BJR

SPECIALTY UPDATE - UPPER LIMB

Trials based on specific fracture configuration and surgical procedures likely to be more relevant for decision making in the management of fractures of the proximal humerus

FINDINGS OF A META-ANALYSIS

S. Sabharwal, N. K. Patel, D. Griffiths, T. Athanasiou, C. M. Gupte, P. Reilly

Imperial College London, London, United Kingdom

Objectives

The objective of this study was to perform a meta-analysis of all randomised controlled trials (RCTs) comparing surgical and non-surgical management of fractures of the proximal humerus, and to determine whether further analyses based on complexity of fracture, or the type of surgical intervention, produced disparate findings on patient outcomes.

Methods

A systematic review of the literature was performed identifying all RCTs that compared surgical and non-surgical management of fractures of the proximal humerus. Meta-analysis of clinical outcomes was performed where possible. Subgroup analysis based on the type of fracture, and a sensitivity analysis based on the type of surgical intervention, were also performed.

Results

Seven studies including 528 patients were included. The overall meta-analysis found that there was no difference in clinical outcomes. However, subgroup and sensitivity analyses found improved patient outcomes for more complex fractures managed surgically. Four-part fractures that underwent surgery had improved long-term health utility scores (mean difference, MD 95% CI 0.04 to 0.28; p = 0.007). They were also less likely to result in osteoarthritis, osteonecrosis and non/malunion (OR 7.38, 95% CI 1.97 to 27.60; p = 0.003). Another significant subgroup finding was that secondary surgery was more common for patients that underwent internal fixation compared with conservative management within the studies with predominantly three-part fractures (OR 0.15, 95% CI 0.04 to 0.63; p = 0.009).

Conclusion

This meta-analysis has demonstrated that differences in the type of fracture and surgical treatment result in outcomes that are distinct from those generated from analysis of all types of fracture and surgical treatments grouped together. This has important implications for clinical decision making and should highlight the need for future trials to adopt more specific inclusion criteria.

Cite this article: ***Bone Joint Res*** **2016;5:470–480.**

Keywords: Proximal humerus fracture; Meta-analysis

- S. Sabharwal, MBBS MRCS MSc BSc, Orthopaedic Registrar,
- D. Griffiths, FRCS (Tr&Orth), Consultant Orthopaedic Surgeon,
- C. M. Gupte, FRCS (Tr&Orth), PhD, Consultant Orthopaedic Surgeon,
- P. Reilly, FRCS (Tr&Orth), MS, Consultant Orthopaedic Surgeon, Department of Trauma and Orthopaedics, St Mary's Hospital, Ground Floor Salton House, South Wharf Road, London, W2 1NY, UK.
- N. K. Patel, FRCS (Tr&Orth), MSc, Orthopaedic Trauma Fellow, Virginia Commonwealth University Medical Centre, Richmond, Virginia, USA.
- T. Athanasiou, FRCS, PhD, MBA, Professor of Cardiac Surgery, Imperial College London, 1022, Queen Elizabeth the Queen Mother Wing (QEQM), St Mary's Campus, London, UK.

Correspondence should be sent to Mr S. Sabharwal; email: sanjeeve.sabharwal@ic.ac.uk

doi:10.1302/2046-3758.510.2000638

Bone Joint Res 2016;5:470–480.
Received: 16 November 2015; Accepted: 8 August 2016

Article focus

- To provide an up-to-date summary of outcomes of RCTs comparing surgical and non-surgical treatment of displaced fractures of the proximal humerus.
- To establish whether increasing fracture complexity or type of surgical intervention produce treatment effects that are distinct from those of the overall meta-analysis.

Key messages

- Differences in the type of fracture and surgical treatment result in outcomes that are distinct from those generated from analysis of all types of fracture and surgical treatments grouped together.

- Further randomised trials in this field should include more specific inclusion criteria to increase their clinical application.

Strengths and limitations

- This is the first meta-analysis of RCTs comparing surgical and non-surgical management of fractures of the proximal humerus that has used subgroup analysis based on complexity of fracture to compare patient outcomes.
- There is a need for further research in this field, which is highlighted by the weakness of estimated effects based on the Grading of Recommendations Assessment, Development, and Evaluation assessment of the included studies.

Introduction

Fractures of the proximal humerus account for approximately 10% of all fractures in patients over the age of 65.[1] Epidemiological research indicates that the incidence of this injury is increasing rapidly, primarily as a result of the growth of an ageing population.[2,3] For displaced fractures which represent approximately 50% of these fractures,[4] there are a myriad of classification systems to describe the different and complex fracture configurations that occur.[5] Interestingly, there is a trend towards increasing surgical management of these fractures[6,7] despite a lack of clinical evidence to support surgical intervention over non-operative management.[8] Over the course of the last five years, there has been emerging Level I evidence[9] within the field. The most recently published clinical study to investigate surgical *versus* non-surgical treatment of displaced fractures of the proximal humerus is the Proximal Fracture of the Humerus Evaluation by Randomization (PROFHER), a multicentre randomised controlled trial (RCT).[10] In an approach similar to previously published RCTs that investigated treatment interventions for fractures of the proximal humerus,[11,12] different types of fracture, as classified by the Neer Classification,[13] were included within their study population.

When clinical heterogeneity between RCTs is associated with statistical heterogeneity amongst their treatment effects, decision making on how to apply research findings to clinical practice may be affected by the relevant patient- or disease-related factors.[14] In the case of a meta-analysis, pooling data from such trials may also be questionable.[14] Identifying the clinical characteristics that underpin statistical differences in treatment effects provides an opportunity for the design of future RCTs with less heterogeneity within their included population, and therefore may potentially have less "noise", that is, problematic factors, affecting the desired objective of identifying a true treatment effect.[15] This is particularly relevant to the management of fractures of the proximal humerus, where the variability in morphology, as well as the different surgical treatments available, may be a reason for the lack of clinical implementation of findings from pragmatically-designed RCTs or meta-analyses, which report that there is no difference in clinical outcomes between surgical and non-surgical management.

The aim of this meta-analysis is to provide an up-to-date summary of outcomes of RCTs comparing surgical and non-surgical treatment of displaced fractures of the proximal humerus, and specifically to establish whether increasing fracture complexity or type of surgical intervention produce treatment effects that are distinct from those of the overall meta-analysis.

Patients and Methods

Review protocol. This study was performed in accordance with the guidelines from the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement.[16]

Search strategy. A systematic search was performed in PubMed, EMBASE and Cochrane databases (January 1 1975 to May 1 2015) using the search term "proximal humerus fracture". The search query used in PubMed was 'Search (proximal humerus fracture) AND ("1975/01/01"[Date - Publication] : "2015/05/01"[Date - Publication])'. The search query used in EMBASE was 'proximal humerus fracture.mp.' and '1975:2015.(sa_year)'. The search query used in the Cochrane Library was 'proximal humerus fracture:ti,ab,kw Publication Year from 1975 to 2015'. All results were combined in an Endnote library and the 'find duplicates' application was used to identify duplicate studies. Studies were selected for further analysis after reading the individual abstracts.

Inclusion and exclusion criteria. All randomised controlled trials that compared surgical and non-surgical management of fractures of the proximal humerus were included. Study inclusion was not restricted by the type of surgical repair performed, or the types of fracture included. Studies that described conservative treatment that involved a closed reduction under general anaesthetic were excluded. Only English language studies were included within the analysis and studies that did not include patient follow-up for at least one year were also excluded. Two reviewers (SS, NKP) independently selected the relevant articles and reviewed their full texts to assess their eligibility. Discrepancies were resolved following assessment by a third reviewer (PR). Inter-rater agreement was calculated using a Kappa statistical analysis.

Outcomes of interest and data extraction. The following information was extracted from each study: authorship, year of publication, sample size, single/multicentre status, mean population age, fracture classification system used, types of fracture included, surgical procedure performed, non-operative protocol, rehabilitation protocol, duration of follow-up and patient outcomes.

Four outcomes of interest were extracted to compare surgical treatment with conservative management:

– Functional outcome based on the Constant Score;
– Quality of life assessment based on the EuroQol or 15-D instrument;
– Additional surgery (secondary surgery following a complication of primary treatment);
– Adverse events or complications: avascular necrosis (AVN), post-traumatic osteoarthritis (OA) and non/malunion.

Risk of bias in individual studies and grading of the evidence. The Cochrane Collaboration's tool for assessing risk of bias in RCTs[17] was used by two reviewers who independently assessed each domain of every study. A third reviewer was used to resolve potential discrepancies (PR). Sensitivity analysis was performed to investigate whether risk of bias within the studies affected the results of the meta-analysis for the pre-determined outcomes of interest. The confidence in effect estimates was evaluated using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach.[18]

Statistical analysis. The meta-analysis was conducted in line with recommendations from the Cochrane Collaboration and PRISMA statement.[16] Analysis was performed in Review Manager Version 5.3 (The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen). The summary statistics for dichotomous data (secondary surgery and adverse events) were odds ratio (OR), whereas the effect measure for continuous data (Constant Score and quality of life assessment) was mean difference (MD), both reported with 95% confidence intervals (CI). The point estimate of the MD was considered statistically significant at $p < 0.05$ if the 95% CI did not include the value 0. An odds ratio was considered statistically significant at $p < 0.05$ if the 95% CI did not include the value 1. Data were analysed using a random effects Mantel-Haenszel model. A random effects analysis was chosen over a fixed effects model for subgroup analysis because the populations included in the different studies were heterogeneous.

Publication bias was assessed by graphical exploration with funnel plots. Heterogeneity was assessed using the I^2 statistic. This represents the proportion of total variation between included studies attributable to differences between studies, rather than sampling error (chance). The degree of heterogeneity was graded as low ($I^2 < 25\%$), moderate ($I^2 = 25\%$ to 75%) or high ($I^2 > 75\%$).

Three strategies were used to investigate potential causes of statistical heterogeneity within the included studies: subgroup analysis of RCTs based on the predominant type of fracture within the study (> 50% have two parts, > 50% have three parts and > 50% have four parts); sensitivity analysis by meta-analysis of studies based on their risk of bias; and sensitivity analysis by meta-analysis of studies based on whether arthroplasty or internal fixation was performed for more than 80% of included cases. For the sensitivity analysis, both fixed and random effects models were adopted and outcome measures were reported for each. This was undertaken because in a sensitivity analysis, presentation of both models provides a more comprehensive evaluation of how differences in datasets or statistical methods influence the observed outcomes.[19] Meta-regression was not performed due to the limited number of RCTs identified.

Results

Seven studies[10-12,20-23] published between 1975 and 2015 met the inclusion criteria for comparing surgical and non-surgical treatment of displaced fractures of the proximal humerus (Fig. 1). The strength of agreement on the included studies between the reviewers was strong with a Kappa value of 0.875. These studies included a total of 528 patients; 50% underwent surgical treatment and 50% were managed conservatively. The results from one study group investigating the same group of patients were published as three separate studies that reported on different outcomes, as well as outcomes at different follow-up periods.[11,24,25] For the purpose of the analysis these are reported as a single study. Six studies adopted the Neer classification system, with the one study that used the AO/OTA classification system, stating that only three or four part fractures were included. Amongst the seven studies, there were three studies that included only four-part fractures, two studies that included three- and four-part fractures, one study that included only three-part fractures, and one study that included all types of fracture (Table I).

Assessment of bias. The Cochrane Collaboration's tool for assessment of bias was used to assess individual domains and overall study bias.[26] Five studies were rated as 'unclear risk of bias,[10,20-22,25] two were rated as 'high risk' of bias[12,23] (Fig. 2). Studies that were rated as unclear risk all had clearly defined methods of randomisation, however, across all seven studies, there was a lack of blinding of participants and clinicians. These were rated as 'unclear risk' instead of 'high risk' because of a perceived impracticality of blinding patients and clinicians when comparing surgical and non-operative treatment. Selective reporting was difficult to assess across most of the trials. However, one study had a published protocol,[10] in which the stated outcomes of interest matched those within the final publication. This was judged to be at a 'low risk' of bias. None of the studies which had an 'unclear risk' of bias had a single domain rated as 'high risk'. The evaluation

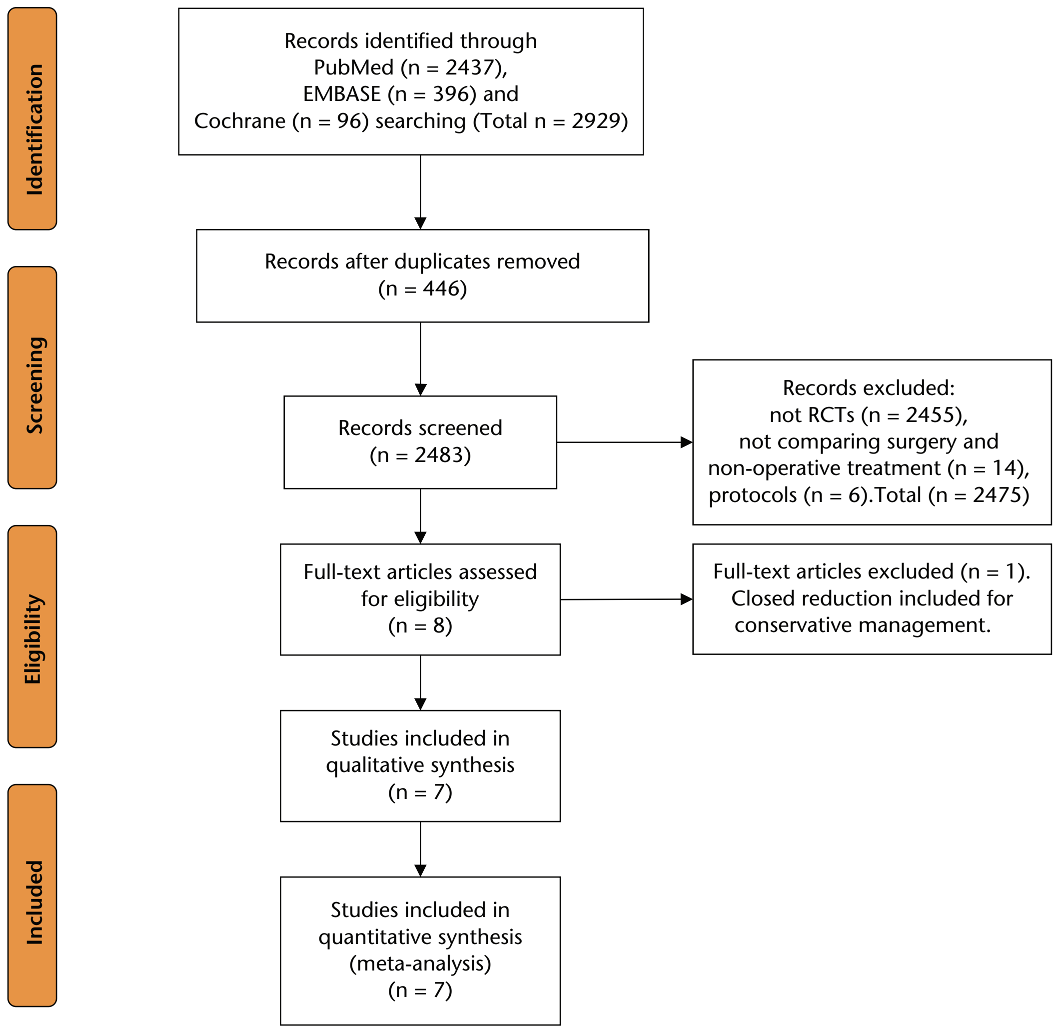

Fig. 1

PRISMA flow diagram showing results of article screening and exclusion

Table I. Characteristics of studies included in the meta-analysis

Author (yr)	Country	Single/ multicentre	Sample size	Mean age surgical group	Mean age conservative group	Surgical procedure (%)	Type of fracture (Neer classification) (%)
Rangan et al (2015)[10]	United Kingdom	Multicentre	250	66.6	65.4	Plate fixation (*82.6*) Hemiarthroplasty (*9.2*) Other (*8.2*)	2 part (*51.2*)/3 part (*37.2*)/4 part (*4.4%*) 1 part (*7.2*)
Fjalestad et al (2012) *[11,24,25]	Norway	Single centre	50	72.2	73.1	Plate fixation	3 part (*52*)/4 part (*48*)
Zyto et al (1997)[12]	Sweden	Single centre	40	73	75	Tension band fixation	3 part (*92.5*)/4 part (*7.5*)
Boons et al (2012)[20]	The Netherlands	Single centre	50	76.4	79.9	Hemiarthroplasty	4 part
Olreud et al (2011) (H)[22]	Sweden	Single centre	55	75.8	77.5	Hemiarthroplasty	4 part
Olreud et al (2011) (F)[21]	Sweden	Single centre	60	72.9	74.9	Plate fixation	3 part
Stableforth (1984)[23]	United Kingdom	Single centre	32	65.6	70.1	Hemiarthroplasty	4 part

of confidence in effect estimates using the GRADE tool is summarised in Table II.

Funnel plots (Fig. 3) based on re-operation and adverse events as outcomes of interest were created to evaluate publication bias. For both evaluations, all the studies were found to lie within a 95% CI as represented by the inverted funnel, suggesting an absence of publication bias.

Outcomes of interest: Constant Score. Four studies[11,20-22] reported a Constant Score at 12-month follow-up, of which there were two subgroups; one that included two studies with > 50% three-part fractures and another that included two studies with > 50% four-part fractures. Meta-analysis of these studies showed comparable functional outcomes at 12-month follow-up based on the

Table II. Grading of Recommendations Assessment, Development and Evaluation summary for meta-analysis of outcomes of interest comparing operative and non-operative management of fractures of the proximal humerus

Outcome of interest	Participant studies (n)	Relative effect (95% CI)	Initial score	Quality score	Consistency score	Directness	Effect size	Overall grade
Constant Score (at 24 mths)	173 (4)	MD -1.24(-7.07 to -4.59)	4	−3	0	0	0	Very low
Health utility (at 24 months)	365 (4)	MD 0.06 (-0.04 to 0.16)	4	−1	−1	0	0	Low
Secondary surgery	534 (7)	OR 0.53 (0.26 to 1.08)	4	−3	−1	0	1	Very low
Adverse outcomes	534 (7)	OR 1.95 (0.95 to 4.00)	4	−3	−1	0	0	Very low

CI, confidence interval; MD, mean difference; OR, odds ratio

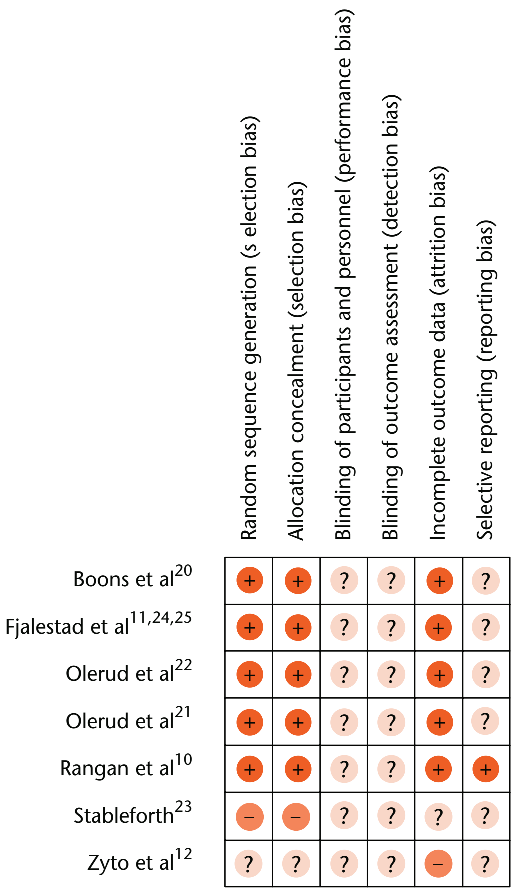

Fig. 2

Cochrane risk of bias summary for seven included studies
(+ represents low risk, ? represents unclear risk, - represents high risk).

Constant Score (MD 2.82, 95% CI 2.17 to 7.82; p = 0.27), with no overall heterogeneity or heterogeneity between the subgroups.

Three studies[21,22,25] reported a Constant Score at 24-month follow-up, of which there were two subgroups; one that included two studies with > 50% three-part fractures and another with one study with > 50% four-part fractures. Meta-analysis of these studies showed comparable functional outcomes at 24-month follow-up based on the Constant Score (MD -0.20, 95% CI - 6.68 to 6.39; p = 0.95), with no overall heterogeneity (I^2 = 0%) or heterogeneity between the subgroups (I^2 = 0%). One study with > 50% three-part fractures reported a Constant Score at 50-month follow up and when included in the meta-analysis of 24-month follow up studies there was no statistically significant difference in outcome (MD -1.24, 95% CI -7.07 to 4.59; p = 0.68).

Outcomes of interest: quality of life assessment. Four studies[10,21,22,25] included quality of life outcomes or health utilities at 12-month follow up. Meta-analysis of these studies revealed that overall there were comparable health utilities at 12-month follow-up for surgically and conservatively managed patients (MD 0.02, 95% CI 0.02 to 0.06; p = 0.38), with low overall heterogeneity (I^2 = 6%) and low heterogeneity between subgroups (I^2 = 16.4%). These studies also reported health utilities at 24-month follow-up and there was also no significant difference between patients treated surgically or conservatively (MD 0.01, 95% CI 0.05 to 0.08; p = 0.69), with moderate overall heterogeneity (I^2 = 55%) and moderate heterogeneity between subgroups (I^2 = 68.2%). There was an observable trend towards favourable outcomes for patients treated surgically as complexity of fracture within the studies increased and the single study that included four-part fractures had a significantly improved health utility score for patients managed surgically at 24-month follow-up (Fig. 4).

Outcomes of interest: additional surgery. All seven studies reported on additional surgery following a failure or complication of the primary treatment. Overall, meta-analysis of all the studies (Fig. 5) showed no statistically significant difference between patients who were initially managed surgically compared with those managed conservatively (OR 1.88, 95% CI 0.92 to 3.85, p = 0.08). Although there was minor heterogeneity (I^2 = 4%) between the seven studies, subgroup analysis between studies with > 50% two-part fractures, > 50% three-part fractures and > 50% four-part fractures revealed moderate heterogeneity (I^2 = 60%). Furthermore, subgroup analysis of the three studies with > 50% three-part fractures produced a statistically significant OR favouring conservative management (OR 6.47, 95% CI 1.59 to 26.38; p = 0.009).

Outcomes of interest: adverse events or complications. All seven studies reported complications or adverse events occurring as a result of surgical or conservative treatment. Overall, meta-analysis of the studies (Fig. 6) showed no statistically significant difference in the incidence of AVN, post-traumatic OA and mal/nonunion between the two groups (OR 0.51, 95% CI 0.25 to 1.05; p = 0.07). There

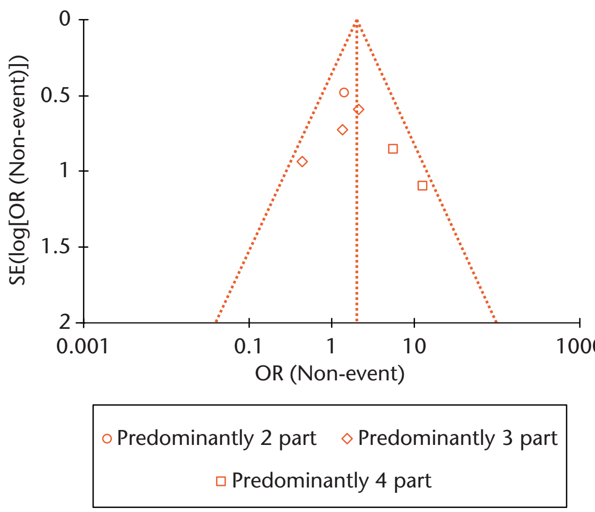

Fig. 3a

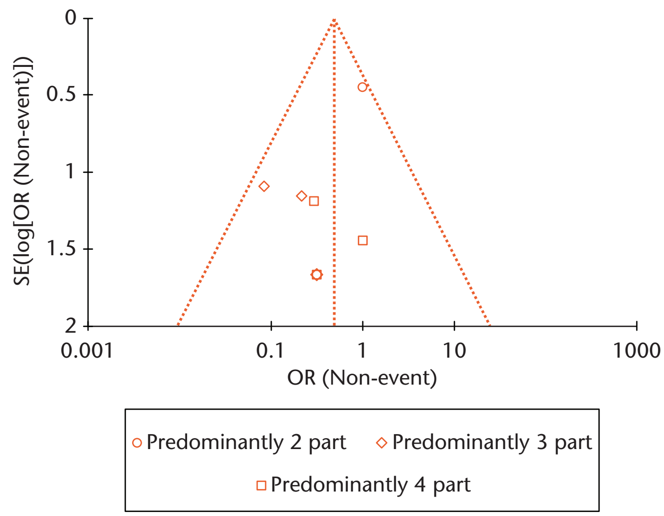

Fig. 3b

Funnel plots demonstrating low risk of publication bias based on a) adverse events as an outcome measure and b) re-operation as an outcome measure.

was moderate overall heterogeneity (I^2 = 34%), which was increased when heterogeneity between subgroups was assessed (I^2 = 60.80%). Furthermore, only a subgroup analysis of the three studies with > 50% four-part fractures demonstrated a statistically significant difference in treatment outcomes with less adverse events likely in patients undergoing surgical treatment (OR 0.14, 95% CI 0.04 to 0.51; p = 0.03).

Sensitivity analysis. This was performed on studies based on their risk of bias in order to determine whether methodological heterogeneity influenced statistical heterogeneity of outcome measures, and also to evaluate the effect of type of surgical treatment on outcomes. Studies at 'low risk of bias' were subjected to meta-analysis, as were the two studies where the risk of bias was 'unclear'/'high risk'.

When sensitivity analysis was performed on studies based on the predominant surgical procedure, a significant treatment effect in quality of life assessment at 24 months was found for the single study that performed hemiarthroplasty (MD 0.16, 95% CI 0.04 to 0.28; p = 0.007). Hemiarthroplasty studies were also less likely to result in the pre-defined adverse events (OR 0.14, 95% CI 0.04 to 0.51; p = 0.03). Hemiarthroplasty studies included only four-part fractures. Internal fixation studies included all fracture types. There was moderate heterogeneity within the internal fixation studies when evaluating re-operation as an outcome of interest (I^2 = 47%). There was no heterogeneity found in any of the meta-analysis performed for hemiarthroplasty studies.

Discussion

The findings of our study update a previous meta-analysis of RCTs comparing surgical and conservative management of displaced fractures of the proximal humerus[8] with the inclusion of two more recent randomised trials[10,20] and outcomes reported at 24 months from a third study.[25] Although recently published meta-analyses have integrated more recent research into their analysis,[27-29] to the best of our knowledge this is the only study that has compared treatment interventions in RCTs based on the complexity of the fracture and the type of surgical procedure performed. The results of the subgroup and sensitivity analyses based on these clinical characteristics offer outcomes based on more specific clinical scenarios. Furthermore, the differences identified between the overall analysis and subgroups underlines the concern that 'comparing apples and oranges' in medical research has its drawbacks.[30]

Subgroup analysis of health utilities based on the predominant types of fracture within each study demonstrated a trend towards improved outcomes as the complexity of the fracture increased. The single study that included patients with four-part fractures had significantly improved outcomes for patients managed surgically. Studies where the main type of fracture was of a three-part nature had a significantly increased risk of additional surgery for patients who underwent surgical intervention as a primary treatment. Although not statistically significant, there was also an observable difference with an increased risk for additional surgery after surgical management of four-part fractures, compared with two-part fractures. Moreover, within the subgroups separating trials by complexity of fracture, there was no statistical heterogeneity. However, there was moderate statistical heterogeneity between those subgroups (I^2 = 60%). This heterogeneity was not eradicated, or reduced from moderate to low, when sensitivity analysis was performed based on risk of bias. This suggests that differences in methodology were not underpinning statistical

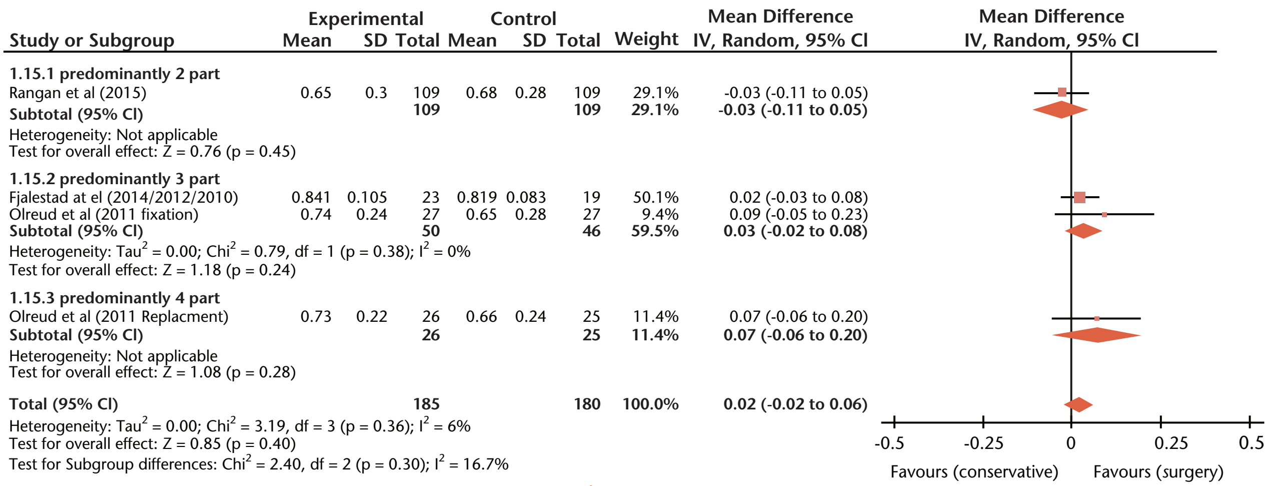

Fig. 4a

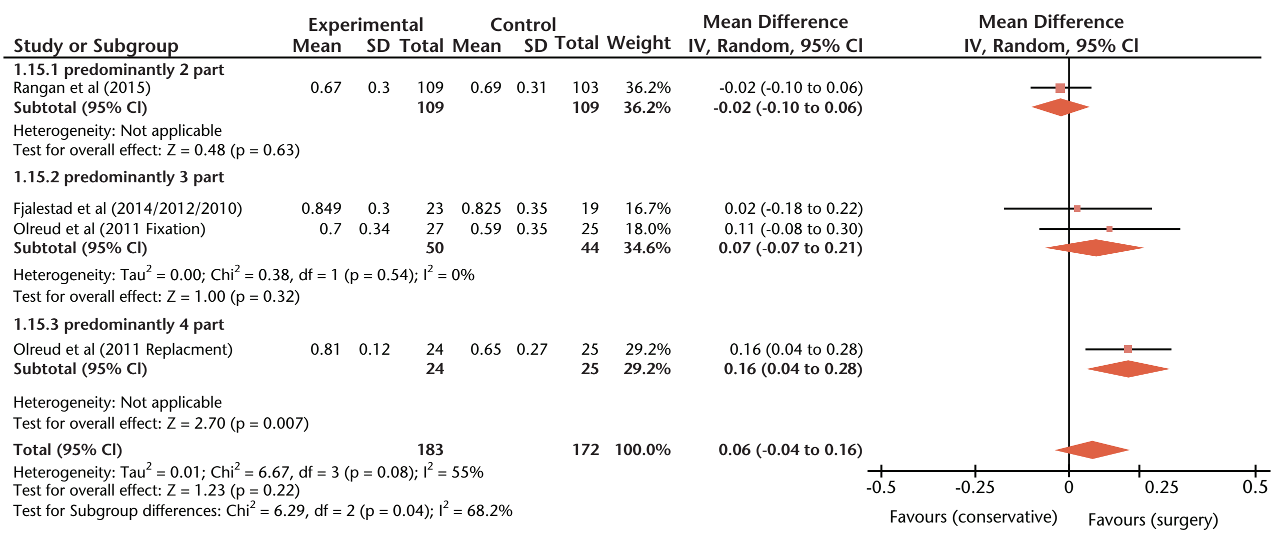

Fig. 4b

Forest plots demonstrating health utilities at 12-month a) and 24-month follow-up b) based on predominant fracture type in randomised controlled trials comparing surgery and conservative management for displaced fractures of the proximal humerus (CI, confidence interval; SD, standard deviation).

heterogeneity of treatment effects when considering additional surgery as an outcome of interest. The finding of the subgroup analysis suggest what is intuitive in clinical practice, that is, more complex fractures are more difficult to manage surgically, hence they are likely to require additional surgery to manage failure or complications following the primary surgical procedure. The statistical significance of a higher rate of re-operation after surgical management in studies with > 50% three-part fractures is likely to be confounded by this point. However, it is important to recognise that surgical management within these studies involved the use of locking plates only. There is an evidence-based view that complications after locking plates are more likely than after hemiarthroplasty in elderly patients.[31] However, more evidence is needed to understand whether the difference in rate of re-operation is also influenced by differences in decision making as surgeons weigh the risks and benefits of revising both implants.

When evaluating the complications occurring after surgical and conservative treatment, we selected adverse events that were plausible with either intervention. Post-traumatic OA, AVN and mal/nonunion are significant indicators of future disability, and have been repeatedly used as markers to compare the incidence of adverse events between the two intervention groups.[5,32] Although our overall analysis suggests there is no statistical difference between surgical and non-operative treatment in relation to these events, subgroup analysis of four-part fractures results in a significant OR that favours surgery (OR 0.14, 95% CI 0.04 to 0.51; p = 0.03) and there was moderate heterogeneity between the subgroups

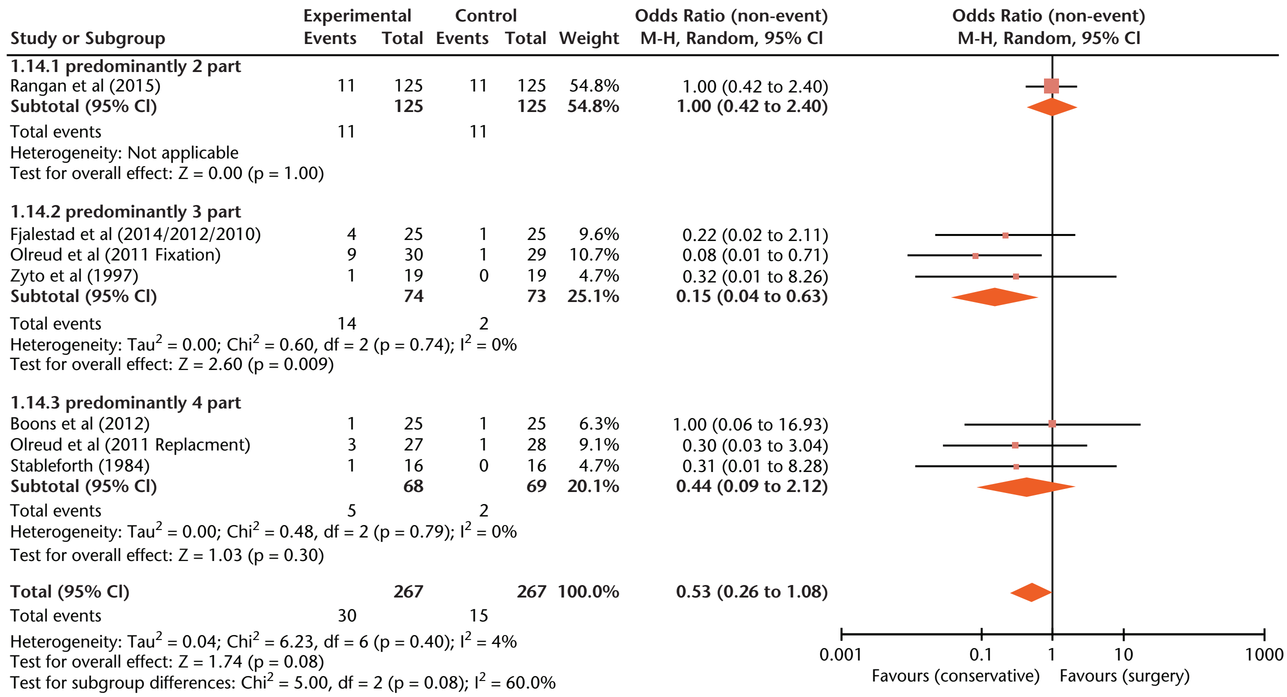

Fig. 5

Forest plot demonstrating secondary surgery following either initial surgical or non-surgical treatment with subgroup analysis based on predominant fracture type, (CI, confidence interval).

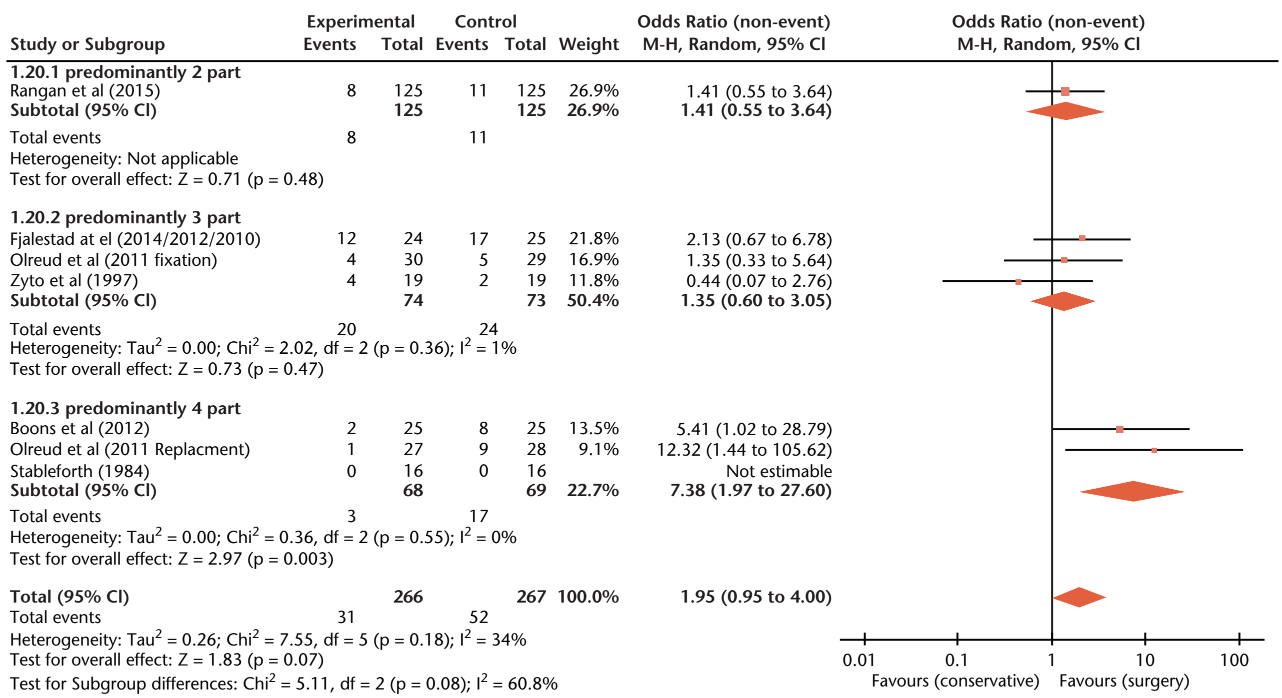

Fig. 6

Forest plot demonstrating adverse events following surgical or conservative management with subgroup analysis based on the predominant type of fracture, (CI, confidence interval).

(I^2 = 60.8%). This result is likely to be entirely confounded by the fact that four-part fracture studies adopted hemi-arthroplasty as the surgical intervention, and this is demonstrated in the sensitivity analysis of this outcome of interest based on the type of surgery performed.

The implications of this meta-analysis for future RCTs investigating the treatment of fractures of the proximal humerus are clear: homogeny of fracture complexity and surgical procedure performed are likely to provide clearer and more clinically applicable findings that inform

Table III. Trials currently registered in the WHO International Clinical Trials Registry Portal investigating surgical *versus* conservative treatment of fractures of the proximal humerus

Study name (ID)	Country	Single centre or multicentre	Sample size	Fracture type included	Interventions
Proximal Humerus Fracture Randomised Control Trial *(NCT02362100)*	Canada	Unknown	155	Neer types: 2,3 &4	Locking plate *versus* conservative
Treatment of Proximal Humeral Fractures (TPHF) *(NCT01246167)*	Finland	Multicentre	290	Neer type: 2,3 &4 (low energy injuries with other specifications described)	Hemiarthroplasty *versus* plate *versus* conservative
Treatment of Comminuted Fractures of the Proximal Humerus *(NCT00999193)*	Finland	Unknown	90	Neer type: 3,4 part with > 5 mm dislocation of anatomical neck AO C1 to 2 for non-luxation fractures and C3 for luxation fractures	Hemiarthroplasty *versus* plate *versus* conservative
Primary hemiarthroplasty versus conservative treatment for comminuted fractures of the proximal humerus in the elderly (ProCon) *(NTR2040)*	Netherlands	Multicentre	80	Selected Hertel types: 9 to 12, selected Hertel types 2, or head split fractures based on the judgement of the attending surgeon.	Hemiarthroplasty *versus* conservative
Comminuted Proximal Humerus Fractures. A Randomised study of Surgical Versus Conservative Treatment *(NCT00863473)*	Norway	Unknown	50	AO B2 or C2 (with further specification on selected types)	Locking plate *versus* conservative
Effect of Osteosynthesis. Primary Hemiarthroplasty, and Non-surgical Management for Fractures of the Proximal Humerus *(NCT00835562)*	Denmark	Multicentre	162	Neer type 4	Hemiarthroplasty *versus* plate *versus* conservative
Operative Versus Non Operative Treatment of Proximal Humerus (Shoulder Joint) Fractures *(NCT0818987)*	Canada	Multicentre	120	Neer types 3 &4	Internal fixation *versus* conservative

WHO, World Health Organization

decision making in this field. There is a potential for misguided interpretation of the results of large well-performed trials, with a view that outcomes for surgery and conservative management are identical for all displaced fractures of the proximal humerus.[33] The diverse range of fracture patterns that numerous systems have attempted to classify are known to predict clinical outcome variably,[34] and therefore, the influence of type of fracture on clinical outcome when hypothesising whether surgical and conservative treatments are comparable, is important. At the time the last search was performed within our systematic review, there were seven ongoing RCTs that included 947 patients, registered in the World Health Organization's International Clinical Trials Registry Platform, that were investigating the difference between surgical and conservative management for fractures of the proximal humerus (Table III).[35] Of the five studies that have used the Neer classification system, two include two-, three- and four-part fractures, two include three- and four-part fractures and only one includes four-part fractures. The other ongoing trials are investigating complex and displaced fractures based on the AO and Hertel's classification system. While it has often been suggested that the reliability of the current classification systems are poor,[5] regardless of the system that is adopted, there is a need for inclusion of more homogenous fracture patterns within future trials. Appropriate study design has a large role to play in the feasibility of such research. Given that PROFHER was the only study out of the seven within our meta-analysis that was performed at multiple centres, and it recruited almost four times as many patients as the next largest study within the shortest period of time, there is a clear benefit in collaboration between research institutions in future RCTs. This will allow appropriate sample size recruitment over a shorter period of time to be conducted for specific types of fracture, which represents smaller cohorts within the general population. Furthermore, conducting multicentre RCTs such as PROFHER may increase the external validity of research findings,[36] which is particularly important in orthopaedic surgery where technical expertise and experience is variable and may be a source of performance bias.[37]

Functional outcome measures reported within the seven RCTs evaluating surgical *versus* conservative treatment for fractures of the proximal humerus were highly variable. Although the majority of these functional assessment measures are valid,[38,39] interpretation and comparison of results between studies that have adopted different functional outcome measures is challenging. In hip fracture research, key opinion leaders have proposed core outcome measures for clinical trials in order to improve synthesis of evidence within their field.[40] Consensus and evidence-based guidance on core outcome measures for future trials that have been developed by recognised professional bodies in upper limb orthopaedics could improve the reporting of outcomes, as well as the ability to interpret the overall evidence base.

Our study has five main limitations. Firstly, the number of studies included was less than ten. As such, our evaluation of publication bias should be treated with caution as funnel plot asymmetry is most accurately performed when there are at least ten studies.[41] Furthermore, because the number of trials included was small, the sample size of certain subgroup analyses was also small, which probably resulted in large CIs that are seen throughout the analysis. This suggests less than precise point estimates, and underlines the need for further information in this field, specifically, larger trials based on each subgroup of fracture complexity. The need for further research is also highlighted by the weakness of estimated effects based on the GRADE assessment performed. Another problem related to the small number of trials included is that when they are separated into subgroups, an I^2 of 0% may not represent true homogeny amongst the trials, rather a lack of power to identify heterogeneity. Secondly, quantifying heterogeneity using I^2 as a metric that has intervals of severity such as those defined in our methods has been subject to methodological debate.[42] Despite this, the definition of moderate heterogeneity that we used when reporting between subgroups of type of fracture is based on values described in recognised standards for meta-analysis research.[41] Thirdly, for the three studies selected and included in sensitivity analysis of internal fixation as a predominant treatment, one involved tension-band wiring[12] and the remaining two included locking-plate fixation.[21,25] Locking plates are a newer technology that are often considered to be advantageous for internal fixation, and therefore, some caution is required when considering the results of sensitivity analysis performed on this group, as there are potential differences in efficacy of these two variations of internal fixation. Fourthly, an absence of RCTs that compare reverse arthroplasty with conservative treatment limits the clinical application of meta-analysis of the current evidence base, particularly with the growing uptake of this procedure for fracture management,[43] and some evidence to suggest that it is superior to hemiarthroplasty for complex fractures of the proximal humerus in the elderly.[44] Finally, it is important to state that subgroup analysis is an observational method that identifies possible variables that drive heterogeneity in meta-analysis. The strength of our findings are increased by our selected subgroup variable having been selected *a priori* and a biological plausibility that fracture morphology affects outcomes. Our results have important implications for the design of future trials, and add weight to an argument that more specific inclusion criteria are required so that their conclusions can better inform clinical decision making.

In conclusion, although there is no overall difference in clinical outcomes for RCTs that compare surgical and conservative treatment of displaced fractures of the proximal humerus, differences in outcomes were detected when subgroup and sensitivity analyses were performed based on type of fracture, as well as the type of surgical intervention performed. Clinical heterogeneity based on these factors appears to be underpinning statistical heterogeneity of treatment effects. Future research within this field may produce more clinically applicable findings if study design adopts an inclusion criteria based on a more homogenous type of fracture, or a single surgical intervention.

Supplementary material

ë A table showing sensitivity analysis of included studies based on risk of bias and the predominant surgical procedure performed within the trial is available alongside this article at www.bjr.boneandjoint.org.uk

References

1. **Baron JA, Karagas M, Barrett J, et al.** Basic epidemiology of fractures of the upper and lower limb among Americans over 65 years of age. *Epidemiology* 1996;7:612-618.
2. **Kannus P, Palvanen M, Niemi S, et al.** Osteoporotic fractures of the proximal humerus in elderly Finnish persons: sharp increase in 1970-1998 and alarming projections for the new millennium. *Acta Orthop Scand* 2000;71:465-470.
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;64:407-414.
4. **Court-Brown CM, Garg A, McQueen MM.** The epidemiology of proximal humeral fractures. *Acta Orthop Scand* 2001;72:365-371.
5. **Majed A, Macleod I, Bull AM, et al.** Proximal humeral fracture classification systems revisited. *J Shoulder Elbow Surg* 2011;20:1125-1132.
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-A:121-131.
7. **Huttunen TT, Launonen AP, Pihlajamäki H, et al.** Trends in the surgical treatment of proximal humeral fractures - a nationwide 23-year study in Finland. *BMC Musculoskelet Disord* 2012;13:261.
8. **Handoll HH, Ollivere BJ, Rollins KE.** Interventions for treating proximal humeral fractures in adults. *Cochrane Database Syst Rev* 2012;12:CD000434.
9. **No authors listed.** Oxford Centre for Evidence-Based Medicine. Oxford Levels of Evidence 2. http://www.cebm.net/ocebm-levels-of-evidence/ (date last accessed 14 September 2016).
10. **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.
11. **Fjalestad T, Hole MO, Hovden IA, et al.** Surgical treatment with an angular stable plate for complex displaced proximal humeral fractures in elderly patients: a randomized controlled trial. *J Orthop Trauma* 2012;26:98-106.
12. **Zyto K, Ahrengart L, Sperber A, et al.** Treatment of displaced proximal humeral fractures in elderly patients. *J Bone Joint Surg [Br]* 1997;79-B:412-417.
13. **Neer CS, II.** Displaced proximal humeral fractures. I. Classification and evaluation. *J Bone Joint Surg [Am]* 1970;52-A:1077-1089.
14. **West SL, Gartlehner G, Mansfield AJ, et al.** Comparative Effectiveness Review Methods: Clinical Heterogeneity. Rockville (MD): Agency for Healthcare Research and Quality (US); 2010. Report No.: 10-EHC070-EF.
15. **Stanley K.** Design of randomized controlled trials. *Circulation* 2007;115:1164-1169.
16. **Moher D, Liberati A, Tetzlaff J, et al.** Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
17. **Higgins JP, Altman DG, Gøtzsche PC, et al.** The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
18. **Guyatt GH, Oxman AD, Vist GE, et al.** GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-926.
19. **Bown MJ, Sutton AJ.** Quality Control in Systematic Reviews and Meta-analyses. *Eur J Vasc Endovasc Surg* 2010;40:669-677.
20. **Boons HW, Goosen JH, van Grinsven S, et al.** Hemiarthroplasty for humeral four-part fractures for patients 65 years and older: a randomized controlled trial. *Clin Orthop Relat Res* 2012;470:3483-3491.
21. **Olerud P, Ahrengart L, Ponzer S, et al.** Internal fixation versus nonoperative treatment of displaced 3-part proximal humeral fractures in elderly patients: a randomized controlled trial. *J Shoulder Elbow Surg* 2011;20:747-755.
22. **Olerud P, Ahrengart L, Ponzer S, et al.** Hemiarthroplasty versus nonoperative treatment of displaced 4-part proximal humeral fractures in elderly patients: a randomized controlled trial. *J Shoulder Elbow Surg* 2011;20:1025-1033.
23. **Stableforth PG.** Four-part fractures of the neck of the humerus. *J Bone Joint Surg [Br]* 1984;66-B:104-108.
24. **Fjalestad T, Hole MO, Jorgensen JJ, et al.** Health and cost consequences of surgical versus conservative treatment for a comminuted proximal humeral fracture in elderly patients. *Injury* 2010;41:599-605.
25. **Fjalestad T, Hole MO.** Displaced proximal humeral fractures: operative versus non-operative treatment–a 2-year extension of a randomized controlled trial. *Eur J Orthop Surg Traumatol* 2014;24:1067-1073.

26. Higgins J, Green S. *Cochrane handbook for systematic reviews of interventions*. Oxford: Wiley-Blackwell; 2008.

27. Xie L, Ding F, Zhao Z, et al. Operative versus non-operative treatment in complex proximal humeral fractures: a meta-analysis of randomized controlled trials. *Springerplus* 2015;4:728.

28. Rabi S, Evaniew N, Sprague SA, et al. Operative vs non-operative management of displaced proximal humeral fractures in the elderly: A systematic review and meta-analysis of randomized controlled trials. *World J Orthop* 2015;6:838-846.

29. Handoll HH, Brorson S. Interventions for treating proximal humeral fractures in adults. *Cochrane Database Syst Rev* 2015;11:CD000434.

30. Fridell JA, Stratta RJ. Program-specific reports for pancreas transplantation: comparing apples and oranges, apples after oranges and apples alone. *Am J Transplant* 2013;13:251-252.

31. Resch H. Proximal humeral fractures: current controversies. *Journal of Shoulder and Elbow Surgery* 2011;20:827-832.

32. Xu J, Zhang C, Wang T. Avascular necrosis in proximal humeral fractures in patients treated with operative fixation: a meta-analysis. *J Orthop Surg Res* 2014;9:31.

33. Wise J. Surgery is no better than a simple sling for displaced fracture of upper arm, study finds. *BMJ* 2015;350:h1304.

34. Foruria AM, de Gracia MM, Larson DR, et al. The pattern of the fracture and displacement of the fragments predict the outcome in proximal humeral fractures. *J Bone Joint Surg [Br]* 2011;93-B:378-386.

35. No authors listed. World Health Organization International Clinical Trials Registry Platform. http://apps.who.int/trialsearch/ (date last accessed 14 September 2016).

36. Bellomo R, Warrillow SJ, Reade MC. Why we should be wary of single-center trials. *Crit Care Med* 2009;37:3114-3119.

37. Boutron I, Ravaud P, Nizard R. The design and assessment of prospective randomised, controlled trials in orthopaedic surgery. *J Bone Joint Surg [Br]* 2007;89-B:858-863.

38. Wylie JD, Beckmann JT, Granger E, et al. Functional outcomes assessment in shoulder surgery. *World J Orthop* 2014;5:623-633.

39. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg [Br]* 1996;78-B:593-600.

40. Haywood KL, Griffin XL, Achten J, et al. Developing a core outcome set for hip fracture trials. *Bone Joint J* 2014;96-B-8:1016-1023.

41. Higgins J, Green SP. *Cochrane handbook for systematic reviews of interventions*. Oxford: Wiley-Blackwell; 2008.

42. Higgins JP. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *Int J Epidemiol* 2008;37:1158-1160.

43. Jobin CM, Galdi B, Anakwenze OA, et al. Reverse shoulder arthroplasty for the management of proximal humerus fractures. *J Am Acad Orthop Surg* 2015;23:190-201.

44. Cuff DJ, Pupello DR. Comparison of hemiarthroplasty and reverse shoulder arthroplasty for the treatment of proximal humeral fractures in elderly patients. *J Bone Joint Surg [Am]* 2013;95-A:2050-2055.

Funding Statement
- None declared

Author Contribution
- S. Sabharwal: Performed study design, Data collection, Data analysis, Writing the paper
- N. K. Patel: Performed data collection, Data analysis, Editing the paper
- D. Griffiths: Performed study design, Editing the paper
- T. Athanasiou: Performed study design, Data analysis, Editing the paper
- C. M. Gupte: Performed study design, data analysis, Editing the paper
- P. Reilly: Performed study design, Data analysis, Editing the paper

ICMJE conflict of interest
- None declared