The ProFHER (PROximal Fracture of the Humerus: Evaluation by Randomisation) trial – a pragmatic multicentre randomised controlled trial evaluating the clinical effectiveness and co...

10 authors, including:

Stephen Brealey
The University of York
104 PUBLICATIONS   3,889 CITATIONS

Amar Rangan
National Health Service
183 PUBLICATIONS   3,323 CITATIONS

Ada Keding
The University of York
114 PUBLICATIONS   1,968 CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Children Learning About Secondhand Smoking (CLASS III): A cluster randomized controlled-trial in Bangladesh and Pakistan View project

Project   Developing a training intervention for staff to improve participant recruitment into randomised controlled trials View project

Original Investigation

Surgical vs Nonsurgical Treatment of Adults With Displaced Fractures of the Proximal Humerus
The PROFHER Randomized Clinical Trial

Amar Rangan, FRCS(Tr&Orth); Helen Handoll, DPhil; Stephen Brealey, PhD; Laura Jefferson, PhD; Ada Keding, MSc; Belen Corbacho Martin, MSc; Lorna Goodchild, MSc; Ling-Hsiang Chuang, PhD; Catherine Hewitt, PhD; David Torgerson, PhD; for the PROFHER Trial Collaborators

IMPORTANCE The need for surgery for the majority of patients with displaced proximal humeral fractures is unclear, but its use is increasing.

OBJECTIVE To evaluate the clinical effectiveness of surgical vs nonsurgical treatment for adults with displaced fractures of the proximal humerus involving the surgical neck.

DESIGN, SETTING, AND PARTICIPANTS A pragmatic, multicenter, parallel-group, randomized clinical trial, the Proximal Fracture of the Humerus Evaluation by Randomization (PROFHER) trial, recruited 250 patients aged 16 years or older (mean age, 66 years [range, 24-92 years]; 192 [77%] were female; and 249 [99.6%] were white) who presented at the orthopedic departments of 32 acute UK National Health Service hospitals between September 2008 and April 2011 within 3 weeks after sustaining a displaced fracture of the proximal humerus involving the surgical neck. Patients were followed up for 2 years (up to April 2013) and 215 had complete follow-up data. The data for 231 patients (114 in surgical group and 117 in nonsurgical group) were included in the primary analysis.

INTERVENTIONS Fracture fixation or humeral head replacement were performed by surgeons experienced in these techniques. Nonsurgical treatment was sling immobilization. Standardized outpatient and community-based rehabilitation was provided to both groups.

MAIN OUTCOMES AND MEASURES Primary outcome was the Oxford Shoulder Score (range, 0-48; higher scores indicate better outcomes) assessed during a 2-year period, with assessment and data collection at 6, 12, and 24 months. Sample size was based on a minimal clinically important difference of 5 points for the Oxford Shoulder Score. Secondary outcomes were the Short-Form 12 (SF-12), complications, subsequent therapy, and mortality.

RESULTS There was no significant mean treatment group difference in the Oxford Shoulder Score averaged over 2 years (39.07 points for the surgical group vs 38.32 points for the non-surgical group; difference of 0.75 points [95% CI, −1.33 to 2.84 points]; P = .48) or at individual time points. There were also no significant between-group differences over 2 years in the mean SF-12 physical component score (surgical group: 1.77 points higher [95% CI, −0.84 to 4.39 points]; P = .18); the mean SF-12 mental component score (surgical group: 1.28 points lower [95% CI, −3.80 to 1.23 points]; P = .32); complications related to surgery or shoulder fracture (30 patients in surgical group vs 23 patients in nonsurgical group; P = .28), requiring secondary surgery to the shoulder (11 patients in both groups), and increased or new shoulder-related therapy (7 patients vs 4 patients, respectively; P = .58); and mortality (9 patients vs 5 patients; P = .27). Ten medical complications (2 cardiovascular events, 2 respiratory events, 2 gastrointestinal events, and 4 others) occurred in the surgical group during the postoperative hospital stay.

CONCLUSIONS AND RELEVANCE Among patients with displaced proximal humeral fractures involving the surgical neck, there was no significant difference between surgical treatment compared with nonsurgical treatment in patient-reported clinical outcomes over 2 years following fracture occurrence. These results do not support the trend of increased surgery for patients with displaced fractures of the proximal humerus.

TRIAL REGISTRATION isrctn.com Identifier: ISRCTN50850043

Proximal humeral fractures account for 5% to 6% of all adult fractures[1]; an estimated 706 000 occurred worldwide in 2000.[2] The majority occur in people older than 65 years.[1,3] Similar to other primarily osteoporotic fractures, the age-specific incidence of these fractures is increasing with a 2.5-fold increase in women and a 3.4-fold increase in men older than 60 years reported between 1970 and 2002 in Finland.[4] Approximately half (51%) of these fractures are displaced, the majority of which involve the surgical neck (77%).[5] Surgical treatment (mainly internal fixation or humeral head replacement) is being increasingly used,[6] contributing to increased treatment costs for upper limb fractures.[7] A Cochrane review found insufficient evidence from randomized clinical trials to conclude whether surgical intervention produces consistently better outcomes than nonsurgical treatment.[8]

In the Proximal Fracture of the Humerus Evaluation by Randomization (PROFHER) trial, we aimed to examine whether surgical treatment compared with nonsurgical treatment resulted in better patient-reported outcomes for displaced fractures of the proximal humerus involving the surgical neck.[9]

Methods

Study Design and Eligibility Criteria

The PROFHER collaborators recruited patients for this pragmatic, open, multicenter, parallel-group, superiority, randomized clinical trial between September 2008 and April 2011 from orthopedic departments (fracture clinics or wards) at 32 acute UK National Health Service hospitals. There was 1 additional National Health Service hospital that screened but failed to recruit patients into the PROFHER trial. All 33 hospitals routinely provide surgical and nonsurgical fracture treatment. Patients received their allocated treatment from the recruiting hospital. Care pathways for all patients included outpatient- and community-based rehabilitation, which primarily comprised 1 or more 1-on-1 sessions with a physiotherapist that focused on restoring function. Follow-up was up to 2 years (until April 2013) for all patients.

Patients were eligible for inclusion if they were aged 16 years or older and presented within 3 weeks after sustaining a displaced fracture of the proximal humerus that involved the surgical neck. The degree of displacement had to be sufficient for the treating surgeon to consider surgical intervention but did not have to meet the displacement criteria of Neer[10] (1 cm or 45° angulation of displaced parts, or both) for inclusion in the trial. This relaxing of the displacement criteria reflected the arbitrariness of these thresholds.[11] Excluded patients had associated dislocation of the injured shoulder joint, open fracture, insufficient mental capacity to understand the trial or instructions for rehabilitation, comorbidities precluding surgery or anesthesia, clear indication for surgery such as severe soft-tissue compromise, multiple injuries (upper limb fractures), pathological fracture (other than osteoporotic), terminal illness, or were not a resident in the hospital catchment area.

The PROFHER protocol[9] (additional details appear in Supplement 1) and all amendments were approved by the York

research ethics committee. Approval was also obtained from all research and development offices of participating hospitals and from 14 primary care trusts to allow for data collection by community physiotherapists. All participants gave written informed consent. The trial was overseen by a trial steering committee, which included an independent chair and 2 former patients treated for proximal humeral fracture, and an independent data monitoring and ethics committee.

Randomization and Blinding

After obtaining informed consent and key baseline information, research associates randomly allocated patients to surgical or nonsurgical treatment using an independent remote randomization service (telephone or online access) provided by the York Trials Unit (University of York). Randomization was performed using a computer program with 1:1 allocation, stratifying by tuberosity involvement (yes or no) and using random block sizes of 4, 8, and 12. There was no blinding of trial participants, clinicians, or assessment of outcomes. There was independent data entry, processing, and analysis. Data (including baseline x-rays) were anonymized before distribution using unique study IDs. Coding was performed by at least 2 independent coders blinded to treatment allocation. The primary clinical analyses were repeated by a second blinded statistician using different statistical software.

Interventions

It was emphasized that good standards of care, both surgical and nonsurgical, should be provided throughout the treatment pathway for the injury, including surgical care or management of the sling, postoperative care, and rehabilitation in both groups. Participating hospitals did not introduce new or experimental interventions for these fractures during the study. To avoid learning curve problems, surgeons and physiotherapists used surgical interventions and procedures with which they were familiar.

Participants allocated to surgery received either internal fracture fixation (eg, with plate and screws) that preserved the humeral head or humeral head replacement (hemiarthroplasty). Participants allocated to nonsurgical treatment were given a sling for the injured arm for as long as deemed necessary (3 weeks was suggested), followed by active rehabilitation. Delivery of care and rehabilitation, which was freely available for all patients, incorporated the following 3 set measures to ensure good standards of care within the National Health Service: provision of an information leaflet on personal care during sling immobilization; a basic treatment protocol to guide physiotherapy; and promotion of home exercises. Rehabilitation care was provided by physiotherapists in inpatient, outpatient, and community settings.

Data Collection and Outcome Measures

Data collection was via hospital forms (baseline characteristics and details of surgery, details of inpatient stay, treatment confirmation at 1 month, physiotherapy and end of physiotherapy, 1-year and 2-year follow-up, and adverse events and review), baseline x-rays, and patient questionnaires at 3, 6, 12, and 24 months. Reflecting known differences in osteoporosis-

related fracture risk among different races, trial participants were asked to identify whether they were white, black, Asian, or other.

The primary outcome was the Oxford Shoulder Score (OSS), which is a shoulder-specific outcome measure validated in a UK population.[12] The OSS provides a total score based on the patient's subjective assessment of pain and function. It contains 12 items, each with 5 categories of response, and a range of total scores of 0 (worst outcome) to 48 (best outcome).[13] The OSS was collected by postal questionnaires at 6, 12, and 24 months. There was no trial-related clinical assessment or radiological follow-up of patients.

Secondary outcome measures were the Short-Form 12 (SF-12) health survey,[14] complications related to surgery and shoulder fracture (eg, surgical site infection, symptomatic malunion, and avascular necrosis of the humeral head), complications requiring secondary surgery or treatment, medical complications during inpatient stay, and mortality. The SF-12, collected along with the primary outcome of OSS, was divided into the physical and mental component scores with a range of 0 (lowest level of health) to 100 (highest level of health). The 3-level version of the EuroQol health status measure (EuroQol 5D)[15] was also collected at baseline and at 3, 6, 12, and 24 months for the economic evaluation of this trial.

Fracture Classification

To describe the study fracture population, 2 independent and blinded shoulder surgeon specialists characterized all study fractures from baseline x-rays using the Neer classification.[10] This was preceded by a piloted training session to improve interrater agreement.[16]

Statistical Analysis

Based on observational data from a cohort of patients who had complication-free surgery (A.R., unpublished data, 2014) and a cohort of patients treated nonsurgically,[17] we found a 5-point difference in OSS (primary outcome). The developers of the OSS agreed this difference represented a minimal clinically important difference.[13] Using an SD of 12,[17] this equated to an effect size of 0.42. One hundred participants in each group were required to detect an effect size of 0.40 with 80% power and a 5% significance level. Allowing for a 20% loss to follow-up, we planned to recruit 250 patients (125 in each group).

Analyses followed a prespecified analysis plan, endorsed by the data monitoring and ethics committee, and were performed using Stata version 12 (StataCorp). All analyses were on an intention-to-treat basis and included all randomized patients in the groups to which they were randomized. Significance tests were 2-sided at the 5% significance level.

The primary analysis compared OSS data from the surgical and nonsurgical treatment groups over all 3 follow-up assessments (6, 12, and 24 months). A multilevel, random-slope model was fitted to the data with time points nested in patients to allow for clustering of data within each patient. This model adjusted for the fixed effects of treatment group, time (6, 12, or 24 months), treatment × time interaction, tuberosity involvement at baseline (yes or no), age (<65 years or ≥65

years), sex, and health status at baseline (EuroQol 5D). An unstructured covariance pattern was selected for the repeated measurements as the least restrictive structure, which resulted in better model fit based on log-likelihood values than more constrained patterns. Estimates of the difference in OSS between treatment groups were assessed overall and at individual time points.

Any response bias was partially minimized by using a multilevel model, which allowed for the inclusion of intermittent responders in the primary analysis. The OSS values for complete and intermittent responders were additionally compared. The effect of missing data also was assessed with a post hoc sensitivity analysis using multiple imputation by chained equations. Missing outcome and covariate data were predicted by age, sex, tuberosity involvement, smoking status, EuroQol 5D index score at baseline, and available OSS data at other follow-up points. In a separate analysis, any demographic and fracture characteristics at baseline that were associated with nonresponse (logistic regression P < .10) were added to the primary analysis model.

Two further sensitivity analyses were conducted: (1) smoking status was added to the primary model, reflecting a chance imbalance at baseline and (2) centers were added as a random effect to the primary model to explore clustering at the center level, which quantified the performance of the surgical team at the hospital rather than on individual surgeons. To explore differences in treatment response according to baseline patient preferences, a treatment group × patient treatment preference interaction was added to the base model.
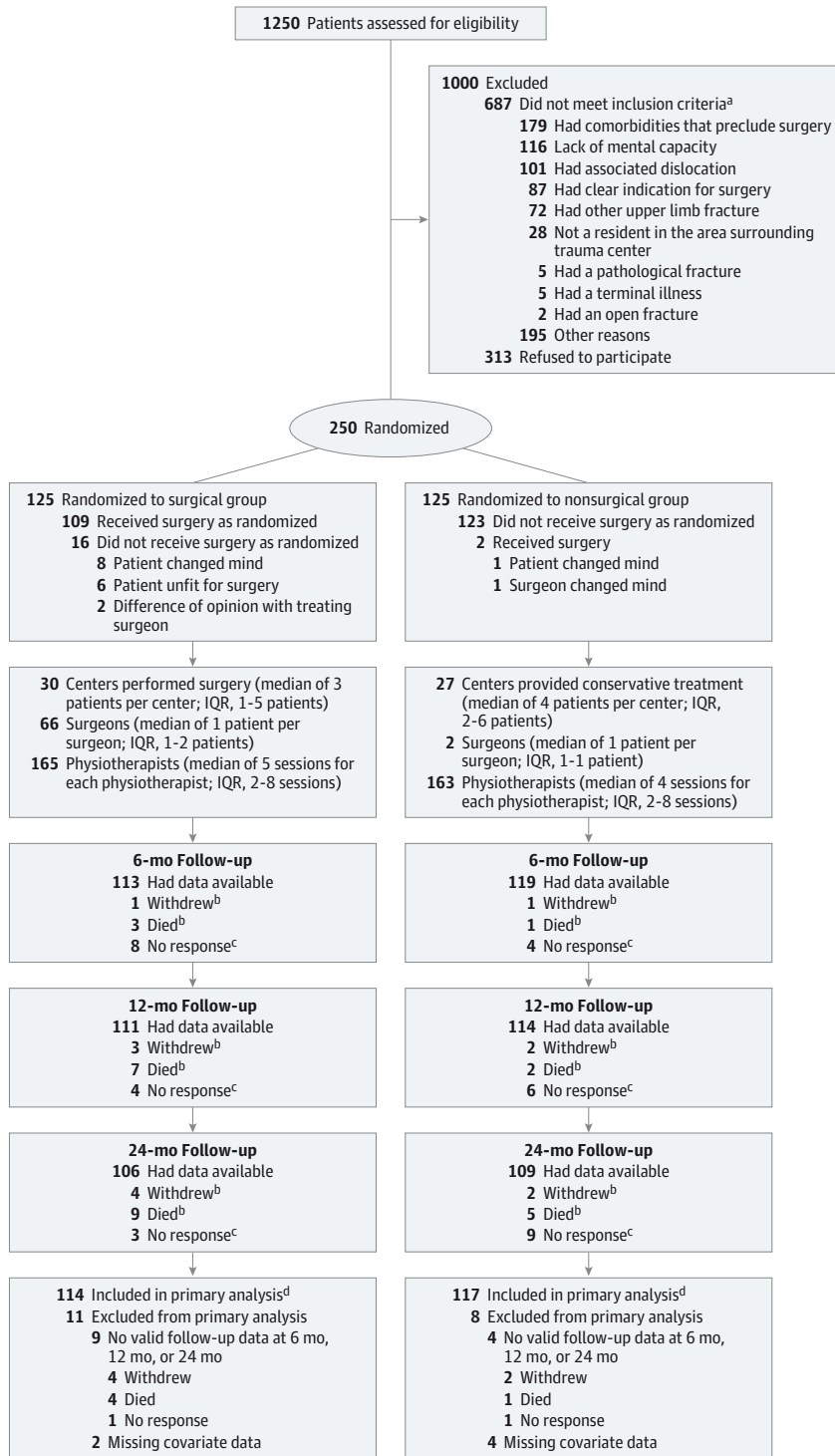
Two planned subgroup analyses were conducted with the expected direction of effect specified a priori in accordance with recent recommendations.[18] A treatment group × age × tuberosity involvement interaction term was added to the base model. Age (≥65 years or <65 years) was chosen because we expected greater benefit of surgery in patients younger than 65 years. Tuberosity involvement (none vs 1 or both tuberosities; sensitivity analysis: Neer 1-part and 2-part fractures vs Neer 3-part and 4-part fractures) was chosen because we expected greater benefit of surgery in patients with displacement of 1 or both tuberosities (Neer 3-part and 4-part fractures) than when neither tuberosity was involved or displaced (Neer 1-part and 2-part fractures). A change of −2 in log likelihood was compared between these models and the base model using the χ^2 test.

The SF-12 physical and mental component scores were analyzed by multilevel modeling using the same fixed effects and covariance structure as the OSS primary model. Frequencies of complications related to shoulder surgery and fracture, any treatments for these complications, and mortality rates were separately compared between treatment groups using the χ^2 test.

Results

Of 1250 patients screened, 563 (45%) were eligible and 687 (55%) were ineligible (**Figure 1**). The most common reason for exclusion (>1 reason per patient possible) was that a patient had co-

Figure 1. Flow of Patients in the Proximal Fracture of the Humerus Evaluation by Randomization Trial

1250 Patients assessed for eligibility

1000 Excluded
  687 Did not meet inclusion criteria[a]
    179 Had comorbidities that preclude surgery
    116 Lack of mental capacity
    101 Had associated dislocation
    87 Had clear indication for surgery
    72 Had other upper limb fracture
    28 Not a resident in the area surrounding trauma center
    5 Had a pathological fracture
    5 Had a terminal illness
    2 Had an open fracture
    195 Other reasons
  313 Refused to participate

250 Randomized

125 Randomized to surgical group
  109 Received surgery as randomized
  16 Did not receive surgery as randomized
    8 Patient changed mind
    6 Patient unfit for surgery
    2 Difference of opinion with treating surgeon

125 Randomized to nonsurgical group
  123 Did not receive surgery as randomized
  2 Received surgery
    1 Patient changed mind
    1 Surgeon changed mind

30 Centers performed surgery (median of 3 patients per center; IQR, 1-5 patients)
66 Surgeons (median of 1 patient per surgeon; IQR, 1-2 patients)
165 Physiotherapists (median of 5 sessions for each physiotherapist; IQR, 2-8 sessions)

27 Centers provided conservative treatment (median of 4 patients per center; IQR, 2-6 patients)
2 Surgeons (median of 1 patient per surgeon; IQR, 1-1 patient)
163 Physiotherapists (median of 4 sessions for each physiotherapist; IQR, 2-8 sessions)

6-mo Follow-up
113 Had data available
  1 Withdrew[b]
  3 Died[b]
  8 No response[c]

6-mo Follow-up
119 Had data available
  1 Withdrew[b]
  1 Died[b]
  4 No response[c]

12-mo Follow-up
111 Had data available
  3 Withdrew[b]
  7 Died[b]
  4 No response[c]

12-mo Follow-up
114 Had data available
  2 Withdrew[b]
  2 Died[b]
  6 No response[c]

24-mo Follow-up
106 Had data available
  4 Withdrew[b]
  9 Died[b]
  3 No response[c]

24-mo Follow-up
109 Had data available
  2 Withdrew[b]
  5 Died[b]
  9 No response[c]

114 Included in primary analysis[d]
11 Excluded from primary analysis
  9 No valid follow-up data at 6 mo, 12 mo, or 24 mo
  4 Withdrew
  4 Died
  1 No response
  2 Missing covariate data

117 Included in primary analysis[d]
8 Excluded from primary analysis
  4 No valid follow-up data at 6 mo, 12 mo, or 24 mo
  2 Withdrew
  1 Died
  1 No response
  4 Missing covariate data

IQR indicates interquartile range.

[a] Patients could be ineligible for more than 1 reason

[b] Missing data are cumulative at each given time point.

[c] Missing data are applicable at that time point only (patients could respond intermittently).

[d] Patients who did not complete follow-up at 1 time point for any reason could still be included in the primary analysis if they had a valid response at another time.

morbidities that precluded surgery or anesthetic. Of the 563 eligible patients, 250 (44%) consented to take part in the trial (Figure 1). These patients were randomized to surgical or nonsurgical treatment between September 2008 and April 2011. The mean age of the trial participants was 66 years (range, 24-92 years), 192 (77%) were female, and 249 (99.6%) were white.

These characteristics were similar to patients who refused consent (mean age, 68 years; 75% female). Ineligible patients tended to be older (mean age, 70 years) and there were slightly fewer females (72%) (eTable 1 in Supplement 2).

Of 125 patients allocated to the surgical group, 109 (87%) received surgery and 16 (13%) were treated nonsurgically. Of

Table 1. Baseline Characteristics and Risk Factors at Randomization and 24-Month Follow-up

	All Randomized Patients		Patients With OSS Data at 24 mo	
	Surgical Group (n = 125)	Nonsurgical Group (n = 125)	Surgical Group (n = 106)	Nonsurgical Group (n = 109)
Sex, No. (%)				
Male	28 (22.4)	30 (24.0)	26 (24.5)	25 (22.9)
Female	97 (77.6)	95 (76.0)	80 (75.5)	84 (77.1)
Age, y				
Mean (SD)	66.60 (11.80)	65.43 (12.09)	66.18 (11.1)	65.79 (11.97)
Median (range)	67.42 (27.04-92.04)	66.12 (24.63-89.02)	66.67 (37.09-87.76)	66.77 (31.33-89.02)
Age group, No. (%)				
<65 y	51 (40.8)	57 (45.6)	45 (42.5)	50 (45.9)
≥65 y	74 (59.2)	68 (54.4)	61 (57.6)	59 (54.1)
Race, No. (%)				
White	124 (99.2)	125 (100.0)	105 (99.1)	109 (100.0)
Black	1 (0.8)	0	1 (0.9)	0
Education, No. (%)				
High school degree	66 (52.8)	68 (54.4)	53 (50.0)	57 (52.3)
Some college	47 (37.6)	43 (34.4)	43 (40.6)	38 (34.9)
≥Bachelor's degree	12 (9.6)	14 (11.2)	10 (9.4)	14 (12.8)
Employment, No. (%)				
Part-time	12 (9.6)	7 (5.6)	12 (11.3)	7 (6.4)
Full-time	17 (13.6)	22 (17.6)	16 (15.1)	19 (17.4)
Self-employed	1 (0.8)	3 (22.4)	1 (0.9)	3 (2.8)
Retired	78 (62.4)	82 (65.6)	64 (60.4)	72 (66.1)
Unemployed but seeking work	3 (2.4)	1 (0.8)	3 (2.8)	1 (0.9)
Other	12 (9.6)	9 (7.2)	9 (8.5)	6 (5.5)
Missing	2 (1.6)	1 (0.8)	1 (0.9)	1 (0.9)
Diabetes, No. (%)				
Yes	18 (14.4)	13 (10.4)	15 (14.2)	11 (10.1)
No	106 (84.8)	111 (88.8)	90 (84.9)	97 (89.0)
Missing	1 (0.8)	1 (0.8)	1 (0.9)	1 (0.9)
Smoking status, No. (%)				
Yes	24 (19.2)	40 (32.0)	20 (18.9)	33 (30.3)
No	96 (76.8)	81 (64.8)	82 (77.4)	72 (66.1)
Missing	5 (4.0)	4 (3.2)	4 (3.8)	4 (3.7)
Steroid use, No. (%)				
Yes	6 (4.8)	7 (5.6)	6 (5.7)	6 (5.5)
No	118 (94.4)	116 (92.8)	100 (94.3)	102 (93.6)
Missing	1 (0.8)	2 (1.6)	0	1 (0.9)

Abbreviation: OSS, Oxford Shoulder Score.

the 125 patients allocated to nonsurgical treatment, 2 (2%) received surgery shortly after allocation.

The baseline characteristics (**Table 1** and **Table 2**) for randomized patients (N = 250) and those providing OSS data at 2 years (n = 215) were well balanced except for smoking status (there were more smokers in the nonsurgical group). Independent characterization of x-rays confirmed that the patients included in the PROFHER trial had sustained injuries that are typically considered for surgical intervention in contemporaneous practice. The assessment based on the Neer classification (eTable 2 in Supplement 2) identified 18 one-part fractures (9 in both groups), 128 two-part fractures (65 in surgical group vs 63 in nonsurgical group), 93 three-part fractures (46 vs 47, respectively), and 11 four-part fractures (5 vs 6).

For 109 participants who were allocated and received surgery, the procedure took place on average 10.4 days (range, 1-33 days) from the date of injury. The 109 operations were performed by 66 surgeons at 30 centers (Figure 1). The majority (82%) of operations were performed by a consultant (attending) surgeon (89 operations). A consultant was present in the operating room during 13 (12%) other operations, 5 of which were performed by a registrar (senior resident surgeon). Senior registrars or specialist fellows performed the remaining 7 operations, but with immediate access to a consultant in the operating suite, if required. Most (82.6%) of the surgeries involved locking plates (90 operations). The remaining patients received hemiarthroplasty (n = 10), intramedullary nails (n = 4), and other surgery (n = 5) (eTable 3 in Supplement 2).

Table 2. Health Status and Other Factors at Randomization and 24-Month Follow-up

	All Randomized Patients		Patients With OSS Data at 24 mo	
	Surgical Group (n = 125)	Nonsurgical Group (n = 125)	Surgical Group (n = 106)	Nonsurgical Group (n = 109)
EuroQol 5D index score[a]				
No. of patients	123	121	104	106
Mean (SD)	0.43 (0.37)	0.38 (0.37)	0.43 (0.35)	0.35 (0.36)
Median (IQR)	0.59 (−0.36 to 1)	0.26 (−0.35 to 1)	0.59 (−0.35 to 1)	0.26 (−0.35 to 1)
Time since injury, d				
Mean (SD)	5.78 (4.90)	5.69 (4.89)	5.81 (5.00)	5.69 (4.82)
Median (IQR)	4.00 (0 to 19.00)	4.00 (0 to 21.00)	4.00 (0 to 19.00)	4.00 (0 to 21.00)
Affected shoulder, No. (%)				
Left	57 (45.6)	68 (54.4)	46 (43.4)	58 (53.2)
Right	68 (54.4)	57 (45.6)	60 (56.6)	51 (46.8)
Tuberosity involvement, No. (%)				
Yes	99 (79.2)	94 (75.2)	85 (80.2)	83 (76.2)
Greater tuberosity	58 (46.4)	61 (48.8)	51 (48.1)	56 (51.4)
Lesser tuberosity	7 (5.6)	3 (2.4)	5 (4.7)	2 (1.8)
Greater and/or lesser tuberosity	34 (20.8)	30 (24.0)	29 (27.4)	25 (22.9)
No or missing	26 (20.8)	31 (24.8)	21 (19.8)	26 (23.9)
Fractures in past 10 y, No. (%)				
Yes	33 (26.4)	33 (26.4)	27 (25.5)	30 (27.5)
No	92 (73.6)	90 (72.0)	79 (74.5)	77 (70.6)
Missing	0	2 (1.6)	0	2 (1.8)
Previous surgery for fractures, No. (%)				
Yes	8 (6.4)	12 (9.6)	6 (5.7)	10 (9.2)
No	23 (18.4)	21 (16.8)	19 (17.9)	20 (18.4)
Missing	2 (1.6)	0	2 (1.9)	0
No previous fractures	92 (73.6)	92 (73.6)	79 (74.5)	79 (72.5)
Shoulder injury located on dominant side, No. (%)				
Yes	67 (53.6)	61 (48.8)	57 (53.8)	55 (50.5)
No	56 (44.8)	62 (49.6)	48 (45.3)	52 (47.7)
Missing	2 (1.6)	2 (1.6)	1 (0.9)	2 (1.8)
Injury mechanism, No. (%)				
Fall or trip from standing height or less	90 (72.0)	96 (76.8)	77 (72.6)	84 (77.1)
Fall down stairs or from a step height	18 (14.4)	17 (13.6)	15 (14.2)	15 (13.8)
Other	15 (12.2)	9 (7.2)	12 (11.3)	7 (6.4)
Missing	2 (1.6)	3 (2.4)	2 (1.9)	3 (2.8)

Abbreviations: IQR, interquartile range; OSS, Oxford Shoulder Score.

[a] Score range is −0.594 to 1 (based on UK reference data), in which 1 represents perfect health; 0, death; and values below 0, a state worse than death. A score of 1 was reported for 11 patients in the surgical group and 13 in the nonsurgical group; however, it is possible that this applied to their prefracture status.

Of 125 patients allocated to nonsurgical treatment, 82 (65.6%) received a broad arm type sling; 35 (28.0%), a collar and cuff; and 3 (2.4%), initially a hanging cast. Data were missing for 5 patients. Consistent with the study protocol, 29 of 32 recruiting centers (91%) recommended 3 or more weeks of sling use.

Physiotherapy treatment log data demonstrated equal access and implementation between groups, with similarly high numbers of participants recorded as performing home exercises (109 in surgical group vs 103 in nonsurgical group) in both groups (eTable 4 in Supplement 2).

Primary Outcome

Valid OSS responses were recorded for 232 patients (93%) at 6 months, 225 (90%) at 12 months, and 215 (86%) at 24 months,

with response rates balanced between treatment groups (details of losses to follow-up by group appear in Figure 1). The OSS data for at least 1 follow-up time point were available for 237 patients (95%), of which 231 (92%) also had complete covariate data and were included in the primary analysis (114 in surgical group vs 117 in nonsurgical group). There were no statistically significant differences between the 2 treatment groups during the 2-year period (difference of 0.75 points in favor of the surgical group [95% CI, −1.33 to 2.84 points]; P = .48) or at individual time points for the OSS (**Table 3**). The adjusted analysis (Table 3) was similar to the unadjusted data (**Figure 2**).

Overall, 41 patients (16%) had missing follow-up data on at least 1 time point. Using complete data derived by multiple

imputation resulted in comparable treatment effect estimates to the primary analysis (Table 3) with no overall statistically significant group difference (P = .48). However, complete responders tended to have better OSS values than

intermittent responders (statistically significant at 12 months, P = .03). Nonresponse (none or intermittent) was not associated with any demographic or fracture characteristics (all P > .10; eTable 5 in Supplement 2).

Table 3. Primary and Secondary Outcomes and Sensitivity Analyses

	Mean (95% CI)[a]		Mean Difference (95% CI)	P Value
	Surgical Group	Nonsurgical Group		
Primary Outcome				
Oxford Shoulder Score[b]				
No. of patients	114	117		
Averaged over 2 y	39.07 (37.30 to 40.76)	38.32 (36.57 to 39.99)	0.75 (−1.33 to 2.84)	.48
At 6 mo	37.84 (35.93 to 39.65)	35.59 (33.62 to 37.45)	2.25 (−0.07 to 4.57)	.06
At 12 mo	39.23 (37.38 to 40.99)	38.80 (36.99 to 40.53)	0.42 (−1.78 to 2.63)	.71
At 24 mo	40.11 (38.24 to 41.90)	40.40 (38.59 to 42.13)	−0.29 (−2.53 to 1.95)	.80
Secondary Outcomes[c]				
SF-12 physical component score				
No. of patients	111	115		
Averaged over 2 y	45.64 (43.44 to 47.84)	43.87 (41.75 to 45.99)	1.77 (−0.84 to 4.39)	.18
At 6 mo	45.73 (43.44 to 48.02)	43.18 (40.97 to 45.39)	2.55 (−0.21 to 5.32)	.07
At 12 mo	45.51 (43.22 to 47.80)	44.22 (42.01 to 46.43)	1.29 (−1.48 to 4.06)	.36
At 24 mo	45.68 (43.28 to 48.08)	44.20 (41.87 to 46.54)	1.48 (−1.48 to 4.43)	.33
SF-12 mental component score				
No. of patients	111	115		
Averaged over 2 y	48.66 (46.55 to 50.77)	49.96 (47.92 to 52.00)	−1.28 (−3.80 to 1.23)	.32
At 6 mo	48.43 (46.07 to 50.80)	48.95 (46.66 to 51.24)	−0.52 (−3.44 to 2.41)	.73
At 12 mo	48.24 (45.96 to 50.53)	50.20 (47.98 to 52.41)	−1.95 (−4.76 to 0.85)	.17
At 24 mo	49.30 (46.97 to 51.64)	50.69 (48.40 to 52.97)	−1.38 (−4.27 to 1.51)	.35
Sensitivity Analyses for Oxford Shoulder Score				
Data derived by multiple imputation[d]				
No. of patients	125	125		
Averaged over 2 y	39.16 (37.42 to 40.81)	38.40 (36.65 to 40.08)	0.75 (−1.32 to 2.83)	.48
At 6 mo	37.96 (36.07 to 39.76)	35.67 (33.71 to 37.54)	2.28 (−0.04 to 4.61)	.05
At 12 mo	39.29 (37.48 to 41.03)	38.84 (37.03 to 40.56)	0.46 (−1.72 to 2.64)	.68
At 24 mo	40.18 (38.36 to 41.93)	40.54 (38.72 to 42.28)	−0.36 (−2.58 to 1.87)	.75
Smoking status (yes or no)[e]				
No. of patients	109	114		
Averaged over 2 y	38.65 (36.74 to 40.48)	38.05 (36.24 to 39.79)	0.60 (−1.57 to 2.77)	.59
At 6 mo	37.40 (35.34 to 39.35)	35.31 (33.26 to 37.24)	2.09 (−0.32 to 4.50)	.09
At 12 mo	38.76 (36.76 to 40.66)	38.53 (36.65 to 40.33)	0.23 (−2.06 to 2.51)	.85
At 24 mo	39.77 (37.75 to 41.69)	40.16 (38.29 to 41.96)	−0.40 (−2.72 to 1.93)	.74
Adjusted for clustering by center[f]				
No. of patients	114	117		
Averaged over 2 y	39.06 (37.18 to 40.86)	38.27 (35.81 to 40.03)	0.79 (−1.30 to 2.88)	.46
At 6 mo	37.83 (35.81 to 39.75)	35.54 (33.48 to 37.49)	2.29 (−0.03 to 4.61)	.05
At 12 mo	39.22 (37.26 to 41.08)	38.76 (36.86 to 40.57)	0.46 (−1.75 to 2.66)	.69
At 24 mo	40.10 (38.13 to 41.99)	40.36 (38.48 to 42.16)	−0.26 (−2.51 to 1.99)	.82

[a] Unless otherwise indicated.

[b] The score range was 0 to 48; higher scores indicate better outcomes. Multilevel model fixed effects: group, time (6, 12, 24 months), group × time, baseline EuroQol 5D index score, sex, age (<65 or ≥65 years), and tuberosity involvement at baseline (yes or no).

[c] Multilevel model fixed effects: group, time (6, 12, 24 months), group × time, baseline EuroQol 5D index score, sex, age (<65 or ≥65 years), and tuberosity involvement at baseline (yes or no).

[d] Missing Oxford Shoulder Score and covariate data derived by multiple

imputation. Multilevel model fixed effects: group, time (6, 12, 24 months), group × time, baseline EuroQol 5D index score, sex, age (<65 or ≥65 years), and tuberosity involvement at baseline (yes or no).

[e] Multilevel model fixed effects: group, time (6, 12, 24 months), group × time, baseline EuroQol 5D index score, sex, age (<65 or ≥65 years), tuberosity involvement at baseline (yes or no), and smoking status (yes or no).

[f] Multilevel model: fixed effects: group, time (6, 12, 24 months), group × time, baseline EuroQol 5D index score, sex, age (<65 or ≥65 years), tuberosity involvement at baseline (yes or no); and random effect: center.

There were no overall statistically significant differences between treatment groups following adjustment for smoking status (*P* = .59) or clustering by center (*P* = .46) (Table 3).

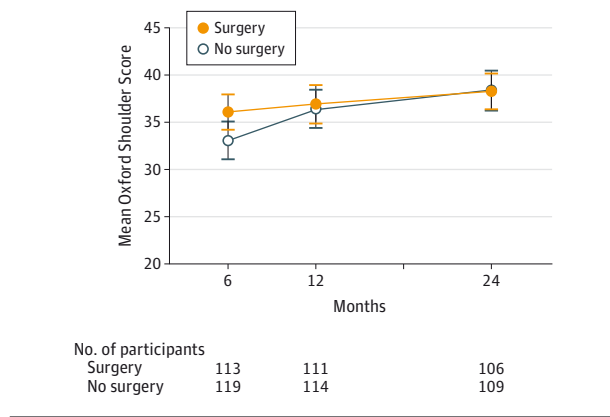For the 2 subgroup analyses, mean OSS values showed substantial overlap between subgroups at all time points (**Figure 3**),

Figure 2. Overall Comparison for Oxford Shoulder Score

No. of participants
	6	12	24
Surgery	113	111	106
No surgery	119	114	109

Error bars indicate 95% confidence intervals.

and there were no statistically significant treatment group × age interactions (*P* = .26) or treatment group × fracture type interactions assessed either by tuberosity involvement at baseline (*P* = .95) or by Neer classification (*P* = .82). Consequently, all model comparisons with the base model were not statistically significant. Similarly, the treatment group × patient preferences interaction was not statistically significant (*P* = .75). All 95% confidence intervals for the treatment effect (main, sensitivity, and subgroup analyses) excluded the prespecified difference of 5 points for the OSS representing clinical significance.

Secondary Outcomes

We found no statistically significant differences between treatment groups during the 2-year follow-up for the mean SF-12 physical component score (1.77 points higher in the surgical group [95% CI, −0.84 to 4.39]; *P* = .18) or the mean SF-12 mental component score (1.28 points lower in the nonsurgical group [95% CI, −3.80 to 1.23], *P* = .32; Table 3).

All 10 medical complications (2 cardiovascular events, 2 respiratory events, 2 gastrointestinal events, and 4 others) occurred in patients in the surgical group during the postoperative hospital stay. Slightly more patients in the surgical group

Figure 3. Comparison for Oxford Shoulder Score by Age Group and Tuberosity Involvement

Comparison by age group

Age <65 y

No. of participants
	6	12	24
Surgery	46	48	45
No surgery	53	53	50

Age ≥65 y

No. of participants
	6	12	24
Surgery	67	63	61
No surgery	66	61	59

Comparison by tuberosity involvement

No tuberosity involvement

No. of participants
	6	12	24
Surgery	24	22	21
No surgery	29	28	26

Involvement of greater and/or lesser tuberosity

No. of participants
	6	12	24
Surgery	89	89	85
No surgery	90	86	83
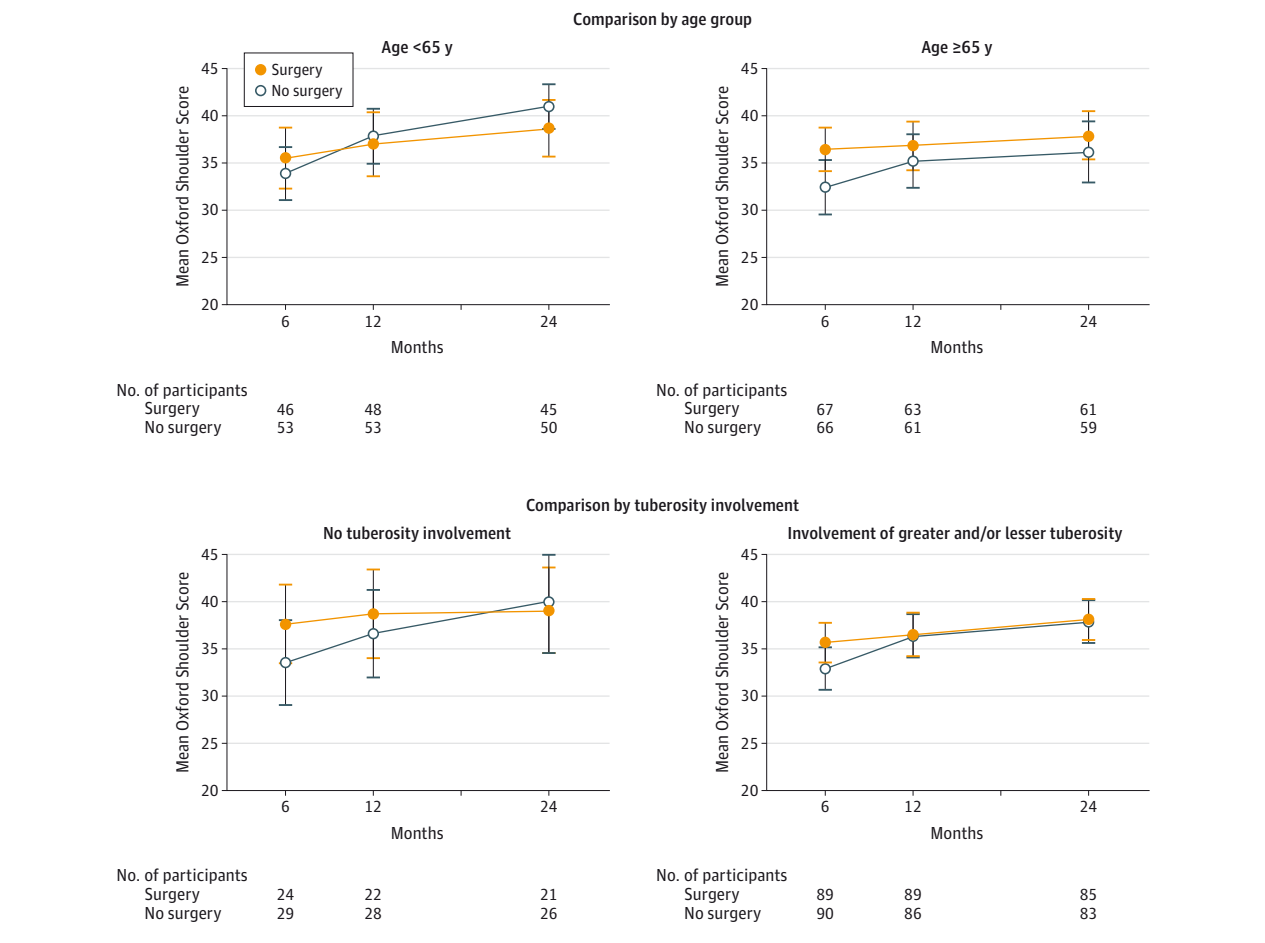
Error bars indicate 95% confidence intervals.

Table 4. Complications Related to Surgery or Shoulder Fracture and Subsequent Treatment

	No. of Complications			
	Surgical Group (n = 125)		Nonsurgical Group (n = 125)	
	Inpatient[a]	Total	Inpatient[a]	Total
Total				
No. of patients with complication	4	30[b]	0	23
No. of complications	4	36	0	23
Type of complication				
Surgical site infection	1	2	0	0
Nerve injury	1	2	0	0
Avascular necrosis	0	4	0	1
Implant-related failure[c]	0	2	0	0
Dislocation or instability	0	0	0	1
Metalwork problems[d]	2	10	0	0
Symptomatic malunion	0	4	0	5
Nonunion	0	0	0	5
Other				
Posttraumatic stiffness	0	6	0	5
Rotator cuff tear	0	3	0	1
Complex regional pain syndrome	0	1	0	0
Severe pain	0	1	0	1
Impingement	0	0	0	1
Unclear	0	1	0	3
Received secondary surgery to shoulder	0	11	0	11
Required increased or new shoulder-related therapy	1	7	0	4

[a] Defined as the end of the initial orthopedic inpatient episode.

[b] Six patients in the surgical group experienced more than 1 complication (2 in each case).

[c] Reported in relation to hemiarthroplasty.

[d] Reported in relation to internal fixation.

(30/125; 24%) experienced a complication related to shoulder fracture or its treatment than in the nonsurgical group (23/125; 18%) over the 2-year follow-up period (*P* = .28) (**Table 4**). In the surgical group, the most common complications were metalwork problems in 10 patients (complications relating to internal fixation) and posttraumatic stiffness in 6 patients. In the nonsurgical group, the most common complications were symptomatic malunion, nonunion, and posttraumatic stiffness (5 patients for each complication; Table 4). Eleven participants (9%) in each group required secondary surgery to the shoulder, whereas slightly more in the surgical group required increased or new shoulder-related therapy (7 [5.6%] vs 4 [3.2%] in the nonsurgical group; *P* = .58). There were slightly more deaths in the surgical group (9 [7.2%]) compared with 5 deaths (4.0%) in the nonsurgical group (*P* = .27). One death (due to venous thromboembolism) in the surgical group was related to the trial. Based on separate reporting of adverse events, there were 28 patients in each group who experienced at least 1 serious adverse event.

Discussion

In this randomized clinical trial of patients with displaced fracture of the proximal humerus, there were no statistically or clinically significant differences between surgical and nonsurgical treatment either overall or at individual time points (at 6, 12, and 24 months) for the OSS, which was our primary outcome. The prespecified clinically significant difference of 5 points for the OSS is appropriate both for general use[13] and for proximal humeral fractures.[19] Our results are supported by the lack of clinically or statistically significant differences in the secondary outcomes, including health-related quality of life (SF-12), complications related to surgery or shoulder fracture, complications requiring secondary surgery or treatment, and mortality.

There was no statistically significant effect of treatment group when including interactions with age or fracture type in the 2 planned subgroup analyses. Because these results do not support our prior expectations of subgroup differences, they strengthen the case for not differentiating treatment (use of surgery) on the basis of these characteristics. Similarly, there were no statistically significant differences between treatment groups after adjustment for smoking status, clustering by center, and patient treatment preference.

An examination of the potential limitations of our trial does not engender concerns that would undermine the validity or applicability of these findings. First, the slightly older ineligible population meets with the expectation that older patients, generally with greater comorbidities, are less likely to be considered for surgery. Inspection of the baseline characteristics of the trial participants and the subsequent independent characterization of x-rays confirms that the trial participants had sustained injuries that are typically considered for surgical intervention. With the exception of smoking status, for which a sensitivity analysis did not significantly affect the

trial findings, patient characteristics were balanced between the treatment groups at both baseline and 2 years. Second, the potential effect of missing data did not affect the results and was reduced in extent through the methods applied for our primary analysis.

Third, although there was some crossover between groups, the greater number of crossovers in the surgical group reflects clinical practice, in which patients can change their mind or be found unsuitable for surgery subsequent to the fracture clinic consultation.

Fourth, both interventions and associated care programs were representative of good practice; this included the majority of operations being undertaken by consultant surgeons. Additionally, locking plates and hemiarthroplasty, as used in PROFHER, are the most commonly used implants in current UK practice as well as in many other countries.[6]

Fifth, our data collection methods were robust, supporting the perception that the number of complications and reoperations for surgery-related complications were not in excess, and indeed lower than reported in the literature.[20] Although lack of blinding of patient-reported outcome assessment is unavoidable, similarities in the 2 groups in patient return of questionnaires and baseline characteristics at 24 months, and the lack of a significant effect of baseline patient preferences on the OSS results suggest this did not introduce a bias.

Sixth, the outcome data were based on patient self-report by mailed questionnaires, rather than clinical or radiological follow-up results. Even though validated, patient-reported measures of function such as the OSS should obviate the need for objective measurement of functional impairment, support for this is found in the finding of an excellent correlation between Constant scores (this widely used tool includes objective assessment of range of motion and strength) and OSS for proximal humeral fractures.[17]

The PROFHER trial has more than doubled the clinical trial evidence available for this question. The Cochrane review (search date January 2012),[8] which included 6 heterogeneous trials[21-26] comparing surgical vs nonsurgical treatment, found no significant difference in patient-reported functional scores at 1 year (standard mean difference, −0.10 [95% CI, −0.42 to 0.22], $P = .40$; data from 3 trials with a total of 153 participants). However, it found significantly more patients in the surgical group required additional or secondary surgery (18/112 vs 5/111; relative risk, 3.36 [95% CI, 1.33 to 8.49], $P = .01$; data from 5 trials). It is noteworthy that the PROFHER trial did not find a difference between groups for this outcome.

Conclusions

Among patients with displaced proximal humeral fractures involving the surgical neck, there was no significant difference between surgical treatment compared with nonsurgical treatment in patient-reported clinical outcomes over 2 years following fracture occurrence. These results do not support the trend of increased surgery for patients with displaced fractures of the proximal humerus.

REFERENCES

1. Court-Brown CM, Caesar B. Epidemiology of adult fractures: a review. *Injury*. 2006;37(8):691-697.

2. Johnell O, Kanis JA. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. *Osteoporos Int*. 2006; 17(12):1726-1733.

3. Kim SH, Szabo RM, Marder RA. Epidemiology of humerus fractures in the United States: nationwide emergency department sample, 2008. *Arthritis Care Res (Hoboken)*. 2012;64(3):407-414.

4. Palvanen M, Kannus P, Niemi S, Parkkari J. Update in the epidemiology of proximal humeral fractures. *Clin Orthop Relat Res*. 2006;442(442): 87-92.

5. Court-Brown CM, Garg A, McQueen MM. The epidemiology of proximal humeral fractures. *Acta Orthop Scand*. 2001;72(4):365-371.

6. Bell JE, Leung BC, Spratt KF, et al. Trends and variation in incidence, surgical treatment, and repeat surgery of proximal humeral fractures in the elderly. *J Bone Joint Surg Am*. 2011;93(2):121-131.

7. Polinder S, Iordens GI, Panneman MJ, et al. Trends in incidence and costs of injuries to the shoulder, arm and wrist in the Netherlands between 1986 and 2008. *BMC Public Health*. 2013;13(531):531.

8. Handoll HHG, Ollivere BJ, Rollins KE. Interventions for treating proximal humeral fractures in adults. *Cochrane Database Syst Rev*. 2012;12:CD000434.

9. Handoll H, Brealey S, Rangan A, et al. Protocol for the ProFHER (PROximal Fracture of the Humerus: evaluation by Randomisation) trial: a pragmatic multi-centre randomised controlled trial of surgical versus non-surgical treatment for proximal fracture of the humerus in adults. *BMC Musculoskelet Disord*. 2009;10(140):140.

10. Neer CS II. Displaced proximal humeral fractures, I: classification and evaluation. *J Bone Joint Surg Am*. 1970;52(6):1077-1089.

11. Neer CSII II. Four-segment classification of proximal humeral fractures: purpose and reliable use. *J Shoulder Elbow Surg*. 2002;11(4):389-400.

12. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br*. 1996;78(4):593-600.

13. Dawson J, Rogers K, Fitzpatrick R, Carr A. The Oxford Shoulder Score revisited. *Arch Orthop Trauma Surg*. 2009;129(1):119-123.

14. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220-233.

15. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53-72.

16. 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(1):7-16.

17. Baker P, Nanda R, Goodchild L, Finn P, Rangan A. A comparison of the Constant and Oxford Shoulder Scores in patients with conservatively treated proximal humeral fractures. *J Shoulder Elbow Surg*. 2008;17(1):37-41.

18. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? updating criteria to evaluate the credibility of subgroup analyses. *BMJ*. 2010;340:c117.

19. van de Water AT, Shields N, Davidson M, Evans M, Taylor NF. Reliability and validity of shoulder function outcome measures in people with a proximal humeral fracture. *Disabil Rehabil*. 2014;36 (13):1072-1079.

20. Thanasas C, Kontakis G, Angoules A, Limb D, Giannoudis P. Treatment of proximal humerus fractures with locking plates: a systematic review. *J Shoulder Elbow Surg*. 2009;18(6):837-844.

21. Kristiansen B, Kofoed H. Transcutaneous reduction and external fixation of displaced fractures of the proximal humerus: a controlled clinical trial. *J Bone Joint Surg Br*. 1988;70(5):821-824.

22. Stableforth PG. Four-part fractures of the neck of the humerus. *J Bone Joint Surg Br*. 1984;66(1): 104-108.

23. Zyto K, Ahrengart L, Sperber A, Törnkvist H. Treatment of displaced proximal humeral fractures in elderly patients. *J Bone Joint Surg Br*. 1997;79(3): 412-417.

24. Fjalestad T, Hole MO, Jørgensen JJ, Strømsøe K, Kristiansen IS. Health and cost consequences of surgical versus conservative treatment for a comminuted proximal humeral fracture in elderly patients. *Injury*. 2010;41(6):599-605.

25. Olerud P, Ahrengart L, Ponzer S, Saving J, Tidermark J. Internal fixation versus nonoperative treatment of displaced 3-part proximal humeral fractures in elderly patients: a randomized controlled trial. *J Shoulder Elbow Surg*. 2011;20(5): 747-755.

26. Olerud P, Ahrengart L, Ponzer S, Saving J, Tidermark J. Hemiarthroplasty versus nonoperative treatment of displaced 4-part proximal humeral fractures in elderly patients: a randomized controlled trial. *J Shoulder Elbow Surg*. 2011;20(7): 1025-1033.