. *Front Genet*. 2021;12:1-21.
12. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J*. 2021;8(2):e188-e194.
13. Oliveira e Carmo L, van den Merkhof A, Olczak J, et al. An increasing number of convolutional neural networks for fracture recognition and classification in orthopaedics: are these externally validated and ready for clinical application? *Bone Jt Open*. 2021;2(10):879-885.
14. Prijs J, Liao Z, Ashkani-Esfahani S, et al. Artificial intelligence and computer vision in orthopaedic trauma: the why, what, and how. *Bone Joint J*. 2022;104-B(8):911-914.
15. Cheng C-T, Wang Y, Chen H-W, et al. A scalable physician-level deep learning algorithm detects universal trauma on pelvic radiographs. *Nat Commun*. 2021;12(1):1066.
16. Weikert T, Noordtzij LA, Bremerich J, et al. Assessment of a deep learning algorithm for the detection of rib fractures on whole-body trauma computed tomography. *Korean J Radiol*. 2020;21(7):891-899.
17. Dankelman LHM, Schilstra S, IJpma FFA, et al. Artificial intelligence fracture recognition on computed tomography: review of literature and recommendations. *Eur J trauma Emerg Surg*. 2023;49(2):681-691.

18. Cha Y, Kim J-T, Park C-H, Kim J-W, Lee SY, Yoo J-I. Artificial intelligence and machine learning on diagnosis and classification of hip fracture: systematic review. *J Orthop Surg Res.* 2022;17(1):520.
19. Health TLD. A digital (r)evolution: introducing The Lancet Digital Health. *Lancet Digit Health*. 2019;1(1):e1.
20. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26(9):1364-1374.
21. Olczak J, Pavlopoulos J, Prijs J, et al. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. *Acta Orthop*. 2021;92(5):513-525.
22. Samilson RL, Prieto V. Dislocation arthropathy of the shoulder. *J Bone Joint Surg Am*. 1983;65(4):456-460.
23. Medixant. RadiAnt DICOM Viewer. Available at: https://www.radiantviewer.com.
24. Mora Guix JM, Pedrós JS, Serrano AC. Updated classification system for proximal humeral fractures. *Clin Med Res*. 2009;7(1-2):32-44.
25. Robinson CM, Wylie JR, Ray AG, et al. Proximal humeral fractures with a severe varus deformity treated by fixation with a locking plate. *J Bone Joint Surg Br*. 2010;92(5):672-678.
26. Hasan AP, Phadnis J, Jaarsma RL, Bain GI. Fracture line morphology of complex proximal humeral fractures. *J Shoulder Elbow Surg*. 2017;26(10):e300-e308.
27. Matsumura N, Furuhata R, Seto T, et al. Reproducibility of the modified Neer classification defining displacement with respect to the humeral head fragment for proximal humeral fractures. *J Orthop Surg Res*. 2020; 15(1):438.
28. Roth KC, Denk K, Colaris JW, Jaarsma RL. Think twice before re-manipulating distal metaphyseal forearm fractures in children. Arch Orthop Trauma Surg. 2014;134(12):1699-707.
29. Labelbox. Labelbox, inc. Available at: https://labelbox.com.
30. ResNet-152. Torch Contributors. Available at: https://pytorch.org/vision/main/models/generated/torchvision.models.resnet152.html.
31. Chung SW, Han SS, Lee JW, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop.* 2018;89(4):468-473.
32. Google Colaboratory. Jupyter Notebook. Available at: https://colab.research.google.com/?utm_source=scs-index.
33. Critchley O, MacLean S, Hasan A, Woodman R, Bain G. Risk factors for intra-articular involvement in proximal humeral fractures. *Arch Orthop Trauma Surg*. 2023;143(3):1341-1351.
34. Mutch J, Laflamme GY, Hagemeister N, Cikes A, Rouleau DM. A new morphological classification for greater tuberosity fractures of the proximal humerus: Validation and clinical implications. *Bone Joint J*. 2014;96 B(5):646-651.
35. Valova I, Harris C, Mai T, Gueorguieva N. Optimization of Convolutional Neural Networks for Imbalanced Set Classification. *Procedia Comput Sci*. 2020;176:660-669.
36. Yoon AP, Lee Y-L, Kane RL, Kuo C-F, Lin C, Chung KC. Development and Validation of a Deep Learning Model Using Convolutional Neural Networks to Identify Scaphoid Fractures in Radiographs. *JAMA Netw Open*. 2021;4(5):e216096-e216096.
37. Meena T, Roy S. Bone Fracture Detection Using Deep Supervised Learning from Radiological Images: A Paradigm Shift. *Diagnostics (Basel).* 2022;12(10):2420.
38. Prijs J, Liao Z, To M-S, et al. Development and external validation of automated detection, classification, and localization of ankle fractures: inside the black box of a convolutional neural network (CNN). *Eur J trauma Emerg Surg*. 2023;49(2):1057-1069.

39. Huang S-T, Liu L-R, Chiu H-W, Huang M-Y, Tsai M-F. Deep convolutional neural network for rib fracture recognition on chest radiographs. *Front Med*. 2023;10:1178798.
40. Anderson PG, Baum GL, Keathley N, et al. Deep Learning Assistance Closes the Accuracy Gap in Fracture Detection Across Clinician Types. *Clin Orthop Relat Res*. 2023;481(3):580-588.
41. Magnéli M, Ling P, Gislén J, et al. Deep learning classification of shoulder fractures on plain radiographs of the humerus, scapula and clavicle. *PLoS One*. 2023;18(8):e0289808.
42. Zhang J, Liu F, Xu J, et al. Automated detection and classification of acute vertebral body fractures using a convolutional neural network on computed tomography. *Front Endocrinol (Lausanne)*. 2023;14:1132725.
43. Kalmet PHS, Sanduleanu S, Primakov S, et al. Deep learning in fracture detection: a narrative review. *Acta Orthop*. 2020;91(2):215-220.

SUPPLEMENTARY MATERIAL

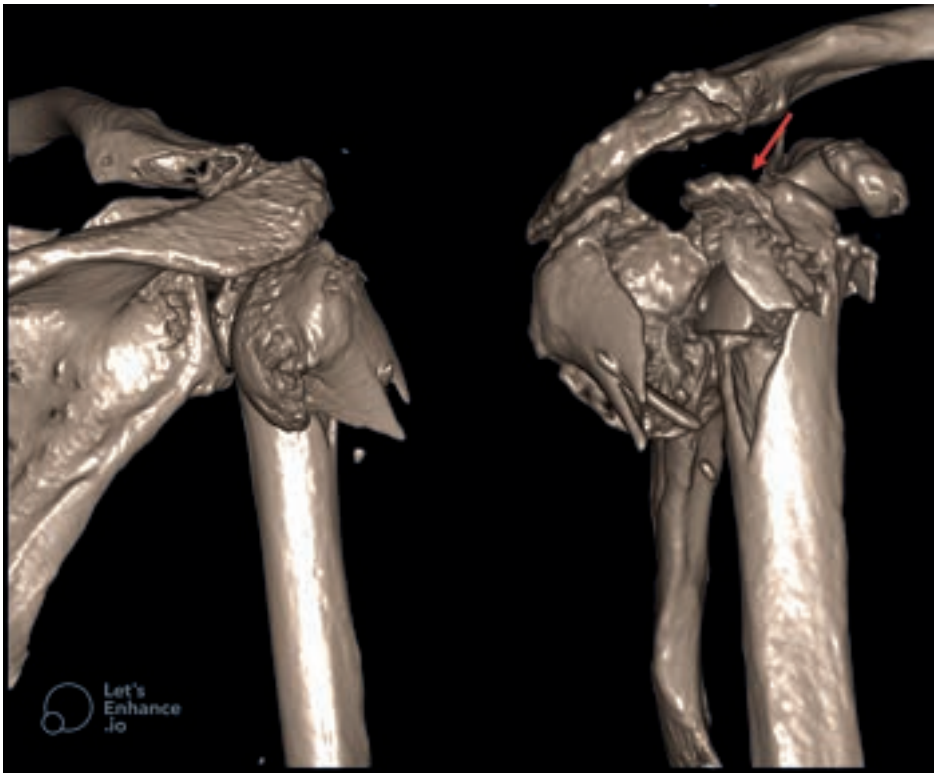


Supplement 1. Breakdown of patient selected for training and internal validation.

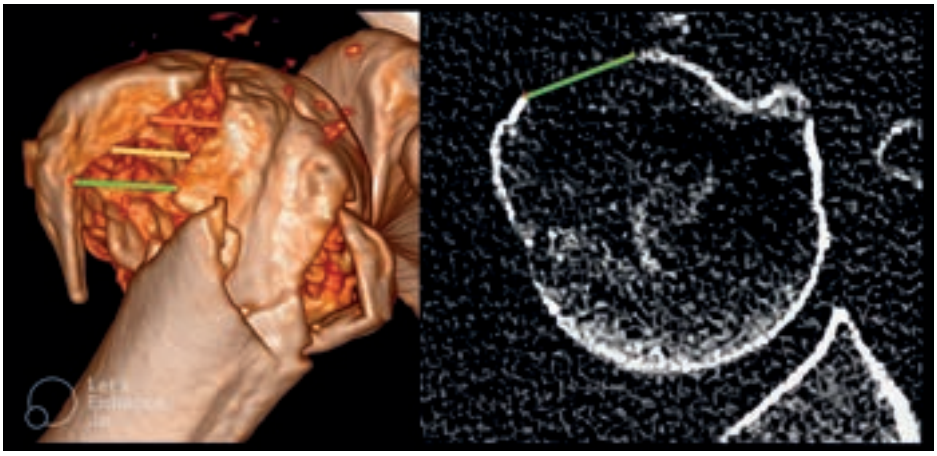
Supplement 2: Additional description on fracture characteristic definitions and methodology

1. Neck-shaft angle ≤100˚. If the variation between both measurements exceeded two times the mean differences (>11.6˚), the NSA was re-measured by the most experienced researcher (R.S.). The value furthest away from this measurement was then discarded and replaced by this new measurement. If the NSA was unmeasurable (due to poor quality of the 3D model), it was categorised by consensus agreement.

2. Extent of articular involvement. First, the 3D virtual model was created from the soft tissue 0.5 mm slices and bony structures other than the proximal humerus were removed. The model was then rotated to generate a frontal view were 50% of the articular segment was visible. On this image, a best fitting circle was overlayed on the humeral head and its diameter recorded. Next, we toggled back to the 3D virtual model to identify the *largest extension* of the fracture across the head. The view perpendicular to this fracture line was created and overlapped with the best-fitting humeral head circle which was generated in the previous step. By doing this, the angle of the fracture could be measured. As the mean articular arc in a normal shoulder is 145˚, we divided the fracture angle by this value and multiplied it by 100%. The population value of 145˚ was used for each fracture. The extent of articular involvement could not exceed 50%, and to ensure accuracy of measurements, zoom was not adjusted once the first best fitting circle was drawn. As RadiAnt Viewer did not have a protractor within the 3D section, we toggled between the snipping tool and radiant viewer using calibration markers to avoid measurement bias (Fig. 7). If the absolute difference between both measurements exceeded two times the mean difference (>12.4%), the fracture was re-measured by the first author (R.S.). The new measurement replaced the most unlikely value of the two. If the % of articular involvement could *not* be measured the correct label was determined in a consensus agreement between two researchers (the quality in some CT scans was insufficient to create smooth 3D models, particularly if the slice-thickness was above 5 mm). Once CT scans were loaded into the software and the learning curve was flattened, measurements took approximately 12 minutes. (Reverse) Hill-Sachs lesions were not considered as intra-articular fractures and were thus assigned to the 0% to <15% group.



Supplement 3. Greater tuberosity fractures were considered from an anatomical point of view: if only the anterior facet was displaced (red arrow), while the middle and posterior facet were intact and non-displaced in relation to the articular fragment, it was still considered a displaced greater tuberosity fracture.



Supplement 4. Greater tuberosity fractures can be measured on different levels which all return different results and can thus introduce bias. In this study, the maximum displacement within the defined boundaries of the greater tuberosity was measured (green line).

Supplement 5: Additional description on image pre-processing for uploading onto Labelbox

To upload the radiographs onto this software, the different views per patients were combined into one single PNG file (this required conversion of the original DICOM files into PNGs). The conversion was done with a script from Matlab version 9.12 (MathWorks, Natick, United States of America), the augmentation with a python script on PyCharm version 2021.2.3 (Python software foundation, Delaware, United States of America). The augmentation script re-sized the images to identical dimensions and produced a NPY file so that information on the distinct views could be retrieved in a later stage.


DARWIN
ALICE SPRINGS

6

What are the patient-reported outcomes, functional limitations, and complications after lesser tuberosity fractures? A systematic review of 172 patients

Reinier W.A. Spek
Bram J.A. Schoolmeesters
Chantal den Haan
Ruurd L. Jaarsma
Job N. Doornberg
Michel P.J. van den Bekerom

JSES International. 2021 Apr;5(4):754-764

ABSTRACT

Aims

Lesser tuberosity fractures are relatively rare, with an incidence of 0.46 per 100,000 persons per year. This systematic review was performed to address patient-reported outcomes (PROMs), shoulder function, and complications after lesser tuberosity fractures in paediatric and adult patients, as well as patients with an associated posterior shoulder dislocation. Within these groups, identical outcomes were evaluated for non-operative, surgical, acute, and delayed treatment.

Methods

A comprehensive search was carried out in multiple databases. Articles were included if patients sustained a lesser tuberosity fracture without a concomitant proximal humerus fracture. There were no restrictions on age, type of treatment, fragment displacement, time to presentation or associated injuries.

Results

One-thousand six hundred forty-four records were screened for eligibility of which 71 studies were included (n = 172). Surgical treatment was provided to 50 of 62 (81%) paediatric patients, 49 of 66 (74%) adults, and 34 of 44 (77%) patients with an associated posterior shoulder dislocation. In the paediatric group, the mean of PROMs was 94 (range 70 - 100) and among adults 89 (range 85 - 100). In the posterior shoulder dislocation group 89% did not regain full range of motion and the complication rate was 17%. In paediatric patients, surgery was associated with fewer complications ($p = 0.021$) compared to non-operative treatment.

Conclusion

Paediatric patients have excellent outcomes after lesser tuberosity fractures and respond well to surgical treatment. Adults have acceptable outcomes but patients with an associated posterior shoulder dislocation have impaired range of shoulder movement and are more likely to develop complications.

INTRODUCTION

The lesser tuberosity (LT) is a bony prominence on the proximal humerus, and important for stability and shoulder internal rotation as it accommodates insertion of the subscapularis tendon. Therefore, a fractured LT may cause shoulder dislocation or restricted internal rotation due to subscapularis insufficiency. LT fractures may occur in the setting of acute trauma –typically with the arm in 90° abduction and external rotation– or indirect, after repetitive stress caused by excessive overhead use of the arm such as in athletes of throwing sports or adolescents [1]. LT fractures are rarely seen in clinical practice, and likely to be missed as they are hard to detect on radiographs [2–4]. Moreover, missed or inadequately treated LT fractures may cause disabilities such as pain, muscle weakness and impaired shoulder movement due to the development of bony exostosis which has been described up to 20 years after the initial trauma [5].

Patients can be treated non-operatively, arthroscopically with suture anchors or via open reduction with internal screw fixation, tension band stabilization, or transosseous sutures. A hazard of non-operative management is secondary fragment dislocation and mal-union, whereas surgical treatment may result in surgery related complications such as infection or implant failure [1]. These options should be discussed with patients; however, there is sparse evidence on optimal management since only case reports and small case series are published to date [6–9]. Therefore, the options of operative *versus* non-operative management remain subject of ongoing debate [8,10–12]. Within this paucity of literature, there seems to be consensus that LT fractures displaced more than 1 centimetre should be treated surgically [3]. However, some studies suggest that surgeons should opt for surgical treatment if the amount of displacement is more than 5 mm, whereas other studies argue surgery for all LT fractures independent of fracture displacement due to concerns for secondary fracture displacement and impingement syndromes [9,10,13–15]. While Vavken et al compared the results of arthroscopic *versus* open surgical treatment and demonstrated the diagnostic importance of physical examination and magnetic resonance imaging in skeletally immature patients, no review has been carried out to ascertain functional and radiographic outcomes after non-operatively *versus* surgically treated paediatric nor adult patients with an LT fracture [2].

Therefore, this systematic review was performed to address the clinically relevant question: what are the patient-reported outcomes, shoulder function and

complications after lesser tuberosity fractures in paediatric and adult patients, as well as patients with an associated posterior shoulder dislocation? Within these groups, identical outcomes were evaluated for non-operative, surgical, acute, and delayed treatment. It was hypothesized that there was no difference in outcomes between paediatric and adult patients, as well as patients with an associated posterior shoulder dislocation.

PATIENTS AND METHODS

This systematic review was written according to the PRISMA guidelines and submitted for registration in PROSPERO on January 14, 2020 (ID number 165241) [16].

Search

A search strategy was created in collaboration with the clinical librarian (C.dH.). Studies were identified by searching Medline/Ovid, Embase.com, Cinahl/Ebsco, the Cochrane Database of Systematic Reviews for Cochrane Central Register of Controlled Trials, SPORTDiscus/Ebsco, Web of Science, Scopus, WHO ICTRP and Clinicaltrials.gov from inception up to and including October 14, 2019. Synonyms of "lesser tuberosity fracture", "subscapularis avulsion fracture" were combined with corresponding index terms and adjusted for every database. Details of the search are supplied in Supplement 1.

Selection

Records were identified with the search specified for each database and duplicates were removed in EndNote X8 (Clarivate Analytics, Boston, MA, USA). Following this identification, 2 authors (R.S. and B.S.) independently performed the screening based on title and abstract using Rayyan – a web and mobile app for systematic reviews (Ouzzani, Doha, Qatar) [17]. Subsequently, full texts were retrieved and were assessed independently for eligibility by the same authors. After each selection phase conflicts were resolved by discussion. If disagreement remained, the last author (M.vdB.) was consulted or the corresponding authors of the articles were contacted. Reference lists of the included articles were manually checked for potential additional relevant articles, and a forward reference check was performed using the Web of Science and Scopus.

Inclusion and exclusion criteria

Randomized trials, observational studies, case reports, letters and conference papers were eligible for this review. Articles were included if patients sustained a LT fracture of the proximal humerus which was managed non-operatively or surgically. A LT fracture was defined as an isolated avulsed bony fragment of the lesser tuberosity independent of the size without a concomitant proximal humerus fracture.

Articles were excluded if no outcome was described, data were not extractable to answer the primary research question after contacting the corresponding authors or if patients presented with a concomitant proximal humerus fracture such as a surgical neck or greater tuberosity fracture. Study protocols, surgical technique reports, editorials and animal or cadaver studies were also excluded.

There were no restrictions on associated injuries (such as shoulder dislocations, biceps tendon ruptures, labral injuries, or glenoid fractures), age, time to presentation, fracture displacement, type of outcome, follow-up length, language, or date of publication.

Quality assessment

The quality of case reports was assessed with the tool suggested by Murad et al and the case series were assessed with the Newcastle-Ottawa scale (NOS) for Cohort studies [18,19]. According to Murad's tool, case reports were evaluated on: 1) selection method, 2) ascertainment of exposure and outcome, 3) causality and 4) reporting. The NOS entailed 1) cohort representativeness, 2) ascertainment of exposure, 3) presence of outcome at start of the study, 4) assessment of outcome, 5) follow-up length, and 6) lost to follow-up rate. The overall quality of each article was judged as poor, fair, or good and was done by 2 authors independently (R.S. and B.S.). Any discrepancies were resolved by discussion in a consensus meeting.

Data extraction and synthesis

Data were collected in Microsoft Excel version 16.35 (Microsoft Corporation, Redmond, WA, USA). Demographic variables were extracted by the first author (R.S.) and the outcome variables were in duplicate extracted by 2 authors independently (R.S. and B.S.). Variables extracted in duplicate were follow-up length, pain, satisfaction, patient-reported outcomes measures (PROMs), range of motion (ROM), strength, complications, radiological assessment and return to sport, work,

and daily life activities. If individual patient data was not extractable but required to answer the research questions the corresponding authors were contacted. If the value of fracture displacement was not reported within an article, computed tomography (CT) or magnetic resonance imaging images presented in the article were appreciated under supervision of a senior author (M.B. and J.D.). If CT or magnetic resonance imaging images were not provided this value was reported as missing. PROMs were combined and expressed as a percentage of 100. The variables pain, strength, range of motion and radiographic assessment were categorized into binary variables. For instance, if a patient reported any pain at follow-up this was reported as "pain" and if a patient reported any muscle weakness at follow-up this was reported as "no full strength". Radiographic outcomes were categorized into union or non-union and outcomes reporting on ROM were categorized into restricted or non-restricted movement according to the cut-off values for elevation, abduction, and internal rotation provided in the Constant Murley Score [20]. External rotation was categorized according to the Rowe score [21].

Statistical analysis

Statistical analysis was performed using IBM SPSS software version 25 (IBM Corp., Armonk, NY, USA). Categorical variables were presented as numbers with percentages, and continuous variables as means with standard deviation or median with range depending on the distribution. To indicate significant differences in outcomes between paediatric patients, adults and posterior shoulder dislocations, a logistic regression analysis was used for categorical dependent variables and a linear regression analysis for continuous dependent variables. Within these subgroups, outcome differences were assessed between acute compared to delayed treatment, and non-operative compared to surgical treatment. Linear regression and logistic regression models were also used for these analyses. An additional regression analysis adjusted for country, was performed to control the models for patients derived from similar cohorts. A *p*-value less than 0.05 was considered to be significant.

RESULTS

A total of 4258 records were identified by the database search, and 1644 records were screened for eligibility after duplicate removal. There were 110 records selected for full-text assessment and broken down to 69 records for quality

assessment (Fig. 1). During full-text retrieval, 3 additional articles were found, and forward reference check revealed 164 articles of which 1 record was included [22–25]. The overall judgement of case reports was categorized as poor in 7 articles, fair in 45 articles and good in 9 articles (Supplement 2). The quality judgement of case series ranged from fair (5 articles) to good (5 articles) (Supplement 3). Given the low level of evidence of case reports and series, no articles were excluded based on the quality assessment. Taken together, 73 articles describing 71 studies were included in the systematic review [5,6,22–31,7,32–41,8,42–51,9,52–61,10,62–71,11,72–81,12,82–84,14,15] (Supplement 4 - 6).

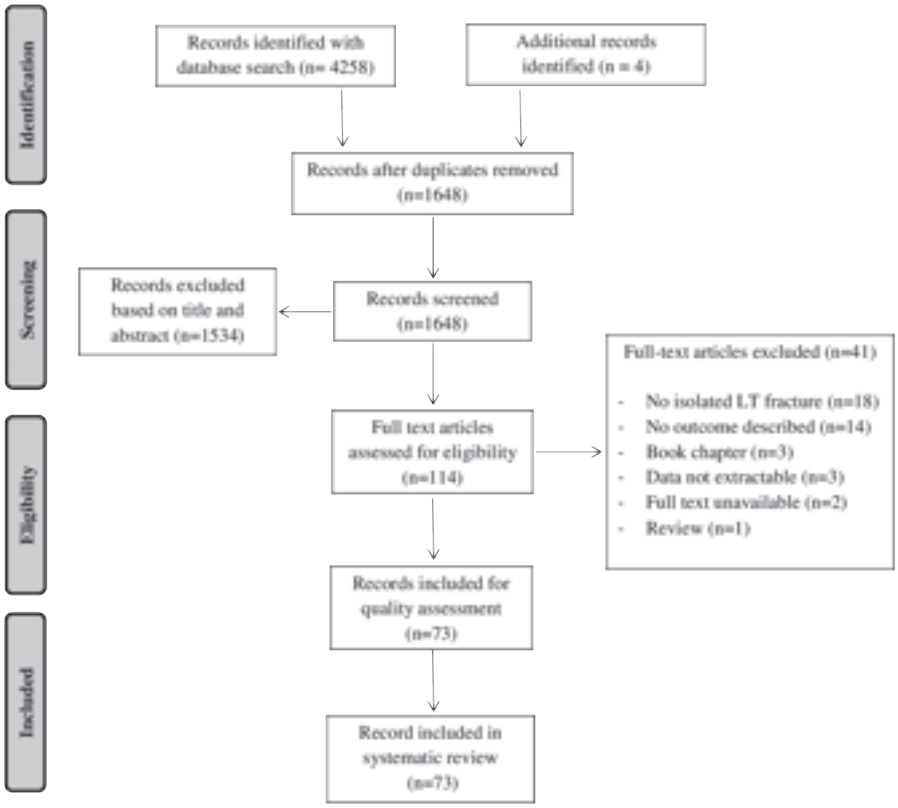


Figure 1. PRISMA breakdown diagram.

Cohort descriptions

Table 1 provides an overview of the final cohort. A total number of 175 shoulders from 172 patients were comprised in this review. There were 144 (82%) male patients of which the majority of the fractures (36%) occurred during sport. Eighty percent

of patients underwent surgery and mean follow-up length was 2.1 years (range 0.08 - 25.0). Surgical treatment was provided to 50 of 62 (81%) paediatric patients, 49 of 66 (74%) adults, and 34 of 44 (77%) patients with an associated posterior shoulder dislocation (PSD).

Table 1. Demographics of included patients (n = 172)

Variable	175 shoulders
Mean age at injury (range)	32.2 (9.0 - 77.0)
Male	144 (82.3%)
Right-sided fracture	88/153 (57.5%)
Dominant side involvement	51/80 (46.0%)
Mechanism of injury	
Sport accident	59/164 (36.0%)
Fall	29/164 (17.7%)
Seizure	25/164 (15.2%)
Traffic accident	20/164 (12.2%)
Fall from height	20/164 (12.2%)
Other *	11/164 (6.7%)
Associated injuries	
Posterior dislocation	47 (26.9%)
BT tear or dislocation	22 (12.6%)
RC pathology	12 (6.9%)
Labrocapsular ligamentous complex injuries	9 (5.1%)
Humeral head defect	9 (5.1%)
Anterior (sub) luxation	5 (2.9%)
Other †	5 (2.9%)
Fracture displacement >5 mm	84/101 (83.2%)
Non-operative treatment	41 (23.4%)
Surgical treatment	134 (76.6%)
Open	119 (88.8%)
Arthroscopic	15 (11.2%)

Table 1. Demographics of included patients (n = 172) (continued)

Variable	175 shoulders
Type of surgical fixation	
Screws	32/125 (25.6%)
Anchors	30/125 (24.0%)
Excision	22/125 (17.6%)
Modified McLaughlin	18/125 (14.4%)
Sutures	12/125 (9.6%)
Other	11/125 (8.8%)
Delayed treatment (>6 weeks)	36/129 (27.9%)
Mean years of follow-up (range)	2.1 (0.08 - 25.0)

Abbreviations: BT, biceps tendon; RC, rotator cuff. Data are expressed as number of shoulders with percentages. If data are missing, the total number of shoulders within a variable is reported after the slash. Age was missing in 1 shoulder, length of follow-up in 6 shoulders.
* Assault (n = 1), no trauma reported (n = 3), syncope (n = 2), hypoglycaemic fit (n = 1), electric shock (n = 4)
† Scapular spine fracture (n = 1), axillary nerve neuropraxia (n = 2), posterior glenoid rim fracture (n = 1), acromion fracture (n = 1).

Subgroup analyses

As Table 2 shows, there were 62 paediatric patients, 66 adults and 44 patients with an associated PSD. In the paediatric group 98% returned to sport, 87% regained full strength, the mean of PROMs was 94 (range 70 - 100) and 80% regained full ROM at follow-up. The mean of PROMs in adults was 89 (range 85 - 100), almost one-third (32%) had impaired range of shoulder movement in at least one plane and the complication rate was 5%. In the PSD group, 89% of shoulders did not regain full ROM and the complication rate was 17%. Unadjusted regression analysis indicated that posterior shoulder dislocations had a significantly lower mean of PROMs (p-value <0.001) compared to adult patients without PSD. When stratified for country the regression analysis indicated no significant association between patients with a PSD and PROMs compared to adults (p-value = 0.10). Results of the sensitivity analysis are supplied in Supplement 7.

Table 2. Outcomes of the paediatric (n = 62), adult (n = 66) and PSD group (n = 44)

	Paediatric	Adults		PSD	
Variable	**62 shoulders**	**66 shoulders**	***p*-value**	**47 shoulders**	***p*-value**
Mean age at injury (range)	13.0 (9.0 - 17.0)	41.3 (18.0 - 71.0)	<0.001	44.8 (28.0 - 77.0)	<0.001
Male	58 (93.5%)	46 (69.7%)	0.002	40 (85.1%)	0.16
Right-sided fracture	33/45 (73.3%)	31/62 (50.0%)	0.02	24/46 (52.2%)	0.04
Number of associated injuries	19 (14 shoulders)	36 (29 shoulders)	0.04	8 (6/9 shoulders)	0.01
Fracture displacement >5 mm	38/42 (90.5%)	43/54 (79.6%)	0.15	3/5 (60.0%)	0.08
Non-operative treatment	12 (19.4%)	17/66 (25.8%)	0.39	12/47 (25.5%)	0.44
Surgical treatment	50 (80.6%)	49/66 (74.2%)	0.39	35/47 (74.5%)	0.44
Open	43 (86.0%)	41 (83.7%)	0.75	35 (100.0%)	0.04†
Arthroscopic	7 (14.0%)	8 (16.3%)	0.75	0 (0.0%)	0.04†
Delayed treatment (>6 weeks)	27/55 (49.1%)	9/49 (18.4%)	0.001	0/25 (0.0%)	<0.001†
Mean years of follow-up (range)	2.6 (0.1 - 25.0)	1.3 (0.1 - 9.5)	0.01	2.6 (0.3 - 3.2)	0.96
Return to sport	40/41 (97.6%)	7/7 (100.0%)	1.00†	*NR*	*n/a*
Return to work	*NR*	6/7 (85.7%)	*n/a*	2/2 (100.0%)	1.00†
Return to daily life activities	10/10 (100.0%)	9/10 (90.0%)	1.00†	4/4 (100.0%)	*n/a*
Pain	3/28 (10.7%)	7/44 (15.9%)	0.54	2/6 (33.3%)	0.18
Mean VAS pain ‡	0.44 ± 0.7	*NR*	*n/a*	0.6 ± 0.8	0.26
Restricted movement	10/50 (20.0%)	12/38 (31.6%)	0.22	40/45 (88.9%)	<0.001
Full strength	27/31 (87.1%)	11/12 (91.7%)	0.68	4/4 (100.0%)	1.00†

Table 2. Outcomes of the paediatric (n = 62), adult (n = 66) and PSD group (n = 44) (continued)

	Paediatric	Adults		PSD	
Variable	**62 shoulders**	**66 shoulders**	***p*-value**	**47 shoulders**	***p*-value**
Satisfaction	8/8 (100.0%)	9/11 (81.8%)	0.49†	2/2 (100.0%)	*n/a*
Mean of PROMs (range)*	93.6 (70.0 - 100.0)	88.6 (85.0 - 100.0)	0.001	82.3 (81.2 - 97.1)	<0.001
Non-union	4/16 (25.0%)	3/13 (23.1%)	0.90	0/4 (0.0%)	0.54†
Complications	4 (6.5%)	3 (4.5%)	0.64	8 (17.0%)	0.09
Secondary surgery	2 (3.2%)	1 (1.5%)	0.53	2 (4.3%)	0.78

Abbreviations: PROMs, patient reported outcome measures; PSD, posterior shoulder dislocation; NR, not reported; VAS, visual analogue scale. Data are expressed as number of shoulders with percentages. If data are missing, the total number of shoulders within a variable is reported after the slash. The *p*-values of the unadjusted regression analysis are presented and are calculated with paediatric patients as the reference group.

* PROMs of 28 shoulders were described in the paediatric group and adult group, and 24 shoulders in the PSD group.

† Data were analysed with a Fisher's exact test.

‡ VAS pain score was described in 8 paediatric patients and 22 PSD patients.

Outcomes of surgical compared to non-operative treatment

The mean of PROMs in non-operatively treated paediatric patients was 84 (70-100) and the complication rate 27%. Complications (n = 3) included only mechanical impingement syndromes due to bony exostosis of the LT of which 2 patients required surgery. The mean of PROMs in surgically treated patients was 96 (85 - 100) and coincided with a 2% complication rate. In addition, 96% of the cases regained full strength after surgical treatment and 25% after non-operative treatment. Adjusted regression confirmed that full strength was significantly different (*p*-value = 0.019), favouring the surgical group (Supplement 8 - 10). Moreover, unadjusted regression analyses revealed that surgery was associated with a significantly higher mean of PROMs (*p*-value = 0.004) and fewer complications (*p*-value = 0.021) compared to non-operative treatment (Table 3).

The mean of PROMs in surgically treated adults was 94 (range 85 - 100), 76% of the cases regained full ROM and the complication rate was 5%. In the non-operatively treated group, 44% regained full shoulder ROM and the complication rate was 8%. Non-union was seen in 3 patients (38%) and only observed in the non-operative group. At follow-up, there was no statistically significant association between surgery and the outcomes as compared to non-operative treatment among adults (Table 4).

Non-operatively and surgically treated patients with PSD had similar complication rates of 17%. In the non-operative group, 70% had impaired shoulder movement and in the surgical group this percentage was 94%. Secondary surgery occurred only in the non-operative group (n = 2; 17%). Results are shown in Table 5.

Outcomes of delayed compared to acute treatment

Results of paediatric and adult patients were displayed in Table 6 to 9. There was no significant difference between the outcomes of acute and delayed treatment (>6 weeks) in paediatric and adult patients, as indicated by both adjusted and unadjusted regression models (Supplement 11 - 14). Regression analysis showed that patients with delayed presentation had significantly more associated injuries in both the surgical (*p*-value = 0.004) and non-operative group (*p*-value = 0.034). The most common reported injuries were biceps tendon (BT) tears, dislocations and labrocapsular ligamentous complex injuries.

Table 3. Outcomes of non-operative and surgical treatment in paediatric patients (n = 56)

Variable	Non-operative 11 shoulders	Surgical 45 shoulders	*p*-value
Mean age at injury (range)	13.3 (12.0 - 17.0)	12.9 (9.0 - 17.0)	0.46
Male	9 (81.8%)	43 (95.6%)	0.14
Right-sided fracture	9 (81.8%)	24/34 (70.6%)	0.47
Number of associated injuries	0	19 (14 shoulders)	0.049
Fracture displacement >5 mm	4/7 (57.1%)	29/29 (100.0%)	0.005†
Open surgical treatment	*n/a*	38 (84.4%)	*n/a*
Arthroscopic surgical treatment	*n/a*	7 (15.6%)	*n/a*
Delayed treatment (>6 weeks)	4/11 (36.4%)	20/38 (44.4%)	0.35
Mean years of follow-up (range)	7.2 (0.13 - 25.0)	1.81 (0.23 - 7.00)	<0.001
Return to sport	4/5 (80.0%)	36/36 (100.0%)	0.12†
Return to daily life activities	2/2 (100.0%)	8/8 (100.0%)	*n/a*
Pain	3/9 (33.3%)	0/13 (0.0%)	0.055†
Mean VAS pain	*NR*	0.44 ± 0.7	*n/a*
Restricted movement	4/8 (50.0%)	6/36 (16.7%)	0.054
Full strength	1/4 (25.0%)	26/27 (96.3%)	0.005
Satisfied	*NR*	8/8 (100.0%)	*n/a*
Mean of PROMs (range)	84.4 (70.0 - 100.0)	95.6 (85.0 - 100.0)	0.004
Non-union	3/8 (37.5%)	1/8 (12.5%)	0.57†
Complications*	3 (27.3%)	1 (2.2%)	0.02
Secondary surgery	2 (18.2%)	0 (0.0%)	0.04†

Abbreviations: PROMs, patient reported outcome measures; NR, not reported; n/a, not applicable. Data are expressed as number of shoulders with percentages. If data are missing, the total number of shoulders within a variable is reported after the slash. The *p*-values of the unadjusted regression analysis are presented.

Studies by Nardo et al and Nové-Josserand et al were excluded since data on initial treatment was not extractable per case.

Follow-up length was reported in 10 shoulders in the non-operative group and in 43 shoulders in the surgical group. PROMs of 5 shoulders were described in the non-operative group and 23 shoulders in the surgical group. The mean VAS was reported in 5 shoulders in the surgical group.

* Mechanical impingements due to bony exostosis (n = 3) was observed after non-operative treatment. Secondary fragment displacement (n = 1) was reported after surgical treatment.

† Data were analysed with a Fisher's exact test.

Table 4. Outcomes of non-operative and surgical treatment in adults (n = 49)

Variable	Non-operative 12 shoulders	Surgical 37 shoulders	*p*-value
Mean age at injury (range)	47 (18.0 - 68.0)	37.9 (18.0 - 71.0)	0.06
Male	6 (50.0%)	25 (67.6%)	0.28
Right-sided fracture	5 (41.7%)	15/33 (45.5%)	0.82
Number of associated injuries	4 (2 shoulders)	17 (12 shoulders)	0.62
Fracture displacement >5 mm	5/9 (55.6%)	32/32 (100.0%)	0.001†
Open surgical treatment	*n/a*	34 (91.9%)	*n/a*
Arthroscopic surgical treatment	*n/a*	3 (8.1%)	*n/a*
Delayed treatment (>6 weeks)	2 (16.7%)	7 (18.9%)	0.86
Mean years of follow-up (range)	1.3 (0.2 - 5.0)	1.2 (0.1 - 9.5)	0.85
Return to sport	2/2 (100.0%)	5/5 (100.0%)	*n/a*
Return to work	1/1 (100.0%)	5/6 (83.3%)	1.00†
Return to daily life activities	4/4 (100.0%)	5/6 (83.3%)	1.00†
Pain	1/8 (12.5%)	4/19 (21.1%)	0.61
Restricted movement	5/9 (55.6%)	7/29 (24.1%)	0.09
Full strength	4/4 (100.0%)	7/8 (87.5%)	1.00†
Satisfied	4/4 (100.0%)	5/7 (71.4%)	0.49†
Mean of PROMs (range)	89.8 (85.0 - 95.0)	94.3 (85.0 - 100.0)	0.20
Non-union	3/8 (37.5%)	0/5 (0.0%)	0.23†
Complications*	1 (8.3%)	2 (5.4%)	0.72
Secondary surgery	0 (0.0%)	1 (2.7%)	1.00†

Abbreviations: PROMs, patient reported outcome measures; NR, not reported; n/a, not applicable. Data are expressed as number of shoulders with percentages. If data are missing, the total number of shoulders within a variable is reported after the slash. The *p*-values of the unadjusted regression analysis are presented.

Studies by Nardo et al and Nové-Josserand et al were excluded since data on initial treatment was not extractable per case. Follow-up length was reported in 33 shoulders in the surgical group. PROMs of 4 shoulders were described in the non-operative group and 7 shoulders in the surgical group.

* Mechanical impingement due to bony exostosis (n = 1) was observed after non-operative treatment. Secondary fragment displacement (n = 1) and frozen shoulder (n = 1) were reported after surgical treatment.

† Data was analysed with a Fisher's exact test.

Table 5. Outcomes of non-operative and surgical treatment in patients with a PSD (n = 44)

Variable	Non-operative 12 shoulders	Surgical 35 shoulders	*p*-value
Mean age at injury (range)	45.9 (29.0 - 77.0)	44.4 (28.0 - 63.0)	0.67
Male	9 (75.0%)	31 (88.6%)	0.27
Right-sided fracture	4 (33.3%)	20 (57.1%)	0.24
Number of associated injuries	3 (3/4 shoulders)	5 (3/5 shoulders)	0.67
Fracture displacement >5 mm	0/2 (0.0%)	3/3 (100.0%)	0.10†
Open surgical treatment	*n/a*	35 (100.0%)	*n/a*
Arthroscopic surgical treatment	*n/a*	0 (0.0%)	*n/a*
Delayed treatment (>6 weeks)	0/12 (0.0%)	0/13 (0.0%)	*n/a*
Mean years of follow-up (range)	2.2 (0.82 - 3.0)	2.7 (0.3 - 3.2)	0.09
Return to work	2/2 (100.0%)	*NR*	*n/a*
Return to daily life activities	2/2 (100.0%)	2/2 (100.0%)	*n/a*
Pain	1/1 (100.0%)	1/5 (20.0%)	0.33†
Mean VAS pain	*NR*	0.6 ± 0.8	*n/a*
Restricted movement	7/10 (70.0%)	33/35 (94.3%)	0.051
Full strength	3/3 (100.0%)	1/1 (100.0%)	*n/a*
Satisfaction	1/1 (100.0%)	1/1 (100.0%)	*n/a*
Mean of PROMs (range)	92.0 (92.0 - 92.0)	81.9 (81.2 - 97.0)	0.007
Non-union	0/2 (0.0%)	0/2 (0.0%)	*n/a*
Complications*	2 (16.7%)	6 (17.1%)	0.97
Secondary surgery	2 (16.7%)	0 (0.0%)	0.06†

Abbreviations: PROMs, patient reported outcome measures; NR, not reported; n/a, not applicable.
Data are expressed as number of shoulders with percentages. If data are missing, the total number of shoulders within a variable is reported after the slash. The *p*-values of the unadjusted regression analysis are presented.
Studies by Nardo et al and Nové-Josserand et al were excluded since data on initial treatment was not extractable per case.
Mean of PROMs of 1 shoulder were described in the non-operative group and 23 shoulders in the surgical group.
*Closed reduction group: iatrogenic fracture (n = 1) and redislocation requiring surgery (n = 1). Surgery group: humeral head necrosis (n = 4) and dorsal suture anchor perforation (n = 1). One patient suffered an iatrogenic brachial plexus injury (n = 1) after initial reduction, before she underwent surgery.
†Data was analysed with a Fisher's exact test.

Table 6. Outcomes of acute and delayed surgery in paediatric patients (n = 38)

	Surgery		
	Acute	Delayed	
Variable	**18 shoulders**	**20 shoulders**	***p*-value**
Number of associated injuries	6 (5 shoulders)	6 (4 shoulders)	0.87
BT tear or dislocation	2 (11.1%)	5 (25.0%)	0.36
LCLC injuries	3 (16.7%)	0 (0.0%)	0.10†
RC pathology	0.0 (0.0%)	1 (5.0%)	1.00†
Anterior (sub) luxation	1 (5.6%)	0 (0.0%)	0.47†
Fracture displacement >5 mm	14/14 (100.0%)	15/15 (100.0%)	*n/a*
Mean years of follow-up (range)	1.6 (0.2 - 6.7)	1.5 (0.4 - 5.0)	0.79
Return to sport	16/16 (100.0%)	15/15 (100.0%)	*n/a*
Return to daily life activities	3/3 (100.0%)	3/3 (100.0%)	*n/a*
Pain	0/7 (0.0%)	0/4 (0.0%)	*n/a*
Mean VAS pain	0.2 (0.0 - 1.0)	0.83 (0.0- 2.0)	0.26
Restricted movement	3/16 (18.8%)	3/15 (20.0%)	0.93
Full strength	14/14 (100.0%)	7/8 (87.5%)	0.36†
Satisfied	5/5 (100.0%)	3/3 (100.0%)	*n/a*
Mean of PROMs (range)	95.7 (85.0 - 100.0)	95.6 (91.0 - 99.6)	0.97
Non-union	1/3 (33.3%)	0/5 (0.0%)	0.38†
Complications	0 (0.0%)	1 (5.0%)	1.00†
Secondary surgery	0 (0.0%)	0 (0.0%)	*n/a*

Abbreviations: PROMs, patient reported outcome measures; NR, not reported; n/a, not applicable. Data are expressed as number of shoulders with percentages. If data are missing, the total number of shoulders within a variable is reported after the slash. The *p*-values of the unadjusted regression analysis are presented.

The mean VAS score was reported in 5 acute and 3 delayed surgically treated patients. The mean of PROMs was reported in 11 acute and 7 delayed surgically treated patients, and 4 acute and 1 delayed non-operative treated patients. Studies by Nardo et al, Nové-Josserand et al, Liu et al, Garrigues et al, Weiss et al were excluded since data on acute and delayed treatment was not extractable per case.

† Data was analysed with a Fisher's exact test.

Table 7. Outcomes of acute and delayed non-operative treatment in paediatric patients (n = 11)

	Non-operative		
	Acute	Delayed	
Variable	**7 shoulders**	**4 shoulders**	***p*-value**
Number of associated injuries	0 (0.0%)	0 (0.0%)	*n/a*
Fracture displacement >5 mm	3/4 (75.0%)	1/3 (33.3%)	0.29
Mean years of follow-up (range)	10.3 (0.13 - 25.0)	13.8 (13.0 - 15.0)	0.20
Return to sport	2/2 (100.0%)	2/3 (66.7%)	1.00[†]
Return to daily life activities	2/2 (100.0%)	*NR*	*n/a*
Pain	2/6 (33.3%)	1/3 (33.3%)	1.00
Restricted movement	3/6 (50.0%)	1/2 (50.0%)	1.00
Full strength	0/2 (0.0%)	1/2 (50.0%)	1.00[†]
Mean of PROMs (range)	81.8 (70.0 - 100.0)	95.0 (95.0 - 95.0)	0.46
Non-union	2/5 (40.0%)	1/3 (33.3%)	0.85
Complications	2 (28.6%)	1 (25.0%)	0.90
Secondary surgery	2.0 (28.6%)	0.0 (0.0%)	0.49[†]

Abbreviations: PROMs, patient reported outcome measures; NR, not reported; n/a, not applicable. Data are expressed as number of shoulders with percentages. If data are missing, the total number of shoulders within a variable is reported after the slash. The *p*-values of the unadjusted regression analysis are presented.

The mean of PROMs was reported in 4 acute and 1 delayed non-operatively treated patient. Studies by Nardo et al, Nové-Josserand et al, Liu et al, Garrigues et al, and Weiss et al were excluded since data on acute and delayed treatment was not extractable per case.

[†] Data was analysed with a Fisher's exact test.

Table 8. Outcomes of acute and delayed surgery in adults (n = 37)

	Surgery		
	Acute	Delayed	
Variable	**30 shoulders**	**7 shoulders**	***p*-value**
Number of associated injuries	9 (7 shoulders)	8 (5 shoulders)	0.004
BT tear or dislocation	4 (13.3%)	4 (57.14%)	0.34
LCLC injuries	0 (0.0%)	2 (28.6%)	0.03†
RC pathology	2 (6.7%)	1 (14.3%)	0.01
Humeral head defect	1 (3.3%)	1 (14.3%)	0.26
Other	2 (6.7%)	0 (0.0%)	1.00†
Fracture displacement >5 mm	26/26 (100.0%)	6/6 (100.0%)	*n/a*
Mean years of follow up (range)	1.3 (0.1 - 9.5)	0.7 (0.3 - 1.0)	0.48
Return to sport	2/2 (100.0%)	3/3 (100.0%)	*n/a*
Return to work	5/5 (100.0%)	0/1 (0.0%)	0.17†
Return to daily life activities	5/6 (83.3%)	*NR*	*n/a*
Pain	2/14 (14.3%)	2/5 (28.6%)	0.24
Restricted movement	6/23 (26.1%)	1/6 (16.7%)	0.63
Full strength	7/8 (87.5%)	*NR*	*n/a*
Satisfied	2/3 (66.7%)	3/4 (75.0%)	0.81
Mean of PROMs (range)	94.3 (85.0 - 100.0)	*NR*	*n/a*
Non-union	0/5 (0.0%)	*NR*	*n/a*
Complications	1 (3.3%)	1 (14.3%)	0.29
Secondary surgery	0.0 (0.0%)	1 (14.3%)	0.19†

Abbreviations: PROMs, patient reported outcome measures; NR, not reported; n/a, not applicable. Data are expressed as number of shoulders with percentages. If data are missing, the total number of shoulders within a variable is reported after the slash. The *p*-values of the unadjusted regression analysis are presented.
The mean of PROMs was described in 7 acute surgically treated shoulders, 3 acute and 1 delayed non-operatively treated shoulder.
† Data was analysed with a Fisher's exact test.

Table 9. Outcomes of acute and delayed non-operative treatment in adults (n = 12)

	Non-operative		
	Acute	Delayed	
Variable	**10 shoulders**	**2 shoulders**	***p*-value**
Number of associated injuries	1 (1 shoulder)	3 (1 shoulder)	0.03
LCLC injuries	0 (0.0%)	2 (100.0%)	0.17†
Humeral head defect	0 (0.0%)	1 (50.0%)	0.17†
Other	1 (10.0%)	0 (0.0%)	1.00†
Fracture displacement >5 mm	4/8 (50.0%)	1/1 (100.0%)	1.00†
Mean years of follow up (range)	0.8 (0.2 - 3.3)	3.5 (1.0 - 5.0)	0.02
Return to sport	2/2 (100.0%)	*NR*	*n/a*
Return to work	1/1 (100.0%)	*NR*	*n/a*
Return to daily life activities	4/4 (100.0%)	*NR*	*n/a*
Pain	0/7 (0.0%)	1/1 (100.0%)	0.13†
Restricted movement	4/8 (50.0%)	0/1 (0.0%)	1.00†
Full strength	4/4 (100.0%)	*NR*	*n/a*
Satisfied	4/4 (100.0%)	*NR*	*n/a*
Mean of PROMs (range)	88.0 (85.0 - 94.0)	95.0 (95.0 - 95.0)	0.36
Non-union	2/5 (71.4%)	1/1 (100.0%)	0.38†
Complications	0 (0.0%)	1 (50.0%)	0.17†
Secondary surgery	0.0 (0.0%)	0.0 (0.0%)	*n/a*

Abbreviations: PROMs, patient reported outcome measures; NR, not reported; n/a, not applicable. Data are expressed as number of shoulders with percentages. If data are missing, the total number of shoulders within a variable is reported after the slash. The *p*-values of the unadjusted regression analysis are presented.
The mean of PROMs was described in 3 acute and 1 delayed non-operatively treated shoulder.
† Data were analysed with a Fisher's exact test.

DISCUSSION

LT fractures are relatively rare, with an incidence of 0.46 per 100,000 persons per year. Moreover, options of operative- *versus* non-operative management of minimally displaced LT fractures remain subject of ongoing debate [3]. To the best of our knowledge, this study identified all reported patients and adds to literature since existing studies have drawn different conclusions on this issue [85]. As illustration, some case series on adult patients report excellent surgical outcomes, whereas others observe comparable outcomes of non-operative treatment, even in the setting of displaced fractures [6,10,30]. In paediatric patients, the majority is treated surgically and data on outcomes of non-operative treatment are limited. This review combines case reports and series to create a relatively large patient cohort aiming to provide an overview to compare these treatment strategies and inform patients about expected results. The objective was to answer the clinical question: what are patient-reported outcomes, shoulder function and complications after lesser tuberosity fractures in paediatric and adult patients, including patients with an associated posterior shoulder dislocation? Within these groups, identical outcomes were evaluated for non-operative, surgical, acute, and delayed treatment in order to guide surgical decision-making: should surgeons opt for surgical treatment in minimally displaced LT fractures?

Paediatric patients have excellent outcomes after LT fractures with almost all patients returned to sport, a high mean of PROMs and a low complication rate. Similarly, this is explained by physiological benefits of children: they have a strong ability to remodel bone, and compared to adults, they have quicker fracture healing [86]. Moreover, they respond well to surgical treatment and show significantly less complications and a higher mean of PROMs compared to non-operative treatment. Adults have acceptable outcomes, but it should be noted that almost one-third did not regain a full ROM. There also seemed to be a trend towards a beneficial effect of surgical treatment; however, this difference was not significant with the numbers available. The complication rate of LT fractures after posterior shoulder dislocations was higher, and almost all patients had limited upper limb function at follow-up. Outcomes after delayed treated patients (>6 weeks) were acceptable but must be interpreted with caution due to the low number of patients within this group.

Consistent with the review of Vavken et al, this study confirmed that surgical treatment of LT fractures provides excellent results in paediatric patients [2]. Additionally, it was found that paediatric patients had better outcomes of surgical treatment compared to non-operative treatment. For this reason, clinicians should strongly consider to treat paediatric patients surgically if LT fractures are displaced more than 5 mm.

In accordance with the well- designed case series of Robinson et al and Cottias et al, this study revealed good outcomes after surgically treated adult patients [3,87]. Moreover, Cottias et al pointed out that almost one-third of the initial non-operatively treated patients had to undergo surgery due to secondary fragment displacement [87]. Therefore, these authors advocated for surgical treatment over non-operative treatment in patients with a displaced LT fracture [3,87]. In this review however, surgical treatment was not associated with better outcomes compared to non-operative treatment and unfortunately both case series were excluded because data were not extractable from patients with and without a PSD. It may be that some non-operatively treated patients in this cohort should have been treated surgically as over half of the patients had more than 5 mm fracture displacement. This was supported by an additional analysis which showed that all adverse outcomes and events occurred in non-operatively treated patients with more than 5 mm of displacement.

In this cohort, almost one-third of all shoulders were dislocated posteriorly, so suspicion should be raised for an LT fracture if patients present with a PSD. Viewed from a biomechanical perspective the fracture is a result of the increased stress of the subscapularis muscle due to posterior luxation. Clinicians should also advise them about the relatively high complication rate and the likelihood that they will not regain full ROM. However, a note of caution is due here since the mean of functional outcome scores were acceptable despite patients did not regain full ROM and that outcomes were not compared between the different types of surgical treatment such as reversed shoulder prosthesis, modified McLaughlin technique or restoration of the humeral head with bone stock [58]. It is important to bear in mind that patients with a PSD are more likely to undergo surgery due to associated reverse Hills-Sachs lesions which are associated with higher risk of recurrent PSD if left untreated [88].

In clinical practice, clinicians should be aware of LT fractures and must assess radiographs carefully [4]. Surgical decision-making should include fracture displacement, symptoms, and demands of the patient. The majority of data is

published on surgical treatment, so clear guidelines on non-operative treatment cannot be provided. However, to the best of our knowledge, we recommend conservative treatment for non-displaced LT fractures and in patients not fit for surgery. If non-operative treatment is chosen patients should be monitored closely and radiographs should be taken regularly and assessed for secondary fragment displacement. If secondary displacement occurs, a low threshold for surgical treatment should be followed, in particular for adults as they have less remodelling capacity than adolescents. Arthroscopic anchor suture fixation of the facture is associated with excellent outcomes and should be performed if fragment size allows this. Alternatively, open reduction with internal screw or anchor suture or transosseous suture fixation can be performed. Cancellous bone screw fixation can be performed by judgement of the surgeon [1].

In some cases, it can be hard to appreciate the size and degree of displacement of LT fractures. It is therefore advised to perform a CT scan when considering surgical treatment. Moreover, patients with a LT fracture may present with associated injuries such as BT dislocations or tears. For this reason, surgeons should visualize the BT during surgery and if BT pathology is suspected an ultrasound can be used in the acute clinical setting [89].

There is an important issue for further research to determine the maximum displacement accepted for non-operative treatment. Preferably, a multicenter, randomized controlled trial will be carried out in which patients with a minimal displaced LT fracture are allocated to either surgical or non-operative management. However, owing to the rarity of LT fractures this is almost unfeasible. Therefore, we advise a nationwide cross-sectional study in which all hospitals document and monitor these patients for 2 years and measure outcomes with PROMs, strength, ROM, and radiologic assessment. This study should also address the following questions: (1) does the shape of the fragment determines outcomes? (2) which fractures associated to PSD need surgery?

There are some important potential drawbacks associated with this review. First, outcome measures had to be merged due to the widespread variation of reported outcomes, so conclusions should be interpreted carefully. Second, there is limited data available since only case reports and case series are reported on this subject and, third, there is a high potential for publication bias given that LT fractures are

rare and that not all patients with an LT fracture worldwide are documented and published. Fourth, regression analysis was adjusted for country as adjusting for 71 different cohorts did not fit the model. Therefore, findings for both the adjusted and unadjusted regression analysis were provided but heterogeneity of population should be taken into account (Supplement 7 - 14). Finally, patients with posterior shoulder dislocations were compared to paediatric and adult patients, but should be considered as the most complex trauma group among these patients. However, within these limits, this review is a collection of the best evidence available.

Conclusion

In clinical practice, this review can be used for patient consultation and provides an overview of expected outcomes after LT fractures. It can be concluded that paediatric patients have excellent outcomes after LT fractures and may benefit more from surgery in comparison to non-operative treatment. While the outcomes of adults are also acceptable, it is clear that the majority of patients with a PSD have lower functional outcomes scores, impaired range of shoulder movements, and are more likely to develop complications compared to adult patients. It also highlights the importance that good outcomes can be achieved in delayed treated patients. However, poor quality of included studies has to be taken into account.

REFERENCES

1. Gruson KI, Ruchelsman DE, Tejwani NC. Isolated tuberosity fractures of the proximal humeral: Current concepts. *Injury*. 2008;39(3):284-298.
2. Vavken P, Bae DS, Waters PM, Flutie B, Kramer DE. Treating Subscapularis and Lesser Tuberosity Avulsion Injuries in Skeletally Immature Patients: A Systematic Review. *Arthroscopy*. 2016;32(5):919-928.
3. Robinson CM, Teoh KH, Baker A, Bell L. Fractures of the Lesser Tuberosity of the Humerus. *J Bone Joint Surg Am*. 2009;91(3):512-520.
4. Harper DK, Craig JG, van Holsbeeck MT. Apophyseal injuries of the lesser tuberosity in adolescents: a series of five cases. *Emerg Radiol*. 2013;20(1):33-37.
5. Goeminne S, Debeer P. The natural evolution of neglected lesser tuberosity fractures in skeletally immature patients. *J Shoulder Elbow Surg*. 2012;21(8):e6-e11.
6. van Laarhoven HAJ, te Slaa RL, van Laarhoven EW. Isolated Avulsion Fracture of the Lesser Tuberosity of the Humerus. *J Trauma*. 1995;39(5):997-999.
7. Garrigues GE, Warnick DE, Busch MT. Subscapularis Avulsion of the Lesser Tuberosity in Adolescents. *J Pediatr Orthop*. 2013;33(1):8-13.
8. Paschal SO, Hutton KS, Weatherall PT. Isolated avulsion fracture of the lesser tuberosity of the humerus in adolescents. A report of two cases. *J Bone Joint Surg Am*. 1995;77(9):1427-1430.
9. Paxinos O, Karavasili A, Manolarakis M, Paxinos T, Papavasiliou A. Neglected lesser tuberosity avulsion in an adolescent elite gymnast. *Shoulder Elbow*. 2014;6(3):178-181.
10. Ogawa K, Takahashi M. Long-term Outcome of Isolated Lesser Tuberosity Fractures of the Humerus. *J Trauma*. 1997;42(5):955-959.
11. Kowalsky MS, Bell JE, Ahmad CS. Arthroscopic treatment of subcoracoid impingement caused by lesser tuberosity malunion: a case report and review of the literature. *J Shoulder Elbow Surg*. 2007;16(6):e10-4.
12. Kunkel SS, Monesmith EA. Isolated avulsion fracture of the lesser tuberosity of the humerus: A case report. *J Shoulder Elbow Surg*. 1993;2(1):43-46.
13. Bono CM, Renard R, Levine RG, Levy AS. Effect of displacement of fractures of the greater tuberosity on the mechanics of the shoulder. *J Bone Joint Surg Br*. 2001;83(7):1056-1062.
14. Levine B, Pereira D, Rosen J. Avulsion fractures of the lesser tuberosity of the humerus in adolescents: review of the literature and case report. *J Orthop Trauma*. 2005;19(5):349-352.
15. Caniggia M, Maniscalco P, Picinotti A. Isolated avulsion fracture of the lesser tuberosity of the humerus. Report of two cases. *Panminerva Med*. 1996;38(1):56-60.
16. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Int J Surg*. 2010;8(5):336-341.
17. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210.
18. Murad MH, Sultan S, Haffar S, Bazerbachi F. Methodological quality and synthesis of case series and case reports. *BMJ Evid Based Med*. 2018;23(2):60-63.
19. Wells G, Shea B, O'Connell D, et al. The Newcastle–Ottawa Scale (NOS) for Assessing the Quality of Non-Randomized Studies in Meta-Analysis. Ottawa Health Research institute, Ottawa, Canada. 2000.
20. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res*. 1987;(214):160-164.
21. Rowe CR, Zarins B. Recurrent transient subluxation of the shoulder. *J Bone Joint Surg Am*. 1981;63(6):863-872.

22. Kim T-H, Hong K-D, Ha S-S, Sim J-C, Sung M-C. Isolated Avulsion Fracture of the Lesser Tuberosity of the Humerus - A Case Report -. *J Korean Fract Soc*. 2014;27(1):72-76.
23. Athanasiadis LP, Petropoulou ET, Neonakis EM. Conservative treatment of isolated avulsion fracture of the lesser tuberosity of the humerus: A case report. *J Res Pract Musculoskelet Syst*. 2017;01(02):38-40.
24. Jaya T, Hadizie D, TMS M. Isolated Avulsion Fracture Lesser Tuberosity of the Humerus, a Rare Presentation Post Seizure. *Trauma Cases Rev*. 2019;5(3):1-5.
25. Sarraf KM, Sadri A, Willis-Owen CA. Rare case of bilateral posterior fracture dislocation of the shoulders secondary to a syncopal episode. *Ortop Traumatol Rehabil*. 2009;11(5):476-480.
26. Zimmermann A, Agneskirchner JD. Isolated complete rupture of the subscapularis tendon in a child. *Arthroskopie*. 2017;31(1):74-77.
27. Zanlungo U, Manetti G, La Cavera F. Isolated avulsion fracture of the lesser tuberosity of the humerus. A case report. *Minerva Ortop Traumatol*. 1994;45(3):97-99.
28. White GM, Riley Jr. LH. Isolated avulsion of the subscapularis insertion in a child. A case report. *J Bone Joint Surg Am*. 1985;67(4):635-636.
29. Weiss JM, Arkader A, Wells LM, Ganley TJ. Rotator cuff injuries in adolescent athletes. *J Pediatr Orthop B*. 2013;22(2):133-137.
30. Vezeridis PS, Bae DS, Kocher MS, Kramer DE, Yen Y-M, Waters PM. Surgical Treatment for Avulsion Injuries of the Humeral Lesser Tuberosity Apophysis in Adolescents. *J Bone Joint Surg Am*. 2011;93(20):1882-1888.
31. Tosun B, Kesemenli C. Isolated avulsion fracture of lesser tuberosity of the humerus: Review of the literature and report of two cases. *Int J Shoulder Surg*. 2011;5(2):50.
32. Thielemann FW, Kley U, Holz U. Isolated injury of the subscapular muscle tendon. *Sportverletz Sportschaden*. 1992;6(1):26-28.
33. Teixeira RP, Johnson AR, Higgins BT, Carrino JA, McFarland EG. Fly Fishing–related Lesser Tuberosity Avulsion in an Adolescent. *Orthopedics*. 2012;35(5):e748-e751.
34. Tabrizi A, Safari MB, Taleb H, Navaeifar N. Locked posterior dislocation of shoulder with fracture of the lesser tuberosity of the humerus: A case report and review of the literature. *Arch Trauma Res*. 2017;6(2):1-4.
35. Sugalski MT, Hyman JE, Ahmad CS. Avulsion fracture of the lesser tuberosity in an adolescent baseball pitcher: a case report. *Am J Sports Med*. 2004;32(3):793-796.
36. Steinbach L, Nardo L, Ma B. Lesser tuberosity avulsions in adolescents. *Pediatr Radiol*. 2011;41:S391.
37. Sikka RS, Neault M, Guanche CA. An avulsion of the subscapularis in a skeletally immature patient. *Trauma Cases Rev*. 2011;31(1):793-796.
38. Shibuya S, Ogawa K. Isolated avulsion fracture of the lesser tuberosity of the humerus. A case report. *Clin Orthop Relat Res*. 1986;(211):215-218.
39. Sharma A, Jindal S, Narula MS, Garg S, Sethi A. Bilateral Asymmetrical Fracture Dislocation of Shoulder with Rare Combination of Injuries after Epileptic Seizure: A Case Report. *Malaysian Orthop J*. 2017;11(1):74-76.
40. Schweighofer F, Schippinger G, Peicha G. Posterior dislocation fracture of the shoulder. *Chirurg*. 1996;67(12):1251-1254.
41. Schweighofer F, Peicha G, Boldin C, Fankhauser F. Posterior fracture-dislocation of the shoulder. *Eur J Trauma*. 2005;31(6):551-556.
42. Schiltz M, De Baere T. Isolated avulsion fracture of the humeral lesser tuberosity in an adolescent tennis player: A case report and review of the literature. *J Traumatol du Sport*. 2015;32(3):133-136.
43. Scheibel M, Martinek V, Imhoff AB. Arthroscopic reconstruction of an isolated avulsion fracture of the lesser tuberosity. *Arthroscopy*. 2005;21(4):487-494.

44. Ross GJ, Love MB. Isolated avulsion fracture of the lesser tuberosity of the humerus: report of two cases. *Radiology*. 1989;172(3):833-834.
45. Reparaz-Padrós FJ, Garbayo-Marturet AJ, Fernández-Hortigüela L. Isolated avulsion fracture of the lesser tuberosity of the humerus: Two cases. *Rev Ortop Traumatol*. 2005;49(5):368-372.
46. Recht J, Docquier J, Soete P, Forthomme JP. Avulsion-fracture of the subscapular muscle. *Acta Orthop Belg*. 1991;57(3):312-316.
47. Provance AJ, Polousky JD. Isolated avulsion fracture of the subscapularis tendon with medial dislocation and tear of biceps tendon in a skeletally immature athlete: a case report. *Curr Opin Pediatr*. 2010;22(3):366-368.
48. Polousky JD, Harms S. Subscapularis tendon injuries in adolescents: a report of 2 cases. *J Pediatr Orthop*. 2011;31(5):e57-9.
49. Patrizio L, Sabetta E. Acute posterior shoulder dislocation with reverse hill-sachs lesion of the epiphyseal humeral head. *ISRN Surg*. 2011;2011:851051.
50. Pace A, Ribbans W, Kim JH. Isolated lesser tuberosity fracture of the humerus. *Orthopedics*. 2008;31(1):94.
51. Ohzono H, Gotho M, Mitsui Y, et al. Isolated Fracture of the Lesser Tuberosity of the Humerus: A Case Report. *Kurume Med J*. 2011;58(4):131-133.
52. Nove-Josserand, Levigne C, Noel E, Walch G. Les fractures du trochin chez l'adulte. Diagnostic, traitement. A propos de 17 cas. Fractures of the lesser tuberosity of the humerus in adults. Diagnosis and treatment in 17 cases. *J Traumatol du Sport*. 1995;12(4):213-217.
53. Nikolaou VS, Chytas D, Tyrpenou E, Babis GC. Two-level reconstruction of isolated fracture of the lesser tuberosity of the humerus. *World J Clin Cases*. 2014;2(6):219-223.
54. Neogi DS, Bejjanki N, Ahrens PM. The consequences of delayed presentation of lesser tuberosity avulsion fractures in adolescents after repetitive injury. *J Shoulder Elbow Surg*. 2013;22(4):e1-e5.
55. Nardo L, Ma BC, Steinbach LS. Lesser Tuberosity Avulsions in Adolescents. *HSS J*. 2014;10(3):201-207.
56. McAuliffe TB, Dowd GS. Avulsion of the subscapularis tendon. A case report. *J Bone Joint Surg Am*. 1987;69(9):1454-1455.
57. Malone T, Mair S. Management of acute shoulder pain in an adolescent lacrosse athlete: a case report. *Int J Sports Phys Ther*. 2014;9(3):383-387.
58. Liu X, Zhu Y, Lu Y, Li F, Wu G, Jiang C. Locked Posterior Shoulder Dislocation Associated With Isolated Fractures of the Lesser Tuberosity: A Clinical Study of 22 Cases With a Minimum of 2-Year Follow-up. *J Orthop Trauma*. 2015;29(6):271-275.
59. Liong MF, Felix LYS, Amir Hamzah MSN. Isolated avulsion fractures of lesser tuberosity humerus: A case report. *Malaysian Orthop J*. 2017;11:Supplement A.
60. Leslie A, Cassar-Pullicino VN. Avulsion of the lesser tuberosity with intra-articular injury of the glenohumeral joint. *Injury*. 1996;27(10):742-745.
61. Le Huec JC, Schaeverbeke T, Moinard M, Kind M, Chauveaux D, Le Rebeller A. Isolated avulsion fracture of the lesser tubercle of the humerus in children. *Acta Orthop Belg*. 1994;60(4):427-429.
62. LaMont LE, Green DW, Altchek DW, Warren RF, Wickiewicz TL. Subscapularis tears and lesser tuberosity avulsion fractures in the pediatric patient. *Sports Health*. 2015;7(2):110-114.
63. LaBriola JH, Mohaghegh HA. Isolated avulsion fracture of the lesser tuberosity of the humerus. A case report and review of the literature. *J Bone Joint Surg Am*. 1975;57(7):1011.
64. Kuroda T, Go G, Ojima S, Nishi S, Mizuno K. Isolated avulsion fracture of the lesser tuberosity of the humerus: A case report. *J Shoulder Elbow Surg*. 1993;2(4):221-224.
65. Kumar V, Al-Couto J, Rangan A. Isolated avulsion fracture of the lesser tuberosity of the humerus associated with delayed axillary nerve neuropraxia. *Injury Extra*. 2006;37(1):31-33.

66. Klasson SC, Vander Schilden JL, Park JP. Late effect of isolated avulsion fractures of the lesser tubercle of the humerus in children. Report of two cases. *J Bone Joint Surg Am*. 1993;75(11):1691-1694.

67. Kato S, Funasaki H, Kan I, Yoshida M, Kasama K, Marumo K. Incomplete joint side tear of the subscapularis tendon with a small fragment in an adolescent tennis player: a case report. *Sport Med Arthrosc Rehabil Ther Technol*. 2012;4(1):24.

68. Kanso I, Bricout JM. Isolated avulsion fracture of the lesser tuberosity of the humerus. Apropos of a case. *Rev Chir Orthop Reparatrice Appar Mot*. 1998;84(6):554-557.

69. Jariwala A, Haines S, McLeod G. "Locked" posterior dislocation of the shoulder with communition of the lesser tuberosity: A stabilisation technique. *Eur J Orthop Surg Traumatol*. 2008;18(5):377-379.

70. Hung LH, Chung KY, Tang N, Leung KS. Isolated Avulsion Fracture of the Lesser Tuberosity of the Humerus: Case Report and Literature Review. *J Orthop Trauma Rehabil*. 2012;16(2):78-81.

71. Heyworth BE, Dodson CC, Altchek DW. Arthroscopic repair of isolated subscapularis avulsion injuries in adolescent athletes. *Clin J Sport Med*. 2008;18(5):461-463.

72. Hayes PR, Klepps S, Bishop J, Cleeman E, Flatow EL. Posterior shoulder dislocation with lesser tuberosity and scapular spine fractures. *J Shoulder Elbow Surg*. 2003;12(5):524-527.

73. Hackl M, Moro F, Durchholz H. Combined displaced fracture of the lesser humeral tuberosity and the scapular spine: A case report. *Int J Surg Case Rep*. 2015;13:106-111.

74. Haas SL. Fracture of the lesser tuberosity of the humerus. *Am J Surg*. 1944;63(2):253-256.

75. Gornitzky AL, Potty AG, Carey JL, Ganley TJ. Repair of Acute-on-Chronic Subscapularis Insufficiency in an Adolescent Athlete. *Orthopedics*. 2015;38(9):e844-848.

76. Fabis J, Kozlowski P. The results of treatment of posterior shoulder dislocation. *Chir Narzadow Ruchu Ortop Pol*. 1998;63(5):455-461.

77. Earwaker J. Isolated avulsion fracture of the lesser tuberosity of the humerus. *Skeletal Radiol*. 1990;19(2):121-125.

78. Dhawan A, Kirk K, Dowd T, Doukas W. Isolated avulsion fracture of the lesser tuberosity of the humerus in an adult: case report and literature review. *Am J Orthop*. 2008;37(12):627-630.

79. Collier SG, Wynn-Jones CH. Displacement of the biceps with subscapularis avulsion. *J Bone Joint Surg Br*. 1990;72(1):145.

80. Biedert RM, Maître T. Conservative treatment of isolated fractures of the lesser tuberosity of the humerus. *Schweizerische Zeitschrift für Sport und Sport*. 2000;48(4):158-160.

81. Berbig R, Keller H, Metzger U. Isolated fracture of the lesser tuberosity of the humerus: case reports and review of the literature. *Z Unfallchir Versicherungsmed*. 1994;87(3):159-168.

82. Becker R, Weyand F. Rare, bilateral posterior shoulder dislocation. A case report. *Unfallchirurg*. 1990;93(2):66-68.

83. Andreasen A. Avulsion Fracture of Lesser Tuberosity of Humerus: report of a case. *Lancet*. 1948;251(6507):750-751.

84. Aagaard K, Lunsjö K. Occult fracture of the lesser tuberosity in a 9-year-old female swimmer. *J Surg Case Reports*. 2017;2017(1):rjw238.

85. Edwards TB. What is the value of a systematic review? *J Shoulder Elbow Surg*. 2014;23(1):1-2.

86. Wilkins KE. Principles of fracture remodeling in children. *Injury*. 2005;36 Suppl 1:A3-11.

87. Cottias P. Fractures of the lesser tuberosity of the humerus. *Rev Chir Orthop Reparatrice Appar Mot*. 1998;84:137-139.

88. Bock P, Kluger R, Hintermann B. Anatomical reconstruction for Reverse Hill-Sachs lesions after posterior locked shoulder dislocation fracture: A case series of six patients. *Arch Orthop Trauma Surg*. 2007;127(7):543-548.

89. Belanger V, Dupuis F, Leblond J, Roy JS. Accuracy of examination of the long head of the biceps tendon in the clinical setting: A systematic review. *J Rehabil Med*. 2019;51(7):479-491.

SUPPLEMENTARY MATERIAL

Supplement 1. Search strategy

Medline/Ovid, 14-10-2019

#	Query	Results
1	Fractures, Avulsion/	122
2	"avulsion*".ab,kf,ti.	9479
3	1 or 2	9494
4	exp humerus/ or exp humeral fractures/ or exp shoulder fractures/	19191
5	"humer*".ab,kf,ti.	25678
6	4 or 5	31743
7	3 and 6	399
8	((Lesser or minor or minus) adj3 (tuberosit* or tubercle* or tuberculum)).ab,kf,ti.	418
9	(subscapular* and avulsion*).ab,kf,ti.	81
10	7 or 8 or 9	804
11	exp editorial/	504717
12	10 not 11	803

Embase.com, 14-10-2019

#	Query	Results
#7	(#3 OR #4 OR #5) NOT [editorial]/lim	1061
#6	#3 OR #4 OR #5	1064
#5	('subscapularis muscle'/exp OR subscapular*:ab,ti) AND avulsion*:ab,ti	86
#4	((lesser OR minor OR minus) NEAR/3 (tuberosit* OR tubercle* OR tuberculum)):ab,ti	503
#3	#1 AND #2	571
#2	'humerus fracture'/exp OR 'shoulder fracture'/exp OR 'humerus'/exp OR humer*:ab,ti	39342
#1	'avulsion injury'/exp OR avulsion:ab,ti	12767

Cinahl/Ebsco, 14-10-2019

#	Query	Results
S11	S10 NOT PT Editorial	235
S10	S7 OR S8 OR S9	236
S9	TI (subscapular* AND avulsion*) OR AB (subscapular* AND avulsion*)	25
S8	TI (((lesser OR minor OR minus) N3 (tuberosit* OR tubercle* OR tuberculum))) OR AB (((lesser OR minor OR minus) N3 (tuberosit* OR tubercle* OR tuberculum)))	123
S7	S3 AND S6	115
S6	S4 OR S5	7,066
S5	TI humer* OR AB humer*	5,672
S4	(MH "Humerus") OR (MH "Humeral Fractures+") OR (MH "Shoulder Fractures+")	4,486
S3	S1 OR S2	2,031
S2	TI avulsion* OR AB avulsion*	1,977
S1	(MH "Avulsion Fractures")	199

The Cochrane Library for CENTRAL and CDSR, 14-10-2019

#	Query	Results
#1	(avulsion* AND humer*):ti,ab,kw	7
#2	subscapular* AND avulsion*:ti,ab,kw	1
#3	(((lesser OR minor OR minus) NEAR/3 (tuberosit* OR tubercle* OR tuberculum))):ti,ab,kw	21
#4	{OR #1-#3} in Cochrane Reviews	0
#5	{OR #1-#3} in Trials	28

SPORTDiscus/Ebsco, 14-10-2019

#	Query	Results
S4	S1 OR S2 OR S3	117
S3	TI (subscapular* AND avulsion*) OR AB (subscapular* AND avulsion*)	16
S2	TI (((lesser OR minor OR minus) N3 (tuberosit* OR tubercle* OR tuberculum))) OR AB (((lesser OR minor OR minus) N3 (tuberosit* OR tubercle* OR tuberculum)))	55
S1	(DE "AVULSION fractures" OR TI avulsion* OR AB avulsion*) AND (DE "HUMERUS" OR DE "SHOULDER injuries" OR DE "HUMERUS injuries" OR TI humer* OR AB humer*)	61

Web of Science, 14-10-2019

#	Query	Results
# 6	(#4) NOT (#5) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*	796
# 5	(#4) *AND* DOCUMENT TYPES: (Editorial Material) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*	11
# 4	#3 OR #2 OR #1 *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*	807
# 3	TOPIC: (((lesser OR minor OR minus) NEAR/3 (tuberosit* OR tubercle* OR tuberculum))) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*	460
# 2	TOPIC: (subscapular* AND avulsion*) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*	78
# 1	TOPIC: (avulsion* AND humer*) *Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=All years*	373

Scopus, 14-10-2019

#	Query	Results
#1	(TITLE-ABS-KEY ((avulsion* AND humer*)) OR TITLE-ABS-KEY ((subscapular* AND avulsion*)) OR TITLE-ABS-KEY (((lesser OR minor OR minus) W/3 (tuberosit* OR tubercle* OR tuberculum)))) AND (EXCLUDE (DOCTYPE , “ed”))	1118

Clinicaltrials.gov, 14-10-2019

#	Query	Results
#1	(tuberosity OR tubercle OR tuberculum) OR (subscapularis AND avulsion) OR (humerus AND avulsion) OR (humeral AND avulsion)	15

WHO ICTRP, 14-10-2019

#	Query	Results
#1	tuberosit* OR tubercle* OR tuberculum	84
#2	avulsion* AND humer*	1
#3	#1 OR #2	84

Supplement 2. Quality assessment of case reports using the tool suggested by Murad et al

Year	First author	1	2	3	4	5	Overall judgement
2017	Aagaard	no	yes	yes	no	yes	Fair
1948	Andreasen	yes	yes	yes	no	yes	Good
2017	Atahnasiadis	no	yes	yes	no	yes	Fair
1990	Becker	no	yes	yes	no	no	Poor
1994	Berbig	no	yes	yes	no	yes	Fair
2000	Biedert	no	yes	yes	no	yes	Fair
1996	Caniggia	no	yes	yes	no	yes	Fair
1990	Collier	no	yes	yes	no	no	Poor
2008	Dhawan	no	yes	yes	no	yes	Fair
1990	Earwaker	no	yes	yes	no	yes	Poor
1998	Fabis	no	yes	yes	no	no	Poor
2012	Goeminne	no	yes	yes	yes	yes	Good
2015	Gornitzky	no	yes	yes	no	yes	Fair
1944	Haas	no	yes	yes	no	yes	Fair
2015	Hackl	no	yes	yes	no	yes	Fair
2003	Hayes	no	yes	yes	yes	yes	Good
2008	Heyworth	no	yes	yes	yes	yes	Good
2012	Hung	no	yes	yes	no	yes	Fair
2008	Jariwala	no	yes	yes	no	yes	Fair
2019	Jaya	no	yes	yes	no	yes	Fair
1998	Kanso	no	yes	yes	no	yes	Fair
2012	Kato	no	yes	yes	yes	yes	Good
2014	Kim	no	yes	yes	no	yes	Fair
1993	Klasson	no	yes	yes	no	yes	Fair
2007	Kowalsky	no	yes	yes	no	yes	Fair
2006	Kumar	no	yes	no	no	yes	Poor
1993	Kunkel	no	yes	yes	no	yes	Fair
1993	Kuroda	no	yes	yes	no	yes	Fair
1975	LaBriola	no	yes	yes	no	yes	Fair
1994	LeHuec	no	yes	yes	no	yes	Fair
1997	Leslie	no	yes	yes	yes	no	Fair
2005	Levine	no	yes	yes	no	yes	Fair
2017	Liong	no	no	no	no	no	Poor
2014	Malone	no	yes	yes	no	yes	Fair

Supplement 2. Quality assessment of case reports using the tool suggested by Murad et al (continued)

Year	First author	1	2	3	4	5	Overall judgement
1987	McAuliffe	no	yes	yes	no	yes	Fair
2013	Neogi	no	yes	yes	no	yes	Fair
2014	Nikolaou	no	yes	yes	no	yes	Fair
2011	Ohzono	no	yes	yes	yes	no	Fair
2008	Pace	no	yes	yes	no	yes	Fair
1995	Paschal	no	yes	yes	no	yes	Fair
2011	Patrizio	no	yes	yes	no	Yes	Fair
2013	Paxinos	no	yes	yes	yes	yes	Good
2011	Polousky	no	yes	yes	no	yes	Fair
2010	Provance	no	yes	yes	no	yes	Fair
1991	Recht	no	yes	yes	no	yes	Fair
2004	Reparaz	no	yes	yes	no	yes	Fair
1989	Ross	no	yes	yes	no	yes	Fair
2009	Sarraf	no	yes	yes	no	yes	Fair
2005	Scheibel	no	yes	yes	no	yes	Fair
2015	Schiltz	no	yes	yes	no	yes	Fair
2017	Sharma	no	yes	yes	no	yes	Fair
1984	Shibuya	no	yes	yes	yes	yes	Good
2004	Sikka	no	yes	yes	no	yes	Fair
2004	Sugalski	no	yes	yes	no	yes	Fair
2017	Tabrizi	no	yes	yes	no	yes	Fair
2012	Teixeira	no	yes	yes	yes	yes	Good
1992	Thieleman	no	yes	yes	no	yes	Fair
2011	Tosun	no	yes	yes	yes	yes	Good
1985	White	no	yes	yes	no	no	Poor
1993	Zanlungo	no	yes	yes	no	yes	Fair
2018	Zimmerman	no	yes	yes	no	yes	Fair

The following domains were assessed: selection (1), ascertainment of exposure (2), ascertainment of outcome (3), length of follow-up (4) and reporting (5). If patients had less than 2 years of follow-up, this domain was judged negative. Records were judged ''good'' if there was 0 or 1 negative answer within all domains. Records were judged ''fair'' if 2 answers were negative and ''poor'' if 3 or more answers were negative.

Supplement 3. Quality assessment of case series using the Newcastle-Ottawa Scale for cohort studies

Year	First author	1	2	3	4	5	6	Overall judgement
2013	Garrigues	1	1	1	1	1	1	Good
2014	LaMont	1	0	0	0	0	1	Fair
2015	Liu	1	1	1	0	1	1	Good
2014	Nardo	1	1	1	0	0	1	Fair
1995	Nové-Josserand	1	1	0	0	1	0	Fair
1997	Ogawa	1	1	1	0	1	1	Good
2005	Schweighofer	1	0	1	0	1	1	Good
1995	van Laarhoven	1	0	0	0	0	1	Fair
2011	Vezeridis	1	1	0	1	1	1	Good
2013	Weiss	1	0	0	0	0	1	Fair

The following domains were assessed: representativeness (1), ascertainment of exposure (2), outcome present at the start of the study (3), assessment of outcome (4), follow-up length (5) and adequacy of follow-up (6).
Records were judged ''good'' if there were 2 or 3 stars in the selection domain and 2 or 3 stars in the outcome domain. Records were judged ''fair'' if there was 1 star in the selection domain and 1 star in the outcome domain. If there were 0 stars in the selection or outcome domain, records were judged ''poor''.

Supplement 4. Studies describing paediatric patients

First author	Year	Country	n
Aagaard	2017	Sweden	1
Garrigues	2013	USA	5
Goeminne	2012	Belgium	3
Gornitzky	2015	USA	1
Heyworth	2008	USA	3
Kato	2012	Japan	1
Klasson	1993	USA	2
Kunkel	1993	USA	1
LaMont	2014	USA	5
LeHuec	1994	France	1
Levine	2005	USA	1
Malone	2014	USA	1
Nardo	2014	USA	6
Neogi	2013	UK	2
Ogawa	1997	Japan	10
Paschal	1995	USA	2
Paxinos	2013	Greece	1
Polousky	2011	USA	2
Provance	2010	USA	1
Ross	1989	USA	2
Schiltz	2015	Belgium	1
Shibuya	1984	Japan	1
Sikka	2004	USA	1
Sugalski	2004	USA	1
Teixeira	2012	USA	1
Vezeridis	2011	USA	8
Weiss	2013	USA	2
White	1985	USA	1
Zimmerman	2018	Germany	1

Supplement 5. Studies describing patients with a PSD

First author	Year	Country	n
Becker	1990	Germany	1
Fabis	1998	Poland	1
Hayes	2003	USA	1
Jariwala	2008	Scotland	1
Liu	2015	China	22
Patrizio	2011	Italy	1
Sarraf	2009	UK	2
Schweighofer	2005	Germany	16
Sharma	2017	India	1
Tabrizi	2017	Iran	1

Supplement 6. Studies describing adult patients

First author	Year	Country	n
Andreasen	1948	India	1
Atahnasiadis	2017	Greece	1
Berbig	1994	Germany	3
Biedert	2000	Switzerland	1
Caniggia	1996	Italy	2
Collier	1990	UK	1
Dhawan	2008	USA	1
Earwaker	1990	Australia	2
Haas	1944	USA	1
Hackl	2015	Germany	1
Hung	2012	China	1
Jaya	2019	Malaysia	1
Kanso	1998	France	1
Kim	2014	Korea	1
Kowalsky	2007	USA	1
Kumar	2006	UK	1
Kuroda	1993	Japan	1
LaBriola	1975	USA	1
LaMont	2014	USA	5
Leslie	1997	UK	1
Liong	2017	Malaysia	1
McAuliffe	1987	UK	1
Nikolaou	2014	Greece	1
Nové-Josserand	1995	France	17
Ogawa	1997	Japan	10
Ohzono	2011	Japan	1
Pace	2008	UK	1
Recht	1991	Belgium	1
Reparaz	2004	Spain	2
Scheibel	2005	Germany	1
Thieleman	1992	Germany	2
Tosun	2011	Turkey	2
van Laarhoven	1995	Netherlands	6
Zanlungo	1993	Italy	1

Supplement 7. Sensitivity analysis for the outcomes of the paediatric, adult and PSD group

	Unadjusted regression		Adjusted regression	
Variable	**B**	***p*-value**	**B**	***p*-value**
Pain				
Adults	0.46	0.54	22.92	1.00
PSD	1.43	0.18	61.95	1.00
Mean VAS pain				
PSD	0.16	0.26	0.16	0.26
Restricted movement				
Adults	0.61	0.22	-0.49	0.60
PSD	3.47	<0.001	1.83	0.09
Full strength				
Adults	0.49	0.68	18.09	1.00
Mean of PROMs				
Adults	-5.00	0.001	-1.63	0.43
PSD	-11.30	<0.001	-6.89	0.03
Non-union				
Adults	-0.11	0.90	19.15	1.00
Complications				
Adults	-0.37	0.64	-1.38	0.27
PSD	1.09	0.09	-1.10	0.41
Secondary surgery				
Adults	-0.77	0.53	-16.89	1.00
PSD	0.29	0.78	14.53	1.00

Results of sensitivity outcome analysis adjusted for country. Paediatric patients were used as reference group for the regression models. Abbreviations: B, unstandardized beta coefficient; PSD, posterior shoulder dislocation; PROMs, patient reported outcome measures.

Supplement 8. Sensitivity analysis for non-operative and surgical treatment in paediatric patients

	Unadjusted regression		Adjusted regression	
Variable	**B**	***p*-value**	**B**	***p*-value**
Restricted movement	-1.61	0.054	-1.18	0.28
Full strength	4.36	0.005	3.80	0.02
Mean of PROMs	11.22	0.004	3.33	0.46
Complications	-2.80	0.02	-19.83	1.00

Results of sensitivity outcome analysis adjusted for country.
Abbreviations: B, unstandardized beta coefficient; PROMs, patient reported outcome measures.

Supplement 9. Sensitivity analysis for non-operative and surgical treatment in adults

	Unadjusted regression		Adjusted regression	
Variable	**B**	***p*-value**	**B**	***p*-value**
Pain	0.62	0.61	18.14	1.00
Restricted movement	-1.37	0.09	-1.53	0.33
Mean of PROMs	4.58	0.20	4.99	0.09
Complications	-0.46	0.72	-19.30	1.00

Results of sensitivity outcome analysis adjusted for country. Abbreviations: B, unstandardized beta coefficient; PROMs, patient reported outcome measures.

Supplement 10. Sensitivity analysis for non-operative and surgical treatment in patients with a PSD

	Unadjusted regression		Adjusted regression	
Variable	**B**	***p*-value**	**B**	***p*-value**
Restricted movement	1.96	0.051	19.26	1.00
Mean of PROMs	-10.10	0.007	-10.10*	0.007*
Complications	0.03	0.97	-0.69	0.60

Results of sensitivity outcome analysis adjusted for country. Abbreviations: B, unstandardized beta coefficient; PROMs, patient reported outcome measures; PSD, posterior shoulder dislocation.
*The model failed to converge when adjusting for country. Therefore, the unadjusted values are presented.

Supplement 11. Sensitivity analysis for acute and delayed surgery in paediatric patients

	Unadjusted regression		Adjusted regression	
Variable	**B**	***p*-value**	**B**	***p*-value**
Mean VAS pain	0.63	0.26	0.63*	0.26*
Restricted movement	0.08	0.93	0.41	0.71
Mean of PROMs	-0.09	0.97	0.91	0.63

Results of sensitivity outcome analysis adjusted for country. Abbreviations: B, unstandardized beta coefficient; PROMs, patient reported outcome measures.
* The model failed to converge when adjusting for country. Therefore, the unadjusted values are presented.

Supplement 12. Sensitivity analysis for acute and delayed non-operative treatment in paediatric patients

	Unadjusted regression		Adjusted regression	
Variable	**B**	***p*-value**	**B**	***p*-value**
Pain	0.00	1.00	20.50	1.00
Restricted movement	0.00	1.00	21.20	1.00
Mean of PROMs	13.25	0.46	10.00	0.71
Non-union	-0.29	0.85	20.50	1.00
Complications	-0.18	0.90	19.62	1.00

Results of sensitivity outcome analysis adjusted for country. Abbreviations: B, unstandardized beta
coefficient: PROMs, patient reported outcome measures.

Supplement 13. Sensitivity analysis for acute and delayed surgery in adults

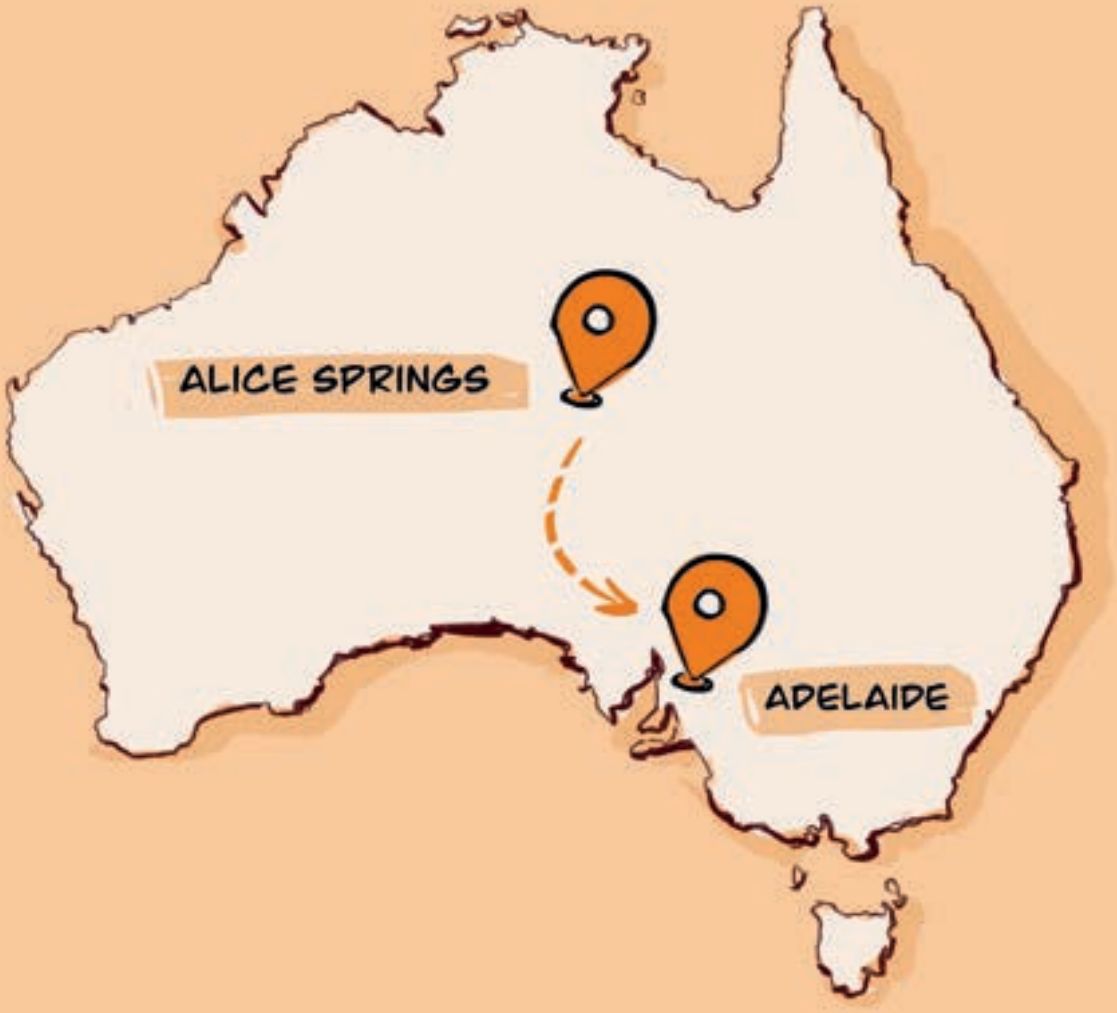
	Unadjusted regression		Adjusted regression	
Variable	**B**	***p*-value**	**B**	***p*-value**
Pain	1.39	0.24	-20.79	1.00
Restricted movement	-0.57	0.63	-1.55	0.33
Satisfaction	0.41	0.81	0.69	0.71
Complications	1.58	0.29	0	1.00

Results of sensitivity outcome analysis adjusted for country. Abbreviations: B, unstandardized beta. Coefficient.

Supplement 14. Sensitivity analysis for acute and delayed non-operative treatment in adults

	Unadjusted regression		Adjusted regression	
Variable	**B**	***p*-value**	**B**	***p*-value**
Mean of PROMs	7.00	0.36	7.00*	0.36*

Results of sensitivity outcome analysis adjusted for country. Abbreviations: B, unstandardized beta coefficient; PROMs, patient reported outcome measures.
* The model failed to converge when adjusting for country. Therefore, the unadjusted values are presented.


ALICE SPRINGS
ADELAIDE

7

Management of displaced humeral surgical neck fractures in daily clinical practice: hanging does not re-align the fracture

Reinier W.A. Spek
Lotje A. Hoogervorst
Michaëla E.C. Elias
Ruurd L. Jaarsma
DirkJan H.E.J. Veeger
Job N. Doornberg
Paul C. Jutte
Michel P.J. van den Bekerom

Archives of Orthopaedic and Trauma Surgery. 2023 Jun;143(6):3119-3128

ABSTRACT

Aims

It is unclear if the collar and cuff treatment improves alignment in displaced surgical neck fractures of the proximal humerus. Therefore, this study evaluated if the neck-shaft angle and extent of displacement would improve between trauma and onset of radiographically visible callus in non-operatively treated surgical neck fractures (Boileau type A, B, C).

Methods

A consecutive series of patients (≥18 years old) were retrospectively evaluated from a Level 1 trauma centre in Australia (inclusion period: 2016 - 2020) and a Level 2 trauma centre in the Netherlands (inclusion period: 2004 to 2018). Patients were included if they sustained a Boileau-type fracture and underwent initial non-operative treatment. The first radiograph had to be obtained within 24 h after the initial injury and the follow-up radiograph(s) 1 week after trauma and before the start of radiographically visible callus. On each radiograph, the maximal medial gap (MMG), maximal lateral gap (MLG), and neck-shaft angle (NSA) were measured. Linear mixed modelling was performed to evaluate if these measurements would improve over time.

Results

Sixty-seven patients were included: 25 type A, 11 type B, and 31 type C fractures. The mean age (range) was 68 years (24 - 93), and the mean number (range) of follow-up radiographs per patient was 1 (1 - 4). Linear mixed modelling on both MMG and MLG revealed no improvement during follow-up among the three groups. Mean NSA of type A fractures improved significantly from 161° at trauma to 152° at last follow-up (*p*-value = 0.004).

Conclusion

Apart from humeral head angulation improvement in type A, there is no increase nor reduction in displacement among the three fracture patterns. Therefore, it is advised that surgical decision-making should be performed immediately after trauma.

INTRODUCTION

In displaced surgical neck fractures of the humerus, it is not well understood which fracture patterns would respond best to non-operative treatment and which ones would require surgical fixation [1]. If non-operative treatment is chosen, patients are advised to wear a collar and cuff with their arm in internal rotation and the humeral shaft in line with the humeral head. In this position, while holding the body upright, traction is generated due to gravity, allowing the shaft to realign with the proximal humerus [2]. However, re-alignment may not occur in each type of fracture and if it fails, surgical management can be required to avoid mal- or non-union.

Besides biomechanical forces (e.g., muscles and bone-on-bone friction) when wearing a collar and cuff, there may be a relationship between fracture pattern and alignment. To date, few studies have evaluated radiographic outcomes in non-surgically treated proximal humerus fractures. One study revealed that radiographic angulation on lateral views after 1 week could predict outcomes in minimally displaced proximal humerus fractures [3]. However, it remained unclear if the collar and cuff treatment would improve angulation and shaft translation in fractures with ≥1 cm of displacement. A French study conducted by Boileau et al. recently classified surgical neck fractures into three categories: type A, B and C (Table 1, Fig. 1). Considering the surgical nature of this work, we hypothesized that hanging down the arm in a collar and cuff (as applied in current clinical practice) would not re-align these three fracture patterns [4]. The aim of this study was to assess (1) if the neck-shaft angle and extent of displacement would improve between trauma and onset of radiographically visible callus in non-operatively treated surgical neck fractures (Boileau type A, B, C), and (2) if there would be a difference in displacement and humeral head tilt between type A, B or C.

Table 1. Fracture patterns according to Boileau's classification

	Humeral shaft translation	Humeral head position
Type A	Partially medial	Valgus
Type B	Entirely medial and/or ventral	Neutral
Type C	Partially lateral	Varus

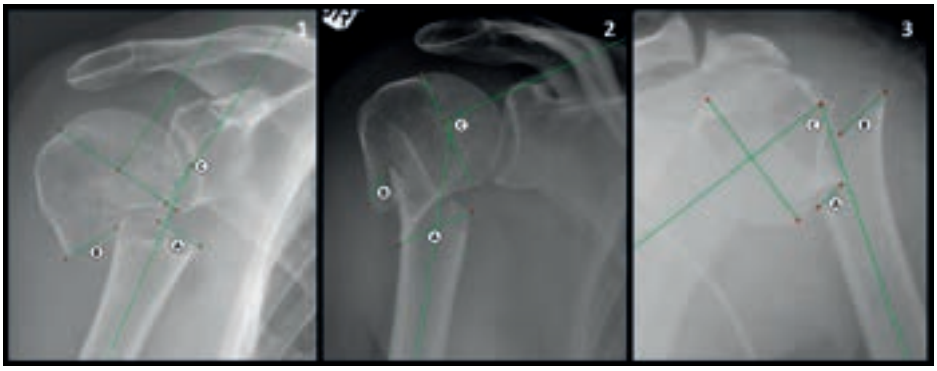


Figure 1. 1 = type A (partial medial shaft displacement with valgus angulation), 2 = type B (entire medial and ventral shaft displacement without humeral head tilt), 3 = type C (lateral shaft displacement with varus angulation). Three parameters were measured on each radiograph: A = maximal medial gap, B = maximal lateral gap, C = neck-shaft angle.

PATIENTS AND METHODS

Setting and study design

This retrospective imaging study was carried out at in a Level 2 trauma centre in the Netherlands and a Level 1 trauma centre in Australia. Ethical approval was received in both centers in compliance with their local institutional review boards.

Screening

In the Dutch Hospital, patients were included between January 1, 2004, and June 30, 2018. The inclusion period from the Australian trauma centre was from March 1, 2016 to July 31, 2020. All surgical neck fractures within this period were screened and categorized according to Boileau's classification system [4]. Screening and classification were performed independently and in duplicate by the first three authors. Discrepancies were resolved by discussion. If consensus could not be achieved, one of the surgeons in our author group was consulted.

Study population

Patients (≥18 years) with an type A, B or C (according to Boileau's classification) isolated displaced surgical neck fracture were included in the study [4,5]. Patients with pathologic surgical neck fractures, undeterminable humeral head angulation on trauma radiographs, concomitant fractures (large Hill-Sachs lesions, greater tuberosity fractures with footprint defects, humeral shaft-, clavicle-, and acromion

fractures), and patients who underwent surgery before day 8 after initial trauma were excluded. Patients were required to have an anteroposterior (AP) radiograph obtained within 24 h after the initial injury and at least one follow-up AP radiograph while following a non-operative treatment protocol. Follow-up radiographs needed to be available at least one week after the initial trauma and before the start of the radiographically visible callus (Supplement 1).

Classification

Boileau et al. classified surgical neck fractures into three categories (Table 1, Fig. 1): (type A): partial medial shaft displacement with valgus angulation of the humeral head (shoulder adductor muscles, predominantly the pectoralis major- and latissimus dorsi muscle, pull the shaft medially resulting in humeral head tilt to the contralateral side), (type B) entire medial and ventral shaft displacement without humeral head tilt, (type C) lateral shaft displacement with varus angulation of humeral head (shoulder abductor muscles, acromial part of deltoid and biceps brachii, pull the shaft laterally and supraspinatus muscle pulls head in further varus tilt). As the original classification article did not specify displacement, we used a displacement cut-off of ≥25% of the humeral shaft diameter. Patients were categorized as type B if they had complete shaft translation in any direction (as opposed to the medial and anterior translation described by Boileau et al.). Dorsal or ventral head angulation was not taken into account: if a patient had medial shaft translation with valgus and dorsal head deformity, the patient was still categorized as type A. If fracture patterns contradicted Boileau's criteria, they were categorized into the miscellaneous category "unclassifiable" and excluded for further analysis. For example, if there was partial shaft translation without humeral head angulation.

Hospital treatment protocol and variables

Routine assessment of patients with a displaced surgical neck fracture in both hospitals included physical examination and radiographic imaging. If non-operative treatment was followed, patients were provided with a collar and cuff with the arm in adduction (elbow and forearm act as a weight to provide traction) and shoulder movements were allowed as tolerated by pain. Surgical decision-making was based on patient comorbidities and fracture patterns. The following variables were collected for each patient: age, gender, date of hospital admission, side of the fracture, days from initial injury to first trauma radiograph, type of treatment, type of surgical treatment, time between injury and surgery, presence of comminution, number of follow-up radiographs, and time from trauma to each radiograph.

Outcome measures

As the Boileau classification is based on deformities in the frontal plane, only radiographic parameters were measured on anteroposterior (AP) radiographic views. The following parameters were measured on each trauma and follow-up radiograph: maximal medial gap (MMG), maximal lateral gap (MLG), and neck-shaft angle (NSA) (Fig. 1). The MMG was defined as the maximal distance between the medial tip of the surgical neck and the edge of the fracture on the inferior humeral head on the medial side, the MLG as the maximal distance between the lateral tip of the surgical neck and the edge of the fracture on the inferior humeral head on the lateral side. MMG, MLG and gap were all evaluated in millimetre (mm) and measured between both outer cortices. The NSA was calculated by drawing a line through the middle of the humeral shaft (bisector), the anatomic neck and a line perpendicular to the anatomic neck. The NSA represented the angle between the bisector and the line perpendicular to the anatomic neck. Measurements of radiographs in the Dutch Hospital were performed using Agfa Health Care (Agfa-Gevaert Group, Mortsel, Belgium) and in the Australian Hospital with RadiAnt DICOM Viewer (Medixant, Poznan, Poland) [6]. All measurements were performed by one assessor (first or second author).

Statistical analysis

Data analyses were performed using IBM SPSS software version 27 (IBM Corp., Armonk, N.Y., USA). Categorical baseline characteristics were presented in numbers with percentages and continuous baseline variables with mean and range depending on the distribution. To assess if MML, MMG and NSA would improve over time in each Boileau type, linear mixed modelling (LMM) was conducted. This model was run separately for each fracture pattern and each outcomes measure, and included time as a co-variate with MML, MMG or NSA as a dependent variable. A random intercept was used, and time slopes were assumed to be fixed. Linear mixed modelling (LMM) was again performed to determine if there was a difference of MML and MMG between the three fracture types. This model contained MML or MLG as dependent variables, time as co-variate and Boileau classification as a factor. Fixed effects estimate (fee) was reported together with a 95% confidence interval (CI) and *p*-value. A *p*-value less than 0.05 was considered significant. Further to this, the MMG, MLG and NSA at trauma were compared between in- and excluded patients using an independent samples t-test (Supplement 2).

RESULTS

A total of 2706 patients were screened for eligibility of which 614 patients had a displaced or undisplaced surgical neck fracture (most common reason for exclusion was the presence of a concomitant proximal humerus fracture such as a tuberosity fracture). Amongst these 614 surgical neck fractures, we identified 121 patients with a Boileau fracture: 41 type A, 20 type B, and 60 type C fractures. Only 1 (2.4%) patient in type A underwent surgery without eligible follow-up radiographs, 5 patients (25.0%) in type B and 7 patients (11.9%) in type C. After assessment against exclusion criteria a cohort of 67 patients was included for further analysis: 25 with type A, 11 with type B, and 31 with type C (Fig. 2). Mean age (range) of the cohort was 68.4 years (24 - 94), the majority were females (62.7%), and the mean number (range) of follow-up radiographs per patient was 1.3 (1 - 4). Surgical intervention was mostly performed in patients with type B fractures (36.4%) and surgical neck comminution did not differ between the three groups (Table 2).

Table 2. Baseline demographics

	All (n = 67)	A (n = 25)	B (n = 11)	C (n = 31)
Age (years)	70 (24 - 93)	78 (42 - 93)	78 (59 - 83)	62 (24 - 91)
Gender				
Female	42 (63%)	20 (80%)	10 (91%)	12 (39%)
Male	25 (37%)	5 (20%)	1 (9%)	19 (61%)
Days to presentation	0 (0 - 1)	0 (0 - 1)	0 (0 - 0)	0 (0 - 1)
Hospital				
Dutch	45 (67%)	17 (68%)	8 (73%)	20 (65%)
Australian	22 (33%)	8 (32%)	3 (27%)	11 (36%)
Right sided fracture	35 (52%)	18 (72%)	5 (46%)	12 (39%)
Comminuted fracture	12 (18%)	5 (20%)	2 (18%)	5 (16%)
Surgical management	12 (18%)	5 (20%)	4 (36%)	3 (10%)
ORIF	9 (13%)	4 (16%)	2 (18%)	3 (10%)
Nail	3 (5%)	1 (4%)	2 (18%)	0 (0%)
Days until surgery	20.5 (12 - 200)	19 (12 - 32)	18 (15 - 36)	55 (30 - 200)
Radiographs per patient	1 (1 - 4)	1 (1 - 3)	1 (1 - 2)	1 (1 - 4)
Days to fu radiograph(s)	14 (8 - 134)	14.5 (8 - 72)	11 (8 - 31)	15 (8 - 134)

Data is presented as median (range) or number (%). Abbreviations; fu, follow-up; ORIF, open reduction and internal fixation.

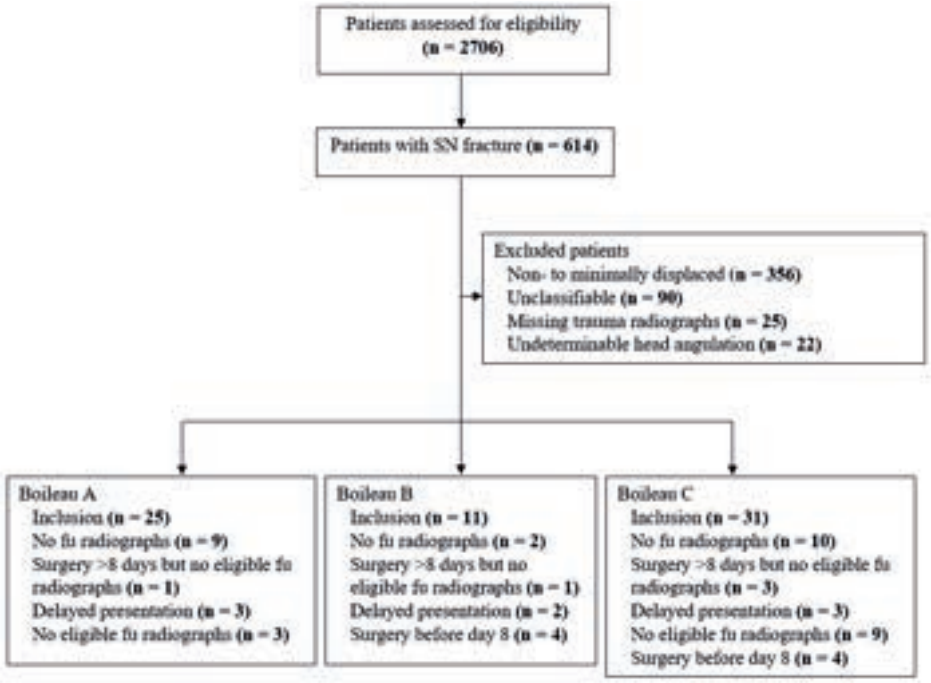


Figure 2. Breakdown of patients screened for a Boileau fracture. Abbreviations: SN, surgical neck, fu, follow-up.

Overall, there was a minimal effect and no significant improvement over time for maximal medial and maximal lateral gap in each fracture type. Mean maximal medial gap in type A fractures changed from 11 mm at trauma to 10 mm at ≥22 days follow-up (fee: 0.004, 95% CI: -0.06 to 0.07, *p*-value = 0.89), in type B fractures from 29 mm at trauma to 35 mm at ≥22 days follow-up (fee: 0.125, 95% CI: -0.36 to 0.61, *p*-value = 0.59), and in type C fractures from 14 mm to 10 mm at ≥22 days follow-up (fee: -0.003, 95% CI: -0.09 to 0.08, *p*-value = 0.94) (Table 3).

Mean maximal lateral gap in type A was 14 mm at trauma and 13 mm at ≥22 days follow-up (fee: 0.012, 95% CI: -0.05 to 0.07, *p*-value = 0.71). MLG in type B improved from 23 mm to 20 mm (fee: 0.126, 95% CI: -0.22 to 0.47, *p*-value = 0.44) and in type C from 9 mm to 6 mm (fee: -0.022, 95% CI: -0.09 to 0.05, *p*-value = 0.51): trauma *versus* ≥22 days follow-up, respectively (Table 4).

Table 3. Displacement (maximal medial gap) over time per fracture type

	A	B	C
Trauma	11.3 (2.1 - 27.9)	29.1 (13.7 - 43.8)	13.6 (3.5 - 31.1)
8 - 14 days	14.5 (0.0 - 30.1)	31.3 (10.6 - 44.7)	12.9 (0.0 - 38.3)
15 - 21 days	14.1 (2.7 - 25.9)	27.6 (18.7 - 41.2)	12.4 (3.4 - 25.5)
≥22 days	9.9 (3.6 - 14.5)	35.1 (29.9 - 40.3)	9.9 (0.0 - 26.2)
	p-value = 0.89	*p*-value = 0.59	*p*-value = 0.94

The maximal medial gap (distance between the medial tip of the surgical neck and medial fracture edge on the humeral head) was presented as the mean (range) in millimetre. *p*-values were obtained from linear mixed modelling with maximal medial gap as a dependent variable and time as co-variate.

Table 4. Displacement (maximal lateral gap) over time per fracture type

	A	B	C
Trauma	13.5 (5.2 - 27.7)	22.9 (11.4 - 41.4)	8.5 (2.0 - 27.0)
8 - 14 days	18.4 (4.0 - 34.7)	26.4 (14.3 - 33.5)	9.7 (0.0 - 30.8)
15 - 21 days	16.0 (1.9 - 29.2)	31.3 (26.9 - 38.2)	6.2 (0.0 - 12.7)
≥22 days	13.3 (3.6 - 26.3)	19.8 (10.8 - 28.8)	6.4 (0.0 - 14.3)
	p-value = 0.71	*p*-value = 0.44	*p*-value = 0.51

The maximal lateral gap (distance between the lateral tip of the surgical neck and lateral fracture edge on the humeral head) was presented as mean (range) in millimetre. *p*-values were obtained from linear mixed modelling with a maximal medial gap as dependent variable and time as co-variate.

Except for type A fractures, neck-shaft angle did not improve significantly over time. Mean neck-shaft angle of type A improved from 161° to 152° at the last follow-up time frame. LMM revealed a significant relationship of NSA over time with a corresponding fee of -0.28 (95% CI: -0.46 to -0.09, *p*-value = 0.004) (Fig. 3). For type B fractures, mean NSA at trauma was 135° and the head remained in a neutral position until callus was visible on radiographs (fee: -0.30, 95% CI: -1.55 to 0.96, *p*-value = 0.62) (Fig. 4). Humeral head of type C remained in varus deformity during follow-up moments (fee: 0.01, 95% CI: -0.12 to 0.13, *p*-value = 0.93) (Table 5) (Fig. 5 - 8).

Table 5. Neck-shaft angle over time per fracture type

	A	B	C
Trauma	161.1 (146.0 - 179.4)	134.8 (110.7 - 150.6)	111.7 (69.4 - 142.0)
8 - 14 days	152.7 (128.9 - 175.4)	127.0 (98.5 - 152.2)	112.9 (83.0 - 151.8)
15 - 21 days	144.8 (124.4 - 178.4)	141.6 (109.5 - 161.6)	109.1 (90.9 - 134.2)
≥22 days	151.9 (133.0 - 171.4)	118.5 (118.5 - 118.5)	119.4 (88.0 - 138.3)
	p-value = 0.004	*p*-value = 0.62	*p*-value = 0.93

Data is presented as mean (range) in degrees. *p*-values were obtained from linear mixed modelling with neck-shaft angle as a dependent variable and time as co-variate.

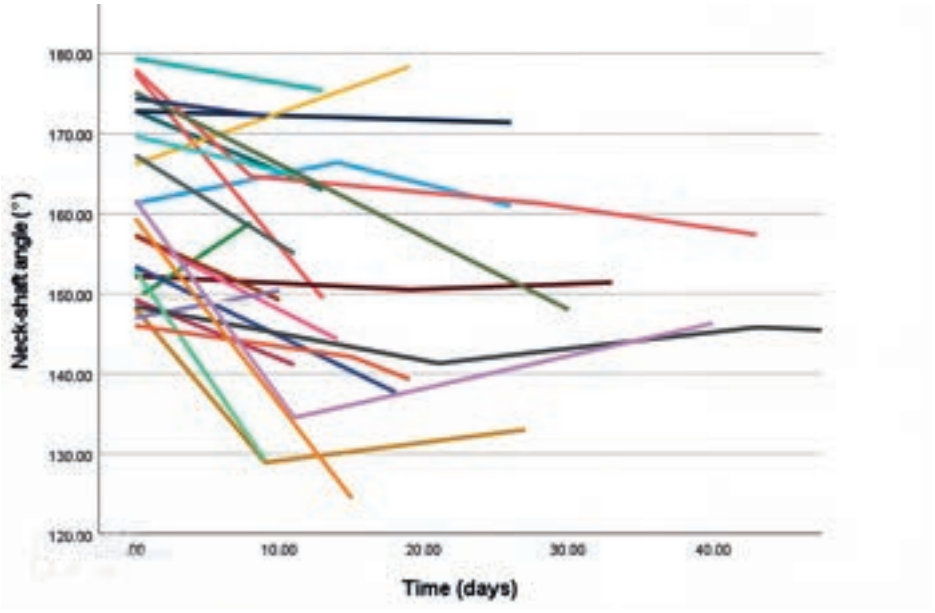


Figure 3. Multiple line graph of neck-shaft angle over time within type A. Each colour represents a patient.

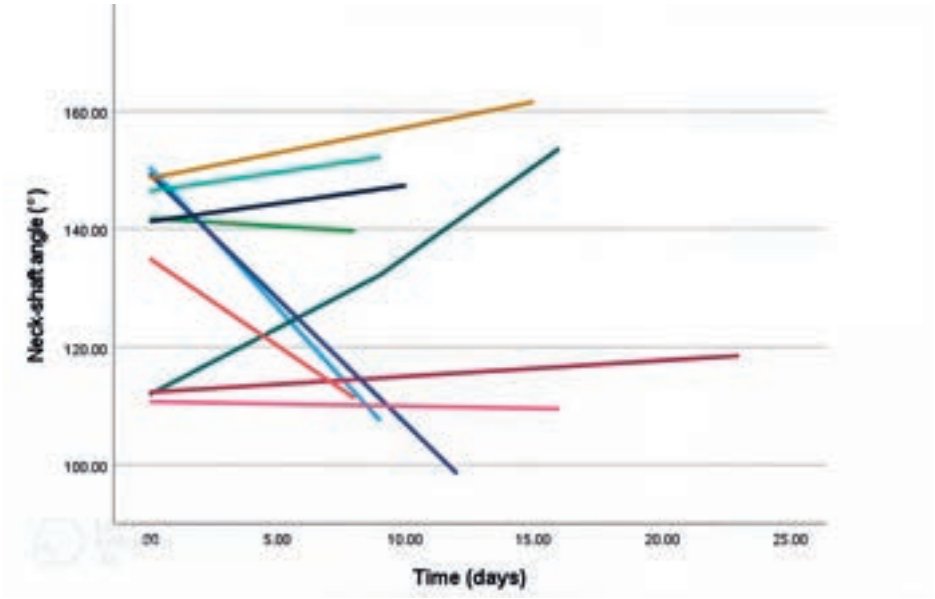


Figure 4. Multiple line graph of neck-shaft angle over time within type B. Each colour represents a patient.

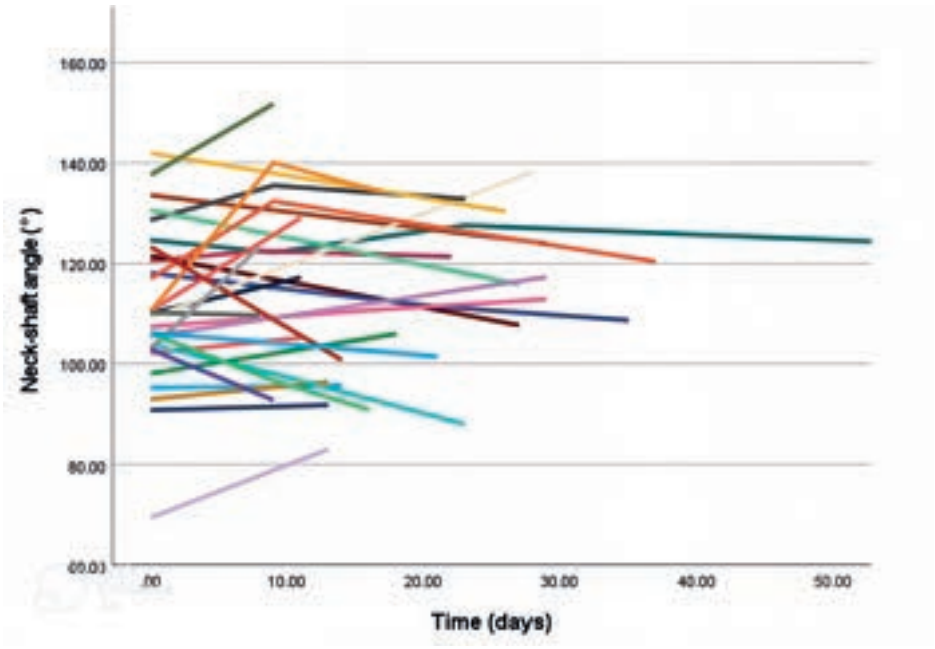


Figure 5. Multiple line graph of neck-shaft angle over time within type C. Each colour represents a patient.

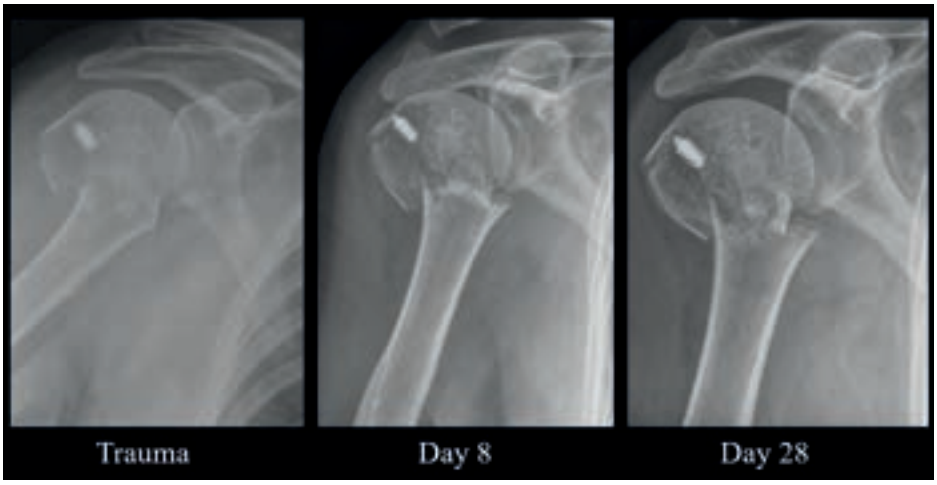


Figure 6. Radiographic follow-up of a type A fracture. Trauma: NSA = 178°, MMG = 5.6 mm, MLG = 10.1 mm. Day 8: NSA = 165°, MMG = 8.5 mm, MLG = 8.4 mm. Day 28: NSA = 161°, MG = 10.3 mm, MLG = 11.2 mm.

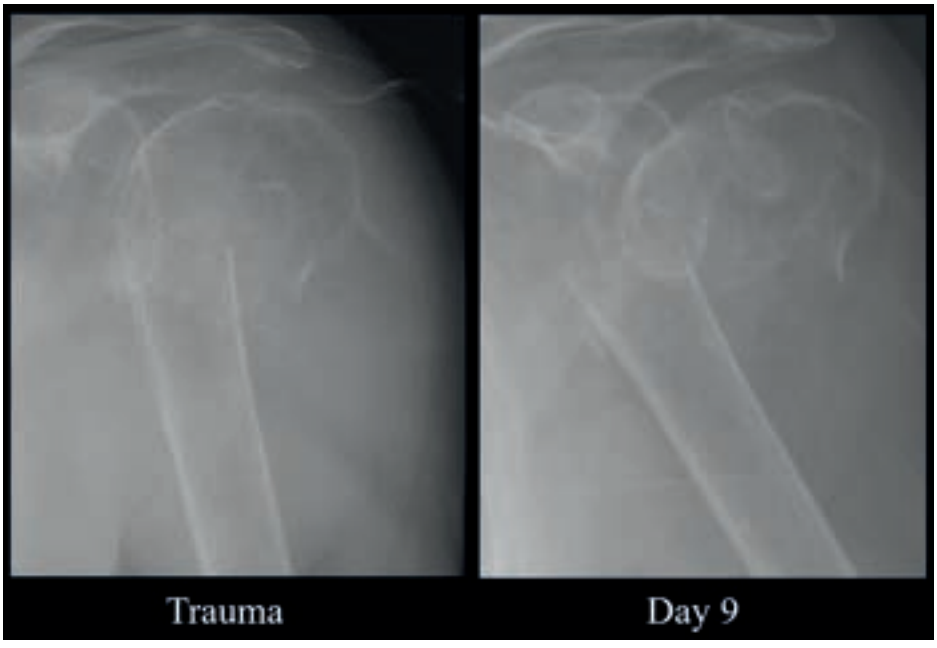


Figure 7. Radiographic follow-up of a type B fracture. Trauma: NSA = 146.4°, MMG = 24.9 mm, MLG = 18.6 mm. Day 9: NSA = 152.2°, MMG = 34.4 mm, MLG = 33.5 mm.

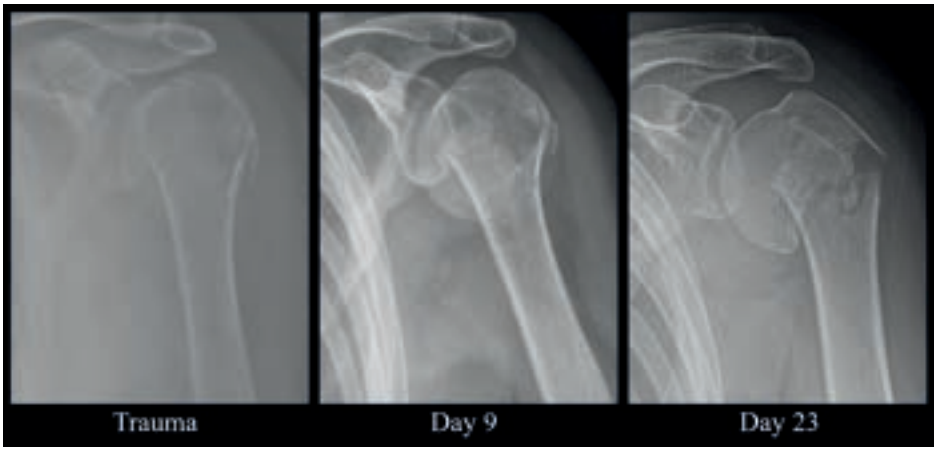


Figure 8. Radiographic follow-up of a type C fracture. Trauma: NSA = 103.8°, MMG = 28.1 mm, MLG = 2.1 mm. Day 9: NSA = 98.1°, MMG = 36.7 mm, MLG = 0.0 mm. Day 23: NSA = 88.0°, MMG = 18.1 mm, MLG = 14.3 mm.

DISCUSSION

Besides establishing actual re-alignment forces when wearing a collar and cuff, evaluating the relationship between fracture patterns and alignment is relevant for clinical decision-making and determining follow-up trajectories of patients. Although we could not control for collar and cuff positioning, patient behaviour, and compliance during management, we postulated that hanging down the arm in a collar and cuff would not re-align the three fracture patterns described by Boileau et al. [4]. In short, we found that apart from valgus head tilt improvement in type A, there was no significant increase nor reduction in displacement among all three Boileau types.

No improvement in the maximal medial gap and the maximal lateral gap was observed within the three fracture patterns. Biomechanically this can be explained by the trade-off between friction, gravity and forces exerted by shoulder muscles. Boileau's fracture types are translated medially or laterally, so a force to the contralateral side may be required for reposition. Most likely the effect of gravity and or muscle activity in the collar and cuff is insufficient to reposition the humeral shaft, and due to muscular atrophy muscle strength also decreases over time [7]. However, studies are lacking on shoulder girdle- and pectoralis major muscle activity during

immobilization. Natural traction force is determined by the gravity of the humerus and surrounding soft tissue which is governed by the weight of the arm. Assuming that patients are compliant and that their collar and cuff instructions were adequate, this natural traction force is apparently not sufficient to relocate the humerus shaft below the humeral head. It should also be considered that repositioning of the shaft requires counteracting forces that are exerted by muscles, tendons, fascia, and friction of bone-on-bone.

Interestingly, neck-shaft angle improved among type A fractures. Re-activation of the supraspinatus muscle during healing may explain this finding but the role of muscle activation during healing of this fracture is unknown. Another theory is that the resolving hematoma contributes (partially) to the head angle restoration. Fracture hematomas in proximal humerus fractures originate from the medial bone arteries and can subluxate the shoulder inferiorly due to accumulation in the glenohumeral joint [8–10]. Therefore, fracture hematomas are more likely to be located on the medial aspect of the humeral head rather than the lateral side. Resolution of hematoma will change the angle of the humeral head, resulting in restoration of traction from the rotator cuff muscles. This restoration might recover the original kinematic balance of the shoulder complex. In type B fractures, no significant NSA improvement was noted over time. Considering that the head is completely separated from the shaft, the forces of rotator cuff muscles are balanced and therefore the head remains in an anatomic position.

Confined to the limits of this study and an unknown quantity of traction, our findings suggest that radiographic re-alignment of type A, B and C fractures should not be expected to improve in clinical practice while managing patients with collar and cuff. Surgical decision-making should therefore be taken upon trauma, in contrast to greater tuberosity fractures where follow-up radiographs could change treatment strategy [11]. Surgical fixation may be required in patients with NSAs ≥160°, so since our data suggest that varus head deformity does not seem to improve over time, surgeons should be aware of this indication [12]. It should be stressed that non-operative treatment could still be valuable considering the high rate of complications in surgically treated patients [13]. For this reason, current guidelines need improvement and a better understanding of the biomechanical concept of this treatment. Further research is required into the actual amount and duration of natural traction provided by collar and cuff over a day using an instrumented collar

and cuff construction to quantify the traction. We also advise to further evaluate the optimal length of immobilization and activity of shoulder girdle muscles while the arm is immobilized and carrying out daily life activities. Additionally, this study should be repeated in a prospective design with patient-reported outcomes and follow-up radiographs at fixed time points. An interobserver study should be carried out to assess the reliability of the Boileau classification to see if it can be incorporated as a subclassification of Neer's two-part fractures [14–17].

There are several shortcomings: first, compliance and collar and cuff instructions in this cohort were unknown. Incorrect collar and cuff positioning may not provide adequate natural traction, so this could have been the case in some patients. Second, the level of activity and general condition of patients were not collected. In bed-bounded patients, for example, natural traction on the fracture is lacking (bisector of the humeral shaft does not points downwards). However, most patients were included from the Level 2 trauma centre which does not treat polytrauma patients. Third, functional outcomes measures were not included, and selection bias may have been introduced as only a limited number of patients had eligible follow-up radiographs and some patients underwent surgery. Fourth, this classification system has not been evaluated in other studies so far and no sample size calculation was performed. However, effect sizes of displacement derived from linear mixed modelling were negligible and even the upper bound and lower bound of the 95% CIs did not exceed 1 mm. Therefore, a clinically relevant improvement is unlikely. Fifth, dorsal humeral head tilt was not considered when classifying the fractures and we included patients to a type B fracture if they had entire medial or ventral displacement. Sixth, analyses were not adjusted for internal or external rotation of the shoulder at trauma and during follow-up radiographs. However, due to pain, it is unlikely that trauma radiographs were taken with the arm in external rotation and radiographers are trained to obtain follow-up radiographs concordantly. Seventh, radiographs were not re-measured by a second researcher so we could not provide the reliability of the measurements. Despite these limitations it should be acknowledged that collar and cuff treatment as applied in current orthopaedic clinical practice was evaluated, and that our results are applicable to a relatively small sample of surgical neck fractures: Boileau fractures comprise only one-fifth of all surgical neck fractures.

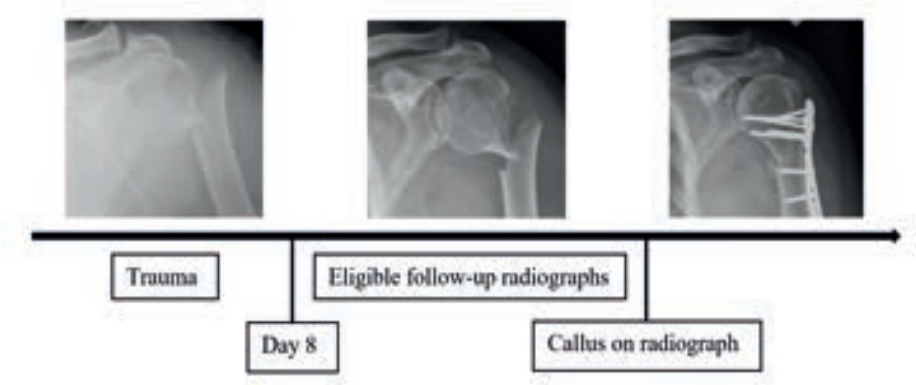
Conclusion

Apart from valgus head tilt improvement in type A, there is no significant increase nor reduction in displacement among all three Boileau types. One may argue that before the radiographically visible callus, there is no effect of hanging on re-alignment of the fracture in the frontal plane. In clinical practice, findings can be used for expectation management of patients and may indicate that re-alignment begins after the onset of callus formation. We advise that surgical decision-making should be performed immediately after trauma.

REFERENCES

1. Launonen AP, Sumrein BO, Reito A, et al. Operative versus non-operative treatment for 2-part proximal humerus fracture: A multicenter randomized controlled trial. *PLoS Med*. 2019;16(7):e1002855.
2. Vachtsevanos L, Hayden L, Desai AS, Dramis A. Management of proximal humerus fractures in adults. *World J Orthop*. 2014;5(5):685.
3. Poeze M, Lenssen AF, Van Empel JM, Verbruggen JP. Conservative management of proximal humeral fractures: Can poor functional outcome be related to standard transscapular radiographic evaluation? *J Shoulder Elbow Surg*. 2010;19(2):273-281.
4. Boileau P, d'Ollonne T, Bessière C, et al. Displaced humeral surgical neck fractures: classification and results of third-generation percutaneous intramedullary nailing. *J Shoulder Elbow Surg*. 2019;28(2):276-287.
5. Meinberg EG, Agel J, Roberts CS, Karam MD, Kellam JF. Fracture and dislocation classification compendium-2018. *J Orthop Trauma*. 2018;32(1):S1-S170.
6. Medixant. RadiAnt DICOM Viewer. Available at: https://www.radiantviewer.com.
7. Deitrick JE, Whedon GD, Shorr E. Effects of immobilization upon various metabolic and physiologic functions of normal men. *Am J Med*. 1948;4(1):3-36.
8. Meyer C, Alt V, Hassanin H, et al. The arteries of the humeral head and their relevance in fracture treatment. *Surg Radiol Anat*. 2005;27(3):232-237.
9. Kolar P, Gaber T, Perka C, Duda GN, Buttgereit F. Human early fracture hematoma is characterized by inflammation and hypoxia. *Clin Orthop Relat Res*. 2011;469(11):3118-3126.
10. Kolar P, Schmidt-Bleek K, Schell H, et al. The early fracture hematoma and its potential role in fracture healing. *Tissue Eng Part B Rev*. 2010;16(4):427-34.
11. van Wier MF, Amajjar I, Hagemeijer NC, Claessen FMAP, van den Bekerom MPJ, van Deurzen DFP. Follow-up radiographs in isolated Greater Tuberosity fractures lead to a change in treatment recommendation; an online survey study. *Orthop Traumatol Surg Res*. 2020;106(2):255-259.
12. Robinson CM, Page RS. Severely impacted valgus proximal humeral fractures results of operative treatment. *J Bone Joint Surg Am*. 2003;85(9):1647-55.
13. Rangan A, Handoll H, Brealey S, et al. Surgical vs nonsurgical treatment of adults with displaced fractures of the proximal humerus the PROFHER randomized clinical trial. *JAMA*. 2015;313(10):1037-1047.
14. Neer CS 2nd. Displaced proximal humeral fractures. I. Classification and evaluation. *J Bone Joint Surg Am*. 1970;52(6):1077-1089.
15. Hertel R, Hempfing A, Stiehler M, Leunig M. Predictors of humeral head ischemia after intracapsular fracture of the proximal humerus. *J Shoulder Elbow Surg*. 2004;13(4):427-433.
16. Russo R, Guastafierro A, Rotonda GD, et al. A new classification of impacted proximal humerus fractures based on the morpho-volumetric evaluation of humeral head bone loss with a 3D model. *J Shoulder Elbow Surg*. 2020;29(10): e374-e385.
17. Sumrein BO, Mattila VM, Lepola V, et al. 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(10):1756-1761.

SUPPLEMENTARY MATERIAL



Supplement 1. Outline of eligible follow-up radiographs.

Supplement 2. Comparison between in- and excluded patients for MMG, MLG and NSA

	A	B	C
MMG incl.	11.3 ± 7.6	29.1 ± 10.2	13.6 ± 8.7
MMG excl.	9.1 ± 5.4	23.1 ± 10.0	16.0 ± 8.7
p-value	0.31	0.21	0.30
MLG incl.	13.5 ± 5.8	22.9 ± 8.7	8.5 ± 5.4
MLG excl.	15.8 ± 6.1	23.3 ± 8.5	9.8 ± 6.4
p-value	0.24	0.92	0.40
NSA incl.	161.1 ± 11.3	134.8 ± 15.8	111.7 ± 15.4
NSA excl.	159.0 ± 12.5	136.7 ± 13.7	114.4 ± 14.8
p-value	0.60	0.78	0.49

MMG and MLG (in millimetre) are presented in means ± standard deviation and NSA (degrees) in means ± standard deviation. Abbreviations: incl., included patients; excl., excluded patients; MMG, maximal medial gap; MLG, maximal lateral gap; NSA, neck-shaft angle.


GRONINGEN
ADELAIDE

8

Pre-operative virtual three-dimensional planning for proximal humerus fractures: a proof-of-concept study

Reinier W.A. Spek
Michel P.J. van den Bekerom
Paul C. Jutte
Frank F.A. IJpma
Ruurd L. Jaarsma
Job N. Doornberg
The Traumaplatform 3D Consortium

Shoulder & Elbow. 2024 Jan;16(4):397-406

ABSTRACT

Aims

To (1) evaluate surgeon agreement on plating features (position and screw length) in virtual 3D planning software, (2) describe outcomes (fracture reduction, plate position, malpositioning of calcar screws and screw lengths) of plate fixations planned with routine pre-operative assessment (2D and 3D CT imaging) and those planned with dedicated virtual 3D software of the same proximal humerus fracture.

Methods

Fourteen proximal humerus fractures were retrospectively reduced and fixed with virtual planning software by eight attending orthopaedic surgeons and compared to the true surgical fixation with post-operative computed tomography (CT) scans. Reduction differences were quantified using CT micromotion analysis.

Results

Intraclass correlation for screw lengths was 0.97 (95% CI: 0.96 - 0.98), and 0.90 (95% CI: 0.79 - 0.96) for plate position. Mean difference in total fracture rotation of the head between the virtual and conventional group was 22.0°. Plate position in the virtual planning group was 3.2 mm more proximal. There were no differences in inferomedial quadrant calcar screw positioning and, apart from the superior posterior converging screw, no significant differences in screw lengths.

Conclusion

Reproducibility on plate position and screw length with virtual planning software is adequate. Apart from fracture reduction, virtual planning yielded similar plate positions, screw malpositioning rates- and lengths compared to routine pre-operative assessment.

INTRODUCTION

Advocates of surgery argue, that suboptimal results of open reduction and internal fixation (ORIF) are due to the specific technical challenges: fracture reduction, calcar screw positioning, and screw lengths are considered the most profound surgical determinants for successful plate fixation [1–3]. Screw lengths should be carefully selected: excessively short screws are associated with an exponential increasing risk for secondary reduction loss whereas overly long screws can cause intra-articular screw penetration demanding revision surgery to avoid progressive cartilage damage [4–7].Interestingly, most of the complications (55%) following proximal humerus plate fixation are iatrogenically induced intra-operatively and could thus be avoided [8]. Among these preventable complications, 62% is caused by primary screw penetration, hence, there is a clear need to optimize these lengths [8]. Virtual three-dimensional (3D) planning software is a relatively new tool to aid in the pre-operative planning of (proximal humerus) fractures [9]. The software allows to perform reduction, plate positioning, and determine the screw lengths on a virtual model. A 2018-study showed promising results: virtual planning leads to a shorter operative time, fewer blood loss and shorter hospital stay compared to conventional planning [10].

More studies are needed to evaluate if this computer software should be implemented in clinical practice, and henceforth can reduce adverse events with associated hospital costs. The aims were to (1) evaluate surgeon agreement on plating features (position and screw length) in virtual 3D software for pre-operative planning of proximal humerus fractures, (2) describe outcomes (fracture reduction, plate position, malpositioning of calcar screws and screw lengths) of plate fixations planned with routine pre-operative assessment (2D- and 3D CT imaging) and those planned with dedicated virtual 3D software of the same proximal humerus fracture.

PATIENTS AND METHODS

Study design and setting

This retrospective study was carried out in a Level 1 trauma centre in Adelaide, Australia. Fourteen proximal humerus fractures were reduced and fixed with virtual planning software (= virtual planning group) and compared to the true surgical fixation as assessed on post-operative computed tomography (CT) scans (= routine pre-operative assessment group). The potential difference in reduction between the groups was quantified using CT micromotion analysis. Surgical treatment to these patients was performed before the virtual planning was carried out (Fig. 1).

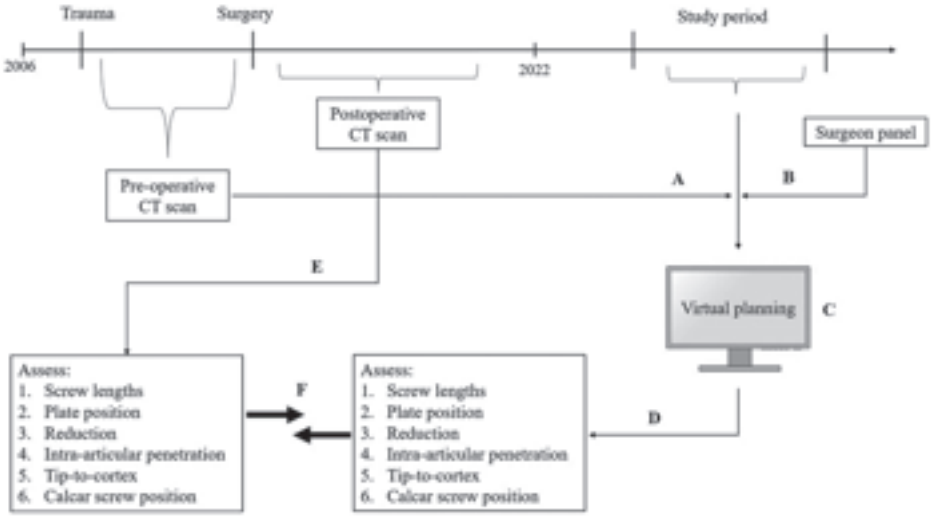


Figure 1. Overview of study design and outcome parameters. All Surgical procedures were carried out before commencement of the study. A = pre-operative CT scans were uploaded into the planning software to create virtual models to fix the plate constructs on. B = A surgeon panel was created, each surgeon was asked to plan the cases, C = the virtual planning was carried out: the reduction was checked, the plate positioned, and screws inserted. D = after the planning the parameters described within the box were measured, E = The post-operative CT scan obtained after the true surgical procedure was assessed for the same parameters listed in the box, F = the virtual and true surgical results were compared.

Study subjects

CT scans with available axial, coronal, and sagittal slices, were included if patients were at least 18 years of age and sustained a proximal humerus fracture where the clinical decision was made to perform open reduction and internal plate fixation

with either a Carbofix (CarboFix Orthopedic Ltd., Herzeliya, Israel) or Philos (Depuy Synthes, Oberdorf, Switzerland) proximal humerus plate. Patients must have had a CT scan upon trauma, a post-operative CT scan, and their fracture had to be reduced anatomically which was defined as a non-retroverted, anteverted, varus or valgus angulated head, with none-to-minimal residual head-shaft translation. Patients with secondary collapsed humeral heads and inadequately positioned plates were also excluded (angulated, too high, or too low). Adequacy of anatomical reduction and plate position was judged on post-operative CT scans and follow-up radiographs according to the AO principles by the first author and an expert orthopaedic surgeon (R.J.). There was no minimum length the CT scans extended down the humerus for inclusion. Screening was performed in a Level 2 Dutch trauma centre and the South Australian medical imaging database (inclusion period: 2006 up to and including 2021) which contains patient images from 11 hospitals across South Australia. Patients with a proximal humerus fracture who underwent ≥2 CT scans were evaluated against the in- and exclusion criteria. An additional search was conducted at our Level 1 trauma centre: each patient who underwent plate fixation between January 2018 and January 2022 was checked for eligibility.

Description of routine pre-operative planning

Routine assessment included evaluation of the radiographs, as well as the two-dimensional (2D) and 3D CT scans. These were also used to establish an indication for the required screw lengths. True lengths were measured intra-operatively using a depth gauge. Intra-operative fluoroscopy was used to check for intra-articular screw penetration: if needed screws were changed accordingly.

Description of virtual planning

Eight attending orthopaedic surgeons were recruited to conduct the virtual planning. Five surgeons finished their training more than one year ago at the start of the study and three surgeons finished their training >15 years ago of which one was dedicated to the upper limb, and the others to both trauma and general orthopaedics. Virtual planning was carried out on a designated laptop provided by the software developer (Sectra, Linköping, Sweden). In this software, the anonymized trauma CT scans (bone slices, 0.5 mm) were uploaded and converted to 3D virtual models which were freely rotatable in space. The model was segmented to create an isolated proximal humerus, with each bone fragment displayed in a separate colour (Fig. 2). The fragments were then reduced by dragging them around. Each reduction was

performed by the first author and checked by each surgeon before they advanced with the plate fixation (Fig. 3). If the surgeons were not satisfied with the reduction, they were allowed to alter it at their discretion. Segmentation and reduction took an estimated 45 minutes per case. After the surgeons approved the reduction, they selected a Philos or Carbofix plate and subsequently inserted the screws (plate brand, screws, and screw holes to be used were dictated by the surgical procedure). Each surgeon placed the locking plate on the proximal humerus and shifted it to their desired position (Fig. 4.). Second, screws were inserted, and adjusted: screw lengths could be altered by dragging the tip of the screw on the 2D slices (shown next to the virtual model) which allowed users to judge their position in relation to the cortex (Fig. 5). Surgeons chose screw lengths they deemed satisfactory for adequate purchase. After screw insertion, a fluoroscopy view was created to check for intra-articular screw penetration and –if needed–, screws could be re-adjusted from there (Fig. 6). Angles of the Philos screws were fixed (0°), while surgeons were allowed to change the Carbofix screws to a maximum angle of 10°. Available screw lengths were identical to the manufacturer guidelines: 2.0 mm increments for Philos screws, 2.5 mm increments for Carbofix screws between 30 and 50 mm (if a size larger was needed 55 or 60 mm had to be chosen).

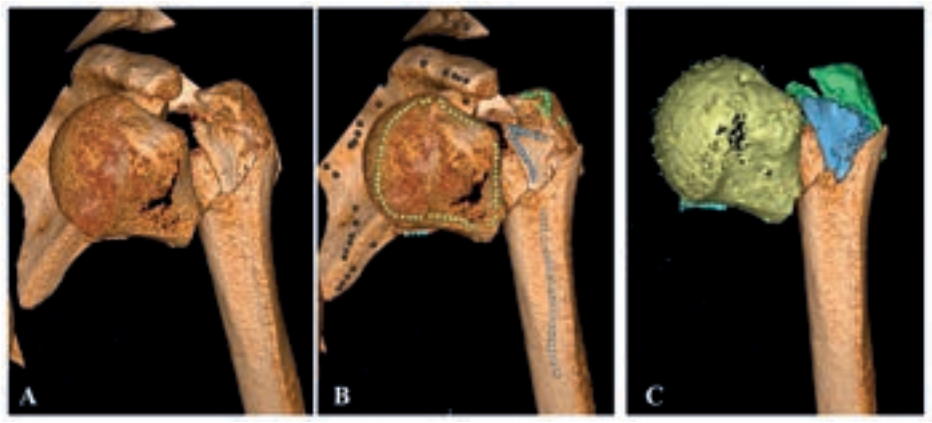


Figure 2. Segmentation of virtual proximal humerus model. A = an adjustable clip box was positioned around the proximal humerus so that all structures outside this area were removed, B = the model was prepared for segmentation by marking them with multiple dots (yellow, blue, green), the non-humeral bones were marked with black dots, C = result of segmentation, the bone fragments are now ready to be reduced.

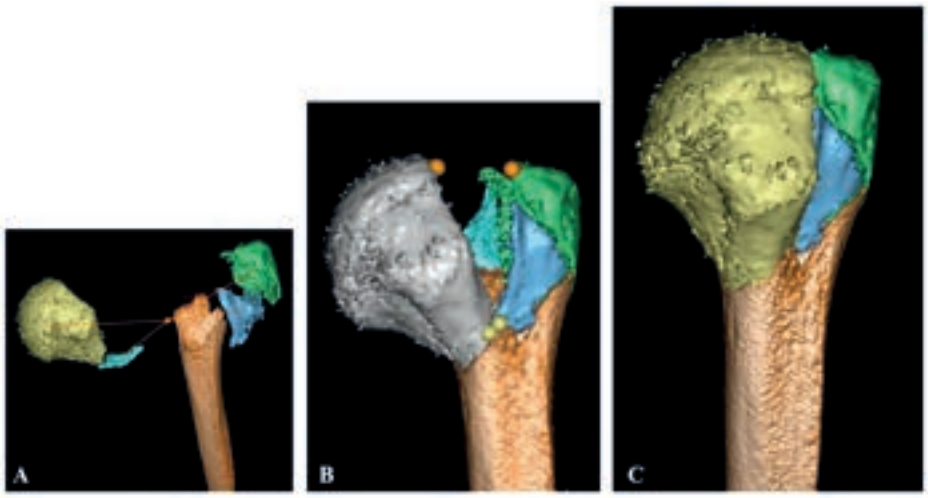


Figure 3. Virtual Reduction of fragments. A = Each fragment could be freely directed in space, B = One could also colour-code the pieces so that the software would connect the fragments (orange to orange, yellow to yellow), C = Final reduction.

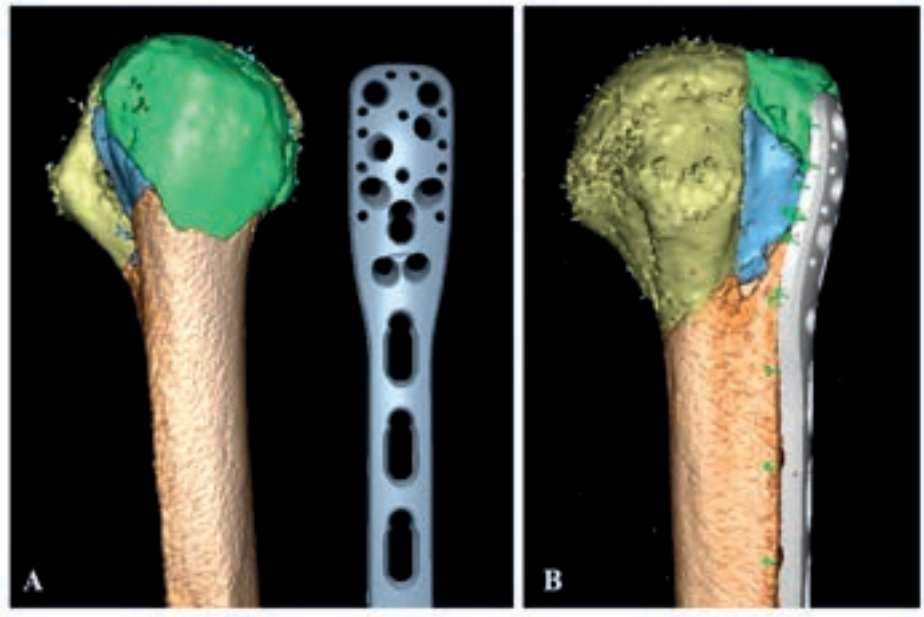


Figure 4. Plate fixation of virtual model. A = the plate was first positioned next to the model, after which it was dragged to the desired position, B = plate is positioned, screws could be inserted by clicking on the green arrows.

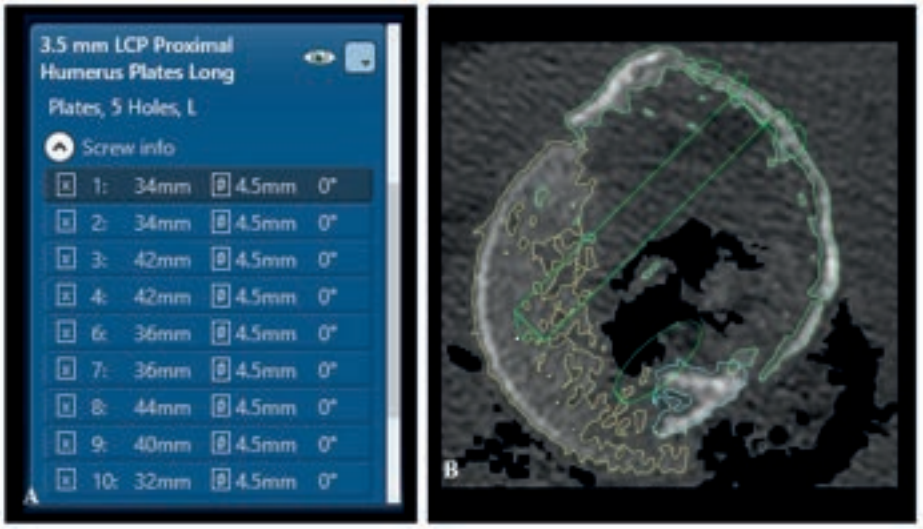


Figure 5. Interactive menu and 2D CT slices displayed next to the virtual model A = One could see each screw lengths as well as their angles, B = One could judge the relationship against the cortex.

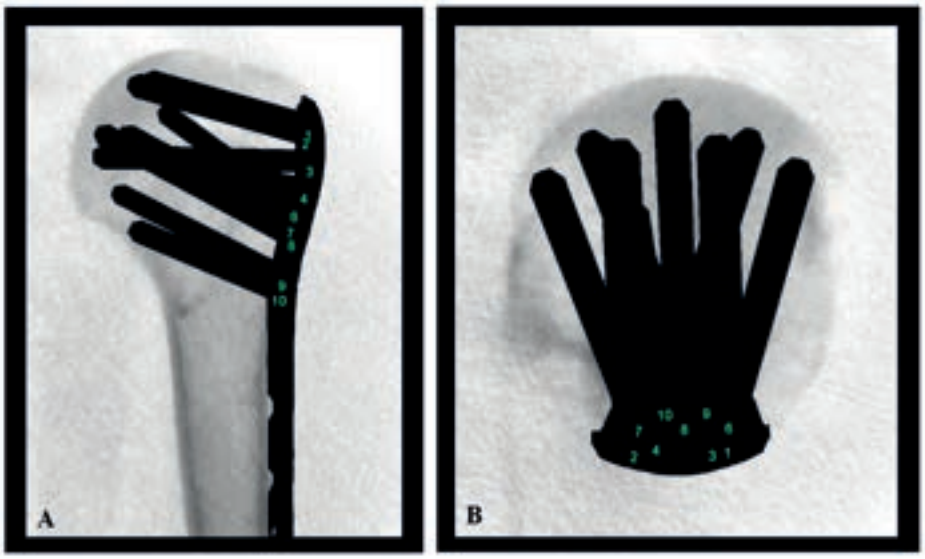


Figure 6. Fluoroscopy to check for intra-articular penetration. A = AP view, B = Superior view.

Variables, outcome measures

Inter-surgeon agreement was evaluated with the intraclass correlation coefficient (ICC). Reduction was measured with the total humeral head rotation, medial hinge displacement, centre of rotation and the neck-shaft angle. The first three were determined by computed tomography micromotion analysis (CTMA) (version 23.1). The neck-shaft angle was determined by measuring the angle between the line parallel to the humeral shaft, and the line perpendicular to the anatomic neck. The inclination angle on the virtual models was performed on the 3D reductions, while the inclination angle on the postsurgical reductions were measured on the 2D CT scans using multiplanar reconstruction (MPR) [11]. The plate-position was determined by measuring the humeral head height: the distance between the top of the humeral head and the tip of the plate, parallel to the line along the base of the plate (Supplement 1) [1]. Adequacy of calcar screw positioning was also evaluated: if they were inserted in the inferomedial quadrant they were judged as adequate, otherwise they were categorized as malpositioned (Supplement 2) [3]. Screw lengths were measured from the head to the tip of the screw on the 2D post-operative CT scans using MPR for each screw (Supplement 3) [12]. As Philos screws seemed larger due to metal artifacts and its head-to-tip distance is longer than labelled (approximately 0.5 mm), they were rounded down if the measurement exceeded a true length (e.g., 38.6 mm was rounded down to 38 mm, 40.5 mm to 40 mm). Carbofix plates did not produce any metal artefacts, so such adjustments were not needed for these screws. Intra-articular screw penetration was defined as a cortex breach with protrusion of the outer cortex layer. The screw tip-to-cortex distance (mm) was defined as the distance from the tip of the screw to the outer cortex (Supplement 3) [13]. Measurements were completed by one assessor (first author).

Computed tomography micromotion analysis

CTMA is an interactive software tool developed to detect micromotion between objects such as bones or implants. In this study, total rotation, translation of the medial hinge (most proximal point of the calcar), and the centre of rotation (exact centre of the best fitting circle overlapping the humeral head) were measured. After uploading the post-operative CT scan and virtual reduction into CTMA, the humeral shaft was chosen as the reference body and marked with coloured dots so that the virtual and surgical model could be semi-automatically overlapped. In the same fashion, the surgically reduced humeral head (moving body) was overlayed on the virtual reduction. By doing this, the software created a colour patterned model

which showed to what extent both models were comparable (Fig. 7). Translations in the x, y, and z direction as well as the rotation around these axes were calculated and combined into one value: the total translation in mm and total rotation in degrees [14–17]. Measurements were performed twice by the same certified software user (the mean of both measurements was reported).

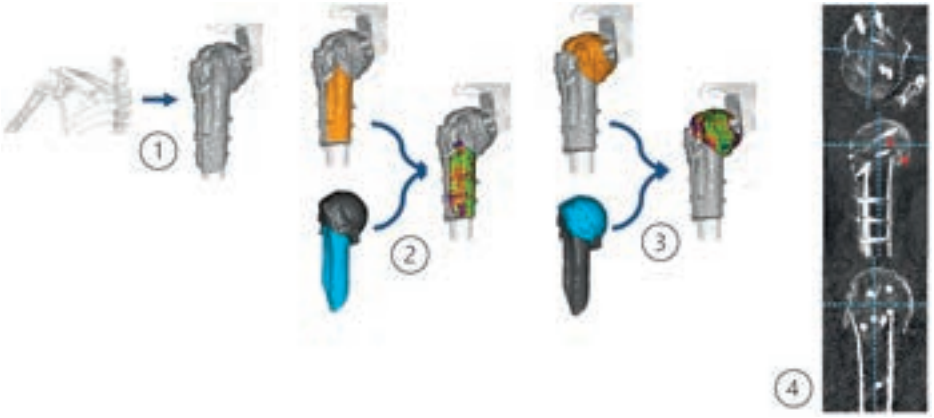


Figure 7. Workflow of CTMA. 1 = Preparation of post-operative CT scans and pre-operative surgical plan, 2 = semi-automatic alignment of reference body (humeral shaft): blue and yellow are merged. The model on the right represents the closeness of fit between both models: green indicates a perfect fit, while red indicates a poor fit. As the dominant colour in this figure is green it can be concluded that the overlap was done well, 3 = Semi-automatic alignment of moving body (humeral head) after which the software provides the total rotation difference of the head. Again, blue and orange were merged, and the right model shows the adequacy of overlap, 4 = To evaluate the centre of rotation and medial hinge displacement these points had to be manually placed on the 2D CT scan. Multiplanar realignment was used to ensure it was measured in the correct plane.

Statistical analysis, study size

To determine the reproducibility of screw lengths and plate position, a 2-way random ICC in absolute agreement was calculated: values ≤0.50 were categorized as poor, 0.51 - 0.75 as moderate, 0.76 - 0.90 as good and >0.90 as excellent [18]. A paired samples T-test was used for comparing the neck-shaft angle, a Welch's T-test for the plate position with the screw tip-to-cortex distance, and a Pearson Chi-Square test for calcar screw malpositioning and intra-articular screw penetration. To analyse the screw lengths, each screw hole was numbered from 1 to 9 similar to Chen et al (Fig. 8) [10]. Screw lengths were averaged for each hole amongst all cases, presented as mean ± standard deviation, and compared with the Welch's T-test. Analyses were not adjusted for the Carbofix screw angle variability. A *p*-value less

than 0.05 was considered as significant. IBM SPSS software version 27 (IBM Corp., Armonk, N.Y., USA) was utilized for analysing the data. The post hoc power analysis was performed in G*Power with an alpha error probability of 0.05 [19,20]. For the neck-shaft angle, a 6° difference could be detected with a power of 86% (dependent t-test, effect size = 0.59) and plate position yielded 88% power (independent t-test, effect size = 0.90). Differences of 18% or higher for adequacy of calcar screw positioning (Chi-Square test, power = 0.84, effect size = 0.25) and 6 mm in screw length could be detected reliable (independent t-test, power = 0.80, effect size = 0.91).

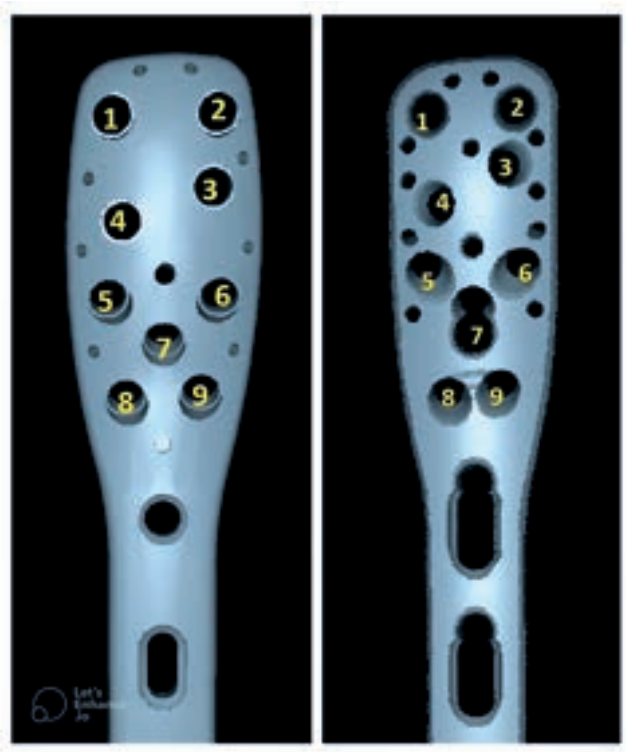


Figure 8. Plate screw hole numbers. The left plate is a Carbofix plate, the right plate the Philos plate. The left side of the image is anterior (1,4,5 and 8 are the anterior screw holes), right posterior (2,3,6 and 9 are the posterior screw holes).

RESULTS

Fifty-four patients who underwent a pre- and post-operative CT scan were identified of which 17 were excluded due to non-anatomical reduction, 13 due to incompatible plate brand (all plates other than Carbofix and Philos), 8 because of inadequate plate position and 2 for secondary head collapse. Thus, after assessment against eligibility criteria, fourteen proximal humerus fractures were included and virtually fixed by every surgeon in the panel. The median age was 66.3 (range: 32 - 76) of which 72% were females. Ten patients were fixed with a Philos plate, four with a Carbofix plate. Included fracture patterns entailed 5 two-part fractures, 5 three-parts, 2 four-parts, and two glenohumeral fracture dislocations (Table 1).

With virtual planning the ICC for screw lengths was 0.97 (95% CI: 0.96 - 0.98) indicating an excellent reliability (number of screws: 107). The ICC for plate position was 0.90 (95% CI: 0.79 - 0.96) indicating good reliability (number of cases: 14). As one surgeon dropped out after planning three cases, his data was excluded for ICC calculations. The mean difference in total rotation between the virtual planning and routine pre-operative planning group was 22.0° ± 11.1°. The medial hinge displacement was 7.9 ± 4.5 mm and the centre of rotation 4.2 ± 1.9 mm (Table 2, Supplement 5 - 7). The mean virtual inclination angle was 136° ± 5° and the inclination angle in the routine assessment group was 132° ± 10° (non-significant difference). Taken together, it can be inferred that the reduction mainly differed with regards to the tuberosities, retro- and anteversion of the head, and position of the humeral shaft. The plate position in the virtual group was 3.2 mm more proximal than the routine surgical assessment group (95% CI delta: 0.8 - 5.7). There were no differences in inferomedial quadrant calcar screw positioning and apart from screw number 3 there were also no significant differences in screw lengths between the virtual and true surgical procedure (Table 3 - 4, Supplement 4). In the standard pre-operative assessment group (true surgical cohort), 14% of all screws penetrated intra-articular. In the virtual group, screws were inserted closer to the cortex (virtual: 3.4 mm *versus* true surgery: 6.8 mm) but only one screw was placed intra-articular. The average time for plate positioning and screw placement per model including the first three fixations (learning cases) was 7 minutes and 11 seconds. When the learning curve was flattened it was 5 minutes and 55 seconds.

Table 1. Demographics of all patients and those with and without screw penetration

	IA screw penetration		
	Without (n = 7)	**With (n = 7)**	**All (n = 14)**
Age	58.4 (32 - 76)	66.3 (56 - 76)	66.3 (32 - 76)
Gender			
Female	4 (57.1%)	6 (85.7%)	10 (71.4%)
Male	3 (42.9%)	1 (14.3%)	4 (28.6%)
Side			
Left	4 (57.1%)	3 (42.9%)	7 (50%)
Right	3 (42.9%)	4 (57.1%)	7 (50%)
Neer classification			
Two-part	3 (42.9%)	2 (28.6%)	5 (35.7%)
Three-part	2 (28.6%)	3 (42.9%)	5 (35.7%)
Four-part	2 (28.6%)	-	2 (14.3%)
Dislocation	-	2 (28.6%)	2 (14.3%)
Plate			
Philos	4 (57.1%)	6 (85.7%)	10 (71.4%)
Carbofix	3 (42.9%)	1 (14.3%)	4 (28.6%)
Screws per fracture	7 (6 - 9)	8 (5 - 9)	7.5 (5 - 9)
Total proximal screws	51	50	101

Categorical variables are presented as number (%), continues variables as median (range). Abbreviations: IA, intra-articular.

Table 2. Rotation and translation differences between virtual and surgical reductions

Case	Rotation	Medial hinge	Centre of rotation
2	19.1°	4.1	3.7
3	16.8°	6.9	2.7
6	38.9°	14.2	3.6
7	7.9°	2.4	2.8
8	38.4°	15.6	5.6
9	10.2°	1.5	2.1
10	28.6°	9.4	3.4
11	26.6°	8.1	8.6
12	7.7°	5.3	5.3
13	27.4°	9.6	2.5
14	20.6°	10.4	5.4
All cases	22.0° ± 11.1°	7.9 ± 4.5	4.2 ± 1.9

Medial hinge and center of rotation displacement are expressed in mm. Mean of all cases is reported ± standard deviation. Three cases (1,4,5) were excluded due to severe metal artifacts. The virtual reduction of case 6 and 8 was adjusted by one of the surgeons. The difference between both reductions in case 6 was 6mm and 20°. For case 8 this was 0.8 mm and 0.3°. The first reduction of case 6 used by six surgeons, the second by one surgeon. The first reduction in case 8 was used by seven surgeons, the second by one surgeon. The ICC was good for measuring the rotation (0.88, 95% CI: 0.61 - 0.98), excellent for the medial hinge (0.93, 95% CI: 0.76 - 0.98) and good for the centre of rotation (0.83, 95% CI: 0.48 - 0.95).

Table 3. Intra-operative *versus* virtual NSA, and chosen screw lengths

	Intra-operative	Virtual	Δ	*p*-value
NSA	132˚ ± 10˚	136˚ ± 5.1˚	-3.7 (-10.0 - 2.6)	0.10
HHH	13.9 ± 4.2	10.6 ± 3.0	3.2 (0.8 - 5.7)	0.01
Screws				
Hole 1	40.4 ± 5.6	38.5 ± 5.2	1.9 (-1.6 - 5.4)	0.27
Hole 2	40.7 ± 5.5	39.0 ± 5.4	1.7 (-1.8 - 5.1)	0.32
Hole 3	41.3 ± 5.9	45.0 ± 4.7	-3.7 (-7.4 - -0.1)	0.046
Hole 4	44.0 ± 6.9	45.4 ± 5.5	-1.4 (-5.5 - 2.7)	0.48
Hole 5	41.4 ± 5.1	39.0 ± 5.2	2.3 (-1.2 - 5.9)	0.18
Hole 6	43.8 ± 7.3	42.7 ± 5.2	1.1 (-4.2 - 6.4)	0.67
Hole 7	47.7 ± 3.4	47.0 ± 3.1	0.7 (-1.8 - 3.2)	0.55
Hole 8	43.8 ± 8.5	40.8 ± 9.5	2.9 (-3.8 - 9.7)	0.36
Hole 9	43.3 ± 10.0	44.1 ± 9.1	-0.9 (-9.4 - 7.6)	0.82

NSA (˚), HHH (mm) and screw lengths (mm) as mean ± standard deviation and Δ with 95% confidence interval. NSA was compared with the paired samples T-test, HHH with the Welch's T-test, and screw lengths with the Welch's T-test.

Table 4. Tip-to-cortex distance, intra-articular penetration, and adequacy of calcar screws

	Non-calcar screws (1 - 7)		Calcar screws (8 - 9)		
	Penetration	Tip-to-cortex	IM quadrant	Penetration	Tip-to-cortex
Intra-operative	12/84 (14%)	6.8 ± 3.7	11/17 (65%)	1/11 (9%)	6.8 ± 3.4
Virtual	1/600 (0%)	3.4 ± 1.4	77/121 (64%)	1/77 (1%)	3.6 ± 1.5
p-value	<0.001	<0.001	0.93	0.11	0.02

IM quadrant positioning of calcar screws was deemed adequate. The penetrating calcar screw in the virtual group was angled too far anterior. Tip-to-cortex was only calculated for the non-penetrating screws and the calcar screws which were angled into the IM quadrant. The Welch's test was used to compare continues variables and the Pearson Chi-Square test for categorical variables. Tip-to-cortex distance was measured in mm.

DISCUSSION

Optimizing screw lengths in plated proximal humerus fractures may be challenging due to the mismatch between intra-operative fluoroscopy and the actual 3D spherical shape of the humeral head: intra-articular screw penetration appears a preventable iatrogenic reason for re-operation. As it is unclear if surgeon-controlled pre-operative virtual planning can reduce this complication, we sought to evaluate the benefits of this software. In short, this study found a satisfactory reproducibility on plate position- and screw lengths with virtual planning. There was also a vast difference in total fracture rotation and translation between both groups without differences in angular malalignment in the anteroposterior plane. Surgeons opted for a somewhat more proximal plate position with virtual planning, but no difference was found in calcar screw positioning, and screw lengths.

For adequate interpretation of this study, it should be highlighted that each fracture was evaluated twice. First retrospectively, using post-operative CT scans as reference for evaluating outcome parameters, and second, prospectively where the same fracture was uploaded into the virtual planning software. The resulting virtually fixed model was assessed for the same outcome variables.

Several shortcomings should be acknowledged. To begin with, only anatomically reduced proximal humerus fractures were selected in this study. If also mal-reduced fractures would be included (like in clinical practice), it is likely that the displacement difference between the virtual and surgical models would have been even larger. Second, CT scans were not performed routinely in the participating institutions (despite thorough multicentre screening, we included only 14 suitable cases) which may have introduced selection bias: the current sample may not cover the full spectrum of fracture patterns which are considered for plate fixation. Third, due to the limited sample size, we were only able to detect marked differences in screw lengths between both groups. Fourth, reductions were completed by a researcher and checked by the surgeons, hence whether surgeons would clinically benefit from doing the virtual reductions themselves was not investigated. Fifth, measurements were not done in duplicate. Finally, the retrospective nature of this work should be realized. The virtual planning was performed years after the surgeries, which means that the surgeons did not have the plan available at the time of surgery to try to replicate. Moreover, the surgeon who carried out the true procedure differed from the

surgeons who carried out the virtual planning. Differences in outcome parameters can be attributed to these factors so prospective studies are needed to evaluate whether virtual planning leads to improved fracture reduction and implant positioning.

The only study which evaluated clinical outcomes after virtual planning, did not include a reliability analysis but this gap in knowledge was filled by our study [10]. Amongst the seven users we detected an excellent reproducibility for screw lengths and an almost excellent (0.90) reproducibility for plate position. This is very satisfactory and even higher than was shown with the same software on distal radius fractures (ICC: 0.77) [9]. It can be concluded that there is a strong consensus on ideal screw lengths and plate position in proximal humerus fractures. The challenge is to mirror the virtual reduction during the true surgical procedure.

Virtual reductions differed from the reductions in the standard pre-operative assessment group with a mean of 22° and 8 mm medial hinge displacement. This can be attributed to several favourable factors present within the virtual software: fragments can be reduced without a time limit or concerns for soft tissues, fragment manoeuvrability is high (not limited by soft tissue attachments), and the visibility is excellent (users are not limited by the view of their surgical approach). Therefore, it is logical to assume that the virtual models were closer to a perfect anatomical reduction and emphasizes that reducing proximal humerus fractures during surgery remains a true challenge. Surgeons picked a 3 mm higher plate position as was done in the surgical procedure. This may have been done to improve positioning of the calcar screws in the Philos plates, but our study did not find any differences between both cohorts (Table 4). Another explanation is that the supraspinatus tendon attachment (which is not present on the virtual model) has restricted the plate height. In our study it was shown that screw number 3 (posterior converging) was 4 mm longer in the virtual planned models. Aside from this finding, the other screws were not significantly different between both groups. In our surgically treated patients, 14% of all screws penetrated intra-articular while in the virtual group this was almost none (1/600). Based on this vast difference in screw penetration, one would expect shorter screws in the virtual group, but this was not the case. Most likely this can be attributed to a difference in reduction, plate position and better screw trajectory. Another clinical study compared virtual planning to 3D printing and conventional planning and did not find any differences between the pre-operative templated screw lengths and the screw lengths during surgery meaning that the

surgeons adhered well to their pre-operative template [10]. In this study, none of the patients who received pre-operative virtual planning had post-operative screw penetration while in the conventional group it occurred in two patients [10]. Despite their outstanding pre-liminary work the authors only included 46 patients and introduced selection bias (patients did not receive virtual planning at random) so whether this (potential) benefit will continue to occur, should be confirmed in larger prospective databases.

Success comes from preparation, so we advise to carry out a multicentre randomized controlled trial with low-dose post-operative CT scans and surgical complications and surgical parameters as outcome measures together with an appropriate cost-effectiveness analysis. The virtually reduced models with pre-planned screw lengths should be displayed on-screen in theatre. Studies can also be carried out on its value as an education tool for training residents and even to familiarize students with surgical procedures. Further to this, one should address the role of virtual planning on fracture classification and characterization, perhaps this new technology can tackle the subjectivity incorporated in the various classification systems (the computer modelling software provides the total rotation and translation from its initial position when fracture fragments are moved) [21].

Conclusion

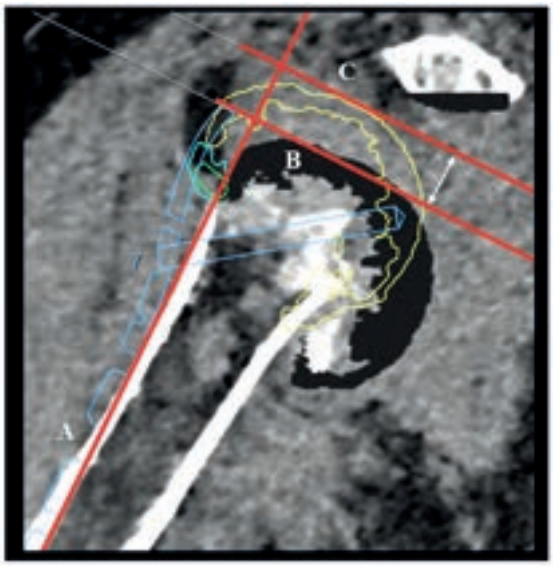
In conclusion, reproducibility on plate position and screw length with virtual planning software is adequate. Apart from fracture reduction, virtual planning yielded similar plate positions, screw malpositioning rates- and lengths compared to routine pre-operative assessment. Surgeons should therefore be careful with replicating the pre-planned screw dimensions during surgery.

REFERENCES

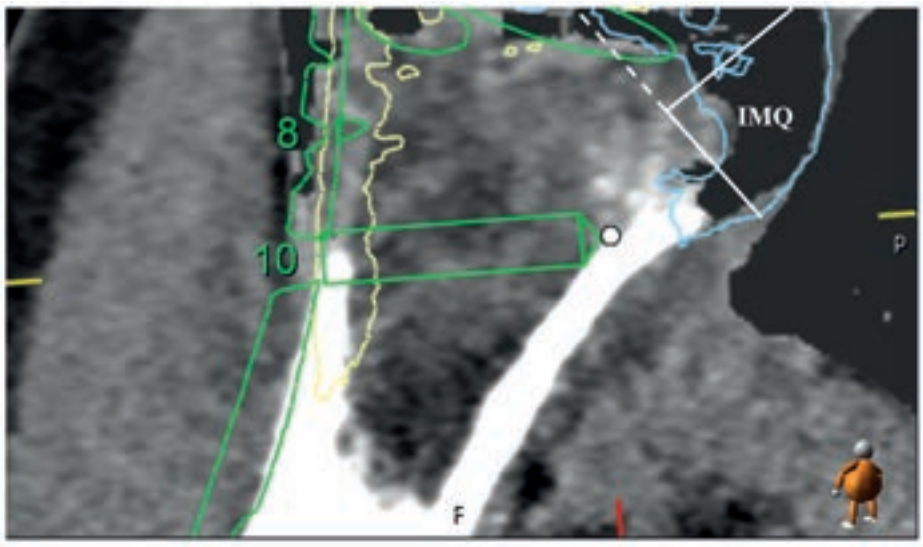
1. Gardner MJ, Weil Y, Barker JU, Kelly BT, Helfet DL, Lorich DG. The importance of medial support in locked plating of proximal humerus fractures. *J Orthop Trauma*. 2007;21(3):185-191.
2. Helfen T, Siebenbürger G, Fleischhacker E, Biermann N, Böcker W, Ockert B. Open reduction and internal fixation of displaced proximal humeral fractures. Does the surgeon's experience have an impact on outcomes? *PLoS One*. 2018;13(11).
3. Wang Q, Sheng N, Rui B, Chen Y. The neck-shaft angle is the key factor for the positioning of calcar screw when treating proximal humeral fractures with a locking plate. *Bone Joint J*. 2020;102-B(12):1629.
4. McMillan TE, Johnstone AJ. Primary screw perforation or subsequent screw cut-out following proximal humerus fracture fixation using locking plates: a review of causative factors and proposed solutions. *Int Orthop*. 2018;42(8):1935-1942.
5. Omid R, Trasolini NA, Stone MA, Namdari S. Principles of Locking Plate Fixation of Proximal Humerus Fractures. *J Am Acad Orthop Surg*. 2021;29(11):E523-E535.
6. Panagiotopoulou VC, Varga P, Richards RG, Gueorguiev B, Giannoudis P V. Late screw-related complications in locking plating of proximal humerus fractures: A systematic review. *Injury*. 2019;50(12):2176-2195.
7. Fletcher JWA, Windolf M, Grünwald L, Richards RG, Gueorguiev B, Varga P. The influence of screw length on predicted cut-out failures for proximal humeral fracture fixations predicted by finite element simulations. *Arch Orthop Trauma Surg*. 2019;139(8):1069-1074.
8. Südkamp N, Bayer J, Hepp P, et al. Open reduction and internal fixation of proximal humeral fractures with use of the locking proximal humerus plate. Results of a prospective, multicenter, observational study. *J Bone Joint Surg Am*. 2009;91(6):1320-1328.
9. Prijs J, Schoolmeesters B, Eygendaal D, et al. 3D virtual pre-operative planning may reduce the incidence of dorsal screw penetration in volar plating of intra-articular distal radius fractures. *Eur J Trauma Emerg Surg*. 2022;48(5):3911-3921.
10. 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(22):1960-1968.
11. Jia X, Chen Y, Qiang M, et al. Compared to X-ray, three-dimensional computed tomography measurement is a reproducible radiographic method for normal proximal humerus. *J Orthop Surg Res*. 2016;11(1):82.
12. Jia X, Chen Y, Qiang M, et al. Detection of Intra-Articular Screw Penetration of Proximal Humerus Fractures: Is Postoperative Computed Tomography the Necessary Imaging Modality? *Acad Radiol*. 2019;26(2):257-263.
13. Totoki Y, Yoshii Y, Kusakabe T, Akita K, Ishii T. Screw Length Optimization of a Volar Locking Plate Using Three Dimensional Preoperative Planning in Distal Radius Fractures. *J Hand Surg Asian Pac Vol*. 2018;23(4):520-527.
14. Angelomenos V, Mohaddes M, Itayem R, Shareghi B. Precision of low-dose CT-based micromotion analysis technique for the assessment of early acetabular cup migration compared with gold standard RSA: a prospective study of 30 patients up to 1 year. *Acta Orthop*. 2022;93:459-465.
15. Sandberg O, Tholén S, Carlsson S, Wretenberg P. The anatomical SP-CL stem demonstrates a non-progressing migration pattern in the first year: a low dose CT-based migration study in 20 patients. *Acta Orthop*. 2020;91(6):654-659.

16. Valstar ER, Gill R, Ryd L, Flivik G, Börlin N, Kärrholm J. Guidelines for standardization of radiostereometry (RSA) of implants. *Acta Orthop*. 2005;76(4):563-572.
17. Brodén C, Giles JW, Popat R, et al. Accuracy and precision of a CT method for assessing migration in shoulder arthroplasty: an experimental study. *Acta radiol*. 2019;61(6):776-782.
18. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155.
19. Erdfelder E, Faul F, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods*. 2009;41(4):1149-1160.
20. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39(2):175-191.
21. Gruson KI. CORR Insights®: 3D-printed Handheld Models Do Not Improve Recognition of Specific Characteristics and Patterns of Three-part and Four-part Proximal Humerus Fractures. *Clin Orthop Relat Res*. 2022;480(1):160-162.

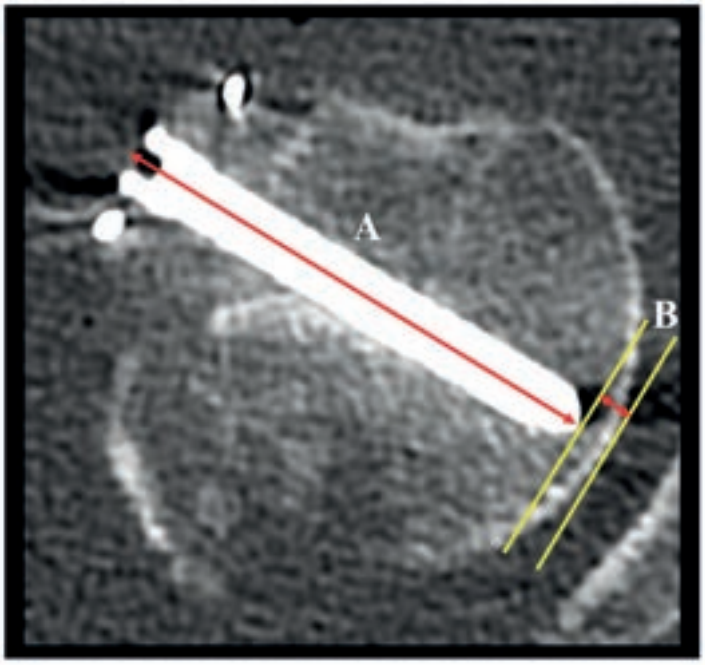
SUPPLEMENTARY MATERIAL



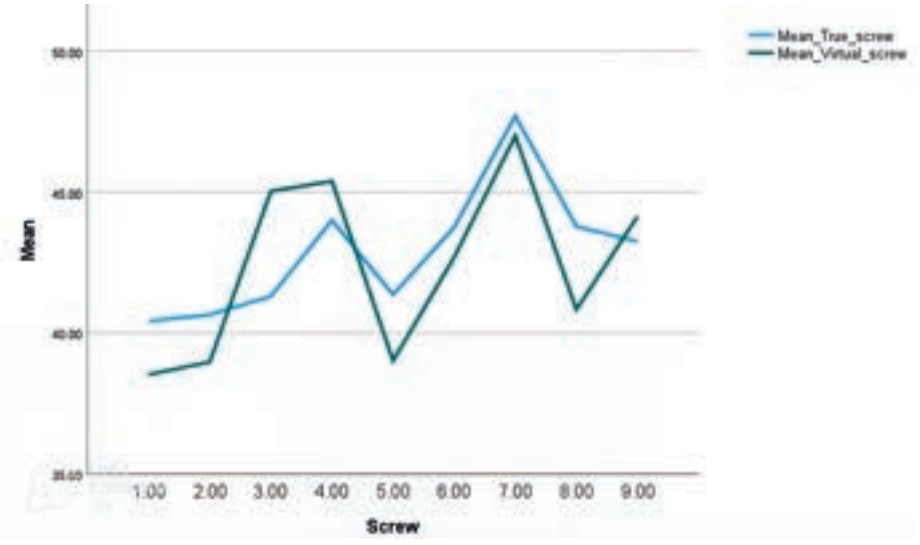
Supplement 1. Measurement of plate position A = line along the base of the plate, B = line perpendicular to A, touching the top of the plate, C = line perpendicular to A, directed along the tip of the head. The humeral head height is the distance between B and C, marked with a bidirectional white arrow.



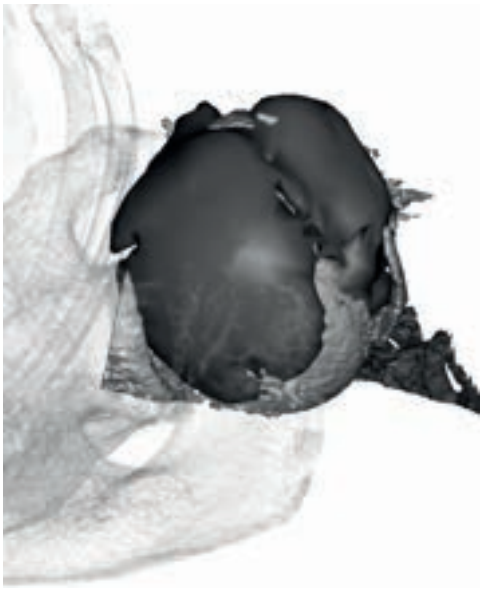
Supplement 2. Assessment of calcar screws. If they were inserted in the inferomedial quadrant they were judged as adequate, otherwise they were categorized as malpositioned which is shown in this example.



Supplement 3. Measurement of screw lengths and tip-to-cortex distance. A = Screw lengths (head-to-tip) were measured on post-operative CT scans. B = the tip-to-cortex was also measured. Multiplanar reconstruction was used to find the correct plane.



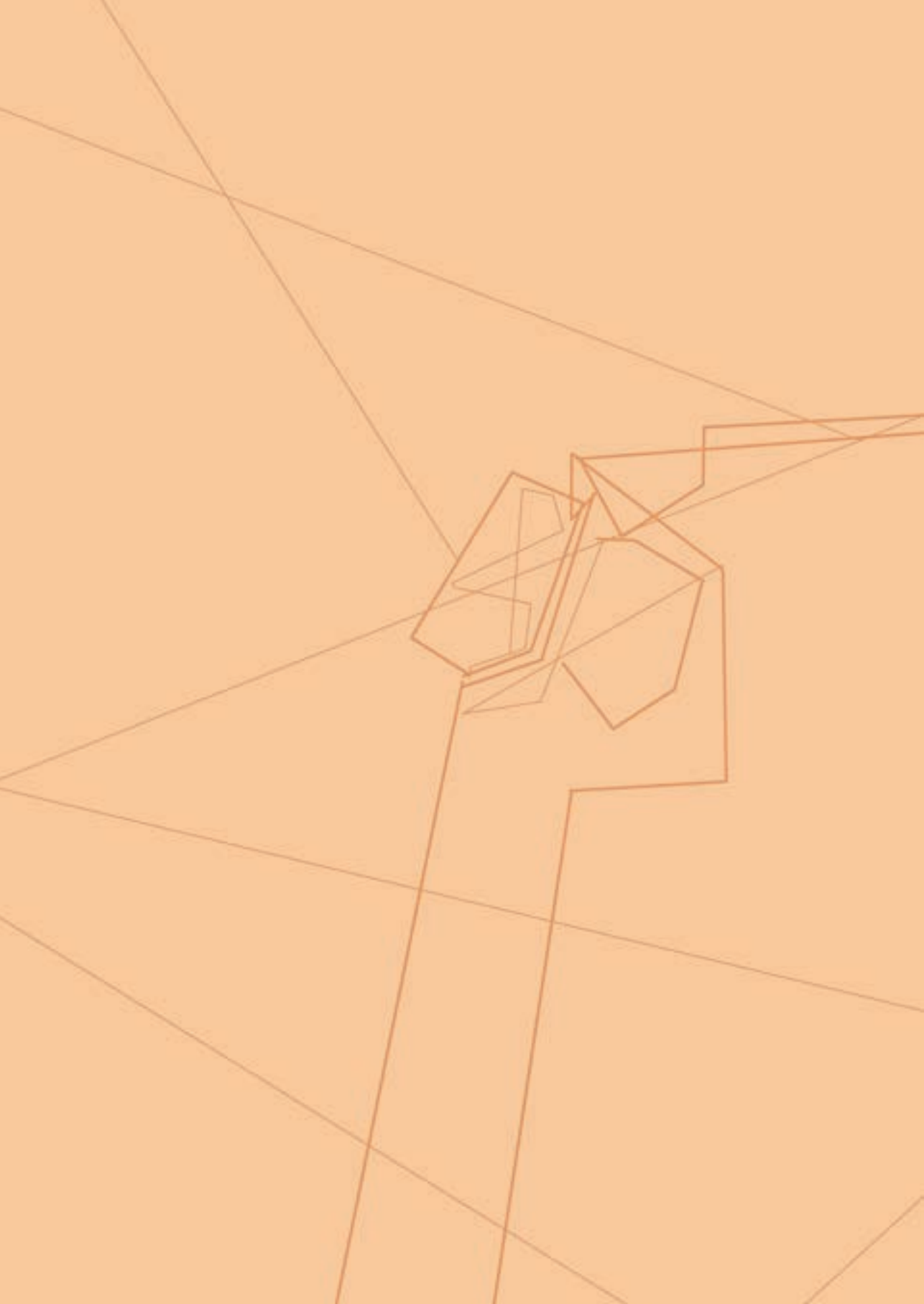
Supplement 4. Virtual screw lengths (green) displayed against the true surgical screw lengths (blue). X-axis: screw holes, Y-axis: mean of screws.



Supplement 5. CTMA results of the model closest to the median rotation of all cases. The difference between the virtual and surgical reduction for this model was 20.6°.

Supplement 6. CTMA results of the model closest to the median medial hinge displacement of all cases. The difference between the virtual and surgical reduction for this model was 8.06 mm.

Supplement 7. CTMA results of the model closest to the median centre of rotation difference amongst all cases: 3.6 mm.



9

Clinical implications, future research, and conclusion

IMPLICATIONS FOR CLINICAL PRACTICE AND FUTURE RESEARCH

Rationale

This thesis explored three phases in the work-up of patients with a proximal humerus fracture: *fracture assessment, patient counselling and decision-making, and pre-operative planning*. We endeavoured to *optimize* workflow and enhance medical knowledge in each of these domains to address the low inter-surgeon agreement on fracture evaluation [1], drive clinical decision-making and improve outcome predictions. First, shortcomings of fracture assessment (i.e. classification and characterisation systems) were quantified and clinical tools were provided to overcome the historically known poor interobserver bias. As such *"new waters"* were explored with 3D printed models and convolutional neural networks (CNNs). Second, treatment outcomes of the uncommon lesser tuberosity fracture, and the effect of collar and cuff on fracture re-alignment were studied to assist in patient counselling and clinical decision-making. Lastly, more *"new waters"* were navigated, by researching virtual 3D planning software. The intention was to see if it could improve adequacy of reduction, plate position and screw lengths to decrease to the contemporary high failure rate of osteosynthesis.

Part 1: fracture assessment

One of the biggest challenges in the field of proximal humerus fractures, is the large bias when interpreting clinical trials in current literature. Patient cohorts are highly heterogenic as they are largely grouped on fracture classifications - which are notoriously subjected to poor interobserver reliability - and not so much on more clinically relevant variables such as co-occurrence of comorbidities, bone quality, functional demand, or specific fracture characteristics [2,3]. We re-emphasized that even simple classification systems are poorly reproducible and showed that 3D printed models are also not the solution to this issue (although they may be useful for other aspects or subspecialties in healthcare such as patient consultation or in the pre-operative planning of extensive bone tumours [4,5]).

CNNs were deployed to objectively assess and classify fractures rather than to recognize subtle, easily overlooked fractures. They showed to be able to accurately detect fractures while its performance to classify them to the simplified Neer's system remained poor [6]. So, despite its global use, our advice is to stop describing

patterns in this way: Neer's classification lacks a good definition, and does not have any clinical consequences. It was also learned that CNNs are not able to outperform humans when tasks become increasingly complex (i.e. recognition of fracture characteristics on plain radiographs). Follow-up studies should therefore incorporate CT imaging. Our library of algorithms requires further development by adding larger (globally acquired) datasets, and by integrating prediction models so that a personalized treatment plan can be created for each patient. They should be tested prospectively, with the eventual goal to publish an open-access algorithm which can group patients reliably for clinical trials but even more to drive surgical decision-making in an objective manner.

Part 2: patient counselling and decision-making

Due to the low incidence of isolated lesser tuberosity fractures, the level of evidence on the best treatment option of this fracture is limited [7]. We therefore collected and synthesized outcomes of patients from the literature and stratified them according to age groups (adolescents, adults), presence of concomitant posterior shoulder dislocation, treatment strategy (non-operative, surgery) and timing of treatment (acute, delayed). Findings can be used to inform patients about complication rates and perceived satisfaction and to aid clinicians in decision-making. Pitfalls were also identified, and should be addressed in further prospective multicentre studies. These include population heterogeneity, degree of displacement and fragment size.

For decades, the concept of collar and cuff is considered an important corner stone in non-operative treatment of surgical neck fractures due to its ability to aid in fracture re-alignment (by the –natural– gravity force of the humerus) [8]. We challenged this concept and revealed that, in the way it is currently applied, it does not provide the desired benefits: there was no significant alignment improvement in the first phase of fracture healing. This highlights that in current form collar and cuff only functions for comfort and pain relief. This study should be followed up with a prospective study design and further data like actual wearing time. It should also be experimented whether weighted collar and cuffs can achieve the desired re-alignment.

Part 3: pre-operative planning

Virtual pre-operative planning is a potential tool to improve surgical outcomes of locking plate fixation [9]. This tool facilitates detailed examination of the fracture, allowing virtual hands-on reduction, plate positioning, and screw sizing prior to

the surgical procedure. The theory is simple: this tool guides surgeons through the technical steps related to the implant, ensuring that they are better prepared and thereby minimizing the likelihood of errors and complications. As a result, this may lead to better implant survival (and thus less re-operations) and a better range of shoulder motion. This proof-of-concept study provided an in-depth instruction on software utilisation demonstrated by multiple figures, revealed an excellent reproducibility on plate position and screw lengths amongst the users, and indicated that it can most likely improve the intra-operative reduction when fixing fractures with a locking plate. Now that the basic principles are summarized, a clinical prospective trial can be conducted where one arm should be assigned to fixing patients with conventional pre-operative planning and the other arm with 3D virtual planning software.

Perspectives on limitations and suggestions for future treatment strategies

Chapter 2 (observer agreement on surgical neck fracture patterns) and Chapter 7 (radiographic outcomes of collar-and-cuff treatment) were designed as hypothesis-generating studies. As such, conclusions should be considered in light of this. In Chapter 2, we assessed a classification system originally designed to guide intramedullary nailing. To expand its applicability to all displaced surgical neck fractures in clinical practice, we modified the system by adding an additional category. Despite this modification, it underscores the inherent subjectivity of radiographic interpretation, which introduces bias and limits consensus on classification systems. Given these challenges, we advocate for moving beyond traditional classification systems and instead focusing on fracture characterisation using clear, reproducible definitions. We also recommend that future studies adhere to the three-phase validation process proposed by Audigé et al [10]. Chapter 7 suggests that immobilization with a collar and cuff is ineffective. However, it is important to acknowledge that certain confounding factors, such as patient compliance with sling use and the presence of comorbidities, were not accounted for in this study. Further research should be conducted prospectively to determine which fracture types are most susceptible to displacement, assess adherence to a collar-and-cuff regimen, and evaluate whether weighted cuffs could enhance fracture alignment. Additionally, future studies should incorporate the remodelling phase to provide a more comprehensive understanding of fracture management.

To improve the treatment of patients, we hope to establish specialized shoulder units staffed with dedicated specialists, ensuring more expert-driven care. Early geriatric involvement in frail and elderly patients, could optimize patient health and enhanced rehabilitation protocols with increased capacity for early surgical intervention (<24 hours) may further improve functional outcomes.

The integration of big data and predictive modelling could aid in patient selection for surgery and in forecasting mortality risks and treatment outcomes. Additionally, strengthening prevention strategies is crucial. Early fall prevention measures, promotion of a healthy lifestyle, and timely initiation of osteoporosis management should be prioritized to reduce fracture incidence and improve long-term health.

Conclusion

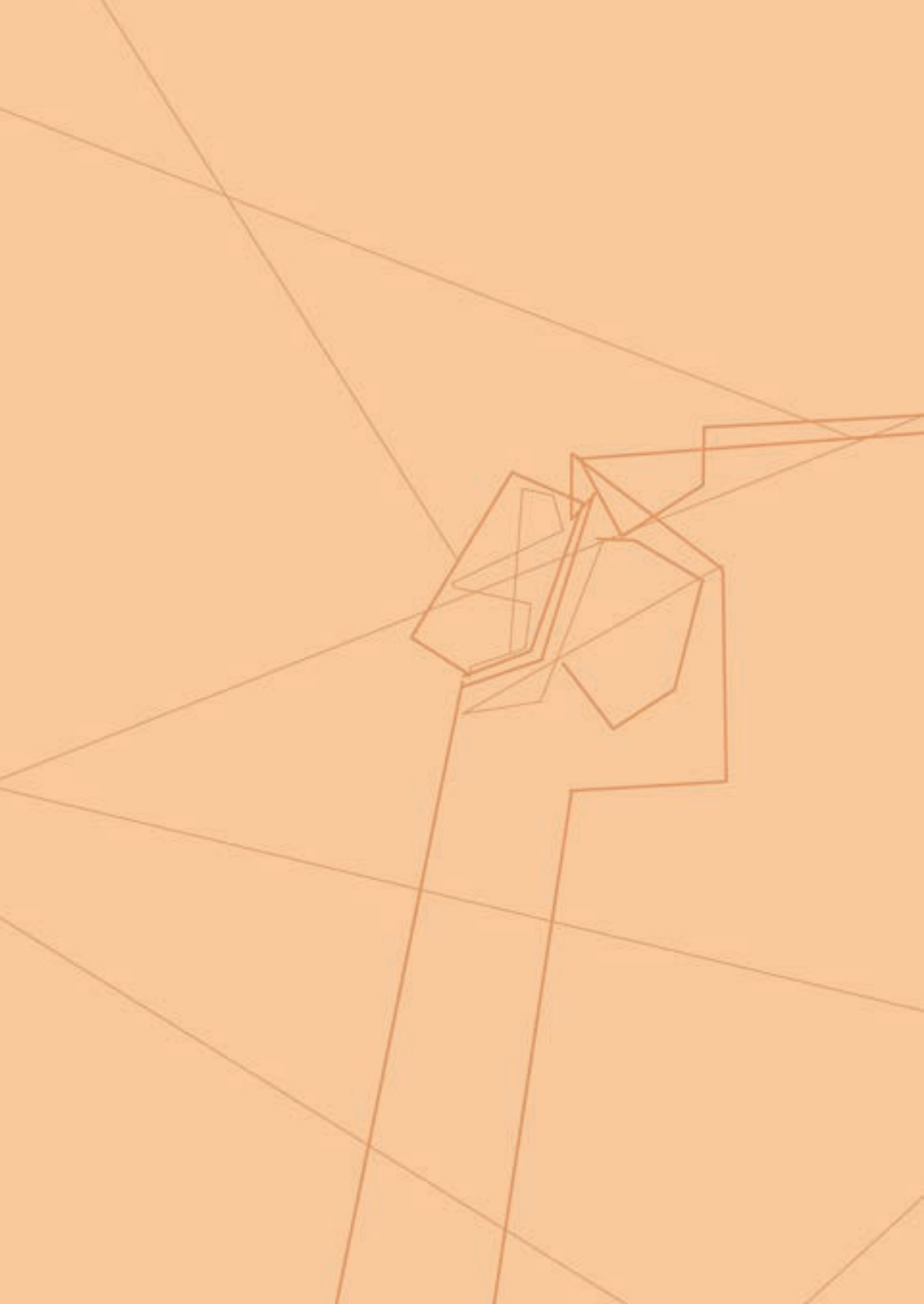
In conclusion, fracture classification systems and characteristics yield a notoriously low interobserver agreement. Even simple classification systems, or adding three-dimensional printed fracture models to conventional assessment methods does not improve reproducibility. The relatively undiscovered waters of CNNs have shown exciting pre-liminary results with regards to ruling out fractures but lack the ability to classify and characterise fractures. Performance of CNNs depend on the nature of the task: the more complex it is for humans the more complex it is for the computer. Further CT-based studies should be deployed before it can be coined as a breakthrough in trauma care with algorithms to be your best man at the emergency department.

Described outcomes of lesser tuberosity serve to inform patients about their expected course of treatment and results. As surgical neck fractures do not re-align in the first phase of bone healing, current insights on collar and cuff treatment should urge clinicians to decide promptly, upon initial presentation, on surgical *versus* non-surgical treatment. Repeat imaging after one to two weeks post trauma, to re-evaluate the decision, is therefore not needed.

Pre-operative virtual planning software has a good inter-surgeon reproducibility and yielded promising results, particularly for enhanced intra-operative fracture reduction. Future trials will point out if this can reduce the high complication rate and whether it should be available by default in surgical theatre.

REFERENCES

1. Bruinsma WE, Guitton TG, Warner JJP, Ring D. Interobserver reliability of classification and characterization of proximal humeral fractures. *J Bone Joint Surg Am*. 2013;95(17):1600-1604.
2. Gomberawalla MM, Miller BS, Coale RM, Bedi A, Gagnier JJ. Meta-analysis of joint preservation versus arthroplasty for the treatment of displaced 3- and 4-part fractures of the proximal humerus. *Injury*. 2013;44(11):1532-1539.
3. Rangan A, Handoll H, Brealey S, et al. Surgical vs nonsurgical treatment of adults with displaced fractures of the proximal humerus: the PROFHER randomized clinical trial. *JAMA*. 2015;313(10):1037-1047.
4. Samaila EM, Negri S, Zardini A, et al. Value of three-dimensional printing of fractures in orthopaedic trauma surgery. *J Int Med Res*. 2019;48(1).
5. Moreta-Martinez R, Pose-Díez-de-la-Lastra A, Calvo-Haro JA, Mediavilla-Santos L, Pérez-Mañanes R, Pascau J. Combining Augmented Reality and 3D Printing to Improve Surgical Workflows in Orthopedic Oncology: Smartphone Application and Clinical Evaluation. *Sensors (Basel)*. 2021;21(4):1-17.
6. Neer CS 2nd. Displaced proximal humeral fractures. I. Classification and evaluation. *J Bone Joint Surg Am*. 1970;52(6):1077-1089.
7. Robinson CM, Teoh KH, Baker A, Bell L. Fractures of the Lesser Tuberosity of the Humerus. *J Bone Joint Surg Am*. 2009;91(3):512-520.
8. Rasmussen S, Hvass I, Dalsgaard J, Christensen BS, Holstad E. Displaced proximal humeral fractures: results of conservative treatment. *Injury*. 1992;23(1):41-43.
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(22):1960-1968.
10. Audigé L, Bhandari M, Hanson B, Kellam J. A concept for the validation of fracture classifications. *J Orthop Trauma*. 2005;19(6):401-406.



10

Summary

SUMMARY

Chapter 2

Displaced surgical neck fractures can be categorized into three patterns: type A (valgus head malalignment due to medial shaft translation), B (humeral shaft separation), and C (varus head malignment due to lateral shaft translation). Our aim was to assess its reproducibility on 30 plain radiographs by a panel of 17 orthopaedic residents and 17 attending orthopaedic trauma surgeons. The interobserver agreement, as well as accuracy, was low. Recognizing surgical neck fracture patterns on radiographs is therefore not reliable.

Chapter 3

The aim was to evaluate if fracture assessments would become more reliable with the adoption of 3D printed fracture models. Twenty proximal humerus 3D printed models were assessed by four orthopaedic residents and four orthopaedic surgeons and compared to conventional assessment with radiographs, 2D and 3D Computed Tomography (CT) scans. Despite their promising theoretical advantages, the models did not improve inter-surgeon reliability. Using these 3D printed models is therefore not recommended for the diagnostic evaluation of a proximal humerus fracture.

Chapter 4

The objective was to develop a convolutional neural network (CNN) for fracture detection and classification on plain radiographs. The algorithm was trained on 1,709 radiographs with multi-rater consensus agreement based on CT imaging as the reference standard. The CNN could adequately rule out proximal humerus fractures, but diagnostic parameters for classification were insufficient for clinical implementation.

Chapter 5

The objective was to develop a CNN to identify greater tuberosity displacement ≥1cm, neck-shaft angle ≤100˚, shaft translation, and articular fracture involvement on plain radiographs. The CNN was trained on 562 radiographs with corresponding CT scans serving as the reference standard, and multi-rater consensus agreement was used as the ground truth. The CNN revealed limited diagnostic ability to detect greater tuberosity displacement ≥1cm and failed to identify any of the other characteristics on plain radiographs. This highlights that applications of artificial

intelligence are not unlimited, and CNNs do not outperform humans when the task becomes increasingly more complex.

Chapter 6

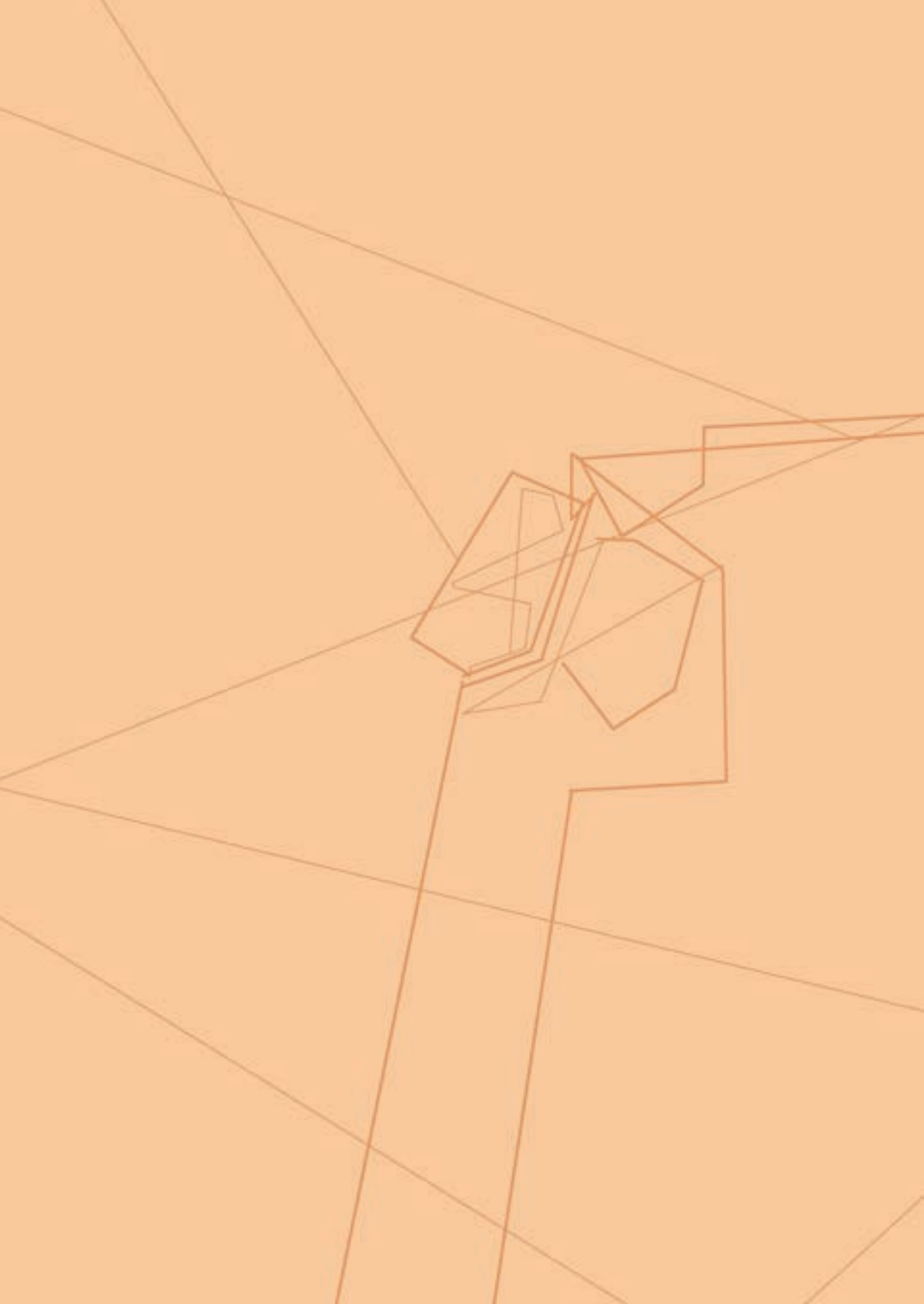
Paediatric and adult patients have acceptable to excellent outcomes after lesser tuberosity fractures and respond well to surgical treatment. Patients who sustained the fracture through a posterior shoulder dislocation were prone to develop more complications during follow-up and had an inferior range of shoulder movements compared to patients who did not have a posterior shoulder dislocation at the time of injury. Data derived from this review can be used for patient consultation and expectation management.

Chapter 7

When a patient sustains a surgical neck fracture of the humerus and is treated non-operatively, surgeons often re-review these patients after a few weeks at the outpatient clinic to assess the alignment of the fracture. The underlying idea of collar and cuff is that it allows the humeral shaft to be pulled down by gravity so that it can re-position below the humeral head. This hypothesis was tested in a cohort of 67 patients with valgus head deformity and medial shaft translation (type A), shaft separation (type B), or varus malalignment with lateral shaft translation (type C). Interestingly, alignment did not improve in any of the types within the first phase of fracture healing (defined as the period in which callus formation was not visible on radiographs). This indicates that hanging down the arm is a dogma, and surgeons should already decide on the need for surgical intervention at the first presentation at the emergency department.

Chapter 8

Complication rates after fixing proximal humerus fractures with locking plates are high. Virtual pre-operative 3D planning is a relatively novel software which can potentially lower this complication rate as surgeons can already practice the reduction, determine the plate position, and measure the required screw lengths before the procedure. Therefore, the potential benefits of 3D virtual planning software were outlined by comparing virtually fixed proximal humerus fractures to the actual surgery. Fourteen fractures were included and planned by eight different orthopaedic surgeons. It was found that the interobserver agreement on plate position and screw length was adequate. Except for fracture reduction, there were no differences with respect to plate position, calcar screw positioning, and screw lengths.



11

Dutch summary

DUTCH SUMMARY

Hoofdstuk 2

Subcapitale humerus fracturen kunnen worden geclassificeerd in drie types: type A (kop in valgus met mediale schachttranslatie), B (volledige translatie van de schacht), C (kop in varus met laterale schachttranslatie). Het doel van deze studie was om de interobserverbetrouwbaarheid van deze classificatie te onderzoeken. Dit werd gedaan op basis van 30 röntgenfoto's die door 17 arts-assistenten in opleiding tot orthopedisch chirurg (AIOS) en 17 orthopedisch chirurgen werden geclassificeerd. Er kon geconcludeerd worden dat de interobserverbetrouwbaarheid en ook de nauwkeurigheid erg laag zijn. Het classificeren van subcapitale humerus fracturen op röntgenfoto's kan dus niet betrouwbaar worden gedaan.

Hoofdstuk 3

Het doel van dit onderzoek was om te bepalen of met 3D-geprinte modellen de lage interobserverbetrouwbaarheid van classificatie- en karakterisatiesystemen verbeterd zou kunnen worden. Twintig 3D-geprinte proximale humerus fracturen werden beoordeeld door vier AIOS orthopedie en vier orthopedisch chirurgen en vervolgens vergeleken met de conventionele methode van fractuurbeoordeling (röntgenfoto's met 2D en 3D CT scans). Ondanks de potentiële theoretische voordelen, droegen de 3D-geprinte modellen niet bij aan een betere inter-beoordelaar betrouwbaarheid. Het gebruik van deze modellen in de klinische praktijk bij het beoordelen van proximale humerus fracturen is daarom niet aan te bevelen.

Hoofdstuk 4

Het doel van deze studie was om een convolutioneel neuraal netwerk (CNN) te trainen en de diagnostische uitkomsten te bepalen voor fractuurherkenning en -classificatie op röntgenfoto's. Het algoritme werd getraind op 1709 röntgenfoto's met CT scans als referentie standaard. De "ground truth" werd bepaald op basis van consensus door meerdere beoordelaars. Het CNN kon uitstekend fracturen uitsluiten, maar niet betrouwbaar classificeren. Klinische implementatie voor fractuurclassificatie is daarom nog niet mogelijk.

Hoofdstuk 5

Het onderzoeksdoel was om een CNN te trainen op herkenning van de volgende fractuurkarakteristieken op röntgenfoto's: tuberculum majus verplaatsing ≥1 cm,

kop-schacht angulatie ≤100˚, schachttranslatie en articulaire betrokkenheid. Het algoritme werd getraind op 562 röntgenfoto's waarbij CT scans werden gebruikt als de referentie standaard, en de "ground truth" werd bepaald middels consensus tussen twee of meer onafhankelijke beoordelaars. Het CNN kon niet nauwkeurig vaststellen of er ≥1 cm tuberculum majus verplaatsing was en was niet in staat om de andere fractuurkarakteristieken te herkennen. Deze uitkomsten benadrukken dat de mogelijkheden van kunstmatige intelligentie niet oneindig zijn en dat computers niet beter zijn dan mensen wanneer de complexiteit van de taak toeneemt.

Hoofdstuk 6

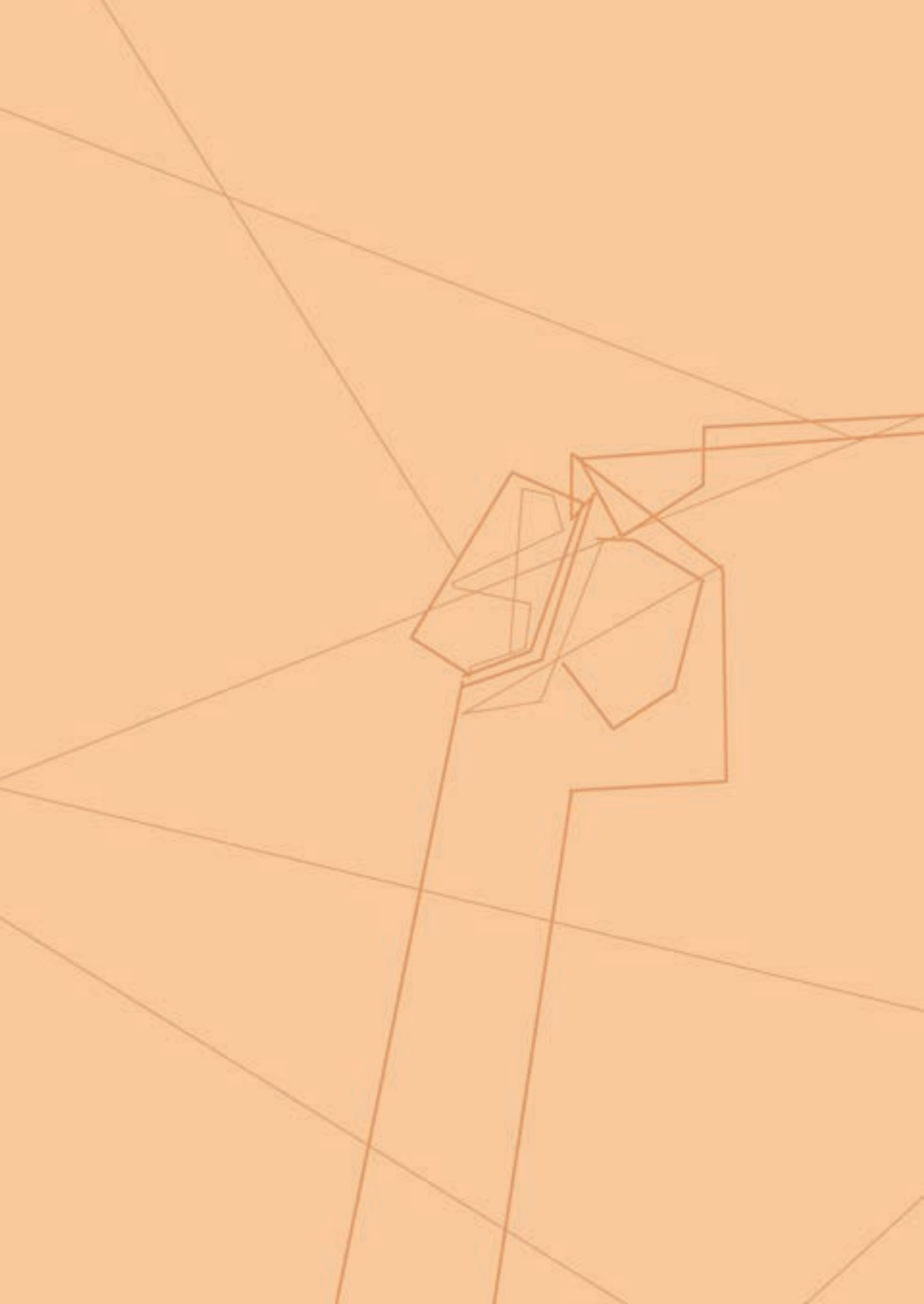
Zowel patiënten jonger en ouder dan 18 jaar hebben goede tot uitstekende uitkomsten na tuberculum minus fracturen, zeker na operatieve ingrepen. Patiënten met een fractuur door een posterieure schouderluxatie hebben een hoger risico om complicaties te ontwikkelen en vaker een minder goede schouderfunctie in vergelijking met patiënten die geen posterieure schouderluxatie hadden. Resultaten uit dit literatuuronderzoek kunnen worden gebruikt om patiënten voorlichting te geven over hun te verwachte uitkomsten.

Hoofdstuk 7

Bij een subcapitale humerus fractuur die niet-operatief behandeld wordt, worden patiënten vaak na één of twee weken teruggezien op de polikliniek om de stand van de fractuur te herbeoordelen. Zo nodig kan er dan alsnog gekozen worden voor een operatieve ingreep. Het idee hierachter is dat patiënten met de sling hun arm laten uithangen, waardoor de humerus na verloop van tijd weer beter onder de kop terechtkomt. Deze hypothese hebben wij getoetst in een groep van 67 patiënten die valgus koprotatie hadden met een partiële mediale schachttranslatie (type A), volledige schachttranslatie (type B) of een laterale schachttranslatie met varuskanteling van de kop (type C). Er kon geconcludeerd worden dat in de eerste weken van botgenezing (gedefinieerd als de periode waarin callusvorming nog niet radiografisch zichtbaar was) de stand van de humerus ten opzichte van de kop niet verbeterde. Dit impliceert dus dat het uithangen van de arm een dogma is en dat artsen direct al kunnen besluiten of een patiënt wel of niet geopereerd dient te worden op basis van het fractuurtype.

Hoofdstuk 8

Complicaties na plaatfixatie van proximale humerus fracturen komen veel voor. Virtuele pre-operatieve 3D planning is een relatief nieuwe software die deze complicatiekansen mogelijk kan verminderen, doordat de chirurg de fractuurrepositie al een keer kan uitvoeren en ook de positie van de plaat kan bepalen (inclusief de lengtes van de schroeven). De opzet van deze studie was dus om de potentiële voordelen van deze plannings tool te beschrijven door virtueel geplande fracturen te vergelijken met de uitkomsten van de echte operaties. Veertien fracturen werden geïncludeerd en gepland door acht orthopedisch chirurgen. De interobserverbetrouwbaarheid van plaatpositie en schroeflengtes was erg goed. Met uitzondering van de fractuurrepositie waren er geen verschillen met betrekking tot de plaatpositie, calcar schroef positionering en lengtes van de schroeven.



12

Supplements

TABLE OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

TABLE OF ACRONYMS

3D	Three-dimensional
AI	Artificial intelligence
AO	Arbeitsgemeinschaft für Osteosynthesefragen
AP	Anteroposterior
AUC	Area under the curve
BT	Biceps tendon
CI	Confidence interval
CNN	Convolutional neural network
CT	Computed tomography
CTMA	Computed tomography micromotion analysis
HHH	Humeral head height
IA	Intra-articular
ICC	Intraclass correlation coefficient
IM	Inferomedial
LCLC	Labrocapsular ligamentous complex
LMM	Linear mixed modelling
LT	Lesser tuberosity
MLG	Maximal lateral gap
MMG	Maximal medial gap
MPR	Multiplanar reconstruction
MRI	Magnetic resonance imaging
NOS	Newcastle-Ottawa scale
NR	Not reported
NSA	Neck-shaft angle
ORIF	Open reduction and internal fixation
OTA	Orthopaedic Trauma Association
PHF	Proximal humerus fracture
PROM	Patient-reported outcome measure
PSD	Posterior shoulder dislocation
RC	Rotator cuff
ROM	Range of motion
VAS	Visual analogue scale

BIBLIOGRAPHY

Reinier Willem Alfred Spek, 31 years of age, grew up in The Hague, the Netherlands. He attended the Paschalisschool for his primary education, and already as a young kid he displayed a passion for sports, particularly tennis and field hockey. After completing primary school, he went to the Vrijzinning Christelijk Lyceum. During his high school time he continued playing field hockey and was selected for the first team of Klein Zwitserland at the age of 16. It also became apparent that he wanted to study medicine, so after graduation, in 2012, he moved to Utrecht to pursue a medical degree. His interest in medicine grew rapidly and soon he knew that he wanted to develop surgical skills to treat patients. Alongside his study he played for MMHC Voordaan Heren 1, became a member of the Utrechtsch Studenten Corps, and was introduced to orthopaedics through research on the meniscal allograft transplantation by Dr. E.R.A. van Arkel. Reinier's studies progressed well, and in 2019 he went to Cape Town for a hands-on trauma surgery internship in the Level-1 Groote Schuur Hospital. Not much later he started his last year of medicine, where he developed a keen interest in shoulders through research collaboration with Dr. D.F.P. van Deurzen. It was also the year where he decided to aspire a future as orthopaedic surgeon, established by his elective at the orthopaedic surgery department at OLVG, Amsterdam. His interest in shoulders was further developed by Prof. Dr. M.P.J. van den Bekerom and marked the beginning of his academic career. Together with Prof. Dr. M.P.J. van den Bekerom, prof. Dr. J.N. Doornberg, Prof. Dr. R.L. Jaarsma, and Prof. Dr. P.C. Jutte, he secured funding and set up a research line focused on emerging innovations for proximal humerus fractures which ultimately formed the current PhD booklet.


olvg


Spek
Spek

Most of his research was done in Adelaide, Australia, where he created life-lasting memories, won two consecutive field hockey titles in the South Australian premier league, and stayed there for almost three years. Upon his return in Amsterdam in 2022, he started working as a general surgical resident (not in training) in OLVG Amsterdam, which was 10 months later followed by a position as an orthopaedic surgery resident (not in training) in the same hospital. Alongside his clinical work, he started with triathlon and in 2024, he completed an Ironman, submitted his PhD thesis, and was accepted into the orthopaedic surgery training program in the Middle-West region of the Netherlands. In January 2025, he began his 1.5-year training in the general surgery department, where he is currently developing his fundamental surgical skills in the trauma unit at OLVG. In July next year, he will start with his 4.5-year orthopaedic surgery training. Being a resident now, he takes great satisfaction in learning the craftsmanship of surgery. He aspires to become an orthopaedic surgeon who makes a meaningful difference for his patients, colleagues, and the rapidly evolving healthcare system.

PUBLICATIONS

1. Spek RWA, Smith WJ, Sverdlov M, et al. Detection, classification, and characterization of proximal humerus fractures on plain radiographs. *Bone Joint J*. 2024;106-B(11):1348-1360
2. Mennes SRJ, Spek RWA, Vorrink S, et al. Long-term Functional Outcomes of Non-operative Treatment in Patients with Humeral Surgical Neck Fractures. Submitted to *European Journal of Trauma and Emergency Surgery*.
3. Spek RWA, Ring D, van den Bekerom MPJ. Letter to the editor: response to "Does primary treatment of proximal humerus fractures show favourable functional outcomes over secondary treatment with reverse shoulder arthroplasty?" *Shoulder Elbow*. 2024;16(5):569-570.
4. Spek RWA, van den Bekerom MPJ, Jutte PC, et al. Pre-operative Virtual Three-Dimensional Planning for Proximal Humerus Fractures: A Proof-of-concept Study. *Shoulder Elbow*. 2024 Jan;16(4):397-406.
5. Spek RWA, Hoogervorst LA, Brink RC, Schoones JW, van Deurzen DFP, van den Bekerom MPJ. Ten technical aspects of baseplate fixation in reverse total shoulder arthroplasty for patients without glenoid bone loss: a systematic review. *Clin Shoulder Elb.* 2024 Mar;27(1):88-107.
6. Spek RWA, Spekenbrink-Spooren A, Vanhommerig JW, et al. Primary reverse total shoulder arthroplasty for fractures requires more revisions than for degenerative conditions one year after surgery: an analysis from the Dutch Arthroplasty Register. *J Shoulder Elbow Surg.* 2023;32(12):2508-2518.
7. Koper MC, Spek RWA, Reijman M, van Es M, Baart SJ, Verhaar JAN, Bos PK. Are serum cobalt and chromium levels predictors for patient-reported outcome measures in the ASR hip resurfacing arthroplasty? *Bone Joint J*. 2023;105-B(7):775-782.
8. Spek RWA, Hoogervorst LA, Elias MEC, et al. Management of Displaced Humeral Surgical Neck Fractures in Daily Clinical Practice: Hanging Does not Re-align the Fracture. *Arch Orthop Trauma Surg*. 2023;143(6):3119-3128.
9. Spek RWA, Kim LJ, Traumaplatform 3D consortium. What Is the Interobserver Agreement of Displaced Humeral Surgical Neck Fracture Patterns? *Clin Shoulder Elb*. 2022;25(4):304-310.
10. Spek RWA, Schoolmeesters BJA, Oosterhoff JHF, et al. 3D-printed Handheld Models Do Not Improve Recognition of Specific Characteristics and Patterns of Three-part and Four-part Proximal Humerus Fractures. *Clin Orthop Relat Res*. 2022;480(1):150-159.
11. Spek RWA, Schoolmeesters BJA, den Haan C, Jaarsma RL, Doornberg JN, van den Bekerom MPJ. What are the patient-reported outcomes, functional limitations, and complications after lesser tuberosity fractures? a systematic review of 172 patients. *JSES Int*. 2021;5(4):754-764.
12. van Deurzen DF, Auw Yang KG, Onstenk R, et al. Long head biceps tenotomy is not inferior compared with suprapectoral tenodesis in arthroscopic repair of non-traumatic rotator cuff tears: A multicenter non-inferiority randomized controlled clinical trial. Mar 2021. *Arthroscopy*. 2021 Jun;37(6):1767-1776.e1.
13. Spek RWA, Smeeing DPJ, van den Heuvel L, et al. Complications after Surgical Treatment of Geriatric Ankle Fractures. *J Foot Ankle Surg.* 2021;60(4):712-717.
14. Hoogervorst LA, Spek RWA, van den Bekerom MPJ. Comments to: "Outcomes of surgical fixation of greater tuberosity fractures: A systematic review" by S.R. Huntley, E.J. Lehtonen, J.X. Robin, A.M. Arguello, D.M. Rouleau, E.W. Brabston, B.A. Ponce, A.M. Momaya published in Orth Traumatol Surg Res. 2020;106(6):1119-1126. *Orthop Traumatol Surg Res*. 2021;107(4):102919.

15. Spek RWA, Doornberg JN, Ring D, van den Bekerom MPJ. Can surgeons differentiate between painful shoulders that grow Cutibacterium acnes and infection benefitting from treatment? *Shoulder Elbow.* 2021;13(2):149-150.

16. van der Wal RJP, Nieuwenhuijse MJ, Spek RWA, Thomassen BJW, van Arkel ERA, Nelissen RGHH. Meniscal allograft transplantation in The Netherlands: long-term survival, patient-reported outcomes, and their association with preoperative complaints and interventions. *Knee Surg Sports Traumatol Arthrosc.* 2020 Nov;28(11):3551-3560.

FINANCIAL SUPPORT

The following organizations are sincerely appreciated for their funding, which facilitated the research conducted in this PhD thesis: Flinders Foundation (Adelaide, Australia), Prins Bernhard Cultuurfonds (Amsterdam, the Netherlands), Stichting Anna Fonds NOREF (Mijdrecht, the Netherlands), Stichting Prof. Michaël-van Vloten Fonds (Venray, the Netherlands), Stichting Wetenschap OLVG (Amsterdam, the Netherlands), and Stichting Zabawas (the Hague, the Netherlands).

ACKNOWLEDGEMENTS

This thesis would not have been possible without my extremely dedicated, involved and motivating supervisory team: **Ruurd Jaarsma, Paul Jutte, Job Doornberg and Michel van den Bekerom.** Your coaching and clinical knowledge inspired me greatly and I am deeply grateful for that. You always gave me complete flexibility and ownership of my time while still pushing me to achieve timely results. I cannot be more grateful for having such an inspiring team around me. Without your guidance, I would not be able to look back on such a joyful time and this most fantastic PhD journey.

Supervisor, **Ruurd**, I have always been impressed by your pragmatic approach to research and clinical work. You have a remarkable ability to break down difficult topics into simple, understandable concepts. Since I had spent a big portion of my time in Australia, I had to privilege of working closely with you. You guided me with many one-on-one meetings, and you were always ready to help when I walked into your office for advice. You gave me a warm welcome from the very beginning I stepped into Flinders Medical Centre, and I am grateful that you helped me bring Adelaide into my life. I appreciate your invitations to your house in Middleton, where we could relax and go surfing. It also gave me a glimpse into your family life, from which I could see that you have built something truly amazing in Adelaide, all the more impressive knowing that you once moved there as a Dutchie to an unfamiliar country. I hope to return one day for a fellowship position to further learn from your surgical expertise.

Supervisor, **Paul**, you have such a calm, approachable yet decisive character. You come across very knowledgeable and as someone you can naturally go to for advice. Hence, when I first met you and heard that you were going to be part of the supervisory team, I was very grateful. I am sure this is something which patients see as well and probably make them feel at ease when they consult you. You also exhibited this to me which re-assured that we would bring this PhD journey to a success. Especially during the final phase of this thesis, you were closely involved and guided me very well with writing the introduction and discussion, combining the chapters, and coordinating the submission to the university and swift assembly of the assessment committee. I hope our paths meet again in the future and that we can collaborate in the clinical setting.

Co-supervisor, **Job**, thanks! That's how you always end your emails and now it is my turn to say thanks to you. In the first year of my PhD, you were still in Adelaide, so I had to privilege to do research with you on a day-to-day basis. It was incredible to see how you made your researchers feel part of the group right away: you invited us to beach barbecues, after-work drinks, and we even road our bikes up to Mount Lofty. I am impressed by your networking skills, the way how you established your innovative research line on a global level, and how you consistently involve the right people in each project. You can see that you are very passionate about research and that also inspired me "to stay hungry". I admire how you combine your clinical work with a vibrant social life, and your enthusiasm for sports. I also appreciate your regular check-ins with encouraging emails on my PhD progress and your genuine interest in sharing photos whether related to travel, research or personal milestones. Thanks for your endless motivation and for making this adventure possible. Since you are now working directly with Paul, I hope that there will be an opportunity to come to Groningen for some time so I can work with both you and learn from your clinical experience in orthopaedic trauma surgery.

Co-supervisor, **Michel**, you were the first person of my supervisory team, and as such, you played a key role in setting up this PhD trajectory. We collaboratively designed my first shoulder projects, secured my initial funding, and you connected me with the other members of the team. Over time we outlined more and more studies based on your feasible and well-thought-out ideas. At every step in the process, you were involved and supportive. I greatly appreciate how you always replied to my emails swiftly and with a clear answer: it was never vague but always a yes or no. That really kept the speed in the research and was therefore truly invaluable. You also gave me helpful advice for my further clinical career and how to stay competitive in the years leading up to the orthopaedic training program interviews. Over time I came to know and appreciate your dry sense of humour, your outstanding ability to retain information, and being up to date with the latest literature. Having already worked with you, I know how swiftly and skilful you perform your surgical procedures. As I will conduct a substantial part of my training program at the OLVG, I cannot wait to learn the ropes of shoulder surgery from you.

Ruurd, I learned from you how to simplify things when they seem complex or overwhelming; **Paul**, I learned from you how exhibition of calmness can make people feel at ease and how important that is in achieving a goal; **Job**, I learned

from you the importance of collaboration and how sharing knowledge may create big opportunities; **Michel**, I learned from you the meaning of clarity and the positive impact of swift but accurate decision-making. **Paul, Ruurd, Job, and Michel**, these are one of your best qualities, which I try to incorporate into in my skillset as person and a doctor. **Thank you so much.**

Over the years, I have collaborated with many great persons who made this research possible and there are many friends who always supported me along the way. I am very grateful to share this work with you, and I would like to say a big thanks to you.

High school friends, **Bas Zaalberg, Maarten van Arkel, Marc Borstlap, Matthijs de Jong, Rafael Uyttendaele, Siebren Posthuma, Tom Hiebendaal.** I feel it is very special how close we are as a group. From the very beginning we have shared countless hangout sessions, deep conversations, and laughs, creating unforgettable memories for almost two decades now. We know each other's parents, and I even have been housemates with many of you. We have travelled together, played sports together, and through all these years you remain the ones I spend most of my time with. You are a constant source energy and inspiration for me. You are one of the reasons I get out of bed every morning. Cheers to all the adventures and awesome times ahead of us. I also cannot refrain from mentioning your girlfriends*, Ida Mae de Waal, Janey Franssen, Maaike van Son, Astrid ten Bosch, Chloé Millo.* Being part of our group for such a long time now, you have become "part of the furniture"*.* It's been wonderful having you around and to see my friends flourish alongside you.

Utrecht friends, **Jaarclub Klapschaats**, **Huize Boyo**. It is wonderful to look back on all the fun we shared during our student days, and I am thankful that I still see many of you on a regular basis. Special gratitude to **Caspar Bosma**, **Pim Hofman**, and **Wessel de Bruijn** for becoming such close friends over time. It's epic to be around you and share amazing moments in life with you. *Pim Hofman,* it was unforgettable

to team up with you and push our unknown limits during our Ironman journey. Also, a big shout out to your amazing girlfriends, *Willemijn Agelink, Chantal Couzijn and Leora de Wit.*

Hockey Club Seacliff, **Premier league team**, **Metro 1 team**, and **coaches**. Thanks for welcoming me, improving my hockey skills, and teaching me the Aussie way of life. You are truly amazing, and I will not forget how special it was to win the title back-to-back together. Let's sing the song one more time.

Ohhhhhh, we're from Tigerland
A fighting fury
We're from Tigerland.
In any weather you will see us with a grin, hey!
Risking head and shin, hey!
If we're behind then never mind
We'll fight and fight and win
For we're from Tigerland
We'll never weaken till the final whistle's blown!
Like the tiger of old
We're strong and we're bold
Oh we're from Tiger,
Yellow and black,
We're from Tigerland.

Orthopaedic fellows and partners, **Henrik Åberg & Ulrika Åberg, Borg Leijtens & Imke Booijink**, **Anne Vonk & Paul van der Voort**, **Femke Hagenmaier & Rob Carpay**, **Jan Louwerens & Nina van Hattum**, **Maarten Koper & Tanja Kanders**. I am grateful for having you around at Flinders Medical Centre, and for the unique moments we shared together ranging from surf sessions to dinners and epic camping trips. You are a fantastic mix of people, and it felt as if you were my Australian family during our time Down Under.

Byron Walton. We already knew each other from hockey at Voordaan, and seeing your familiar face really helped me feel at home in Adelaide. Thanks for the good conversations, the laughs, and for making fun of "datumprikker". You have life well

sorted, and it was wonderful to meet your wife *Yves Charlesworth* as well. You two make a fantastic couple and I wish you all the best.

Joint Research, **Nienke Willigenburg** and **Sigrid Vorrink**. You taught me the fundamentals of research and helped me tremendously with my first research protocols, funding acquisitions and ethical request. Thank you.

Department of orthopaedic surgery, Flinders Medical Centre. Thank you for your commitment, engagement in discussions, and for critically reviewing my projects. A special word of thanks to **Sylvia McAndrew** for welcoming me and for making my time in Adelaide as smooth as possible.

Department of orthopaedic surgery, OLVG. In 2019, during my elective as a sixth-year medical student in your department, it was your enthusiasm and passion that inspired me to choose the path of orthopaedic surgery. You are a unique and fantastic group of persons. Thanks for motivating and teaching me. I very much look forward to work together again.

Department of general surgery, OLVG. Thanks to all residents and staff for making work enjoyable, and for teaching me the fundamentals of surgery. A special thanks to clinical directors **Carel Goslings** and **Hille Witjes**, whose support made this possible.

The Traumaplatform Consortium. It has been great to work in such a big research group. This really improved all our projects and allowed us to always think bigger and brainstorm new ideas. Special thanks to **Marouska van Boxtel** and **Jasper Prijs**, as my direct colleagues in Flinders Medical Centre. Also credits to talented researcher **Stijn Mennes** who is continuing the current line of research with impressive follow-up studies.

Co-authors, **Frank IJpma**, **Nini Jonkman, Arthur van Noort, William Smith, Marat Sverdlov, Joost Vanhommerig and Hugo van Veen.** Thanks for your clear thoughts and thorough edits on my papers, some of which extended even beyond the scope of this PhD.

Orthopaedic surgeon and clinical director, **Derek van Deurzen**. You sparked my enthusiasm for shoulders during our research collaboration in 2019 and connected

me with the right person to begin my PhD. Especially in clinical practice, you have pushed me further, inspired, and motivated me. Thanks for creating opportunities and for allowing me to work in this amazing field.

Orthopaedic surgeon and mentor, **Ewoud van Arkel**. It is an absolute privilege to know you. Not only because of your warm hospitality every time I came over to see *Maarten*. But also, for guiding me in choosing the right paths at many key moments of my medical career. Your mentorship is invaluable.

Friend, cover designer, and illustrator, **Siebren Posthuma**. Your illustrations are truly epic. Thanks for your creativity and dedication you put into this.

Paranymphs, **Rafael Uyttendaele**, **Maarten van Arkel**. Thanks for being such amazing friends. Your trustworthiness, humoristic approach to life and infinite support in both good and more difficult times have been unwavering. Probably without even realizing it yourselves, you taught me presence, creative thinking, letting go of control, and making small things count in life.

Brother-in-law, **Arno de Wolf,** my brother's girlfriend**, Denise Campbell.** Arno, thanks for being such an amazing husband to my sister and for being a reliable, steady presence in our family. Denise, a warm welcome to our family, it is wonderful to have you around.

My niece, **Julie de Wolf**, you have only just been born and have a lifetime of adventures ahead of you. I cannot wait to see you grow up.

My siblings, **Hugo Spek, Leonie Spek**. I am blessed with both of you and really foster our unique relationship. I can be completely myself around you, and I just wanted to say that I am extremely proud of you.

My parents, **Margo van Holthe, Martijn Spek**. This journey simply would not have been possible without you, and I wouldn't be the person I am right now, without you. You mean the world to me.