BJR

TRAUMA

Defining the fracture population in a pragmatic multicentre randomised controlled trial

PROFHER AND THE NEER CLASSIFICATION OF PROXIMAL HUMERAL FRACTURES

H. H. G. Handoll, S. D. Brealey, L. Jefferson, A. Keding, A. J. Brooksbank, A. J. Johnstone, J. J. Candal-Couto, A. Rangan

University of York, York, United Kingdom

- H. H. G. Handoll, DPhil, Reader in Orthopaedics, Health and Social Care Institute, School of Health and Social Care, Teesside University, Middlesbrough, Tees Valley, TS1 3BA, UK.
- S. D. Brealey, BSc, PhD, Senior Trial Manager/Research Fellow in Health Sciences,
- L. Jefferson, BSc, MSc, PhD, Trial Co-ordinator/Research Fellow,
- A. Keding, BSc, MSc, Statistician, York Trials Unit, Department of Health Sciences, University of York, Heslington, York, YO10 5DD, UK.
- A. J. Brooksbank, FRCS Ed (Tr&Orth), Consultant Trauma and Orthopaedic Surgeon, Department of Orthopaedics and Trauma, Glasgow Royal Infirmary, 84 Castle Street, Glasgow, G4 0SF, UK.
- A. J. Johnstone, FRCS Ed (Orth), Consultant Orthopaedic Surgeon, Orthopaedic Trauma Unit, Aberdeen Royal Infirmary, Foresterhill, Aberdeen, AB25 2ZN, UK.
- J. J. Candal-Couto, MBChB, BMSC, FRCS(Tr&Orth), Consultant Orthopaedic Surgeon, Northumbria Healthcare NHS Trust, Wansbeck General Hospital, Woodhorn Lane, Ashington, Northumberland, NE63 9JJ, UK.
- A. Rangan, ChM, FRCS (Tr&Orth), Professor of Orthopaedic Surgery, Department of Trauma and Orthopaedics, James Cook University Hospital, South Tees Hospitals NHS Trust, Marton Road, Middlesbrough, TS4 3BW, UK.

Correspondence should be sent to A. Rangan;
email: amar.rangan@york.ac.uk

doi: 10.1302/2046-3758.510.BJR-2016-0132.R1

Bone Joint Res 2016;5:481–489.
Received: 11 May 2016;
Accepted: 24 August 2016

Objectives

Accurate characterisation of fractures is essential in fracture management trials. However, this is often hampered by poor inter-observer agreement. This article describes the practicalities of defining the fracture population, based on the Neer classification, within a pragmatic multicentre randomised controlled trial in which surgical treatment was compared with non-surgical treatment in adults with displaced fractures of the proximal humerus involving the surgical neck.

Methods

The trial manual illustrated the Neer classification of proximal humeral fractures. However, in addition to surgical neck displacement, surgeons assessing patient eligibility reported on whether either or both of the tuberosities were involved. Anonymised electronic versions of baseline radiographs were sought for all 250 trial participants. A protocol, data collection tool and training presentation were developed and tested in a pilot study. These were then used in a formal assessment and classification of the trial fractures by two independent senior orthopaedic shoulder trauma surgeons.

Results

Two or more baseline radiographic views were obtained for each participant. The independent raters confirmed that all fractures would have been considered for surgery in contemporaneous practice. A full description of the fracture population based on the Neer classification was obtained. The agreement between the categorisation at baseline (tuberosity involvement) and Neer classification as assessed by the two raters was only fair (kappa 0.29). However, this disparity did not appear to affect trial findings, specifically in terms of influencing the effect of treatment on the primary outcome of the trial.

Conclusions

A key reporting requirement, namely the description of the fracture population, was achieved within the context of a pragmatic multicentre randomised clinical trial. This article provides important guidance for researchers designing similar trials on fracture management.

Cite this article: *Bone Joint Res* **2016;5:481–489**

Keywords: Fracture classification; Proximal humeral fractures; Pragmatic randomised controlled trials

Article focus

- This article considers the practicalities of defining the fracture population, based on the Neer classification, within a pragmatic multicentre randomised controlled trial that compared surgical with non-surgical treatment of adults with displaced fractures of the proximal humerus involving the surgical neck.

Key messages

- Reporting involvement of either or both of the tuberosities as a proxy for Neer three- or four-part fractures was a suitable approach for surgeons assessing trial eligibility as it reflected assessment of fractures in clinical practice.
- Obtaining transferrable electronic versions of anonymised baseline radiographs of

participants in a multicentre trial is achievable, but can be a challenge. Exact linear measurement, as required for classification systems such as Neer's, requires the use of a calibration object at the fracture site.

- A piloted process involving training and a detailed proforma resulted in the best possible description of the fracture population according to the Neer classification.

Strengths and limitations

- Strengths: A thorough, systematic and informed approach was taken to accommodate what happens in clinical practice and the known limitations of the chosen fracture classification system.
- Limitations: The main limitations are those inherent in the Neer classification, in fully characterising the fracture population in a reproducible way.

Introduction

Characterisation of fractures is essential in order to define the study population and facilitate the scope, performance and interpretation of fracture management trials. This is usually hampered by imperfect fracture classification systems, which are frequently further undermined by poor inter-observer agreement. The latter is even more problematic for multicentre trials. This article focuses on the practicalities of defining the fracture population within a pragmatic multicentre randomised controlled trial comparing surgical with non-surgical treatment of adults with displaced fractures of the proximal humerus involving the surgical neck.

The PROximal Fracture of the Humerus Evaluation by Randomisation (PROFHER) trial recruited 250 adults from the orthopaedic departments (fracture clinics or wards) of 32 acute care NHS hospitals in the United Kingdom between September 2008 and April 2011. The results of the trial do not support the trend of increased surgery for patients with these fractures.[1]

Formal fracture classification in PROFHER was based on the commonly used Neer classification.[2,3] Central to this classification are the relative positions of the four main segments of the proximal humerus: the humeral head, the greater tuberosity, the lesser tuberosity and the humeral shaft. Although these may be delineated by fracture lines, a segment is only considered a 'part' if there is displacement of > 1 cm or 45° angulation. A 'minimally displaced' fracture, often referred to as a one-part fracture, occurs when the displacement criteria are not met for any of the four segments. Two-part, three-part and four-part fractures involve the relative displacement of two, three or all four segments, respectively. Other Neer categories involve fractures associated with an anterior or posterior humeral head dislocation and fractures involving the articular surface of the humeral head. Figure 1[4] shows the Neer classification,[2] with numbering of the 16 categories by Sidor et al.[5]

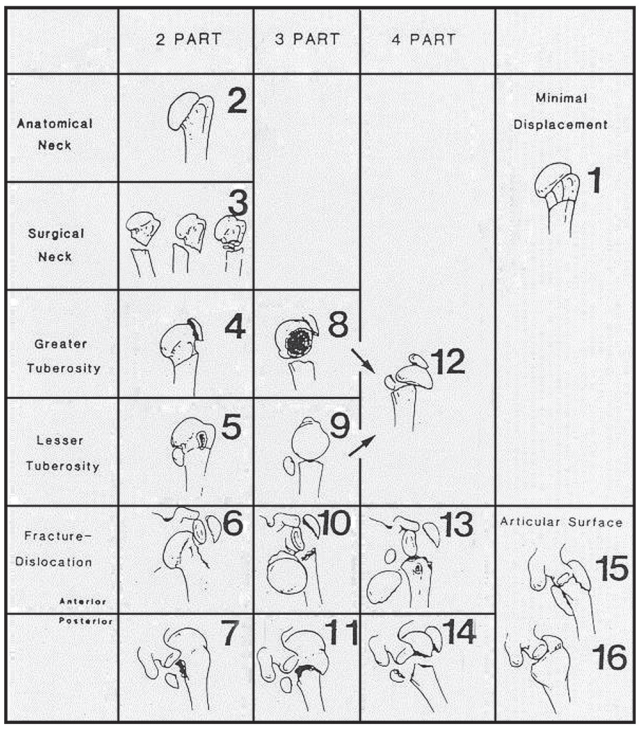

Fig. 1

The Neer classification for proximal humeral fractures. Modified figure reproduced, with permission from publisher Wolters Kluwer, from **Neer CS II.** Displaced proximal humeral fractures. I. Classification and evaluation. *J Bone Joint Surg [Am]* 1970;52-A6:1077-1089,[2] and from **Brorson S, Eckardt H, Audigé L, Rolauffs B, Bahrs C.** Translation between the Neer- and the AO/OTA classification for proximal humeral fractures: do we need to be bilingual to interpret the scientific literature? *BMC Res Notes* 2013;6:69.[4]

The Neer classification provides "a useful framework for clinical assessment of and research for proximal humeral fractures".[3] However, its limitations include the arbitrary definition of displacement, difficulties in assessing the extent of displacement of fracture parts from plain radiographs, and poor inter-observer agreement.[3,6] Poor agreement was also found for CT images.[7] However, a randomised controlled trial[8] found that agreement between surgeons increased with training comprising a 45-minute session on Neer classification.

Awareness of the limitations of the Neer classification informed the PROFHER trial design and the procedures for obtaining a definitive description of the fracture population in patients for whom there was uncertainty over whether surgery was required. This article reports the practical measures taken to ensure the inclusion of the intended fracture population within the context of undertaking a pragmatic trial that aimed to reflect good standard clinical practice and, in consequence, maximise the relevance and applicability of the trial findings. We report on the processes undertaken to optimise achievement of a valid formal description of the fracture population via an independent and blinded assessment of baseline radiographs of all randomised patients in PROFHER. After reporting the results of these endeavours, we discuss

these in terms of the study design and applicability of the PROFHER trial results.

The overarching aim of this article is to describe, examine and discuss the methods relating to the characterisation of fractures used in the PROFHER trial in order to inform researchers designing future pragmatic multicentre clinical trials on fracture management.

Patients and Methods

This extended section serves both to describe our methods and to highlight some of the practical issues we encountered when performing the trial. Figure 2 presents a summary of the fracture classification pathway.

Trial methods: baseline radiographs and assessing study eligibility. The trial manual provided to participating sites included a PowerPoint (Microsoft Corporation, Redmond, Washington) presentation illustrating the recommended full shoulder trauma series (three 'perpendicular' views: anteroposterior view; scapular Y-lateral view; and axillary (modified axillary) view)[9] for assessing fracture eligibility. However, we felt it was necessary to adhere to local guidelines for radiographic assessment. We thus stipulated that a minimum of two radiographic views/projections were required for the assessment of study eligibility. The radiographic views taken were recorded on the study eligibility form for all patients who met the primary inclusion criteria (Table I). No additional imaging was required for trial purposes.

An introduction to the Neer classification was also included in the trial manual. In all trial information, the study eligibility criteria were expressed in terms of the Neer classification and the displacement criteria stated. However, in keeping with normal practice in busy fracture clinics, there was no expectation that the recruiting surgeons would classify the fractures and thus judge whether or not displaced parts met the Neer criteria. Instead, surgeons were asked to indicate on the eligibility form if the fracture involved either tuberosity.

Anonymised electronic versions of baseline radiographs, identified only via the patient's unique four-digit trial number used in the assessment of trial eligibility for each randomised patient, were sent on CDs to the York Trials Unit (YTU). Early on, we found several sites could not send us Digital Imaging and Communication in Medicine (DICOM) images. Therefore, Joint Photographic Experts Group (JPEG) files were requested, preferably with a resolution of 300 dpi, as these could be provided by all hospitals and can also be read easily on all computer platforms.

Initial processing of the radiographic images involved checks on anonymisation, correct labelling and ensuring that the images could be accessed and transferred electronically. For blinding purposes, the sets of radiographs were renumbered using a three-digit code, with letters used to label each of the radiographs within a set (e.g., 038a, 038b).

Preparations for the independent Neer classification. Preparations for the independent assessment and classification of trial baseline radiographs based on the Neer classification comprised an interim quality assessment of the radiographs; information gathering; removal of excess baseline radiographs; developing a protocol, training presentation and data collection tool (proforma); and testing these in a pilot study.

A protocol-specified monitoring and audit process was established to check for clear breaches of the main inclusion criterion (i.e., fractures should involve the surgical neck) and to assess the quality of the copies of the radiographic images available for each patient in terms of the suitability for fracture classification. Initially, the audit involved two consultant shoulder surgeons (AR and JJC-C), each of whom independently assessed the first five sets of images from each participating centre. Assessment was carried out for subsequent participants from each centre by one surgeon only (AR). The quality of the images was assessed according to three criteria: at least two projections in planes perpendicular to each other; proximal humeral and glenohumeral joint visible on each projection; and across all views, the shaft, greater tuberosity, lesser tuberosity, head of the humerus, and the glenohumeral joint could be identified. A failure to meet at least one criterion was considered an indication of potential difficulty for the Neer classification. A third surgeon acted as an arbiter to resolve disagreement at the final stage of the audit.

Information gathering included a review of the rater reliability studies on the Neer classification, with a particular focus on the approaches taken for assessing displacement and training. After trial recruitment, a postal survey was also sent to the radiographers representing the 33 participating hospitals that had screened patients for eligibility. This was done to obtain their specialist feedback on the imaging relating to fractures of the proximal humerus. The survey included questions on recommended views, assessing image quality, provision of specialist reports for use at fracture clinics, perceived difficulties in the interpretation of radiographs and the routine use of CT scans. Responses were received from 26 radiographers (79%).

An in-depth discussion with one radiographer at the lead site revealed that the accurate measurement of linear and angular displacement of bony parts using plain radiographs alone is unrealistic. The following was established: without the use of a calibration object, such as a scaling ball placed in an appropriate location at image acquisition, accurate measurement of scale is not possible. An approximation of scale is provided via picture archiving and communications systems (PACS), although the inbuilt scale is lost when patient details are removed upon anonymising the images. The problem is worse where there has been image manipulation, such as enlargement, prior to file closure. Thus,

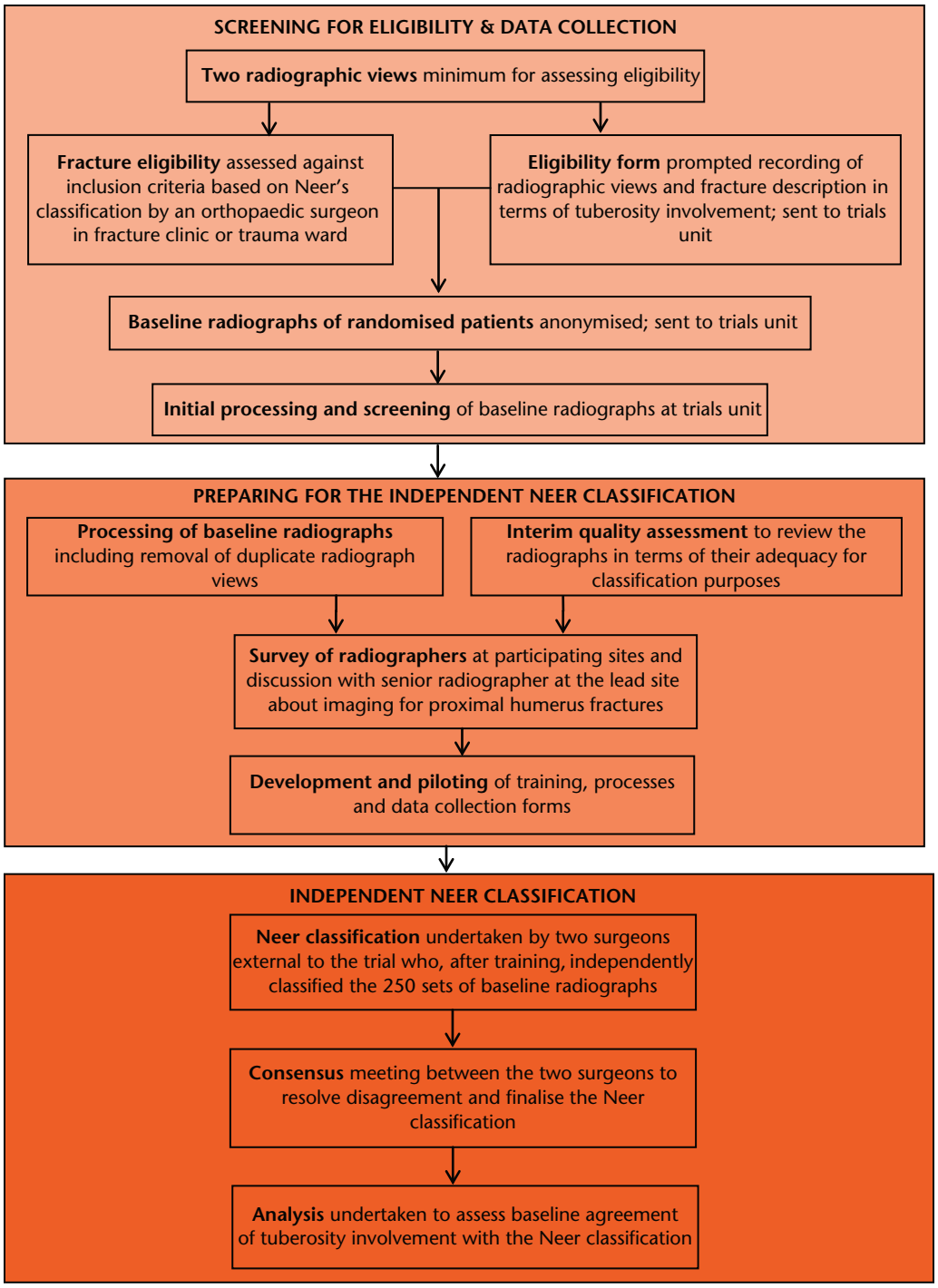

Fig. 2

Flow chart showing the fracture classification pathway.

assessing linear displacement is more challenging when there are anonymised copies of radiographs. One sub-optimal approach of limited applicability to assist the assessment of linear displacement is present in those radiographs where standard left/right markers (added to the cassette at image acquisition) are used, as the stems of the 'L' and 'R' are 9 mm. Judging angular displacement is also problematic as it depends on image orientation and the clear delineation of the edges of the bone parts.

Rather than use illustrations from Neer's original articles[2,10] for training purposes as undertaken by Brorson et al,[8] actual PROFHER trial images were used for facilitating the rater training. This decision ensured continuity in the images available for the decision-making process – from those seen by the recruiting surgeon to

Table I. PROximal Fracture of the Humerus Evaluation by Randomisation (PROFHER) inclusion and exclusion criteria

PROFHER trial
Inclusion criteria
Adults (aged 16 or above) presenting within three weeks of their injury with a radiologically confirmed displaced fracture of the humerus involving the surgical neck. This should include all two-part surgical neck fractures, as well as three-part (including surgical neck) and four-part fractures of the proximal humerus (Neer Classification). It may also include displaced surgical neck fractures that do not meet the exact displacement criteria of the Neer classification (> 1 cm or/and 45° angulation of displaced parts) where this reflects an individual surgeon's equipoise (i.e., whether or not the surgical neck fracture should be treated surgically).
Exclusion criteria
Associated dislocation of the injured shoulder joint
Open fracture
Mentally incompetent patient: unable to understand trial procedure or instructions for rehabilitation; significant mental impairment that would preclude compliance with rehabilitation and treatment advice
Comorbidities precluding surgery/anaesthesia
A clear indication for surgery such as severe soft-tissue compromise requiring surgery/emergency treatment (nerve injury/dysfunction)
Multiple injuries: same limb fractures; other upper limb fractures
Pathological fractures (other than osteoporotic) and terminal illness
Participant not resident in trauma centre catchment area

Table II. Protocol objectives for the independent classification of study fractures

The primary objective of this study is for two independent shoulder surgeons with training in the trial procedures and Neer's classification to assess the radiographs of all randomised patients in the proximal fracture of the humerus evaluation by randomisation trial in order to categorise the trial fractures reliably. This will enable completion of the following:

- A description of the study population to inform on the generalisability of the study findings.
- Assessment of the potential difference between the actual and the originally intended trial population in terms of:
  - The original categories of fracture types (i.e. 3, 8, 9 and 12; see Fig. 1) *versus* those resulting from a pragmatic application of the displacement criteria and
  - Protocol violations (i.e. non-involvement of the surgical neck; dislocation at the shoulder joint).
- Quantification of the differences between 'involvement' of the tuberosities as recorded by recruiting surgeons on study eligibility forms compared with that of the two independent surgeons and also for the latter surgeons' assessment of whether (a) the surgical neck is involved and (b) there is 'displacement' of parts according to Neer's categorisation; this will enable some insight as to whether what we did for the trial is a suitable proxy of Neer's classification.
- Identification, and thus quantification, of the numbers of fractures in the two subgroup categories (2-part *versus* 3- and 4-part) for secondary subgroup analyses.
- Assessment of the correspondence between the actual population in the distribution of the four original fracture categories and that based on the prospective epidemiological study by Court-Brown et al.[11]

those seen by the independent assessor. Similarly, in keeping with the trial's pragmatic design, we recruited two independent United Kingdom-based consultant orthopaedic surgeons who were experienced in treating fractures of the proximal humerus and who had experience comparable with that of surgeons in the PROFHER trial. From the literature and our experience, it was clear that involving more raters would not improve agreement.

The number of radiographs received for each trial participant ranged from two to seven. A maximum of four images per set of radiographs was arranged for the Neer classification. The chief investigator (AR) selected the best-quality projection (based on pre-specified criteria) in each perpendicular plane. The 62 poorer-quality duplicates in these planes were removed, with the reasons for exclusion being recorded and checked by two other authors (SDB and HHGH).

The objectives listed in our protocol are summarised in Table II. Together with our protocol, which included both a pictorial and a verbal description of the 16 categories of the Neer classification,[2,5] we prepared a Neer training presentation and a proforma to aid in the description and assessment of the quality of each set of radiographs for classification purposes and in the assessment of the Neer classification of the study fractures. The proforma facilitated the recording of data on the displacement of structures according to the Neer classification, as well as data on 'involvement' (any indication of a fracture), and a 'no contact' surgical neck fracture (bone parts/fragments do not overlap). As well as 'undisplaced' fractures (or clearly insufficiently displaced to meet Neer's criteria), and displaced fractures according to Neer, the proforma allowed for an element of doubt with the addition of an extra category: 'displaced (unclear if Neer displacement criteria met)'. Other characteristics collected included whether or not the head segment was in varus or valgus.

The training presentation and data collection process were piloted with the help of two senior orthopaedic registrars using ten sets of patient radiographs selected according to pre-specified criteria to give an adequate range in fracture type, view type and number. The pilot resulted in some adjustments and additions, including a briefing document on the Neer classification of radiographs, and a realistic timetable for the main assessment process. For each independent assessor, this comprised a training day, 20 hours to assess 250 sets of radiographs and up to a day to achieve consensus.

Table III. Radiographic views for study participants (eligibility form and Neer's raters)

Radiographic view (eligibility form and Neer's proforma)	Eligibility form (n = 250)		Rater 1 (n = 250)		Rater 2 (n = 250)	
	n	%	n†	%	n†	%
Anteroposterior only*	22	*8.8*	10[1]	*4.0*	7[1]	*2.8*
Axillary only	0	*0.0*	1[1]	*0.4*	0	*0.0*
Scapular Y-lateral only	0	*0.0*	1[1]	*0.4*	0	*0.0*
Anteroposterior + axillary	85	*34.0*	87[4]	*34.8*	82	*32.8*
Anteroposterior + scapular Y-lateral	61	*24.4*	91[3]	*36.4*	95	*38.0*
Axillary + scapular Y-lateral	0	*0.0*	1[1]	*0.4*	1	*0.4*
Anteroposterior + axillary + scapular Y-lateral	76	*30.4*	58[1]	*23.2*	65[2]	*26.0*
Missing	6	*2.4*	1	*0.4*	0	*0.0*

*Combines scapular and coronal plane views (Neer classification only)
†Superscript numbers (1 to 4) indicate how many patients had named radiograph view(s) plus one 'other' view (Neer classification only)

Independent Neer classification. The two consultant orthopaedic surgeons (AJB and AJJ), who were from non-trial-participating hospitals, signed a formal agreement to commit to the requirements for participation, including arranging protected time. The same format and sets of radiographs were used for the main study training day as in the pilot. Subsequently, the surgeons returned copies of their completed data collection forms (to SDB). The results were collated and a table returned indicating where there were differences in the Neer classification for individual fractures. The two surgeons met to resolve these differences and document the decisions behind each of the final verdicts.

Assessment of tuberosity involvement at baseline (no/yes) was cross-tabulated against Neer's classification (one- and two-part, three- and four-part fractures). The Kappa statistic was used to measure agreement in the classification of the fractures for the two assessments.

Results

Baseline radiographic views. The study eligibility forms completed at the time of patients being considered to enter the trial showed that the minimum requirement of two named radiographic views (from anteroposterior, axillary (or modified axillary), and scapular Y-lateral) was achieved for 1104 (88%) of the 1250 screened patients. The anteroposterior view was recorded in 1219 patients (98%), the axillary view in 708 patients (57%) and the scapular Y-lateral view in 525 patients (50%). The standard trauma series (all three named views) was recorded in 224 (18%) of screened patients. These findings are compatible with the results of the survey of radiographers at participating centres. Responses confirmed that two views (always including the anteroposterior view with either the axillary or scapular Y-lateral views depending on patient's condition and other practicalities) were required at 25 hospitals and three views at one hospital.

Table III shows the breakdown of the radiographic views reported on the eligibility forms at the time of baseline assessment for the randomised patients, together with the assessment by the two independent persons rating the views available to them for Neer classification. There were two radiographic images for 165 patients, three for 68 patients and four for 17 patients for this assessment. Compared with the baseline assessment, the raters judged that there were a greater number of anteroposterior plus scapular Y-lateral views, and fewer anteroposterior only, and anteroposterior plus axillary plus scapular Y-lateral views. There was little difference in assessment between the two raters (91% agreement, kappa 0.87, $p < 0.001$), whereas agreement compared with baseline had a weaker rate of agreement (rater 1: 56% agreement, kappa 0.39, $p < 0.001$; rater 2: 59% agreement, kappa 0.43, $p < 0.001$). All ten radiographic sets, for which either rater indicated a single plane only (excluding any additional 'other' views), had all been rated to show a minimum of two planes on the eligibility form.

Assessment of radiographic quality. Although repeated requests were sometimes required to hospital radiology departments because of problems with file transfer and anonymisation, ultimately JPEG files were available for all baseline radiographs. All images were below the requested 300 dpi resolution; where checked, the majority were at 96 dpi. The initial quality assessment involving three surgeons identified 46 radiograph sets (18%) that were likely to present future difficulties for the independent Neer classification of fractures. Feedback on image quality from the hospital radiographers suggested exposure and patient positioning were key components in their judgement of image quality.

The two independent raters were asked to evaluate the quality of the radiographs available for each patient in terms of their adequacy for classification purposes. There was good agreement (94%) between the two raters for the key question, namely 'Considering all the available views together, can you visualise the location of all five structures (the humeral shaft, greater tuberosity, lesser tuberosity, head of humerus and glenohumeral joint) sufficiently to determine the position and displacement of the fractured segments?' The answer was 'no' for two sets for rater 1, and for 16 sets for rater 2. Although our proforma for Neer's classification specifically asked the raters to indicate the structures that were visible on each radiograph, this global assessment is more likely to reflect clinical practice.

Table IV. Initial and consensus Neer's classification by the two independent raters

Neer's classification categories identified during independent assessment		Initial classification by the two raters				Consensus classification	
		Rater 1 (n = 250)		Rater 2 (n = 250)		(n = 250)	
		n	%	n	%	n	%
1*	Neer 1 part: undisplaced† surgical neck	8	*3.2*	21	*8.4*	18	*7.2*
3	Neer 2 part: surgical neck	**127**	***50.8***	**129**	***51.6***	**119**	***47.6***
4*	Neer 2 part: greater tuberosity	2	*0.8*	14	*5.6*	8	*3.2*
5*	Neer 2 part: lesser tuberosity	1	*0.4*	0	*0.0*	1	*0.4*
8	Neer 3 part: surgical neck + greater tuberosity	**81**	***32.4***	**81**	***32.4***	**90**	***36.0***
9	Neer 3 part: surgical neck + lesser tuberosity	**0**	***0.0***	**0**	***0.0***	**1**	***0.4***
10*	Neer 3 part: anterior dis-location + greater tuberosity	0	*0.0*	0	*0.0*	2	*0.8*
12	Neer 4 part: surgical neck + greater + lesser tuberosity	**30**	***12.0***	**4**	***1.6***	**11**	***4.4***
13*	Fracture-dislocation – anterior (4 part)	1	*0.4*	0	*0.0*	-	-
15*	Fracture-dislocation – anterior (articular surface)	0	*0.0*	1	*0.4*	-	-

* Categories outside the expected categories for the trial
†'Undisplaced' according to Neer's definition

Other sources of information that could inform surgeons' assessment of study eligibility. The majority of radiographers in our survey (n = 24) provided specialist reports for use at the fracture clinic either all, or most, of the time. However, the feedback also showed that the availability of the specialist report was unlikely to inform the interpretation of the radiographs by surgeons considering patient eligibility for the trial. There was no indication of problems in interpretation that made reporting of these fractures routinely difficult for radiographers. It was clear from feedback that few radiographers were familiar with Neer's linear displacement and angulation criteria; none used the classification system in their reports.

Only one radiographer, from a non-recruiting hospital, reported the routine use of CT scanning for these fractures. This is consistent with the reports of CT scans being used to help assess study eligibility in six randomised patients (three in each treatment allocation group), each from a different hospital.

Neer's classification of baseline fractures. Using assessments of individual anatomical features for each radiograph and arriving at a verdict for the involvement/displacement of each feature, the two raters independently assigned a Neer's classification value to each patient, which could take one of 16 possible categories (Fig. 1). Agreement between the two raters was moderate (68% agreement, kappa 0.48, $p < 0.001$). Overall, the two raters independently assigned the same category in 169 cases, the greatest between-rater difference was in category 12 (Table IV). After a consensus meeting, the two raters arrived at the final agreed Neer's classification of baseline fractures shown in Table IV. Classifications were very well balanced between trial arms, as had been tuberosity involvement reported on the study eligibility forms.[1]

Monitoring during recruitment and independent assessment confirmed that all fractures met the main inclusion criterion on involvement of a surgical neck fracture. However, as shown in Table IV, some categories were 'unexpected' fractures (categories 1, 4, 5 and 10), particularly those that were not associated with substantial displacement (Neer's criteria) of the surgical neck. When relaxing the criteria for assessing displacement of the surgical neck to include 'displaced but unclear if Neer displacement criteria met', the surgical neck fractures were clearly not displaced enough to meet the Neer criteria in far fewer cases. Notably, of the 18 fractures in category 1, rater 1 reported four fractures and rater 2 reported only one fracture that absolutely did not meet the Neer displacement criteria. The variation between the raters in assessing displacement was also manifest for 'no contact' surgical neck fractures and other characteristics and is shown in Table V.

When designing the trial, the estimates for the distribution of the Neer categories in the trial were derived from the epidemiological study by Court-Brown et al.[11] The relative proportions of the 221 fractures in the four expected categories in the trial show a greater proportion of three-part fractures (Table VI).

Agreement between baseline assessment of tuberosity involvement and Neer's classification. The agreed Neer's classifications by the two raters were compared against radiographic assessments at baseline of tuberosity involvement. This grouping of the two assessments was used for pre-specified subgroup analysis (tuberosity involvement at baseline no or yes) and subgroup sensitivity analysis (Neer's one-part plus two-part fractures *versus* three-part plus four-part factures). Table VII shows that the majority of baseline assessments of no tuberosity involvement (93%) were identified as Neer's one-part and two-part fractures. Conversely, half (48%) of patients with reported involvement of one or both tuberosities at baseline were associated with Neer one-part and two-part fractures, and the other half (52%) with three-part and four-part fractures. Accordingly, the agreement between the two assessments was fair (61% agreement, kappa 0.29, $p < 0.001$).

Table V. Separate fracture characteristics on Neer's proforma. Individual judgements of raters

Fracture characteristic	Rater 1 (n = 250)		Rater 2 (n = 250)	
	n	%	n	%
Surgical neck - no contact fracture	23	*9.2*	27	*10.8*
Surgical neck – impacted	59	*23.6*	76	*30.4*
Anterior fracture dislocation	1	*0.4*	0	*0.0*
Posterior fracture dislocation	0	*0.0*	0	*0.0*
Articular surface fracture	11	*4.4*	10	*4.0*
Head segment in varus	62	*24.8*	49	*19.6*
Head segment in valgus	69	*27.6*	77	*30.8*

Table VI. Relative proportions of expected Neer fracture categories for the proximal fracture of the humerus evaluation by randomisation (PROFHER) trial

Source	Fracture categories (%)			
	3	8	9	12
Court-Brown[11] Proportions of overall population (categories 1 to 16)	*28.0*	*9.0*	*0.3*	*2.0*
Court-Brown[11] Proportions of 3, 8, 9 and 12	*71.2*	*22.9*	*0.8*	*5.1*
PROFHER Proportions of 3, 8, 9 and 12	*53.8*	*40.7*	*0.5*	*5.0*

Table VII. Agreed Neer's classification and tuberosity assessment at baseline

Agreed Neer's classification	Tuberosity involvement (eligibility form)			
	None	(n = 57)	Greater and/or lesser tuberosity*	(n = 193)
	n	%	n	%
Neer 1 + 2 part	53	*93.0*	93	*48.2*
Neer 3 + 4 part	4	*7.0*	100	*51.8*

*The distribution was split into 119 greater tuberosity only, 10 lesser tuberosity only and 64 both tuberosities

Discussion

This article describes the systematic approach taken to define the fracture population of the multicentre PROFHER trial, both at recruitment, and via an independent and blinded assessment in terms of the Neer classification. The approach taken and methods used accommodate the known limitations of the Neer classification system and maximise external validity (applicability of trial findings) and accuracy in the formal definition of the fracture population. The key product is the description of the study fracture population according to the Neer classification.

Both the intended and actual fracture population represent the collective and individual uncertainty as to whether surgery provided a better outcome for patients with these fractures. Both raters independently confirmed that all patients had sustained injuries typically considered for surgery in contemporaneous practice. Although a very few fractures of the surgical neck were 'minimally displaced', the actual distribution tended more towards complex fractures (Table VI). Indeed, there were proportionally more fractures without tuberosity fractures amongst those that were ineligible than in the trial population: 40.2% *versus* 22.8%.[1] Notably, the Neer classification resulted in just 11 four-part fractures (4.4%), but 64 fractures (25.6%) were recorded as involving both tuberosities on the trial eligibility forms. This disparity may be explained in part by the observation in a study by Brorson et al[12] that there is substantial confusion as to whether all involved segments for three- and four-part fractures should be displaced according to Neer's definition. Additionally, lesser tuberosity displacement is harder to gauge, perhaps contributing to the 10.4% disagreement between the two raters for these fractures (Table IV).

Although the agreement between the assessments (tuberosity involvement at baseline *versus* Neer parts) was only fair, good balance was maintained between groups in the Neer categories. Moreover, this disparity did not result in any important changes to the findings of the fracture subgroup analyses based on either grouping for the primary outcome, the Oxford Shoulder Score.[1] Neither subgroup analysis supported differentiating treatment (use of surgery) on the basis of these characteristics.

Insights from the hospital radiographer's survey and other sources point to the assessment of the fracture in terms of suitability for surgery in PROFHER being equivalent to that in the practice employed in the United Kingdom. This includes the availability of generally two views rather than the full trauma series, the assessment being undertaken by the surgeon using images of plain radiographs, and CT scans rarely being used. The latter is no drawback;[7] even sophisticated physical models of fractures do not improve inter-observer agreement among shoulder specialists.[13] Crucially, although the four-part aspect of the Neer classification was used in assessing eligibility, there was no expectation of a rigorous application of the Neer displacement criteria. This too is likely to reflect clinical practice, in which judgements are often made with incomplete evidence and with suboptimal images. Considerations of image quality also relate to our decision to obtain JPEG files rather than DICOM files. This reflected the practicalities of obtaining a standardised set of portable images at the time, even if it was at the expense of loss of resolution. In the event, all images could be classified, and thus compressing the images to JPEG files for ease of use proved practical and effective.

The lack of a calibration object is especially noteworthy when an exact linear threshold is to be met. Yet, as confirmed by Neer, his displacement criteria are arbitrary,[9] and introduced in response to a stipulation by the journal editor prior to publication in the *JBJS [Am]*.[3] Furthermore, there is evidence that surgeons agree more

on treatment options (conservative treatment, locking plate fixation, hemiarthroplasty) than on fracture classification.[12] Lastly, we cannot rule out that some surgeons may consider other fracture characteristics, such as varus/valgus positioning of the head segment, not explicitly covered in the Neer classification. Nonetheless, we confirmed that fractures with these characteristics were present in the study population.

The involvement of two independent raters helped us to obtain as good a summary of the fractures characterised according to Neer as was feasible, and thus fulfil a key reporting requirement. As expected, there was disparity between how fractures were assessed during busy clinical practice for tuberosity involvement compared with the independent Neer classification. This did not, however, affect the relative distributions between treatment groups with respect to fracture type. Nor did the fracture classification at baseline *versus* the Neer classification influence the effect of treatment on the primary patient outcome. This has endorsed the pragmatic approach we took to classify fractures realistically when screening for patient eligibility during clinical practice. These insights are likely to apply to other fracture classifications, including those with displacement thresholds, which typically have similar problems of poor inter-observer agreement. The process we have described in this article should help mitigate these difficulties and provide an important guide for researchers designing pragmatic multicentre clinical trials on fracture management. Finally, we note that our key underlying philosophy reflected normal clinical practice, where emphasis on the surgeon's judgement, rather than observation of exacting fracture classification criteria, was used for recruitment. This approach tallies with the aforementioned findings of Brorson et al.[12] Similar research to identify the basis for surgeons' decisions in the treatment of other fractures is also likely to provide valuable insights.

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;313:1037-1047.
2. **Neer CS II.** Displaced proximal humeral fractures. I. Classification and evaluation. *J Bone Joint Surg [Am]* 1970;52-A:1077-1089.
3. **Carofino BC, Leopold SS.** Classifications in brief: the Neer classification for proximal humerus fractures. *Clin Orthop Relat Res* 2013;471:39-43.
4. **Brorson S, Eckardt H, Audigé L, Rolauffs B, Bahrs C.** Translation between the Neer- and the AO/OTA classification for proximal humeral fractures: do we need to be bilingual to interpret the scientific literature? *BMC Res Notes* 2013;6:69.
5. **Sidor ML, Zuckerman JD, Lyon T, et al.** The Neer classification system for proximal humeral fractures. An assessment of interobserver reliability and intraobserver reproducibility. *J Bone Joint Surg [Am]* 1993;75-A:1745-1750.
6. **Brorson S, Hróbjartsson A.** Training improves agreement among doctors using the Neer system for proximal humeral fractures in a systematic review. *J Clin Epidemiol* 2008;61:7-16.
7. **Sjödén GO, Movin T, Güntner P, et al.** Poor reproducibility of classification of proximal humeral fractures. Additional CT of minor value. *Acta Orthop Scand* 1997;68:239-242.
8. **Brorson S, Bagger J, Sylvest A, et al.** Improved interobserver variation after training of doctors in the Neer system. A randomised trial. *J Bone Joint Surg [Br]* 2002;84-B:950-954.
9. **Neer CS II.** Four-segment classification of proximal humeral fractures: purpose and reliable use. *J Shoulder Elbow Surg* 2002;11:389-400.
10. **Neer CS II.** Displaced proximal humeral fractures. II. Treatment of three-part and four-part displacement. *J Bone Joint Surg [Am]* 1970;52-A:1090-1103.
11. **Court-Brown CM, Garg A, McQueen MM.** The epidemiology of proximal humeral fractures. *Acta Orthop Scand* 2001;72:365-371.
12. **Brorson S, Olsen BS, Frich LH, et al.** Surgeons agree more on treatment recommendations than on classification of proximal humeral fractures. *BMC Musculoskelet Disord* 2012;13:114.
13. **Majed A, Macleod I, Bull AM, et al.** Proximal humeral fracture classification systems revisited. *J Shoulder Elbow Surg* 2011;20:1125-1132.

Funding Statement

- This work, as part of the PROFHER trial funding, was funded by the National Institute for Health Research (NIHR) Health Technology Assessment Programme (Project number: 06/404/53) and is published in full in UK Health Technology Assessment Vol.19, Issue 24. See the HTA Programme website for further project information.
- Mr A. Brooksbank reports providing expert testimony for Medicolegal; this is outside the submitted work and did not influence the trial or this report.
- Professor A. Johnstone reports grants and personal fees from Clearsurgical Ltd and Invibio Ltd; both are outside the submitted work. In addition, Professor Johnstone has a number of international patents. None of these influenced the trial or this report.
- Professor A. Rangan reports grants and personal fees from De Puy Ltd, and grants from JRI Ltd; all are outside the submitted work. None of these influenced the trial or this report.
- None relevant declared for the other authors

Acknowledgements

- The authors wish to thank Mr B. Cox for his advice about collecting radiographic images and measuring displacement; Professor A. Carr for his help with the audit; and Mr M. Ismail and Mr W. Eardley for their assistance with piloting the training in the Neer classification of the radiographs. We are indebted to the many hospital radiographers who contributed to the PROFHER trial.

Author Contribution

- H. H. G. Handoll: Input into all aspects of classification-related trial methods, forms and survey design, Writing the paper
- S. D. Brealey: Trial management, Input into all aspects of classification-related trial methods, forms and survey design, Advice on content, Revising the paper
- L. Jefferson: Trial management, including processing of baseline images and radiographers survey, Revising the paper
- A. Keding: Advised on and performed analysis of the associated datasets, Revising the paper
- A. J. Brooksbank: Independent rater of the baseline images, Revising the paper
- A. J. Johnstone: Independent rater of the baseline images, Revising the paper
- J. J. Candal-Couto: Monitoring and quality assessment of baseline images, Revising the paper
- A. Rangan: Chief investigator, Monitoring and quality assessment of baseline images, Input into all aspects including training, Advice on content, Revising the paper

ICMJE conflict of interest

- None declared