#### **ORIGINAL ARTICLE**



# New technologies for the classification of proximal humeral fractures: Comparison between Virtual Reality and 3D printed models—a randomised controlled trial

Rafael Almirón Santa-Bárbara<sup>1,2</sup> · Francisco García Rivera<sup>3</sup> · Maurice Lamb<sup>3,4</sup> · Rodrigo Víquez Da-Silva<sup>5</sup> · Mario Gutiérrez Bedmar<sup>6,7,8</sup>

Received: 25 June 2022 / Accepted: 17 January 2023 / Published online: 4 February 2023 © The Author(s) 2023

#### **Abstract**

Correct classification of fractures according to their patterns is critical for developing a treatment plan in orthopaedic surgery. Unfortunately, for proximal humeral fractures (PHF), methods for proper classification have remained a jigsaw puzzle that has not yet been fully solved despite numerous proposed classifications and diagnostic methods. Recently, many studies have suggested that three-dimensional printed models (3DPM) can improve the interobserver agreement on PHF classifications. Moreover, Virtual Reality (VR) has not been properly studied for classification of shoulder injuries. The current study investigates the PHF classification accuracy relative to an expert committee when using either 3DPM or equivalent models displayed in VR among 36 orthopaedic surgery residents from different hospitals. We designed a multicentric randomised controlled trial in which we created two groups: a group exposed to a total of 34 3DPM and another exposed to VR equivalents. Association between classification accuracy and group assignment (VR/3DPM) was assessed using mixed effects logistic regression models. The results showed VR can be considered a non-inferior technology for classifying PHF when compared to 3DPM. Moreover, VR may be preferable when considering possible time and resource savings along with potential uses of VR for presurgical planning in orthopaedics.

**Keywords** Proximal humeral fracture  $\cdot$  Three-dimensional printed models  $\cdot$  Virtual Reality  $\cdot$  Interobserver agreement  $\cdot$  Shoulder surgery planning

- Rafael Almirón Santa-Bárbara ralmiron@uma.es
- Department of Orthopaedic Surgery and Traumatology, Hospital de Antequera, Malaga, Spain
- School of Medicine, Universidad de Málaga, Málaga, Spain
- School of Engineering Science, University of Skövde, Skövde, Sweden
- School of Informatics, University of Skövde, Skövde, Sweden
- Department of Orthopaedic Surgery and Traumatology, Hospital Universitario Virgen de la Victoria, Málaga, Spain
- Preventive Medicine and Public Health Department, School of Medicine, University of Málaga, 29010 Málaga, Spain
- Biomedical Research Institute of Malaga-IBIMA, 29010 Málaga, Spain

#### **Abbreviations**

RSA Reverse shoulder arthroplasty

HA Hemi-arthroplasty

ORIF Open reduction with internal fixation CT-Scan Computerized tomography scan 3DPM Three-Dimensional printed models

VR Virtual Reality

PHF Proximal humeral fracture

OR Odds-Ratio PGY Post-grade year

# 1 Introduction

Proximal humerus fractures (PHF) are the seventh most frequent fractures in adults and the third in patients over 65 years old (Roux et al. 2012), which means it is a condition that will increase parallel to population ageing. In Sweden, for example, the incidence of PHF increased 31% from 92.7 per 100,000 person-years in 2001 to 121.9 per

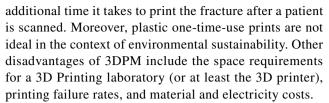


100,000 person-years in 2012, according to the National Register (Sumrein et al. 2017). Moreover, PHF can involve high costs for the healthcare systems. For example, it is estimated that the total cost of PHF was 48.5 M€ for the Netherlands between 1986 and 2012, including both direct and indirect costs (Mahabier et al. 2015). While the cost can be quite high, it also highly depends on the treatment provided. Among the main treatment options, Reverse Shoulder Arthroplasty (RSA) is the most expensive 90-days post-surgery (\$16,151), followed by hemiarthroplasty (HA) (\$9348), Open Reduction with Internal Fixation (ORIF) (\$6745) and non-operative treatment (\$1932) (Levy et al. 2017). Due to these cost differences and resources required for different treatments, some authors emphasize the need to select the correct treatment for each patient. Further, in medicine, assessment of treatment cost-effectiveness must include the trade-offs involved in choosing among alternative interventions to maximize overall health outcomes given the available resources. As a result, cost-effectiveness includes considerations of immediate patient treatment expenses, available healthcare resources, economic and societal impacts of long-term injury, and time spent in therapies (Neuman and Sanders 2017). Thus, some have argued that ORIF might sometimes be a more cost-effective choice than HA or that surgical treatments are not cost-effective versus non-operative treatments on PHF involving the humerus neck (Corbacho et al. 2016; Thorsness et al. 2016).

Given the likely increasing incidence of PHF and the range of treatment costs, it is critical to have effective methods for determining appropriate treatment. Commonly, the number of fragments and their displacement are two important variables relevant to PHF treatment decisions. However, how exactly these variables relate to treatment selection remains controversial, and there is significant heterogeneity about how to weigh the variables relative to different fracture patterns. Generally, it is accepted that minimally displaced fractures are treated conservatively, while displaced, comminuted or angulated fractures are usually treated with surgical methods (Schumaier and Grawe 2018).

Three-dimensional printed models (3DPM) have constituted a solid alternative to traditional methods (i.e. CT-Scan and conventional radiographs). They have shown a higher interobserver and intraobserver agreement in the classification of PHF using Neer classification (Bougher et al. 2021; Iordens et al. 2016). By providing a physical 3D model of the fracture, 3DPM allows observers to classify the fractures properly, using both visual and tactile interactions with the 3DPM to diagnose fractures accurately. A further benefit of 3DPM is that the models can be used for surgical planning and for practising the procedure after selecting an initial diagnostic route.

While 3DPM has proven effective, it can be limited by the need for specialized printing technologies and the



Virtual reality (VR) can be considered a possible new tool in the daily workflow of Orthopaedic Surgery. This technology has been used for various purposes in shoulder and elbow surgery, such as surgical planning or educative aims (Lohre et al. 2020). However, to the best of our knowledge, no previous literature has presented a study of the effectiveness of VR in PHF classification. Given that VR does not require a physical 3D model to be made, as in 3DPM, it may provide a more sustainable alternative to 3DPM.

In the present work, we propose VR as a new implement for achieving accurate PHF classification among orthopaedic surgery residents without the disadvantages of 3DPM noted above. This study aims to assess the accuracy of evaluations when using VR as compared to classifications using 3DPM. Evaluations were considered accurate or correct when they agreed with classifications provided by a committee of experts, and incorrect when these evaluations differed. To accomplish our aim, we performed a study with residents of orthopaedic surgery from different hospitals who were randomised into two groups (3DPM versus VR models).

# 2 Materials and methods

# 2.1 Sample obtention and participants

## 2.1.1 Sample: patients with PHF

We analysed the entire population of patients of Hospital de Antequera (Malaga, Andalusia, Spain) diagnosed with PHF from 2012 to 2022 with a CT-Scan (n = 63).

Three inclusion criteria were established: (1) patients must have had a CT-Scan on the acute moment of the fracture with at least 80 slices uploaded to the image database of the hospital, (2) all treatment and follow-up of the patients was done at Hospital de Antequera and (3) patients signed an informed consent permitting the use of their images for medical or investigation purposes. Further, a single exclusion criterion was used: any previous fractures or pathology on the fractured shoulder or fracture-dislocation.

A total of 34 patients met all the previous conditions, and their images were used as the sample for this study. The CT-Scan model used was SOMATOM Emotion 16-slice 2007 (Siemens), and DICOM files of the mentioned CT-Scans were processed.



Of the patients who met the established criteria, 16 had fractures on their right humerus and the other 18 on their left humerus.

## 2.1.2 Participants: orthopaedic surgery residents

The study involved all 37 residents from three hospitals of different levels of attention in Andalusia. One of these participants was excluded due to technical issues, leaving 36 participants included in the presented analyses (n = 36; male = 25, female = 11; average age = 27.75). All participants signed an informed consent form allowing the anonymized use of their responses.

Hospital levels are defined according to the definitions adopted by WHO-CHOICE: primary-level hospitals have few specialities and under 200 beds, secondary-level hospitals are highly differentiated by function with 200 to 800 beds, and tertiary-level hospitals are highly specialized and equipped with 300 to 1500 beds (Mulligan et al. 2003).

Each of the 36 orthopaedic surgery residents surveyed classified a total of 34 PHF using both Neer and HGLS classifications and selected a potential treatment.

# 2.2 Gold standard classification

While many accepted classifications exist for PHF, Neer classification is probably the most common. Traditional Neer classification identifies a fragment as an independent entity when there is a split over 1 cm, or it is angulated over 45°, although Neer has noted that he specified these values somewhat arbitrarily (Carofino and Leopold 2013; Schumaier and Grawe 2018). Other widely accepted classifications for PHF are AO/OTA (Arbeitsgemeinshaft für Osteosynsthesefragen/Orthopaedic Trauma Foundation) and Hertel's HGLS (Head-Greater-Lesser-Shaft) classifications. HGLS shows higher interobserver and intraobserver agreement than the other mentioned classifications on plain radiographs and CT-Scans (Sukthankar et al. 2013; Iordens et al. 2016). All the classifications mentioned above specify that the humerus can be fractured in different patterns that mainly affect four different pieces: head, shaft, greater tuberosity and lesser tuberosity. Otherwise, each classification method is subtly but meaningfully distinctive. Regardless of the classification method, the delimitation of the morphological characteristics of each fracture line and fragment is essential for the correct selection of a patient's treatment. Nevertheless, there is a need to improve the interobserver and intraobserver agreement, and recently many new diagnostic methods have appeared in pursuit of improving understanding of PHF mechanics and classification.

In order to specify an accurate initial classification of the sampled PHF, a committee of three senior Orthopaedic Surgeons specialized in upper limb surgery was constituted, with an average experience of over 20 years in superior limb surgery. We provided all available images for each of the 34 selected patients (simple radiographs, CT-Scan slices, conventional 3D reconstructions, 3DPM and VR models) to each committee member for an assessment. Committee members worked independently to provide initial classification. Each expert provided a classification for each fracture by Neer and HGLS classifications. In this first round, the experts gave a concordant classification of 30 of the 34 patients. The remaining four were discussed by the committee until there was consensus on a unique classification for each. The classifications of the expert committee were considered the correct classifications for the remainder of the study, and, as such, a PHF classification in this study is considered accurate or correct when it is the same as the classification provided by the expert committee for the presented PHF model.

The appropriate treatment was determined through a review of the clinical histories, and the best treatment was defined as the course of treatment provided to each patient. We provided the three main treatment options available for any fracture on orthopaedic surgery, offering conservative treatment (also known as non-operative, typically performed with a sling), osteosynthesis (classically as ORIF performed with wires, plates and/or screws) and arthroplasty (including hemiarthroplasty, total shoulder arthroplasty and reversed shoulder arthroplasty).

#### 2.3 Graphical design: classifications and models

#### 2.3.1 Classifications and treatment forms

Classification forms for both the 3DPM and VR environment were designed by the lead author using a Huion H1060P Graphics Drawing Tablet and GIMP software (v2.10.30). The classification forms replicate the standard classification forms used for Neer and HGLS classifications and are implemented in a way that ensures that the forms in both conditions are as similar in appearance as possible. The classification forms were printed in colour on two A4 plasticised paper (one for the left humeri and another for the right ones) for the 3DPM condition (Fig. 4) and presented as UI images in the VR environment (Fig. 2).

## 2.3.2 Design of the 3D models

3D models were obtained from DICOM files. These files were processed by InVesalius 3.1 (CTI Renato Archer, Brazil), distinguishing and delimiting different fracture fragments and excluding irrelevant elements until a useful STL file was obtained (Moldovan et al. 2021; Ejnisman et al. 2021). Images were then processed on Meshmixer v3.5 (Autodesk, 2018). Tissue and bone defects unrelated to the



PHF and scan artefacts were excluded from the 3D model of the humerus to make it printable (Raeker-Jordan et al. 2021). A refined STL file was obtained for each model, which was used as the final model for both the 3DPM and the VR model. Model names were enumerated with a random index from 1 to 34.

# 2.4 Experiment in VR

#### 2.4.1 Hardware and software

The VR experiment was run on an Oculus Quest 2. This system consists of a standalone head-mounted display with a resolution of 1832 by 1920 pixels per eye, a 90 Hertz refresh rate, and two controllers. The virtual environment was created in Blender 2.93 and exported into Unity 2020.3.26 LTS. The VR framework used to handle the hardware interaction

was the XR interaction toolkit 1.0. This framework is publicly available in the Unity repository and provides tools to manage the hardware input and translate them into actions in the virtual environment. The cleaned STL files were imported into the Unity environment for presentation in VR as virtual 3D objects.

#### 2.4.2 Virtual environment

Navigation and interaction with the virtual environment were performed with the Oculus Quest 2 controllers (Fig. 1). Participants could grab the humeri with any controller by pressing the grab button. Using the joystick, they could navigate the 3D environment with the right controller. The left controller was used to interact with the PHF visualization and classification system (Fig. 2). Participants could switch between interface and bone interaction

Fig. 1 Controllers bindings

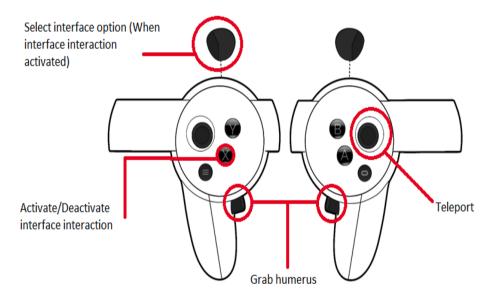
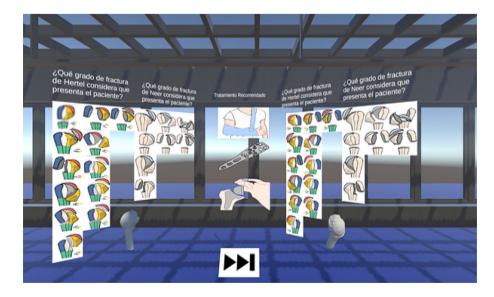


Fig. 2 Set up from the centre. Right Neer and HGLS classifications can be seen in the left, left HGLS and Neer classifications can be seen in the right (like exploring a patient or an image of the shoulder in front of the observer), recommended treatment and button "next" are in the centre. Note only one Humerus was presented at a time, but left and right are visualized here for demonstration





modes by pressing the 'X' button on the left controller. When the interface interaction mode was activated, a line (ray cast) extended from the controller to facilitate aiming. When pointing at interactable interface elements, the ray cast would be green when the user pointed towards them and red when pointing at something else. Participants could select an option by pressing the trigger button on the left controller. When pointing at individual elements in the interface, they were highlighted green. When a participant selected an interface element, the option was highlighted in red to let the participant know which option was selected.

When in bone interaction mode, participants could grab and move the humerus models by pressing the grab button while hovering over the model. To avoid unnecessary distractions, there was no gravity affecting the humeri, meaning that when participants grabbed the humeri, these would follow the speed of the hand that the participant used to grab the humeri, and when they released the humeri, it would stay in the position that it was released.

#### 2.4.3 VR experiment mechanics

When the experiment started, the participants were immersed in a virtual room. The main elements were the interface containing HGLS and Neer classifications criteria and the humeri models (see Fig. 2 for the virtual room layout). Since each humerus could be either left or right, there was a diagram of HGLS and Neer classifications on the right side corresponding to the left humerus, and a diagram of HGLS and Neer classifications on the left side corresponding to the right humerus. To make it easier for the participants, the right humeri were presented on the left side, and the left humeri were presented on the right side (like seeing a patient or an image in front of the observer). The models were presented near the participants, so they did not have to navigate or walk to see them well. In the middle of the interface, the participants could choose the recommended treatment from a list of options (conservative, osteosynthesis, and arthroplasty). The button to render the next humerus was placed at the bottom centre of the interface (see Fig. 2).

The humeri were shown to the participants one by one. The presentation order of the humeri was randomised and distributed to avoid order bias. For each humerus, the participants were asked to select one of three treatment options (conservative, osteosynthesis, or arthroplasty) and classify the fracture by both the Neer and HGLS systems. These selections could be made in any order. Once a treatment option and both classifications answers were selected, participants pressed a button to move on to the next humerus. The responses were saved in a.CSV file for further analysis after the experiment.

#### 2.5 Experiment in 3DPM

#### 2.5.1 3D models printing and storage

The cleaned STL files were also used for printing the 3DPM (Fig. 3). Each model was processed for 3D printing in Cura 4.5 (Ultimaker 2020) and sliced by the specific predefined parameters of this software for a Creality Ender 3 printer. Processing for 3D printing did not visibly alter the external structure or presentation of the PHF.

3D Printing was performed on a Creality Ender 3 v2, with an average time of around 7 h per model. Models were printed using 1:1 scale. The material used was 1.75 mm white PLA, with a total amount of 1.3 kg for the whole experiment. Printing supports were removed post-print, and no further post-printing treatment was needed.

Models were stacked on a foam mould and randomly identified with a number from 1 to 34. Each humerus was marked with its number and a letter 'I' (Izquierdo, Left in Spanish) or 'D' (Derecho, Right in Spanish).

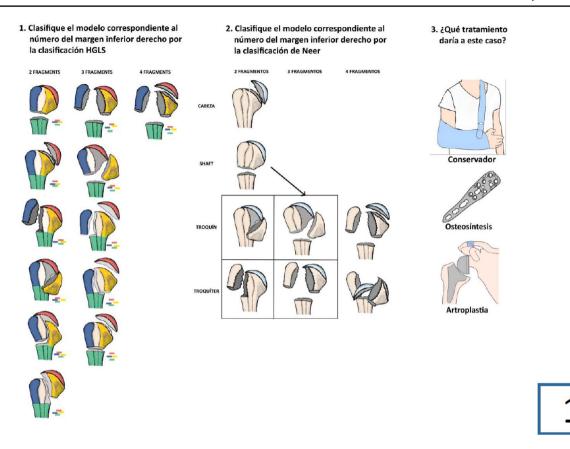
#### 2.5.2 3DPM experiment environment and materials

Each participant received two reference cards, one for left humeri and another for the right ones, printed in High Quality with Neer and HGLS classifications and treatment options (Fig. 4). Then, a pack of 34 cards numbered from 1 to 34 was randomly ordered and given to each participant in order to determine model classification order. The residents were asked to classify each 3DPM according to both the Neer and HGLS classifications and to indicate a proper



**Fig. 3** Proximal humerus fracture Three-Dimensional Printed Model. Each model was printed on white PLA of 1.75 mm on 1:1 scale. Thirty-four of these models were evaluated by each participant





**Fig. 4** Original card for right-side fractures shown on 3DPM questionnaire. Left-side fractures had analogue cards with a representation of left humerus fractures. Number '1' on the right bottom corner alludes to the corresponding model. Cards were shown on a ran-

domised fashion. Translation: '1. Classify the corresponding model by HGLS classification' '2. Classify the corresponding model by Neer classification' '3. Provide a proper treatment on this patient'

treatment approach from a list of 3 options (conservative, osteosynthesis, or arthroplasty).

#### 2.6 Study design and data collection

The study structure was based on a randomised controlled trial design. Participants were run individually and randomly assigned to a condition, VR or 3DPM. Each participant received a brief seminar about PHF and the different classification approaches as well as a description of the experiment. All participants were individually randomised and had a 50% of probability of belonging to any of the two groups. After the experiment was introduced, the participants signed an informed consent form and classified the humeri using the method (3DPM or VR) corresponding to their assigned condition.

Age, sex and previous experience with the technologies (3DPM or VR) used were registered, as well as the year of residency, the WHO-CHOICE hospital each participant belongs to, and time spent (in minutes) completing the 34 humeri questionnaire. Experience with VR/3DPM was specified in 3 levels: no previous experience, low experience

(less than 10 previous uses of VR or 3DPM for any purpose) or high experience (10 or more previous uses of VR or 3DPM for any purpose). After completing the classifications, participants were asked to complete a utility and confidence VAS (Visual Analogue Scale) and a questionnaire to assess the degree of satisfaction with the experiment's performance and the technology used.

#### 2.7 Statistical analysis

Characteristics of participants are presented according to trial group assignment. We performed descriptive analyses on outcomes and covariates. Numerical variables are described by means and standard deviations, and categorical variables are described by absolute and relative frequencies.

Group comparisons were carried out using Mann–Whitney's test, Student's t-test, or Fisher's exact test as appropriate. We assessed associations between the binary outcome variable (correct classification) and group assignment (VR/3DPM) using mixed-effects logistic regression models adjusted by age, sex and potential confounding variables to ensure that additional variables do not meaningfully impact



statistical results, obtaining results as Odds Ratios (ORs). Potential confounders were included in the model when they were associated with the outcome variable at a level of statistical significance p < 0.25. A random intercept was included to account for correlation among measurements in the same subject (Hosmer et al. 2013). All statistical tests were two-sided, and p values < 0.05 were considered statistically significant. All statistical analyses were conducted using Stata 17.0 (StataCorp LLC).

#### 3 Results

# 3.1 Homogeneity of the groups

3DPM and VR groups were compared by age, gender, time of residency (years), hospital level, residency level, previous experience with the given technologies, and time spent completing the evaluations (descriptive statistics can be found in Table 1). Within the 3DPM group (n=18), 15 identified as male, and 3 as female, with an average age of 27.9 years with a Standard Deviation (SD) of 2.4 years. Within the VR group (n=18), 10 identified as male and 8 as female, with an average age of 27.6 years (SD=1.9 years). The average amount of time in residency was 2.8 years (SD=1.4 years) for the 3DPM condition and 3.1 years (SD=1.5 years) for the VR condition, with a detailed breakdown of Post-Grade Year (PGY) and hospital type by group presented in Table 1.

Table 1 Characteristics of the participants according to the trial group assignment

Characteristic n	Virtual Reality participants	3D Printed Models partici- pants 18	p value
Age (years)	27.6 (1.9)	27.9 (2.4)	0.858 <sup>a</sup>
Time of residency (years)	3.1 (1.5)	2.8 (1.4)	$0.572^{a}$
Time spent (minutes)	31.5 (11.6)	32.5 (13.4)	$0.801^{b}$
Sex (% men)	55.6	83.3	$0.146^{c}$
Hospital Level (%)			
Primary-level	16.7	11.1	$0.817^{c}$
Secondary-level	44.4	38.9	
Tertiary-level	38.9	50.0	
Previous experience (%)			
None	66.6	50.0	0.271 <sup>c</sup>
Low (< 10 times)	27.8	22.2	
High (> 10 times)	5.6	27.8	

Data given as mean (standard deviation) or %

No significant differences were found on any of the analysed variables (Table 1). So, we can conclude that participants of both groups are homogenous for these variables. Notably, both groups spent a similar amount of time analysing all 34 humeri, with an average total time of 32.5 (SD=13.4 min) for the 3DPM group and 31.5 min (SD=11.6 min) for the VR group. This difference was also not statistically significant.

# 3.2 Association between different characteristics and correct classification

To study the impact of 3DPM and VR on fracture classification accuracy, we investigated the correct classification rate (or accuracy) relative to the expert classifications according to two classification methods (Neer and HGLS). Along with a comparison of accuracy between 3DPM and VR, we also considered how participant characteristics might impact classification accuracy to get a more complete picture of any differences or similarities observed between conditions (see Table 2 for summary).

Table 2 Association between participant's characteristics and correct classification

Characteristic	Correct HGLS classification		Correct Neer classification	
	N (%)	$p^{\mathrm{a}}$	N (%)	$p^{\mathrm{a}}$
Group				'
Virtual Reality	317 (51.8)	0.169	397 (64.9)	0.365
3d Printed Models	342 (55.9)		413 (67.5)	
Sex				
Men	465 (54.7)	0.384	569 (66.9)	0.395
Women	194 (51.9)		241 (64.4)	
Hospital Level				
Primary-level	96 (56.5)	0.193	108 (63.5)	0.601
Secondary-level	259 (50.8)		335 (65.7)	
Third-level	304 (55.9)		367 (67.5)	
Residency year				
First	115 (48.3)	< 0.001	146 (61.34)	0.005
Second	119 (43.8)		161 (59.2)	
Third	144 (60.5)		168 (70.6)	
Fourth	137 (57.6)		163 (68.5)	
Fifth	144 (60.5)		172 (72.3)	
Previous experience				
None	356 (49.9)	0.001	454 (63.6)	0.026
Low (< 10 times)	171 (55.9)		206 (67.3)	
High (≥ 10 times)	132 (64.7)		150 (73.5)	

Statistically significant results are shown in bold (p < 0.05)

Data given as a mean (standard deviation) or %. HGLS: Head-Greater-Lesser-Shaft



a Mann-Whitney's test

b Student's t-test

c Fisher's exact test

aFisher's exact test

Regarding technologies, 3DPM showed higher fracture classification accuracy than VR for both the HGLS classification (51.8% vs. 55.9%) and Neer classification (64.9% vs. 67.5%). However, these differences were not statistically significant (p = 0.169 for HGLS and p = 0.365 for Neer).

Unsurprisingly, participants with more professional experience performed classifications more accurately than those with less. Professional experience, as measured by PGY, had a significant impact on the classification accuracy for both Neer (p < 0.001) and HGLS (p = 0.005) classifications, with higher PGY associated with improved accuracy for each classification method.

Previous experience with the technology (3DPM or VR, depending on the condition) also showed a statistically significant association with the classification accuracy, with more experience significantly improving accuracy for both classification methods. For HGLS, the difference on accuracy between no previous experience (49.9%), low experience (55.9%), and high experience (64.7%) had a span of 14.8% from no to high experience (p<0.001). The difference between the same experience characteristics for Neer was 9.9% (no=63.6%, low=67.3%, high=73.5%, p=0.026). These results indicate that familiarity with the technology used to display the fracture plays a role in one's ability to accurately classify the fracture.

Across all participants and conditions, a total of 1224 assessments were provided for both Neer and HGLS classifications. Table 3 shows age and time spent comparing right and wrong answers on Neer and HGLS using both technologies. We did not find statistically significant differences.

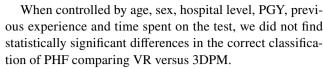
# 3.3 Clinical significance and comparison of VR and 3DPM

Multivariate adjusted Odds Ratios (ORs) for PHF classification accuracy adjusted by gender, age, hospital level, PGY, previous experience and time spent on the test were estimated. Multivariate adjusted associations between the assigned group and the classification accuracy are shown in Table 4. These results are largely in line with the results presented in Sect. 3.2.

**Table 3** Association between numerical covariates and correct classification

	HGLS			Neer		
	$\overline{\text{Correct } N = 659}$	Incorrect $N = 565$	p <sup>a</sup>	$\overline{\text{Correct } N = 810}$	Incorrect N=414	p <sup>a</sup>
Age in years	27.8 (1.9)	27.7 (2.3)	0.566	27.8 (2.0)	27.7 (2.3)	0.473
Time for whole test in min- utes	32.5 (12.2)	31.4 (12.2)	0.117	32.2 (12.3)	31.7 (12.0)	0.480

Data given as a mean (standard deviation). HGLS: Head-Greater-Lesser-Shaft



Time of residency showed a direct relationship with fracture classification accuracy for both HGLS and Neer classification (OR = 1.24; 95% Confidence Interval (95% CI) 1.04-1.48, and OR = 1.21; 95% CI 1.05-1.40, respectively). This implies an improvement of approximately 24% for HGLS and 21% for Neer classification per year of residency.

Time spent on the whole test also showed a statistically significant direct relationship with classification accuracy for both HGLS and Neer (OR = 1.02; 95%CI 1.01-1.04, and OR = 1.02; 95%CI: 1.00-1.03, respectively). Thus, for each extra minute spent on analysis of the 34 humeri there was an observed accuracy improvement of approximately 2%.

#### 4 Discussion

New technologies that allow for rapid 3D interactive visualizations and classification of humeral factures are extremely valuable for assessing and determining treatment plans. Fürnstahl et al. proposed a computer-assisted method for identification of humeral fragments based on a CT-Scan and comparative analysis of the 3D models of a fractured humerus and the contralateral humerus (Fürnstahl et al. 2012). While the value of these technologies is not in doubt, there is still a need for identifying which technologies are best suited for a given situation and user group. Currently, 3DPM and VR are prominent and rapidly developing technologies for interacting with 3D models whose potentials are not yet fully explored in the context of fracture classification.

It is clear from previous literature that using 3DPM has become a Gold Standard method for the diagnosis of PHF (Bougher et al. 2021; Iordens et al. 2016). The use of 3DPM in preoperative planning has improved shoulder surgery processes and outcomes, reducing operation time, radiation exposure, pain, and enhance patient outcomes while improving anatomic healing and increasing post-operative mobility (Li et al. 2022; You et al. 2016). The use of 3DPM has also been tested and validated for fracture classification



a Student's t-test

**Table 4** Multivariate adjusted odds ratios for correct classification of proximal humeral fractures

Variable	HGLS classification		Neer classification	
	OR (95% CI)	$p^{\mathrm{a}}$	OR (95% CI)	$p^{\mathrm{a}}$
Group				
Virtual Reality	Reference	0.513	Reference	0.630
3D Printed Models	1.14 (0.77–1.67)		1.08 (0.79–1.48)	
Age (years)	0.99 (0.89-1.11)	0.852	0.98 (0.90-1.07)	0.676
Sex				
Women	Reference		Reference	
Men	0.99 (0.63-1.55)	0.384	1.09 (0.76–1.56)	0.654
Hospital Level				
Primary-level	Reference		Reference	
Secondary-level	0.96 (0.55-1.68)	0.885	1.29 (0.82-2.02)	0.273
Tertiary-level	0.89 (0.50-1.56)	0.676	1.14 (0.72–1.80)	0.581
Time of residency (years)	1.24 (1.04–1.48)	0.016	1.21 (1.05–1.40)	0.008
Previous experience				
None	Reference		Reference	
Low (<10 times)	1.47 (0.84–2.56)	0.173	1.28 (0.81-2.00)	0.290
High (≥ 10 times)	1.70 (0.85-3.38)	0.134	1.47 (0.80-2.51)	0.233
Time spent (minutes)	1.02 (1.01-1.04)	0.008	1.02 (1.00-1.03)	0.019

Statistically significant results are shown in bold (p < 0.05). HGLS: Head–Greater–Lesser–Shaft

and diagnosis in other regions of the body (Yammine et al. 2021). In that current context, we have reported the results from a non-inferiority randomised controlled trial to determine if there could be any significant differences between using 3DPM and VR for PHF classification and diagnosis. We observed that there are no statistically significant differences among classifications using these technologies, indicating that diagnostic outcomes using either technology are likely equivalent. Observed fracture classification accuracy was 51.8% (VR) vs. 55.9% (3DPM) for HGLS and 64.9% (VR) vs. 67.5% (3DPM) for Neer classification, which is consentient with interobserver agreement results reported in the literature (Iordens et al. 2016). In the present work, we also directly compared PHF classification accuracy using VR and 3DPM. We obtained an adjusted OR for VR versus 3DPM of 1.14 (0.77–1.67) on HGLS classification and 1.08 (0.79–1.48) on Neer classification, both favouring 3DPM though without statistical significance. Thus, according to our study VR is non-inferior to 3DPM for PHF classification and we can consider both technologies as equivalent.

To our knowledge, this is the first study evaluating the use of VR on PHF classification. In at least one case, a comparison of Augmented Reality (AR) combined with other methods for diagnosing PHF, including 3DPM, showed the highest interobserver agreement using 3DPM versus other technologies (Cocco et al. 2020). There are also some works that argue for the use of both AR and VR separately in orthopaedic trauma surgery, highlighting specifically the combination of AR with 3DPM, though in this case there is no

direct empirical comparison of the technologies (Negrillo-Cárdenas et al. 2020). Some other works evaluate the utility of VR technologies for pre-surgical tasks, demonstrating that VR can be an appropriate technology (Negrillo et al. 2020; Verhey et al. 2020; Zhang et al. 2021). For example, Zhang and colleagues (2021) investigated the utility of VR for the evaluation and understanding of craniofacial structures.

#### 4.1 Comparison between 3DPM and VR

While the non-inferiority design can indicate equivalent diagnostic outcomes between 3DPM and VR, there are obviously additional differences between these technologies that may be considered. The cost and time of making 3DPM has diminished considerably in recent years along with a reduction in the expertise required to prepare and produce models from 3D scan data. However, production of a 3DPM still requires a 3D printer, often a fixed location for the printer, and a regular supply of materials. This can often mean that valuable office or equipment space must be dedicated to the process. Perhaps the most notable challenge for 3DPM is the time it takes to produce a 3DPM of sufficient quality. For this study, we spent an average of about seven hours printing each model and typically one model is printed on a 3D printer at a time. Adding in the possibility for print failures, a 3DPM could take up to a full day to produce, though failure rates are constantly improving. Finally, 3DPM may be made of one-time use plastics which can have negative climate impacts, though bioplastics can reduce this impact.



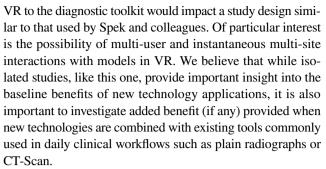
<sup>&</sup>lt;sup>a</sup>Models are from mixed effect logistic regression analyses with a random intercept

Because VR does not require a physical production step, once the digital model is converted and cleaned for presentation it can be directly imported and viewed in VR. This results in a notable reduction in time for model availability in VR compared to 3DPM, which requires a similar conversion and cleaning process. With VR, time-to-presentation can be reduced to minutes instead of hours and no additional space is required as the same computer used for preparing the model can export it to the VR headset. Modern VR headsets can be deployed without much setup and often require only the headset and hand controllers to function. The VR scenario used for the current test required minimal movement and could be designed for seated or standing workspaces. An additional advantage of VR is that it can be used for instantaneous multi-site and multi-user evaluations, making remote consultations more robust and reducing ecological impacts of traveling or sending the 3DPM to another hospital. Also, the decrease in plastic use with VR may mean that VR has a lower environmental impact over time. It is precisely because of these advantages that we devised this study as one of non-inferiority. In our study, we also observed a shorter time for PHF classifications using VR when compared to 3DPM, although this difference was not statistically significant.

While there is a lot of promise around VR, there are a few possible downsides. Some users may find VR head-sets uncomfortable and while simulator sickness is much less common, some users may still experience dizziness or nausea in VR. Moreover, not all users with glasses will be able to wear their glasses while using the headset, possibly resulting in reduced visual acuity in VR. Finally, there are currently no off-the-shelf solutions for the workflow presented in this study, making implementation with current systems more technically demanding than more mature 3D printing technologies. However, this is something we expect to change rapidly in coming years.

# 4.2 Future questions and work

Potentially countering the demonstrated benefits of 3DPM, Spek and colleagues (2022) reported that the use of 3DPM did not improve agreement among observers while identifying fracture characteristics and patterns (Spek et al. 2022). While observer agreement is important, this paper should be analysed carefully, both because of the small sample size and because the possibility of anchoring bias due to the combination of onscreen images and 3DPM was not considered. Thus, it is not clear if 3DPM did not improve agreement because it did not add anything to the classification process or because participants primarily relied on more familiar non-3DPM data when it was available. However, building on the current research demonstrating that VR produced outcomes equivalent to 3DPM, it would be interesting to see how the introduction of



VR is widely accepted as a useful technology for learning during medical residency and is often more effective than benchtop trainers or synthetic models (Verhey et al. 2020). However, there is a lack of evidence supporting the usefulness of VR technology for fracture classifications. This is due, in part, to a focus in the Orthopaedics field on using VR for rehabilitation and training technical surgical skills (Hasan et al. 2021; Berton et al. 2020; Matamala-Gómez et al. 2022). When used for surgical pre-planning, but not diagnosis, for shoulder fractures, VR demonstrated better outcomes than conventional training methods and faster time-to-surgery than 3DPM used for this purpose, making preoperative planning better and shorter (Chen et al. 2018). The current research suggests that including VR methods earlier in the planning process, e.g. diagnostic classification, may further streamline pre-operative planning processes, allowing for consistent tool and visualization throughout the process. Future research into using VR and 3DPM throughout the entire diagnostic and treatment process could be valuable in this regard.

One potentially useful insight from the current work is that there seemed to be a significant, and unsurprising, correlation between time in residency and classification outcomes. Future works comparing these kinds of technological interventions for treatment determinations should consider and report training time and/or years of experience. More importantly, care should be taken when comparing results between studies when participant training level or experience may be different. Additionally, while the sample size is fairly large for this type of study, especially given the difficulty of accessing a concentrated population of orthopaedic surgical specialists and residents, the width of the confidence interval on both Neer and HGLS classifications suggests that future work should consider an even bigger sample. Though for the non-inferiority study the confidence intervals may not be important, it still worth future work considering the benefit of identifying an even larger participant population.

# **5 Conclusions**

Our findings support that VR can be a valuable alternative to 3DPM for the classification of PHF, providing both noninferior classification power and advantages with regard to



time and resource costs. This work demonstrates a significant direct relationship between PHF classification accuracy, PGY, and time spent analysing PHF. Importantly, obtained data demonstrate that while there is no significant difference in PHF classification outcomes between VR and 3DPM, the time needed to analyse a fracture using VR is marginally, but not significantly, shorter than using 3DPM. When these results are considered in the context of the difference in time to prepare a 3DPM for viewing, these strongly favour VR, suggesting that VR should be included as an additional weapon in the presurgical armoury of orthopaedic surgeons for PHF.

The main strengths of the current work are the high number of expert participants, the use of validated models from actual patient CT-Scans, and the randomised study design. This work represents the first paper in the literature that evaluates the feasibility of classifying PHF in virtual reality. Moreover, the work is a prospective, multicenter and multidisciplinary work incorporating engineers, statisticians, and orthopaedic surgeons in the development and implementation of the study. Further investigation is needed into the effects of mixing VR with more conventional radiological techniques and the utility of VR for preoperative planning in other surgical contexts. Moreover, user-friendly and accessible software tools for VR-based preoperative planning are needed for the broader adoption of the methods presented in this paper. Given the results of this paper, we believe efforts in this direction will be valuable for both doctors and patients, improving presurgical planning and reducing treatment time.

Acknowledgements This work would not have been possible without the altruistic and priceless help of all the orthopaedic surgery residents that took part on this study fulfilling the questionnaires (Hospital Virgen de las Nieves, Granada Spain; Hospital Virgen de la Victoria, Malaga, Spain; Hospital de Antequera, Malaga, Spain). Either, all our gratefulness to our Experts Committee (R.L. Egozcue-Folgueras, A. López-García and A. López-Camacho).

Author contributions RA made substantial contribution to the conceptualization of the work, artwork, collection of data, 3D printing process and fracture STL models design. FG made substantial contributions to conceptualization of the work and development of the Unity environment. ML made substantial contributions to development of the Unity environment, text editing, and translation correction. RV made substantial contribution on collection of data. MG made substantial contribution on statistical analysis and exposure of the results. All authors provided their consent on the final version to be published and took part on the resolution of any remaining issue to be cleared up.

**Funding** Funding for open access publishing: Universidad Málaga/CBUA.

## **Declarations**

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

**Ethical approval** This study has been approved by the Ethics Committee of Investigation of Malaga (reference #0171-N-22).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

#### References

- Berton A, Longo UG, Candela V et al (2020) Virtual Reality, Augmented Reality, gamification, and telerehabilitation: psychological impact on orthopedic patients' rehabilitation. J Clin Med. https://doi.org/10.3390/jcm9082567
- Bougher H, Buttner P, Smith J et al (2021) Interobserver and intraobserver agreement of three-dimensionally printed models for the classification of proximal humeral fractures. JSES Int 5:198–204. https://doi.org/10.1016/j.jseint.2020.10.019
- Carofino BC, Leopold SS (2013) Classifications in brief: the Neer classification for proximal humerus fractures. Clin Orthop Relat Res 471:39–43. https://doi.org/10.1007/s11999-012-2454-9
- Chen Y, Jia X, Qiang M et al (2018) 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 100:1960–1968. https://doi.org/10.2106/JBJS.18.00477
- Cocco LF, Aihara AY, Franciozi C et al (2020) Three-dimensional models increase the interobserver agreement for the treatment of proximal humerus fractures. Patient Saf Surg 14:33. https://doi.org/10.1186/s13037-020-00258-2
- Corbacho B, Duarte A, Keding A et al (2016) Cost effectiveness of surgical versus non-surgical treatment of adults with displaced fractures of the proximal humerus. Bone Joint J 98:152–159. https://doi.org/10.1302/0301-620X.98B2.36614
- Ejnisman L, Gobbato B, de França Camargo AF, Zancul E (2021) Three-dimensional printing in orthopedics: from the basics to surgical applications. Curr Rev Musculoskelet Med 14:1–8. https://doi.org/10.1007/s12178-020-09691-3
- Fürnstahl P, Székely G, Gerber C et al (2012) Computer assisted reconstruction of complex proximal humerus fractures for preoperative planning. Med Image Anal 16:704–720. https://doi.org/10.1016/j.media.2010.07.012
- Hasan LK, Haratian A, Kim M et al (2021) Virtual reality in orthopedic surgery training. Adv Med Educ Pract 12:1295–1301. https://doi.org/10.2147/AMEP.S321885
- Hosmer DW, Lemeshow S, Sturdivant RX (2013) Applied logistic regression, 3rd edn. John Wiley & Sons, New York. https://doi.org/10.1002/9781118548387
- Iordens GIT, Mahabier KC, Buisman FE et al (2016) The reliability and reproducibility of the Hertel classification for comminuted proximal humeral fractures compared with the Neer classification. J Orthop Sci 21:596–602. https://doi.org/10.1016/j.jos. 2016.05.011
- Levy J, Rosas S, Law TY et al (2017) Ninety-day cost of treatment for proximal humerus fractures: an acute and 90-day cost



- evaluation. Orthop Proc 99:92. https://doi.org/10.1302/1358-992X.99BSUPP\_4.ISTA2016-092
- Li K, Liu Z, Li X, Wang J (2022) 3D printing-assisted surgery for proximal humerus fractures: a systematic review and metaanalysis. Eur J Trauma Emerg Surg. https://doi.org/10.1007/ s00068-021-01851-5
- Lohre R, Warner JJP, Athwal GS, Goel DP (2020) The evolution of virtual reality in shoulder and elbow surgery. JSES Int 4:215–223. https://doi.org/10.1016/j.jseint.2020.02.005
- Mahabier KC, Den HD, Van Veldhuizen J et al (2015) Trends in incidence rate, health care consumption, and costs for patients admitted with a humeral fracture in The Netherlands between 1986 and 2012. Injury 46:1930–1937. https://doi.org/10.1016/j.injury.2015.07.025
- Matamala-Gomez M, Slater M, Sanchez-Vives MV (2022) Impact of virtual embodiment and exercises on functional ability and range of motion in orthopedic rehabilitation. Sci Rep 12:5046. https://doi.org/10.1038/s41598-022-08917-3
- Moldovan F, Gligor A, Bataga T (2021) Structured integration and alignment algorithm: a tool for personalized surgical treatment of Tibial plateau fractures. J Pers Med 11:190
- Mulligan J-A, Fox-Rushby J (2003) Unit costs of health care inputs in low and middle income regions. Dis Control Priorities Proj Work Pap 9, Bethesda
- Negrillo-Cárdenas J, Jiménez-Pérez J-R, Feito FR (2020) The role of virtual and augmented reality in orthopedic trauma surgery: from diagnosis to rehabilitation. Comput Methods Progr Biomed 191:105407. https://doi.org/10.1016/j.cmpb.2020.105407
- Neumann PJ, Sanders GD (2017) Cost-effectiveness analysis 2.0. N Engl J Med 376:203–205. https://doi.org/10.1056/NEJMp16126
- Raeker-Jordan EA, Martinez M, Aziz KT et al (2021) High-fidelity wrist fracture phantom as a training tool to develop competency in orthopaedic surgical trainees. JAAOS Glob Res Rev 5:e20
- Roux A, Decroocq L, El Batti S et al (2012) Epidemiology of proximal humerus fractures managed in a trauma center. Orthop Traumatol Surg Res 98:715–719. https://doi.org/10.1016/j.otsr.2012.05.013
- Schumaier A, Grawe B (2018) Proximal humerus fractures: evaluation and management in the elderly patient. Geriatr Orthop Surg Rehabil 9:2151458517750516. https://doi.org/10.1177/2151458517750516

- Spek RWA, Schoolmeesters BJA, Oosterhoff JHF et al (2022) 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 480:150–159. https://doi.org/10.1097/CORR.0000000000001921
- Sukthankar AV, Leonello DT, Hertel RW et al (2013) A comprehensive classification of proximal humeral fractures: HGLS system. J Shoulder Elb Surg 22:e1-6. https://doi.org/10.1016/j.jse.2012. 09.018
- Sumrein BO, Huttunen TT, Launonen AP et al (2017) Proximal humeral fractures in Sweden-a registry-based study. Osteoporos Int 28:901–907. https://doi.org/10.1007/s00198-016-3808-z
- Thorsness R, Shields E, Iannuzzi JC et al (2016) Cost drivers after surgical management of proximal humerus fractures in medicare patients. J Orthop Trauma 30:262–268. https://doi.org/10.1097/BOT.0000000000000513
- Verhey JT, Haglin JM, Verhey EM, Hartigan DE (2020) Virtual, augmented, and mixed reality applications in orthopedic surgery. Int J Med Robot 16:e2067. https://doi.org/10.1002/rcs.2067
- Yammine K, Karbala J, Maalouf A et al (2021) Clinical outcomes of the use of 3D printing models in fracture management: a metaanalysis of randomized studies. Eur J Trauma Emerg Surg. https:// doi.org/10.1007/s00068-021-01758-1
- You W, Liu LJ, Chen HX et al (2016) Application of 3D printing technology on the treatment of complex proximal humeral fractures (Neer3-part and 4-part) in old people. Orthop Traumatol Surg Res 102:897–903. https://doi.org/10.1016/j.otsr.2016.06.009
- Zhang H, He Y, Chen Y et al (2021) Virtual Reality and three-dimensional printed models improve the morphological understanding in learning mandibular sagittal split ramus osteotomy: a randomized controlled study. Front Surg 8:705532. https://doi.org/10.3389/fsurg.2021.705532

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

