# Clinical Trials in Low Volume Orthopaedic Surgery:

## Management of Unreconstructible Distal Humerus Fractures

**Adam Charles Watts MBBS BSc FRCS(T&O)**

PhD

University of York
Health Sciences

September 2024

## Abstract

Multi-fragment intra-articular distal humerus fractures in adults are treated with elbow replacement when surgery is indicated, there are no serious co-morbidities, and where the elbow fracture cannot be treated with plates and screws. It is not clear if patients should be treated with replacement of both sides of the elbow, total elbow replacement (TER), or just replacement of the broken humeral part, distal humerus hemi-replacement (DHH). This type of injury has very low incidence and that poses questions about the best methods to evaluate the different surgical procedures. The overall aim of this thesis is to explore outcomes that could be used in a clinical trial of low volume surgical procedures of the management of acute unreconstructible distal humerus fractures in adults. A scoping review was conducted to map the outcomes, trial methods and funding sources used for elbow replacement research. There were 362 published studies identified that reported 583 outcomes. A real-time Delphi study identified 9 mandatory core domains for elbow replacement that should be used in future research, and patients identified pain as the most important outcome domain for them. A systematic review of the literature on pain outcomes following elbow replacement for trauma found that there may be differences in the distribution of pain outcomes from TER and DHH, but direct comparison could not be made because the outcome instruments were not validated for elbow replacement and the quality of included studies was low. The patient rated elbow evaluation is the only pain instrument validated for use in elbow replacement that captures the frequency of pain. An expert elicitation study was conducted to produced prior probability density functions of the likely difference in pain outcomes for a comparative randomised trial of TER and DHH in adult acute distal humerus fracture, using this pain instrument.

# Table of Contents

# List of Tables

**Tables for Paper 4 (Chapter 2.6)**

**Table 1** Individual expert elicited Median PREE scores at 12 months following TER for acute distal humerus fracture.

**Table 2** Individual expert elicited difference in the median PREE score at 12 months between TER and DHH for distal humerus fracture.

**Table 3** Individual expert elicited distribution of the standard deviations of the PREE pain score at 12 months following TER and DHH

# List of Figures

# Acknowledgements

I would like to thank Professor Catherine Hewitt and Professor Catriona McDaid, my supervisors through this PhD for their invaluable support and insights. They have generously given their time to encourage my studies, and patiently guided my thought processes through some challenging areas of exploration. They have undoubtedly improved my scientific thinking and critical analysis along the way, and thankfully restrained some of my more exuberant aspirations, to help deliver what I hope is a coherent work that will provide some useful advances for the orthopaedic community of which I am part.

I would also like to thank my Thesis Advisory Panel, Professor David Torgerson and Miss Ada Keding from the University of York, and Dr Lisa Hampson, from Pharmaceutical Company Novartis for their valuable support and guidance. The insights throughout and particularly into the world of Bayesian methods were invaluable. I would also like to thank Dr Jeremy Oakley from the University of Sheffield for taking the time to discuss the expert elicitation methods used.

There have been many patients and clinicians who have given their time freely to contribute to the studies included here and I am grateful to all for their generosity and support, without which this would not have been possible.

Lastly, I would like to thank my very long-suffering family, Scott, Bailey and Hayden for putting up with me and the time taken to complete these studies. Thanks for tolerating the head down, tapping on the computer or in books and the general sense of distraction. I could not do any of what I do without the love and happiness you bring. To my parents Vanessa and Alan, thanks for all you have given me, all the sacrifices you have made, and all your love and belief.

## Author's declaration

The four papers that make up the thesis are original works written by me in collaboration with other authors, and I have described my contribution to each in the relevant sections. I declare that the introductory chapter, integrative sections, discussion and conclusions of the thesis are a presentation of original work and I am the sole author. A journal style-thesis has been employed with chapters in a format suitable for publication with surrounding supporting commentaries. This has led to repetition of arguments through the thesis, but is necessary to make the individual publications coherent. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

*“The ordinary patient goes to his doctor because he is in pain or some other discomfort and wants to be comfortable again; he is not in pursuit of the ideal of health in any direct sense. The doctor on the other hand wants to discover the pathological condition and control it if he can. The two are thus to some degree at cross purposes from the first, and unless the affair is brought to an early and happy conclusion this diversion of aims is likely to become more and more serious as the case goes on.”*

**Wilfred Trotter, Surgeon, University College Hospital, London (1872-1939)**

# 1 Introduction

When seeing a patient in clinic, typically an older female, with a multi-fragment fracture of the humerus bone entering the elbow joint that, based on the plain radiographs and computer tomography (CT) scan assessment, is not thought to be amenable to fixation, the clinician is left with a dilemma. What is the best treatment for this patient? Should we just replace the broken part of the elbow (distal humerus hemi-replacement) or should we replace the whole joint (total elbow replacement)? This uncertainty is reflected in my own practice.

## 1.1 Distal humerus fractures

Distal humerus fractures account for approximately 0.5% of adult fractures and with a reported incidence of 5.8 fractures per 100,000 people per year in the UK and a 2:1 female to male ratio.(Court-Brown and Caesar, 2006) The incidence of distal humerus fractures increases with age and it is reported that 80% occur in people over the age of 50 years.(Bergdahl *et al.*, 2016) Most of these types of fracture occur as a result of a low energy fall and the overall number of fractures has been reported to be increasing, possibly due to the ageing demographic in developed nations.(Palvanen *et al.*, 2010; Bergdahl *et al.*, 2016) In the population over 65 years of age approximately 50% of distal humerus fractures have been reported to include the whole of the articular surface with varying degrees of fragmentation of the bone.(Charissoux *et al.*, 2013)

The treatment of a distal humerus fracture is determined by patient factors, fracture personality, surgeon factors and resources. The treatment options for this injury are cast immobilisation without an operation, fixation with plates and screws (osteosynthesis), and elbow prosthetic replacement (elbow arthroplasty).(Stoddart *et al.*, 2022) Where the bone cannot be fixed there are two options for elbow replacement for trauma. These are distal humerus hemi-replacement (DHH) in which just the broken elbow end of the humerus is replaced, and total elbow replacement (TER) in which both the broken humerus and intact proximal ulna are replaced. In a prospective study of 87 patients treated in 19 hospitals in France, 25% underwent non-surgical treatment, 61% osteosynthesis and 14% elbow replacement.(Charissoux *et al.*, 2013)

The best treatment for fractures of the distal part of the humerus in adults is not known. In younger patients, fixation with plates and screws (osteosynthesis) is the preferred treatment because it preserves the patient's own skeleton and because joint replacement has a finite lifespan. In older patients, over the age of 60, there is still some uncertainty whether fixation or replacement is better, and this is currently under investigation in the UK in the NIHR funded FOREST trial (NIHR154915) that will compare the outcome of osteosynthesis and joint replacement. This phase III, multicentre, parallel-group, superiority randomised controlled trial will compare the clinical and cost-effectiveness of elbow replacement or fracture fixation with plates and screws in 320 participants over the age of 60 years with intra-articular distal humerus fractures using the primary outcome of Disabilities of Arm Shoulder and Hand (DASH) score at 4 months. There are circumstances, however, where a high degree of fracture fragmentation, poor bone quality or pre-existing painful joint damage (arthritis) mean that successful restoration of pain free function with fixation is not going to be possible, and in this situation elbow replacement is indicated.

Total elbow replacement has been used in orthopaedic practice for more than 20 years as a treatment in older patients where fixation is not possible, but studies indicate high rates of complication when patients are followed up for a minimum of 10 years.(Cobb and Morrey, 1997; Barco *et al.*, 2017) Total elbow replacement continues to be the treatment of choice in patients with pre-existing symptomatic elbow arthritis where hemi-replacement would cause pain, or in those with co-existing fractures of the proximal forearm bones (radial head and coronoid process of the ulna) that could lead to instability of a hemi-replacement. Hemi-replacement has been used successfully in the treatment of fractures of the hip and shoulder, and has rapidly grown to become the most common treatment for unreconstructible elbow fractures since the development of an anatomical prosthesis in the first decade of this century.(Kontakis *et al.*, 2008; Ekhtiari *et al.*, 2020; Achakri *et al.*, 2023) This change in practice is not supported by scientific evidence, and the effectiveness and cost-effectiveness of the different treatment options remains unanswered.

## 1.2 Elbow Anatomy

The upper limb consists, from proximal to distal, of the shoulder girdle, the arm, forearm, wrist and hand. The only bone in the arm is the humerus that articulates with the glenoid of the scapula (shoulder blade) proximally and the greater sigmoid notch of the ulna and radial head distally. Distally the humerus bone consists of the humeral head, neck, shaft, medial and lateral columns, medial and lateral epicondyles and the articular surface that is made from the spool shaped trochlea medially and the hemispherical capitellum laterally. Just proximal to the trochlea spool are two fossae, anteriorly the coronoid fossa that accommodates the coronoid process of the ulna in deep flexion of the elbow, and posteriorly the olecranon fossa that accommodates the olecranon process of the ulna in full elbow extension. The joint capsule has thickenings medially and laterally that form the medial and lateral ligaments complexes. The medial ligament complex is divided into an anterior band from the anteroinferior portion of the medial epicondyle to the sublime tubercle of the ulna, the posterior band from the inferior surface of the medial epicondyle to the medial margin of the greater sigmoid notch, and the transverse band that connects two points on the ulna. The laterally ligament complex consists, from anterior to posterior, of the radial collateral ligament arising from the lateral epicondyle and blending distally with the annular ligament that encircles the radial head, the lateral ulna collateral ligament from the lateral epicondyle to the supinator crest of the ulna, the accessory lateral ulna collateral ligament, and posteriorly the posterolateral ligament of Osborne that arises from the inferior portion of the lateral epicondyle and fans out to insert onto the lateral rim of the greater sigmoid notch. The medial epicondyle is the common origin for the flexor tendons of the forearm muscles, the lateral epicondyle is the common origin for the extensor tendons of the forearm muscles.

## 1.3 Definition of distal humerus fracture

There is no formal definition of the distal humerus but it can be described as the lower half of the humerus furthest from the shoulder. The distal humerus consists of the distal half of the humeral diaphysis and the distal metaphysis, defined as that portion of the lower end of the bone that is bounded by a square with sides determined by the widest part of the bone at that level, which includes the medial and lateral humeral column, the medial and lateral epicondyles, the olecranon and coronoid fossa, the trochlea spool that articulates with the

greater sigmoid notch of the ulna to form the ulnohumeral joint and the capitellum that articulates with the radial head to form the radiocapitellar joint. The ulnohumeral and radiocapitellar joint constitute the elbow joint, a complex spiral hinge joint.

The term distal humeral fracture includes a fracture of any or several of these parts. Fractures can enter the elbow joint (intra-articular) or can be outside the elbow joint (extra-articular). Fracture of the distal humerus in the immature skeleton (paediatric distal humerus fracture) has different considerations for anatomy, diagnosis, classification and treatment and will not be considered further.

## 1.4 Classification of distal humerus fracture

A number of classifications of distal humerus fractures have been described. Riseborough and Radin considered patterns of intra-articular "T" shaped fracture with displacement, rotation and fragmentation as discriminators.(Riseborough and Radin, 1969) Jupiter and Mehne described a classification that separated fracture into intra-articular, extra-articular intra-capsular fractures and extra-capsular categories with subclassification of intra-articular fractures by the number of columns and the plane of the injury, extra-articular intra-capsular by the level and plane of injury and extra-capsular as medial or lateral epicondyle.(Jupiter and Mehne, 1992) In 1990 Muller et al. proposed a system for classification for all principle bones of the human body.(Müller *et al.*, 1990) In this system the humerus is assigned the number "1", and is then divided into proximal (1), middle (2) and distal (3) parts and then assigned the letter (A) extra-articular, (B) partial articular and (C) complete articular fracture. Further subcategorization based on the degree of fragmentation is possible in this comprehensive system. In patients over 65 years there has been reported to be a roughly equal split between AO C1,C2,C3 fracture configurations of the distal humerus in the population.(Charissoux *et al.*, 2013) However, none of these classifications have been shown to have more than moderate agreement on inter- and intra-observer assessment based on kappa statistics categorised according to the guidelines proposed by Landis and Koch, and their utility has been called in to question.(Landis and Koch, 1977; Wainwright, Williams and Carr, 2000)

The Dubberley classification considered only fractures of the articular surface involving the capitellum and trochlea, dividing them according to the amount of involvement of the trochlea and the degree of fragmentation with subclassification depending on the presence of dorsal fragmentation.(Dubberley, 2006) It has been shown that the degree of fragmentation is underestimated from pre-operative plain radiographic assessment.(Watts, Morris and Robinson, 2007)

## 1.5 Treatment of distal humerus fractures

### 1.5.1 Non-surgical treatment

Non-surgical (non-operative) treatment of distal humerus fracture is generally reserved for some extra-articular fractures, fractures without significant displacement and fractures in frail adults in whom the risks of surgery outweigh the benefits. In this latter group non-surgical treatment has been used even for displaced multi-fragment fractures in what is colloquially known as “bag of bones” treatment, with moderate functional outcomes in the medium term but mortality of 39% at 5 years, reflecting the frailty of this population.(Aitken, Jenkins and Rymaszewski, 2015)

### 1.5.2 Osteosynthesis (fixation with plates and screws)

In younger patients osteosynthesis is the preferred treatment for distal humerus fractures, where this is achievable, as it retains native bone stock and articular cartilage. However, a 2021 systematic review and meta-analysis of 83 clinical outcome studies published in the English language found that distal humerus osteosynthesis is associated with an overall reported complication rate of 53% and a re-operation rate of 21%.(Yetter, Weatherby and Somerson, 2021) Salvage of failed osteosynthesis has inferior outcome, lower implant survival and greater risk of more, and more serious, complications than acute primary total elbow for trauma.(Prasad and Dent, 2008)

### 1.5.3 Elbow replacement

Prosthetic joint replacement of all or part of the elbow has been performed for a number of indications since the first "modern" elbow replacement design of Dee in the 1960's, including inflammatory arthropathy, osteoarthritis, acute trauma, trauma sequelae, instability and (rarely) tumour.(Dee, 1975)

The term elbow replacement is often taken to mean total elbow replacement, that is replacement of the distal humerus and proximal ulna with or without the radial head, but also includes distal humerus replacement in isolation (distal humerus hemi-replacement), replacement of the capitellum of the humerus and radial head (radio-capitellar replacement) and radial head replacement in isolation. Total elbow replacement and distal humerus hemi-replacement are used for the treatment of unreconstructible distal humerus fractures. The purpose of elbow arthroplasty is to relieve pain, restore function and restore quality of life.

#### *1.5.3.1 Total elbow replacement*

Total elbow replacement can be used for treatment of unreconstructible distal humerus fracture where osteosynthesis is not possible. The first reported series came from the Mayo clinic in 1997.(Cobb and Morrey, 1997) A number of studies, including randomised trials, have compared total elbow replacement to osteosynthesis for acute distal humerus fracture.(Frankle *et al.*, 2003; McKee *et al.*, 2009) This assumes that both treatments are appropriate for all multi-fragmentary distal humerus fractures but as demonstrated by high rates of crossover (25% in one trial) osteosynthesis is not a viable option if the bone fragmentation is too great or the bone quality insufficient to achieve fixation.(McKee *et al.*, 2009)

A number of case series from different countries have reported on the outcome of elbow replacement for acute fracture with sample sizes from 7 to 87 patients with an average age from 69 to 84 years and average follow up from 17 to 120 months.(Cobb and Morrey, 1997; Ray *et al.*, 2000; Gambirasio *et al.*, 2001; Garcia, Mykula and Stanley, 2002; Kamineni and Morrey, 2004; Lee, Lai and Singh, 2006; Prasad and Dent, 2008; Mansat *et al.*, 2013; Barco

*et al.*, 2017) All reported outcomes using the Mayo Elbow Performance Score (MEPS) with average scores between 84 and 95(out of 100). The incidence of reported complication was high ranging from 12 to 52% of cases, and patients had a flexion contracture of the elbow ranging from an average of 20 degrees to 41 degrees. This loss of elbow extension can limit how far an individual can reach and may affect function, gait velocity and balance.(Akhtar, Hughes and Watts, 2021)(Trehan *et al.*, 2015) In one series over one third of patients reported ongoing pain in the elbow.(Mansat *et al.*, 2013)

One study compared the outcome of acute elbow replacement to replacement performed for salvage of failed osteosynthesis and has reported no significant differences in the Mayo Elbow Performance Score and range of movement between the two groups.(Logli *et al.*, 2020) It should be noted however that the groups were quite heterogeneous with regards to age, make of implant used and surgical approach. The rate of re-operation was comparable at 36% at an average follow up of 4.4 years for the acute group and 39% at an average follow up of 5.5 years for the salvage group.

#### *1.5.3.2 Hemi-replacement*

Use of hemi-replacement to replace the distal humerus has been documented from the 1940's with custom implants created for trauma sequelae or tumour reconstruction.(Mellex and Phalen, 1947; MacAusland, 1954; Street and Stevens, 1974; Shifrin and Johnson, 1990; Barr and Eaton, no date) The first acute trauma cases where treated with commercially available humeral components intended for use as part of a total elbow replacement.(Adolfsson and Hammer, 2006; Smith and Hughes, 2013) It was not until 2011 that the use of a commercially available implant specifically designed to replace the anatomy of the distal humerus for acute trauma reconstruction was reported.(Burkhart *et al.*, 2011) Since then a number of mainly small case series have been reported.(Argintar *et al.*, 2012; Nestorson *et al.*, 2015; Phadnis, *et al.*, 2015; Schultzel *et al.*, 2017; Al-Hamdani *et al.*, 2019) A systematic review of 116 elbow hemi-replacement cases for acute trauma reported a mean patient age of 62 years and 72% were female.(Dunn, Kusnezov and Pirela-Cruz, 2014) The authors reported that 79% of the injuries occurred as a result of a fall from standing height. The average flexion-extension arc of movement was reported as 98°, a loss of range

of movement from the normal 140°, with high rates of re-operation (33%) and complications (28%).

The stated advantages of hemi-replacement over total elbow replacement are an eliminated risk of polyethylene wear because the polyethylene ulna component used in total elbow replacement is absent, shorter surgical time, lower risk of implant loosening and unrestricted loading through the elbow.(Dunn, Kusnezov and Pirela-Cruz, 2014) Whilst statements regarding polyethylene wear may be well founded, there is no evidence to support the other stated benefits, and concerns have been raised regarding wear of the native greater sigmoid notch and the risk of joint dislocation.(Lee, Adams and Morrey, 2005)(Smith and Hughes, 2013) There is also a conflict in the literature regarding the most suitable patient cohort in whom to use a hemi-replacement and total elbow replacement. Some prefer using hemi-replacement in the elderly "to reduce the duration of surgery and the risk of potential complications in a population of elderly and multimorbid patients".(Burkhart *et al.*, 2011) Others prefer to use total elbow replacement "for elderly patients who have moderate functional demands and are able to accept lifelong activity restriction to prevent mechanical complications".(Al-Hamdani *et al.*, 2019) Personal experience suggests that the more unhappy patients following hemi-replacement are functionally active patients, who complain of elbow stiffness, functional impairment and ongoing low level pain. There is some limited support for these observations from the published literature.(Argintar *et al.*, 2012; Rotini *et al.*, 2020; Heifner *et al.*, 2024)

## 1.6 Surgical volume – data from the National Joint Registry

Elbow replacement is not frequently performed. The annual incidence in Scotland has been estimated as 1.4 per 100,000 population. (Jenkins *et al.*, 2013) Analysis of the National Consultant Information Portal (NCIP) and National Census data indicates that the annual incidence in England in more recent years is lower at 0.7/100 000 (Appendix B). This figure is consistent with the number of episodes recorded in the National Joint Registry of England, Wales and Northern Ireland (NJR), a national database of replacement procedures, that has collected data on elbow replacement since 2012.(Hamoodi *et al.*, In press) While it is a mandatory requirement for surgeons to submit data on all elbow replacement cases

performed, anomalies in the data indicate that capture of cases has been incomplete but may be improving. Between 480 to 1,052 elbow replacement procedures have been recorded annually in the NJR since 2012 and it already contains the largest dataset of elbow replacements worldwide with 8,940 cases included in the latest annual report.(Achakri *et al.*, 2023) The NJR annual report separates cases into "confirmed" when there is a match between the description of the procedure submitted and the implants used, and "unconfirmed" where there is a mismatch. Replacement for acute trauma constitutes 54% of all elbow replacements in the NJR dataset from 2012 to 2022 but as a proportion there has been an increase in elbow replacement for trauma, which accounts for 59% of all elbow replacements in the most recent year recorded (2022).(Achakri *et al.*, 2023) The majority of these, however, are radial head replacements (328 of 472 confirmed cases) which are not used for treating distal humerus fractures. Distal humeral hemi-replacement has only been formally captured by the NJR since 2018 but in the last three years of the NJR dataset (2020-2022) there were 247 confirmed hemi-replacement procedures recorded for acute trauma and 239 confirmed total elbow replacement procedures performed in over 90 hospitals.(Achakri *et al.*, 2023) It is too early for the NJR to report reliable implant survival data for hemi-replacement, but at 1 year follow up of 385 implants the revision rate is reported to be 2.75% (95% Confidence Interval (CI) 1.49 to 5.06). The revision rate for total elbow replacement used in acute trauma is reported to be 1.28% (95% CI 0.73-2.25), and appears to increase by approximately 1% per annum up to 5 years. The 12-month mortality after hemi-replacement is reported to be 2.34% (95% CI 1.17 to 4.62), which is lower than for total elbow replacement for acute trauma at 6.39% (95% CI 5.00 to 8.13), but the median age of patients undergoing hemi-replacement is 5 years younger than those undergoing total elbow replacement for trauma (hemi-replacement median 72 years, IQR 65-79; total elbow replacement median 77 years, IQR 71-83). Implant survival and mortality are the only outcome data collected and reported for elbow replacement by the NJR.

## 1.7 Clinical trials in low volume orthopaedic surgery

Well conducted trials have been recognised as the best way to evaluate the effectiveness of interventions since the first trial famously conducted by James Lind on a British Naval vessels into the effect of fresh citrus fruit on scurvy in 1753.(Lind, 1753) The introduction of

randomisation was attributed to a group from Nuremberg in 1836, but the modern era of randomised controlled trials (RCTs) began with Colebrook in 1929.(Colebrook, 1929; Jamison, 2019)

It is recognised that surgical trials are challenging to undertake because surgical interventions are complex, with multiple social, organisational, physiological, physical, technical and economic variables that may alter the outcome, either independently or in concert.(Blencowe *et al.*, 2015) Research is made even more challenging when the intervention is not commonly performed. Research in these low volume interventions is particularly important because individual surgeon experience may not highlight concerns regarding harms. One or two failures may be dismissed as just bad luck. The scale of a problem may only come to light when data is pooled, as was seen with an implant called the LPM proximal interphalangeal replacement (Van Straten Medical, Netherlands), that was found to have a failure rate of 50% within 6 years of implantation.(Hobby *et al.*, 2008) Failure of an elbow replacement may mean additional surgery for a patient, with approximately 50% risk of complications, and substantial direct healthcare costs that have been estimated to be nearly three times the cost of a primary procedure.(Geurts *et al.*, 2019)(Wagner *et al.*, 2017)

Drawing inference from RCTs typically employs classical approaches for testing research data, usually a combination of hypothesis testing and testing of significance developed by Neyman and Pearson and Fisher. These classical approaches require achieving statistical power in trials with sufficient participants or events to detect a meaningful difference between interventions.(Altman, 1980; Cook *et al.*, 2018)(Spiegelhalter *et al.*, 2000) The null hypothesis assumes that there is no difference between the two interventions. The P value is the probability that we could have obtained the data or more extreme data, when the null hypothesis is true. The $\alpha$ value is the probability of rejecting the null hypothesis when it is in fact true (type 1 error), and this is normally predefined, often at 0.05 or 5%. The $\beta$ value is the probability of accepting the null hypothesis when it is false (type II error), and we can set $\beta$ in advance of a study by calculating a sample size that will detect a true effect of a given magnitude, this is expressed as 1-$\beta$ (or 100 x (1-$\beta$)%), the power of the study.(Altman, 1980)

A smaller effect size will require a larger sample to achieve the desired power. There has been strong criticism in the past of the practice of underpowered trials, that is trials with a high risk of type II error, that may be viewed as unethical as they expose participants to the risk of the trial, and demand their time, but are of limited clinical value.(Halpern, 2002) These underpowered trials can be combined in meta-analysis to draw more robust conclusions, but heterogeneity in methods and measurement, and the risk of publication bias, may prevent useful pooling of data.(Bailar, 1997) When you have low-volume surgical procedures this poses challenges in using traditional approaches.

## 1.8 Learning from research into rare diseases

The European Union defines rare diseases as those with a prevalence of less than 50 per 100 000 population for the purposes of orphan drugs, where normal economic models prevent development.('Orphan Drug Act', 1983) Whilst individual conditions may be rare, the total number of rare conditions mean that large numbers of the population (approximately 30 million people in Europe) may be affected by one of these conditions, and in fact rare diseases may be viewed as commonplace.(Eurordis, 2005; Elliott and Zurynski, 2015)

It has been argued that in rare disease research, where a multicentre trial is unlikely to recruit sufficient participants to achieve statistical power, it may be ethical to conduct underpowered studies, on the basis that some knowledge is better than none.(Edwards et al., 1997) This approach is considered problematic because it is most likely, unless the effect size is very large, that the null hypothesis will be accepted when it may be false.(Halpern, 2002) There may be some value in early phase trials (phase 1 and 2) because they provide useful information on secondary outcomes, such as safety, despite being underpowered, and may be used to inform power calculations of phase 3 trials.(Halpern, 2002)

The concept of rarity also has applicability for low-volume surgical interventions. The same economic and scientific challenges apply, with low investment in research and challenges to conventional frequentist inference to investigate clinical outcomes due to feasibility of conducting large-scale randomised controlled trials (RCTs). Combining Office for National Statistics population data for the period April 2019 to March 2020 (prior to the disruption of

the COVID-19 pandemic) with data from the National Health Service of England (NHS) National Consultant Information Portal (NCIP), that reports data on hospital episodes in England based on the mandatory Hospital Episode Statistics (HES) database, shows us than in orthopaedics very few interventions have an annual incidence exceeding 50/100 000 population (Appendix B). Total knee replacement has the highest incidence at 140/100,000 population per annum, followed by total hip replacement (131/100,000), fracture neck of femur surgery (111/100,000), knee arthroscopy (82/100,000) and carpal tunnel release (76/100,000). Carpal tunnel surgery is performed three times more frequently than the next most common orthopaedic intervention reported by NCIP, Dupuytren's surgery, that had an annual incidence of 27/100,000 in that period. Whilst it has been feasible to conduct well-funded large scale randomised controlled trials over many centres for interventions of this incidence, as was achieved in the NIHR funded DISC trial that recruited 673 participants in just under four years, when the incidence is very low there are legitimate questions about feasibility and costs.(Dias *et al.*, 2021) Evidence of the challenges faced can be observed from NIHR funded trials in orthopaedics such as UK-FROST investigating treatments for frozen shoulder (incidence 6.4/100,000) that required a 5 month extension for recruitment and increased the number of recruitment centres by 40%, CSAW investigating subacromial decompression (incidence 6.39/100,000) that recruited to target but increased recruitment centres by 130% against plan, and TARVA that investigated ankle replacement (incidence 1.54/100,000) that took 7.25 years to reach recruitment target against a plan of 4.25 years.(Rangan *et al.*, 2020)(Beard *et al.*, 2018)(Goldberg *et al.*, 2023) The incidence of elbow replacement is even lower (0.7/100,000) and only 40% of these are undertaken for acute trauma.(Achakri *et al.*, 2023) The challenge of conducting research is compounded by the fragmentation of service delivery, with elbow replacement procedures being undertaken by over ninety providers in England, Wales, Northern Ireland and Guernsey.(Achakri *et al.*, 2023)

## 1.9 Trial methods for rare disease

In RCTs comparing the effects of interventions, two broad approaches are used to address the challenge of a limited pool of potential participants; the first, to reduce the number of participants required and the second, to increase the proportion of participants receiving

treatment.(Gagne *et al.*, 2014) The techniques employed in reducing the required sample size include increasing the period of observation to capture more events, focusing on high-risk patients who are more likely to experience an outcome of interest, using genetic testing to reduce variability between individuals or factorial designs in which more than one treatment comparison is undertaken together.(Gagne *et al.*, 2014) More practical approaches for surgical trials can include using continuous outcome variables rather than binary measures to reduce the sample size, the use of surrogate or composite endpoints and the use of repeated measure outcomes.(Van der Lee *et al.*, 2008; Shurin, Krischer and Groft, 2012) Alternative approaches include building research networks and the use of international trials to draw participants from a number of countries.(Mundi *et al.*, 2014) Many of the approaches used, while effective, may increase research costs prohibitively in fields that may not attract much funding precisely because they are rare.(Reinecke, Rommel and Schmidtke, 2011) This may hamper rigorous assessment of the clinical benefits and new developments in rare procedures such as elbow arthroplasty.

Different trial designs have been explored in rare disease research, including crossover designs (participants receive both treatments in variable order), N-of-1 trials ( participant spends variable periods of and on treatment), adaptive designs (including response-adaptive randomisation or ranking) and selection designs (either “pick-the-winner” or “drop-the-loser” designs), internal pilot and sequential trials all of which require early determination of the outcome of the intervention.(Gupta *et al.*, 2011) The irreversible nature of surgical interventions, and the need to demonstrate durability of outcomes, particularly when the trial involves a joint replacement, means that these approaches may not be applicable in the surgical setting.(Whicher, Philbin and Aronson, 2018) Others have changed $\alpha$ to raise the threshold for rejecting the null-hypothesis, or have argued that the only ethical approach is to have a pre-planned meta-analysis of data emanating from multiple small trials, but it is difficult to gain funding for these approaches.(Halpern, Karlawish and Berlin, 2002; Gallin *et al.*, 2003) Parmar et al. have suggested a stepwise approach to trial design in small populations.(Parmar, Sydes and Morris, 2016) A power analysis is conducted based on standard parameters. If it is estimated that it would not be feasible to achieve the target sample in a reasonable timeframe consideration is given to enhancing feasibility by increasing accrual or follow-up time, broadening eligibility criteria, or considering

collaboration with other centres nationally or internationally. If this is still not considered feasible then consideration is given to commonly considered approaches to reducing sample size, such as changing the experimental arm to one with a bigger difference to the control arm, changing the outcome instrument to one that offers more precision, increasing the target difference if this can be justified, or reducing the power of the study to the lower limits traditionally accepted. If these measures still fail to address the feasibility then attention is paid to less common approaches, such as relaxing $\alpha$ beyond traditional levels, consideration of one-sided significance testing or use of external information. The last step, if feasibility has not been achieved, is to consider designs for very rare diseases.

One example of an alternative approach that makes use of external information, is to employ Bayesian methodology to trial design and analysis. This is controversial because it does not rely solely on observed data, but also considers prior data to produce prior probability distribution that may be subjective. The 'prior' can be updated using the trial data (likelihood) to produce a posterior distribution for the outcome of interest. The benefit of this approach is that it does not require recruitment of a fixed number of patients to a trial to achieve a predetermined power as in a frequentist paradigm. This has clear advantages in low volume conditions or procedures, for example in research of implantable devices, but has been found to be underused.(Pibouleau and Chevret, 2011) A Bayesian statistical approach cannot replace the frequentist methods to provide a definitive statement of comparative outcomes between two interventions, but where the required sample size is unlikely to be reached even through broad collaboration, a Bayesian approach can bring us closer to understanding the probabilities of chosen outcomes.(Hampson *et al.*, 2014) Bayesian methods can also be combined with standard frequentist randomised controlled trial design, most of which are not powered to permit adequate assessment of subgroups within the trial, to improve analysis of secondary outcomes and interpretation of outcomes.(Wijeysundera *et al.*, 2009; Hampson *et al.*, 2014)

## 1.10 Bayesian methods in clinical trials

Bayesian inference is the use of Bayes theorum to update our beliefs about a quantity of interest based on what we knew previously about that quantity, and updating our knowledge based on observations made.(Donovan and Mickey, 2019) As stated by Berry "Bayes' theorum is formalised learning: that's what I thought before, this is what I just saw, so here's what I now think – and I may change my views tomorrow.(Berry, 1993) Bayes' theorum is expressed most simply as a proportion(Bayes, 1763):

| Posterior probability of hypothesis A, given data B | $\propto$ | Likelihood of data B, given hypothesis A | X | Prior probability of hypothesis A |
|---|---|---|---|---|

The Bayesian inference differs from classical, or frequentist, inference in the way that data are used. In the classical method investigators assess the probability of the observations given the hypothesis, whereas in the Bayesian method the investigator considers the probability of the hypothesis given the observations.(Lewis and Wears, 1993) This is much more closely aligned to the natural way of thinking of a clinician, who might consider what is the probability of a particular diagnosis given the observations of the clinical examination of the patient, or the probability of success of a particular intervention given what has been observed about that intervention in the past.(Gill, Sabin and Schmid, 2005)

When considering the use of Bayes' theorum in clinical trials Berry identified important differences between Bayesian and classical theory. (Berry, 1993) Bayesian probabilities are direct probabilities. Bayesian inference, like clinical practice, can use indictive reasoning and provide the probability of a specified treatment effect, based on the observations that have been made. We can state that the probability of no difference between treatment A and treatment B is 5% given what we knew before and what we have observed. In the classic method the same would be stated indirectly. The latter frequently causes confusion when clinicians are interpreting the results of trials. Classical inference uses deductive reasoning, that starts with a hypothesis and tests whether our observations fit with that

hypothesis.(Wijeysundera *et al.*, 2009) When comparing two interventions a p value of 0.05 tells us that the probability of observing a result as, or more extreme, than that seen is 5%, assuming that the null-hypothesis (typically of no difference between the interventions) is correct. Clinicians, however, will often erroneously interpret $p < 0.05$ as "there was less than 5% probability of finding this difference by chance". Bayesian inference can allow greater flexibility too. Whereas classic theory is applied to a set sample size, the Bayesian model can be updated as more data is added. Bayesian inference can produce probabilities of future outcomes, the probability that my patient will benefit from treatment A. It can also be used to aid decision making; the costs and benefits of a various decisions can be weighed by the predictive probabilities to help choose the best outcome or utility.(Claxton, Lacey and Walker, 2000)

The NHS Health Technology Assessment programme published a review in 2000 that assessed the role of Bayesian methodology in health technology assessment.(Spiegelhalter *et al.*, 2000). This reviewed the literature on Bayesian inference in health technology assessment, providing guides for its use in the conduct of randomised trials, observational studies, evidence synthesis and policy making. This review introduces standards for reporting of Bayesian research, BayesWatch, which like the Consolidated Standards of Reporting Trials (CONSORT) guidelines in classical designs, recommends standard reporting of technology, design and results but emphasises additional elements crucial to transparent Bayesian analysis.(Schulz *et al.*, 2010)(Spiegelhalter *et al.*, 2000)

An important element of the planning and reporting is to establish priors. These do not have to be established before trial observations are collected, as they simply reflect the state of knowledge before the trial, but as in classical methods, to avoid a risk of observations influencing priors, it is helpful to consider these a priori. There is no correct prior to choose and in most cases a family of priors is chosen that can be used to assess how robust the findings are, and consider how our starting beliefs will be affected by the observations made.(Spiegelhalter *et al.*, 2000) Possible priors include non-informative priors or informative priors. Informative priors can take several forms. They can be probability density functions based on prior evidence or on expert views (expert elicitation), they can be sceptical (low probability of large effect size), enthusiastic (centred on the alternative

hypothesis) or 'lump-and-smear' priors, with high probability for outcomes around the null hypothesis and then a spread of probabilities across the range of alternatives.(Spiegelhalter *et al.*, 2000)(Cornfield, 1969) It might be believed, for example, that treatment B has no benefit over treatment A, and a sceptical prior could be adopted, with a distribution based around the null hypothesis with less than 5% chance of an important treatment effect, or an enthusiastic prior could be adopted where there is a believe that B is superior to A, with a treatment difference based on meta-analysis of existing data, or on expert opinion. Where the observed data is strong to indicate a superiority of treatment B over A, the choice of prior will have little effect on the posterior and each will show similar probabilities. Where the observed effect is weak, we can consider the posteriors in relation to possible prior beliefs, reflecting the true uncertainty in the probability of an expected outcome.(Wijeysundera *et al.*, 2009) Non-informative priors will typically assign equal probability to all outcomes across a carefully chosen scale, or consider a normal distribution across all possible outcomes centred on the null hypothesis. Applying Bayesian transformation to a non-informative prior with an even distribution will mean that the posterior distribution will take the form of the likelihood.

## 1.11 Limitations of existing research

The National Joint Registry recorded 486 elbow replacement procedures for acute trauma over the latest three years reported (2020 to 2022) undertaken in over 90 hospitals, with approximately half being total elbow replacement and half hemi-replacement.(Achakri *et al.*, 2023) There is no clear reason why some surgeons are choosing to use total elbow replacement and others hemi-replacement, and no clear evidence to guide patients and clinicians which intervention to choose. To date, there is only one randomised controlled trial that has been published comparing total elbow replacement and distal humerus hemi-replacement for acute distal humerus fractures.(Jonsson *et al.*, 2023) This trial from Sweden recruited 40 patients, 20 in each trial arm, over a period of 10 years. Using a frequentist approach the authors reported no difference in outcome between the two interventions, with a primary outcome of function measured using the Disabilities of the Arm Shoulder and Hand (DASH) Score. It is not clear how the outcomes were chosen for this trial, and there is no defined core outcome set for elbow replacement research. A clearer understanding is

needed of the outcomes that are important to patients and clinicians before further trials can be conducted. The authors concluded that there was no difference between the interventions, however, given the sample size there is a risk of type II error, accepting the null hypothesis when it may be false, due to inadequate sampling of population data. A larger randomised trial and cost effectiveness study could provide valuable information to guide practice, but with the number of hospitals delivering the few interventions undertaken it would be extremely challenging to successfully complete a fully powered trial using a frequentist paradigm in an economically viable way. A Bayesian trial may be a pragmatic solution in this situation.

Bayesian methodology has been almost completely ignored in orthopaedic research. Whilst in recent years there has been interest in the use of Bayesian network meta-analysis when pooling data from randomised trials, the use of Bayesian methods in primary research is almost absent.(Palmer *et al.*, 1997; Wang *et al.*, 2018; Cheng and Sheng, 2020; Migliorini *et al.*, 2020; Wu *et al.*, 2020) A Bayesian adaptive approach has been used to improve efficiency in a trial of treatment for rotator cuff deficiency of the shoulder (start:REACTS trial), where interim analysis was permitted to assess futility or efficacy against strict rules, leading to early completion of the trial.(Metcalfe *et al.*, 2022) A retrospective investigation of the utility of Bayesian adaptive methods using the results from a randomised trial concluded that there was the potential to generate earlier results and allocate more patients to better performing arms, and that these methods should be more widely used.(Ryan *et al.*, 2020) Bayesian methodology has been used in a trauma trial of resuscitative endovascular occlusion of the aorta (REBOA), but the authors chose to use a non-informative prior due to the conflicting prior evidence related to REBOA, which results in a trial that more closely resembles a frequentist paradigm.(Jansen *et al.*, 2017)

Techniques have been described for generating informative priors from previously published research, with scoring and weighting of the data according to the relevance to the intended trial, data validity and precision.(Tan, Dear, *et al.*, 2003) Where there is insufficient published data to generate a prior probability density, techniques of expert elicitation have been used to exploit expert knowledge to extrapolate on related published data to produce informative priors, such as used in a study of chronic nonbacterial osteomyelitis.(Ramanan

*et al.*, 2019)(O'Hagan, 2006) Before undertaking any trial, we need a better understanding of the outcomes that should be measured to compare the interventions, and both classical and Bayesian trial planning require data synthesis and estimation of current knowledge for the primary outcome to inform sample size calculations with the frequentist approach or for informative priors in a Bayesian trial.

## 1.12 Aims and objectives

The aim of this thesis is to explore outcomes that could be used in a clinical trial of low-volume surgical procedures (elbow replacement) for the management of acute unreconstructible distal humerus fractures in adults.

The objectives are:

A: to map the current literature on elbow replacement to describe the populations investigated, the research methods, funding sources and outcomes described to inform future research. (Paper 1)

B: to define the core outcome domains for elbow replacement research and identify a key outcome that is important to patients. (Paper 2)

C: to synthesise the current evidence related to the key outcome for patients having elbow replacement for acute trauma. (Paper 3)

D: to generate an informative prior for the key outcome using expert elicitation techniques informed by the evidence synthesis. (Paper 4)

# 2 Research papers

## 2.1 Overview of research

To address these objectives, the research was conducted in four stages, which are presented in turn. The output from each stage informed the next (Figure 2.1).

First a scoping review of the literature on elbow replacement was undertaken to describe the research methods that have been used, the outcomes that have been reported and the funding sources that have been used to conduct the research. This exploratory review sought to map the current research landscape for elbow replacement, and to identify the outcome domains currently reported by researchers.

These outcomes were then used to inform a real-time Delphi study to determine a core outcome domain set for elbow replacement, engaging the views of all key stakeholders. This sought to provide the research community with a core set of domains and time points for data collection that should be reported consistently within studies. The patient representatives were invited to identify which of the domains was the most important from their perspective, and this was designated the key outcome domain.

Armed with this insight from stakeholders a systematic review of the literature was undertaken to determine what instruments had been used for the primary outcome domain in research into elbow replacement for acute trauma, and to describe the distribution of data for the key outcome for the instruments described. This systematic review found that there was a lack of direct evidence to enable generation of probability density functions that could be used based on the published literature.

All the data that was available, and that was related to the key outcome, was used to inform an expert elicitation study to determine an informative prior probability distribution for the key outcome that could be used in a comparative study of total elbow replacement and distal humerus hemi-replacement for unreconstructible distal humerus fractures in adults.

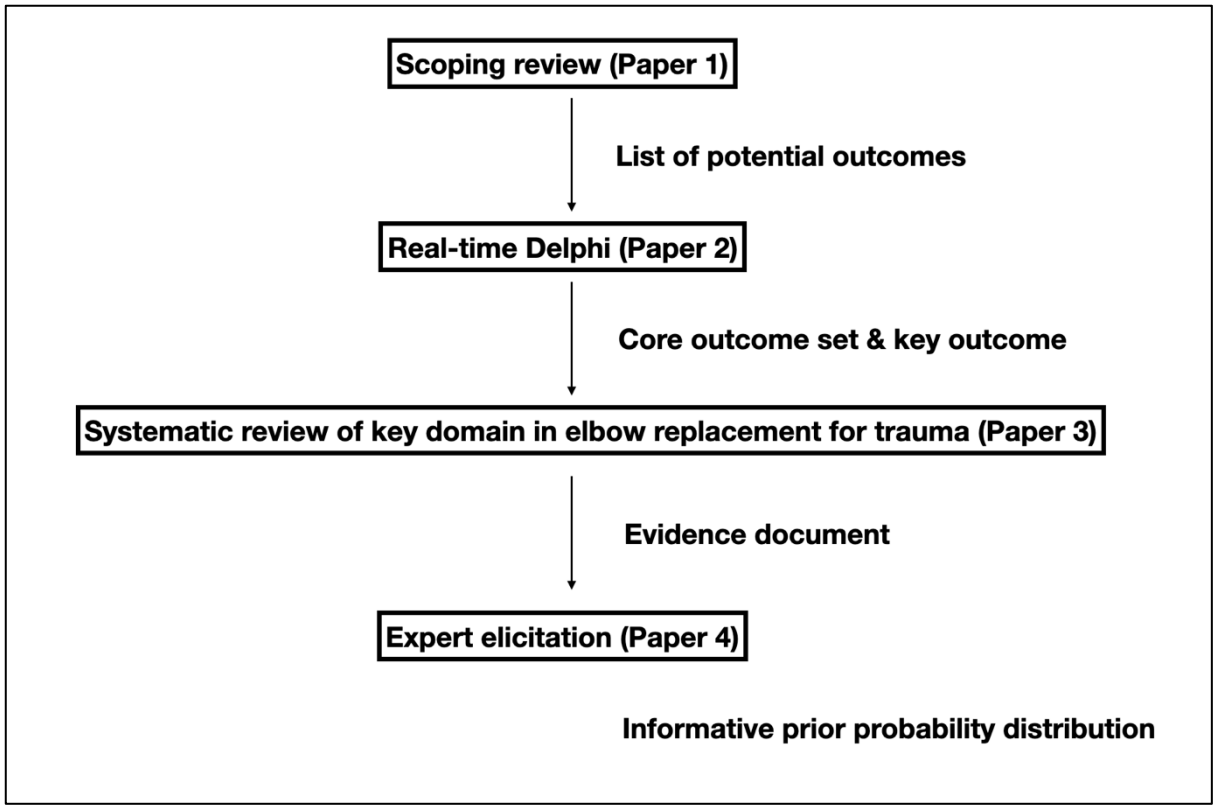



Figure 2.1 Overview of research plan

## 2.2 Patient and public involvement

The purpose of clinical research is to improve understanding to provide better care for patients. When making decisions regarding choice of treatment the role of the clinician is to outline the different treatment options with sufficient information to enable a decision to be made that is suitable for the individual.(Klifto, Klifto and Slover, 2017) To facilitate this discussion it is important that we have the best available data to share with patients, arising from high quality research. It is equally important that we involve patients and carers in research to ensure that our investigations are relevant to their needs.

The Wrightington Wigan and Leigh (WWL) Teaching Hospital NHS Foundation Trust has a Patient and Public Involvement and Engagement (PPIE) Group made of patients, carers and healthcare advocates. The group have an established background in musculoskeletal research, and for the purposes of this research additional patients and carers with personal experience of elbow disorders and interventions, in particular elbow replacement were recruited to the PPIE by the PPIE management team.

To determine the core outcome domains (Paper 2) that should be measured in elbow replacement research key stakeholders including patients, carers, and the PPIE group from WWL were involved. Ethical approval was obtained from National Health Service (NHS) research ethics committee (REC) and from the research ethics committee of the University of York. Patients who had undergone elbow replacement under my care were invited to take part in the real-time Delphi survey, alongside other key-stakeholders, to produce a shortlist of outcome domains. The PPIE group were then asked to review and comment on the short-listed outcome domains and time points for their collection and to determine the choice of key outcome. The group did not identify any missing outcome domains that were important to them, and were satisfied with the shortlisted domains. Notes recorded by an independent facilitator at the PPIE meeting are presented in Appendix C.

## 2.3 Paper 1: Elbow replacement research methods, outcome domains and instruments in clinical outcome studies: a scoping review

### 2.3.1 Publication status

Published in the Journal of Bone and Joint Surgery, 2022.

### 2.3.2 Co-author statement

**Watts, A.C., Hamoodi, Z., McDaid, C. and Hewitt, C., 2022. Elbow arthroplasty research methods, outcome domains, and instruments used in clinical outcome studies: a scoping review. The Bone & Joint Journal, 104(10), pp.1148-1155.**

**Candidate contribution: Conceived the idea; designed the study; wrote the protocol; wrote the search strategy; carried out the search, screening and data extraction; undertook the analysis; wrote the manuscript; submitted for publication.**

| Signed: | Signed: | Signed: |
|---|---|---|
|  |  |  |
| (Adam Watts) | (Catriona McDaid) | (Catherine Hewitt) |

### 2.3.3 Introduction to the paper

To gain an understanding of the current landscape of elbow replacement research it was important to undertake a scoping review of the literature on elbow replacement. This aligns with objective A of the thesis. This scoping review describes the demographics of populations studied, the study types and funding sources, and provides a comprehensive list of all the outcomes that have been reported for elbow replacement between January 1990 and February 2021 to inform the development of future research. The published manuscript is presented in Appendix A with a full list of authors.

### 2.3.4 Accepted manuscript

**Title**

Elbow replacement research methods, outcome domains and instruments in clinical outcome studies: a scoping review


Abstract

Aim

Prosthetic joint replacement of the elbow including total elbow replacement, hemi-replacement, radial head replacement and radio-capitellar replacement, are rare procedures. This scoping review aims to map current research to inform the development of future research.

Materials and Methods

A scoping review was undertaken adhering to the Joanna Briggs Institute (JBI) guidelines using Medline, Embase, CENTRAL and trial registries limited to studies published between 1st January 1990 and 7th February 2021. Endnote software was used for screening and selection and was limited to randomised trials, non-randomised controlled trials, prospective and retrospective cohort studies, case-control studies, analytical cross-sectional studies and case series of ten patients or more reporting clinical outcomes of elbow replacement. The results

are presented as frequency counts of types of studies reported, sample size, length of follow up, clinical outcome domains and instruments used, funding sources and a narrative review.

Results

362 studies met the inclusion criteria. The majority were of total elbow replacement (246, 68%), followed by radial head replacement (100, 28%), distal humerus hemi-replacement (11, 3%) and radio-capitellar replacement (5, 1%). Most studies were retrospective (326, 90%) and most were observational (315, 87%).

The median study sample size for all implant types across all studies was 36 implants. The median length of follow up for all study types was 56 months. A total of 583 unique outcome descriptors were used that were categorised into 18 domains. A total of 105 outcome instruments were used to measure 39 outcomes.

Discussion

This review has found the majority of published research into elbow replacement consists of retrospective observational studies with small sample sizes and short follow up. A large number of outcome descriptors were used with a high number of different outcome instruments employed indicating a need to define a core outcome set for elbow replacement.

## Introduction

Prosthetic joint replacement (elbow replacement) of all or part of the elbow has been performed routinely for a number of indications since the cemented replacement design of Dee in the 1960's, including inflammatory arthropathy, osteoarthritis, acute trauma, trauma sequelae, instability and (rarely) tumour.(Dee, 1975) The term elbow replacement is often taken to mean total elbow replacement, that is replacement of the distal humerus and proximal ulna with or without the radial head, but also includes distal humerus replacement in isolation (distal humerus hemi-replacement), replacement of the capitellum of the humerus and radial head (radiocapitellar replacement) and radial head replacement in

isolation. The purpose of elbow replacement is to relieve pain, restore function and quality of life for the patient.

Clinical practice should be underpinned by robust scientific evidence, and randomised controlled trials are viewed as the gold standard method for assessing interventions. Undertaking evaluations of the effectiveness of elbow replacement techniques and devices is challenging as the intervention is relatively rare. The National Joint Registry Annual Report for England, Wales and Northern Ireland reported a total of 834 cases of elbow replacement in the year 2019 (prior to the Covid-19 outbreak) divided between acute trauma (455 cases) and elective indications (379 cases) across 172 units and 249 surgeons.(Ben-Shlomo *et al.*, 2021) These low numbers mean that innovative approaches to the design and delivery of trials are required to ensure surgical practice is underpinned by high quality evidence.

To inform the development of future research assessing the effects of elbow replacement it is important to have an understanding of the research methods that have been employed. In addition, with uncertainty as to the most appropriate outcomes to use in evaluation of elbow replacement, work is required to determine the outcome domains that have been assessed and the outcome instruments used to support development of a core outcome set as outlined in the COMET handbook.(Williamson *et al.*, 2017) Further, it is important to map the traditional sources of funding for elbow replacement research in order to understand the limitations and opportunities available. This mapping is best undertaken by a scoping review of evidence sources to examine how research is undertaken in this field.(Munn *et al.*, 2018)

A preliminary search of Pubmed conducted on 28th December 2020 for previous systematic or scoping reviews on elbow arthroplasty or replacement identified eight relevant articles. One of the systematic reviews had been retracted leaving one scoping review and six systematic reviews.(Chen, Shao and Li, 2019; Kunutsor *et al.*, 2019; Kyriacou *et al.*, 2019; Wang *et al.*, 2019; Watts *et al.*, 2019; Kholinne *et al.*, 2020; Macken *et al.*, 2020) Two studied the clinical outcomes of primary elbow replacement, two compared the clinical outcome of radial head replacement with osteosynthesis, one analysed the trends in indication for total elbow replacement and one was a review of revision of infected primary

elbow replacement. The scoping review was also on the diagnosis and management of infected elbow replacement.(Watts *et al.*, 2019) An additional internet search of Google Scholar conducted on 29th December 2020 identified a further seven systematic reviews, six of which reviewed the outcome of total elbow replacement and one reviewed the causes for failure of elbow replacement.(van der Lugt and Rozing, 2004; Little, Graham and Carr, 2005; Githens *et al.*, 2014; Prkic *et al.*, 2017; Welsink *et al.*, 2017; Chou *et al.*, 2020; Samdanis *et al.*, 2020) No systematic or scoping reviews have been identified that map the research methods, outcome domains and instruments, and funding sources used in elbow replacement research.

We therefore undertook a scoping review to identify and map the research methods, domains and outcome instruments used in elbow replacement research of clinical outcomes to inform future research in this field and to describe the sources of funding used for published elbow replacement research.

This scoping review addressed the following research questions:

a) What research methods are used to study the clinical effectiveness of elbow replacement surgery?
b) What outcome domains are assessed and which instruments have been used to evaluate clinical outcomes of elbow replacement surgery?
c) What funding sources are identified in clinical outcome studies of elbow replacement?

## Methods

The protocol was developed in accordance with the JBI guidelines(M. Peters *et al.*, 2020) and is available online at Open Science Framework (https://osf.io/t6qyh/)

### *Eligibility criteria*

Studies of adults with a diagnosis of inflammatory arthropathy, osteoarthrosis, post-trauma sequelae, and acute trauma undergoing primary elbow replacement were eligible for inclusion. Reports of replacement for tumour or other rare indications were excluded. In-

vitro studies and studies of surgical approaches, biomechanics, health economics and revision elbow replacement were excluded. Studies of populations with heterogeneous diagnoses were included as long as at least 90% of the participants had one of the eligible diagnoses.

The context for the scoping review was all primary elbow replacement studies published between 1st January 1990 and 7th February 2021 and any trials on international registries that met the inclusion criteria. The types of evidence included were randomised trials, non-randomised controlled trials, prospective and retrospective cohort studies, case-control studies, analytical cross-sectional studies and case series of ten patients or more published in the English language. Review articles, surveys, case reports and conference abstracts were excluded.

*Search strategy*

An initial limited search of Medline was undertaken to perform an analysis of text words in titles and abstracts and index terms to inform the full search strategy. A full search was conducted, with support from an information specialist on 7th February 2021 of Medline, Embase, and CENTRAL using the terms "Elbow Prosthesis", "Arthroplasty, Replacement, Elbow", "Hemiarthroplasty", "Radial head arthroplasty or replacement" and "Radiocapitellar arthroplasty or replacement". This was adapted for the other databases. A search was conducted of the reference lists of reports selected for inclusion in the review to identify any additional studies. The reference lists of prior reviews in elbow replacement were also searched. The ISRCTN Registry and Clinicaltrials.gov websites were reviewed to identify any ongoing trials meeting the inclusion criteria. No search of grey literature was undertaken. The full search strategy for Medline is included in table 1.

*Evidence selection*

Endnote software Version X9; Clarivate Analytics, Philadelphia, PA, USA was used for management of the results of the search. Duplicates were excluded before initial screening based on title and abstract was undertaken by two reviewers (AW, ZH) with independent selection of evidence based on the pre-specified inclusion criteria. The full article of

potentially relevant records were obtained and screened by two reviewers (AW, ZH) to identify eligible studies. Where there was disagreement the two reviewers reviewed the manuscripts together to reach consensus.

| | | |
|---|---|---|
| Ovid MEDLINE(R) ALL <1946 to February 04, 2021> | | |
| 1 | Elbow Prosthesis/ or Arthroplasty, Replacement, Elbow/ | 485 |
| 2 | (elbow* adj3 (arthroplast* or replacement* or hemiarthroplast* or hemireplacement*)).tw. | 1217 |
| 3 | ((radial head or capitell*) adj3 (arthroplast* or replacement* or hemiarthroplast* or hemireplacement*)).tw. | 453 |
| 4 | 1 or 2 or 3 | 1721 |
| 5 | ((interposition or osteocapsular or arthroscop*) adj arthroplasty).tw. | 430 |
| 6 | 4 not 5 | 1671 |
| 7 | exp animals/ not humans.sh. | 4787332 |
| 8 | 6 not 7 1 | 653 |
| 9 | limit 8 to yr="1990 -Current" | 1480 |

**Table 1** Full search string Medline search. (Search Conducted 7th February 2021)

*Data extraction*

Pilot testing of a customised excel data extraction tool was undertaken using 25 articles selected at random and screened by AW and ZH. A meeting was held to review the screening results to determine whether any changes needed to be made, with 75% agreement required before the data extraction tool was accepted. Full text screening of the remaining articles was conducted by AW including citation details (author/s, date, title, journal, volume, issue, pages), the country in which research has been undertaken, further details of the research methodology (RCT, non-randomised controlled trials, prospective and retrospective cohort studies, case-control studies, analytical cross-sectional studies and case series of ten patients or more), implant studied, population (diagnosis, age and sex), setting,

sample size, length of follow up (minimum, maximum, mean/median), outcome assessed, instruments used to assess outcomes and funding sources for the research.

*Analysis of the evidence*

Data were tabulated and key study characteristics described. The planned analyses were frequency counts of types of studies reported, length of follow up, domains and outcome instruments used. Where possible, findings were stratified by diagnosis (inflammatory arthropathy, osteoarthrosis, post-trauma sequelae, and acute trauma) and type of elbow replacement (total elbow replacement, distal humerus hemi-replacement, radial head replacement and radio-capitellar replacement). Outcomes were recorded verbatim from source and categorised into domains after extractions using the taxonomy described by Dodd et al.(Dodd *et al.*, 2018) For ease of reporting the most common outcomes for each domain were rationalised into a common term. For example, in the adverse events domain the outcomes "pain", "residual pain", "proximal forearm pain", "severe pain" and "post-operative pain" were recorded in the table as post-operative pain. The instruments were categorised according to the outcome that the source reported they were being used to assess. No assessment of the quality of the studies or reporting was undertaken, in keeping with guidelines for scoping reviews, as the review is not designed to inform clinical decision making.(M. D. J. Peters *et al.*, 2020) A full list of included studies and outcomes is available from the corresponding author.

## Results

The findings of the literature search are reported in a flow chart adhering to the PRISMA-ScR statement and PRISMA-S extension (Figure 1).(Tricco *et al.*, 2018; PRISMA-S Group *et al.*, 2021) From a total of 2,197 de-duplicated titles identified from the searches, 402 full text articles were reviewed, of which 40 were excluded for the reasons stated in Figure 1, leaving 362 studies for final inclusion in the scoping review.

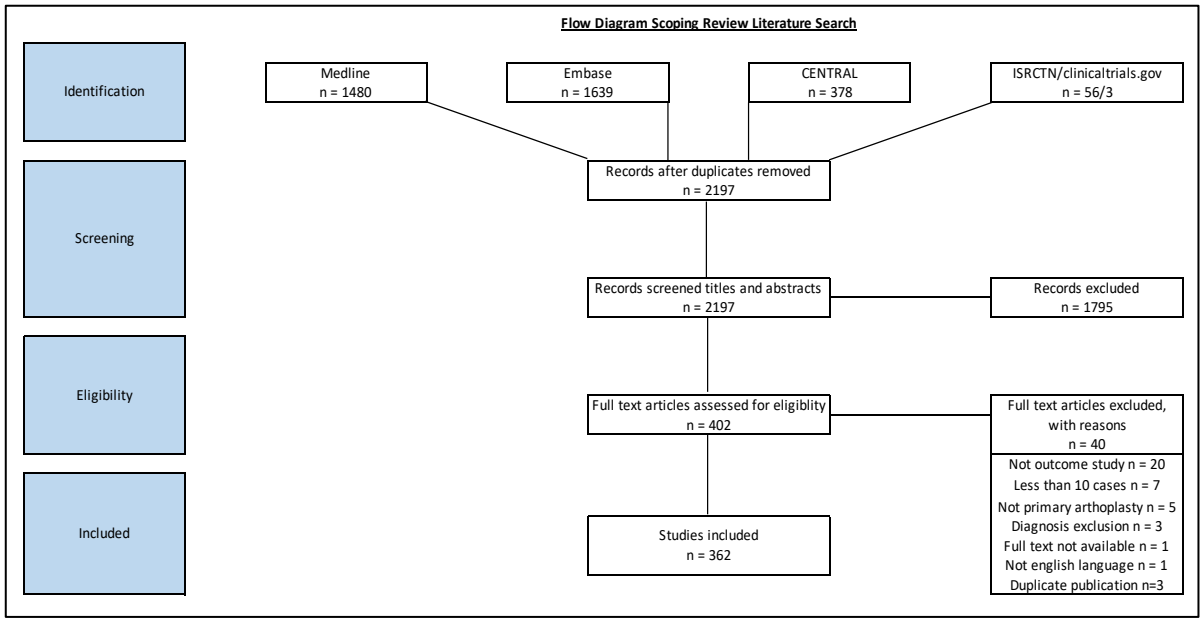



**Figure 1** PRISMA Flow Chart

*Scope of included studies*

Of the 362 studies published between 1st January 1990 and 7th February 2021 the subject of the study was total elbow replacement in 246(68%), radial head replacement in 100(28%), distal humerus hemi-replacement in 11(3%) and radiocapitellar replacement in 5(1%). Over this time period there has been an overall increase in the number of publications per annum, although a decrease in the annual number of publications on total elbow replacement has been observed since 2015 (Figure 2).

Studies were reported from 34 countries, with one international collaboration. The top ten locations for published English language elbow replacement studies by country are listed in table 2.

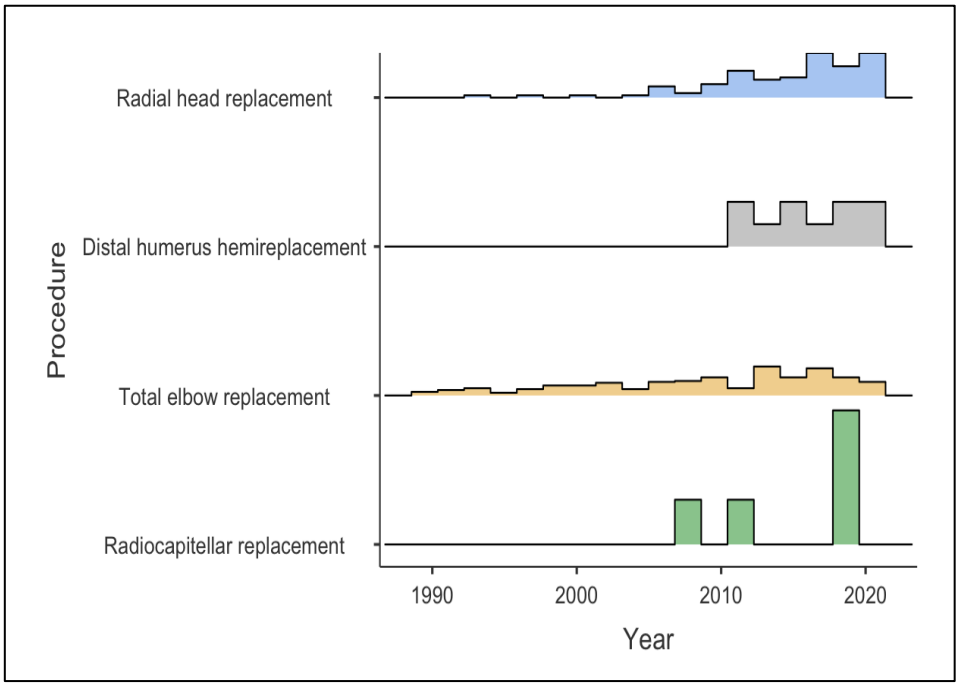



**Figure 2** Histogram of the number of publications per year categorised by type of procedure.

The majority of studies were conducted in a hospital setting (329, 91%) with few community-based studies (32, 9%). Most studies were retrospective (326, 91%) and most were observational (321, 89%). There were 23 administrative database studies, 21 of which were conducted in the USA and 8 national implant registry studies from Australia, Denmark (2 studies), Norway, Finland, Sweden and New Zealand. There were 6 prospective randomised controlled trials (RCT),(Tanaka *et al.*, 2006; McKee *et al.*, 2009; Ruan *et al.*, 2009; Chen *et al.*, 2011; Yan *et al.*, 2015; Singh *et al.*, 2019) one retrospective review of patients from a previous RCT(Dehghan *et al.*, 2019) and three RCT protocols.(Adolfsson, 2011; Al-Hamdani *et al.*, 2020; Smith, 2020)

Two RCT sources described comparison of total elbow replacement to osteosynthesis of distal humerus fractures, one reporting study results at two years and one retrospective review of the same cohort at an average of 12.5 years.(McKee *et al.*, 2009; Dehghan *et al.*, 2019)

| Country | Number | % |
|---|---|---|
| USA | 91 | 25% |
| UK | 52 | 14% |
| France | 25 | 7% |
| Japan | 19 | 5% |
| Italy | 18 | 5% |
| Netherlands | 18 | 5% |
| Canada | 16 | 4% |
| Germany | 16 | 4% |
| Sweden | 12 | 3% |
| Finland | 11 | 3% |

**Table 2** The 10 countries with the highest number of recorded publications on elbow replacement (frequency, % of total).

One source reports the outcome of total elbow replacement with two different ulna components.(Tanaka *et al.*, 2006) Three sources compare the outcome of radial head replacement to osteosynthesis for acute radial head fracture, and one compared radial head replacement to radial head excision.(Ruan *et al.*, 2009; Chen *et al.*, 2011; Yan *et al.*, 2015; Singh *et al.*, 2019) Two protocols described comparison of total elbow replacement to distal humerus hemi-replacement for distal humerus fracture, and one protocol was for a study comparing distal humerus hemi-replacement to osteosynthesis. (Adolfsson, 2011; Al-Hamdani *et al.*, 2020; Smith, 2020) None of the RCTs compare elbow replacement to a non-surgical intervention. Only one RCT had a stated source of funding which was from a commercial source.(McKee *et al.*, 2009) In five RCTs it was stated that there was no funding or conflict of interest for the trial,(Tanaka *et al.*, 2006; Chen *et al.*, 2011; Dehghan *et al.*, 2019; Singh *et al.*, 2019; Al-Hamdani *et al.*, 2020, p. ) and for four it was not stated if there was any funding or conflict of interest.(Ruan *et al.*, 2009; Adolfsson, 2011; Yan *et al.*, 2015; Smith, 2020) The median sample size by study design is given in table 3.

| | Mean | Median | Min | Max |
|---|---|---|---|---|
| **Observational studies** | 67 | 32 | 10 | 1441 |
| **Randomised trial** | 38 | 40 | 20 | 60 |
| **Administrative database** | 7819 | 1625 | 176 | 56379 |
| **Registry study** | 692 | 584 | 126 | 1457 |

**Table 3** Sample size of studies by study type.

*Properties of included studies*

The median study sample size by implant type was 17 (range 10 to 44) implants for distal humerus hemi-replacement, 20 (range 10 to 31) for radio-capitellar replacement, 32 (range 11 to 528) for radial head replacement and 41 (range 10 to 56,379) for total elbow replacement. The indication for surgery by procedure type is presented in table 4.

| | RA | OA | Post Trauma | Acute Trauma | Other |
|---|---|---|---|---|---|
| **Total elbow replacement** | 61% | 5% | 16% | 18% | 2% |
| **Distal humerus hemi-replacement** | 0% | 0% | 15% | 85% | 0% |
| **Radial head replacement** | <1% | <1% | 12% | 87% | <1% |
| **Radio-capitellar replacement** | 11% | 58% | 27% | 1% | 3% |

**Table 4** Primary diagnosis at the time of surgery by type of procedure.

Differences were found, between procedure types, in the mean percentage of female study participants and mean age of participants (Table 5).

| | Mean % female | Mean of minimum age in years (range) | Mean of maximum age in years (range) | Mean of mean age in years (range) |
|---|---|---|---|---|
| **Total elbow replacement** | 76 | 36 (5-75) | 81 (40-97) | 61 (28-85) |
| **Distal humerus hemi-replacement** | 85 | 47 (16-62) | 81 (63-90) | 67 (45-79) |
| **Radio-capitellar replacement** | 51 | 31 (25-40) | 74 (69-82) | 55 (53-61) |
| **Radial head replacement** | 46 | 23 (14-62) | 75 (50-93) | 49 (31-67) |

**Table 5** Sex and Age of patients included in studies by type of procedure.

The median of the mean length of follow up for all study types was 56 months (range 1 to 216 months). It was longest for registry studies with a median follow up of 96 months (range 67 to 126 months). The median follow-up in observational case series, which was the largest group, was 57 months (range 6 to 216 months). For randomised trials the median of the mean follow-up was 29 months (range 15 to 151 months). The median of the mean reported follow up for studies of total elbow replacement was 60 months (range 1 to 216 months), for radial head replacement 42 months (range 10 to 145 months), distal humerus hemi-replacement 35 months (range 12 to 82 months) and radio-capitellar replacement 59 months (range 23 to 100 months).

*Reported outcomes*

A total of 583 unique outcome descriptors were used across 362 included studies and were categorised into 18 domains (Table 6). Many of these outcome descriptors reported the same outcome using different terms and in 76 cases the outcome instrument was reported without specification of what was being assessed. The largest group of outcomes are categorised in the adverse events domain (311, 53%) despite the fact that complications

were not a pre-specified outcome in most studies. The physical function domain contains the second largest number of outcome descriptors (93, 16%). These can be grouped into four outcomes: function/disability, range of movement, strength and activities of daily living. Strength was often poorly defined but most commonly described an assessment of strength of elbow extension as a measure of triceps function. Radiographic appearance of the elbow was assessed by 19 separate outcomes in the musculoskeletal domain including implant alignment, congruency, fixation, success, lengthening, head size, stem size, stem positioning, implant positioning, prosthetic sizing, quality of cementing technique, bone graft integration or incorporation, cortical fit, cement mantle, valgus tilting, congruence of the proximal radio-ulnar joint, cement technique, and joint congruity.

| **Domain** | **Outcome** | |
|---|---|---|
| Adverse events (311) | Post-operative pain<br>Radiographic complications<br>Infection<br>Triceps weakness<br>Neurological<br>Effusion/Synovitis | Implant related<br>Wound problems<br>Stiffness<br>Bone problems<br>Transfusion<br>Medical complications |
| Physical function (93) | Function/disability<br>Range of movement | Strength<br>Activities of daily living |
| MSK connective tissue (37) | Radiographic appearance<br>Elbow stability | |
| Need for further intervention (34) | Re-operation<br>Implant revision/survival | |
| Nervous system (30) | Pain<br>Neurological status | |
| Delivery of care (15) | Satisfaction | |
| Social | Sport participation | |

| | | |
|---|---|---|
| function (13) | Social-psychological status | |
| Hospital resources (11) | Length of stay<br>Re-admission | Return to operating room<br>Length of surgery |
| Role function (8) | Return to work | |
| Perceived health (7) | General health | |
| Economic resource (7) | Hospital costs<br>Non-routine discharge costs | |
| Mortality /Survival (6) | Mortality | |
| Psychiatric (3) | Mental health | |
| Quality of life (2) | Quality of life | |
| Emotional well-being (2) | Well-being<br>Role emotional | |
| Cognitive function (1) | Cognitive status | |
| Personal circumstances (1) | Patient autonomy | |
| Societal carer burden (1) | Non-homebound discharge | |

**Table 6** Reported outcomes categorised by domain (number of outcomes in each domain).

Outcome instruments

A total of 105 outcome instruments were used to measure 39 outcomes, of which 26 were clinical and 13 were radiographic. The average number of instruments used per outcome where an instrument was described was 5 (range 1 to 26). The list of instruments used to assess each of the 39 outcomes is provided in table 7. For implant survival the listed

instruments are methods of analysis but are included for completeness. Some instruments are included in more than one outcome category and may be listed for outcomes and domains that may seem inappropriate, but these are taken verbatim from the included studies.

| Outcome | Instrument |
|---|---|
| Activities of daily living | Liverpool elbow score, Mayo elbow performance index, Mayo elbow performance score, UCLA Activity score, Wrightington score |
| Cognitive status | MMSE |
| Complications | Voloshin |
| Cost utility analysis | QALY |
| Deformity | Ewald scoring system, JOA score |
| Function | Andrews score, ASES score, Broberg Morrey score, DASH score, Elbex score, Elbow functional assessment, Ewald score, Inglis score, JOA score, Khatri score, likert scale, Liverpool elbow score, Mayo elbow performance index, Mayo elbow performance score, numerical rating scale, Oxford score, PREE, QuickDASH, SANE, SECEC, Souter, Stanford health assessment questionnaire, VAS |
| Global Health | EQ5D |
| Grip strength | ASES, Broberg Morrey, Dynamometer, NK hand evaluation system |
| Implant Survival | Dobbs, Kaplan-Meier, Life table, Murray |
| Joint position sense | Propriometer |
| Locomotion | Linkoping |
| Mental health | SF-12 Mental |
| Pain | ASES, Broberg Morrey, DASH, Elbow functional assessment, Ewald score, Inglis score, JOA score, Khatri score, likert scale, Mayo elbow performance score, Modified Andrews score, numeric rating scale, Oxford elbow score, Liverpool elbow score, Pain intensity scle, PREE, VAS |
| Patient Autonomy | Katz |
| Physical health | SF-12 physical |
| Quality of life | RAND-36, SF-36 |
| Radial Nerve Palsy | Hirachi, Electrophysiology |
| Range of motion | ASES, Broberg Morrey, Cassebaum, Elbow functional assessment, Ewald score, HSS score, JOA score, Inglis score, Khatri score, likert, Liverpool elbow socre, Mayo elbow performance index, Mayo elbow performance score, Modified Andrews, NK hand evaluation system, Propriometer |
| Return to Work | binary |
| Satisfaction | ASES, binary, Jungbluth, likert scale, numerical rating scale, SANE, VAS |
| Sport participation | Allain, Liverpool elbow score |
| Stability | ASES, Broberg Morrey, Elbow functional assessment, JOA score, likert, Mayo elbow performance index, Mayo elbow performance score, Modified Andrews score |
| Strength | ASES, Broberg Morrey, HSS score, Isobex, Kindeyn Dynamometer, Lido workset, Mayo elbow performance index, Mayo elbow performance score, MRC scale, likert, Liverpool elbow score |
| Success of treatment | VAS |
| Tenderness | ASES |
| Ulnar nerve function | Electophysiology, Liverpool elbow score, McGowan grade |
| | |
| Radiographic alignment | Figgie, H index, O'Driscoll, RSA analysis, Storen's line, U index, Wrightington |
| Radiographic bushing wear | Gill, Llamas, Lee, Mayo, Ramsey, Schneeberger |
| Radiographic capitellum erosion | Broberg Morrey, likert scale, Llamas |
| Radiographic heterotopic ossification | binary, Brooker, Hastings and Graham, Ilahi, likert |
| Radiographic loosening | Berschback, binary, Cil, Gill and Morrey, Goldberg, Grewal, Harris, King, likert scale, Madsen, Mayo, Morrey and Adams, Popovic, Schneeberger, Wagener, Wrightington |
| Radiographic lucency | binary, Fehringer, Gruen, Kodde, Kudo and Iwano, Morrey, Souter, Tanaka, Wrightington |
| Radiographic medial collateral ligament healing | Ultrasound |
| Radiographic osteoarthrosis | binary, Broberg Morrey, Kellgren Lawrence, Knirk and Jupiter, Lindenhovius, likert scale |
| Radiographic overstuffing | Rowland, Van Riet |
| Radiographic quality cement technique | Schneeberger |
| Radiographic radial head prominence | Doornberg |
| Radiographic synovitis | Forster |
| Radiographic ulna wear | Smith and Hughes |

**Table 7** Outcome instruments used in included studies (no assessment of psychometric properties has been conducted). *(ASES – American Shoulder and Elbow Score; DASH – Disabilities of the Arm Shoulder and Hand; HSS – Hospital for Special Surgery; JOA – Japanese Orthopaedic Association; PREE – Patient Rated Elbow Evaluation; SANE – Single Assessment Numeric Evaluation; VAS – visual analogue score; RAND-36 – early version of SF-36; MMSE – Mini-Mental State Evaluation)*

Discussion

Randomised controlled trials (RCTs) are considered the gold-standard for evaluation of healthcare interventions but in orthopaedic surgery the number of RCTs undertaken is consistently low. A systematic review of orthopaedic literature found that only 20% of procedures had at least one low risk of bias RCT supporting the operative intervention when compared to non-operative treatment.(Lim *et al.*, 2014) We did not identify any for elbow replacement. Many different reasons have been proposed for this and the cause is likely to be multifactorial.(Solomon and McLeod, 1995; McCulloch, 2002) One of the challenges of performing a randomised trial in some areas of orthopaedics where the condition is rare is the feasibility of recruiting sufficient participants to address the research question.

The European Union defines rare diseases as those with a prevalence of less than 50 per 100,000 population for the purposes of orphan drugs, where normal economic models prevent research and development.('Orphan Drug Act', 1983) Elbow replacement is a "rare" procedure with an annual incidence estimated from Scottish data as 1.4 per 100,000 population and the same economic and scientific challenges apply, with low investment in research and barriers to conventional frequentist approaches to investigate clinical outcomes due to the feasibility of randomised controlled trials (RCTs). (Jenkins *et al.*, 2013)

This scoping review has found that the literature on elbow replacement consists largely of unfunded retrospective observation studies with small sample sizes. It is interesting to note that the number of studies on total elbow replacement has declined in recent years. The cause of this is unknown, but it may be due to the contracting usage resulting from improved medical management of inflammatory joint disease. Only 9 prospective randomised controlled trials were identified, three of which are protocols of ongoing trials rather than reports of trial results. The sample size of these RCTs ranges from 20 to a maximum of 60 participants. Only one randomised trial had a stated source of funding, and that funding was from a commercial body. This review has not identified any RCTs comparing joint replacement to non-operative treatment in the elbow. It is beyond the scope of this review to determine the causes for the quality of the evidence but there remain many questions regarding elbow replacement that require robust unbiased scientific

investigation. Researchers designing future RCTs in elbow replacement will need to explore solutions for investigation of rare diseases with multi-centre and possibly international collaborations to ensure sufficient statistical power and exploration of alternative trial designs. The moral hazard of enrolment of patients into underpowered trials could be avoided through a planned meta-analysis of duplicate studies.(Halpern, 2002) However, this would require co-ordination and establishment of a musculoskeletal rare intervention trials registry and a mechanism to ensure data sharing and individual patient data (IPD) meta-analysis.(Medley *et al.*, 2021; Rydzewska, Stewart and Tierney, 2022) Alternative trial designs have been explored for investigation of rare diseases including crossover designs, N-of-1 trials and adaptive designs, but most are not suitable for surgical research due to the irreversible nature of surgical interventions and the time period required to determine the outcome of the intervention.(Gupta *et al.*, 2011; Cornu *et al.*, 2013) Bayesian analysis methods, however, could be used to exploit all available data to strengthen the findings from smaller RCTs.(Tan, 2003) Appropriate funding will be required to ensure the successful delivery of these more complicated investigations and qualitative research may be needed to understand the barriers and facilitators within the elbow replacement community that may be affecting research practice.

Alternative approaches to tackling the paucity of information may mean pooling data from multiple sources. The combination of data for meta-analysis is hindered, however, by the diversity of outcome domains and instruments used to measure outcomes. The 583 individual descriptors identified in this scoping review for clinical outcomes of elbow replacement can be rationalised to 41 outcomes over 18 domains using the taxonomy adopted by the COMET initiative, however this is still a large number of domains and it is unclear which of these might be considered most important by patients, carers and treating clinicians.(Dodd *et al.*, 2018) An average of five instruments have been used for the 39 outcomes where an outcome instrument has been described. This review has not sought to analyse the psychometric properties of the instruments used but rather to map the domains, outcomes and instruments without an assessment of quality or validity. Once core outcomes have been defined it will be necessary to undertake an assessment of relevant instruments to determine if any meet the criteria of truth, discrimination and feasibility.(Boers *et al.*, 2014)

This review has highlighted a clear need to define the core outcome domains for elbow replacement research that can then lead to the development of a set of core outcome instruments.

### 2.3.5 Concluding remarks

This scoping review has found that current literature on elbow replacement consists of retrospective case series studies, that are likely to have a high risk of bias inherent in case series and retrospective observational studies. The literature does not allow us to draw inference regarding the relative treatment effect of one intervention, alternative intervention, or control. More high quality randomised clinical trials are needed in the field of elbow replacement, particularly to address the question regarding the superiority of hemi-replacement over total elbow replacement for acute trauma in adults. There is no consensus regarding the outcomes that should be assessed in trials of elbow replacement. There is a clear need for a defined core outcome domain set for elbow replacement research, that reflects the views of stakeholders, so that future research data can be easily more compared between studies or combined in meta-analysis. The next chapter reports the rationale, methods and results of a real-time Delphi study undertaken to define core outcome domains in elbow replacement.

.

## 2.4 Paper 2: Core Outcome Domains in Elbow Replacement (CODER)

### 2.4.1 Publication status

Accepted for publication in the Journal of Bone and Joint Surgery in 2024 (in press).

### 2.4.2 Co-author statement

**Watts A.C., McDaid C., Hewitt C., Bateman M., Evans J.P., Higgs D., Hughes B., Luokkala T., Smith C, Uppal E. Core Outcome Domains in Elbow Replacement (CODER). The Bone & Joint Journal (in press)**

**Candidate contribution: Conceived the idea; designed the study; wrote the protocol; developed the online questionnaire; undertook the analysis; wrote the manuscript; submitted for publication.**

| Signed: | Signed: | Signed: |
|---|---|---|
|  |  |  |
| (Adam Watts) | (Catriona McDaid) | (Catherine Hewitt) |

### 2.4.3 Introduction to the paper

The scoping review of the literature (Paper 1) identified a lack of consensus about what outcomes should be measured when undertaking clinical assessment of elbow replacement interventions. The following paper addresses objective B, to define the core outcome domains for elbow replacement research and identify a key outcome that is important to patients. The full published manuscript is presented in Appendix A, including a list of all co-authors and their roles.

Clinical participants (surgeons and physiotherapists) were recruited through advertisement of the study on Cubitus Mundi, a network of elbow specialists on the social media application Whatsapp (Whatsapp Ltd, Dublin, Ireland), that has 1023 members from around the World. All respondents were sent links to complete the survey. Patient participants were identified from the National Health Service (NHS) clinical practice of the author (AW) at Wrightington Hospital, UK. Seventy patients were identified and contacted by email with ethics committee approval. Patient participants were asked to register their interest on Forms (Microsoft Teams) using an NHS account and to provide consent for use of their contact details for the purpose of e-mailing a personalised link to the survey, so that their contact details did not have to be shared with a third party.

### 2.4.4 Accepted manuscript

## Title

Core Outcome Domains for Elbow Replacement (CODER).

Abstract

Background

A review of the literature on elbow replacement found no consistency in clinical outcome measures used to assess the effectiveness of interventions. The aim of this study was to define core outcome domains for elbow replacement.

Methods

A real-time Delphi survey was conducted over four weeks using outcomes from a scoping review of 362 studies on elbow replacement published between 1990 to 2021. 583 outcome descriptors were rationalised to 139 unique outcomes. The survey consisted of 139 outcomes divided into 18 domains. Readability and clarity of the survey was determined by a survey advisory group including a patient representative. Participants were able to view aggregated responses from other participants in real-time and to revisit their responses as many times as they wished during the study period. Participants were able to propose additional items for inclusion. A Patient Public Involvement and Engagement (PPIE) panel considered the consensus findings.

Results

Forty-five respondents completed the survey. Nine domains were identified as core mandatory domains; “return to work or normal daily role”, delivery of care measured in the domains “patient satisfaction with the outcome of surgery” and “would the patient have the same operation again”, “pain”, “revision”, “elbow function”, “independence in activities of daily living”, “health related quality of life” and “adverse events”. “Elbow range of movement”, was identified as important by consensus but was felt to be less relevant by a Patient Public Involvement and Engagement (PPIE) panel. The PPIE panel unanimously stated that pain should be used as the primary outcome domain.

Conclusions

The study has defined core outcome domains for the clinical outcomes of elbow replacement obtained by consensus from patients, carers and healthcare professionals. Pain may be used as the primary outcome in future studies where appropriate. Further work is required to define the instruments that should be used.

## Introduction

Where quality of life is significantly impacted by elbow pain, or limited function, an elbow joint replacement has been a surgical treatment option for over five decades.(Dee, 1975) Initially used solely for the treatment of inflammatory joint disease, the indications have expanded to include acute trauma, trauma sequelae, degenerative joint diseases and

reconstruction after tumour resection. Elbow replacemet can be categorised as: total elbow replacement (replacement of both sides of the ulno-humeral joint) with or without replacement of the radial head; distal humerus hemi-replacement that replaces the trochlea and capitellum; radial head replacement in isolation; or radio-capitellar replacement, replacement of the radial head and capitellum. A recent scoping review of the published literature on elbow replacement has identified differences in the indications and demographics of patients receiving the different types of joint replacement.(Watts *et al.*, 2022)

Measuring the effect of these interventions and the possible harms, is important to determine their place as a treatment for the clinical conditions outlined. However, to enable comparison of one intervention to another and between different settings it is important to have consistency in what is measured, and to ensure that what is measured is relevant to users. In the 362 clinical outcome studies on elbow replacement, reported or registered between 1st January 1990 and 7th February 2021, a total of 583 unique outcome descriptors were identified. In addition to the heterogeneity in what was measured, there was no consistency in how the outcomes were measured with 105 outcome instruments described for 39 of the outcomes and no clearly defined tool for the remaining outcomes.(Watts *et al.*, 2022) This supports a previous study that identified a need for a core outcome set for elbow replacement.(Riedel and Beaton, 2013)

Determination of what is measured, the core outcome set, requires an understanding of the healthcare outcome domains that are important to the patient, their care givers and the clinicians with whom they engage to help in the management of their condition. The authors are not aware of any studies undertaken to elicit a consensus for the core outcome domains from key stakeholders for elbow replacement. A search of the COMET database (https://www.comet-initiative.org/Studies) conducted on 19th June 2022 identified no other studies on this topic.

This study sought to reach a consensus on the core outcome domains using a real-time Delphi (RT-Delphi) process with internal pilot. The protocol was developed with consideration of the COS-STAD and COS-STAP consensus statements for standards of core

outcome set survey design.(Kirkham *et al.*, 2017, 2019) The research does not attempt to determine the best outcome instruments to measure the domains.

## Aims

A. To identify the core outcome domains for the clinical outcome of elbow replacement from key stakeholders including patients who have undergone elbow replacement, their care givers and healthcare professionals engaged in providing guidance on treatment.
B. To determine the best time points to capture outcome measurements after elbow replacement.

## Methods

A Real-time Delphi (RT-Delphi) consensus survey was conducted online using Surveylet software (Calibrum, USA). This is an online survey tool that provides anonymous survey completion, with participants able to score outcome domains, view the aggregate scores of previous respondents and provide text to explain why they may take a different view to the aggregate response to inform other respondents. All respondents can return to the survey as many times as they like while the survey is open to review or amend responses. This approach replaces several rounds of survey used in Delphi studies, that can be associated with attrition and respondent fatigue.(Trevelyan and Robinson, 2015)

### *Scope of survey*

#### Health conditions

The survey investigated consensus opinions for the core outcome domains following elbow replacement for the treatment of inflammatory joint disease, degenerative joint disease, acute trauma and trauma sequelae. Elbow replacement for reconstruction following tumour resection were not included as the outcome priorities may be different due to the burden of disease.

Population

The COS will apply to adult patients over the age of 16 years undergoing elbow replacement for the health conditions outlined.

Interventions

Elbow replacement including total elbow replacement with or without a radial head replacement, distal humerus hemi-replacement, radial head replacement alone or radio-capitellar replacement, also known as lateral resurfacing elbow replacement.

Timing

The survey sought consensus on the time points to capture the core domains, once defined.

*Context for application*

The core outcome domain for elbow replacement (CODER) set is being developed to be used in research to assess the effects of elbow replacement procedures and in monitoring of individual patient outcomes as part of implant surveillance strategies.

*Stakeholders*

Survey advisory group (SAG)

The survey group consisted of the chief investigator, a patient with experience of elbow replacement, four surgeons and two physiotherapists from different institutions and countries with experience of elbow replacement and research into elbow replacement. The SAG attended study management meetings online over the duration of the study.

Survey Participants

A minimum of 30 key-stakeholders including adults over the age of 18 years who have had an elbow replacement (total elbow replacement, distal humerus hemi-replacement, radio-capitellar replacement or radial head replacement), other “expert” patients familiar with research processes, carers of patients with elbow joint replacements and healthcare professionals who assist in the management of elbow problems (including surgeons,

physicians and physiotherapists). Participants were drawn from countries around the world. Ethical approval was obtained to approach patients who have had an elbow replacement to invite them to participate in the survey. Patient participants and their carers were recruited by the Chief Investigator through a letter of invitation including the patient information sheet. Patients without the mental capacity to participate in an online survey were excluded. Surgeons and allied health professionals were recruited through research networks and social media.

*Ethical approval*

Ethical approval was obtained from the Health Sciences Research Committee, University of York (HSRGC/2022/258/G, 30/09/2022) and, for patient recruitment, through the integrated research approval system (IRAS ID 320243, 30/03/2023) from the Health Research Authority in the UK. Participants were invited to join the study by letter and asked to register their interest and provide consent for use of their anonymised responses using the online registration tool.

*Information sources*

The list of outcomes presented in the RT-Delphi were taken from a scoping review of elbow replacement clinical outcomes literature published between 1st January 1990 and 7th February 2021.(Watts *et al.*, 2022) This identified 583 unique outcome descriptors from 362 publications. To make the process more feasible for participants two members of the research team (AW, ZH) condensed the outcome descriptors and combined outcomes that are recording the same thing but using different terminology under a common term. For example “stability”, “instability”, “implant stability”, and “instability - subjective” were described as “elbow stability”. The two investigators conducted this separately and then discussed any discrepancies to reach agreement. No possible outcomes were removed in this process. Where appropriate, outcome domains were grouped into categories to reduce the question burden for participants. For example, mortality was used as an outcome category, with in-hospital mortality, 30-day mortality, 90-day mortality as domains within that category. The clinical members of the SAG were asked to consider the readability and

clarity of each outcome to determine if any additional text was required to explain it to participants in the RT-Delphi. The patient member of the SAG was then asked to review the description of each outcome to ensure that terms were appropriate and understandable to a lay-reader. The members of the SAG piloted the online RT-Delphi tool prior to publication of the survey.

*Consensus process*

Participants in the CODER survey registered their interest and gave electronic consent for their responses to be used using a Microsoft Forms questionnaire in a host institution Teams account (Microsoft corp). A personalised and anonymised link was sent to consenting participants from a host institution email account. Participants were able to return to the RT-Delphi at any time and as many times as they wished during the four-week period that the study was open, to review and amend their responses. A reminder email was sent after two weeks to encourage engagement and for those who had participated to return to review their responses.

Participants were asked to record their sex, age in ten-year age bands and role (patient, carer, general practitioner, commissioner, allied health professional or surgeon).

In keeping with the OMERACT handbook version 2.1, participants were asked to rate each outcome on a scale from 1-9 where nine is the most important and one the least important outcome in the participant's view.('The OMERACT Handbook', 2021) Participants were provided with a dashboard of the responses for that outcome from prior participants (using the responses from the SAG as the baseline response at the start of the survey) and the dashboard was updated in real time by the online software. Where the researchers had grouped outcomes into categories (as in the mortality example) the respondents were asked to rate the outcome category question and their response determined whether they were asked to rate the domains within that category. Participants scoring between 7 and 9 on any of the outcome categories were asked to rate each domain falling within that category using the same scale. If the participant scores less than 7 for any outcome categories they were not asked to rate the domains within that category and were automatically taken to the next

domain question. This was intended to reduce respondent burden for participants. Participants scoring more than 3 points above or below the aggregate score for each outcome were given the option of recording the reason for their response and their anonymised response was made available for other participants to view. All participants were given the opportunity to suggest additional outcomes not already listed if they felt that they were important to the assessment of the outcome of elbow replacement, these would have been reviewed by the SAG if it was felt that they were not covered by other outcomes already included, but no additional outcomes were suggested by participants.

Consensus Meeting and final PPI engagement

All participants were invited to attend an online meeting to review the outcomes that passed the threshold for inclusion and to reach consensus on the domains into which these could be grouped to produce a final list of included core outcome domains. The results of the survey were presented and each outcome achieving the threshold was presented to the group one at a time for consideration. For each core outcome domain attendees were given timing options to select (baseline, 6 weeks, 3 months, 6 months, 12 months, 18 months and 24 months) and the option to suggest alternative time points for the collection of the outcome.

To ensure the patients' voice was well represented in the consensus process a separate face-to-face patient and public inclusion and engagement (PPIE) meeting was held, following the survey period, to review the core outcome domains and time points, and to invite their comment.

*Analysis*

A descriptive analysis of the responses was undertaken. All outcomes given a score 7 or higher by a minimum of 70% of participants were included in the final consensus meeting. All outcomes were rationalised into domains by the consensus group and have been included in the core outcome domain set. Outcomes were categorised into domains using the taxonomy of Dodds.(Dodd *et al.*, 2018) For the core timings for core outcome domain capture, each time point chosen by a minimum of 70% of the group has been included.

*Registration*

The CODER study was registered on the Core Outcome Measures in Effectiveness Trials (COMET) database (Oct 2022) and the protocol made available on the open science framework platform.(Watts, 2023a)

Results

Forty-five people registered to complete the survey and provided electronic consent for use of their responses. The self-reported descriptors of the respondents are given in table 1. An additional three patients and two surgeons registered interest but did not go on to complete the survey.

The four-week real-time Delphi survey identified 52 outcomes across 13 domains that reached the pre-specified threshold (70% or more respondents scoring 7/9 or higher) for consideration at the final consensus meeting (Appendix E).

| | | n | % |
|---|---|---|---|
| Sex | Female | 9 | 20 |
| | Male | 36 | 80 |
| Age | 30-39 | 12 | 27 |
| | 40-49 | 19 | 42 |
| | 50-59 | 9 | 20 |
| | 60-69 | 5 | 11 |
| Role | AHP | 6 | 14 |
| | Carer | 2 | 4 |
| | GP/Commissioner | 1 | 2 |
| | Patient | 3 | 7 |
| | Surgeon | 33 | 73 |

**Table 1** Demographic characteristics of survey participants (AHP: allied health professional, GP: General practitioner)

All 45 respondents were invited to an online consensus meeting on 13/07/2023, and 14 attended, including one patient, one carer, three allied health professionals (AHP) and nine surgeons. It was noted that where categories were selected as important, every question within that domain category reached the threshold for inclusion with the exception of "minor adverse event" under the category "adverse events". The group consensus was that outcome categories could be retained as core domains and those outcomes within the category removed without loss of fidelity. In addition, the group felt that "return to work" and "return to normal daily role" could be combined into a single domain of "return to work or normal daily role". The consensus group took the view that "patient autonomy" used in the way it had been used in the source material to assess independence with daily activities as measured by the Katz instrument, could be combined with "activities of daily living" under a single domain of "independence in activities of daily living." Whilst "Patient satisfaction with the outcome of surgery" and "Would the patient have the same operation again" both fall within the domain of "delivery of care" according to Dodd's taxonomy, it was agreed that both should be retained as they assess different aspects of satisfaction or acceptability. This resulted in ten core outcome domains that were taken to the PPIE group meeting for consideration, along with the core timings for each outcome domain to be assessed (Table 2).

| Dodd Classification[8] | Core Domain | Baseline | 3 months | 6 months | 12 months | 24 months |
|---|---|---|---|---|---|---|
| Role function | Return to work or normal daily role | | X | X | X | |
| Delivery of care | Patient satisfaction with outcome of surgery | | | | X | X |
| | Would patient have the same operation again? | | | | X | X |
| Nervous system | Pain | X | X | X | X | X |
| Need for further intervention | Revision | | | | | X |
| Physical function | Elbow function | X | | X | X | X |
| | Independence in activities of daily living | X | | X | X | X |
| | Active range of elbow movement | X | X | | X | X |
| Quality of life | Health related quality of life | X | | X | X | X |
| Adverse events | Adverse event including death | | X | X | | |

**Table 2** Core outcome domains for elbow replacement. Note "active range of elbow movement" and "six-month" time point for data capture were felt to be less relevant by the PPIE group and should be considered as important but not mandatory.

The PPIE group attended a face-to-face meeting with the chief investigator and a facilitator. The group consisted of 8 female and one male participant with personal experience of orthopaedic care. Two members of the group had personal experience of elbow replacement, one was a carer for someone who had an elbow replacement, one was a disability group representative and one member of the group is actively involved in clinical research as a lay representative. The group considered and discussed each core domain and the timings for capture of core outcomes. The group felt strongly that efforts should be made to reduce the burden of research for participants. The group therefore questioned the value of data capture at six months.

The PPIE group discussed at length whether range of movement of the elbow was important to them independent of elbow function. The majority consensus of the PPIE group (>70% approval) was that what mattered to them was the function of the elbow and the ability to undertake tasks rather than the measured range of elbow movement, and there was agreement within the group that this was better captured by other domains such as "elbow function" and "independence in activities of daily living". The group were asked to consider if there were any outcomes that were not included that were important from their

perspective but they were happy that the list represented the key elements. When asked which of the outcomes was most important the group unanimously offered "pain" as the most important outcome for patients after elbow replacement. The group was challenged to consider alternative outcomes but were very clear that this was the primary outcome from their perspective. There was no clear consensus about the best timing for this primary outcome domain but it was felt that it should be far enough from the time of surgery for post-surgical pain symptoms to have resolved but close enough to ensure patients were not lost from follow-up.

Discussion

Inconsistency in the "what" and "how" of outcome assessment in elbow surgery research has been reported by other investigators.(Riedel and Beaton, 2013; Evans *et al.*, 2018) This heterogeneity leads to costly redundancy for investigators, makes difficult or even impossible to compare the outcomes of different studies and creates futile additional burden for patients. Defining a consensus on core outcome domains is important to increase efficiency and enable comparison of results between studies and meta-analysis.

The nine mandatory core domains defined by the CODER study, "return to work or normal daily role", delivery of care measured in the domains "patient satisfaction with the outcome of surgery" and "would the patient have the same operation again", "pain", "revision", "elbow function", "independence in activities of daily living", "health related quality of life" and "adverse events", are feasible and appear proportionate. The four mandatory core time points for data collection, baseline, 3 months, 12 months and 24 months post intervention, balance the social burden on patients and financial burden to researchers against the needs of scientific study. Additional desirable data points may provide added value to address specific research questions. The core domain set and time points defined in the CODER study do not prevent collection of additional outcome data where necessary to address research questions but should be included in future studies of elbow replacement. The domains identified for elbow replacement in this study show good concordance with other areas of upper limb orthopaedic research. The OMERACT Shoulder working group identified

pain, function, patient global – shoulder, and adverse events including death as mandatory domains for clinical trials of shoulder disorders.(Ramiro *et al.*, 2019)

Although key stakeholder groups were represented, the findings of this study are limited by the low participation in the consensus meeting, and the relatively low patient and carer representation in the Real-Time Delphi study. Extensive efforts were made to try to recruit patients and carers to the survey and a number did register interest but did not go on to complete the survey. This may be a reflection of the relatively cumbersome process required to access the survey electronically, having to complete a registration form to be sent an anonymised link through which to access the survey, which was necessary due to General Data Protection Regulation (GDPR) compliance requirements. We attempted to mitigate the imbalance of representation through engagement with a PPIE group following the online survey, and there was strong support for the domains included. The PPIE group felt that measuring range of movement provided no useful additional information. They reported that they were more interested in what could be done with the arm, which they felt could be captured in the function domain, than how much movement there was at the elbow. The PPIE group were keen to reduce patient burden and questioned the added value gained by data collection at 6 months. The optimum time point for the primary outcome was not determined but this is likely to depend on the research question under investigation. A minimum follow-up of two years is required for clinical outcomes research for joint replacement surgery by some orthopaedic journals however that does not necessarily define the primary outcome time point. Time periods for follow up beyond two years were not considered within the CODER study. Embedding trials with joint registries can facilitate long term outcome research as long as criteria for relevance, robustness, reliability and confidentiality are in place, but will not capture many of the core domains identified here without the use of linked or add-on data sets.(Mikita *et al.*, 2021) Due to the number of patient participants completing the online survey, a secondary survey within the research protocol to determine mean clinician and patient scores for adverse events has not been reported as a reliable estimate of the mean could not be calculated.(Watts, 2023a)

Conclusion

Future studies of elbow replacement should include the nine mandatory core domains at four time points agreed in the CODER study to ensure consistency between studies and facilitate comparison of research findings. Pain can be used as the primary outcome in future studies of elbow replacement where it is appropriate to do so. Further work is required to define the instruments that should be used to measure these outcome domains

### 2.4.5 Concluding remarks

Consensus from key stakeholders using a real-time Delphi methodology has identified nine mandatory core domains, "return to work or normal daily role", delivery of care measured in the domains "patient satisfaction with the outcome of surgery" and "would the patient have the same operation again", "pain", "revision", "elbow function", "independence in activities of daily living", "health related quality of life" and "adverse events", that should be captured in future studies of elbow replacement. A further domain, "range of movement", was not prioritised by patients who felt that function was more important. Outcomes should be captured at least at baseline pre-operatively and post-operatively at 3 months, 12 and 24 months. The six-month post operative time point was not prioritised by patients, citing reduced participant burden. This study has identified that patients prioritise pain as the outcome domain of greatest importance to them. This has not been reported previously, and should change our priorities as researchers when determining the primary outcome in future trials. The next paper reports a systematic review investigating current knowledge of pain outcomes following elbow replacement for acute trauma in adults, to inform future trial planning.

## 2.5 Paper 3: A systematic review, and meta-analysis of pain outcomes following total elbow replacement and hemi-replacement for unreconstructible acute distal humerus fractures in adults.

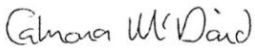
### 2.5.1 Publication status

The manuscript has been submitted to the Shoulder and Elbow Journal and has been accepted for publication.

### 2.5.2 Co-author statement

**Watts A.C., Hamoodi Z., Wright A.C, McDaid C., Hewitt C. A systematic review, and meta-analysis of pain outcomes following total elbow replacement and hemi-replacement for unreconstructible acute distal humerus fractures in adults.(Submitted to Journal of Shoulder and Elbow Surgery)**

**Candidate contribution: Conceived the idea; designed the study; wrote the protocol; wrote the search strategy; carried out the search, screening and data extraction; undertook the analysis; wrote the manuscript; submitted for publication.**

| Signed: | Signed: | Signed: |
|---|---|---|
|  |  |  |
| (Adam Watts) | (Catriona McDaid) | (Catherine Hewitt) |

### 2.5.3 Introduction to the paper

The CODER study (paper 2) identified 9 core outcome domains for studies investigating the clinical outcome of elbow replacement. Of these, pain was identified as the domain of key interest to patients and carers. To plan a comparative trial of DHH and TER for management of acute trauma in adults it is important to have a clear understanding of the pain outcomes following these two interventions. Systematic reviews that have assessed the outcomes of TER and DHH for this indication do not report any evidence synthesis regarding pain outcomes. To meet objective C of this thesis, to synthesise the current evidence related to the key outcome for patients having elbow replacement for acute trauma, a systematic review was conducted with meta-analysis of pain outcomes following total elbow replacement and distal humerus hemi-replacement for unreconstructible acute distal humerus fractures in adults.

### 2.5.4 Submitted manuscript

Title

A systematic review, and meta-analysis of pain outcomes following total elbow replacement and hemi-replacement for unreconstructible acute distal humerus fractures in adults.

Abstract

Background

Unreconstructible distal humerus fractures have been managed with total elbow replacement (TER), but there has been increasing use of hemi-replacement (DHH). Pain has been identified as the most important outcome domain for patients having elbow replacement. This study aims to systematically review the literature on reported pain outcomes in adult distal humerus fractures treated with TER or DHH.

Materials and methods

Medline, Embase and Central were searched using OVID (from January 2000 to September 2023) using the MeSH headings “Elbow Prosthesis”, “Arthroplasty, Replacement, Elbow”, “Hemiarthroplasty”, and “Humeral Fractures”. Studies in adults undergoing DHH or TER using implants that are currently available for acute closed distal humerus fractures, were

included. The primary outcome of interest was patient-rated pain measured on a numerical rating scale. Secondary outcomes were patient-rated pain scores on a Likert scale, pain measurement from function scores or clinician- and patient-rated outcome scores. A quantitative summary of reported outcomes is provided stratified by intervention type (DHH or TER) with an assessment of clinical, methodological and statistical heterogeneity.

Results

Twenty-three studies met the inclusion criteria, including one published randomised controlled trial comparing TER and DHH. There were 17 case series reporting pain outcomes for TEA; 3 reported pain using numerical rating scales (NRS pain) only and 12 used a Likert scale for pain only and 2 used both. One of these reporting pain on NRS used a different scale from other studies and was not used in the meta-analysis. There were 5 case series reporting pain outcomes for DHH; 1 using NRS pain only and 3 using Likert pain only and 1 using both scales. Meta-analysis indicated a pooled average NRS pain from 4 studies including 129 patients for TER of 1.7/10 (95% CI 0.44 to 2.99) and 1.5/10 for DHH (95% CI 0.001 to 3.56) from two studies including 39 patients. On a four-point Likert scale the pooled probability of no pain, mild pain, moderate pain, severe pain from 14 studies including 283 patients for TER was 0.75, 0.21, 0.02, 0.00 and for DHH was 0.76, 0.11, 0.12, 0.00 from 4 studies including 111 patients.

Discussion

This systematic review describes the reported data on pain outcomes following TER and DHH for treatment of distal humerus fractures. The available evidence does not enable comparison of pain outcomes between the two interventions, which should be assessed in an appropriately powered and stratified randomised trial.

Ethical approval was not required for this systematic review of published literature.

Introduction

Fractures of the distal humerus account for approximately 0.5% of all fractures in adults, with an incidence of approximately 5.8/100 000 people reported over 1 year in a high volume trauma centre.(Court-Brown and Caesar, 2006) The distribution of fractures by age is unimodal with a low risk in young adults that increases from the age of 50 years and rises markedly in those over 80 years, with a higher incidence in the female population.(Sheps, Kemp and Hildebrand, 2011) These are osteoporotic fractures and the number of cases is thought to be increasing in developed nations due to the ageing population even though the age adjusted incidence may be consistent or falling slowly.(Palvanen *et al.*, 2010; Kannus *et al.*, 2017) Distal humerus fractures are classified using the AO/OTA classification system.(Müller *et al.*, 1990)

There are a number of treatment strategies for distal humerus fractures in the older population and the choice of treatment should be decided on an individual basis through joint decision-making with non-operative treatment generally reserved for completely undisplaced stable fractures or for patients in whom the risks of surgery outweigh the benefits.(Aitken, Jenkins and Rymaszewski, 2015) In cases where surgery is indicated, osteosynthesis may be preferred where this is achievable, however in some cases, particularly in older patients with poor bone quality, consideration must be given to elbow replacement.(McKee *et al.*, 2009) The last decade has seen increasing use of hemi-replacement (DHH, replacement of only the distal humerus) as opposed to total elbow replacement (TER, replacement of the distal humerus and the proximal ulna) for fractures that are not suitable for osteosynthesis, to the point in 2022 where the number of DHH performed for acute trauma exceeded the number of TER in the National Joint Registry.(Achakri *et al.*, 2023) It is not clear which of these treatment options provides a better outcome for patients.

A search of the Prospero registry using the term “elbow”, conducted on 21st August 2023, identified one systematic review comparing surgical and non-surgical options for the treatment of distal humerus fractures in older adults (CRD42021250811) and one comparing DHH to TER for unreconstructible distal humerus fractures in the elderly

(CRD42021228329).(Burden *et al.*, 2022; Stoddart *et al.*, 2022) An additional search of Pubmed and Google Scholar conducted on 21st August 2023 identified a further systematic review comparing DHH to osteosynthesis for acute AO13C multi-fragmentary distal humerus fractures.(Nielsen *et al.*, 2022) The three reviews were assessed using AMSTAR-2, a critical appraisal tool for systematic reviews that contain randomised or non-randomised studies or both, and all were rated critically low (Appendix F).(Shea *et al.*, 2017)

In a recent real-time Delphi study to determine the core outcome domains for studies of elbow replacement, patients identified pain as the primary outcome domain.(Watts, 2023a) Current knowledge of pain outcomes from elbow replacement for distal humerus fractures is not well described. None of the systematic reviews identified in our search reported analysis of pain scores, without explanation given for excluding this outcome. Knowledge of likely pain outcomes is needed if a future trial is undertaken to compare the two interventions using pain as the primary outcome as preferred by patient representatives. It was, therefore, appropriate to conduct a systematic review of pain outcomes in this population as this is of high importance to patients and has not been included in the scope of previous reviews.

## Objective

To undertake a systematic review of the published literature to report the outcome of distal humerus hemi-replacement and total elbow replacement for the treatment of unreconstructible acute distal humerus fractures in adult patients using patient rated pain instruments.

## Methods

The protocol was developed with consideration of the PRISMA-P reporting guideline.(Shamseer *et al.*, 2015; Campbell *et al.*, 2020) The protocol was published on the Open Science Framework (d.o.i.:10.17605/OSF.IO/3TKWH) and registered on Prospero (Registration number CRD42023474360). Ethical approval was not required. The PRISMA checklist has been completed (Appendix G).

## Inclusion Criteria

The criteria for studies to be included in the review are provided in table 1.

| | |
|---|---|
| Participants | Adult human participants (18 years and over) treated with elbow replacement for acute (within 6 weeks of injury) closed distal humerus fracture classified using the AO classification system as 13A3, 13B3 or 13C type fractures.(Müller *et al.*, 1990) |
| Intervention | Participants undergoing distal humerus hemi-replacement using the Latitude prosthesis (Stryker). Studies using humeral components from total elbow replacement systems or custom distal humerus hemi-replacement implants will not be included. |
| Comparison | Participants undergoing total elbow replacement using implants that are identified by the National Joint Registry as in use in the UK at the time of the most up to date report.(Achakri *et al.*, 2023) This includes Coonrad-Morrey (Zimmer-Biomet), Nexel (Zimmer-Biomet), Discovery (Lima) and Latitude (Stryker) total elbow replacement systems. |
| Outcome (Primary) | The outcome of interest is patient reported pain scores recorded using a numerical rating scale (NRS) for pain. |
| Outcome (Secondary) | Patient rated pain scores measured using a Likert scale. Where available pain scores derived from functional outcome scores such as the Disabilities of the Arm Shoulder and Hand (DASH) – Item 24, short form version (QuickDASH) – item 9, Oxford Elbow Score (OES) – item 12 and combined physician and patient rated outcome scores such as the Mayo Elbow Performance Score (MEPS) were included if pain specific instruments had not been reported. |
| Timing | The outcome data at a minimum follow up of 3 months was included. If data was available for individual studies or patients at different time points the longest follow up data was used for analysis. |

**Table 1** Inclusion criteria for systematic review.

## Eligibility

Studies reporting in English between January 2000 and September 2023 were eligible for inclusion in the systematic review. This time period was chosen to maximise the available data and to ensure only modern implants were included. Where studies included patients with heterogeneous indications for surgery they were only included if the data on patients treated for acute distal humerus fractures were described separately.

Randomised trials, non-randomised controlled trials, prospective and retrospective cohort studies, case-control studies, analytical cross-sectional studies and case series of ten patients or more were eligible for inclusion. Review articles, surveys, case reports and conference abstracts were excluded. Studies comparing one of the interventions of interest to another intervention not eligible for inclusion in the review were included where the data for the eligible intervention were disaggregated (for example osteosynthesis).

The NRS pain score was selected as the primary outcome of interest as this could be most easily used to inform power calculations for any future randomised controlled trial. It is recognised that pain scores may change over time following arthroplasty surgery, and it is unlikely that a standardised time point for recording pain has been used. To avoid inclusion of pain from the immediate post-operative phase (wound healing) studies had to have a minimum follow-up of 3 months.

## Information sources

A database search was conducted on 24$^{th}$ September 2023 of Medline (Ovid), Embase and CENTRAL from January 2000 to September 2023. The citation list of all included studies, systematic reviews and review articles were searched for any additional articles. If trials were identified in CENTRAL for which there was no published report, investigators were contacted to ask for available data. There was no search of the grey literature. Multiple reports of the same study were collated. The study selection process is shown in the PRISMA flow chart (Figure 1).

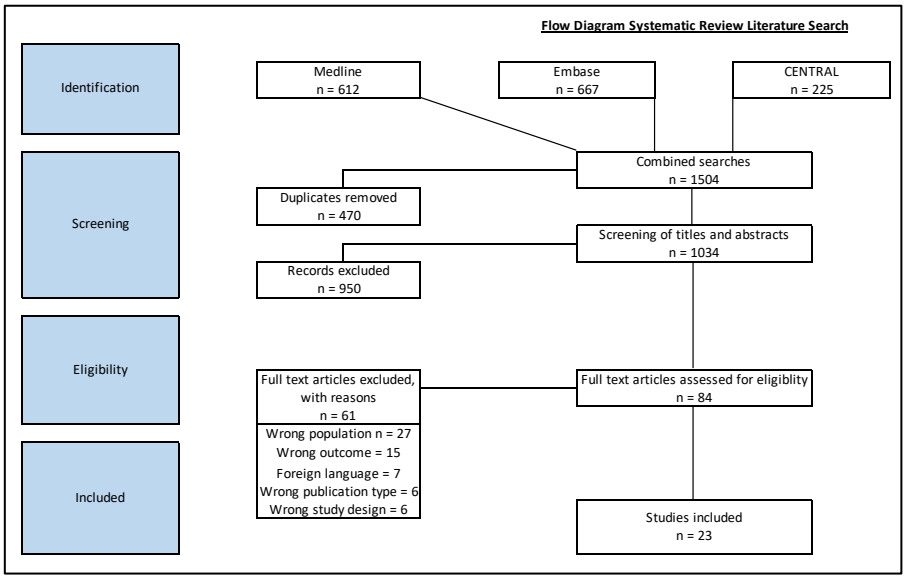



**Figure 1** Prisma flow diagram.

## Search strategy

The database search was conducted with support from an information specialist. Medline, Embase, and CENTRAL were searched using the MeSH terms “arthroplasty, replacement, elbow”, “humeral fractures” “hemiarthroplasty” and key words “hemireplacement” and “distal humerus”. The full search strategy used for Ovid MEDLINE® is detailed in Figure 2.

| | |
|---|---|
| 1. | Elbow Prosthesis/ or Arthroplasty, Replacement, Elbow/ |
| 2. | (elbow* adj3 (arthroplast* or replacement* or hemiarthroplast* or hemireplacement*or hemi-arthroplast* or hemi-replacement*)).tw. |
| 3. | ((distal adj humer*) adj3 (arthroplast* or replacement* or hemiarthroplast* or hemireplacement* or hemi-arthroplast* or hemi-replacement*)).tw. |
| 4. | 1 or 2 or 3 |
| 5. | ((interposition or osteocapsular or arthroscop*) adj arthroplasty).tw. |
| 6. | 4 not 5 |
| 7. | (trauma or fractur*).tw. |
| 8. | 6 and 7 |
| 9. | exp animals/ not humans.sh. |
| 10. | 8 not 9 |
| 11. | limit 10 to yr="2000 -Current" |

**Figure 2** Medline search strategy.

## Study records

### *Data management*

Searches were imported into Endnote (Version X9.3.3, Clarivate, Philadelphia, PEN, USA). Full text articles of short-listed citations were obtained from multiple sources and stored in Endnote.

### *Selection process*

Duplicates were excluded in Endnote before initial screening based on title and abstract using Rayyan.ai (Rayyan, Cambridge, MA, USA).(Ouzzani *et al.*, 2016) The screening was undertaken independently by two reviewers (AW, ZH) using the pre-specified inclusion criteria. Full text screening of shortlisted records was conducted independently by two reviewers (AW, ZH) with disagreements resolved through discussion.

### *Data collection process*

Data collection was undertaken using a data collection tool in Excel (Microsoft). Pilot testing of the data collection tool was not undertaken given the small number of studies included. Data were extracted independently by two reviewers (AW, ZH) and cross-referenced to identify any transcription errors which were corrected by discussion and re-examination of the source data.

Aggregate results for the post operative pain scales from included studies were extracted where available. For studies including heterogeneous groups of patients which prevented extraction of aggregate data, for example studies using a variety of different implants or including indications other than acute trauma, exploration was undertaken of extraction of individual patient data to generate aggregate results, using an excel spreadsheet, from included published studies by AW and checked for accuracy by ZH. Pre-operative pain scores were not available as these interventions are undertaken for acute trauma.

### Data items

Data items extracted included citation details, study design, the age and sex of included participants, fracture type according to AO classification, time from injury to surgery, implant studied, sample size, and length of follow up (minimum, maximum, mean/median). For instruments measuring pain with a discrete or continuous numerical scale, mean, standard deviation, minimum and maximum scores were extracted. For instruments measuring pain using categorical scales, counts and sample size were extracted to calculate proportions.

### Prioritisation

Overall priority was given to pain outcomes recorded with a patient rated pain instrument (NRS Pain, Likert scales). Pain scores derived from combined patient and physician rated outcome instruments (e.g. Mayo Elbow Performance Score) were used only where an individually reported patient pain instrument had not been reported.

### Risk of Bias

Based on initial searches and a previous scoping review of the literature it was anticipated that the majority of the evidence available would be case series.(Watts *et al.*, 2022) Risk of bias assessment was undertaken using the JBI Critical Appraisal Checklist for Case Series.(Munn *et al.*, 2020) The included randomised controlled trials was assessed using the Cochrane ROB-2 tool.(Sterne *et al.*, 2019) Each study was assessed independently by two reviewers (AW,ZH) and where there were differences in rating, agreement was reached through discussion.

### Data extraction and synthesis

The data synthesis was conducted in line with the Synthesis Without Meta-analysis (SWiM) guidelines that are intended to guide the conduct of systematic reviews of quantitative effects where meta-analysis of effect size is not possible due to the nature of the available data.(Campbell *et al.*, 2020) The data synthesis aimed to obtain an estimate of the pain

outcome for each treatment without comparison between interventions, due to the expected paucity of comparative studies.

### Assessment of clinical and methodological heterogeneity

Clinical heterogeneity included a comparison of patients age, sex and AO classification between included studies. The assessment of methodological heterogeneity was based on study design and risk of bias or critical appraisal assessment.

### Primary Outcome: Numerical rating scale for pain

The primary outcome was the NRS pain score. Mean and standard deviation values were extracted where available. Where the mean and standard deviation was not available it was imputed from the median, minimum and maximum values, and sample size as described by Hozo et al.(Hozo, Djulbegovic and Hozo, 2005)

### Secondary Outcome: Likert pain scale

A four-point scale was used for analysis as this is a commonly used scale; none, mild, moderate, severe. Data transformation was conducted for studies that use a scale with a different number of pain descriptors. For scales that reported a five-point scale: “none”, “mild”, “moderate”, “severe” and “extreme”, the “severe” and “extreme” categories were combined into “severe”.

### Meta-analysis of pain outcomes

Means were combined using standard formulae to provide an estimate of the average weighted mean NRS pain score for DHH and TER separately.(‘Cochrane Handbook for Systematic Reviews of Interventions version 6.4’, 2023) A meta-analysis of single average weighted means was conducted using a random effects model in R (Version 2023.06.2) Metamean package that uses the inverse variance method for pooling, for each intervention without comparison of interventions. Statistical heterogeneity was assessed using the Q statistic, degrees of freedom and p-value, $tau^2$ as a measure of variance, $I^2$ statistic to quantify the proportion of variance due to true heterogeneity and the H-statistic as a

measure of excess of Q over degrees of freedom, where homogeneous treatment effects within a meta-analysis would give a value of 1 and values greater than 1 indicate a degree of heterogeneity in the data.(Higgins and Thompson, 2002) Categorisation from the Cochrane Handbook was used, where an $I^2$ value of 0% to 40% might indicate the proportion of variance attributable to heterogeneity is not important, 30% to 60% may represent moderate importance, 50% to 90% may represent substantial importance and 75% to 100% considerable importance.('Cochrane Handbook for Systematic Reviews of Interventions version 6.4', 2023) It should be noted that the statistical tests for heterogeneity can be misleading where the number of included studies is low.(Hardy and Thompson, 1998)

For the Likert pain score, proportions falling within each category were combined in a random effects meta-analysis in R (Version 2023.06.2) Metaprop package, with the variances stabilised using the Freeman-Tukey double arcsine transformation. Whilst this data could be split into (no pain)/(pain) or (no pain, mild pain)/ (moderate pain, severe pain) these divisions are arbitrary and there is no clear way to choose between the approaches. The authors felt that it will be more informative to clinicians to synthesise an estimate of the weighted proportions within each category (no pain, mild pain, moderate pain and severe pain).

## Results

### Included studies

A total of 1034 abstracts were identified from searches after duplicates were removed. Nine-hundred and forty were excluded after screening of titles and abstracts. Eight-four full texts articles were reviewed and 23 studies met the inclusion criteria (Appendix H). The reasons for exclusion are provided in Figure 1.

Two randomised controlled trials were identified that compare DHH to TER for acute unreconstructible distal humerus fractures. One pilot study identified from CENTRAL has not yet been published.(Smith, 2020) The authors were contacted and kindly shared their data but it did not meet the criteria for inclusion in this systematic review due to the small

sample size and was excluded. The other published trial reported no difference in pain or function outcomes between the two interventions.(Jonsson *et al.*, 2023) The sample size was 40 patients and there is some uncertainty regarding the findings due to a potential risk of bias.(Jonsson *et al.*, 2023) The TER and DHH data from this trial was extracted separately and included in the meta-analysis of proportions.

One case series met the inclusion criteria measuring pain in patients having a TER for acute distal humerus fracture on a numerical rating scale, but the scale was from 1-10 which meant the results could not be included in the meta-analysis with other NRS pain scales measuring pain from 0-10.(Schiavi *et al.*, 2020) In this study 12 female participants (mean age 79 years) with AO 13C acute distal humerus fractures treated with a Coonrad-Morrey TEA had a median pain score 3.2/10 (range 1 to 7) at an average follow up of 91 months (range 60 to 120 months). There were 16 further case series reporting pain outcomes of TEA.(Ali, Shahane and Stanley, 2010) (Antuña *et al.*, 2012; Baik *et al.*, 2020) (Gambirasio *et al.*, 2001; Garcia, Mykula and Stanley, 2002; Frankle *et al.*, 2003; Kamineni and Morrey, 2004; Jost, Adams and Morrey, 2008; Prasad and Dent, 2008; Sørensen, 2014; Gallucci *et al.*, 2016; Pogliacomi *et al.*, 2016; Prasad, Ali and Stanley, 2016; Barco *et al.*, 2017; Celli *et al.*, 2021; Strelzow *et al.*, 2021) There were 5 case series reporting pain outcomes of DHH.(Nestorson *et al.*, 2015; Al-Hamdani *et al.*, 2019; Rotini *et al.*, 2020; Celli *et al.*, 2022; Dirckx *et al.*, 2023) The demographic details, fractures types and length of follow up are provided in table 2. For all case series included in the meta-analysis, pain data was only available for one time point (the last follow up) and this data was used in the analysis.

### Risk of Bias of Included Studies

The demographic data suggest similar age cohorts and similar proportion of female patients, however there was heterogeneity in the included fracture types. The methodological risk of bias of the included randomised trial was assessed as high using the RoB-2 tool (Appendix I). For case series the JBI Critical Appraisal Checklist for Case Series was used (Figure 3). Whilst most of the case series met the majority of the checklist criteria, case series are at risk of bias due to their uncontrolled design. In addition, we found that there was uncertainty

regarding complete inclusion of cases and a lack of reporting of the presenting sites demographic information, that could increase the risk of bias of included studies.

| Author and Year | Sample size | Fracture type (AO classification) | Implant | Mean age in years (range) | % female patients | Mean follow up in months (range) |
|---|---|---|---|---|---|---|
| **Total elbow arthroplasty** | | | | | | |
| Gambirasio 2001* | 10 | 13B1/13C | Coonrad-Morrey | 85 (57-95) | 100 | 18 (12-34) |
| Garcia 2002* | 16 | 13A3/13B3/13C3 | Coonrad-Morrey | 73 (61-95) | 75 | 36 (12-66) |
| Frankle 2003* | 12 | 13C | Coonrad-Morrey | 72 (65-88) | 100 | 45 (24-72) |
| Kamineni 2004 | 43 | 13A/13B/13C | Coonrad-Morrey | 69 (34-92) | 72 | 84 (24-180) |
| Prasad 2008* | 15 | 13A3/13B3/13C3 | Coonrad-Morrey | 78 (61-89) | 73 | 52 (24-88) |
| Jost 2008 | 10 | 13A/13B/13C | Coonrad-Morrey | 65 (47-83) | NA | 66 (24-132) |
| Ali 2010* | 20 | 13A3/13B3/13C3 | Coonrad-Morrey | 72 (62-92) | 80 | 63 (36-108) |
| Antuna 2012* | 14 | 13B3/13C2/13C3 | Coonrad-Morrey | 77 (63-89) | NA | 54 (24-82) |
| Sorensen 2014 | 20 | 13B3/13C3 | Coonrad-Morrey | 77 (55-95) | 90 | 21 (4-52) |
| Prasad 2016 | 19 | 13A3/13B/13C | Coonrad-Morrey | 68 (40-80) | 63 | 156 (120-210) |
| Pogliacomi 2016 | 20 | 13C | Coonrad-Morrey / Latitude | 74 (68-86) | 95 | 60 (24-168) |
| Gallucci 2016* | 23 | 13A2/13C | Coonrad-Morrey | 76 (43-87) | 91 | 40 (13-96) |
| Barco 2017 | 44 | NA | Coonrad-Morrey | 70 (38-93) | 75 | NA (120-NA) |
| Baik 2020 | 43 | 13C | Coonrad-Morrey | 79 (65-91) | 79 | 34 (6-116) |
| Strelzow 2021 | 40 | NA | Latitude | 79 (NA) | 88 | 70 (24-156) |
| Celli 2021* | 13 | 13C2/13C3 | Coonrad-Morrey | 64 (61-67) | 85 | 153 (144-170) |
| Jonsson 2023 | 17 | 13B3/13C1/13C3 | Latitude / Discovery | 77 (NA) | 88 | 26 (NA) |
| **Hemiarthroplasty** | | | | | | |
| Nestorson 2015* | 42 | 13C3 | Latitude Hemiarthroplasty | 72 (56-84) | 93 | 34 (24-61) |
| Al-hamdani 2019* | 24 | 13C2,13C3 | Latitude Hemiarthroplasty | 65 (47-80) | 88 | 25 (12-70) |
| Celli 2022* | 24 | 13C2/13C3/13B3 | Latitude Hemiarthroplasty | 64 (45-78) | 75 | 91 (24-151) |
| Rotini 2023* | 27 | 13C3/13C2/13B3 | Latitude Hemiarthroplasty | 64 (45-78) | 78 | 82 (12-151) |
| Dirckx 2023* | 15 | 13C3/13B3 | Latitude Hemiarthroplasty | 75 (62-86) | 93 | 31 (16-56) |
| Jonsson 2023 | 18 | 13B3/13C1/13C3 | Latitude Hemiarthroplasty | 74 (NA) | 89 | 32 (NA) |

**Table 2** Demographics of included studies (NA = not available, * indicates individual patient data available)

| Author and Year | Clear criteria for inclusion | Condition measured in standard reliable way | Valid methods for identification of cases | Consecutive inclusion of participants | Complete inclusion of participants | Clear reporting of demographics | Clear reporting clinical information | Outcomes clearly reported | Reporting presenting sites demographic information | Appropriate statistical analysis |
|---|---|---|---|---|---|---|---|---|---|---|
| **Total elbow arthroplasty** | | | | | | | | | | |
| Gambirasio 2001 | yes | yes | yes | unclear | unclear | yes | yes | yes | no | yes |
| Garcia 2002 | yes | yes | yes | yes | no | yes | yes | yes | no | yes |
| Frankle 2003 | yes | yes | yes | unclear | yes | no | yes | yes | no | yes |
| Kamineni 2004 | yes | yes | yes | yes | yes | yes | yes | yes | yes | yes |
| Prasad 2008 | yes | yes | yes | unclear | unclear | yes | yes | yes | no | yes |
| Jost 2008 | Yes | Yes | Yes | Yes | no | yes | yes | yes | no | yes |
| Ali 2010 | yes | yes | yes | yes | no | yes | yes | yes | no | no |
| Antuna 2012 | yes | yes | yes | unclear | unclear | no | yes | yes | no | yes |
| Sorensen 2014 | yes | yes | yes | yes | no | yes | no | yes | no | yes |
| Prasad 2016 | yes | yes | yes | yes | no | yes | yes | yes | no | yes |
| Pogliacomi 2016 | yes | yes | yes | unclear | unclear | yes | yes | yes | no | yes |
| Gallucci 2016 | yes | yes | yes | unclear | no | yes | yes | yes | no | yes |
| Barco 2017 | yes | yes | yes | yes | no | yes | yes | yes | unclear | no |
| Baik 2020 | yes | yes | yes | unclear | unclear | yes | yes | no | no | no |
| Strelzow 2021 | no | yes | unclear | no | no | no | yes | no | no | yes |
| Celli 2021 | yes | yes | yes | yes | yes | yes | yes | yes | no | yes |
| **Hemiarthroplasty** | | | | | | | | | | |
| Nestorson 2015 | yes | yes | yes | yes | no | yes | yes | yes | no | yes |
| Al-hamdani 2019 | yes | yes | yes | yes | no | yes | yes | yes | yes | yes |
| Celli 2022 | yes | yes | unclear | unclear | no | yes | yes | yes | no | yes |
| Rotini 2023 | yes | yes | yes | unclear | no | yes | yes | yes | no | yes |
| Dirckx 2023 | yes | yes | yes | yes | no | yes | yes | yes | no | yes |

**Figure 3** Outcome of JBI critical appraisal tool for case series.(Munn *et al.*, 2020)

Pain Outcomes

Pain outcome data was obtained from the following outcome instruments; Numerical Rating Scale pain (NRS pain) score, Likert pain scale, Disabilities of the Arm Shoulder and Hand (DASH) item 24 or and short form DASH (QuickDASH) item 9 (pain in the last week), Oxford Elbow Score item 12 (average daily pain), Mayo Elbow Performance Score Likert pain scale.

The included randomised controlled trial reported pain for TER and DHH on a Likert scale extracted from the Mayo elbow performance score.(Jonsson *et al.*, 2023) Sixteen case series reported pain outcomes for TER; 3 reported pain using numerical rating scales (NRS pain) only, 12 used a Likert scale for pain only and 1 reported both.(Table 3) In 2 of the 4 studies reporting NRS pain the mean or median was provided with minimum and maximum values but with no estimate of variance, in one study the median, minimum, maximum and interquartile range was reported.(Prasad, Ali and Stanley, 2016; Barco *et al.*, 2017; Baik *et al.*, 2020) These values were used to calculate means and standard deviations.

| Author and Year | Sample size | Mean NRS pain score (SD) | Minimum NRS pain score | Maximum NRS pain score | Number with no pain | Number with mild pain | Number with moderate pain | Number with severe pain |
|---|---|---|---|---|---|---|---|---|
| **Total elbow arthroplasty** | | | | | | | | |
| Gambirasio 2001 | 10 | NA | NA | NA | 8 | 2 | 0 | 0 |
| Garcia 2002 | 16 | NA | NA | NA | 11 | 5 | 0 | 0 |
| Frankle 2003 | 12 | NA | NA | NA | 10 | 2 | 0 | 0 |
| Kamineni 2004 | 43 | NA | NA | NA | 35 | 8 | 0 | 0 |
| Prasad 2008 | 15 | NA | NA | NA | 8 | 5 | 2 | 0 |
| Jost 2008 | 10 | NA | NA | NA | 10 | 0 | 0 | 0 |
| Ali 2010 | 20 | NA | NA | NA | 14 | 6 | 0 | 0 |
| Antuna 2012 | 14 | NA | NA | NA | 6 | 3 | 4 | 1 |
| Sorensen 2014 | 20 | NA | NA | NA | 15 | 4 | 1 | 0 |
| Prasad 2016 | 19 | 1.8 (2.2)* | 0 | 5 | NA | NA | NA | NA |
| Pogliacomi 2016 | 20 | NA | NA | NA | 18 | 2 | 0 | 0 |
| Gallucci 2016 | 23 | 1.0 (1.3) | 0 | 5 | NA | NA | NA | NA |
| Barco 2017 | 44 | 0.6 (1.0) | 0 | 4 | 27 | 6 | 0 | 0 |
| Baik 2020 | 43 | 3.5 (2.0) | 0 | 8 | NA | NA | NA | NA |
| Strelzow 2021 | 40 | NA | NA | NA | 28 | 10 | 2 | 0 |
| Celli 2021 | 13 | NA | NA | NA | 7 | 6 | 0 | 0 |
| Jonsson 2023 | 17 | NA | NA | NA | 12 | 3 | 1 | 1 |
| **Hemiarthroplasty** | | | | | | | | |
| Nestorson 2015 | 42 | NA | NA | NA | 31 | 6 | 4 | 1 |
| Al-hamdani 2019 | 24 | 2.6 (2.4) | 0 | 7 | NA | NA | NA | NA |
| Celli 2022 | 24 | NA | NA | NA | 20 | 1 | 3 | 0 |
| Rotini 2023 | 27 | NA | NA | NA | 23 | 1 | 3 | 0 |
| Dirckx 2023 | 15 | 0.5 (0.8) | NA | NA | NA | NA | NA | NA |
| Jonsson 2023 | 18 | NA | NA | NA | 10 | 5 | 3 | 0 |

**Table 3** Data for pain outcomes (* indicates imputed data).

In addition to the included randomised trial reporting pain using a Likert scale for DHH, there were 5 case series reporting pain outcomes for DHH; 3 reporting Likert pain scores only and 2 reporting NRS pain only (Table 3).

*Pain measured on Numerical Rating Scale*

All studies included in the meta-analysis reported pain on an 11-point NRS scale, from 0 = no pain to 10 = worst pain. Four cohort studies measured pain following TER including 129 participants.(Gallucci *et al.*, 2016; Prasad, Ali and Stanley, 2016; Barco *et al.*, 2017; Baik *et al.*, 2020) The NRS pain score from the four studies of TEA had a mean ranging from 0.6 (SD 1.0) to 3.5 (SD 2.0). Meta-analysis estimated a weighted mean NRS pain for TER of 1.7/10 (95% CI 0.4 to 3.0). The statistical tests indicate substantial variation in the outcomes, a large proportion of which is due to true heterogeneity with an $I^2$ of 95.9%.

The two studies of DHH reporting NRS pain scores on an 11-point scale include 39 participants with means of 0.5 (SD 0.8) and 2.6 (SD 2.4).(Al-Hamdani *et al.*, 2019; Dirckx *et al.*, 2023) Meta-analysis estimated a weighted mean NRS pain for DHH of 1.5/10 (95% CI 0.0 to 3.6). Again, statistical tests indicate substantial variation in the reported outcomes, a large proportion of which is probably due to true heterogeneity ($I^2$ 93.6%), although with only two studies this should be interpreted with caution (Figure 4).

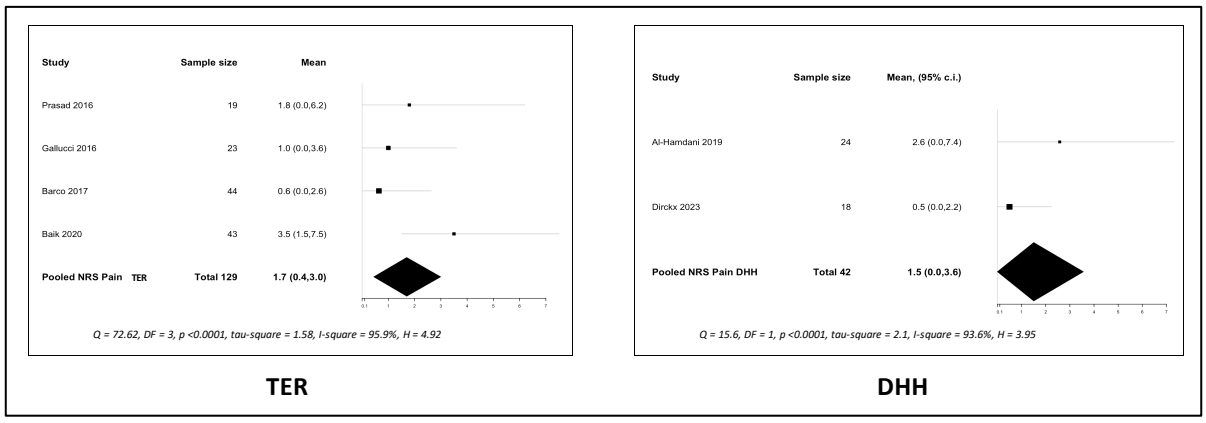



**Figure 4** Forest plots NRS pain for TER and DHH (DF = degrees of freedom).(Gallucci *et al.*, 2016; Prasad, Ali and Stanley, 2016; Barco *et al.*, 2017; Al-Hamdani *et al.*, 2019; Baik *et al.*, 2020; Dirckx *et al.*, 2023).

*Pain measured on a 4-point Likert scale*

Meta-analysis of the probability estimate of each response on the 4-point Likert pain scale was undertaken for TER and DHH.

For TER the weighted probability of no pain from 14 studies including 283 participants was 0.75(95% c.i. 0.67 to 0.82), mild pain 0.21(95% c.i 0.16 to 0.26), moderate pain 0.02 (95% c.i. 0.00 to 0.05), severe pain 0.00 (95% c.i. 0.00 to 0.01). (Ali, Shahane and Stanley, 2010) (Antuña *et al.*, 2012) (Frankle *et al.*, 2003; Barco *et al.*, 2017; Celli *et al.*, 2021) (Gambirasio *et al.*, 2001; Garcia, Mykula and Stanley, 2002; Kamineni and Morrey, 2004; Jost, Adams and Morrey, 2008; Prasad and Dent, 2008; Pogliacomi *et al.*, 2016; Jonsson *et al.*, 2023) (Sørensen, 2014; Strelzow *et al.*, 2021)

For DHH the weighted probability of no pain from 4 studies including 111 participants was 0.76 (95% CI 0.64 to 0.86), mild pain 0.10(95% CI 0.03 to 0.21), moderate pain 0.12 (95% CI 0.06 to 0.18) and severe pain 0.00(95% CI 0.00 to 0.03).(Nestorson *et al.*, 2015; Rotini *et al.*, 2020; Celli *et al.*, 2022; Jonsson *et al.*, 2023) (Figure 5)

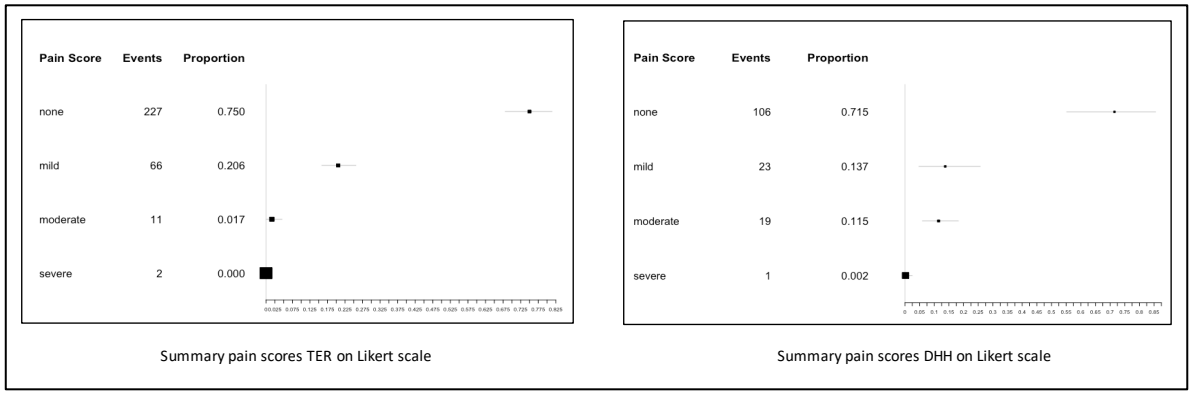



**Figure 5** Forest plots of pooled probabilities of Likert pain scores for TER and DHH.

The results of statistical test for heterogeneity for the pain outcomes and sensitivity analysis are provided in Appendix J. Sensitivity analysis indicated no change in the findings that would alter the interpretation when imputed data was removed from the analysis of the NRS pain score for TER.

Discussion

Adults with multi-fragmentary fractures of the distal humerus, who are considered to be well enough to undergo surgery, but whose fracture configuration or bone quality does not allow for fracture fixation (osteosynthesis), may consider either total elbow replacement or distal humerus hemi-replacement to restore function. Current NJR data suggests community uncertainty amongst elbow surgeon between TER or DHH in this situation with approximately equal numbers of each procedure recorded each year in the National Joint Registry of England, Wales, Northern Ireland and Guernsey.(Achakri *et al.*, 2023) The perceived advantage of DHH is preservation of bone stock, unlimited weight restriction, avoidance of polyethylene particulate debris and reduced risk of component loosening, as ulna stem loosening is a common problem in total elbow arthroplasty.(Figgie *et al.*, 2006; Goldberg *et al.*, 2008; Argintar *et al.*, 2012) TER on the other hand could result in better function with less pain, when compared to hemiarthroplasty, as has been observed in other joints.24,25 There may also be differences in the rates of revision following these two interventions.(Achakri *et al.*, 2023)

Two systematic reviews published in 2022 comparing the functional outcome of TER and DHH indicate no clinical difference in outcome when function was measured using the Mayo Elbow Performance Score, short form version of the Disabilities of the Arm Shoulder and Hand Score (quickDASH), and measured range of flexion and extension and pronosupination of the forearm.(Burden *et al.*, 2022; Stoddart *et al.*, 2022) One of the reviews reported that the full version of the Disabilities of the Arm Shoulder and Hand Score (DASH score) may be lower in the DHH group, with a weighted mean difference of 18.4/100 compared to TEA.(Burden *et al.*, 2022) However the other systematic review using broader inclusion criteria reported minimal difference in DASH scores.(Stoddart *et al.*, 2022) None of the published systematic reviews have investigated the pain outcome for TER or DHH for acute distal humerus fractures.

The only published randomised controlled trial reporting pain outcomes found no significant differences between the TER and DHH group with 20 participants in each group, when pain was measured on a Likert scale.(Jonsson *et al.*, 2023) Through this systematic review we

have described in more detail the available data on pain outcomes following TER and DHH for treatment of distal humerus fractures. Overall average pain scores are low following TER and DHH for acute distal humerus fractures, but this data synthesis indicates that some patients do report ongoing pain following these interventions. Data from studies reporting pain measured using a numerical rating scale suggest similar weighted mean pain scores between the two interventions, but this was based on only a few small studies with 129 observations for TER and 39 for DHH, which may explain the wider confidence interval for DHH. Differences in the distribution of pain scores were observed from a larger group of studies including more patients (283 observations TER and 111 DHH) measuring pain using 4-point Likert scales, with evidence indicating that more than 1 in 10 patients will report moderate pain after DHH, whereas less than 2/100 report moderate pain after TER. There is substantial heterogeneity in the reported data from included studies and therefore there is uncertainty regarding this finding. The quality of the available evidence does not enable direct comparison of pain outcomes between the two interventions. There is concern regarding clinical heterogeneity between studies (different fracture types) and the majority of studies are case series. The only randomised trial comparing outcomes between TER and DHH showed no difference in pain outcomes using a Likert instrument, but the trial was not powered to address this question.(Jonsson *et al.*, 2023)

Currently there is insufficient information to guide clinicians or patients regarding the better treatment option. This systematic review has added to the uncertainty regarding the equivalence of outcomes of TER and DHH for treatment of acute unreconstructible distal humerus fractures in adults. A prospective randomised trial is needed to compare the outcome of TER and DHH in this population. This systematic review provides data to inform an estimate of the distribution for the pain outcomes that can be used to help the trial design. Further work is required, however, to determine acceptable parameters for power analysis.

### 2.5.5 Concluding remarks

To my knowledge this is the first systematic review to report pain outcomes, the primary outcome identified by patients, following elbow replacement for acute distal humerus

fractures in adults. Whilst it is not possible to draw direct comparison between the pain outcome following total elbow replacement and hemi-replacement, the pooled results indicate that over 10% of patients report moderate or severe pain after hemi-replacement. Pooled pain results for total elbow replacement indicate that reports of moderate or severe pain are not common.

Pain has been measured entirely using either numerical rating scales or Likert pain scales, often as part of a combined function and pain instrument. Both instruments only assess pain in a single domain (pain over a specified time period) and neither of these pain instruments have been validated for use in elbow arthroplasty. The American Shoulder and Elbow Society pain (ASES pain) subscale and the Patient Rated Elbow Evaluation pain (PREE pain) subscale, have been validated for use in this setting, but currently there is a lack of data on the expected pain outcomes from these two instruments.(MacDermid, 2001)(King *et al.*, 1999)(King *et al.*, 1999; MacDermid, 2001; Angst *et al.*, 2005) Neither instrument provides weighting to the pain scales captured, and the total score for each instrument is used to report outcomes. Of the two instruments, only the PREE assesses the frequency of pain symptoms. This is a useful measure of the impact of pain, that is used in clinical practice, because it has been shown that increases in the frequency of pain can have a negative effect on quality of life.(Katz, 2002)

In the absence of prior data on PREE pain scores following elbow replacement for trauma, other techniques have to be employed to establish probability distributions. Expert elicitation, where experts are asked to express their view of the likely distribution of values of a quantity of interest based on their own experience of a subject and a summary of current knowledge, is a technique for generating probability distributions for unknown quantities. The next chapter will explore estimates of PREE pain outcomes following elbow arthroplasty for acute trauma in adults, using Sheffield Expert Elicitation Framework.(Oakley and O'Hagan, Anthony, 2019)

## 2.6 Paper 4: Pain score distribution Expert elicitation study for Distal humerus hemi-replacement And total elbow replacement using a validated Numerical patient rated outcome-measure in Trauma (PEDANT)

### 2.6.1 Publication status

Submitted for publication to the Bone and Joint Journal (in review).

### 2.6.2 Co-author statement

**Watts A.C., Hamoodi Z., McDaid C., Hewitt C., PEDANT expert panel. Pain score distribution Expert elicitation study for Distal humerus hemi-replacement And total elbow replacement using a validated Numerical patient rated outcome-measure in Trauma (PEDANT). (Submitted to Bone and Joint Journal)**

**Candidate contribution: Conceived the idea; designed the study; wrote the protocol; arranged and facilitated the workshops; undertook the analysis; wrote the manuscript; submitted for publication.**

| Signed: | Signed: | Signed: |
|---|---|---|
|  |  |  |
| (Adam Watts) | (Catriona McDaid) | (Catherine Hewitt) |

### 2.6.3 Introduction to the paper

The previous paper described the weighted pooled distribution of pain outcomes following TER and DHH for acute trauma from the published literature. The reported outcomes have been measured using numerical rating scales for pain or Likert scales, both of which are unidimensional and neither has been validated for use in elbow replacement research. These pooled distributions can, however, be used by elbow surgery experts to elicit probability distributions for an instrument that has been validated for use after elbow replacement, but for which parameter properties are not known, namely the Patient Rated Elbow Evaluation pain (PREE pain) score. There is a rich literature on the psychology behind expert elicitation and techniques to manage bias, anchoring and group behaviour.(O'Hagan, 2006) The Sheffield elicitation framework is a structured approach to expert elicitation that has been used in the next paper to develop prior probability distribution functions to address objective D of the thesis, to generate an informative prior for the key outcome using expert elicitation techniques informed by the evidence synthesis.(Oakley and O'Hagan, Anthony, 2019) The experts were selected from an international group of known elbow replacement experts who are active in research and who were willing to take part in the study, balancing those who have published on the use of total elbow replacement and hemi-replacement in trauma, and therefore have an understanding of probability distributions for pain outcomes. Other key-stakeholders, such as patients and physiotherapists were not included.

### 2.6.4 Submitted manuscript

Title

Pain score distribution expert elicitation study for distal humerus hemi-replacement and total elbow replacement using a validated numerical patient rated outcome-measure in trauma (PEDANT)

## Abstract

### Background

When distal humerus fractures in adults cannot be fixed, total elbow replacement (TER) and distal humerus hemi-replacement (DHH) are used for joint salvage in approximately equal numbers in England. Patients have reported that pain is the primary outcome of this intervention from their perspective. The Patient Rated Elbow Evaluation (PREE) pain subscale has been validated for use as a pain instrument for elbow replacement but there is no data on PREE pain outcomes for patients treated for acute trauma. Expert elicitation is a method for producing probability distributions for unknown parameters.

### Aim

The aim of this study is to conduct expert elicitation using the Sheffield Expert Elicitation Framework (SHELF) methodology to produce probability distributions for the PREE pain score at 12 months in adult patients with unreconstructible distal humerus fractures treated with TER or DHH.

### Methods

SHELF methodology was used to elicit probability distributions. Seven experts were recruited from several countries selecting a balance of those favouring TER and those favouring DHH. The experts were provided with a summary of known pain outcomes for elbow replacement for trauma from a systematic review, and were asked to consider their median of the Patient Rated Elbow Evaluation (PREE) pain score at 12 months after a TER for acute trauma, their difference in the median PREE pain scores at 12 months between TER and DHH, and the standard deviation of the PREE pain score at 12 months after a TER and DHH. The Rational Independent Observer (RIO) concept was used to achieve consensus. Where elicited distributions follow a normal distribution the standard distribution (SD) is used to describe distribution of data, for other distributions the interquartile range (IQR) is used.

Results

The elicited median value for the PREE pain at 12 months after TER for acute trauma was 13.6/50 (IQR 10.6 to 16.8). The experts median estimate of the difference in PREE pain scores between TER and DHH at 12 months for acute trauma was 3.4 (IQR -0.3 to 7.1). The elicited standard distribution of PREE pain scores at 12 months after TER for trauma was 9 (standard deviation (SD) 2.2), and after DHH was 12 (SD 3.0).

Conclusion

Having considered the available evidence and using standard elicitation techniques seven experts reached a consensus that pain scores measured using the PREE pain instrument at 12 months in patients treated with elbow arthroplasty for acute trauma would be higher following DHH than TER, and that the pain scores for DHH patients would exhibit greater variance than patients having TER. This difference is estimated to be small and modelling of trial outcome suggests it would not be possible to show a statistically significant difference in pain outcome between the two interventions using this instrument. This data provides prior distributions that can be updated using Bayesian methodology in a future randomised controlled trial to provide more accurate estimates of population probability distributions.

## Introduction

Fractures of the distal humerus are increasing due to an ageing population, and elbow arthroplasty for trauma has doubled in the ten years that the National Joint Registry has been collecting data for England, Wales, Northern Ireland and Guernsey with 144 confirmed elbow arthroplasty operations for acute trauma in 2022.(Palvanen *et al.*, 2010; Achakri *et al.*, 2023) The treatment options for this injury are cast immobilisation without an operation, fixation with plates and screws (osteosynthesis), and elbow prosthetic replacement.(Stoddart *et al.*, 2022) Where the bone cannot be fixed there are two options for elbow replacement for trauma, distal humerus hemi-replacement (DHH) in which just the broken elbow end of the humerus is replaced, and total elbow replacement (TER) in which both the broken humerus and intact proximal ulna are replaced. TER has been the standard

care but there are concerns about limitations in loading the operated joint, the risk of the implant loosening from the bone over time leading to pain and a need for revision surgery.(Barco *et al.*, 2017) DHH has been proposed as an alternative solution with the hypothesis that the risks caused by joint loading activities and of implant loosening and revision is reduced because only one part of the joint being replaced.(Phadnis, Banerjee, Adam C Watts, *et al.*, 2015) Hemi-replacement has been used in other joints for trauma, but has been shown to be associated with higher risk of joint pain and inferior functional outcomes when compared to total joint arthroplasty.(Zi-Sheng *et al.*, 2012; Shukla *et al.*, 2016) There are no large randomised controlled trials comparing the outcome of DHH to TER in acute elbow trauma and it is not known which approach is better for patients.

One of the barriers to conducting a randomised trial is the incidence relatively low incidence of distal humerus fractures requiring arthroplasty, which makes it challenging to recruit and randomise sufficient participants to adequately power the study in an economically viable time period. Elbow arthroplasty is a rare intervention with an annual incidence of less than 1 per 100,000 population, and elbow arthroplasty for acute trauma accounts for only 40% of these, with approximately equal numbers having TER and DHH in the NJR data from 2020 to 2022. Whilst it may be viable to conduct a suitably powered randomised trial using a classical or frequentist approach if all of these patients could be randomised, these procedures are spread over approximately 90 hospitals, creating a logistical and economic challenge.

An alternative approach is to use Bayesian approaches, whereby we start with an initial probability distribution for the outcome (prior distribution), collect data (likelihood) and apply Bayes theorum to update the probability distribution (posterior distribution).(Donovan and Mickey, 2019) The posterior distribution should be closer to the true population distribution than the prior, based on the evidence collected. When using this approach we can use an uninformative prior, where every outcome has the same probability, which is akin to the frequentist approach, where we ignore any previous knowledge on the subject in the analysis other than for conducting power calculations. In most cases, however, we have some knowledge of the expected outcomes, and it may be possible to produce an

informative prior based on published data and expert knowledge, whereby some outcomes are considered to have a higher probability than others.

Trial planning requires identification of a primary outcome which should be valid, feasible to measure and relevant to the participants and end-users.(McLeod *et al.*, 2019) In the CODER Real-Time Delphi consensus study patients identified pain as the primary outcome domain for elbow arthroplasty.(Watts *et al.*, 2024) Of the validated pain instruments for elbow arthroplasty, the Patient Rated Elbow Evaluation(PREE) best fulfils the criteria for the primary outcome instrument.(MacDermid, 2001) The PREE measures the average amount of pain over the last week using an 11-point numerical rating scale from no pain (0) to worst pain ever (10) in five domains; worst pain, rest pain, pain when lifting a heavy object, pain with repeated movement, and frequency of pain from never (0) to always (10). However, a lack of data on PREE scores in the population undergoing elbow arthroplasty for acute trauma is a barrier to trial planning.

A systematic review of the published literature since 2000 has shown that the estimate of pain outcomes after elbow arthroplasty for trauma is not well described.(Watts, 2023b) The literature consists largely of case series with implicit high risk of bias. The identified studies use either an NRS or Likert scale for pain. None report the pain scores using the PREE. Expert elicitation is a scientific model used to produce probability distributions for unknown parameters. A number of different approaches can be used such as the Classic Method and the Sheffield Expert Elicitation Framework (SHELF).(Cooke, R.M., 1991; Flandoli *et al.*, 2011; O'Hagan, 2019) In the elicitation process experts are asked to consider what is currently known on a subject, and their own experience as an expert, in order to elicit a probability distribution for quantities of interest for which the value is not known. This elicited information can be used to inform prior distributions for Bayesian approaches or to inform sample size calculations with a frequentist approach.(Tan, Chung, *et al.*, 2003)(Oremus *et al.*, 2002; Tan, Chung, *et al.*, 2003)

The aim of this study is to undertake expert elicitation using the SHELF methodology to produce prior probability distributions for the PREE pain score at 12 months in adult patients with unreconstructible distal humerus fractures treated with TER or DHH.

## Materials and Methods

The protocol adheres to the reference protocol for expert elicitation studies described by Bojke et al. in 2021 which was funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme.(Bojke *et al.*, 2021)

### Experts

The experts were recruited from specialists in orthopaedic elbow surgery. The experts were chosen on the basis of known expertise in elbow replacement for distal humeral fracture with evidence of research on this topic and from specialist shoulder and elbow surgeons with a strong background in research methods evidenced by publication. Representation was balanced between those predominantly using DHH and those predominantly using TER for acute trauma reconstruction. Experts were drawn from a number of different countries to ensure a breadth of opinion was sampled. Experts were required to be able to communicate in English and understand written English.

Letters of invitation were sent via email outlining the protocol and time commitment required. Experts had to be willing to participate in one-to-one sessions and commit to attending a group elicitation session. The experts were asked to declare any personal or financial conflicts of interest.

### Approach to elicitation

The Sheffield Expert Elicitation Framework (SHELF) was used.(Oakley and O'Hagan, Anthony, 2019) The experts were provided with a standardised SHELF expert briefing document that explained the process of expert elicitation and introduced the key concepts that the expert needed to understand for the workshop (Appendix K). Additional training was available online via the SHELF website which participants were signposted to (https://shelf.sites.sheffield.ac.uk/e-learning-course).

The experts were provided with an evidence dossier developed by AW based on a systematic review summarising current evidence regarding pain outcomes for TER and DHH in trauma from studies published since 2000 (Appendix L). The evidence consists largely of observational studies of case series, and pain was measured entirely with either 11-point numerical rating scales for pain (NRS pain) or 4- or 5-point Likert scales. The summary evidence from the systematic review reporting a high risk of bias was provided for the experts, showing summary mean NRS pain scores and 95% confidence intervals for TER and DHH, as well as summary Likert pain scale outcomes reporting the proportion reporting pain as none, mild, moderate, and severe.

To help the experts convert the summary evidence using the Likert scale into PREE outcomes, data from a separate study by Angst et al. of a non-trauma population was presented that had measured pain six months after TER using PREE pain and a Likert pain scale from the SF-36.(Angst *et al.*, 2005) The mean PREE pain score was 14/50 (SD 13.3) and the mean SF-36 pain score was 60 which equates to "mild pain" on the 6-item SF-36 Likert scale. The experts were invited to comment on the data presented and suggest changes prior to starting the elicitation workshops to produce a final evidence dossier that was shared with all reviewers.

One-to-one elicitation was conducted by AW as facilitator and ZH as recorder with the experts to elicit the quantities of interest.
Once all of the experts had completed their one-to-one elicitation exercises, they were brought together online for a group elicitation meeting to discuss the anonymised individual results and to elicit a distribution for each quantity of interest as a group using the principle of the "rational independent observer" (RIO).

### Elicitation method

Variable interval methods (VIM) were used to elicit probability distributions for the median NRS pain scores for DHH and TER.(Bojke *et al.*, 2021) The median value was elicited, rather than the mean, as it is easier for experts to consider the value below or above which they would expect 50% of the values to lie. In normally distributed data the median will

approximate to the mean. The quartile method was used for initial elicitation by individual experts. The probability method was used for the group elicitation. Standard deviations were elicited using the method described by Alhussain and Oakley.(Alhussain and Oakley, 2020)

### Quantities of interest

A: Experts were asked to consider their median of the Patient Rated Elbow Evaluation (PREE) pain score at 12 months after a TER.

B: Experts were asked to consider their difference in the median PREE pain scores at 12 months in patients having TER or DHH for acute trauma.

C: Experts were asked to consider their standard deviation of the PREE pain score at 12 months after a TER and DHH.

### Fitting

The fitting was guided by the expert's beliefs using SHELF software and distribution parameters generated, the beta distribution was chosen as the default because this produces distribution bounded by the lower and upper plausible limits provided by the expert.(Oakley, 2015) The individual fitted distribution were shown to the expert without commentary or revision. For the group stage the quantity of interest was elicited using the RIO model and the fitted consensus RIO distribution was shared with the experts for consideration and adjustment if it did not adequately reflect their consensus view. A beta distribution was generated from the parameters elicited from the experts, bounded by the parameter limits.

### Aggregation

Mathematical aggregation was not undertaken as there was insufficient agreement between the individually elicited fitted distributions. Behavioural aggregation was used with experts attending an online consensus meeting to agree a fitted distribution for the RIO.

### Documentation

Standardised SHELF documents were used to collect data from the experts before and during the elicitation process. The "Expert Enquiry" form was completed by the experts prior to the elicitation meeting and their responses collated (Appendix M). This included the expert's name and title, their background and expertise and their declarations of interest. The experts were asked to review the evidence dossier and to list any additional sources of evidence that they felt should be considered by the experts, and to list any questions, clarifications or corrections pertaining to the evidence dossier.

During each workshop the SHELF elicitation record documents were used. The purpose of the elicitation was clearly defined and a record of the pre-elicitation training received by each expert was documented, including any training elicitations. The expertise of all participants was recorded alongside declarations of interest. The evidence document was reviewed and clarifications made without discussion of the quality or interpretation of the evidence. The structure of the elicitation was discussed with the expert, with clear definitions of the quantities of interest provided.

### One-to-one elicitation

The SHELF elicitation record was completed for each quantity of interest. This captured the date, start time and finish time of the meeting, a record of the discussion and parameters elicited. The facilitator (AW), an elbow surgeon trained in SHELF methodology, led the discussion. To ensure anonymity of the experts' elicitations at the group feedback stage, each was assigned a unique identifier that was used in feedback.

For each quantity of interest, the definition and the evidence regarding only that quantity was reviewed. The expert was asked to record their lower and upper plausible limits for that quantity. The expert was then asked to consider and record their median (M) and lower and upper quartiles (Q1, Q2) for the quantity of interest using pre-prepared slide sets from the SHELF package to explain these concepts. The recorder (ZH) documented the experts'

responses and then fitted a distribution using SHELF software to the expert's assessment which was captured in the form of density functions. The expert was shown the fitted distribution but not invited to revise it and no commentary was provided by the facilitator at this stage.

Standard deviations for the PREE pain at 12 months for TER and DHH were elicited by asking the expert to consider their lower and upper estimates of the proportion of patients scoring between their elicited median score for PREE pain after TER and their median plus 10 points. For the standard deviation of DHH their median PREE pain score was calculated from the sum of their median for TEA and their difference in PREE pain score between TER and DHH at 12 months. The expert was asked to consider their lower and upper estimates of the proportion of patients scoring between their elicited median score for PREE pain after DHH and their median plus 10 points. The assurance shiny app from the SHELF website was used to calculate their upper and lower limit of the standard deviations from the proportions estimated.(Alhussain and Oakley, 2020)

### Group Elicitation

The group elicitation record from SHELF was used to document the group elicitation exercise. The Facilitator computed an equally-weighted average of the density functions (linear pool) using SHELF software. This was not revealed to the experts, but available as a benchmark by the facilitator if group discussions lead to an unwarranted narrowing of the uncertainty, that was not supported by the group elicitations. The group were shown each of the individually synthesised fitted distributions for each quantity of interest in anonymised format to stimulate discussion. The experts were given the opportunity to give their opinion about the quality and interpretation of the evidence.

For the group stage elicitation, the limits were set to the parameter limits for the quantity of interest; the median PREE at 12 months after TER (0 to 50) and the RIO difference in PREE scores at 12 months between TER and DHH (-50 to 50). The group elicitation method and judgements were recorded. The group consensus judgements for RIO using the probability

method were recorded for the probabilities (P1, P2, P0) of three values of the quantity of interest (X) chosen by the facilitator (X1, X2, X0).

To elicit the standard deviation for PREE pain scores for TER and DHH at 12 months the group were asked to reach a consensus on what a rational independent observer (RIO) would consider to be a plausible upper and lower limit for the standard deviation of PREE pain scores at 12 months after TER and DHH, based on the experts lower and upper limits from the individual elicitation stage, which was recorded as the "group plausible range". The quartile method was used to elicit the RIO probability distribution of the quantities of interest.

The parameters for the fitted distribution for the consensus RIO judgement was recorded in the SHELF documentation, and the fitted distribution shared with the experts alongside some implied probabilities such as the $10^{th}$ and $90^{th}$ percentile. The experts were given the opportunity to adjust the elicited parameters with further rounds of fitting and feedback if required until the final fitted distribution was accepted and recorded in the chosen distribution field.

### Delivery

The structured elicitation exercises and group elicitation stage was conducted online using Zoom (Zoom Video Communications Inc.) to accommodate expert surgeons from Belgium, France, Finland, Sweden and the United Kingdom.

### Training and piloting

Online training videos for self-directed learning were made available ( https://shelf.sites.sheffield.ac.uk/e-learning-course) and experts were encouraged to review these before the one-to-one elicitation meetings. To familiarise the expert to the methods of expert elicitation a trial elicitation was conducted to elicit parameters for a quantity of interest unrelated to the study aims at the start of the one-to-one elicitation workshop. This allowed the Facilitator to highlight common errors of elicitation such as overconfidence

(bounds between upper and lower plausible limits too narrow) and imprecision (for example placing upper and lower quartiles too far from the median).

The protocol was piloted with a local elbow surgeon expert (AWr) to ensure that the procedures function as intended and could be easily followed by the participants.

Results

Seven experts were contacted and agreed to take part in the expert elicitation. Prior to participation in the elicitation workshop all experts read the SHELF elicitation guide and one expert accessed the online SHELF training material. The experts declared any conflicts of interest (Appendix M). All experts were happy with evidence dossier provided and none provided any additions, corrections or sought any clarifications.

*Median PREE pain score at 12 months following TER for distal humerus fracture*

The experts were asked to consider their distribution of the median score on the PREE pain scale from 0 to 50 (were 0 represents no pain) at 12 months after TER. They were asked to consider what the distribution would look like if 100 experiments were conducted and all the medians were plotted. The parameters elicited from the experts are given in table 1.

At the group stage this elicited data was considered and there was discussion of the available evidence. The parameter limits for the PREE are 0 and 50. The experts were asked to agree RIO cumulative probabilities for a value of the median PREE of 10, 16 and 20 and agreed values of 0.2, 0.7, and 0.9. The fitted beta distribution curve (parameters a = 6.69, ß = 17.4) had a median value of 13.6/50 (inter-quartile range (IQR) 10.6 to 16.8). This was fed back to the experts, with the implications that 10% of the median values would lie below 7.9 and 10% would lie above 19.6, and the experts felt that this reasonably represented the RIO distribution. No further elicitation was conducted for this quantity.

| Expert | Lower plausible limit | Lower quartile | Median | Upper quartile | Upper plausible limit | Fitted Beta distribution parameters | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | $\alpha$ | $\beta$ | $\mu$ |
| A | 10 | 12 | 15 | 16 | 18 | 0.953 | 0.825 | 14.40 |
| B | 10 | 12 | 16 | 20 | 26 | 0.711 | 1.070 | 15.70 |
| C | 5 | 8 | 10 | 12 | 16 | 1.670 | 1.960 | 9.97 |
| D | 10 | 12 | 15 | 17 | 20 | 0.908 | 1.000 | 14.60 |
| E | 7 | 12 | 14 | 16 | 26 | 3.650 | 6.090 | 14.00 |
| F | 10 | 13 | 16 | 18 | 22 | 1.310 | 1.410 | 15.70 |
| G | 12 | 14 | 15 | 16 | 18 | 2.160 | 2.160 | 15.00 |

**Table 1** Individual expert elicited Median PREE scores at 12 months following TER for acute distal humerus fracture.

*Difference in the median PREE pain score at 12 months between TER and DHH for distal humerus fracture*

The experts were asked to consider their distribution of the difference in the median PREE pain scores in 100 trials comparing TER and DHH at 12 months. The parameters elicited are provided in table 2.

The parameter limits for the median difference of the PREE pain scores were -50 and 50. The experts were asked to agree RIO cumulative probabilities for a value of the median PREE of 0, 4 and 10 and agreed values of 0.3, 0.5, and 0.925. The fitted beta distribution curve (parameters a = 44.5, ß = 38.9) had a median value of 3.4 (IQR -0.3 to 7.1). This was fed back to the experts, with the implications that 10% of the difference values would lie below -3.6 and 10% would lie above 10.4, and the experts felt that this reasonably represented the RIO distribution and no further elicitation was required.

| Expert | Lower plausible limit | Lower quartile | Median | Upper quartile | Upper plausible limit | Fitted Beta distribution parameters | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | $\alpha$ | $\beta$ | $\mu$ |
| A | 1 | 2 | 3 | 4 | 5 | 1.000 | 1.000 | 3.00 |
| B | 0 | 2 | 5 | 8 | 12 | 0.824 | 1.080 | 4.86 |
| C | -5 | 2 | 5 | 7 | 10 | 2.480 | 1.450 | 4.82 |
| D | 5 | 7 | 10 | 15 | 20 | 0.655 | 0.980 | 10.30 |
| E | -5 | -2 | 0 | 2 | 8 | 1.870 | 2.850 | -0.06 |
| F | 0 | 3 | 5 | 7 | 10 | 1.530 | 1.530 | 5 |
| G | -6 | -2 | 0 | 2 | 6 | 2.160 | 2.160 | 0 |

**Table 2** Individual expert elicited difference in the median PREE score at 12 months between TER and DHH for distal humerus fracture.

*Distribution of the Standard Deviations of the PREE pain score at 12 months following TER and DHH*

To elicit the standard deviations the experts were asked to consider their estimate of the proportion of patients that would score between their median and their median plus 10 points. Table 3 shows the 5% and 95% probability estimates of the proportion scoring in this interval for each expert.

In the group stage the experts were asked to consider the proportions elicited and from these to elicit a RIO distribution using the quartile method. For TER the experts agreed a lower limit for the standard distribution of 7 and an upper limit of 20. The RIO median was 12, lower quartile 10 and upper quartile 20. These parameters were fitted to a normal distribution eliciting a mean standard deviation for PREE pain scores at 12 months after TER of 9 (SD 2.2). For DHH the experts agreed a lower limit of 7 and upper limit of 20. The RIO median was 12, lower quartile 10 and upper quartile 14. These parameters were fitted to a normal distribution eliciting a mean standard deviation for PREE pain scores at 12 months

after DHH of 12 (SD 3.0). The experts were satisfied that the elicited distributions represented their views.

| Expert | Probability of scores falling between median, median+10 TER | | Elicited Standard Distribution PREE pain for TER | | Probability of scores falling between median, median+10 DHH | | Elicited Standard Distribution PREE pain for DHH | |
|---|---|---|---|---|---|---|---|---|
| | 5% | 95% | 5% | 95% | 5% | 95% | 5% | 95% |
| A | 0.30 | 0.45 | 11.9 | 6.1 | 0.25 | 0.40 | 14.8 | 7.8 |
| B | 0.20 | 0.40 | 19.1 | 7.8 | 0.20 | 0.40 | 19.1 | 7.8 |
| C | 0.20 | 0.45 | 19.0 | 6.1 | 0.10 | 0.40 | 39.5 | 7.8 |
| D | 0.25 | 0.45 | 14.8 | 6.1 | 0.25 | 0.45 | 14.8 | 6.1 |
| E | 0.40 | 0.48 | 7.8 | 4.9 | 0.35 | 0.45 | 9.7 | 6.1 |
| F | 0.30 | 0.499 | 11.9 | 4.3 | 0.15 | 0.499 | 26.0 | 4.3 |
| G | 0.40 | 0.45 | 7.8 | 6.1 | 0.35 | 0.40 | 9.7 | 7.8 |

**Table 3** Individual expert elicited distribution of the standard deviations of the PREE pain score at 12 months following TER and DHH

Discussion

Elbow prosthetic replacement is used for unreconstructible intra-articular distal humerus fractures in patients who are fit for surgical intervention. According to the National Joint Registry data, in the most recent three years reported, when patients undergo elbow replacement for trauma, approximately half receive a TER and half DHH.(Achakri *et al.*, 2023) There are some apparent differences in the demographics with 22% of those undergoing DHH under 65 years of age and 43% over 74 years, compared to 9% <65 years for TER and 62% >74 years.(Achakri *et al.*, 2023) The proportion of female patients is 84% for both interventions from the latest report.(Achakri *et al.*, 2023) This data indicates some preference for use of DHH in younger patients, which is consistent with a concern regarding long term risks of ulna implant loosening that is present with TER. There is a lack of quality

evidence to compare TER to DHH to help the clinician and patient choose the best intervention. One randomised trial with 20 patients in each treatment arm, has reported equivalent functional outcomes from TER and DHH for acute fractures with no difference in pain outcomes, but the trial has been found to be at high risk of bias using the Cochrane ROB-2 criteria and may be underpowered to detect a difference.(Sterne *et al.*, 2019; Jonsson *et al.*, 2023) From this, some may conclude that it does not matter which intervention is used, but there are differences in the reported revision rate for TER and DHH at 2 years in favour of TER reported in the NJR, and systematic reviews have raised concerns regarding functional outcomes in younger patients, and the overall levels of pain after DHH.(Achakri *et al.*, 2023; Heifner *et al.*, 2024)(Achakri *et al.*, 2023; Watts, 2023b; Heifner *et al.*, 2024) Others have reported that function scores may be higher after DHH compared to TER.(Burden *et al.*, 2022) This uncertainty regarding the best treatment option for patients with unreconstructible intra-articular distal humerus fractures could be addressed by a suitably powered prospective randomised controlled trial stratified by age, but this may not be feasible due to low incidence. Bayesian trial methodology may be a more feasible approach for this type of rare intervention.

A recent Delphi consensus study with patient and public involvement has reported that the most important outcome for patients after elbow replacement is pain. This was felt to be more important than elbow function.(Watts *et al.*, 2024) The choice of instruments to measure pain after total elbow replacement is limited. Many instruments combine pain assessment with function assessment (Oxford elbow score, DASH score, Liverpool elbow score) and pain items are rarely reported separately by investigators.(Dawson *et al.*, 2008)(Hudak, Amadio and Bombardier, 1996)(Sathyamoorthy, 2004;Dawson *et al.*, 2008) Pain scales such as the visual analogue scale, numerical rating scale or Likert scale have been validated as tools for measuring pain, but we are not aware of any investigations of their psychometric properties in elbow replacement studies and they are unidimensional.(Herr *et al.*, 2004) The PREE and American Shoulder and Elbow Society (ASES-e) scores measure and report pain independently, over several dimensions, and have been validated in general elbow conditions and in total elbow replacement studies.(Angst *et al.*, 2005, 2012; Vincent *et al.*, 2021) The ASES-e does not capture frequency of pain symptoms and includes

physician assessment of strength, range of movement and stability making it more cumbersome, therefore the PREE has been selected for use in this study.

Angst et al. have reported that in patients undergoing total elbow replacement the mean PREE improved from a score of 30/50 (SD 9.2) pre-operation to 15/50 (SD 13.7) at 6 months after surgery. However the study population was not representative of those undergoing replacement for trauma with 54% of the 79 patients included having a diagnosis of rheumatoid arthritis.(Angst *et al.*, 2012) Acute fracture patients were excluded from the analysis and this data can only guide the expert as to the likely outcome in the trauma population.

Expert elicitation is a tool to produce probability distributions for unknown quantities of interest. The experts in the PEDANT study have provided their consensus based on a synthesis of current knowledge, their own expertise and the views of other experts in the field, to provide estimated median PREE pain scores for patients 12 months after TER, the difference between median PREE pain scores for TER and DHH and estimates of the standard deviations for both. The expert elicitation process was conducted in two phases, individual elicitation and group elicitation. In the individual elicitation phase the experts were asked to comment on the evidence but there was no discussion of the evidence provided. The experts were given a practice elicitation exercise in order to address common elicitation errors of overconfidence and precision.(O'Hagan, 2006) The experts have expressed their view that a rational independent observer, considering the evidence provided, observing their elicited quantity probability distributions and listening to the considerations made by the experts in the group discussion would conclude that in a randomised trial comparing pain outcomes at 12 months in patients undergoing TER or DHH for acute distal humerus fractures the median pain score for TER would be 14/50 with a variance of 81, for DHH the median score would be 17/50 with a variance of 144. Applying these parameters to the Bayesian assurance application developed by Alhussain and Oakley, it is unlikely that a randomised trial will show benefit of one intervention over the other if pain is the primary outcome.(Alhussain and Oakley, 2020) It is possible that the background data is misleading, or that the parameters elicited by the experts underestimate the treatment effect. It is also important to note that behavioural aggregation was used to agree a fitted distribution for

the RIO between the experts. Even if well-managed this type of aggregation may be a source social and cognitive biases such as group polarisation, overconfidence and groupthink and hence is a possible limitation of this study. Nevertheless, this data provides prior distributions that can be updated using Bayesian methodology in a future randomised controlled trial for this rare intervention, to provide more accurate estimates of population probability distributions.

### 2.6.5 Additional material

To conduct a trial using Bayesian inference, priors are required. Using the SHELF shiny apps in R enthusiastic, sceptical and uninformative priors can be generated based on the parameters elicited from the experts in PEDANT.(Oakley, 2015) The minimal clinically important difference (MCID) would often be used to generate the enthusiastic prior, but this has not yet been determined for PREE in patients undergoing elbow replacement. The enthusiastic prior is therefore based on the probability density function (PDF) of the difference in mean PREE scores elicited (Figure 2.2).

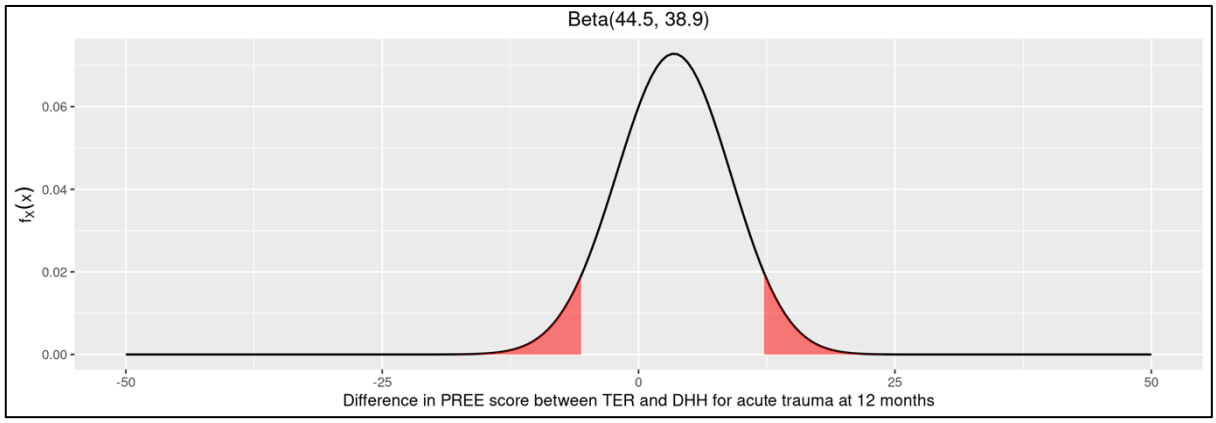



**Figure 2.2** Enthusiastic probability density function for the difference in PREE score between TER and DHH for acute trauma at 12 months.

The sceptical prior has a mean difference of zero and a 5% probability of exceeding the median difference from the enthusiastic prior, as specified by Wijeysundera et al. (Figure 2.3). (Wijeysundera *et al.*, 2009)

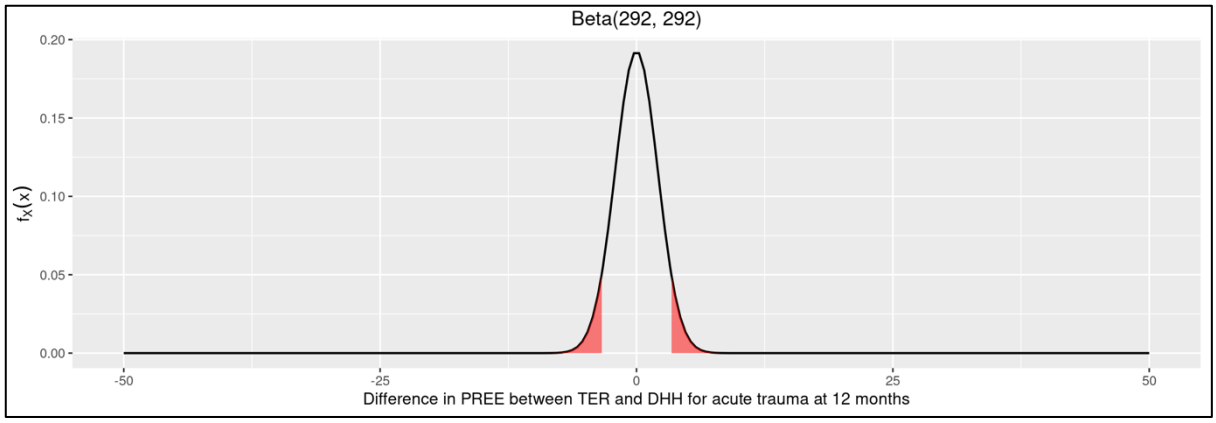



**Figure 2.3** Sceptical probability density function for the difference in PREE score between TER and DHH for acute trauma at 12 months.

A non-informative prior can take a number of different forms. Wijeysundera advise using a mean of zero and standard deviation of 10 on a logarithmic scale.(Wijeysundera *et al.*, 2009) An alternative, and that chosen here, is to use a non-informative prior with equal probability of all possible outcomes, where the likelihood will define the posterior (Figure 2.4).

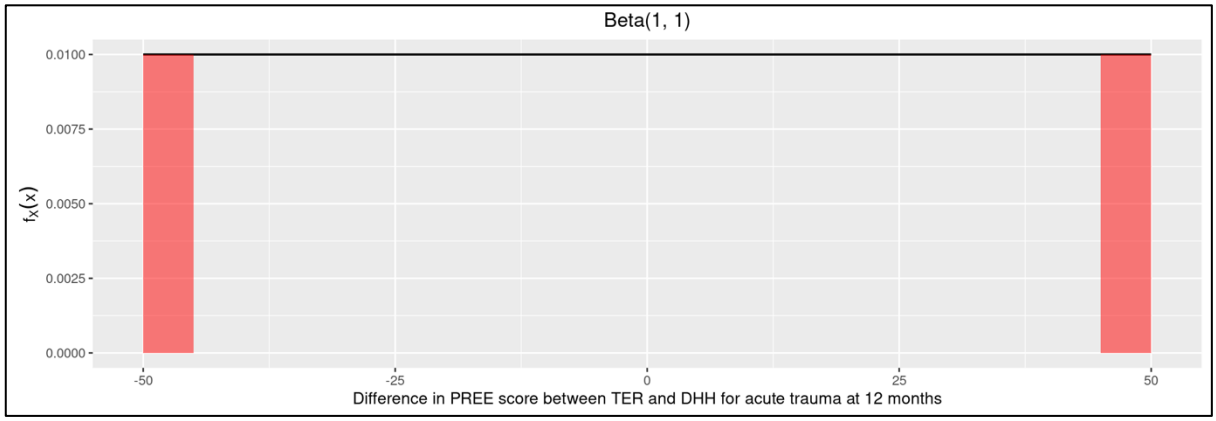



**Figure 2.4** Non-informative probability density function for the difference in PREE score between TER and DHH for acute trauma at 12 months.

### 2.6.6 Concluding remarks

This expert elicitation study has generated probability distributions that can be used as priors in a future Bayesian designed randomised trial to compare pain outcomes following

total elbow arthroplasty and hemiarthroplasty for acute distal humerus fractures in adults. Assurance testing using the methodology of Alhussain and Oakley suggest that if the median difference probability distribution generated by the experts is correct it is unlikely that a trial will find a statistically significant difference between the two interventions with an $\alpha$ of 5%, when pain is measured using the PREE.(Alhussain and Oakley, 2020) It might be concluded that there is little value is conducting a trial to compare the two interventions, however these estimations regarding PREE pain outcome may be incorrect and the higher revision rate at two years for DHH reported by the National Joint Registry raises questions regarding the cost effectiveness of the two interventions and the surgical burden placed in the patient that warrant further investigation.

# 3 Discussion

## 3.1 Summary and contribution to the field

The aim of this thesis was to explore outcomes that could be used in a clinical trial of a low-volume surgical intervention (elbow replacement) for the management of unreconstructible acute distal humerus fractures in adults. Assessment of new surgical interventions is important to ensure that they are safe and effective. For rare interventions, it is even more important to undertake structured evaluation, because the experience of an individual clinician may not highlight high rates of implant failure, and harms, such as re-operation, that may affect a larger group of patients if early evaluation is not undertaken.(Hobby *et al.*, 2008)

In a staged approach I have demonstrated that there is substantial inconsistency in the outcome measured employed in the existing literature on elbow replacement, and I have worked with key stakeholders using a structured real-time Delphi approach to define a core outcome domain set that is relevant to patients and should be included in the outcome set by researchers investigating the clinical outcomes of elbow replacement for any indication. Patient representatives have made it clear that from their perspective the most important outcome is a measure of pain, rather than function scores that are most commonly used. This has a degree of face validity, as the majority of patients that I see who are unhappy after a total elbow replacement for trauma, are complaining of pain, not loss of function. This insight is valuable to the design of any future trials of elbow replacement interventions, and would be a key outcome in a future trial of elbow replacement for management of acute unreconstructible distal humerus fractures in adults. Core time points for outcome collection after the intervention were identified as 3 months, 12 months and 24 months. It is too early to assess outcomes following elbow replacement at 3 months, because patients are still recovering and outcomes are unlikely to have peaked, and longer follow-up may lead to higher attrition. Therefore, a future trial might use 12 months for the primary outcome when comparing the efficacy of the interventions. I have synthesised existing knowledge on pain outcomes for elbow replacement in acute trauma in a systematic review and meta-analysis, identifying differences in pain distribution but uncertainty about the findings given the small sample size and heterogeneity of included data. I found that the instruments

commonly used to measure pain have not been validated for elbow replacement. There is an outcome instrument to measure pain that has been validated for use following elbow replacement (Patient Rated Elbow Evaluation pain (PREE pain)), and having determined that there is no direct data to determine probability distributions for this outcome instrument in the setting of replacement for acute elbow trauma, I have completed an expert elicitation exercise, using the weighted pooled outcome data from this evidence synthesis to inform the deliberations of elbow replacement experts, to generate prior probability distribution functions that can be used to plan a future trial.

A review of existing evidence is vital before embarking on primary research to avoid wasting resources in research that will not provide any additional knowledge.(Chalmers *et al.*, 2014) A systematically conducted scoping review following current best methodological practice and reporting standards mapped existing evidence based on 362 included studies on elbow replacement.(M. Peters *et al.*, 2020)(Tricco *et al.*, 2018) The population of interest for the scoping review and the real-time Delphi study to define core outcome domains was patients undergoing any form of elbow replacement, including total elbow replacement, distal humerus hemi-replacement, radial head replacement and radio-capitellar replacement. The scoping review identified that the majority of the literature on elbow replacement, on which we base our decisions and information for patients, is from relatively small case series, with the inherent risk of biased observations.(Watts *et al.*, 2022) Where randomised trials have been conducted in elbow replacement, the sample sizes were found to be small and the studies are likely to be underpowered. The assessment of interventions occurs in an unstructured way and could be improved by the adoption of stepwise approaches to the development of surgical interventions as described by the IDEAL framework.(McCulloch *et al.*, 2009) Most studies are unfunded, which is not unique to elbow replacement research, but the low incidence of elbow replacement is likely to be the main reason for these small trials.(Rankin *et al.*, 2014) From background reading into clinical trials in rare diseases outlined in the introduction, it became clear that the majority of techniques used in drug trials to mitigate the problem of a small available sample size cannot be used in surgical trials. Surgical interventions are irreversible in nature and usually require prolonged follow up to determine success or failure.(Jansen *et al.*, 2017) This means that any design that requires withdrawal of treatment or early outcome, such as cross-over designs, N-of-one or

adaptive randomisation are not possible to apply.(Whicher, Philbin and Aronson, 2018) National research collaborative networks in the UK have supported recruitment to multicentre studies in orthopaedics and have led to successful research delivery.(Rangan, Brealey and Carr, 2012) For rare interventions it may be necessary to explore international collaboration to recruit sufficient participants over a reasonable recruitment period, but there are challenges with data sharing, aligning protocols to local practice, and gaining agreement on intellectual property.(Costa, Brealey and Perry, 2023) When the disease or intervention is very rare, even this approach may be challenging as it may be necessary to involve many countries to achieve the required sample size, with an accompanying administrative burden, but here initiatives such as the NIHR Global Health Research Unit may provide a useful framework for researchers (https://www.globalsurgeryunit.org/).

The scoping review of the literature demonstrated that there was substantial inconsistency in both what is measured (583 outcome descriptors across 362 studies) and how outcomes are measured, with 105 instruments used to measure 39 outcomes.(Watts *et al.*, 2022) Many studies did not identify a primary outcome, and most outcomes reported were not patient rated. Given the clear lack of consensus about which outcomes should be assessed, in advance of the proposed randomised trial to compare outcomes of TER and DHH, it was essential to establish which outcomes are important to patients and clinicians. A core outcome study was, therefore, conducted using best practice guidelines for core outcome set surveys and engaging key stake-holders to identify what should be measured in elbow replacement research.(Kirkham *et al.*, 2017)(Kirkham *et al.*, 2019) The reporting was aligned to the best practice guidelines for reporting.(Kirkham *et al.*, 2016) Delphi consensus techniques can suffer from attrition and respondent fatigue so a real-time Delphi approach was adopted, to enable participants to engage with the survey in a single round.(Trevelyan and Robinson, 2015) Participants could return to the survey as many times as they wanted over a period of the survey and could see in real-time the aggregate responses of other participants for each question, and where they felt that an outcome was substantially more, or less, important (a pre-determined 2 point difference on a 9-point scale) than the real-time aggregate, they were able to inform other participants of the reasons for their response. The nine mandatory core domains identified for elbow replacement were “return to work or normal daily role”, delivery of care measured in the domains “patient satisfaction with the

outcome of surgery" and "would the patient have the same operation again", "pain", "revision", "elbow function", "independence in activities of daily living", "health related quality of life" and "adverse events", and there was an additional outcome "range of movement", that was considered important by clinicians but less important by patients so that was considered a desirable but not mandatory domain. To direct research resources more efficiently, these domains should be used by all researchers investigating elbow replacement, rather than outcomes that are not considered important by stakeholders, such as radiographic outcomes. Physical function and adverse events were found to be the most commonly described outcomes in the scoping review (paper 1), but the finding from the CODER study that patients consider the measurement of pain outcome most important should result in researchers considering this key outcome as the primary outcome for future trials where appropriate for the research question. The collaboration determined that data should be collected at baseline before the intervention, 3 months, 12 months and 24 months post intervention as a minimum. Data collection at 6 months post intervention was considered important by clinicians, but patients felt this could be avoided to reduce burden on trial participants, and it is therefore considered a desirable but not mandatory time-point. Other data time-points may be useful depending on the clinical question and time points for data collection beyond two-years were not considered.

The CODER core outcome set (Paper 2) is new and reflects the outcome domains that patients and clinicians consider to be important for elbow replacement. If these outcomes are measured consistently by researchers, it will enable reporting of data that is relevant to stakeholders, will aid comparison of outcomes between trials and facilitate meta-analysis, which can be helpful in rare disease research. There are issues with adoption of core outcome sets by the research community, with evidence that take-up can be low.(Matvienko-Sikar *et al.*, 2022) The most frequently identified barriers to the use of a potentially relevant core outcome set identified in a survey were researcher preference of alternative outcomes, a lack of awareness of available core outcome sets, and barriers to use of recommended outcome instruments.(Matvienko-Sikar *et al.*, 2022) The CODER outcome set describes core domains that can be readily assessed with patient rated instruments, and the manuscript has been published in a high impact orthopaedic journal and presented at the European Society for Surgery of the Shoulder and Elbow annual scientific meeting.

Further work is need to define the best instruments to measure the core domains and to continue to raise awareness of the core outcome set, emphasising the importance of incorporation of the outcomes into future trials to help reduce research waste.

The CODER study found that the pain domain is the key outcome of interest to patients and carers following elbow replacement surgery. It has also been shown previously by others that pain disproportionately affects composite outcome scores for patients after elbow trauma surgery.(Doornberg, 2005) The most recently published systematic reviews exploring outcomes following TER and DHH have not reported pain outcomes at all, focusing instead on composite outcome instruments such as the Disabilities of the Arm Shoulder (DASH) and Hand or Mayo Elbow Performance Score (MEPS) that combine assessment of pain, function and other domains into a single score, or on physician assessed outcomes such as range of movement.(Stoddart *et al.*, 2022)(Burden *et al.*, 2022) The only randomised trial comparing TER and DHH did not report pain as a separate domain from elbow function, reporting instead the composite outcome of the DASH Score, as the primary outcome, and MEPS.(Jonsson *et al.*, 2023)(Hudak, Amadio and Bombardier, 1996)(Morrey and Adams, 1992) It was therefore necessary to undertake an evidence synthesis of pain outcomes for elbow replacement conducted for acute elbow trauma in a systematic review of the published literature. From the scoping review (Paper 1) it was already apparent that it would not be possible to undertake a comparative meta-analysis of effect estimates of the available data for the two different types of elbow replacement used in acute trauma (total elbow replacement and distal humerus hemi-replacement) due to a lack of comparative trials, instead the guidelines for systematic review without meta-analysis (SWiM) were followed allowing transparent reporting and descriptive meta-analysis of outcome data for each intervention without comparison.(Campbell *et al.*, 2020) The majority of the literature reporting pain outcomes for elbow replacement for trauma, consisting largely of retrospective observational studies, used either a numerical rating scale or Likert scale for measurement of pain outcomes, either in isolation or as part of a combined pain and function score as reported in paper 3. Both of these instruments are unidimensional, measuring only one characteristic of pain (e.g. worst pain, average pain, etc.) and neither have been validated for use in elbow replacement. The evidence synthesis from this systematic review should, therefore, not be used directly to inform future trial planning, but

may help in the synthesis of expected pain outcomes using instruments that have been validated for used after elbow replacement. The finding that more than 10% of patients report moderate pain after distal humerus hemi-replacement needs to be interpreted with caution due to the small sample sizes of included studies and heterogeneity of reported results and populations studied, but does justify further investigation of effect estimates of pain outcomes following TER and DHH given the importance that patients attach to this outcome domain.

The patient rated elbow evaluation (PREE) is the only patient rated instrument that could be identified that has been validated for the assessment of pain after elbow replacement, and that measures the frequency of pain symptoms.(MacDermid, 2001)(Angst *et al.*, 2005)(Angst *et al.*, 2012) The PREE could be used as the primary outcome in a future trial. It measures pain across five domains from a score of zero for no pain in any situation to a maximum pain score of 50. The PREE also measures function separately on a scale from zero to fifty, and the two can be combined to give a total score out of 100. It is recommended however that pain and function are reported separately. PREE scores following elbow replacement for trauma have not been reported. The expert elicitation exercise reported in paper 4, used the evidence synthesis of pain outcomes from paper 3, to inform experts in elbow replacement for trauma to provide an estimate of the anticipated pain outcomes using the PREE for patients undergoing elbow replacement for trauma at 12 months after the intervention. A group of seven experts in elbow replacement, with a balance of users of TER and DHH for acute trauma, were guided through an expert elicitation exercise using SHELF methodology to produce estimated parameters for the PREE pain outcomes.(Oakley and O'Hagan, Anthony, 2019) They reached consensus that in a trial measuring the pain outcome at 12 months in a trial comparing TER to DHH for acute trauma, there would be a median difference in the PREE pain scale of 3.4 (IQR -0.3 to 7.1). The estimated standard deviation of the PREE pain scores after TER at 12 months was 9 (SD 2.2), giving an estimated effect size of 0.38. Classical power analysis using the parameters elicited in PEDANT performed in G*Power (Version 3.1.9.6) with $\alpha$ = 0.05 and 1-$\beta$ = 0.8 using a-priori mean (SD) PREE pain score for TER of 14(9) and DHH 17(12) with a two-tailed T-test with 1:1 allocation indicates a sample size of 396, with 198 patients completing each arm of the trial (Appendix O).(Faul *et al.*, 2007) Assuming an attrition rate of 20% between recruitment and primary

follow up at 12 months 495 participants would have to be randomised. If an assumption is made, based on the SOFF Trial that compares two interventions of similar intensity in elbow trauma, that 70% of eligible patients, that is adult patients with an unreconstructible distal humerus fracture consenting to participate, will agree to randomisation, a future study would need 707 participants to be eligible from screening.(Cook *et al.*, 2023) There were 144 elbow replacement for trauma in 2022 undertaken in 90 hospitals recorded in the NJR.(Achakri *et al.*, 2023) Even if it were possible to approach all these patients, and assuming that all surgeons had the equipoise to randomise patients to the study, it would still take at least five years to reach the sample size required if recruiting within England, Wales and Northern Ireland. An international multicentre trial could be designed to achieve target more quickly, if funding could be secured and agreements reached with researchers in other countries.

Whilst it is recognised that Bayesian inference cannot replace the degree of certainty about an effect size that a suitably powered and well conducted classical frequentist design trial provides, it may be a pragmatic solution for interventions that are very rare such as elbow arthroplasty for trauma. From the PEDANT study (paper 4) non-informative, sceptical and enthusiastic priors have been described that can be used in a future trial using Bayesian inference to compare the outcome of TER and DHH for the management of acute unreconstructible distal humerus fractures in adults as measured by the PREE pain scale at 12 months. It is clear that these estimates are unlikely to precisely represent the true population probability density functions, but collection of likelihood observations from a well-designed randomised trial and application of Bayesian transformation to produce posterior distributions, will enable investigators to use the available resources to obtain a better estimate of the true population parameters, with a variety of probability estimates, some sceptical, some enthusiastic, that can be used to inform balanced future decision making. With a Bayesian approach we can use the knowledge that we already have from existing data, and update our knowledge from new data through the application of Bayes theorum. This approach has been criticised for being too subjective because the prior is based on our beliefs about an outcome, but it has been argued that this is no more subjective than the calculations used in power analysis to inform classical trial design.(Berry, 1993) A Bayesian approach has the added benefit that the output is likely to be more

intuitive to the clinician because the Bayesian inference of outcome probabilities aligns more closely with the way clinicians think in practice.(Gill, Sabin and Schmid, 2005)

## 3.2 Strengths and limitations

In this thesis a stepwise approach has been taken to address gaps in knowledge regarding the best method to assess an important clinical question, whether patients should be treated with TER or DHH for acute unreconstructible distal humerus fracture. The research landscape for elbow replacement was mapped using established guidelines for scoping reviews, indicating that there was a lack of consistency in outcome domains being assessed, and that the majority of the literature to frame evidence-based practice is small retrospective case series with inherent risks of bias.(M. D. J. Peters *et al.*, 2020)(Tricco *et al.*, 2018)(Watts *et al.*, 2022) There is clear evidence of a lack of funding in elbow replacement research, which may be related to the low volumes of surgery, and therefore low economic priority. As with rare, or orphan diseases, the low incidence of the pathology makes the clinical decision making no less important for the individual patient.(Elliott and Zurynski, 2015) Key stakeholders, including patients, carers, primary care practitioners, physiotherapists and orthopaedic surgeons were engaged in the process to identify nine core outcome domains that should be captured at baseline and three months, 12 and 24 months after the intervention, to ensure that the limited resources available for elbow replacement research are used as efficiently as possible. However, the number of patients participating in the survey was lower than anticipated, in part due to the multi-stage process to register their interest and access the software, required because of data protection and the use of a United States based survey software company. In order to ensure the patients voice was adequately represented additional work was undertaken to review the core domain set with a PPI group. Whilst the PPI group participating in the CODER study included a representation of important stakeholders, including patients, carers, representative of patients with disabilities and patients with experience of research, only two participants had lived experience of total elbow replacement. None-the-less, the unanimity amongst the group regarding the most important core outcome, pain domain, was notable and should be

given further consideration by researchers and clinicians who tend to focus on function or composite outcomes.

If adopted by the research community this core domain set would be of significant benefit to enable comparison between studies and pooling of data. This impact is not guaranteed however, and despite efforts to disseminate the core outcome set, more work is needed to ensure that it is adopted by the research community.(Matvienko-Sikar *et al.*, 2022) The patients made it clear their determination that the key outcome domain from their point of view is pain. This was particularly important because as a research community we have tended to prioritise functional outcomes or instruments that combine a number of domains into one overall score, as in the only published randomised trial to compare TER to DHH for acute fractures.(Jonsson *et al.*, 2023) We should conclude that, where appropriate to the clinical question, pain should be the primary outcome domain used to assess elbow replacement.

Pain outcomes have not be assessed in previous systematic reviews of elbow replacement outcomes for acute trauma, which have focused on function or multi-domain outcome instruments, range of movement, adverse events and re-operation.(Stoddart *et al.*, 2022)(Burden *et al.*, 2022) Paper 3 was the first systematic review to describe pain outcomes following elbow replacement for trauma. This review was limited by the quality of the available literature which consists largely of retrospective case series. Clinical and methodological heterogeneity prevented comparison between DHH and TER outcome data. All data was given weighting depending on the sample size, rather than assessing their direct relevance to the research question. Tan et al. have proposed that for the establishment of priors, data should be assessed for pertinence (how close the information is to that being collected in a trial based, for example, on disease, demographic or intervention characteristics), validity (quality of trials based on their design) and precision (the reliability of the data obtained). Data can be downrated using a correction factor described in their paper based on the evaluation of these three factors.(Tan, 2003) The review of published literature identified that previous studies have used pain instruments, such as the numerical rating scales or Likert scales for pain, neither of which has been validated for use in elbow replacement research. No formal systematic review was undertaken to identify pain

instruments with suitable psychometric properties, but only one patient rated pain instrument, the PREE, could be identified that had been validated for use in this setting.(MacDermid, 2001)(Vincent and MacDermid, 2012) The ASES elbow score could be an alternative pain instrument for a future trial, but requires physician assessment of some elements of the score and does not measure frequency of pain.(King *et al.*, 1999) Neither of these instruments have been subjected to a detailed analysis using the COSMIN or EMPRO instrument to objectively determine the psychometric properties and suitability for this outcome domain, and this work should be completed before widespread adoption in a core outcome set.(Terwee *et al.*, 2018)(Valderas *et al.*, 2008)

The use of subjective priors, such as those produced from expert elicitation studies, has raised concerns amongst the statistical community.(Whitehead, 1993) Whilst it must be acknowledged that if the prior probability density function is markedly wrong, it will draw the likelihood data away from the true population probability, this is in some way mitigated by the use of multiple priors, no-informative, sceptical and enthusiastic that can be assessed to draw balanced inferences. It is expected of course that suitably informed, knowledgeable and experienced experts will elicit a probability density function that is more informative than a neutral prior. There are risks, however, in the elicitation process with experts subject to particular biases including anchoring, where an expert is influenced by a previous number, availability or ease of recall of an event may influence the expert's perception of frequency of events and overconfidence, where the confidence interval limits are too narrow is a common phenomenon.(O'Hagan, 2006) To mitigate these risks I attempted to achieve a balance amongst the experts of level of experience and preference for use of TER or DHH for acute trauma in their own practice. There is also a concern in any group consensus exercise that strong personalities may influence the outcome. I minimised these risks by closely adhering to the SHELF methodology. I provided the experts with a summary of the evidence on pain outcomes based on paper 3 to tackle the availability problem. At the start of each individual elicitation, I conducted a trial elicitation, asking the experts to elicit the median weight of a male Southern African Lion, to highlight the problems of overconfidence and placing quartiles too far from the median. I carefully constructed the questioning to avoid anchoring, and to address the challenges of group psychology the experts were first asked to express their own beliefs in PDFs before coming together as a

group to consider what a rational independent observer would consider to be the truth, based on the experts' individual PDFs and the arguments put forward in discussion.(Oakley and O'Hagan, Anthony, 2019) This approach may more easily allow experts to consider a position that is discordant to their own beliefs. The facilitator, too, can have an influence on the success of the expert elicitation exercise. Facilitation is a challenging undertaking that requires good knowledge of the background material but also of the psychological principles of elicitation, and like all things experience can improve the quality of facilitation. As the facilitator in the PEDANT study, I was inexperienced as a facilitator, and my expertise in elbow replacement surgery could have influenced the outcome of the PEDANT study. I had attended the SHELF methodology training course to learn SHELF techniques to avoid introducing bias, but this lack of experience could have influenced the outcome of the elicitation exercise. Great care was taken, however, to avoid influencing the experts when eliciting their own beliefs and during the group exercise, in particular to avoid any anchoring. The experts were all orthopaedic surgeons, and the results of the PEDANT study may not be representative of all stakeholders. The inclusion of patients would have been challenging, given the understanding of probability distributions required for expert elicitation, however inclusion of physiotherapists in the expert panel may have increased the diversity of opinion regarding likely pain outcomes.

## 3.3 Implications for patients and carers

The work presented in this thesis has importantly taken great care to engage patients and carers in the selection of outcomes that should be measured. The use of the core outcome set should accelerate scientific discovery in the field of elbow replacement research, and importantly reduce waste of resources in trials that use outcomes that are not relevant to stakeholders, which will be to the benefit of patients. Patients have told us that we need to prioritise the assessment of pain outcomes when evaluating the effectiveness of elbow replacement interventions, which is an important message for researchers in this field.

The thesis presents evidence that there may be differences in the pain outcomes between TER and DHH conducted for acute trauma, but that there is substantial uncertainty regarding the size and significance of any differences. The four studies described in this thesis have

brought us closer to a trial that can evaluate the effectiveness of these rare interventions for patients with acute distal humerus fractures.

## 3.4 Implications for healthcare providers and policy makers

Medical devices, unlike pharmaceuticals, can be released into the market without evidence of effectiveness. Distal humerus hemi-replacement has overtaken total elbow replacement as the most commonly used treatment for unreconstructible distal humerus fractures in adults in England since design and development of a specific hemi-replacement prosthesis in the first decade of this century. The scoping review (paper 1) indicates that up until 2023, and the publication of the first randomised trial to compare the effectiveness to total elbow replacement, the only evidence on which to base the use of this device was small retrospective observational case series studies.(Watts *et al.*, 2022) The adoption of structured approaches to the assessment of interventions, as outlined in the IDEAL framework, could accelerate development of devices and techniques, and could improve the evidence on which clinical decisions should be made.(McCulloch *et al.*, 2009)

Policy makers, funders and researches should recognise that the majority of orthopaedic interventions are rare, but the number of rare interventions being offered means that there are many patients that are being impacted by a lack of good quality scientific evidence on which to inform healthcare decision making. These challenges are likely to apply to other surgical specialties and beyond, into other medical disciplines. Policy makers should consideration whether there is a need to develop specialised funding strategies to address this underserved population, as has been done with the orphan drug act for research into pharmaceuticals.('Orphan Drug Act', 1983) The low incidence of many orthopaedic interventions means that trials conducted using classic frequentist methods are may not be feasible without considerable resource. Bayesian trial methodology and inference may have particularly utility in scientific research where there are barriers to alternative techniques to deal with the low incidence of interventions or where the population under study is very heterogeneous.(Jansen *et al.*, 2017) The main barriers to adoption of Bayesian trials are reported to be a lack of knowledge about the Bayesian approach, and lack of guidance from

regulators.(The Medical Outreach Subteam of the Drug Information Association Bayesian Scientific Working Group *et al.*, 2023)

Healthcare researchers should adopt the core domains identified in CODER (Paper 2) in future elbow replacement research to improve consistency of outcome reporting, to ensure that they are assessing outcomes of importance to patients and clinicians, and reduce waste in research and the burden for patients. Funding panels should be looking for inclusion of the core domains and outcome time points when assessing future funding applications in this area. Healthcare providers and policy makers should consider whether the increased use of DHH over TER can be justified based on the current evidence, and whether more detailed scientific assessment of these interventions in a prospective randomised trial is required to determine which intervention results in the best outcome for our patients. Although outside the scope of this thesis cost effectiveness analysis should also be undertaken to compare these interventions and their associated additional costs related to rehabilitation resources, adverse events or re-operations.

## 3.5 Implications for research and future research priorities

A systematic review of the instruments that can be used to measure the core domains identified in CODER and formal assessment of candidate instruments using the COSMIN or EMPRO tool should be undertaken to determine the core outcome instrument set for elbow replacement research.(Terwee *et al.*, 2018)(Valderas *et al.*, 2008) The results for the core outcome instruments should be reported in future research into elbow replacement, at the specified time points of baseline, 3 months, 12 and 24 months post intervention. The need for additional outcome assessment beyond two years has not been considered in this thesis. In line with the requirement of the Orthopaedic Data Evaluation Panel (ODEP), it would be sensible for long term outcome analysis, to collect data at 3, 5, 7 and 10 years.(Orthopaedic Data Evaluation Panel, 2021)

The FOREST trial (NIHR154915) is in set-up, with sites preparing to begin recruitment. This will randomise 320 adult patients, aged 60 years and over, with distal humerus fractures to either osteosynthesis or replacement. The choice of replacement, total elbow replacement

or hemi-replacement, will be decided by the treating surgeon. There is no intention to randomise the patients in the replacement group because of concern that this could impact the feasibility of the trial, as some surgeons are not comfortable performing a hemi-replacement. The primary outcome for this trial is the DASH score, but the PREE has been adopted as a secondary outcome measure, which will provide valuable prospective data on the outcome distributions for total elbow replacement and hemi-replacement in acute trauma using this instrument.

Until this data is available, the prior probability distributions provided in this thesis can be used to inform a future randomised controlled trial of TER and DHH for acute unreconstructible distal humerus fracture in adults. Given the low incidence of this intervention, this is likely to require a classical trial design with a prolonged recruitment period, or establishment of an international trial to establish sufficient recruiting sites to more rapidly reach the required sample size. Alternatively, Bayesian trial methods could be used based on the established priors, with the support of a statistician with expertise in this field.

Further research should be conducted to establish the minimally clinically important difference for the PREE after elbow replacement which can be used to inform future trials and to develop an alternative informative prior.(Vincent and MacDermid, 2012)

# 4 Conclusions

A randomised trial is needed to determine the effectiveness, safety and cost effectiveness of the different elbow replacement approaches for acute unreconstructible distal humerus fracture. There are barriers to evaluating elbow replacements, including inconsistencies in the outcomes that have been measured in historical studies, which has been addressed by the development of a core outcome domain set for elbow replacement. This core outcome set includes return to work or normal daily role, patient satisfaction with the outcome of surgery, would the patient have the same operation again, pain, revision, elbow function, independence in activities of daily living, health related quality of life and adverse events including death. These core outcomes should be assessed at baseline and 3, 12 and 24 months after the intervention. Pain was considered by patients to be a key outcome. Evidence synthesis of pain outcomes confirms that there is a need for a well conducted randomised trial to compare pain outcomes using validated instruments such as the PREE pain score. Estimates for the distribution of PREE pain scores at 12 months after elbow replacement for acute trauma have been synthesised, that can inform trial planning. A future trial could be conducted using a Bayesian framework.

# 5 References


Achakri, H. *et al.* (2023) *The National Joint Registry 20th Annual Report 2023*. London: National Joint Registry (National Joint Registry Annual Reports). Available at: http://www.ncbi.nlm.nih.gov/books/NBK601075/ (Accessed: 5 June 2024).

Adolfsson, L. (2011) 'Elbow Hemiarthroplasty Versus Total Elbow Arthroplasty for Irreparable Distal Humeral Fractures'. Available at: https://ClinicalTrials.gov/show/NCT03596736.

Adolfsson, L. and Hammer, R. (2006) 'Elbow hemiarthroplasty for acute reconstruction of intraarticular distal humerus fractures: A preliminary report involving 4 patients', *Acta Orthopaedica*, 77(5), pp. 785–787. Available at: https://doi.org/10.1080/17453670610012999.

Aitken, S.A., Jenkins, P.J. and Rymaszewski, L. (2015) 'Revisiting the "bag of bones": functional outcome after the conservative management of a fracture of the distal humerus', *The Bone & Joint Journal*, 97-B(8), pp. 1132–1138. Available at: https://doi.org/10.1302/0301-620X.97B8.35410.

Akhtar, A., Hughes, B. and Watts, A.C. (2021) 'The post-traumatic stiff elbow: A review', *Journal of Clinical Orthopaedics and Trauma*, 19, pp. 125–131. Available at: https://doi.org/10.1016/j.jcot.2021.05.006.

Al-Hamdani, A. *et al.* (2019) 'Good outcome after elbow hemiarthroplasty in active patients with an acute intra-articular distal humeral fracture', *Journal of Shoulder and Elbow Surgery*, 28(5), pp. 925–930. Available at: https://doi.org/10.1016/j.jse.2018.10.018.

Al-Hamdani, A. *et al.* (2020) 'Elbow hemiarthroplasty versus open reduction and internal fixation for AO/OTA type 13 C2 and C3 fractures of distal humerus in patients aged 50 years or above: a randomized controlled trial', *Trials*, 21(1), p. 497. Available at: https://doi.org/10.1186/s13063-020-04418-8.

Alhussain, Z.A. and Oakley, J.E. (2020) 'Assurance for clinical trial design with normally distributed outcomes: Eliciting uncertainty about variances', *Pharmaceutical Statistics*, 19(6), pp. 827–839. Available at: https://doi.org/10.1002/pst.2040.

Ali, A., Shahane, S. and Stanley, D. (2010) 'Total elbow arthroplasty for distal humeral fractures: indications, surgical approach, technical tips, and outcome', *Journal of Shoulder and Elbow Surgery*, 19(2 Suppl), pp. 53–58. Available at: https://doi.org/10.1016/j.jse.2009.12.013.

Altman, D.G. (1980) 'Statistics and ethics in medical research. How large a sample?', *BMJ*, 281, pp. 1336–1338.

Angst, F. *et al.* (2005) 'Comprehensive assessment of clinical outcome and quality of life after total elbow arthroplasty', *Arthritis Care & Research*, 53(1), pp. 73–82. Available at: https://doi.org/10.1002/art.20911.

Angst, F. *et al.* (2012) 'Responsiveness of five outcome measurement instruments in total elbow arthroplasty', *Arthritis Care & Research*, 64(11), pp. 1749–1755. Available at: https://doi.org/10.1002/acr.21744.

Antuña, S.A. *et al.* (2012) 'Linked total elbow arthroplasty as treatment of distal humerus fractures', *Acta Orthopaedica Belgica*, 78(4), pp. 465–472.

Argintar, E. *et al.* (2012) 'Hemiarthroplasty for the Treatment of Distal Humerus Fractures: Short-term Clinical Results', *Orthopedics*. Edited by B.G.P. Lindeque, 35(12), pp. 1042–1045. Available at: https://doi.org/10.3928/01477447-20121120-06.

Baik, J.S. *et al.* (2020) 'Comparison of open reduction and internal fixation with total elbow arthroplasty for intra-articular distal humeral fractures in older age: a retrospective study', *Clinics in Shoulder and Elbow*, 23(2), pp. 94–99. Available at: https://doi.org/10.5397/cise.2020.00052.

Bailar, J.C. (1997) 'The Promise and Problems of Meta-Analysis', *New England Journal of Medicine*, 337(8), pp. 559–561. Available at: https://doi.org/10.1056/NEJM199708213370810.

Barco, R. *et al.* (2017) 'Total Elbow Arthroplasty for Distal Humeral Fractures: A Ten-Year-Minimum Follow-up Study', *Journal of Bone and Joint Surgery*, 99(18), pp. 1524–1531. Available at: https://doi.org/10.2106/JBJS.16.01222.

Barr, J.S. and Eaton, R.G. (no date) 'Elbow Reconstruction with a New Prosthesis to Replace the Distal End of the Humerus: A CASE REPORT', *JBJS*, 47(7), pp. 1408–1413.

Bayes, T. (1763) 'LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S.', *Philosophical transactions of the Royal Society of London*, 53, pp. 370–418.

Beard, D.J. *et al.* (2018) 'Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicentre, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial', *The Lancet*, 391(10118), pp. 329–338. Available at: https://doi.org/10.1016/S0140-6736(17)32457-1.

Ben-Shlomo, Y. *et al.* (2021) *The National Joint Registry 18th Annual Report 2021*. London: National Joint Registry (National Joint Registry Annual Reports). Available at: http://www.ncbi.nlm.nih.gov/books/NBK576858/ (Accessed: 22 March 2022).

Bergdahl, C. *et al.* (2016) 'Epidemiology and patho-anatomical pattern of 2,011 humeral fractures: data from the Swedish Fracture Register', *BMC Musculoskeletal Disorders*, 17(1), p. 159. Available at: https://doi.org/10.1186/s12891-016-1009-8.

Berry, D.A. (1993) 'A case for bayesianism in clinical trials', *Statistics in Medicine*, 12(15–16), pp. 1377–1393. Available at: https://doi.org/10.1002/sim.4780121504.

Blencowe, N.S. *et al.* (2015) 'Interventions in randomised controlled trials in surgery: issues to consider during trial design', *Trials*, 16(1), p. 392. Available at: https://doi.org/10.1186/s13063-015-0918-4.

Boers, M. *et al.* (2014) 'Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0', *Journal of Clinical Epidemiology*, 67(7), pp. 745–753. Available at: https://doi.org/10.1016/j.jclinepi.2013.11.013.

Bojke, L. *et al.* (2021) 'Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study', *Health Technology Assessment*, 25(37), pp. 1–124. Available at: https://doi.org/10.3310/hta25370.

Burden, E.G. *et al.* (2022) 'Hemiarthroplasty or total elbow arthroplasty for unreconstructable distal humeral fractures in patients aged over 65 years: a systematic review and meta-analysis of patient outcomes and complications', *The Bone & Joint Journal*, 104-B(5), pp. 559–566. Available at: https://doi.org/10.1302/0301-620X.104B5.BJJ-2021-1207.R2.

Burkhart, K.J. *et al.* (2011) 'Distal Humerus Hemiarthroplasty of the Elbow for Comminuted Distal Humeral Fractures in the Elderly Patient', *Journal of Trauma: Injury, Infection & Critical Care*, 71(3), pp. 635–642. Available at: https://doi.org/10.1097/TA.0b013e318216936e.

Campbell, M. *et al.* (2020) 'Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline', *BMJ*, p. l6890. Available at: https://doi.org/10.1136/bmj.l6890.

Celli, A. *et al.* (2021) 'Total elbow arthroplasty for acute distal humeral fractures with humeral condyle resection or retention: a long-term follow-up study', *JSES International*, 5(4), pp. 797–803. Available at: https://doi.org/10.1016/j.jseint.2021.03.006.

Celli, A. *et al.* (2022) 'Elbow hemiarthroplasty for acute distal humeral fractures and their sequelae: medium- and long-term follow-up of 41 cases', *Journal of Shoulder and Elbow Surgery*, 31(5), pp. 1015–1025. Available at: https://doi.org/10.1016/j.jse.2021.12.030.

Chalmers, I. *et al.* (2014) 'How to increase value and reduce waste when research priorities are set', *The Lancet*, 383(9912), pp. 156–165. Available at: https://doi.org/10.1016/S0140-6736(13)62229-1.

Charissoux, J.-L. *et al.* (2013) 'Epidemiology of distal humerus fractures in the elderly', *Orthopaedics & Traumatology: Surgery & Research*, 99(7), pp. 765–769. Available at: https://doi.org/10.1016/j.otsr.2013.08.002.

Chen, X. *et al.* (2011) 'Comparison between radial head replacement and open reduction and internal fixation in clinical treatment of unstable, multi-fragmented radial head fractures', *International Orthopaedics*, 35(7), pp. 1071–1076. Available at: https://doi.org/10.1007/s00264-010-1107-4.

Cheng, Y. and Sheng, X. (2020) 'Optimal surgical methods to treat intertrochanteric fracture: a Bayesian network meta-analysis based on 36 randomized controlled trials', *Journal of Orthopaedic surgery and research*, 15(1), pp. 1–14.

Chou, T.-F.A. *et al.* (2020) 'Total elbow arthroplasty in patients with rheumatoid arthritis: a systematic review and meta-analysis', *The Bone & Joint Journal*, 102-B(8), pp. 967–980. Available at: https://doi.org/10.1302/0301-620X.102B8.BJJ-2019-1465.R1.

Claxton, K., Lacey, L.F. and Walker, S.G. (2000) 'Selecting Treatments: A Decision Theoretic Approach', *Journal of the Royal Statistical Society Series A: Statistics in Society*, 163(2), pp. 211–225. Available at: https://doi.org/10.1111/1467-985X.00166.

Cobb, T.K. and Morrey, B.F. (1997) 'Total Elbow Arthroplasty as Primary Treatment for Distal Humeral Fractures in Elderly Patients*':, *The Journal of Bone & Joint Surgery*, 79(6), pp. 826–32. Available at: https://doi.org/10.2106/00004623-199706000-00004.

'Cochrane Handbook for Systematic Reviews of Interventions version 6.4' (2023). Cochrane. Available at: www.training.cochrane.org/handbook.

Colebrook, D.C. (1929) 'Irradiation and Health'. H.M.Stationary Office.

Cook, E. *et al.* (2023) 'A randomized controlled trial to compare clinical and cost-effectiveness of suture fixation versus tension band wiring for simple olecranon fracture fixation in adults: The Simple Olecranon Fracture Fixation Trial (SOFFT) protocol', *Bone & Joint Open*, 4(1), pp. 27–37. Available at: https://doi.org/10.1302/2633-1462.41.BJO-2022-0132.R1.

Cook, J.A. *et al.* (2018) 'DELTA $^{2}$ guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial', *BMJ*, p. k3750. Available at: https://doi.org/10.1136/bmj.k3750.

Cooke, R.M. (1991) *Experts in uncertainty: opinion and subjective probability in science.* USA: Oxford University Press.

Cornfield, J. (1969) 'The Bayesian Outlook and Its Applications', *Biometrics*, 25(4), pp. 617–657.

Cornu, C. *et al.* (2013) 'Experimental designs for small randomised clinical trials: an algorithm for choice', *Orphanet Journal of Rare Diseases*, 8(1), p. 48. Available at: https://doi.org/10.1186/1750-1172-8-48.

Costa, M.L., Brealey, S.D. and Perry, D.C. (2023) 'International orthopaedic trials: global reach for a global burden of disease', *The Bone & Joint Journal*, 105-B(7), pp. 719–722. Available at: https://doi.org/10.1302/0301-620X.105B7.BJJ-2023-0070.

Court-Brown, C.M. and Caesar, B. (2006) 'Epidemiology of adult fractures: A review', *Injury*, 37(8), pp. 691–697. Available at: https://doi.org/10.1016/j.injury.2006.04.130.

Dawson, J. *et al.* (2008) 'The development and validation of a patient-reported questionnaire to assess outcomes of elbow surgery', *The Journal of Bone and Joint Surgery. British volume*, 90-B(4), pp. 466–473. Available at: https://doi.org/10.1302/0301-620X.90B4.20290.

Dee, R. (1975) 'Elbow replacement with the R. Dee prosthesis', *Acta Orthopaedica Belgica*, 41(4), pp. 477–483.

Dehghan, N. *et al.* (2019) 'Long-term outcomes of total elbow arthroplasty for distal humeral fracture: results from a prior randomized clinical trial', *Journal of Shoulder and Elbow Surgery*, 28(11), pp. 2198–2204. Available at: https://doi.org/10.1016/j.jse.2019.06.004.

Dias, J. *et al.* (2021) 'Dupuytren's interventions surgery versus collagenase (DISC) trial: study protocol for a pragmatic, two-arm parallel-group, non-inferiority randomised controlled trial', *Trials*, 22(1), p. 671. Available at: https://doi.org/10.1186/s13063-021-05595-w.

Dirckx, M. *et al.* (2023) 'Hemiarthroplasty versus open reduction internal fixation for intra-articular distal humerus fractures in older patients', *Shoulder & Elbow*, 15(1), pp. 83–92. Available at: https://doi.org/10.1177/17585732221093004.

Dodd, S. *et al.* (2018) 'A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery', *Journal of Clinical Epidemiology*, 96, pp. 84–92. Available at: https://doi.org/10.1016/j.jclinepi.2017.12.020.

Donovan, T. and Mickey, R.M. (2019) *Bayesian Statistics for Beginners: a step-by-step approach*. 1st edn. Oxford University PressOxford. Available at: https://doi.org/10.1093/oso/9780198841296.001.0001.

Doornberg, J.N. (2005) 'Pain Dominates Measurements of Elbow Function and Health Status', *The Journal of Bone and Joint Surgery (American)*, 87(8), p. 1725. Available at: https://doi.org/10.2106/JBJS.D.02745.

Dubberley, J.H. (2006) 'Outcome After Open Reduction and Internal Fixation of Capitellar and Trochlear Fractures', *The Journal of Bone and Joint Surgery (American)*, 88(1), p. 46. Available at: https://doi.org/10.2106/JBJS.D.02954.

Dunn, J., Kusnezov, N. and Pirela-Cruz, M. (2014) 'Distal Humeral Hemiarthroplasty: Indications, Results, and Complications. A Systematic Review', *HAND*, 9(4), pp. 406–412. Available at: https://doi.org/10.1007/s11552-014-9681-3.

Edwards, S. *et al.* (1997) 'Why "underpowered" trials are not necessarily unethical', *The Lancet*, 350(9080), pp. 804–807. Available at: https://doi.org/10.1016/S0140-6736(97)02290-3.

Ekhtiari, S. *et al.* (2020) 'Total Hip Arthroplasty Versus Hemiarthroplasty for Displaced Femoral Neck Fracture: A Systematic Review and Meta-Analysis of Randomized Controlled Trials', *Journal of Bone and Joint Surgery*, 102(18), pp. 1638–1645. Available at: https://doi.org/10.2106/JBJS.20.00226.

Elliott, E.J. and Zurynski, Y.A. (2015) 'Rare diseases are a common problem for clinicians', *Australian Family Physician*, 44(9), pp. 630–633.

Eurordis (2005) 'Rare Diseases: Understanding this public health priority'. European Organisation for Rare Diseases. Available at: www.eurordis.org.

Evans, J.P. *et al.* (2018) 'Clinical rating systems in elbow research—a systematic review exploring trends and distributions of use', *Journal of Shoulder and Elbow Surgery*, 27(4), pp. e98–e106. Available at: https://doi.org/10.1016/j.jse.2017.12.027.

Faul, F. *et al.* (2007) 'G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences', *Behavior Research Methods*, 39(2), pp. 175–191. Available at: https://doi.org/10.3758/BF03193146.

Figgie, M.P. *et al.* (2006) 'Locking mechanism failure in semiconstrained total elbow arthroplasty', *Journal of Shoulder and Elbow Surgery*, 15(1), pp. 88–93. Available at: https://doi.org/10.1016/j.jse.2005.05.007.

Flandoli, F. *et al.* (2011) 'Comparison of a new expert elicitation model with the Classical Model, equal weights and single experts, using a cross-validation technique', *Reliability Engineering & System Safety*, 96(10), pp. 1292–1310. Available at: https://doi.org/10.1016/j.ress.2011.05.012.

Frankle, M.A. *et al.* (2003) 'A Comparison of Open Reduction and Internal Fixation and Primary Total Elbow Arthroplasty in the Treatment of Intraarticular Distal Humerus Fractures in Women Older Than Age 65':, *Journal of Orthopaedic Trauma*, 17(7), pp. 473–480. Available at: https://doi.org/10.1097/00005131-200308000-00001.

Gagne, J.J. *et al.* (2014) 'Innovative research methods for studying treatments for rare diseases: methodological review', *BMJ*, 349(nov24 15), pp. g6802–g6802. Available at: https://doi.org/10.1136/bmj.g6802.

Gallin, J.I. *et al.* (2003) 'Itraconazole to prevent fungal infections in chronic granulomatous disease', *New England Journal of Medicine*, 348(24), pp. 2416–2422.

Gallucci, G.L. *et al.* (2016) 'Total elbow arthroplasty for the treatment of distal humeral fractures', *Revista Española de Cirugía Ortopédica y Traumatología (English Edition)*, 60(3), pp. 167–174. Available at: https://doi.org/10.1016/j.recote.2016.03.003.

Gambirasio, R. *et al.* (2001) 'Total elbow replacement for complex fractures of the distal humerus: AN OPTION FOR THE ELDERLY PATIENT', *The Journal of Bone and Joint Surgery. British volume*, 83-B(7), pp. 974–978. Available at: https://doi.org/10.1302/0301-620X.83B7.0830974.

Garcia, J.A., Mykula, R. and Stanley, D. (2002) 'Complex fractures of the distal humerus in the elderly: THE ROLE OF TOTAL ELBOW REPLACEMENT AS PRIMARY TREATMENT', *The Journal of Bone and Joint Surgery. British volume*, 84-B(6), pp. 812–816. Available at: https://doi.org/10.1302/0301-620X.84B6.0840812.

Geurts, E.J. *et al.* (2019) 'Outcomes after revision total elbow arthroplasty: a systematic review', *Journal of Shoulder and Elbow Surgery*, 28(2), pp. 381–386. Available at: https://doi.org/10.1016/j.jse.2018.08.024.

Gill, C.J., Sabin, L. and Schmid, C.H. (2005) 'Why clinicians are natural bayesians', *BMJ*, 330(7499), pp. 1080–1083. Available at: https://doi.org/10.1136/bmj.330.7499.1080.

Githens, M. *et al.* (2014) 'Open Reduction and Internal Fixation Versus Total Elbow Arthroplasty for the Treatment of Geriatric Distal Humerus Fractures: A Systematic Review and Meta-Analysis', *Journal of Orthopaedic Trauma*, 28(8), pp. 481–488. Available at: https://doi.org/10.1097/BOT.0000000000000050.

Goldberg, A.J. *et al.* (2023) 'Total ankle replacement versus ankle arthrodesis for patients aged 50–85 years with end-stage ankle osteoarthritis: the TARVA RCT', *Health Technology Assessment*, 27(05), pp. 1–80. Available at: https://doi.org/10.3310/PTYJ1146.

Goldberg, S.H. *et al.* (2008) 'Modes of Wear After Semiconstrained Total Elbow Arthroplasty':, *The Journal of Bone and Joint Surgery-American Volume*, 90(3), pp. 609–619. Available at: https://doi.org/10.2106/JBJS.F.01286.

Gupta, S. *et al.* (2011) 'A framework for applying unfamiliar trial designs in studies of rare diseases', *Journal of Clinical Epidemiology*, 64(10), pp. 1085–1094. Available at: https://doi.org/10.1016/j.jclinepi.2010.12.019.

Halpern, S.D. (2002) 'The Continuing Unethical Conduct of Underpowered Clinical Trials', *JAMA*, 288(3), p. 358. Available at: https://doi.org/10.1001/jama.288.3.358.

Halpern, S.D., Karlawish, J.H. and Berlin, J.A. (2002) 'The continuing unethical conduct of underpowered clinical trials', *JAMA*, 288(3), pp. 358–362.

Hamoodi, Z. *et al.* (In press) 'Total elbow replacement in England: analysis of National Joint Registry and Hospital Episode Statistics data', *Bone & Joint Journal* [Preprint].

Hampson, L.V. *et al.* (2014) 'Bayesian methods for the design and interpretation of clinical trials in very rare diseases', *Statistics in Medicine*, 33(24), pp. 4186–4201. Available at: https://doi.org/10.1002/sim.6225.

Hardy, R.J. and Thompson, S.G. (1998) 'Detecting and describing heterogeneity in meta-analysis', *Statistics in Medicine*, 17(8), pp. 841–856. Available at: https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<841::AID-SIM781>3.0.CO;2-D.

Heifner, J.J. *et al.* (2024) 'The age specific performance of elbow hemiarthroplasty: A systematic review', *Shoulder & Elbow*, p. 17585732241246756. Available at: https://doi.org/10.1177/17585732241246756.

Herr, K.A. *et al.* (2004) 'Pain Intensity Assessment in Older Adults: Use of Experimental Pain to Compare Psychometric Properties and Usability of Selected Pain Scales With Younger Adults', *The Clinical Journal of Pain*, 20(4), pp. 207–219. Available at: https://doi.org/10.1097/00002508-200407000-00002.

Higgins, J.P.T. and Thompson, S.G. (2002) 'Quantifying heterogeneity in a meta-analysis', *Statistics in Medicine*, 21(11), pp. 1539–1558. Available at: https://doi.org/10.1002/sim.1186.

Hobby, J.L. *et al.* (2008) 'A Report on the Early Failure of the LPM Proximal Interphalangeal Joint Replacement', *Journal of Hand Surgery (European Volume)*, 33(4), pp. 526–527. Available at: https://doi.org/10.1177/1753193408094920.

Hozo, S.P., Djulbegovic, B. and Hozo, I. (2005) 'Estimating the mean and variance from the median, range, and the size of a sample', *BMC Medical Research Methodology*, 5(1), p. 13. Available at: https://doi.org/10.1186/1471-2288-5-13.

Hudak, P.L., Amadio, P.C. and Bombardier, C. (1996) 'Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG)', *American Journal of Industrial Medicine*, 29(6), pp. 602–608. Available at: https://doi.org/10.1002/(SICI)1097-0274(199606)29:6<602::AID-AJIM4>3.0.CO;2-L.

Jamison, J.C. (2019) 'The Entry of Randomized Assignment into the Social Sciences', *Journal of Causal Inference*, 7(1), p. 20170025. Available at: https://doi.org/10.1515/jci-2017-0025.

Jansen, J.O. *et al.* (2017) 'Bayesian clinical trial designs: Another option for trauma trials?', *Journal of Trauma and Acute Care Surgery*, 83(4), pp. 736–741. Available at: https://doi.org/10.1097/TA.0000000000001638.

Jenkins, P.J. *et al.* (2013) 'Total elbow replacement: outcome of 1,146 arthroplasties from the Scottish Arthroplasty Project', *Acta orthopaedica*, 84(2), pp. 119–123.

Jonsson, E.Ö. *et al.* (2023) 'Elbow hemiarthroplasty and total elbow arthroplasty provided a similar functional outcome for unreconstructable distal humeral fractures in patients aged 60 years or older: a multicenter randomized controlled trial', *Journal of Shoulder and Elbow Surgery*, p. S105827462300705X. Available at: https://doi.org/10.1016/j.jse.2023.08.026.

Jost, B., Adams, R.A. and Morrey, B.F. (2008) 'Management of Acute Distal Humeral Fractures in Patients with Rheumatoid Arthritis: A Case Series', *The Journal of Bone and Joint Surgery-American Volume*, 90(10), pp. 2197–2205. Available at: https://doi.org/10.2106/JBJS.G.00024.

Jupiter, J.B. and Mehne, D.K. (1992) 'FRACTURES OF THE DISTAL HUMERUS', *Orthopedics*, 15(7), pp. 825–833. Available at: https://doi.org/10.3928/0147-7447-19920701-07.

Kamineni, S. and Morrey, B.F. (2004) 'Distal Humeral Fractures Treated with Noncustom Total Elbow Replacement':, *The Journal of Bone & Joint Surgery*, 86(5), pp. 940–947. Available at: https://doi.org/10.2106/00004623-200405000-00009.

Kannus, P. *et al.* (2017) 'Continuous decline in incidence of distal humeral fracture of older women in Finland', *Aging Clinical and Experimental Research*, 29(3), pp. 467–471. Available at: https://doi.org/10.1007/s40520-016-0594-6.

Katz, N. (2002) 'The Impact of Pain Management on Quality of Life', *Journal of Pain and Symptom Management*, 24(1), pp. S38–S47. Available at: https://doi.org/10.1016/S0885-3924(02)00411-6.

King, G.J. *et al.* (1999) 'A standardized method for assessment of elbow function', *Journal of Shoulder and Elbow Surgery*, 8(4), pp. 351–354. Available at: https://doi.org/10.1016/S1058-2746(99)90159-3.

Kirkham, J.J. *et al.* (2016) 'Core Outcome Set–STAndards for Reporting: The COS-STAR Statement', *PLOS Medicine*, 13(10), p. e1002148. Available at: https://doi.org/10.1371/journal.pmed.1002148.

Kirkham, J.J. *et al.* (2017) 'Core Outcome Set-STAndards for Development: The COS-STAD recommendations', *PLOS Medicine*, 14(11), p. e1002447. Available at: https://doi.org/10.1371/journal.pmed.1002447.

Kirkham, J.J. *et al.* (2019) 'Core Outcome Set-STAndardised Protocol Items: the COS-STAP Statement', *Trials*, 20(1), p. 116. Available at: https://doi.org/10.1186/s13063-019-3230-x.

Klifto, K., Klifto, C. and Slover, J. (2017) 'Current concepts of shared decision making in orthopedic surgery', *Current Reviews in Musculoskeletal Medicine*, 10(2), pp. 253–257. Available at: https://doi.org/10.1007/s12178-017-9409-4.

Kontakis, G. *et al.* (2008) 'Early management of proximal humeral fractures with hemiarthroplasty: A SYSTEMATIC REVIEW', *The Journal of Bone and Joint Surgery. British volume*, 90-B(11), pp. 1407–1413. Available at: https://doi.org/10.1302/0301-620X.90B11.21070.

Kyriacou, S. *et al.* (2019) 'Radial head replacement versus reconstruction for the treatment of the terrible triad injury of the elbow: a systematic review and meta-analysis', *Archives of Orthopaedic and Trauma Surgery*, 139(4), pp. 507–517. Available at: https://doi.org/10.1007/s00402-019-03111-z.

Landis, J.R. and Koch, G.G. (1977) 'An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers', *Biometrics*, 33(2), p. 363. Available at: https://doi.org/10.2307/2529786.

Lee, B.P., Adams, R.A. and Morrey, B.F. (2005) 'Polyethylene Wear After Total Elbow Arthroplasty':, *The Journal of Bone & Joint Surgery*, 87(5), pp. 1080–1087. Available at: https://doi.org/10.2106/JBJS.D.02163.

Lee, K.T., Lai, C.H. and Singh, S. (2006) 'Results of Total Elbow Arthroplasty in the Treatment of Distal Humerus Fractures in Elderly Asian Patients':, *The Journal of Trauma: Injury, Infection, and Critical Care*, 61(4), pp. 889–892. Available at: https://doi.org/10.1097/01.ta.0000215421.77665.7a.

Lewis, R.J. and Wears, R.L. (1993) 'An introduction to the Bayesian analysis of clinical trials', *Annals of Emergency Medicine*, 22(8), pp. 1328–1336. Available at: https://doi.org/10.1016/S0196-0644(05)80119-2.

Lim, H.C. *et al.* (2014) 'Randomised Trial Support for Orthopaedic Surgical Procedures', *PLoS ONE*. Edited by M. Elsalanty, 9(6), p. e96745. Available at: https://doi.org/10.1371/journal.pone.0096745.

Lind, J. (1753) 'A Treatise of the Scurvy. In Three Parts. Containing an Inquiry into the Nature, Causes and Cure, of that Disease. Together with a Critical and Chronological View of what has been Published on the Subject'. Kincaid and Donaldson, Edinburgh.

Little, C.P., Graham, A.J. and Carr, A.J. (2005) 'Total elbow arthroplasty: a systematic review of the literature in the English language until the end of 2003', *The Journal of Bone and Joint Surgery. British Volume*, 87(4), pp. 437–444. Available at: https://doi.org/10.1302/0301-620X.87B4.15692.

Logli, A.L. *et al.* (2020) 'Total Elbow Arthroplasty for Distal Humerus Fractures Provided Similar Outcomes When Performed as a Primary Procedure or After Failed Internal Fixation', *Journal of Orthopaedic Trauma*, 34(2), pp. 95–101. Available at: https://doi.org/10.1097/BOT.0000000000001631.

van der Lugt, J.C.T. and Rozing, P.M. (2004) 'Systematic review of primary total elbow prostheses used for the rheumatoid elbow', *Clinical Rheumatology*, 23(4). Available at: https://doi.org/10.1007/s10067-004-0884-9.

MacAusland, W.R. (1954) 'Replacement of the lower end of the humerus with a prosthesis; a report of four cases.', *Western journal of surgery, obstetrics, and gynecology*, 62(11), pp. 557–566.

MacDermid, J.C. (2001) 'Outcome evaluation in patients with elbow pathology: Issues in instrument development and evaluation', *Journal of Hand Therapy*, 14(2), pp. 105–114. Available at: https://doi.org/10.1016/S0894-1130(01)80040-5.

Mansat, P. *et al.* (2013) 'Total elbow arthroplasty for acute distal humeral fractures in patients over 65 years old – Results of a multicenter study in 87 patients', *Orthopaedics & Traumatology: Surgery & Research*, 99(7), pp. 779–784. Available at: https://doi.org/10.1016/j.otsr.2013.08.003.

Matvienko-Sikar, K. *et al.* (2022) 'Use of core outcome sets was low in clinical trials published in major medical journals', *Journal of Clinical Epidemiology*, 142, pp. 19–28. Available at: https://doi.org/10.1016/j.jclinepi.2021.10.012.

McCulloch, P. (2002) 'Randomised trials in surgery: problems and possible solutions', *BMJ*, 324(7351), pp. 1448–1451. Available at: https://doi.org/10.1136/bmj.324.7351.1448.

McCulloch, P. *et al.* (2009) 'No surgical innovation without evaluation: the IDEAL recommendations', *The Lancet*, 374(9695), pp. 1105–1112. Available at: https://doi.org/10.1016/S0140-6736(09)61116-8.

McKee, M.D. *et al.* (2009) 'A multicenter, prospective, randomized, controlled trial of open reduction—internal fixation versus total elbow arthroplasty for displaced intra-articular distal humeral fractures in elderly patients', *Journal of Shoulder and Elbow Surgery*, 18(1), pp. 3–12. Available at: https://doi.org/10.1016/j.jse.2008.06.005.

McLeod, C. *et al.* (2019) 'Choosing primary endpoints for clinical trials of health care interventions', *Contemporary Clinical Trials Communications*, 16, p. 100486. Available at: https://doi.org/10.1016/j.conctc.2019.100486.

Medley, N. *et al.* (2021) 'Developing a topic-based repository of clinical trial individual patient data: experiences and lessons learned from a pilot project', *Systematic Reviews*, 10(1), p. 162. Available at: https://doi.org/10.1186/s13643-021-01717-2.

Mellex, R.H. and Phalen, G.S. (1947) 'ARTHROPLASTY OF THE ELBOW BY REPLACEMENT OF THE DISTAL PORTION OF THE HUMERUS WITH AN ACRYLIC PROSTHESIS', *JBJS*, 29(2). Available at: https://journals.lww.com/jbjsjournal/Fulltext/1947/29020/ARTHROPLASTY_OF_THE_ELBOW_BY_REPLACEMENT_OF_THE.10.aspx.

Metcalfe, A. *et al.* (2022) 'Subacromial balloon spacer for irreparable rotator cuff tears of the shoulder (START:REACTS): a group-sequential, double-blind, multicentre randomised controlled trial', *The Lancet*, 399(10339), pp. 1954–1963. Available at: https://doi.org/10.1016/S0140-6736(22)00652-3.

Migliorini, F. *et al.* (2020) 'Comparison between intra-articular infiltrations of placebo, steroids, hyaluronic and PRP for knee osteoarthritis: a Bayesian network meta-analysis', *Archives of Orthopaedic and Trauma Surgery*, pp. 1–18.

Mikita, J.S. *et al.* (2021) 'Determining the Suitability of Registries for Embedding Clinical Trials in the United States: A Project of the Clinical Trials Transformation Initiative', *Therapeutic Innovation & Regulatory Science*, 55(1), pp. 6–18. Available at: https://doi.org/10.1007/s43441-020-00185-5.

Morrey, B.F. and Adams, R.A. (1992) 'Semiconstrained arthoplasty for the treatment of rheumatoid arthitis of the elbow', *The Journal of Bone & Joint Surgery*, 74(4), pp. 479–490.

Müller, M.E. *et al.* (1990) *The Comprehensive Classification of Fractures of Long Bones*. Berlin, Heidelberg: Springer Berlin Heidelberg. Available at: https://doi.org/10.1007/978-3-642-61261-9.

Mundi, R. *et al.* (2014) 'Design and execution of clinical trials in orthopaedic surgery', *Bone & Joint Research*, 3(5), pp. 161–168. Available at: https://doi.org/10.1302/2046-3758.35.2000280.

Munn, Z. *et al.* (2018) 'Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach', *BMC Medical Research Methodology*, 18(1), p. 143. Available at: https://doi.org/10.1186/s12874-018-0611-x.

Munn, Z. *et al.* (2020) 'Methodological quality of case series studies: an introduction to the JBI critical appraisal tool', *JBI Evidence Synthesis*, 18(10), pp. 2127–2133. Available at: https://doi.org/10.11124/JBISRIR-D-19-00099.

Nestorson, J. *et al.* (2015) 'Hemiarthroplasty for irreparable distal humeral fractures: medium-term follow-up of 42 patients', *The Bone & Joint Journal*, 97-B(10), pp. 1377–1384. Available at: https://doi.org/10.1302/0301-620X.97B10.35421.

Nielsen, A.F. *et al.* (2022) 'Elbow hemiarthroplasty vs. open reduction internal fixation for acute Arbeitsgemeinschaft für Osteosynthesefragen/Orthopaedic Trauma Association (AO/OTA) type 13C fractures—A systematic review', *JSES International*, 6(5), pp. 713–722. Available at: https://doi.org/10.1016/j.jseint.2022.06.002.

Oakley, J. (2015) 'SHELF: Tools to Support the Sheffield Elicitation Framework'. Available at: https://doi.org/10.32614/CRAN.package.SHELF.

Oakley, J.E. and O'Hagan, Anthony (2019) 'SHELF: the Sheffield Elicitation Framework (version 4)'. School of Mathematics and Statistics, University of Sheffield, UK. Available at: https://shelf.sites.sheffield.ac.uk/home.

O'Hagan, A. (ed.) (2006) *Uncertain judgements: eliciting experts' probabilities*. London ; Hoboken, NJ: John Wiley & Sons (Statistics in practice).

O'Hagan, A. (2019) 'Expert Knowledge Elicitation: Subjective but Scientific', *The American Statistician*, 73(sup1), pp. 69–81. Available at: https://doi.org/10.1080/00031305.2018.1518265.

Oremus, M. *et al.* (2002) 'A survey of physician efficacy requirements to plan clinical trials', *Pharmacoepidemiology and Drug Safety*, 11(8), pp. 677–685. Available at: https://doi.org/10.1002/pds.750.

'Orphan Drug Act' (1983). Available at: https://www.govinfo.gov/content/pkg/STATUTE-96/pdf/STATUTE-96-Pg2049.pdf.

Orthopaedic Data Evaluation Panel (2021) 'ODEP Elbow Clinical Reference Data Summary'. Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.odep.org.uk/wp-content/uploads/2022/07/ODEP_Criteria_Elbows_2021.pdf.

Ouzzani, M. *et al.* (2016) 'Rayyan—a web and mobile app for systematic reviews', *Systematic Reviews*, 5(1), p. 210. Available at: https://doi.org/10.1186/s13643-016-0384-4.

Palmer, A.J. *et al.* (1997) 'Efficacy and safety of low molecular weight heparin, unfractionated heparin and warfarin for thrombo-embolism prophylaxis in orthopaedic surgery: a meta-analysis of randomised clinical trials', *Pathophysiology of Haemostasis and Thrombosis*, 27(2), pp. 75–84.

Palvanen, M. *et al.* (2010) 'Secular trends in distal humeral fractures of elderly women', *Bone*, 46(5), pp. 1355–1358. Available at: https://doi.org/10.1016/j.bone.2009.11.025.

Parmar, M.K.B., Sydes, M.R. and Morris, T.P. (2016) 'How do you design randomised trials for smaller populations? A framework', *BMC Medicine*, 14(1), p. 183. Available at: https://doi.org/10.1186/s12916-016-0722-3.

Peters, M. *et al.* (2020) 'Chapter 11: Scoping Reviews', in E. Aromataris and Z. Munn (eds) *JBI Manual for Evidence Synthesis*. JBI. Available at: https://doi.org/10.46658/JBIMES-20-12.

Peters, M.D.J. *et al.* (2020) 'Updated methodological guidance for the conduct of scoping reviews', *JBI Evidence Synthesis*, 18(10), pp. 2119–2126. Available at: https://doi.org/10.11124/JBIES-20-00167.

Phadnis, J., Banerjee, S., Watts, Adam C., *et al.* (2015) 'Elbow hemiarthroplasty using a "triceps-on" approach for the management of acute distal humeral fractures', *Journal of Shoulder and Elbow Surgery*, 24(8), pp. 1178–1186. Available at: https://doi.org/10.1016/j.jse.2015.04.010.

Phadnis, J., Banerjee, S., Watts, Adam C, *et al.* (2015) 'Elbow hemiarthroplasty using a "triceps-on" approach for the management of acute distal humeral fractures', *Journal of Shoulder and Elbow Surgery*, 24(8), pp. 1178–1186.

Pibouleau, L. and Chevret, S. (2011) 'Bayesian statistical method was underused despite its advantages in the assessment of implantable medical devices', *Journal of Clinical Epidemiology*, 64(3), pp. 270–279. Available at: https://doi.org/10.1016/j.jclinepi.2010.03.018.

Pogliacomi, F. *et al.* (2016) 'Total elbow arthroplasty following complex fractures of the distal humerus: results in patients over 65 years of age', *Acta Bio-Medica: Atenei Parmensis*, 87(2), pp. 148–155.

Prasad, N., Ali, A. and Stanley, D. (2016) 'Total elbow arthroplasty for non-rheumatoid patients with a fracture of the distal humerus: a minimum ten-year follow-up', *The Bone & Joint Journal*, 98-B(3), pp. 381–386. Available at: https://doi.org/10.1302/0301-620X.98B3.35508.

Prasad, N. and Dent, C. (2008) 'Outcome of total elbow replacement for distal humeral fractures in the elderly: A COMPARISON OF PRIMARY SURGERY AND SURGERY AFTER FAILED INTERNAL FIXATION OR CONSERVATIVE TREATMENT', *The Journal of Bone and Joint Surgery. British volume*, 90-B(3), pp. 343–348. Available at: https://doi.org/10.1302/0301-620X.90B3.18971.

PRISMA-S Group *et al.* (2021) 'PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews', *Systematic Reviews*, 10(1), p. 39. Available at: https://doi.org/10.1186/s13643-020-01542-z.

Prkic, A. *et al.* (2017) 'Why does total elbow arthroplasty fail today? A systematic review of recent literature', *Archives of Orthopaedic and Trauma Surgery*, 137(6), pp. 761–769. Available at: https://doi.org/10.1007/s00402-017-2687-x.

Ramanan, A.V. *et al.* (2019) 'Defining consensus opinion to develop randomised controlled trials in rare diseases using Bayesian design: An example of a proposed trial of adalimumab versus pamidronate for children with CNO/CRMO', *PLoS One*, 14(6), p. e0215739.

Ramiro, S. *et al.* (2019) 'The OMERACT Core Domain Set for Clinical Trials of Shoulder Disorders', *The Journal of Rheumatology*, 46(8), pp. 969–975. Available at: https://doi.org/10.3899/jrheum.181070.

Rangan, A. *et al.* (2020) 'Management of adults with primary frozen shoulder in secondary care (UK FROST): a multicentre, pragmatic, three-arm, superiority randomised clinical trial', *The Lancet*, 396(10256), pp. 977–989. Available at: https://doi.org/10.1016/S0140-6736(20)31965-6.

Rangan, A., Brealey, S. and Carr, A. (2012) 'Orthopaedic Trial Networks', *Journal of Bone and Joint Surgery*, 94(Supplement_1), pp. 97–100. Available at: https://doi.org/10.2106/JBJS.L.00241.

Rankin, K.S. *et al.* (2014) 'The orthopaedic research scene and strategies to improve it', *The Bone & Joint Journal*, 96-B(12), pp. 1578–1585. Available at: https://doi.org/10.1302/0301-620X.96B12.33990.

Ray, P.S. *et al.* (2000) 'Total elbow arthroplasty as primary treatment for distal humeral fractures in elderly patients', *Injury*, 31(9), pp. 687–692. Available at: https://doi.org/10.1016/S0020-1383(00)00076-0.

Reinecke, M., Rommel, K. and Schmidtke, J. (2011) 'Funding of rare disease research in Germany: a pilot study', *Journal of Community Genetics*, 2(2), pp. 101–105. Available at: https://doi.org/10.1007/s12687-011-0045-1.

Riedel, K. and Beaton, D.E. (2013) 'Update on the State of Outcome Measurement in Total Elbow Arthroplasty Research: Identifying a Need for Consensus', *The Journal of Bone and Joint Surgery-American Volume*, 95(14), pp. e97-1–8. Available at: https://doi.org/10.2106/JBJS.K.01420.

Riseborough, E. and Radin, E. (1969) 'Intercondylar T Fractures of the Humerus in the Adult', *Journal of Bone and Joint Surgery*, 51(1), pp. 130–141.

Rotini, R. *et al.* (2020) 'Elbow hemiarthroplasty in distal humeral fractures: Indication, surgical technique and results', *Injury*, p. S0020138320309463. Available at: https://doi.org/10.1016/j.injury.2020.11.020.

Ruan, H.-J. *et al.* (2009) 'A comparative study of internal fixation and prosthesis replacement for radial head fractures of Mason type III', *International Orthopaedics*, 33(1), pp. 249–253. Available at: https://doi.org/10.1007/s00264-007-0453-3.

Ryan, E.G. *et al.* (2020) 'Bayesian adaptive designs for multi-arm trials: an orthopaedic case study', *Trials*, 21(1), p. 83. Available at: https://doi.org/10.1186/s13063-019-4021-0.

Rydzewska, L.H.M., Stewart, L.A. and Tierney, J.F. (2022) 'Sharing individual participant data: through a systematic reviewer lens', *Trials*, 23(1), p. 167. Available at: https://doi.org/10.1186/s13063-021-05787-4.

Samdanis, V. *et al.* (2020) 'Indications and outcome in total elbow arthroplasty: A systematic review', *Shoulder & Elbow*, 12(5), pp. 353–361. Available at: https://doi.org/10.1177/1758573219873001.

Sathyamoorthy, P. (2004) 'Development and validation of an elbow score', *Rheumatology*, 43(11), pp. 1434–1440. Available at: https://doi.org/10.1093/rheumatology/keh367.

Schiavi, P. *et al.* (2020) 'Survival and outcome of total elbow arthroplasty for distal humeral fracture at long term follow-up.', *Acta Bio Medica Atenei Parmensis*, 91(14-S), p. e2020031. Available at: https://doi.org/10.23750/abm.v91i14-S.11112.

Schultzel, M. *et al.* (2017) 'Hemiarthroplasty for the treatment of distal humeral fractures: midterm clinical results', *Journal of Shoulder and Elbow Surgery*, 26(3), pp. 389–393. Available at: https://doi.org/10.1016/j.jse.2016.09.057.

Schulz, K.F. *et al.* (2010) 'CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials', *BMJ*, 340(mar23 1), pp. c332–c332. Available at: https://doi.org/10.1136/bmj.c332.

Shamseer, L. *et al.* (2015) 'Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation', *BMJ*, 349(jan02 1), pp. g7647–g7647. Available at: https://doi.org/10.1136/bmj.g7647.

Shea, B.J. *et al.* (2017) 'AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both', *BMJ*, p. j4008. Available at: https://doi.org/10.1136/bmj.j4008.

Sheps, D.M., Kemp, K.A. and Hildebrand, K.A. (2011) 'Population-based incidence of distal humeral fractures among adults in a Canadian urban center', *Current Orthopaedic Practice*, 22(5), pp. 437–442.

Shifrin, P.G. and Johnson, D.P. (1990) 'Elbow hemiarthroplasty with 20-year follow-up study. A case report and literature review.', *Clinical orthopaedics and related research*, (254), pp. 128–133.

Shukla, D.R. *et al.* (2016) 'Hemiarthroplasty versus reverse shoulder arthroplasty for treatment of proximal humeral fractures: a meta-analysis', *Journal of Shoulder and Elbow Surgery*, 25(2), pp. 330–340. Available at: https://doi.org/10.1016/j.jse.2015.08.030.

Shurin, S., Krischer, J. and Groft, S.C. (2012) 'Clinical trials In BMT: ensuring that rare diseases and rarer therapies are well done', *Biology of Blood and Marrow Transplantation*, 18(1), pp. S8–S11.

Singh, A.K. *et al.* (2019) 'Functional Outcome of Elbow Kinematics in Radial Head Excision Versus Radial Head Replacement: A Comparative Study', *Open Access Macedonian Journal of Medical Sciences*, 7(9), pp. 1505–1508. Available at: https://doi.org/10.3889/oamjms.2019.305.

Smith, C. (2020) ‘Hemiarthroplasty Or Total Elbow Arthroplasty in the Elderly’. Available at: https://clinicaltrials.gov/show/NCT04646798.

Smith, G.C.S. and Hughes, J.S. (2013) ‘Unreconstructable acute distal humeral fractures and their sequelae treated with distal humeral hemiarthroplasty: a two-year to eleven-year follow-up’, *Journal of Shoulder and Elbow Surgery*, 22(12), pp. 1710–1723. Available at: https://doi.org/10.1016/j.jse.2013.06.012.

Solomon, M.J. and McLeod, R.S. (1995) ‘Should we be performing more randomized controlled trials evaluating surgical operations?’, *Surgery*, 118(3), pp. 459–467. Available at: https://doi.org/10.1016/S0039-6060(05)80359-9.

Sørensen, B.W. (2014) ‘Primary total elbow arthroplasty in complex fractures of the distal humerus’, *World Journal of Orthopedics*, 5(3), p. 368. Available at: https://doi.org/10.5312/wjo.v5.i3.368.

Spiegelhalter, D.J. *et al.* (2000) ‘Bayesian methods in health technology assessment: a review’.

Sterne, J.A.C. *et al.* (2019) ‘RoB 2: a revised tool for assessing risk of bias in randomised trials’, *BMJ*, p. l4898. Available at: https://doi.org/10.1136/bmj.l4898.

Stoddart, M.T. *et al.* (2022) ‘A systematic review of the treatment of distal humerus fractures in older adults: A comparison of surgical and non-surgical options’, *Shoulder & Elbow*, p. 175857322210998. Available at: https://doi.org/10.1177/17585732221099845.

Street, D.M. and Stevens, P.S. (1974) ‘A humeral replacement prosthesis for the elbow: results in ten elbows.’, *The Journal of bone and joint surgery. American volume*, 56(6), pp. 1147–1158.

Strelzow, J.A. *et al.* (2021) ‘Results of Linked Convertible Total Elbow Arthroplasty for the Management of Distal Humeral Fractures in the Elderly’, *The Journal of Hand Surgery*, 46(5), pp. 396–402. Available at: https://doi.org/10.1016/j.jhsa.2020.10.034.

Tan, S.-B., Chung, Y.-F.A., *et al.* (2003) ‘Elicitation of prior distributions for a phase III randomized controlled trial of adjuvant therapy with surgery for hepatocellular carcinoma’, *Controlled Clinical Trials*, 24(2), pp. 110–121. Available at: https://doi.org/10.1016/S0197-2456(02)00318-5.

Tan, S.-B., Dear, K.B., *et al.* (2003) ‘Strategy for randomised clinical trials in rare cancers’, *Bmj*, 327(7405), pp. 47–49.

Tan, S.-B. (2003) ‘Strategy for randomised clinical trials in rare cancers’, *BMJ*, 327(7405), pp. 47–49. Available at: https://doi.org/10.1136/bmj.327.7405.47.

Tanaka, N. *et al.* (2006) ‘Comparison of two types of ulnar component in type-5 Kudo total elbow arthroplasty in patients with rheumatoid arthritis: a long-term follow-up’, *The Journal of Bone and Joint Surgery. British Volume*, 88(3), pp. 341–344. Available at: https://doi.org/10.1302/0301-620X.88B3.17017.

Terwee, C.B. *et al.* (2018) 'COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study', *Quality of Life Research*, 27(5), pp. 1159–1170. Available at: https://doi.org/10.1007/s11136-018-1829-0.

The Medical Outreach Subteam of the Drug Information Association Bayesian Scientific Working Group *et al.* (2023) 'Why are not There More Bayesian Clinical Trials? Perceived Barriers and Educational Preferences Among Medical Researchers Involved in Drug Development', *Therapeutic Innovation & Regulatory Science*, 57(3), pp. 417–425. Available at: https://doi.org/10.1007/s43441-021-00357-x.

'The OMERACT Handbook' (2021). Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://omeract.org/wp-content/uploads/2023/09/OMERACT-Handbook-Chapter-Final-June_21_2021.pdf.

Trehan, S.K. *et al.* (2015) 'The effect of simulated elbow contracture on temporal and distance gait parameters', *Gait & Posture*, 41(3), pp. 791–794. Available at: https://doi.org/10.1016/j.gaitpost.2015.02.010.

Trevelyan, E.G. and Robinson, P.N. (2015) 'Delphi methodology in health research: how to do it?', *European Journal of Integrative Medicine*, 7(4), pp. 423–428. Available at: https://doi.org/10.1016/j.eujim.2015.07.002.

Tricco, A.C. *et al.* (2018) 'PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation', *Annals of Internal Medicine*, 169(7), p. 467. Available at: https://doi.org/10.7326/M18-0850.

Valderas, J.M. *et al.* (2008) 'Development of EMPRO: A Tool for the Standardized Assessment of Patient-Reported Outcome Measures', *Value in Health*, 11(4), pp. 700–708. Available at: https://doi.org/10.1111/j.1524-4733.2007.00309.x.

Van der Lee, J.H. *et al.* (2008) 'Efficient ways exist to obtain the optimal sample size in clinical trials in rare diseases', *Journal of clinical epidemiology*, 61(4), pp. 324–330.

Vincent, J. and MacDermid, J.C. (2012) 'The Patient-Rated Elbow Evaluation (PREE)', *Journal of Physiotherapy*, 58(4), p. 274. Available at: https://doi.org/10.1016/S1836-9553(12)70134-0.

Vincent, J.I. *et al.* (2021) 'The Patient-Rated Elbow Evaluation and the American Shoulder and Elbow Surgeons—Elbow form capture aspects of functioning that are important to patients with elbow injuries', *Journal of Hand Therapy*, 34(3), pp. 415–422. Available at: https://doi.org/10.1016/j.jht.2020.02.001.

Wagner, E.R. *et al.* (2017) 'Comparison of the hospital costs for two-stage reimplantation for deep infection, single-stage revision and primary total elbow arthroplasty', *Shoulder & Elbow*, 9(4), pp. 279–284. Available at: https://doi.org/10.1177/1758573217706364.

Wainwright, A.M., Williams, J.R. and Carr, A.J. (2000) 'Interobserver and intraobserver variation in classification systems for fractures of the distal humerus', *The Journal of Bone*

*and Joint Surgery. British volume*, 82-B(5), pp. 636–642. Available at: https://doi.org/10.1302/0301-620X.82B5.0820636.

Wang, Ji *et al.* (2018) 'Bayesian network meta-analysis of the effectiveness of various interventions for nontraumatic osteonecrosis of the femoral head', *BioMed research international*, 2018.

Watts, A. *et al.* (2019) 'Scoping review: Diagnosis and management of periprosthetic joint infection in elbow arthroplasty', *Shoulder & Elbow*, 11(4), pp. 282–291. Available at: https://doi.org/10.1177/1758573218789341.

Watts, A. (2023a) 'Core Outcome Domains Elbow Replacement'. Available at: https://doi.org/10.17605/OSF.IO/JV2CB.

Watts, A. (2023b) 'Protocol for a systematic review and meta-analysis of pain outcomes following hemiarthroplasty and total elbow arthroplasty for unreconstructible acute distal humerus fractures in adults.' Available at: https://doi.org/10.17605/OSF.IO/3TKWH.

Watts, A., Morris, A. and Robinson, C. (2007) 'Fractures of the distal humeral articular surface', *Bone & Joint Journal*, 89(4), pp. 510–515.

Watts, A.C. *et al.* (2022) 'Elbow arthroplasty research methods, outcome domains, and instruments used in clinical outcome studies: a scoping review', *The Bone & Joint Journal*, 104-B(10), pp. 1148–1155. Available at: https://doi.org/10.1302/0301-620X.104B10.BJJ-2022-0570.R1.

Watts, A.C. *et al.* (2024) 'Core Outcome Domains in Elbow Replacement (CODER)', *Bone & Joint Journal* [Preprint].

Welsink, C.L. *et al.* (2017) 'Total Elbow Arthroplasty: A Systematic Review', *JBJS Reviews*, 5(7), pp. e4–e4. Available at: https://doi.org/10.2106/JBJS.RVW.16.00089.

Whicher, D., Philbin, S. and Aronson, N. (2018) 'An overview of the impact of rare disease characteristics on research methodology', *Orphanet Journal of Rare Diseases*, 13(1), p. 14. Available at: https://doi.org/10.1186/s13023-017-0755-5.

Whitehead, J. (1993) 'The case for frequentism in clinical trials', *Statistics in Medicine*, 12(15–16), pp. 1405–1413. Available at: https://doi.org/10.1002/sim.4780121506.

Wijeysundera, D.N. *et al.* (2009) 'Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials', *Journal of Clinical Epidemiology*, 62(1), pp. 13-21.e5. Available at: https://doi.org/10.1016/j.jclinepi.2008.07.006.

Williamson, P.R. *et al.* (2017) 'The COMET Handbook: version 1.0', *Trials*, 18(S3), p. 280. Available at: https://doi.org/10.1186/s13063-017-1978-4.

Wu, Z. *et al.* (2020) 'A Bayesian network meta-analysis of orthopaedic treatment in Class III malocclusion: Maxillary protraction with skeletal anchorage or a rapid maxillary expander', *Orthodontics & craniofacial research*, 23(1), pp. 1–15.

Yan, M. *et al.* (2015) 'Radial head replacement or repair for the terrible triad of the elbow: which procedure is better?', *ANZ journal of surgery*, 85(9), pp. 644–648. Available at: https://doi.org/10.1111/ans.13060.

Yetter, T.R., Weatherby, P.J. and Somerson, J.S. (2021) 'Complications of articular distal humeral fracture fixation: a systematic review and meta-analysis', *Journal of Shoulder and Elbow Surgery*, 30(8), pp. 1957–1967. Available at: https://doi.org/10.1016/j.jse.2021.02.017.

Zi-Sheng, A. *et al.* (2012) 'Hemiarthroplasty vs Primary Total Hip Arthroplasty For Displaced Fractures of the Femoral Neck in the Elderly', *The Journal of Arthroplasty*, 27(4), pp. 583–590. Available at: https://doi.org/10.1016/j.arth.2011.07.009.

■ SHOULDER & ELBOW

# Elbow arthroplasty research methods, outcome domains, and instruments used in clinical outcome studies

A SCOPING REVIEW

**A. C. Watts,**
**Z. Hamoodi,**
**C. McDaid,**
**C. Hewitt**

*From University of York, York, UK*

## Aims

**Arthroplasties of the elbow, including total elbow arthroplasty, radial head arthroplasty, distal humeral hemiarthroplasty, and radiocapitellar arthroplasty, are rarely undertaken. This scoping review aims to outline the current research in this area to inform the development of future research.**

## Methods

**A scoping review was undertaken adhering to the Joanna Briggs Institute guidelines using Medline, Embase, CENTRAL, and trial registries, limited to studies published between 1 January 1990 and 7 February 2021. Endnote software was used for screening and selection, and included randomized trials, non-randomized controlled trials, prospective and retrospective cohort studies, case-control studies, analytical cross-sectional studies, and case series of ten or more patients reporting the clinical outcomes of elbow arthroplasty. The results are presented as the number of types of studies, sample size, length of follow-up, clinical outcome domains and instruments used, sources of funding, and a narrative review.**

## Results

**A total of 362 studies met the inclusion criteria. Most were of total elbow arthroplasty (246; 68%), followed by radial head arthroplasty (100; 28%), distal humeral hemiarthroplasty (11; 3%), and radiocapitellar arthroplasty (5; 1%). Most were retrospective (326; 90%) and observational (315; 87%). The median sample size for all types of implant across all studies was 36 (interquartile range (IQR) 21 to 75). The median length of follow-up for all studies was 56 months (IQR 36 to 81). A total of 583 unique outcome descriptors were used and were categorized into 18 domains. A total of 105 instruments were used to measure 39 outcomes.**

## Conclusion

**We found that most of the literature dealing with elbow arthroplasty consists of retrospective observational studies with small sample sizes and short follow-up. Many outcomes have been used with many different instruments for their measurement, indicating a need to define a core set of outcomes and instruments for future research in this area.**

**Cite this article: *Bone Joint J* 2022;104-B(10):1148–1155.**

## Introduction

Arthroplasty of all or part of the elbow has been performed routinely for various indications since the cemented design of Dee was introduced in the 1960s, including for patients with inflammatory arthropathy, osteoarthritis (OA), trauma and its sequelae, instability, and (rarely) tumour.[1] The term 'elbow arthroplasty' is often taken to mean total elbow arthroplasty, involving replacement of the distal humerus and proximal ulna with or without replacing the radial head, but also may include replacement of the radial head or distal humerus in isolation and arthroplasty of the capitellum of the humerus and radial head (radiocapitellar arthroplasty). The aim of these procedures is to relieve pain and restore function and quality of life to the patient.

Clinical practice should be based on robust scientific evidence, and randomized controlled

Correspondence should be sent to A. C. Watts; email: adam.watts@elbowdoc.co.uk


doi:10.1302/0301-620X.104B10.BJJ-2022-0570.R1 $2.00

*Bone Joint J* 2022;104-B(10):1148–1155.

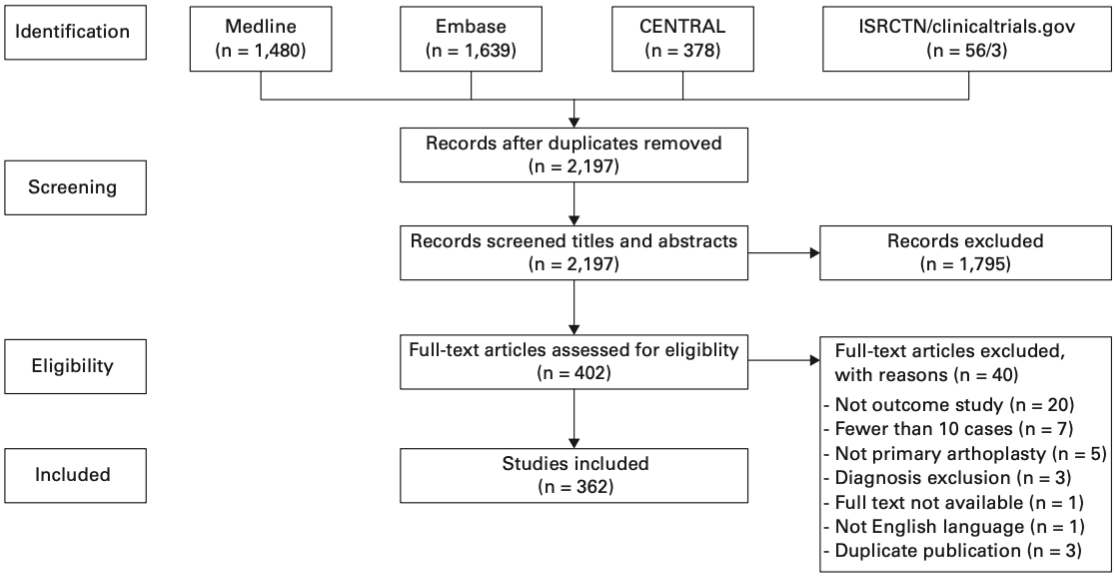



Fig. 1

Preferred Reporting Items for Systematic reviews and Meta-Analyses flowchart.

**Table I.** Medline search string (search conducted 7 February 2021).

| Terms for Medline search (Ovid MEDLINE(R) ALL <1946 to February 04, 2021>) | Search results retrieved |
|---|---|
| Elbow Prosthesis/ or Arthroplasty, Elbow/ | 485 |
| (elbow* adj3 (arthroplast* or arthroplasty* or hemiarthroplast* or hemireplacement*)).tw. | 1,217 |
| ((radial head or capitell*) adj3 (arthroplast* or arthroplasty* or hemiarthroplast* or hemireplacement*)).tw. | 453 |
| 1 or 2 or 3 | 1,721 |
| ((interposition or osteocapsular or arthroscop*) adj arthroplasty).tw. | 430 |
| 4 not 5 | 1,671 |
| exp animals/ not humans.sh. | 4,787,332 |
| 6 not 7 | 1,653 |
| limit 8 to yr="1990 -Current" | 1,480 |

trials (RCTs) are viewed as the gold-standard methods for assessing interventions. Undertaking evaluation of the effectiveness of elbow arthroplasties is difficult, as these operations are rarely undertaken. The National Joint Registry Annual Report for England, Wales and Northern Ireland reported a total of 834 elbow arthroplasties in 2019 (prior to the COVID-19 outbreak) including 455 undertaken for acute trauma and 379 for elective indications in 172 units and by 249 surgeons.[2] These low numbers mean that innovative approaches to the design and delivery of trials are required to ensure the highest possible level of evidence.

In order to inform the development of future research into the evaluation of the outcome after elbow arthroplasty, it is important to understand the research methods that have so far been used. As the most appropriate outcomes that should be used in this evaluation remain uncertain, the domains that have been assessed and the instruments used to measure outcomes need to be determined, in order to support the development of a core set of outcomes, as outlined in the Core Outcome Measures in Effectiveness Trials (COMET) initiative handbook.[3] It is also important to identify the traditional sources of funding for research involving elbow arthroplasty in order to understand the limitations and opportunities which are available. This is best undertaken by a scoping review of the sources of evidence to establish how research is undertaken in this field.[4]

A preliminary search of PubMed conducted on 28 December 2020, for previous systematic or scoping reviews on elbow arthroplasty or associated forms of arthroplasty, identified eight relevant articles. One systematic review had been retracted, leaving one scoping review and six systematic reviews.[5–11] Two studied the clinical outcomes of primary elbow arthroplasty, two compared the clinical outcome of radial head arthroplasty with osteosynthesis, one analyzed trends in the indications for total elbow arthroplasty, and one was a review of revision of infected primary elbow arthroplasties. The scoping review also involved the diagnosis and management of infected elbow arthroplasty.[5] An additional search of Google Scholar conducted on 29 December 2020 identified a further seven systematic reviews; six reviewed the outcome of total elbow arthroplasty and one reviewed the causes of failure of elbow arthroplasty.[12–18] No

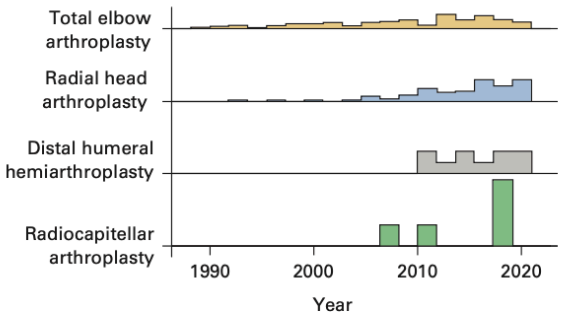



Fig. 2

Histogram of the number of publications per year categorized by type of procedure.

systematic or scoping reviews have been identified that outline the research methods, outcome domains and instruments, and sources of funding used in research in this area.

We therefore undertook a scoping review to identify and outline the research methods, domains, and outcome instruments used in research into the clinical outcomes of elbow arthroplasty to inform future work in this area and to describe the sources of funding which were used. This review addressed the following questions: what methods have been used to study the clinical effectiveness of elbow arthroplasty?; what outcome domains have been assessed, and which instruments have been used to evaluate the clinical outcomes?; and what sources of funding have been identified in these studies?

## Methods

The protocol was developed in accordance with the guidelines of the Joanna Briggs Institute (JBI)[19] and is available online at Open Science Framework.[20]

Studies of adults undergoing primary elbow arthroplasty with a diagnosis of inflammatory arthropathy, OA, acute trauma, and post-traumatic sequelae were eligible for inclusion. Studies involving arthroplasty undertaken for tumour or other rare indications were excluded, as were in vitro studies and studies of surgical approaches, biomechanics, health economics, and revision arthroplasty. Studies of groups of patients with heterogeneous diagnoses were included as long as at least 90% of them had one eligible diagnosis.

The review was to include all relevant studies of primary elbow arthroplasty published between 1 January 1990 and 7 February 2021, and any trials involving international registries that met the inclusion criteria. The types of evidence included were those from randomized trials, non-RCTs, prospective and retrospective cohort studies, case-control studies, analytical cross-sectional studies, and case series of ten patients or more, published in the English language. Review articles, surveys, case reports, and conference abstracts were excluded.

An initial limited search of Medline was undertaken in order to analyze the words used in titles and abstracts and index terms, to inform the search strategy. A full search was conducted, with support from an information specialist, on 7 February 2021 of Medline, Embase, and CENTRAL using the terms "Elbow Prosthesis", "Arthroplasty, Elbow", "Radial head arthroplasty", "Hemiarthroplasty", and "Radiocapitellar arthroplasty". This was adapted for the other databases. A search was conducted of the reference lists of studies selected for inclusion to identify any additional studies. The reference lists of previous reviews in elbow arthroplasty were also searched. The International Standard Randomised Controlled Trial Number Registry and Clinicaltrials.gov websites were reviewed to identify ongoing trials meeting the inclusion criteria. No search of grey literature was undertaken. The full strategy for Medline is shown in Table I.

Endnote software v. X9 (Clarivate Analytics, USA) was used for management of the search results. Duplicates were excluded before initial screening based on title and abstract. This was undertaken by two reviewers (ACW, ZH), with independent selection of evidence based on the inclusion criteria. The full articles of potentially relevant studies were obtained and screened by the same reviewers to identify those which were eligible. Where there was disagreement, the reviewers reviewed the manuscripts together to reach a consensus.

Pilot testing of a customized excel data extraction tool was undertaken using 25 studies selected at random and screened by ACW and ZH. A meeting was held to review the results of the screening to determine whether any changes needed to be made, with 75% agreement required before the extraction tool was accepted. Full-text screening of the remaining studies was conducted by ACW including citation details (author/s, date, title, journal, volume, issue, pages), the country in which the research was undertaken, further details of the methodology (RCT, non-RCTs, prospective and retrospective cohort studies, case-control studies, analytical cross-sectional studies, and case series of > ten patients), implant studied, patients (diagnosis, age, and sex), setting, sample size, length of follow-up (minimum, maximum, mean/median), outcome assessed, instruments used to assess outcome, and the sources of funding.

Data were tabulated and key characteristics of each study were described. The planned analyses were the number of types of study reported, length of follow-up, and domains and outcome instruments used. Where possible, findings were stratified by diagnosis (inflammatory arthropathy, OA, post-traumatic sequelae, and acute trauma) and type of arthroplasty (total elbow arthroplasty, radial head arthroplasty, distal humeral hemiarthroplasty, and radiocapitellar arthroplasty). Outcomes were recorded verbatim from the source and categorized into domains after extractions using the taxonomy described by Dodd et al.[21] For ease of reporting, the most common outcomes for each domain were rationalized into a common term. For example, in the adverse events domain the outcomes "pain", "residual pain", "proximal forearm pain", "severe pain", and "postoperative pain" were recorded as postoperative pain. The instruments were categorized according to the outcome that was assessed. No assessment of the quality of the studies or reporting was undertaken, in keeping with guidelines for scoping reviews, which do not inform clinical decision-making.[22] A full list of the studies included and outcomes is available from the corresponding author.

**Table II.** The ten countries with the highest number of recorded publications on elbow arthroplasty.

| Country | n (%) |
|---|---|
| USA | 91 (25) |
| UK | 52 (14) |
| France | 25 (7) |
| Japan | 19 (5) |
| Italy | 18 (5) |
| Netherlands | 18 (5) |
| Canada | 16 (4) |
| Germany | 16 (4) |
| Sweden | 12 (3) |
| Finland | 11 (3) |

## Results

The findings of the literature search are shown in a flowchart adhering to the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for scoping reviews (PRISMA-ScR) statement and PRISMA-S extension (Figure 1).[23,24] From a total of 2,197 de-duplicated titles identified from the searches, 402 full-text articles were reviewed, of which 40 were excluded for the reasons stated in Figure 1, leaving 362 studies published between 1 January 1990 and 7 February 2021, which were included in the review.

The subject of the study was total elbow arthroplasty in 246 (68%), radial head arthroplasty in 100 (28%), distal humeral hemiarthroplasty in 11 (3%), and radiocapitellar arthroplasty in five (1%). There was an increase in the number of publications per annum during this time, although there has been a decrease in the annual number of publications on total elbow arthroplasty since 2015 (Figure 2). Studies were reported from 34 countries, with one international collaboration. The top ten locations for published English-language elbow arthroplasty studies by country are shown in Table II.

Most studies were conducted in a hospital setting (329; 91%) with few community-based studies (32; 9%). Most were retrospective (326; 91%) and most were observational (321; 89%). There were 23 administrative database studies, 21 of which were conducted in the USA, and eight national registry studies from Australia, Denmark (two studies), Norway, Finland, Sweden, and New Zealand. There were six prospective RCTs,[25–30] one retrospective review of patients from a previous RCT,[31] and three RCT protocols.[32–34] The authors of two RCTs compared total elbow arthroplasty with osteosynthesis of distal humeral fractures, one reporting results at two years and one retrospective review of the same cohort at a mean follow-up of 12.5 years.[26,31] One study reported the outcome of total elbow arthroplasty with two different ulnar components.[29] Three studies compared the outcome of radial head arthroplasty with osteosynthesis for an acute radial head fracture, and one compared radial head arthroplasty to excision of the radial head.[25,27,28,30] Two protocols compared total elbow arthroplasty with distal humeral hemiarthroplasty for distal humeral fracture, and one protocol compared distal humeral hemiarthroplasty with osteosynthesis.[32–34] None of the RCTs compared elbow arthroplasty with non-operative management. Only one RCT had a stated source of funding which was from a commercial source.[26] In five RCTs, it was stated that there was no funding or conflict of interest for the trial,[25,28,29,31,33] and this was not reported in four studies.[27,30,32,34] The median sample size by study design is shown in Table III.

**Table III.** Sample size by the type of study.

| Type of study | Mean (range) | Median (IQR) |
|---|---|---|
| Observational studies | 67 (10 to 1,441) | 32 (20 to 57) |
| Randomized trial | 38 (20 to 60) | 40 (37 to 44) |
| Administrative database | 7,819 (176 to 56,379) | 1,625 (526 to 5,400) |
| Registry study | 692 (126 to 1,457) | 584 (386 to 934) |

IQR, interquartile range.

The median sample size by type of procedure was 41 (interquartile range (IQR) 21 to 99) for total elbow arthroplasty, 32 (IQR 22 to 56) for radial head arthroplasty, 17 (IQR 13 to 35) for distal humeral hemiarthroplasty, and 20 (IQR 19 to 30) for radiocapitellar arthroplasty. The indication for surgery by type of procedure is shown in Table IV. Differences were found between the types of procedure and the mean percentage of female patients and the mean age of the patients (Table V).

The median of the mean length of follow-up for all types of study was 56 months (IQR 36 to 80). It was longest for registry studies with a median follow-up of 96 months (IQR 81 to 110). The median follow-up in observational case series, which was the largest group, was 57 months (IQR 36 to 80). For randomized trials the median of the mean follow-up was 29 months (IQR 24 to 36). The median of the mean follow-up for studies of total elbow arthroplasty was 60 months (IQR 42 to 89), for radial head arthroplasty 42 months (IQR 25 to 68), for distal humeral hemiarthroplasty 35 months (IQR 21 to 68), and for radiocapitellar arthroplasty 59 months (IQR 32 to 87).

A total of 583 unique outcome descriptors were used in the 362 studies, and were categorized into 18 domains (Table VI). Many reported the same outcome using different terms, and in 76 cases the instrument was reported without a description of what outcome was being assessed. The largest group of outcomes were categorized as adverse events (311; 53%) despite the fact that complications were not a pre-specified outcome in most studies. The physical function domain contained the second largest number of descriptors (93; 16%). These could be grouped into four outcomes: function/disability, range of motion, strength, and activities of daily living. Strength was often poorly defined but most commonly described an assessment of the strength of extension of the elbow as a measure of triceps function. The radiological appearance of the elbow was assessed by 19 separate outcomes including implant alignment, congruency, fixation, success, lengthening, head size, stem size, stem positioning, implant positioning, prosthetic sizing, quality of cementing, bone graft incorporation, cortical fit, cement mantle, valgus tilting, congruence of the proximal radioulnar joint, cement technique, and joint congruity.

A total of 105 instruments were used to measure 39 outcomes, of which 26 were clinical and 13 were radiological. The mean number of instruments used per outcome when an instrument was described was five (1 to 26). The list of instruments used to assess each outcome is shown in Table VII. For implant survival, the instruments were methods of analysis, but are included for completeness. Some instruments are included in more than one category of outcome and may be listed for

**Table IV.** Primary diagnosis at the time of surgery by type of procedure.

| Type of procedure | RA | OA | Post-trauma | Acute trauma | Other |
|---|---|---|---|---|---|
| Total elbow arthroplasty | 61% | 5% | 16% | 18% | 2% |
| Radial head arthroplasty | < 1% | < 1% | 12% | 87% | < 1% |
| Distal humeral hemiarthroplasty | 0% | 0% | 15% | 85% | 0% |
| Radiocapitellar arthroplasty | 11% | 58% | 27% | 1% | 3% |

OA, osteoarthritis; RA, rheumatoid arthritis.

**Table V.** Sex and age of patients included in studies by type of procedure.

| Type of procedure | Mean % female | Mean of minimum age, yrs (range) | Mean of maximum age, yrs (range) | Mean of mean age, yrs (range) |
|---|---|---|---|---|
| Total elbow arthroplasty | 76 | 36 (5 to 75) | 81 (40 to 97) | 61 (28 to 85) |
| Radial head arthroplasty | 46 | 23 (14 to 62) | 75 (50 to 93) | 49 (31 to 67) |
| Distal humeral hemiarthroplasty | 85 | 47 (16 to 62) | 81 (63 to 90) | 67 (45 to 79) |
| Radiocapitellar arthroplasty | 51 | 31 (25 to 40) | 74 (69 to 82) | 55 (53 to 61) |

**Table VI.** Outcomes categorized by domain.

| Domain (n) | Outcome |
|---|---|
| Adverse events (311) | Postoperative pain, implant-related, radiological complications, wound problems, infection, stiffness, triceps weakness, bone problems, neurological, transfusion, effusion/synovitis, medical complications |
| Physical function (93) | Function/disability, strength, range of motion, activities of daily living |
| Musculoskeletal/connective tissue (37) | Radiological appearance, elbow stability |
| Need for further intervention (34) | Reoperation, implant revision/survival |
| Nervous system (30) | Pain, neurological status |
| Delivery of care (15) | Satisfaction |
| Social function (13) | Sport participation, social-psychological status |
| Hospital resources (11) | Length of stay, return to operating room, readmission, length of surgery |
| Role function (8) | Return to work |
| Perceived health (7) | General health |
| Economic resource (7) | Hospital costs, non-routine discharge costs |
| Mortality/survival (6) | Mortality |
| Psychiatric (3) | Mental health |
| Quality of life (2) | Quality of life |
| Emotional wellbeing (2) | Wellbeing, role emotional |
| Cognitive function (1) | Cognitive status |
| Personal circumstances (1) | Patient autonomy |
| Societal carer burden (1) | Non-homebound discharge |

outcomes and domains that seem inappropriate, but these were taken verbatim from the studies.

## Discussion

The number of RCTs undertaken in orthopaedic surgery is consistently low. A systematic review of the orthopaedic literature found that only 20% of operations had at least one low risk of bias RCT supporting its use when compared with non-operative treatment.[35] We did not identify any for elbow arthroplasty. Many different reasons have been proposed for this; the cause is likely to be multifactorial.[36,37] One of the challenges of performing a RCT in some areas of orthopaedics in which the condition is rare is the feasibility of recruiting sufficient patients to address the research question.

The European Union defines rare diseases as those with a prevalence of < 50 per 100,000 population for the purposes of orphan drugs (a pharmaceutical product developed to treat a rare condition, which would not be profitable without additional support), where normal economic models prevent research and development.[38] Elbow arthroplasty is a rare procedure with an annual incidence estimated from Scottish data of 1.4 per 100,000 of the population, and the same economic and scientific challenges apply, with low investment in research and barriers to conventional statistical frequentist approaches to the investigation of clinical outcomes due to the feasibility of RCTs.[39]

We found that the literature on elbow arthroplasty consists largely of unfunded retrospective observational studies with small sample sizes. It is interesting to note that the number of studies on total elbow arthroplasty has declined in recent years. The cause of this is unknown, but it may be due to a decreasing use of this operation due to the improved medical management of inflammatory joint disease. Only nine prospective RCTs were identified, three of which were protocols of ongoing trials rather than reports of results. The sample size of these RCTs ranged from 20 to a maximum of 60 patients. Only one randomized trial had a stated source of funding, and that was from a commercial body. We were not able to identify any RCTs comparing arthroplasty with non-operative treatment. It is beyond the scope of this review to determine the causes for the poor quality of the evidence, but there remain many questions

**Table VII.** Instruments used to assess outcome in the studies (no assessment of psychometric properties was conducted).

| Outcome | Instrument |
|---|---|
| Activities of daily living | Liverpool Elbow Score, Mayo Elbow Performance Index, Mayo Elbow Performance Score, UCLA Activity Score, Wrightington Score |
| Cognitive status | MMSE |
| Complications | Voloshin |
| Cost utility analysis | QALY |
| Deformity | Ewald scoring system, JOA score |
| Function | Andrews score, ASES score, Broberg Morrey score, DASH score, Elbex score, Elbow functional assessment, Ewald score, Inglis score, JOA score, Khatri score, Likert scale, Liverpool Elbow Score, Mayo Elbow Performance Index, Mayo Elbow Performance Score, numerical rating scale, Oxford score, PREE, QuickDASH, SANE, SECEC, Souter, Stanford health assessment questionnaire, VAS |
| Global health | EQ-5D |
| Grip strength | ASES, Broberg Morrey, Dynamometer, NK hand evaluation system |
| Implant survival | Dobbs, Kaplan-Meier, Life table, Murray |
| Joint position sense | Propriometer |
| Locomotion | Linkoping |
| Mental health | SF-12 Mental |
| Pain | ASES, Broberg Morrey, DASH, Elbow functional assessment, Ewald score, Inglis score, JOA score, Khatri score, Likert scale, Mayo Elbow Performance Score, Modified Andrews score, numerical rating scale, Oxford Elbow Score, Liverpool Elbow Score, Pain intensity scale, PREE, VAS |
| Patient autonomy | Katz |
| Physical health | SF-12 physical |
| Quality of life | RAND-36, SF-36 |
| Radial nerve palsy | Hirachi, Electrophysiology |
| Range of motion | ASES, Broberg Morrey, Cassebaum, Elbow functional assessment, Ewald score, HSS score, JOA score, Inglis score, Khatri score, Likert scale, Liverpool Elbow Score, Mayo Elbow Performance Index, Mayo Elbow Performance Score, Modified Andrews, NK hand evaluation system, Propriometer |
| Return to work | Binary |
| Satisfaction | ASES, binary, Jungbluth, Likert scale, numerical rating scale, SANE, VAS |
| Sport participation | Allain, Liverpool Elbow Score |
| Stability | ASES, Broberg Morrey, Elbow functional assessment, JOA score, Likert scale, Mayo Elbow Performance Index, Mayo Elbow Performance Score, Modified Andrews score |
| Strength | ASES, Broberg Morrey, HSS score, Isobex, Kindeyn Dynamometer, Lido workset, Mayo Elbow Performance Index, Mayo Elbow Performance Score, MRC scale, Likert scale, Liverpool Elbow Score |
| Success of treatment | VAS |
| Tenderness | ASES |
| Ulnar nerve function | Electrophysiology, Liverpool Elbow Score, McGowan grade |
| Radiological alignment | Figgie, H index, O'Driscoll, RSA, Storen's line, U index, Wrightington |
| Radiological bushing wear | Gill, Llamas, Lee, Mayo, Ramsey, Schneeberger |
| Radiological capitellum erosion | Broberg Morrey, Likert scale, Llamas |
| Radiological heterotopic ossification | Binary, Brooker, Hastings and Graham, Ilahi, Likert scale |
| Radiological loosening | Berschback, binary, Cil, Gill and Morrey, Goldberg, Grewal, Harris, King, Likert scale, Madsen, Mayo, Morrey and Adams, Popovic, Schneeberger, Wagener, Wrightington |
| Radiological lucency | Binary, Fehringer, Gruen, Kodde, Kudo and Iwano, Morrey, Souter, Tanaka, Wrightington |
| Radiological medial collateral ligament healing | Ultrasound |
| Radiological osteoarthritis | Binary, Broberg Morrey, Kellgren Lawrence, Knirk and Jupiter, Lindenhovius, Likert scale |
| Radiological overstuffing | Rowland, Van Riet |
| Radiological quality cement technique | Schneeberger |
| Radiological radial head prominence | Doornberg |
| Radiological synovitis | Forster |
| Radiological ulna wear | Smith and Hughes |

ASES, American Shoulder and Elbow Score; DASH, Disabilities of the Arm, Shoulder and Hand questionnaire; EQ-5D, EuroQol five-dimension index; HSS, Hospital for Special Surgery; JOA, Japanese Orthopaedic Association; MMSE, Mini-Mental State Evaluation; PREE, Patient Rated Elbow Evaluation; QALY, quality-adjusted life year; RAND-36, early version of SF-36; RSA, radiostereometric analysis; SANE, Single Assessment Numeric Evaluation; SECEC, European Society for Surgery of the Shoulder and Elbow; SF-12, 12-Item Short-Form Health Survey questionnaire; SF-36, 36-Item Short-Form Health Survey questionnaire; VAS, visual analogue score.

about elbow arthroplasty that require robust unbiased scientific investigation. Researchers designing future RCTs in this area will need to explore solutions for the investigation of rare conditions using multicentre and possibly international collaborations to ensure sufficient statistical power, and the exploration of alternative trial designs. The moral hazard of enrolment of patients into underpowered trials can be avoided by using a meta-analysis of duplicate studies.[40] However, this would require the establishment of a registry of trials involving rare musculoskeletal interventions and a mechanism to ensure data sharing and individual patient data meta-analysis.[41,42] Alternative trial designs have been explored for the investigation of rare diseases including crossover designs, N-of-1 trials, and adaptive designs, but most are not suitable for surgical research due to the irreversible nature of operations, and the time period required to determine the outcome of surgery.[43,44] Bayesian analytical methods, however, could be used to exploit all available data to strengthen the findings from smaller RCTs.[45] Appropriate funding would be required to ensure the successful delivery of these more complicated investigations, and qualitative research may be needed to understand the barriers and facilitators within the elbow arthroplasty community that may be affecting research practice.

Alternative approaches to tackling the paucity of information may mean pooling data from several sources. The combination of data for meta-analysis is hindered, however, by the diversity of outcome domains and instruments used to measure outcomes. The 583 individual descriptors identified in this review of clinical outcomes of elbow arthroplasty can be rationalized to 41 outcomes over 18 domains using the taxonomy adopted by the COMET initiative. However, this is still a large number of domains, and it is unclear which of these might be considered most important by patients, carers, and clinicians.[21] An average of five instruments have been used for the 39 outcomes when an instrument has been described. We have not sought to analyze the psychometric properties of the instruments which have been used, but rather to identify the domains, outcomes, and instruments without an assessment of quality or validity. Once core outcomes have been defined, it will be necessary to undertake an assessment of relevant instruments to determine if any meet the criteria of truth, discrimination, and feasibility.[46]

In summary, this review has highlighted a clear need to define core outcome domains for research involving elbow arthroplasty that can lead to the development of outcome instruments to be used in further work in this area.

### Take home message

- Decision-making in elbow arthroplasty is based on retrospective observational studies with no consistency in the outcomes assessed, which makes comparison challenging.
- This study has highlighted the need to define a core outcome set for elbow arthroplasty research, and to explore alternative research methods.

### Twitter

Follow A. C. Watts @myelbowdoc
Follow Z. Hamoodi @Zaidahamoodi
Follow C. McDaid @Catriona_McDaid



## References


1. **Dee R**. Elbow replacement with the R. Dee prosthesis. *Acta Orthop Belg*. 1975;41(4):477–483.
2. **Ben-Shlomo Y**, **Blom A**, **Boulton C**, **et al**. The National Joint Registry 18th Annual Report 2021. National Joint Registry. 2021. http://www.ncbi.nlm.nih.gov/books/NBK576858/ (date last accessed 22 July 2022).
3. **Williamson PR**, **Altman DG**, **Bagley H**, **et al**. The COMET Handbook: version 1.0. *Trials*. 2017;18(Suppl 3):280.
4. **Munn Z**, **Peters MDJ**, **Stern C**, **Tufanaru C**, **McArthur A**, **Aromataris E**. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol*. 2018;18(1):143.
5. **Watts AC**, **Duckworth AD**, **Trail IA**, **Rees J**, **Thomas M**, **Rangan A**. Scoping review: Diagnosis and management of periprosthetic joint infection in elbow arthroplasty. *Shoulder Elbow*. 2019;11(4):282–291.
6. **Wang J-H**, **Ma H-H**, **Chou T-FA**, **et al**. Outcomes following total elbow arthroplasty for rheumatoid arthritis versus post-traumatic conditions: a systematic review and meta-analysis. *Bone Joint J*. 2019;101-B(12):1489–1497.
7. **Kholinne E**, **Altamimi LA**, **Aldayel A**, **et al**. Primary linked total elbow arthroplasty for acute distal humerus fracture management: a systematic review of clinical outcome. *Clin Orthop Surg*. 2020;12(4):503–513.
8. **Macken AA**, **Prkic A**, **Kodde IF**, **Lans J**, **Chen NC**, **Eygendaal D**. Global trends in indications for total elbow arthroplasty: a systematic review of national registries. *EFORT Open Rev*. 2020;5(4):215–220.
9. **Chen H**, **Shao Y**, **Li S**. Replacement or repair of terrible triad of the elbow: A systematic review and meta-analysis. *Medicine (Baltimore)*. 2019;98(6):e13054.
10. **Kunutsor SK**, **Beswick AD**, **Whitehouse MR**, **Blom AW**. One- and two-stage surgical revision of infected elbow prostheses following total joint replacement: a systematic review. *BMC Musculoskelet Disord*. 2019;20(1):467.
11. **Kyriacou S**, **Gupta Y**, **Bains HK**, **Singh HP**. Radial head replacement versus reconstruction for the treatment of the terrible triad injury of the elbow: a systematic review and meta-analysis. *Arch Orthop Trauma Surg*. 2019;139(4):507–517.
12. **Samdanis V**, **Manoharan G**, **Jordan RW**, **et al**. Indications and outcome in total elbow arthroplasty: A systematic review. *Shoulder Elbow*. 2020;12(5):353–361.
13. **van der Lugt JCT**, **Rozing PM**. Systematic review of primary total elbow prostheses used for the rheumatoid elbow. *Clin Rheumatol*. 2004;23(4):291–298.
14. **Chou TFA**, **Ma HH**, **Wang JH**, **et al**. Total elbow arthroplasty in patients with rheumatoid arthritis: a systematic review and meta-analysis. *Bone Joint J*. 2020;102-B(8):967–980.
15. **Githens M**, **Yao J**, **Sox AHS**, **Bishop J**. Open reduction and internal fixation versus total elbow arthroplasty for the treatment of geriatric distal humerus fractures: a systematic review and meta-analysis. *J Orthop Trauma*. 2014;28(8):481–488.
16. **Prkic A**, **Welsink C**, **The B**, **van den Bekerom MPJ**, **Eygendaal D**. Why does total elbow arthroplasty fail today? A systematic review of recent literature. *Arch Orthop Trauma Surg*. 2017;137(6):761–769.
17. **Little CP**, **Graham AJ**, **Carr AJ**. Total elbow arthroplasty: a systematic review of the literature in the English language until the end of 2003. *J Bone Joint Surg Br*. 2005;87-B(4):437–444.
18. **Welsink CL**, **Lambers KTA**, **van Deurzen DFP**, **Eygendaal D**, **van den Bekerom MPJ**. Total elbow arthroplasty: A systematic review. *JBJS Rev*. 2017;5(7):e4.
19. **Aromataris E**, **Munn Z**. Chapter 11: Scoping Reviews. In: Aromataris E, Munn Z, eds. *JBI Manual for Evidence Synthesis*. JBI, 2020. https://jbi-global-wiki.refined.site/space/MANUAL/4687342/Chapter+11%3A+Scoping+reviews (date last accessed 8 August 2022).
20. **No authors listed**. Elbow arthroplasty research methods and outcome domains and instruments in clinical outcome studies: a scoping review. Open Science Framework. https://osf.io/t6qyh/ (date last accessed 25 July 2022).
21. **Dodd S**, **Clarke M**, **Becker L**, **Mavergames C**, **Fish R**, **Williamson PR**. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. *J Clin Epidemiol*. 2018;96:84–92.
22. **Peters MDJ**, **Marnie C**, **Tricco AC**, **et al**. Updated methodological guidance for the conduct of scoping reviews. *JBI Evid Synth*. 2020;18(10):2119–2126.
23. **Tricco AC**, **Lillie E**, **Zarin W**, **et al**. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467–473.
24. **Rethlefsen ML**, **Kirtley S**, **Waffenschmidt S**, **et al**. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Syst Rev*. 2021;10(1):39.
25. **Chen X**, **Wang S**, **Cao L**, **Yang G**, **Li M**, **Su J**. Comparison between radial head replacement and open reduction and internal fixation in clinical treatment of unstable, multi-fragmented radial head fractures. *Int Orthop*. 2011;35(7):1071–1076.

26. **McKee MD**, **Veillette CJH**, **Hall JA**, **et al**. A multicenter, prospective, randomized, controlled trial of open reduction--internal fixation versus total elbow arthroplasty for displaced intra-articular distal humeral fractures in elderly patients. *J Shoulder Elbow Surg*. 2009;18(1):3–12.
27. **Ruan HJ**, **Fan CY**, **Liu JJ**, **Zeng BF**. A comparative study of internal fixation and prosthesis replacement for radial head fractures of Mason type III. *Int Orthop*. 2009;33(1):249–253.
28. **Singh AK**, **Jidge A**, **Ramteke U**, **et al**. Functional outcome of elbow kinematics in radial head excision versus radial head replacement: A comparative study. *Open Access Maced J Med Sci*. 2019;7(9):1505–1508.
29. **Tanaka N**, **Sakahashi H**, **Ishii S**, **Kudo H**. Comparison of two types of ulnar component in type-5 Kudo total elbow arthroplasty in patients with rheumatoid arthritis: a long-term follow-up. *J Bone Joint Surg Br*. 2006;88-B(3):341–344.
30. **Yan M**, **Ni J**, **Song D**, **Ding M**, **Liu T**, **Huang J**. Radial head replacement or repair for the terrible triad of the elbow: which procedure is better? *ANZ J Surg*. 2015;85(9):644–648.
31. **Dehghan N**, **Furey M**, **Schemitsch L**, **et al**. Long-term outcomes of total elbow arthroplasty for distal humeral fracture: results from a prior randomized clinical trial. *J Shoulder Elbow Surg*. 2019;28(11):2198–2204.
32. **Adolfsson L**. Elbow Hemiarthroplasty Versus Total Elbow Arthroplasty for Irreparable Distal Humeral Fractures. ClinicalTrials.gov. 2018. https://ClinicalTrials.gov/show/NCT03596736 (date last accessed 22 July 2022).
33. **Al-Hamdani A**, **Rasmussen JV**, **Holtz K**, **Olsen BS**. Elbow hemiarthroplasty versus open reduction and internal fixation for AO/OTA type 13 C2 and C3 fractures of distal humerus in patients aged 50 years or above: a randomized controlled trial. *Trials*. 2020;21(1):497.
34. **Smith C**. Hemiarthroplasty Or Total Elbow Arthroplasty in the Elderly. ClinicalTrials.gov. 2020. https://clinicaltrials.gov/show/NCT04646798 (date last accessed 22 July 2022).
35. **Lim HC**, **Adie S**, **Naylor JM**, **Harris IA**. Randomised trial support for orthopaedic surgical procedures. *PLoS One*. 2014;9(6):e96745.
36. **Solomon MJ**, **McLeod RS**. Should we be performing more randomized controlled trials evaluating surgical operations? *Surgery*. 1995;118(3):459–467.
37. **McCulloch P**, **Taylor I**, **Sasako M**, **Lovett B**, **Griffin D**. Randomised trials in surgery: problems and possible solutions. *BMJ*. 2002;324(7351):1448–1451.
38. **No authors listed**. Orphan Drug Act (1983). U.S. Government Publishing Office. 1983. https://www.govinfo.gov/content/pkg/STATUTE-96/pdf/STATUTE-96-Pg2049.pdf (date last accessed 22 July 2022).
39. **Jenkins PJ**, **Watts AC**, **Norwood T**, **Duckworth AD**, **Rymaszewski LA**, **McEachan JE**. Total elbow replacement: outcome of 1,146 arthroplasties from the Scottish Arthroplasty Project. *Acta Orthop*. 2013;84(2):119–123.
40. **Halpern SD**, **Karlawish JHT**, **Berlin JA**. The continuing unethical conduct of underpowered clinical trials. *JAMA*. 2002;288(3):358–362.
41. **Medley N**, **Cuthbert A**, **Crew R**, **Stewart L**, **Smith CT**, **Alfirevic Z**. Developing a topic-based repository of clinical trial individual patient data: experiences and lessons learned from a pilot project. *Syst Rev*. 2021;10(1):162.
42. **Rydzewska LHM**, **Stewart LA**, **Tierney JF**. Sharing individual participant data: through a systematic reviewer lens. *Trials*. 2022;23(1):167.
43. **Cornu C**, **Kassai B**, **Fisch R**, **et al**. Experimental designs for small randomised clinical trials: an algorithm for choice. *Orphanet J Rare Dis*. 2013;8(1):48.
44. **Gupta S**, **Faughnan ME**, **Tomlinson GA**, **Bayoumi AM**. A framework for applying unfamiliar trial designs in studies of rare diseases. *J Clin Epidemiol*. 2011;64(10):1085–1094.
45. **Tan S-B**, **Dear KBG**, **Bruzzi P**, **Machin D**. Strategy for randomised clinical trials in rare cancers. *BMJ*. 2003;327(7405):47–49.
46. **Boers M**, **Kirwan JR**, **Wells G**, **et al**. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *J Clin Epidemiol*. 2014;67(7):745–753.



**Author information:**
A. C. Watts, BSc, FRCS (T&O), Orthopaedic and Trauma Elbow Surgeon, Wrightington, Wigan and Leigh Teaching Hospitals NHS Foundation Trust, Wigan, UK; Department of Health Sciences, University of York, York, UK.

Z. Hamoodi, MRCS, Orthopaedics and Trauma Specialty Trainee, Wrightington, Wigan and Leigh Teaching Hospitals NHS Foundation Trust, Wigan, UK.

C. McDaid, PhD, Reader in Trials, York Trials Unit
C. Hewitt, PhD, Deputy Director, York Trials Unit
Department of Health Sciences, University of York, York, UK.



**Author contributions:**
A. C. Watts: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft.
Z. Hamoodi: Investigation.
C. McDaid: Methodology, Writing – review & editing.
C. Hewitt: Methodology, Writing – review & editing.

**Funding statement:**
The authors received no financial or material support for the research, authorship, and/or publication of this article.

**Acknowledgements:**
The authors would like to acknowledge the contribution of David Brown, Academic Liaison Team, University of York Library for support with the literature search.

**Ethical review statement:**
Ethics approval was not required for this literature review.

This article was primary edited by J. Scott.

■ SHOULDER & ELBOW

# Core Outcome Domains for Elbow Replacement (CODER)

**A. C. Watts,**
**C. McDaid,**
**C. Hewitt,**
**The CODER Consensus group**

*From University of York, York, UK*

## Aims

**A review of the literature on elbow replacement found no consistency in the clinical outcome measures which are used to assess the effectiveness of interventions. The aim of this study was to define core outcome domains for elbow replacement.**

## Methods

**A real-time Delphi survey was conducted over four weeks using outcomes from a scoping review of 362 studies on elbow replacement published between January 1990 and February 2021. A total of 583 outcome descriptors were rationalized to 139 unique outcomes. The survey consisted of 139 outcomes divided into 18 domains. The readability and clarity of the survey was determined by a advisory group including a patient representative. Participants were able to view aggregated responses from other participants in real time and to revisit their responses as many times as they wished during the study period. Participants were able to propose additional items for inclusion. A Patient and Public Inclusion and Engagement (PPIE) panel considered the consensus findings.**

## Results

**A total of 45 respondents completed the survey. Nine core mandatory domains were identified: 'return to work or normal daily role'; delivery of care was measured in the domains 'patient satisfaction with the outcome of surgery' and 'would the patient have the same operation again'; 'pain'; 'revision'; 'elbow function'; 'independence in activities of daily living'; 'health-related quality of life'; and 'adverse events'. 'Elbow range of motion' was identified as important by consensus but was felt to be less relevant by the PPIE panel. The PPIE panel unanimously stated that pain should be used as the primary outcome domain.**

## Conclusion

**This study defined core domains for the clinical outcomes of elbow replacement obtained by consensus from patients, carers and healthcare professionals. Pain may be used as the primary outcome in future studies, where appropriate. Further work is required to define the instruments that should be used.**

**Cite this article:** ***Bone Joint J*** **2024;106-B(11):xxx–xxx.**

## Introduction

When the quality of life is significantly affected by pain in the elbow or limited function, an elbow replacement has been a surgical option for more than five decades.[1] Although initially used solely for the treatment of inflammatory joint disease, the indications have expanded to include the sequelae of trauma, degenerative joint disease and reconstruction after resection of a tumour. Elbow arthroplasty can include: total elbow arthroplasty involving replacement of both sides of the ulnohumeral joint with or without replacement of the radial head, distal humeral hemiarthroplasty with replacement of the trochlea and capitellum, radial head arthroplasty in isolation or radiocapitellar arthroplasty with replacement of the radial head and capitellum. A recent review of the literature on elbow arthroplasty identified differences in the indications and demographics of patients treated with different types of arthroplasty.[2]

When considering the effect of these procedures and the possible harms, it is important

Correspondence should be sent to A. C. Watts; email: adam.watts@elbowdoc.co.uk


doi:10.1302/0301-620X.106B11.BJJ-2024-0352.R1 $2.00

*Bone Joint J* 2024;106-B(11):xxx–xxx.

**Table I.** The demographic characteristics of participants in the survey.

| Variable | n (%) |
|---|---|
| **Sex** | |
| Female | 9 (20) |
| Male | 36 (80) |
| **Age, yrs** | |
| 30 to 39 | 12 (27) |
| 40 to 49 | 19 (42) |
| 50 to 59 | 9 (20) |
| 60 to 69 | 5 (11) |
| **Role** | |
| AHP | 6 (14) |
| Carer | 2 (4) |
| GP/commissioner | 1 (2) |
| Patient | 3 (7) |
| Surgeon | 33 (73) |

AHP, allied health professional; GP, general practitioner.

to determine their place as a treatment for particular clinical conditions. However, in order to compare one intervention with another and between different clinical settings, it is important to have consistency in what is measured and to ensure that what is measured is relevant to users. In the 362 studies dealing with the clinical outcomes of elbow replacement reported or registered between 1 January 1990 and 7 February 2021, a total of 583 unique outcome descriptors were identified. In addition to the heterogeneity of what was measured, there was no consistency in how the outcomes were measured, with 105 instruments for reporting outcomes being described for 39 of the outcomes, and no clearly defined tool for the remaining outcomes.[2] This supports a previous study, whose authors identified a need for a core set of outcomes to be defined for elbow replacement.[3]

Determination of what is measured, the core set of outcomes, requires an understanding of the domains which are important to the patient, their caregivers and the clinicians with whom they engage, to help in the management of their condition. We are not aware of any studies which have been undertaken to elicit a consensus for the core outcome domains from key stakeholders for elbow replacement. A search of the COMET database conducted on 19 June 2022 identified no other studies on this topic.

Thus, the aim of this study was to reach a consensus on the core outcome domains using a real-time Delphi (RT-Delphi) process with internal pilot. The protocol was developed with consideration of the Core Outcome Set-STAndards for Development (COS-STAD) and Core Outcome Set-STAndards Protocol Items (COS-STAP) consensus statements for standards of the design of a survey of the core set of outcomes.[4,5] We did not attempt to determine the best instruments to measure outcome but sought to identify the core domains for the clinical outcome of elbow replacement from key stakeholders, including patients who have undergone elbow replacement, their caregivers and healthcare professionals engaged in providing guidance on treatment. We also sought to determine the best times to record the measurements of outcome after elbow replacement.

## Methods

A RT-Delphi consensus survey was conducted online using Surveylet software (Calibrum, USA). This tool provides anonymous completion of surveys, with participants able to score outcome domains, view the aggregate scores of previous respondents and provide text to explain why they may take a different view to the aggregate response in order to inform other respondents. All respondents can return to the survey as many times as they like while the survey is open, to review or amend responses. This approach replaces several rounds of survey used in Delphi studies, which can be associated with attrition and respondent fatigue.[6]

The survey investigated opinions for the core outcome domains following elbow replacement for the treatment of inflammatory joint disease, degenerative joint disease, acute trauma and the sequelae of trauma. Replacement for reconstruction following the resection of a tumour was not included, as the priorities may be different due to the burden of disease. The core outcome set will apply to adult patients aged > 16 years undergoing elbow replacement for the conditions outlined. Operations include total elbow replacement with or without a radial head replacement, distal humeral hemiarthroplasty, radial head replacement alone, and radiocapitellar replacement, also known as lateral resurfacing elbow replacement. The survey also sought consensus on the times at which the measurements of the core domains, once defined, should be recorded.

The Core Outcome Domain for Elbow Replacement (CODER) set is being developed to be used in research to assess the effects of elbow replacements, and in monitoring individual patient's outcomes as part of implant surveillance.

The survey advisory group (SAG) consisted of the chief investigator (ACW), a patient who had undergone elbow replacement, four surgeons and two physiotherapists from different institutions and countries who had experience of treating patients who had undergone elbow replacement or who undertook research into this area. The group attended meetings online during the study.

A minimum of 30 key stakeholders, including adults aged > 18 years who have had any replacement involving the elbow, other 'expert' patients familiar with research processes, carers of patients who have undergone elbow replacement, and healthcare professionals who assist in the management of elbow problems (including surgeons, physicians, and physiotherapists) were included. Participants were drawn from countries around the world. Ethical approval was obtained to approach patients who have undergone elbow replacement in order to invite them to participate. Participants who were patients and their carers were recruited by the chief investigator through a letter of invitation and an information sheet. Patients who were unable to participate in an online survey were excluded. Surgeons and allied health professionals were recruited through research networks and social media.

Ethical approval was obtained from the Health Sciences Research Committee, University of York (HSRGC/2022/258 /G, 30/09/2022) and, for the recruitment of patients, through the integrated research approval system (IRAS ID 320243, 30/03/2023) from the Health Research Authority in the UK. Participants were invited to register their interest and provide consent for the use of their anonymized responses using the online registration tool.

The list of outcomes presented in the RT-Delphi was taken from a review of the literature about the outcomes of elbow replacement published between 1 January 1990 and 7 February 2021.[2] As indicated, this identified 583 unique outcome descriptors from 362 publications. In order to make the process more manageable for participants, two members of the research team (ACW, ZH) condensed the outcome descriptors and combined outcomes that record the same thing using different terminology under a common term. For example 'stability', 'instability', 'implant stability', and 'instability – subjective' were described as 'elbow stability'. These two investigators conducted this separately and discussed any discrepancies to reach agreement. No possible outcomes were removed in this process. Where appropriate, outcome domains were grouped into categories to reduce the burden for the participants. For example, mortality was used as a category of outcome, with in-hospital mortality, 30-day mortality and 90-day mortality as separate domains within the category. The clinical members of the SAG were asked to consider the readability and clarity of each outcome to determine if any additional text was required to explain it to participants. The patient member of the SAG was then asked to review the description of each outcome to ensure that the terms were appropriate and understandable to a lay reader. The members of the SAG piloted the online RT-Delphi tool prior to publication of the survey.

Participants in the CODER survey registered their interest and gave electronic consent for their responses to be used, using a Microsoft Forms questionnaire in a host institution Microsoft Teams account (Microsoft, USA). A personalized and anonymized link was sent to consenting participants from a host email account. Participants were able to return to the RT-Delphi at any time and as many times as they wished during the four-week period that the study was open, to review and amend their responses. A reminder email was sent after two weeks to encourage engagement, and for those who had participated to review their responses.

Participants were asked to record their sex, age in ten-year age bands, and role (patient, carer, general practitioner, commissioner, allied health professional or surgeon).

In keeping with the OMERACT handbook version 2.1, participants were asked to rate each outcome on a scale from 1 to 9 where 9 is the most important and 1 the least important in the participant's view.[7] Participants were provided with a dashboard of the responses for that outcome from previous participants, using the responses from the SAG as the baseline response at the start of the survey, and the software updated the dashboard in real time. When the researchers had grouped outcomes into categories, as in the mortality example, the respondents were asked to rate the outcome category and their response determined whether they were asked to rate the domains within that category. Participants scoring between 7 and 9 on any of the categories of outcome were asked to rate each domain falling within that category using the same scale. If the participant scored < 7 for any category, they were not asked to rate the domains within that category and were automatically taken to the next domain question. This was intended to reduce the burden for participants. Participants scoring > three points above or below the aggregate score for each outcome were given the option of recording the reason for their response, and their anonymized response was made available for other participants to view. All participants were given the opportunity to suggest additional outcomes not already listed if they felt that they were important to the assessment of the outcome of elbow replacement. These would have been reviewed by the SAG if it was felt that they were not covered by other outcomes already included, but no additional outcomes were suggested by participants.

All participants were invited to attend an online meeting to review the outcomes that passed the threshold for inclusion, and to reach consensus on the domains into which these could be grouped to produce a final list of included core outcome domains. The results of the survey were presented, and each outcome achieving the threshold was presented to the group one at a time for consideration. For each core outcome domain, attendees were given timing options to select (baseline, six weeks, and three, six, 12, 18 and 24 months) and the option to suggest alternative times for the collection of the outcome.

In order to ensure that the patients' voice was well represented in the consensus process, a separate face-to-face public inclusion and engagement panel (PPIE) meeting was held, following the period of survey, to review the core outcome domains and timepoints, and to invite their comment.

The CODER study was registered on the Core Outcome Measures in Effectiveness Trials (COMET) database (October 2022) and the protocol was made available on the Open Science Framework platform.[8]

**Statistical analysis.** A descriptive analysis of the responses was undertaken. All outcomes given a score of ≥ 7 by a minimum of 70% of participants were included in the final meeting. All outcomes were rationalized into domains by the consensus group and have been included in the core outcome domain set. Outcomes were categorized into domains using the taxonomy of Dodd et al.[9] For the timings of the recording of the measurements of the core outcome domains, each time chosen by a minimum of 70% of the group was included.

## Results

A total of 45 people registered to complete the survey and provided electronic consent for use of their responses. The self-reported descriptors of the respondents are given in Table I. An additional three patients and two surgeons registered interest but did not complete the survey.

The four-week RT Delphi survey identified 52 outcomes across 13 domains that reached the pre-specified threshold (≥ 70% respondents scoring 7/9 or higher) for consideration at the final consensus meeting (Supplementary Table i).

All 45 respondents were invited to an online consensus meeting on 13 July 2023, and 14 attended. These included one patient, one carer, three allied health professionals and nine surgeons. It was noted that where categories were selected as important, every question within that category reached the threshold for inclusion, with the exception of 'minor adverse event' under the category 'adverse events'. The consensus view was that the categories of outcome could be retained as core domains and outcomes within each category removed without loss of fidelity. In addition, the group felt that 'return to work'

**Table II.** Core outcome domains for elbow replacement. Note "active range of elbow movement" and "six-month" timepoint for data capture were felt to be less relevant by the public inclusion and engagement panel, and should be considered as important but not mandatory.

| Dodd classification[9] | Core domain | Baseline | 3 months | 6 months | 12 months | 24 months |
|---|---|---|---|---|---|---|
| Role function | Return to work or normal daily role | | X | X | X | |
| Delivery of care | Patient satisfaction with outcome of surgery | | | | X | X |
| | Would patient have the same operation again? | | | | X | X |
| Nervous system | Pain | X | X | X | X | X |
| Need for further intervention | Revision | | | | | X |
| Physical function | Elbow function | X | | X | X | X |
| | Independence in activities of daily living | X | | X | X | X |
| | Active range of elbow movement | X | X | | X | X |
| Quality of life | Health related quality of life | X | | X | X | X |
| Adverse events | Adverse event including death | | X | X | | |

and 'return to normal daily role' could be combined into a single domain of 'return to work or normal daily role'. The group took the view that 'patient autonomy', used in the way it had been used in the source material to assess independence with daily activities as measured by the Katz instrument, could be combined with 'activities of daily living' under a single domain of 'independence in activities of daily living'. While 'patient satisfaction with the outcome of surgery' and 'would the patient have the same operation again?' both fall within the domain of 'delivery of care' according to Dodd's taxonomy, it was agreed that both should be retained as they assess different aspects of satisfaction or acceptability. This resulted in ten core outcome domains which were taken to the PPIE panel meeting for consideration, along with the core timings for the recording of the assessment of each outcome domain (Table II).

The PPIE group attended a face-to-face meeting with the chief investigator and a facilitator. The group consisted of eight female participants and one male participant with personal experience of orthopaedic care. Two members had undergone elbow replacement, one was a carer for someone who had an elbow replacement, and one represented a disability group and was active in clinical research as a lay representative. The group discussed each core domain and the timings for the capture of core outcomes. They felt strongly that efforts should be made to reduce the burden of research for the participants. They therefore questioned the value of recording data at six months postoperatively. They discussed at length whether the range of motion of the elbow was important independent of elbow function. The majority view (> 70% approval) was that what mattered was the function of the elbow and the ability to undertake tasks rather than the ROM, and they agreed that this was better captured by other domains such as 'elbow function' and 'independence in activities of daily living'. They were asked to consider if there were any outcomes that were not included which might be important from their perspective, but they were happy that the list represented the key elements. When asked which outcome was most important, they unanimously offered 'pain' as the most important outcome for patients after elbow replacement. They were asked to consider alternative outcomes but were very clear that this was the primary outcome from their perspective. There was no clear consensus about the best timing for recording this primary outcome domain, but it was felt that it should be far enough from the time of surgery for post-surgical pain to have resolved, while being close enough to ensure patients were not lost to follow-up.

## Discussion

Inconsistency in the 'what' and 'how' of the assessment of outcome in research dealing with aspects of elbow surgery has been reported by other investigators.[3,10] This heterogeneity in the assessment of outcome leads to costly redundancy for investigators through the collection of futile data, makes it difficult or impossible to compare the outcomes of different studies and creates an unnecessary additional burden for patients. Defining a consensus on core outcome domains is important in order to increase efficiency and enable comparison of results between studies, as well as meta-analysis.

The nine mandatory core domains defined by the CODER study, 'return to work or normal daily role', delivery of care measured in the domains 'patient satisfaction with the outcome of surgery' and 'would the patient have the same operation again?', 'pain', 'revision', 'elbow function', 'independence in activities of daily living', 'health-related quality of life', and 'adverse events', are feasible and appear proportionate. The four mandatory core timepoints for data collection, baseline and three, 12, and 24 months postoperatively, balance the social burden on patients and the financial burden on research workers. Additional desirable data points may provide added value to address specific research questions. The core domain set and timepoints defined in the CODER study do not prevent the collection of additional outcome data when necessary to address research questions, but should be included in future studies about elbow replacement. The domains identified for elbow replacement in this study show good concordance with other areas of upper limb orthopaedic research. The OMERACT Shoulder Working Group identified 'pain', 'function', 'patient global – shoulder', and 'adverse events' including death as mandatory domains for clinical trials dealing with shoulder conditions.[11]

Although key stakeholder groups were represented, a limitation of this study was the low participation in the consensus meeting, and the relatively low patient and carer representation in the RT-Delphi study. Extensive efforts were made to try to recruit patients and carers, and several registered interest but did not complete the survey. This may reflect the relatively cumbersome process required to access the survey

electronically, having to complete a registration form in order to be sent an anonymized link through which to access the survey, which was necessary due to General Data Protection Regulation (GDPR) compliance requirements. We attempted to mitigate the imbalance of representation through engagement with a PPIE group following the online survey and there was strong support for the domains which were included. The PPIE group felt that measuring ROM provided no useful additional information. They reported that they were more interested in what could be done with the arm, which they felt could be captured in the function domain, than how much movement there was at the elbow. The PPIE group were keen to reduce the burden on patients and questioned the added value gained by data collection at six months postoperatively. The optimum timepoint for the recording of the primary outcome measure was not determined, but this is likely to depend on the research question under investigation. A minimum follow-up of two years is required for reporting the outcome of joint replacement by some orthopaedic journals.[12,13] This does not, however, define the time when the measure of the primary outcome should be recorded. Periods of time beyond two years were not considered in the CODER study. Embedding trials within joint registries can facilitate research into long-term outcomes as long as criteria for relevance, robustness, reliability and confidentiality are in place, but will not capture many of the core domains identified here without the use of linked or add-on datasets.[14] Due to the number of patients who participated in the online survey, a secondary survey within the research protocol to determine the mean clinician and patient scores which should be recorded for adverse events has not been reported as a reliable estimate of the mean could not be calculated.[8]

Future studies dealing with elbow replacement should include the nine mandatory core domains at four timepoints agreed in the CODER study to ensure consistency between studies and allow comparison of research findings. Pain can be used as the primary outcome in future studies dealing with elbow replacement when it is appropriate to do so. Further work is required to define the instruments that should be used to measure these outcome domains.

### Take home message

- The Core Outcome Domains for Elbow Replacement (CODER) study provides a consensus from key stakeholders on the core domains that should be used for the assessment elbow arthroplasty.
- This information can be used in future clinical research into elbow arthroplasty outcomes to improve consistency and allow comparison between studies.
- Further work is required to define the instruments that should be used to measure these domains.

## Social media

Follow A. C. Watts on X @myelbowdoc

## Supplementary material

Outcomes domains, domain descriptions, and participant responses from the real-time Delphi questionnaire.

## References


1. **Dee R**. Elbow replacement with the R. Dee prosthesis. *Acta Orthop Belg*. 1975;41(4):477–483.
2. **Watts AC**, **Hamoodi Z**, **McDaid C**, **Hewitt C**. Elbow arthroplasty research methods, outcome domains, and instruments used in clinical outcome studies: a scoping review. *Bone Joint J*. 2022;104-B(10):1148–1155.
3. **Riedel K**, **Beaton DE**. Update on the state of outcome measurement in total elbow arthroplasty research: identifying a need for consensus. *J Bone Joint Surg Am*. 2013;95-A(14):e97.
4. **Kirkham JJ**, **Davis K**, **Altman DG**, **et al**. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. *PLoS Med*. 2017;14(11):e1002447.
5. **Kirkham JJ**, **Gorst S**, **Altman DG**, **et al**. Core Outcome Set-STAndardised Protocol Items: the COS-STAP Statement. *Trials*. 2019;20(1):116.
6. **Trevelyan EG**, **Robinson PN**. Delphi methodology in health research: how to do it? *Eur J Integr Med*. 2015;7(4):423–428.
7. **No authors listed**. The OMERACT Handbook. OMERACT. 2021. https://omeract.org/wp-content/uploads/2023/09/OMERACT-Handbook-Chapter-Final-June_21_2021.pdf (date last accessed 14 August 2024).
8. **Watts A**. Core Outcome Domains Elbow Replacement. Open Science Framework. 2023. https://osf.io/jv2cb/ (date last accessed 14 August 2024).
9. **Dodd S**, **Clarke M**, **Becker L**, **Mavergames C**, **Fish R**, **Williamson PR**. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. *J Clin Epidemiol*. 2018;96:84–92.
10. **Evans JP**, **Smith CD**, **Fine NF**, **et al**. Clinical rating systems in elbow research-a systematic review exploring trends and distributions of use. *J Shoulder Elbow Surg*. 2018;27(4):e98–e106.
11. **Ramiro S**, **Page MJ**, **Whittle SL**, **et al**. The OMERACT core domain set for clinical trials of shoulder disorders. *J Rheumatol*. 2019;46(8):969–975.
12. **No authors listed**. Guide for authors. Journal of Shoulder and Elbow Surgery. 2024. https://www.sciencedirect.com/journal/journal-of-shoulder-and-elbow-surgery/publish/guide-for-authors (date last accessed 4 September 2024).
13. **No authors listed**. Instructions for Authors. The Journal of Bone and Joint Surgery. 2016. https://sites.jbjs.org/oa/oa_instructions-authors.pdf (date last accessed 4 September 2024).
14. **Mikita JS**, **Mitchel J**, **Gatto NM**, **et al**. Determining the suitability of registries for embedding clinical trials in the United States: a project of the Clinical Trials Transformation Initiative. *Ther Innov Regul Sci*. 2021;55(1):6–18.



**Author information:**

A. C. Watts, MBBS, BSc, FRCS, Professor of Orthopaedics and Honorary Consultant Orthopaedic Surgeon, Edge Hill University, Ormskirk, UK; Wrightington, Wigan and Leigh Teaching Hospitals NHS Foundation Trust, Wigan, UK; Health Sciences, University of York, York, UK.

C. McDaid, BSc, MSc, PhD, Professor of Applied Health Research, Department of Health Sciences, University of York, York, UK.

C. Hewitt, BSc, MSc, PhD, Director, York Trials Unit, University of York, York, UK.



**Author contributions:**

A. C. Watts: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft.
C. McDaid: Methodology, Supervision, Writing – review & editing.
C. Hewitt: Methodology, Supervision, Writing – review & editing.

**Funding statement:**

The authors received no financial or material support for the research, authorship, and/or publication of this article.

**ICMJE COI statement:**

A. C. Watts reports grants or contracts from NIHR, Zimmer Biomet, TRB Chemedica, and Stryker, royalties or licenses and patents planned, issued or pending from Adler, and payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Xiros, Medartis, and Arthrex, all of which are unrelated to this study.
C. McDaid reports grants or contracts from NIHR HTA and i4i panels, research funding from Hull Teaching Hospitals NHS Trust, and is the co-director of NIHR RSS, all of which are unrelated to this study.

**Data sharing:**
The data that support the findings for this study are available to other researchers from the corresponding author upon reasonable request.

**Acknowledgements:**
**CODER Consensus Group**
The authors wish to thank the following for their valuable contribution to the Consensus Group Meeting:
Pam Boekel, Marcus Bateman, Katy Boland, Jon Evans, Jan Herregodts, Deborah Higgs, Val Jones, Matthew Kent, Dani Rotman, Elaine Uppal, Sam Uppal, Roger van Riet, Alexander van Tongel, Adam Watts.
The authors wish to thank the following for their valuable contribution to the survey:
Michel van den Bekerom, Ville Äärimaa, Juan Ameztoy Gallego, Jenna Bardsley, Tomithy Batten, Enrico Bellato, Robert Bjorkenheim, Katy Boland, Jill Bown, Sarah Briggs, Pieter Caekebeke, Helen Conroy, Barbara Cooke, Lindsay Cunningham, Margo Drickx, Tapio Flinkkilä, Michael Goldring, Enrico Guerra, Andrew Hodson, Richard Jamieson, Val Jones, Kapil Kumar, Tuomas Lähdeoja, Antti Launonen, Joideep Phadnis, Dani Rotman, Anssi Ryösä, Pjotr Sarantsin, Metwally Sayedahmad, Marco Scacchi, Harvinder Singh, Neil Smith, Bakir Omar, Jill Thomas, Helen Tunnicliffe, Sam Uppal, Alexander Van Tongel, Sam Vollans.
**PPIE group**
The authors wish to thank the members of the Wrightington Wigan and Leigh Teaching Hospitals NHS Foundation Trust Patient Public Involvement and Engagement group for their valuable contribution.

**Ethical review statement:**
Ethical approval was obtained from the Health Sciences Research Committee, University of York (HSRGC/2022/258 /G, 30/09/2022) and, for patient recruitment, through the integrated research approval system (IRAS ID 320243, 30/03/2023) from the Health Research Authority in the UK. Participants were invited to join the study by letter and asked to register their interest and provide consent for use of their anonymized responses using the online registration tool.

**Trial registration number:**
The CODER study was registered on the Core Outcome Measures in Effectiveness Trials (COMET) database (October 2022) and the protocol was made available on the Open Science Framework platform.

This article was primary edited by J. Scott.

# Appendix B Annual Incidence of Orthopaedic Procedures based on National Consultant Information Programme (Hospital Episode Statistics) and Office for National Statistics data.

| | England Count (Apr 2019-Mar 2020) | Annual incidence/100 000 (Based on England Population n = 56550000 in 2020 (Source: NCIP and Office of National Statistics)) |
|---|---|---|
| Total knee replacement | 78993 | 139.69 |
| Hip replacement | 74077 | 130.99 |
| Fracture neck of femur surgery | 62748 | 110.96 |
| Arthroscopic knee | 46272 | 81.82 |
| Carpal tunnel release | 42716 | 75.54 |
| Dupuytren's | 15413 | 27.26 |
| Rotator cuff repair | 11788 | 20.85 |
| Cruciate ligament reconstruction | 10720 | 18.96 |
| 1st metatarsal osteotomy | 10048 | 17.77 |
| Acromioclavicular joint surgery | 8327 | 14.73 |
| Unicompartmental knee replacement | 8174 | 14.45 |
| Extensor tendon repair | 7705 | 13.63 |
| Shoulder replacement | 7688 | 13.60 |
| Trigger finger release | 7289 | 12.89 |
| 1st Metatarsophalangeal joint fusion | 5658 | 10.01 |
| Ganglion excision | 4985 | 8.82 |
| Revision knee replacement | 4975 | 8.80 |
| Revision hip replacement | 4968 | 8.79 |
| Shoulder stabilisation | 3789 | 6.70 |
| Frozen shoulder | 3619 | 6.40 |
| Subacromial decompression | 3618 | 6.40 |
| Revision hip replacement - non elective | 2808 | 4.97 |
| Ankle fusion | 1815 | 3.21 |
| Revision carpal tunnel | 1523 | 2.69 |
| Small joint replacement hand | 1415 | 2.50 |
| Patellofemoral replacement | 1174 | 2.08 |
| Excision calcific tendinopathy | 900 | 1.59 |
| Revision shoulder replacement | 874 | 1.55 |
| Ankle replacement | 871 | 1.54 |
| Wrist fusion | 808 | 1.43 |
| Revision cruxiate ligament recon | 721 | 1.27 |
| Elbow replacement | 411 | 0.73 |
| Arthroscopic biceps | 361 | 0.64 |
| Revision elbow replacement | 251 | 0.44 |

Horizontal line indicating the boundary for European Union Definition of Rare Disease 50/100 000). (NCIP: https://gettingitrightfirsttime.co.uk/associated_projects/ncip/) (ONS: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2020)

# Appendix C Minutes of CODER PPIE Meeting

Wrightington, Wigan and Leigh Teaching Hospitals NHS Foundation Trust

**PPI&EE Meeting**
Date: Wednesday 30th August 2023

Time: 10:00

Venue: Ashton Clinical Research Hub

## Minutes

| Item No | Item | Action |
|---|---|---|
| | Professor Adam Watts (AW) – WWL Consultant Orthopaedic Surgeon<br>Helen Spickett (HS) – WWL Research Sponsorship Manager<br>Zaid Hamoodi (ZH) – Research Fellow<br>Alison Hegarty (AH2) – WWL Research PA<br><br>PPI&E Group:<br>Jean Peet (JP)<br>Joy Winstanley (JW)<br>Maggie Skilling (MS)<br>Ann Heaton (AH)<br>Debra Baxter (DB)<br>Pauline Gregory (PG)<br>Malcolm Ryding (MR)<br>Renee Mellis (RM)<br><br>**Minutes**: Alison Hegarty (AH2) – WWL Research PA | |
| 1 | **Welcome and introduction.**<br><br>AW welcomed attendees and brief introductions from attendees were made. | |
| 2 | AW introduced CODER to the group, gave an overview of the study and demonstrated an example of an implant used in elbow replacement.<br><br>Presented slideshow (included with minutes, please see slides for further information).<br><br>**What are Outcome domains?**<br>Outcome domains are the key aspects of a condition that matter to patients and clinicians and can be measured to assess the effect of a treatment.<br><br>AW explained the mechanics of the elbow replacement, and how force is distributed through the replacement and bone it is connected to. Too much force may make it dislocate.<br><br>AH noted what mattered to her was being able to pick up her grandchildren, so waited quite a long time before having her elbow replacement. | |

| | | |
|---|---|---|
| | DB asked if there is any data to show how much pain a patient would be in before a replacement is dislocated.<br><br>The first part of the study was a scoping exercise, taking outcomes identified in 362 studies published between 1990-2021. This demonstrated that overall, a total of 583 measures across 18 domains were being used across these studies, making it difficult to compare study results.<br><br>An advisory group was formed made up of patients, carers and healthcare professionals and through a series of meetings identified a list of key outcomes to create a core set to be measured.<br>The proposal is that these core set of measures are the minimum measures that any future study should report on and therefore, they can be compared across all studies.<br><br>MS asked if there was no one older than 69.<br>AW advised that this was the age of the oldest responder to the survey.<br><br>AW concerned the patient representation during this process was small and is looking for further representation from patients from this group.<br><br>PG was concerned that she doesn't have lived experience.<br>AW advised that even if she doesn't, her opinion as a patient/public is still very valuable.<br><br>DB asked about the outcome: return to work after 12 months. Why not longer such as 24 months?<br>AW advised this was suggested by the patient responses and is measured from surgery to return to work. Most would be back in work at 12 months; therefore 24 months would not be relevant.<br>AW advised although some may deteriorate, this would not be measured by return-to-work but by another outcome.<br><br>Patient autonomy is patient being in control of their care. But the measurement was for patients' independence. Have now changed it to return to work or normal daily life.<br><br>AH noted work can be different for each person, such as a bar tender who constantly uses their elbow motion to pour drinks from the cask.<br><br>DB advised from experience that pushing yourself through pain to move can have a faster recovery time.<br><br>AW advised that this is why we report averages. More focus today on qualitative work on individual.<br><br><u>Question to Group: Is the return-to-work outcome important?</u><br>Group agreed this is important.<br>PG suggested to include 18 months as well as 12 months. | |

| | | |
|---|---|---|
| | AH advised that by the time people get to the 12-month mark, many ppl drop out after 12 months unless there are complication.<br>PG suggested adding 18 months instead of 24 months if there is a percentage of responses not going back to work.<br>AW advised that each individual study is different and in the design of future studies, researchers can add additional data point/measures, we are trying to establish what will be measured as a minimum standard.<br>AW was asked how many months he thinks. He suggested that if a patient doesn't return to work by 12 months, then the operation would be considered a failure so the 18 or 24 months would be redundant.<br><br>MS asked if carers were involved?<br>AW confirmed carers were involved.<br><br><u>Question from AW, if condition deteriorates after 12 months, would that change the patient's perspective?</u><br><br>Group responded that yes it would, although you would expect to know by 12 months if it is deteriorating.<br><br>AW asked if the group if the 24-month measure should therefore be removed?<br>Group agreed the 24-month measure would not be necessary.<br><br>Same would apply for 'would the patient have the same operation again' outcome.<br><br><u>Pain outcome</u><br>AH noted this was very important to her during her elbow replacement.<br>Rheumatoid Arthritis (RA) patients are used to managing pain, noted trauma patients may not.<br><br>MR asked what the trigger would be to revise the baseline outcome.<br>AW: dislocation, infection examples. Pain may be a trigger for revision, as may be a marker for infection.<br>MR asked how severe would pain need to be to revise the surgery?<br><br>HS at 24 months, look back and see if patient has had their operation.<br><br>AW asked if it is valid to have 3- and 6-months measure or are these too soon?<br>Group felt pain at every time interval is import.<br><br>AH noted that as a RA patient, you must come off meds when measuring pain.<br><br>JP asked that during physio, pain may be increased, would this count?<br><br>AW advised there are many variations of pain (such as pain at rest, pain at night etc), deliberately avoided the intricacies.<br><br>MS asked why the pain outcome is measured so often?<br>AW to identify if we can get rid of pain more quickly using one intervention over another. | |

| | | |
|---|---|---|
| | AW asked which times group would suggest removing.<br><br>MS would question 24 months pain measure; would this be necessary?<br>RM asked that if you are still having pain after 12 months and 24 months then that would suggest there is something wrong.<br>AW advised that if the pain is getting worse than this is a sign it is deteriorating.<br><br>AW asked the group to consider if all time points are necessary?<br><br>AW explained the following measures and the differences between them:<br>• Physical Function - Elbow function<br>• Independence in activities<br>• Active range of Movement<br><br>AW asked, does it matter as a patient if you do not have a full range of movement following surgery?<br><br>AH mentioned that she could not wash her hair but could do other things.<br>DB noted that it comes down to the individual as she would be ok for someone to wash her hair but with use of only one arm, losing full use of her other arm due to elbow surgery would be of great concern, different circumstances for different people.<br><br>AW added, does it matter if you have 10 degrees less range of movement than before your surgery?<br><br>DB noted it would be down to patient expectations.<br><br>AW asked if we wish to measure by activities rather than degrees of movement.<br>Group agreed activities, degrees are not important; elbow function is important.<br>Therefore, should we remove the active range of movement measure?<br>Agreed.<br><br><u>Health related quality of life</u> (cost related – important).<br>Is this a cost-effective treatment, related to obtaining grants for funding.<br><br><u>Adverse events</u> (complications etc related to the implant itself).<br>3 and 6 months suggested. Does the group feel there should be more time included such as 12 and/or 24 months.<br>Group suggested including 24 months.<br><br>JW noted you could have a fall and cause issue to implant.<br>DB added that should still be recorded.<br>AW advised there are mechanisms in place to record this data in trials.<br><br>AW asked if there is anything we should be capturing that isn't on the slide?<br>Group happy with suggestions on slide.<br><br>AW explained that we don't want to over burden research patients with unnecessary measures so again consider if all the time points are essential for the core data set. | |

| | | |
|---|---|---|
| | AH suggested that trauma patients may have higher expectations and wish for all time points.<br>AW suggested that the 6-month set is potentially burdensome on the patient and if we have data from 3 months and again at 12 months, then the 6 months may be irrelevant.<br><br>Group agreed with AW and agreed that these are a core set of measures may not need 6 months. Each individual study may offer an additional 6 month if required.<br><br>DB asked if patient understand how important the completion of outcome measure questionnaires is? Others have mentioned high dropout rates at 24 months.<br>AW advised this is something given consideration in all trials.<br><br>AW asked what is the most important question?<br>Group unanimously agreed, pain is the most important. | |

# Appendix D CODER questionnaire

CODER outcome questionaire version 1.0 21st March 23 AW

| Outcome Domain | Outcome | Outcome subcategory | Description | Implications for patient of adverse event | Please rate the importance of each outcome for total elbow replacement from 1 (least important) to 9 (most important). We are interested in your own view, there is no right or wrong response. If your rating is more than 2 points away from the median aggregate response from previous survey participants you can povide your reason for your response. |
|---|---|---|---|---|---|
| Societal Carer Burden | Non-homebound discharge | | Patient discharged to nursing/residential home or step down care | | 1 2 3 4 5 6 7 8 9 |
| Personal Circumstances | **Patient autonomy** | | Patient in control of their own circumstances and/or activities of daily living | | 1 2 3 4 5 6 7 8 9 |
| Cognitive function | **Cognitive status** | | An individual's level of awareness with perception, reasoning, judgment, intuition, and memory | | 1 2 3 4 5 6 7 8 9 |
| Economic resource | **Hospital costs** | | Costs incurred by hospital for treatment | | 1 2 3 4 5 6 7 8 9 |
| | **Discharge disposition** | | Patient discharged home, to nursing home or step down care | | 1 2 3 4 5 6 7 8 9 |
| | **Cost-utility analysis** | | Ratio between cost of treatment and number of years lived in full health | | 1 2 3 4 5 6 7 8 9 |
| Psychiatric | **Mental Health** | | Emotional, psychological, behavioral and social health | | 1 2 3 4 5 6 7 8 9 |
| Emotional wellbeing | **Emotional wellbeing** | | A patient's emotional state | | 1 2 3 4 5 6 7 8 9 |
| | **Confident and in control of your life** | | Patient's emotional state | | 1 2 3 4 5 6 7 8 9 |
| Social function | **Social-psychological function** | | Patient's ability to carry out social tasks and/or develop and maintain social relationships | | 1 2 3 4 5 6 7 8 9 |
| | **Time to return to sport** | | Period of time from surgery to return to sport | | 1 2 3 4 5 6 7 8 9 |
| | **Participation in sport** | | Patient able to return to sport after surgery | | 1 2 3 4 5 6 7 8 9 |
| | **Grade of sport participation** | | The level of sport that the patient has been able to return to | | 1 2 3 4 5 6 7 8 9 |
| Perceived health | General Health Status | | Patient's overall assessment of their health | | 1 2 3 4 5 6 7 8 9 |
| Mortality survival | **In hospital mortality** | | Death occuring during the hospital admission for elbow replacement | | 1 2 3 4 5 6 7 8 9 |
| | **30 day mortality** | | Death occuring within 30 days of elbow replacement | | 1 2 3 4 5 6 7 8 9 |
| | **90 day mortality** | | Death occuring within 90 days of elbow replacement | | 1 2 3 4 5 6 7 8 9 |
| | **Mortality** | | Death occuring at any time point between surgery and final follow-up | | 1 2 3 4 5 6 7 8 9 |
| Role function | **Return to work** | | Patient able to return to work after surgery | | 1 2 3 4 5 6 7 8 9 |
| | **Time to return to work** | | Period of time from surgery to return to work | | 1 2 3 4 5 6 7 8 9 |
| | **Return to unrestricted function** | | Patient able to return to normal activities without restriction | | 1 2 3 4 5 6 7 8 9 |
| Delivery of care | **Satisfaction with surgery/treatment** | | Patient's self reported happiness with the surgery or treatment / acceptable state of symptoms | | 1 2 3 4 5 6 7 8 9 |
| | **Satisfaction with results of procedure** | | Patient's self reported happiness with the outcome of the procedure / acceptable state of symptoms | | 1 2 3 4 5 6 7 8 9 |
| | **Patient satisfaction** | | Patient's self reported happiness with care and outcome in general | | 1 2 3 4 5 6 7 8 9 |
| | **Satisfaction with elbow function** | | Patient's self reported happiness with the use of the elbow | | 1 2 3 4 5 6 7 8 9 |
| | **Success of treatment** | | Patient's assessment of the outcome of the intervention | | 1 2 3 4 5 6 7 8 9 |
| | **Room for improvement** | | Patient's assessment of how much the intervention could be improved | | 1 2 3 4 5 6 7 8 9 |
| | **Would patient have the same operation again?** | | Patient's assessment of how likely it is they would have the same intervention if they had the same problem in the future | | 1 2 3 4 5 6 7 8 9 |
| | **Patient assessment of outcome** | | Patient's assessment of the outcome of the intervention | | 1 2 3 4 5 6 7 8 9 |
| Nervous system | **Pain** | | Patient's assessment of their level of pain in general | | 1 2 3 4 5 6 7 8 9 |
| | **Pain at rest** | | Patient's assessment of their level of pain when not active | | 1 2 3 4 5 6 7 8 9 |
| | **Pain on movement** | | Patient's assessment of their level of pain when moving the arm | | 1 2 3 4 5 6 7 8 9 |
| | **Pain with activity** | | Patient's assessment of their level of pain with activity | | 1 2 3 4 5 6 7 8 9 |
| | **Night pain** | | Patient's assessment of their level of pain at night | | 1 2 3 4 5 6 7 8 9 |
| | **Average daily pain** | | Patient's assessment of their average level of pain each day | | 1 2 3 4 5 6 7 8 9 |
| | **Pain at worst** | | Patient's assessment of their level of pain when it is at its worst | | 1 2 3 4 5 6 7 8 9 |
| | **Tenderness** | | Patient's experience of pain when the elbow is touched | | 1 2 3 4 5 6 7 8 9 |
| | **Pain medication used** | | Measurement of the quantity of pain relieving drugs used by the patient | | 1 2 3 4 5 6 7 8 9 |
| | **Neurological status** | | Physician assessment of the nerve function in the arm and hand | | 1 2 3 4 5 6 7 8 9 |
| | **Ulnar nerve function** | | Physician assessement of the function of the ulnar nerve | | 1 2 3 4 5 6 7 8 9 |
| | **Radial nerve function** | | Physician assessement of the function of the radial nerve | | 1 2 3 4 5 6 7 8 9 |
| Hospital resource | **Operation duration** | | How long the operation takes in minutes | | 1 2 3 4 5 6 7 8 9 |
| | **Length of stay** | | Number of days that the patient stays in hospital | | 1 2 3 4 5 6 7 8 9 |

| | | | | |
|---|---|---|---|---|
| | Re-admission | | Patient admitted to the same or another healthcare institution after discharge (excludes step-down care) | 1 2 3 4 5 6 7 8 9 |
| | Re-admission within 30 days | | Patient admitted to the same or another healthcare institution up to 30 days after discharge (excludes step-down care) | 1 2 3 4 5 6 7 8 9 |
| | Re-admission within 90 days | | Patient admitted to the same or another healthcare institution up to 90 days after discharge (excludes step-down care) | 1 2 3 4 5 6 7 8 9 |
| | Return to operating room | | A need to take the patient back to the operating room due to post operative problems | 1 2 3 4 5 6 7 8 9 |
| | Unplanned intubation | | A need to put a ventilation tube back in to the patient due to problems after surgery | 1 2 3 4 5 6 7 8 9 |
| | Healthcare utilisation | | The patient's total use of healthcare resources related to the episode of care | 1 2 3 4 5 6 7 8 9 |
| Need for further intervent | Re-operation | Re-operation | The need for any additional surgery after the first intervention excluding surgery that involves the elbow replacement | 1 2 3 4 5 6 7 8 9 |
| | | Amputation | Removal of a limb - for example through upper arm to control infection in elbow | 1 2 3 4 5 6 7 8 9 |
| | | Arthrodesis | Attempt to fuse bones above and below elbow to make one continuous bone | 1 2 3 4 5 6 7 8 9 |
| | | Lateral ligament reconstruction | Rebuilding the soft tissue that holds the bones together at the elbow | 1 2 3 4 5 6 7 8 9 |
| | | Secondary wound suture | Using thread to close a wound if it has opened up after surgery | 1 2 3 4 5 6 7 8 9 |
| | | Ulnar nerve decompression | Release of the soft tissues around the nerve that gives feeling to the little and ring finger to prevent damage | 1 2 3 4 5 6 7 8 9 |
| | | Return to operating room | Taking a patient back to the operating theatre for any reason | 1 2 3 4 5 6 7 8 9 |
| | | Incision and drainage | Opening a wound in the operating theatre usually for treatment of infection or a collection of blood | 1 2 3 4 5 6 7 8 9 |
| | | Irrigation and debridement | Washing out of a woud in the operating theatre and removing dead tissue usually for the treatment of infection | 1 2 3 4 5 6 7 8 9 |
| | | Manipulation under anaesthesia | Trying to move the joint with the patient asleep in the operating theatre | 1 2 3 4 5 6 7 8 9 |
| | | Lysis of adhesions | Opening the joint to remove scar tissue causing a restriction of movement of the elbow | 1 2 3 4 5 6 7 8 9 |
| | Time to re-operation | | Time period between the first surgery to implant the elbow replacement and any additional surgery (needs specification of units of measurement (hours, days, weeks, months, years)) | 1 2 3 4 5 6 7 8 9 |
| | Revision | Revision surgery | Any further operation involving the elbow replacement. This includes adjustments, additions, exchanges or removal of all or part of elbow implant components | 1 2 3 4 5 6 7 8 9 |
| | | Reason for revision | The explanation for why the implant has to be altered or removed | 1 2 3 4 5 6 7 8 9 |
| | | Implant removal | Taking the implant out of the patient | 1 2 3 4 5 6 7 8 9 |
| | | Implant survival | Whether the elbow replacement is still in the patient | 1 2 3 4 5 6 7 8 9 |
| | | Implant exchange | Changing of all or part of the elbow replacement | 1 2 3 4 5 6 7 8 9 |
| | | Implant survival - aseptic loosening | A measure of the number of elbow replacements that have not been removed or altered for coming loose from the bone without infection over time | 1 2 3 4 5 6 7 8 9 |
| | | Implant survival - Revision mechanical failure | A measure of the number of elbow replacements that have not been removed or altered because of something going wrong with the implant itself (such as fracture, component wear or loosening of the components) over time | 1 2 3 4 5 6 7 8 9 |
| | | Implant survival - Revision mechanical failure or loosening | A measure of the number of elbow replacements that have not been removed or altered because of something going wrong with the implant itself (such as fracture or component wear) or for coming loose from the bone for any reason over time | 1 2 3 4 5 6 7 8 9 |
| | | Implant survival - loosening | A measure of the number of elbow replacements that have not been removed or altered for coming loose from the bone for any reason over time | 1 2 3 4 5 6 7 8 9 |
| | | Implant survival - Revision | A measure of the number of elbow replacements that have not been removed or altered for any reason over time | 1 2 3 4 5 6 7 8 9 |
| | | Implant survival - definite loosening and bushing wear | A measure of the number of elbow replacements that have not been removed or altered for clear evidence of coming loose from the bone for any reason or wear of the plastic components over time | 1 2 3 4 5 6 7 8 9 |
| | | Implant survival - possible loosening and bushing wear | A measure of the number of elbow replacements that have not been removed or altered for a suspicion of coming loose from the bone for any reason or wear of the plastic or metal components over time | 1 2 3 4 5 6 7 8 9 |
| | | Implant survival - revision and loosening | A measure of the number of elbow replacements that have not been removed or altered for any reason for coming loose from the bone for any reason over time | 1 2 3 4 5 6 7 8 9 |
| | | Cumulative success rate | An alternative description of implant survival | 1 2 3 4 5 6 7 8 9 |
| MSK connective tissue | Stability | Stability | The ability of the elbow joint to maintain a fixed position with a force applied . No clinical evidence of joint dislocation or subluxation. Excludes radiographic evidence. | 1 2 3 4 5 6 7 8 9 |
| | | Stability on varus/valgus stress | The ability of the elbow joint to maintain a fixed position with a force applied inwards or outwards. No clinical evidence of joint dislocation or subluxation. Excludes radiographic evidence. | 1 2 3 4 5 6 7 8 9 |
| | | Instability - subjective | The patient's own perception of the presence of the elbow not maintaining correct position in rest or during movement | 1 2 3 4 5 6 7 8 9 |
| | | Stability of prosthesis in axial compression | The ability of the elbow joint to maintain a fixed position with a force applied through forearm towards elbow. No clinical evidence of joint dislocation or subluxation. Excludes radiographic evidence. | 1 2 3 4 5 6 7 8 9 |
| | Radiographic appearance | Radiographic appearance | Changes seen on plain X-ray images of the elbow | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic alignment | Radiographic assessement of the alignment of the implant with the bone | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic fixation | Fixation method (for example cementing) and quality of fixation seen in radiographs (No radiographic evidence of loosening of the implant) | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic success | No apparent radiographic problem identified | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic lengthening | Radiographic evidence that the implant has been positioned in a way to lengthen the segment | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic head size | Radiographic assessment of the size of the radial head | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic congruence PRUJ | Radiographic assessement of the way the radial head fits into the socket on the ulna | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic stem size | Radiographic assessment of the size of the stem of a prosthesis | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic stem positioning | Radiographic assessment of the position of an implant within the bone | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic quality of cementing technique | Radiographic assessment of the quality of the cementation between the implant and the bone | 1 2 3 4 5 6 7 8 9 |

| | | | | |
|---|---|---|---|---|
| | | Radiographic implant positioning | Radiographic assessement of the position of the implant within the bone | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic prosthetic sizing | Radiographic assessment of the size of the implant | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic bone graft integration | Radiographic assessment of the degree to which any bone graft has attached to the host bone | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic cortical fit | Radiographic assessment of the space between the implant and the bone | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic valgus tilting | Radiographic assessment of the alignment of the implant within the bone | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic joint congruity | Radiographic assessment of the conformity of the two sides of the joint | 1 2 3 4 5 6 7 8 9 |
| | Other | Synovitis | Inflammation of the joint (lining) | 1 2 3 4 5 6 7 8 9 |
| | | Deformity | Change in the appearance of the elbow | 1 2 3 4 5 6 7 8 9 |
| | | Cubitus valgus | Alignment of the forearm relative to the arm with the elbow in a fully straightened position | 1 2 3 4 5 6 7 8 9 |
| | | Swelling | Increase in the size of the elbow. May be caused by oedema(fluid in the soft tissues). | 1 2 3 4 5 6 7 8 9 |
| | | Joint position sense | The ability to determine how bent the elbow is without looking at it | 1 2 3 4 5 6 7 8 9 |
| | | Deficient bone stock | The amount of bone missing around the elbow | 1 2 3 4 5 6 7 8 9 |
| | | MCL healing | A measure of the amount of healing of the medial ligament of the elbow | 1 2 3 4 5 6 7 8 9 |
| | | Implant size | The choice of size of the implant used | 1 2 3 4 5 6 7 8 9 |
| Physical function | Function/disability | Function/Disability | A measure of the use of the arm or impairment of use | 1 2 3 4 5 6 7 8 9 |
| | Range of movement | Range of movement | A measure of the amount of movement of the elbow or forearm | 1 2 3 4 5 6 7 8 9 |
| | | Range of pronosupination | A measure of the amount of rotational movement of the forearm from palm up to palm down | 1 2 3 4 5 6 7 8 9 |
| | | Range of flexion/extension | A measure of the amount of movement of the elbow from straight to bent | 1 2 3 4 5 6 7 8 9 |
| | | Range of flexion/extension (active) | A measure of the amount of movement of the elbow from straight to bent by the patient only | 1 2 3 4 5 6 7 8 9 |
| | | Range of flexion/extension (passive) | A measure of the amount of movement of the elbow from straight to bent with a clinician helping to move the limb | 1 2 3 4 5 6 7 8 9 |
| | | Range of pronosupination (active) | A measure of the amount of rotational movement of the forearm from palm up to palm down by the patient only | 1 2 3 4 5 6 7 8 9 |
| | | Range of pronosupination (passive) | A measure of the amount of movement of the forearm from palm up to palm down with a clinician helping to move the limb | 1 2 3 4 5 6 7 8 9 |
| | Strength | Strength | A measure of the power of the muscles around the elbow | 1 2 3 4 5 6 7 8 9 |
| | | Grip strength | A measure of the power of the muscles to grip an object in the hand | 1 2 3 4 5 6 7 8 9 |
| | | Flexion Strength | A measure of the power of the muscles around the elbow to move the elbow from straight to bent | 1 2 3 4 5 6 7 8 9 |
| | | Extension strength | A measure of the power of the muscles around the elbow to move the elbow from bent to straight | 1 2 3 4 5 6 7 8 9 |
| | | Pronation strength | A measure of the power of the muscles around the elbow to move the forearm from palm up to palm down | 1 2 3 4 5 6 7 8 9 |
| | | Supination strength | A measure of the power of the muscles around the elbow to move the forearm from palm down to palm up | 1 2 3 4 5 6 7 8 9 |
| | | Maximum grip strength | A measure of the maximum power of the muscles to grip an object in the hand | 1 2 3 4 5 6 7 8 9 |
| | | Pinch strength | A measure of the power of the muscles to pinch an object between finger and thumb | 1 2 3 4 5 6 7 8 9 |
| | | Isometric flexion strength | A measure of the power of the muscles around the elbow when the elbow is held in a fixed position and the patient tries to move the elbow from straight to bent | 1 2 3 4 5 6 7 8 9 |
| | | Isometric extension strength | A measure of the power of the muscles around the elbow when the elbow is held in a fixed position and the patient tries to move the elbow from bent to straight | 1 2 3 4 5 6 7 8 9 |
| | | Isometric pronation strength | A measure of the power of the muscles around the elbow when the elbow is held in a fixed position and the patient tries to move the forearm from palm up to palm down | 1 2 3 4 5 6 7 8 9 |
| | | Isometric supination strength | A measure of the power of the muscles around the elbow when the elbow is held in a fixed position and the patient tries to move the forearm from palm down to palm up | 1 2 3 4 5 6 7 8 9 |
| | | Triceps strength | A measure of the power of the triceps muscle to move the elbow from bent to straight | 1 2 3 4 5 6 7 8 9 |
| | Activities of daily living | Activities of daily living | A measure of the ability to do everyday tasks | 1 2 3 4 5 6 7 8 9 |
| | | Autonomy on ADL | A measure of the patients independence for everyday tasks | 1 2 3 4 5 6 7 8 9 |
| | | Self-care activities | A measure of the ability of the patient to do personal care such as hygiene | 1 2 3 4 5 6 7 8 9 |
| | | Housekeeping ability | A measure of the ability of the patient to do household activities such as cleaning | 1 2 3 4 5 6 7 8 9 |
| | Locomotion | Ambulation | A measure of the patients ability to move around | 1 2 3 4 5 6 7 8 9 |
| Quality of life | Health related quality of life | Health related quality of life | A measure of the patients own perception of the impact of their health on their life | 1 2 3 4 5 6 7 8 9 |
| Adverse events | Adverse event | All adverse events (no time limit) | A measure of complications occuring after surgery | 1 2 3 4 5 6 7 8 9 |
| | | 30-day adverse event | A measure of complications occuring within 30 days following surgery | 1 2 3 4 5 6 7 8 9 |
| | | Intra-operative complication | A measure of complications occuring during an operation | 1 2 3 4 5 6 7 8 9 |
| | | peri-operative complication | A measure of complications occuring around the time of an operation | 1 2 3 4 5 6 7 8 9 |
| | | post-operative complication | A measure of complications occuring after an operation | 1 2 3 4 5 6 7 8 9 |

| | | | | | |
|---|---|---|---|---|---|
| | | Major complications | A measure of serious adverse events or complications of surgery such as death or those needing additional surgery | | 1 2 3 4 5 6 7 8 9 |
| | | Minor complications | A measure of less serious adverse events or complications of surgery such temporary problems that get better by themselves | | 1 2 3 4 5 6 7 8 9 |
| | Pain | Post operative pain | The presence of persistent post-operative pain | Relatively uncommon. May increase hospital stay, may require additional pain treatment | 1 2 3 4 5 6 7 8 9 |
| | Infection | Superficial infection | The occurrence of superficial infection at the surgical site after surgery | Relatively uncommon. May require antibiotic treatment or readmission to hospital and further surgery | 1 2 3 4 5 6 7 8 9 |
| | | Deep infection | The occurrence of deep infection at the surgical site after surgery | Relativly uncommon. May require antibiotic treatment or readmission to hospital and further surgery. | 1 2 3 4 5 6 7 8 9 |
| | | Systemic sepsis | The occurrence of infection causing a whole body reaction after surgery | Uncommon. Will require readmission to hospital for antibiotics and possibly further surgery. In the worst case there is a risk of death. | 1 2 3 4 5 6 7 8 9 |
| | | Septic shock | The occurrence of infection causing a whole body reaction resulting in a decrease in blood pressure after surgery | Uncommon. Will require readmission to hospital for antibiotics and possibly further surgery. In the worst case there is a risk of death. | 1 2 3 4 5 6 7 8 9 |
| | Tendon | Triceps insufficiency | Loss of function of triceps muscle after surgery | Relatively uncommon. Affects function of the arm, may require prolonged physiotherapy or further surgery | 1 2 3 4 5 6 7 8 9 |
| | Nerve | Ulnar nerve dysfunction | The occurrence of an alteration of ulnar nerve function after surgery (numbness little and ring finger or weakness in hand) | Relatively common. Affects function of the limb or hand, may require prolonged physiotherapy or further surgery | 1 2 3 4 5 6 7 8 9 |
| | | Radial nerve dysfunction | The occurrence of an alteration of radial nerve function after surgery (wrist and finger drop) | Uncommon. Affects function of the limb or hand, may require prolonged physiotherapy or further surgery | 1 2 3 4 5 6 7 8 9 |
| | | Posterior interosseous nerve dysfunction | The occurrence of an alteration of posterior interosseous nerve function after surgery (finger drop) | Uncommon. Affects function of the limb or hand, may require prolonged physiotherapy or further surgery | 1 2 3 4 5 6 7 8 9 |
| | | Transient neurological deficit | The occurrence of an alteration of nerve function after surgery (numbness or weakness) with evidence of complete recovery | Relatively common. Temporary affects on function of the limb or hand, may require prolonged physiotherapy. | 1 2 3 4 5 6 7 8 9 |
| | | Permanent neurological deficit | The occurrence of an alteration of nerve function after surgery (numbness or weakness) without evidence of complete recovery | Relatively uncommon. Affects function of the limb or hand, may require prolonged physiotherapy or further surgery | 1 2 3 4 5 6 7 8 9 |
| | | Neurological - sensory | The occurrence of an alteration of nerve function affecting feeling or touch after surgery | Relatively common. May affect function of the limb or hand. May require further surgery. | 1 2 3 4 5 6 7 8 9 |
| | | Neurological - motor | The occurrence of an alteration of nerve function causing weakness after surgery | Relatively uncommon. May affect function of the limb or hand. May require further surgery. | 1 2 3 4 5 6 7 8 9 |
| | Wound | Bleeding | The occurrence of abnormal bleeding after surgery | Relatively uncommon. May require repeat hospital admission and further surgery. | 1 2 3 4 5 6 7 8 9 |
| | | Haematoma | The presence of an abnormal collection of blood in the wound after surgery | Relatively uncommon. May cause pain and swelling. May require drainage with a needle. May increase the risk of infection. May require further surgery. | 1 2 3 4 5 6 7 8 9 |
| | | Wound healing problems | The presence of a wound that has not healed in the usual time or manner after surgery | Relatively uncommon. May require prolonged treatmet with dressings. May require further surgery. | 1 2 3 4 5 6 7 8 9 |
| | | Scar related problems | The occurrence of problems related to the scar at the site of surgery | Uncommon. May require additional out-patient visits. Might affect the cosmetic appearance of the elbow. | 1 2 3 4 5 6 7 8 9 |
| | | Skin necrosis | The occurrence of death of an area of skin at the site of the surgery | Uncommon. May require additional out-patient visits. May require additional surgical intervention | 1 2 3 4 5 6 7 8 9 |
| | | Pressure sore | The occurrence of a defect in the skin due to a localised pressure effect | Uncommon. May require additional out-patient visits. May require additional surgical intervention | 1 2 3 4 5 6 7 8 9 |
| | Range of movement | Range of movement deficit | The presence of a loss of range of movement below a specified number of degrees | Relatively common. May require additional out-patient and physiotherapy visits. May require additional surgery to improve range of movement | 1 2 3 4 5 6 7 8 9 |
| | Bone | Intra-operative fracture | The occurrence of a fracture of a bone during the surgical procedure | Relatively uncommon. May prolong the surgery and increase risks of implant loosening or infection or pain | 1 2 3 4 5 6 7 8 9 |
| | Radiological | Radiographic loosening | The presence of X-ray changes consistent with the implant or fixation material coming loose within the bone | Relatively common. This may require additional out-patient visits and investigations and may need revision surgery if causing pain or bone loss. | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic Osteoarthritis | The presence of X-ray changes consistent with wear and tear arthritis | Relatively common after some types of elbow replacement such as radial head replacement. This may cause pain and may lead to a need for additional surgery | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic Heterotopic Ossification | The presence of X-ray changes consistent with abnormal bone forming in the soft tissues | Relatively common. This may not cause a problem for the patient but can be related to elbow stiffness and may require additional surgery | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic periprosthetic fracture | The presence of X-ray changes consistent with fracture of the bone around the implant | Relatively common. This is likely to be painful. This may require treatment in a cast or additional surgery to fix the bone or replace the implant. | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic dislocation | The presence of X-ray changes of the implants moving out of proper positioning relative to each other | relatively uncommon. This may require revision surgery | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic polyethylene wear | The presence of X-ray changes consistent with wear of the plastic parts of the implant | Relatively common. Risk increases with time from operation. This may require revision surgery | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic implant failure | X-ray evidence of failure of an implant usually by breakage | Relatively uncommon. Is likely to require revision surgery to replace the broken implant. | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic cement leakage | X-ray evidence of bone cement escape into the soft tissues | Relatively common. May cause no problem for the patient but may cause stiffness or rarely nerve problems | 1 2 3 4 5 6 7 8 9 |
| | | Radiographic bone perforation | X-ray evidence of a hole being made in the side of the bone during surgery | Relatively common. May not cause a problem for the patient but may lead to early implant failure and revision surgery. | 1 2 3 4 5 6 7 8 9 |
| | Clotting Events | Deep venous thrombosis | Blood clot, usually in the blood vessels of the leg | treatment with anticoagulant drugs. May require hospital re-admission. Rarely may lead to stroke. In the worst case there is a rare risk of death. | 1 2 3 4 5 6 7 8 9 |
| | | Pulmonary embolism | Blood clot in the blood vessels of the lung | require prolonged treatment wth anticoagulant drugs. May require readmission to hospital. In the worst case there is a rare risk of death. | 1 2 3 4 5 6 7 8 9 |
| | Bleeding | Haemorrhage | Bleeding during or after surgery | Uncommon. May lead to a drop in blood pressure and circulating blood cells leading to blood transfusion. In the worst case there is a rare risk of death. | 1 2 3 4 5 6 7 8 9 |
| | Respiratory | Respiratory failure | Inability to get enough air exchange through the lungs to support life | Uncommon. Will cause shortness of breath. May require readmission to hospital or prolonged hospital stay for respiratory support. In the worst case there is a rare risk of death. | 1 2 3 4 5 6 7 8 9 |
| | | Pneumonia | Infection in the lung | require antibiotic treatment by mouth or through a drip. May require readmission to hospital. In the worst case there is a rare risk of death. | 1 2 3 4 5 6 7 8 9 |
| | Cardiac | Cardiac arrest | The heart stopping | Uncommon. Prolong hospital stay or require re-admission. May require additional drug treatment or surgical procedures. In the worst case there is a risk of death. | 1 2 3 4 5 6 7 8 9 |
| | | Myocardial infarct | Blockage in the heart affecting the heart pump | Uncommon. May cause chest pain, shortness of breath, palpitations. May prolong hospital stay or require re-admission. May require additional drug treatment or surgical procedures. | 1 2 3 4 5 6 7 8 9 |
| | Central nervous system disorder | Cerebrovascular accident | A blood clot or bleeding in the brain blood vessel | Uncommon. May lead to altered conciousness, weakness, altered brain function. In the worst case there is a rare risk of death. | 1 2 3 4 5 6 7 8 9 |
| | GI disorder | Ileus | Loss of normal bowel movement | Uncommon. May cause abdominal pain, sickness, bloating, constipation. May prolong hospital stay or readmission. | 1 2 3 4 5 6 7 8 9 |
| | Renal | Urinary tract infection | Infection affecting the kindeys, bladder or connecting pipes | Uncommon. May cause pain, fever, sickness, confusion. May prolong hospital stay or require readmission. May require treatment with antibiotics with tablets or through a drip. | 1 2 3 4 5 6 7 8 9 |
| | | Acute Renal Failure | New damage to the kidney function | Uncommon. May cause pain. May prolong hospital stay or require readmission. May require medical treatment or rarely dialysis. | 1 2 3 4 5 6 7 8 9 |
| | Vascular | Perithperal vascular disorder | Blockage or narrowing of the blood vessels in the arms or more usually the legs | Uncommon. May cause pain or ulceration. May require surgical intervention. May lead to amputation. | 1 2 3 4 5 6 7 8 9 |
| | Compartment syndrome | | Increased pressure in a muscle compartment of the arm or forearm | Uncommon. May cause pain and muscle damage. May require further surgery. May lead to kidney damange or to loss of a limb. | 1 2 3 4 5 6 7 8 9 |
| | Complex regional pain syndrome | | A condition that causes abnormal prolonged pain, swelling and stiffness of the upper limb | Uncommon. May cause prolonged pain and stiffness and affect function of the arm. May need additonal physiotherapy treatment and pain killers. | 1 2 3 4 5 6 7 8 9 |

| | | | | | |
|---|---|---|---|---|---|
| | Altered mental status | | A change in the mental function of the patient after surgery | Uncommon. May require additional supportive care. May prolong hospital stay. | 1 2 3 4 5 6 7 8 9 |

# **Appendix E** CODER real-time DELPHI respondent's scores by question.

CODER Appendix 1

| Outcome Dom | Outcome | Outcome subcategory | Description | % respondents for each score | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | % score >6 |
| Societal Carer Burde | Non-homebound discharge | | Patient discharged to nursing/residential home or step down care | 15 | 19 | 4 | 23 | 15 | 15 | 4 | 4 | 0 | 8 |
| Personal Circumstan | Patient autonomy | | Patient in control of their own circumstances and/or activities of daily living | 0 | 0 | 0 | 0 | 16 | 4 | 28 | 40 | 12 | 80 |
| Cognitive function | Cognitive status | | An individual's level of awareness with perception, reasonin | 4 | 0 | 4 | 4 | 16 | 16 | 32 | 20 | 4 | 56 |
| Economic resource | Hospital costs | | Costs incurred by hospital for treatment | 12 | 8 | 12 | 28 | 12 | 4 | 8 | 12 | 4 | 24 |
| | Discharge disposition | | Patient discharged home, to nursing home or step down care | 16 | 4 | 4 | 8 | 36 | 8 | 4 | 16 | 4 | 24 |
| | Cost-utility analysis | | Ratio between cost of treatment and number of years lived in full health | 12 | 0 | 0 | 4 | 20 | 12 | 24 | 20 | 8 | 52 |
| Psychiatric | Mental Health | | Emotional, psychological, behavioral and social health | 0 | 0 | 9 | 4 | 17 | 9 | 26 | 17 | 17 | 60 |
| Emotional wellbeing | Emotional wellbeing | | A patient's emotional state | 4 | 9 | 0 | 0 | 9 | 22 | 22 | 26 | 9 | 57 |
| | Confident and in control of your life | | Patient's emotional state | 4 | 0 | 0 | 4 | 9 | 17 | 30 | 26 | 9 | 65 |
| Social function | Social-psychological function | | Patient's ability to carry out social tasks and/or develop and maintain social relationships | 0 | 5 | 5 | 5 | 5 | 14 | 27 | 24 | 18 | 69 |
| | Time to return to sport | | Period of time from surgery to return to sport | 13 | 0 | 13 | 13 | 38 | 13 | 13 | 0 | 0 | 13 |
| | Participation in sport | | Patient able to return to sport after surgery | 9 | 5 | 14 | 9 | 5 | 27 | 23 | 0 | 9 | 32 |
| | Grade of sport participation | | The level of sport that the patient has been able to return to | 13 | 0 | 25 | 0 | 50 | 0 | 13 | 0 | 0 | 13 |
| Perceived health | General Health Status | | Patient's overall assessment of their health | 0 | 5 | 9 | 5 | 9 | 27 | 23 | 5 | 18 | 46 |
| Mortality survival | Mortality | | Death occuring at any time point between surgery and final follow-up | 5 | 5 | 9 | 5 | 5 | 27 | 9 | 23 | 14 | 46 |
| | 30 day mortality | | Death occuring within 30 days of elbow replacement | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 7 | 71 | 99 |
| | 90 day mortality | | Death occuring within 90 days of elbow replacement | 0 | 0 | 0 | 0 | 0 | 7 | 36 | 29 | 29 | 94 |
| | In hospital mortality | | Death occuring during the hospital admission for elbow replacement | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 0 | 79 | 100 |
| Role function | Return to work | | Patient able to return to work after surgery | 5 | 0 | 0 | 0 | 9 | 9 | 23 | 36 | 18 | 77 |
| | Time to return to work | | Period of time from surgery to return to work | 9 | 0 | 5 | 14 | 0 | 18 | 23 | 23 | 5 | 51 |
| | Return to unrestricted function | | Patient able to return to normal activities without restriction | 0 | 5 | 0 | 5 | 9 | 9 | 27 | 27 | 18 | 72 |
| Delivery of care | Satisfaction with surgery/treatment | | Patient's self reported happiness with the surgery or treatment / acceptable state of symptoms | 0 | 0 | 0 | 0 | 5 | 10 | 5 | 24 | 57 | 86 |
| | Satisfaction with results of procedure | | Patient's self reported happiness with the outcome of the procedure / acceptable state of symptoms | 5 | 0 | 0 | 0 | 0 | 5 | 10 | 19 | 62 | 91 |
| | Patient satisfaction | | Patient's self reported happiness with care and outcome in g | 0 | 0 | 0 | 0 | 0 | 0 | 32 | 18 | 50 | 100 |
| | Satisfaction with elbow function | | Patient's self reported happiness with the use of the elbow | 0 | 0 | 0 | 0 | 5 | 5 | 19 | 14 | 57 | 90 |
| | Success of treatment | | Patient's assessment of the outcome of the intervention | 0 | 0 | 0 | 0 | 0 | 5 | 14 | 38 | 43 | 95 |
| | Room for improvement | | Patient's assessment of how much the intervention could be improved | 5 | 0 | 5 | 0 | 10 | 38 | 10 | 29 | 5 | 44 |
| | Would patient have the same operation again? | | Patient's assessment of how likely it is they would have the same intervention if they had the same problem in the future | 0 | 0 | 5 | 0 | 0 | 10 | 24 | 29 | 33 | 86 |
| | Patient assessment of outcome | | Patient's assessment of the outcome of the intervention | 0 | 0 | 0 | 0 | 0 | 0 | 24 | 24 | 52 | 100 |
| Nervous system | Pain | | Patient's assessment of their level of pain in general | 0 | 0 | 0 | 0 | 0 | 0 | 29 | 38 | 33 | 100 |
| | Pain at rest | | Patient's assessment of their level of pain when not active | 0 | 0 | 0 | 0 | 0 | 5 | 14 | 33 | 48 | 95 |
| | Pain on movement | | Patient's assessment of their level of pain when moving the arm | 0 | 0 | 0 | 0 | 0 | 5 | 38 | 29 | 29 | 96 |
| | Pain with activity | | Patient's assessment of their level of pain with activity | 0 | 0 | 0 | 0 | 5 | 5 | 29 | 33 | 29 | 91 |
| | Night pain | | Patient's assessment of their level of pain at night | 0 | 0 | 0 | 0 | 5 | 5 | 24 | 38 | 29 | 91 |
| | Average daily pain | | Patient's assessment of their average level of pain each day | 0 | 5 | 0 | 5 | 0 | 5 | 33 | 38 | 14 | 85 |
| | Pain at worst | | Patient's assessment of their level of pain when it is at its worst | 0 | 5 | 0 | 0 | 5 | 14 | 19 | 33 | 24 | 76 |
| | Tenderness | | Patient's experience of pain when the elbow is touched | 0 | 5 | 10 | 14 | 24 | 10 | 29 | 10 | 0 | 39 |
| | Pain medication used | | Measurement of the quantity of pain relieving drugs used by the patient | 0 | 5 | 0 | 0 | 19 | 24 | 24 | 14 | 14 | 52 |
| | Neurological status | | Physician assessment of the nerve function in the arm and hand | 0 | 0 | 0 | 5 | 5 | 0 | 33 | 33 | 24 | 90 |
| | Ulnar nerve function | | Physician assessement of the function of the ulnar nerve | 0 | 0 | 0 | 5 | 0 | 10 | 29 | 24 | 33 | 86 |
| | Radial nerve function | | Physician assessement of the function of the radial nerve | 5 | 0 | 0 | 5 | 0 | 15 | 24 | 33 | 19 | 76 |
| Hospital resource | Operation duration | | How long the operation takes in minutes | 5 | 14 | 14 | 14 | 19 | 10 | 0 | 14 | 10 | 24 |
| | Length of stay | | Number of days that the patient stays in hospital | 5 | 14 | 10 | 5 | 19 | 19 | 24 | 5 | 0 | 29 |
| | Re-admission | | Patient admitted to the same or another healthcare institution after discharge (excludes step-down care) | 5 | 10 | 0 | 5 | 10 | 20 | 24 | 19 | 10 | 53 |
| | Re-admission within 30 days | | Patient admitted to the same or another healthcare institution up to 30 days after discharge (excludes step-down care) | 0 | 0 | 0 | 5 | 10 | 18 | 23 | 45 | 0 | 68 |
| | Re-admission within 90 days | | Patient admitted to the same or another healthcare institution up to 90 days after discharge (excludes step-down care) | 0 | 0 | 0 | 0 | 18 | 45 | 27 | 0 | 9 | 36 |
| | Return to operating room | | A need to take the patient back to the operating room due to post operative problems | 0 | 5 | 5 | 10 | 10 | 5 | 24 | 24 | 19 | 67 |
| | Unplanned intubation | | A need to put a ventilation tube back in to the patient due to problems after surgery | 10 | 5 | 10 | 14 | 14 | 10 | 5 | 24 | 10 | 39 |
| | Healthcare utilisation | | The patient's total use of healthcare resources related to the episode of care | 10 | 5 | 5 | 5 | 29 | 10 | 20 | 10 | 10 | 40 |
| Need for further inte | Re-operation | Re-operation | The need for any additional surgery after the first intervention excluding surgery that involves the elbow replacement | 5 | 0 | 10 | 0 | 10 | 14 | 19 | 19 | 4 | 42 |
| | | Amputation | Removal of a limb - for example through upper arm to control infection in elbow | 8 | 0 | 0 | 0 | 0 | 25 | 0 | 17 | 50 | 67 |
| | | Arthrodesis | Attempt to fuse bones above and below elbow to make one continuous bone | 17 | 0 | 8 | 8 | 8 | 8 | 17 | 8 | 25 | 50 |
| | | Lateral ligament reconstruction | Rebuilding the soft tissue that holds the bones together at the elbow | 8 | 8 | 0 | 0 | 25 | 33 | 17 | 0 | 8 | 25 |
| | | Secondary wound suture | Using thread to close a wound if it has opened up after surgery | 8 | 0 | 8 | 17 | 17 | 8 | 25 | 17 | 0 | 42 |
| | | Ulnar nerve decompression | Release of the soft tissues around the nerve that gives feeling to the little and ring finger to prevent damage | 8 | 0 | 0 | 0 | 8 | 25 | 33 | 17 | 8 | 58 |
| | | Return to operating room | Taking a patient back to the operating theatre for any reason | 8 | 8 | 8 | 0 | 0 | 25 | 25 | 17 | 8 | 50 |
| | | Incision and drainage | Opening a wound in the operating theatre usually for treatment of infection or a collection of blood | 8 | 0 | 8 | 0 | 0 | 33 | 8 | 42 | 0 | 50 |
| | | Irrigation and debridement | Washing out of a woud in the operating theatre and removing dead tissue usually for the treatment of infection | 8 | 0 | 8 | 0 | 0 | 25 | 0 | 42 | 17 | 59 |
| | | Manipulation under anaesthesia | Trying to move the joint with the patient asleep in the operating theatre | 8 | 0 | 0 | 0 | 0 | 42 | 17 | 8 | 25 | 50 |
| | | Lysis of adhesions | Opening the joint to remove scar tissue causing a restriction of movement of the elbow | 8 | 0 | 0 | 0 | 0 | 50 | 8 | 25 | 8 | 41 |
| | Time to re-operation | | Time period between the first surgery to implant the elbow replacement and any additional surgery (needs specification of units of measurement (hours, days, weeks, months, years)) | 0 | 0 | 8 | 0 | 8 | 25 | 17 | 25 | 17 | 59 |
| | Revision | Revision surgery | Any further operation involving the elbow replacement. This includes adjustments, additions, exchanges or removal of all or part of elbow implant components | 0 | 0 | 0 | 0 | 5 | 0 | 15 | 25 | 55 | 95 |
| | | Reason for revision | The explanation for why the implant has to be altered or removed | 0 | 0 | 0 | 0 | 11 | 11 | 21 | 16 | 42 | 79 |
| | | Implant removal | Taking the implant out of the patient | 0 | 0 | 0 | 5 | 11 | 5 | 11 | 21 | 47 | 79 |
| | | Implant survival | Whether the elbow replacement is still in the patient | 0 | 0 | 0 | 5 | 0 | 16 | 21 | 5 | 53 | 79 |
| | | Implant exchange | Cha ging of all or part of the elbow replacement | 0 | 0 | 0 | 11 | 0 | 16 | 11 | 21 | 42 | 74 |
| | | Implant survival - aseptic loosening | A measure of the number of elbow replacements that have not been removed or altered for coming loose from the bone without infection over time | 0 | 0 | 0 | 0 | 0 | 5 | 42 | 11 | 42 | 95 |
| | | Implant survival - Revision mechanical failure | A measure of the number of elbow replacements that have not been removed or altered because of something going wrong with the implant itself (such as fracture, component wear or loosening of the components) over time | 5 | 0 | 0 | 0 | 0 | 5 | 26 | 26 | 37 | 89 |
| | | Implant survival - Revision mechanical failure or loosening | A measure of the number of elbow replacements that have not been removed or altered because of something going wrong with the implant itself (such as fracture or component wear) or for coming loose from the bone for any reason over time | 0 | 0 | 5 | 0 | 5 | 5 | 37 | 11 | 37 | 85 |
| | | Implant survival - loosening | A measure of the number of elbow replacements that have not been removed or altered for coming loose from the bone for any reason over time | 5 | 5 | 0 | 0 | 0 | 16 | 26 | 5 | 42 | 73 |
| | | Implant survival - Revision | A measure of the number of elbow replacements that have not been removed or altered for any reason over time | 5 | 5 | 0 | 0 | 0 | 5 | 32 | 11 | 42 | 85 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Implant survival - definite loosening and bushing wear | A measure of the number of elbow replacements that have not been removed or altered for clear evidence of coming loose from the bone for any reason or wear of the plastic components over time | 5 | 0 | 0 | 0 | 0 | 21 | 32 | 5 | 37 | 74 |
| | | Implant survival - possible loosening and bushing wear | A measure of the number of elbow replacements that have not been removed or altered for a suspicion of coming loose from the bone for any reason or wear of the plastic or metal components over time | 5 | 0 | 0 | 5 | 5 | 11 | 42 | 11 | 21 | 74 |
| | | Implant survival - revision and loosening | A measure of the number of elbow replacements that have not been removed or altered for any reason for coming loose from the bone for any reason over time | 5 | 0 | 0 | 0 | 0 | 11 | 32 | 16 | 37 | 85 |
| | | Cumulative success rate | An alternative description of implant survival | 0 | 0 | 5 | 0 | 5 | 5 | 11 | 47 | 26 | 84 |
| MSK connective tissu | Stability | Stability | The ability of the elbow joint to maintain a fixed position with a force applied . No clinical evidence of joint dislocation or subluxation. Excludes radiographic evidence. | 5 | 0 | 0 | 0 | 10 | 0 | 30 | 40 | 15 | 85 |
| | | Stability on varus/valgus stress | The ability of the elbow joint to maintain a fixed position with a force applied inwards or outwards. No clinical evidence of joint dislocation or subluxation. Excludes radiographic evidence. | 6 | 0 | 0 | 18 | 6 | 18 | 18 | 18 | 18 | 54 |
| | | Instability - subjective | The patient's own perception of the presence of the elbow not maintaining correct position in rest or during movement | 6 | 0 | 0 | 6 | 12 | 12 | 35 | 12 | 18 | 65 |
| | | Stability of prosthesis in axial compression | The ability of the elbow joint to maintain a fixed position with a force applied through forearm towards elbow. No clinical evidence of joint dislocation or subluxation. Excludes radiographic evidence. | 12 | 12 | 6 | 6 | 18 | 6 | 30 | 6 | 6 | 42 |
| | Radiographic appearance | Radiographic appearance | Changes seen on plain X-ray images of the elbow | 0 | 5 | 5 | 5 | 10 | 20 | 25 | 15 | 15 | 55 |
| | | Radiographic alignment | Radiographic assessement of the alignment of the implant with the bone | | | | | | | | | | 0 |
| | | Radiographic fixation | Fixation method (for example cementing) and quality of fixation seen in radiographs (No radiographic evidence of loosening of the implant) | 0 | 0 | 0 | 0 | 0 | 9 | 27 | 36 | 27 | 90 |
| | | Radiographic success | No apparent radiographic problem identified | 9 | 0 | 0 | 0 | 9 | 9 | 27 | 45 | 0 | 72 |
| | | Radiographic lengthening | Radiographic evidence that the implant has been positioned in a way to lengthen the segment | 18 | 0 | 0 | 18 | 0 | 27 | 18 | 0 | 18 | 36 |
| | | Radiographic head size | Radiographic assessment of the size of the radial head | | | | | | | | | | 0 |
| | | Radiographic congruence PRUJ | Radiographic assessement of the way the radial head fits into the socket on the ulna | 18 | 0 | 0 | 9 | 18 | 18 | 18 | 9 | 9 | 36 |
| | | Radiographic stem size | Radiographic assessment of the size of the stem of a prosthesis | 9 | 0 | 18 | 18 | 36 | 9 | 0 | 0 | 9 | 9 |
| | | Radiographic stem positioning | Radiographic assessment of the position of an implant within the bone | 0 | 0 | 0 | 0 | 9 | 18 | 27 | 27 | 18 | 72 |
| | | Radiographic quality of cementing technique | Radiographic assessment of the quality of the cementation between the implant and the bone | 0 | 0 | 0 | 0 | 0 | 9 | 27 | 36 | 27 | 90 |
| | | Radiographic implant positioning | Radiographic assessement of the position of the implant within the bone | 9 | 0 | 0 | 0 | 0 | 0 | 36 | 36 | 18 | 90 |
| | | Radiographic prosthetic sizing | Radiographic assessment of the size of the implant | 9 | 0 | 9 | 18 | 9 | 45 | 0 | 9 | 0 | 9 |
| | | Radiographic bone graft integration | Radiographic assessment of the degree to which any bone graft has attached to the host bone | 0 | 0 | 18 | 9 | 0 | 0 | 18 | 36 | 18 | 72 |
| | | Radiographic cortical fit | Radiographic assessment of the space between the implant and the bone | 9 | 0 | 0 | 9 | 18 | 27 | 27 | 9 | 0 | 36 |
| | | Radiographic valgus tilting | Radiographic assessment of the alignment of the implant within the bone | 9 | 9 | 0 | 9 | 45 | 9 | 0 | 9 | 9 | 18 |
| | | Radiographic joint congruity | Radiographic assessment of the conformity of the two sides of the joint | 0 | 0 | 0 | 0 | 9 | 9 | 9 | 36 | 36 | 81 |
| | Other | Synovitis | Inflammation of the joint (lining) | 5 | 15 | 5 | 20 | 20 | 20 | 10 | 0 | 5 | 15 |
| | | Deformity | Change in the appearance of the elbow | 5 | 15 | 5 | 25 | 15 | 5 | 15 | 10 | 5 | 30 |
| | | Cubitus valgus | Alignment of the forearm relative to the arm with the elbow in a fully straightened position | 10 | 5 | 10 | 20 | 10 | 20 | 10 | 10 | 5 | 25 |
| | | Swelling | Increase in the size of the elbow | 15 | 10 | 0 | 10 | 25 | 20 | 10 | 10 | 0 | 20 |
| | | Joint position sense | The ability to determine how bent the elbow is without looking at it | 10 | 0 | 5 | 15 | 30 | 20 | 10 | 5 | 5 | 20 |
| | | Deficient bone stock | The amount of bone missing around the elbow | 5 | 0 | 5 | 15 | 10 | 10 | 25 | 20 | 10 | 55 |
| | | MCL healing | A measure of the amount of healing of the medial ligament of the elbow | 10 | 5 | 25 | 30 | 10 | 10 | 5 | 0 | 5 | 10 |
| | | Implant size | The choice of size of the implant used | 15 | 10 | 10 | 20 | 5 | 10 | 15 | 10 | 5 | 30 |
| Physical function | Function/disability | Function | A measure of the use of the arm or impairment of use | 0 | 0 | 0 | 0 | 5 | 10 | 15 | 50 | 20 | 85 |
| | | Disability | | 0 | 0 | 0 | 0 | 10 | 10 | 40 | 25 | 15 | 80 |
| | Range of movement | Range of movement | A measure of the amount of movement of the elbow or forearm | 0 | 0 | 0 | 0 | 10 | 20 | 10 | 40 | 20 | 70 |
| | | Range of pronosupination | A measure of the amount of rotational movement of the forearm from palm up to palm down | 0 | 0 | 0 | 0 | 0 | 7 | 20 | 40 | 33 | 93 |
| | | Range of flexion/extension | A measure of the amount of movement of the elbow from straight to bent | 0 | 0 | 0 | 0 | 0 | 13 | 27 | 33 | 27 | 87 |
| | | Range of flexion/extension (active) | A measure of the amount of movement of the elbow from straight to bent by the patient only | 0 | 0 | 0 | 0 | 0 | 7 | 27 | 33 | 33 | 93 |
| | | Range of flexion/extension (passive) | A measure of the amount of movement of the elbow from straight to bent with a clinician helping to move the limb | 7 | 0 | 7 | 7 | 13 | 7 | 20 | 7 | 33 | 60 |
| | | Range of pronosupination (active) | A measure of the amount of rotational movement of the forearm from palm up to palm down by the patient only | 0 | 0 | 0 | 7 | 0 | 7 | 27 | 33 | 27 | 87 |
| | | Range of pronosupination (passive) | A measure of the amount of movement of the forearm from palm up to palm down with a clinician helping to move the limb | 7 | 0 | 7 | 20 | 0 | 13 | 13 | 13 | 27 | 53 |
| | Strength | Strength | A measure of the power of the muscles around the elbow | 0 | 5 | 0 | 5 | 15 | 20 | 45 | 5 | 5 | 55 |
| | | Grip strength | A measure of the power of the muscles to grip an object in the hand | 0 | 0 | 0 | 8 | 17 | 8 | 50 | 0 | 17 | 67 |
| | | Flexion Strength | A measure of the power of the muscles around the elbow to move the elbow from straight to bent | 0 | 0 | 17 | 8 | 25 | 0 | 33 | 8 | 8 | 49 |
| | | Extension strength | A measure of the power of the muscles around the elbow to move the elbow from bent to straight | 0 | 8 | 8 | 8 | 17 | 8 | 17 | 25 | 8 | 50 |
| | | Pronation strength | A measure of the power of the muscles around the elbow to move the forearm from palm up to palm down | 0 | 8 | 8 | 25 | 8 | 25 | 17 | 0 | 8 | 25 |
| | | Supination strength | A measure of the power of the muscles around the elbow to move the forearm from palm down to palm up | 0 | 8 | 8 | 17 | 0 | 33 | 25 | 0 | 8 | 33 |
| | | Maximum grip strength | A measure of the maximum power of the muscles to grip an object in the hand | 0 | 0 | 8 | 8 | 25 | 17 | 25 | 8 | 8 | 41 |
| | | Pinch strength | A measure of the power of the muscles to pinch an object between finger and thumb | 0 | 42 | 8 | 0 | 8 | 17 | 0 | 25 | 0 | 25 |
| | | Isometric flexion strength | A measure of the power of the muscles around the elbow when the elbow is held in a fixed position and the patient tries to move the elbow from straight to bent | 0 | 8 | 17 | 17 | 17 | 8 | 25 | 0 | 8 | 33 |
| | | Isometric extension strength | A measure of the power of the muscles around the elbow when the elbow is held in a fixed position and the patient tries to move the elbow from bent to straight | 0 | 8 | 8 | 25 | 8 | 8 | 25 | 8 | 8 | 41 |
| | | isometric pronation strength | A measure of the power of the muscles around the elbow when the elbow is held in a fixed position and the patient tries to move the forearm from palm up to palm down | 0 | 8 | 17 | 17 | 25 | 17 | 8 | 0 | 8 | 16 |
| | | Isometric supination strength | A measure of the power of the muscles around the elbow when the elbow is held in a fixed position and the patient tries to move the forearm from palm down to palm up | 0 | 8 | 17 | 8 | 17 | 25 | 17 | 0 | 8 | 25 |
| | | Triceps strength | A measure of the power of the triceps muscle to move the elbow from bent to straight | 0 | 8 | 8 | 8 | 17 | 25 | 25 | 8 | 0 | 33 |
| | Activities of daily living | Activities of daily living | A measure of the ability to do everyday tasks | 0 | 0 | 0 | 5 | 0 | 5 | 20 | 35 | 35 | 90 |
| | | Autonomy on ADL | A measure of the patients independence for everyday tasks | 0 | 0 | 0 | 0 | 11 | 16 | 5 | 21 | 47 | 73 |
| | | Self-care activities | A measure of the ability of the patient to do personal care such as hygiene | 5 | 0 | 0 | 0 | 11 | 5 | 11 | 32 | 37 | 80 |
| | | Housekeeping ability | A measure of the ability of the patient to do household activities such as cleaning | 0 | 0 | 5 | 5 | 5 | 21 | 21 | 26 | 17 | 64 |
| | Locomotion | Ambulation | A measure of the patients ability to move around | 0 | 10 | 5 | 10 | 10 | 30 | 20 | 5 | 10 | 35 |
| Quality of life | Health related quality of life | Health related quality of life | A measure of the patients own perception of the impact of their health on their life | 0 | 0 | 0 | 5 | 0 | 25 | 20 | 25 | 25 | 70 |
| Adverse events | Adverse event | All adverse events (no time limit) | A measure of complications occuring after surgery | 0 | 0 | 0 | 10 | 0 | 5 | 10 | 35 | 40 | 85 |
| | | 30-day adverse event | A measure of complications occuring within 30 days following surgery | 0 | 0 | 0 | 0 | 6 | 17 | 11 | 33 | 33 | 77 |
| | | intra-operative complication | A measure of complications occuring during an operation | 6 | 0 | 0 | 0 | 6 | 1 | 22 | 17 | 39 | 78 |
| | | peri-operative complication | A measure of complications occuring around the time of an operation | 0 | 0 | 0 | 0 | 6 | 17 | 22 | 39 | 17 | 78 |
| | | post-operative complication | A measure of complications occuring after an operation | 0 | 0 | 0 | 0 | 6 | 6 | 17 | 44 | 28 | 89 |
| | | Major complications | A measure of serious adverse events or complications of surgery such as death or those needing additional surgery | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 11 | 72 | 100 |

# Appendix F Assessment of systematic reviews using AMSTAR-2 criteria.

(Shea et al., 2017)

| AMSTAR 2 | Research Q and Inclusion PICO Q1 | Review methods established before review and deviation justified? Q2* | Explanation of study designs included Q3 | Comprehensive search strategy Q4* | Selection in duplicate Q5 | Data extraction in duplicate Q6 | Excluded studies list and justification Q7* | Adequate description included studies Q8 | RoB adequate Q9* | Sources of funding for studies Q10 | If meta-analysis - appropriate stats Q11* | Impact of RoB on meta-analysis Q12 | Impact of RoB on interpretation Q13* | Explanation for heterogeneity Q14 | Publication bias Q15* | Conflicts of interets Q16 | RATING |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| STODDART | NO | YES | YES | NO | YES | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | YES | CRITICALLY LOW |
| EVANS | NO | YES | NO | NO | YES | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | YES | CRITICALLY LOW |
| NIELSEN | NO | NO | NO | NO | YES | YES | NO | PARTIAL | PARTIAL | NO | NA | NA | YES | NO | NA | YES | CRITICALLY LOW |

# Appendix G Prisma checklist for systematic review

**PRISMA 2020 Checklist**

| Section and Topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| **TITLE** | | | |
| Title | 1 | Identify the report as a systematic review. | Page 1 |
| **ABSTRACT** | | | |
| Abstract | 2 | See the PRISMA 2020 for Abstracts checklist. | Page 2 |
| **INTRODUCTION** | | | |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge. | Page 4-5 |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | Page 5-6 |
| **METHODS** | | | |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | Page 6-7 |
| Information sources | 6 | Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | Page 8-9 |
| Search strategy | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | Page 9-10 |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process. | Page 10 |
| Data collection process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process. | Page 11 |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | Page 11 |
| | 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information. | Page 12 |
| Study risk of bias assessment | 11 | Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process. | Page 12 |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | Page 12 |
| Synthesis methods | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)). | Page 13 |
| | 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | Page 13-14 |
| | 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | Page 13-14 |
| | 13d | Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | Page 13-14 |
| | 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | NA |
| | 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesized results. | Page 20 |
| Reporting bias assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | Page 16 |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | Page 14 |

## PRISMA 2020 Checklist

| Section and Topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| **TITLE** | | | |
| Title | 1 | Identify the report as a systematic review. | Page 1 |
| **ABSTRACT** | | | |
| Abstract | 2 | See the PRISMA 2020 for Abstracts checklist. | Page 2 |
| **INTRODUCTION** | | | |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge. | Page 4-5 |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | Page 5-6 |
| **METHODS** | | | |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | Page 6-7 |
| Information sources | 6 | Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | Page 8-9 |
| Search strategy | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | Page 9-10 |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process. | Page 10 |
| Data collection process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process. | Page 11 |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | Page 11 |
| | 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information. | Page 12 |
| Study risk of bias assessment | 11 | Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process. | Page 12 |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | Page 12 |
| Synthesis methods | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)). | Page 13 |
| | 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | Page 13-14 |
| | 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | Page 13-14 |
| | 13d | Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | Page 13-14 |
| | 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | NA |
| | 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesized results. | Page 20 |
| Reporting bias assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | Page 16 |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | Page 14 |

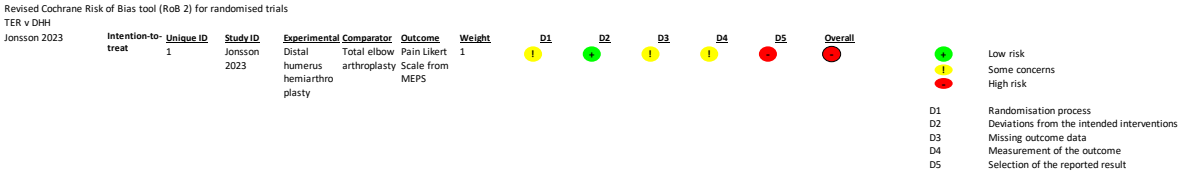
# Appendix H Screened titles for systematic review of pain outcomes

| authors | title | journal | year | volume | issue | pages |
|---|---|---|---|---|---|---|
| Amin, A. and Suresh, S. and Sanghrajka, A. and Cannon, S. R. and Briggs, T. W. R. and Unwin, P. | Custom-made endoprosthetic reconstruction of the distal humerus for non-tumorous pathology | Acta Orthopaedica Belgica | 2008 | 74 | 4 | 446-450 |
| Argintar, Evan and Berry, Micah and Narvy, Steven J. and Kramer, Jonathan and Omid, Reza and Itamura, John M. | Hemiarthroplasty for the treatment of distal humerus fractures: short-term clinical results | Orthopedics | 2012 | 35 | 12 | 1042-5 |
| Barthel, P. Y. and Mansat, P. and Sirveaux, F. and Dap, F. and Mole, D. and Dautel, G. | Is total elbow arthroplasty indicated in the treatment of traumatic sequelae? 19 cases of Coonrad-Morrey reviewed at a mean follow-up of 5.2years | Orthopaedics and Traumatology: Surgery and Research | 2014 | 100 | 1 | 113-118 |
| Bhashyam, Abhiram R. and Jupiter, Jesse B. | Patients with Poor Bone Quality or Bone Loss - Is This Viable as a Long-term Treatment Option? | The archives of bone and joint surgery | 2019 | 7 | 3 | 251-257 |
| Bornu, B. Chedal and Clement, X. and Kempf, J. F. and Clavert, P. | Arthroscopic elbow joint release with radial head resection arthroplasty: 12 cases | Orthopaedics & traumatology, surgery & research : OTSR | 2015 | 101 | 6 | 735-9 |
| Burden, Eleanor Grace and Batten, Timothy and Smith, Christopher and Evans, Jonathan P. | Hemiarthroplasty or total elbow arthroplasty for unreconstructable distal humeral fractures in patients aged over 65 years : a systematic review and meta-analysis of patient outcomes and complications | The bone & joint journal | 2022 | 104 | 5 | 559-566 |
| Celli, A. and Deluise, G. and Celli, L. | Patients with Osteoarthritis Secondary to Fracture: Clinical Experience | Journal of Shoulder and Elbow Surgery | 2021 | 30 | 7 | e441 |
| Chantelot, C. and Feugas, C. and Ala Eddine, T. and Migaud, H. and Gueguen, G. and Fontaine, C. | Kudo non-constrained elbow prosthesis for inflammatory joint disease: Analysis in 30 cases | Revue de Chirurgie Orthopedique et Reparatrice de l'Appareil Moteur | 2002 | 88 | 4 | 398-405 |
| Chen, Chen and Jiang, Xie-Yuan and Gong, Mao-Qi | [Review and selection of the approach of total elbow arthroplasty] | Zhongguo gu shang = China journal of orthopaedics and traumatology | 2014 | 27 | 1 | 79-84 |
| Cottias, Pascal and Leclerc, Philippe and Zaoui, Amine and Abouchaaya, Abdel Massih and Khallouk, Raphiq and Anract, Philippe | Digastric olecranon osteotomy a new approach to the elbow: retrospective study of 24 Coonrad-Morrey R total elbow arthroplasty at 30-month follow-up | European journal of orthopaedic surgery & traumatology : orthopedie traumatologie | 2020 | 30 | 3 | 485-491 |
| DeFroda, Steven and Henry, Havalee and Paxton, E. Scott | Treatment of a Complex Interprosthetic Humerus Fracture | Journal of orthopaedic trauma | 2019 | 33 | | S3-S4 |
| Desai, Sagar J. and Lalone, Emily and Athwal, George S. and Ferreira, Louis M. and Johnson, James A. and King, Graham J. W. | Hemiarthroplasty of the elbow: the effect of implant size on joint congruency | Journal of shoulder and elbow surgery | 2016 | 25 | 2 | 297-303 |
| Ding, Jian and Yin, Wen-Jing and Zhang, Bao-Kun and Yu, Xin-Gang and Ruan, Hong-Jiang and Zhang, Wei | Bilateral triceps tendon approach is flexible and efficient in the treatment of type C distal humerus fractures | Chinese journal of traumatology = Zhonghua chuang shang za zhi | 2022 | 25 | 3 | 145-150 |
| Duparc, Fabrice and Merlet, Marie-Caroline | Prevention and management of early treatment failures in elbow injuries | Orthopaedics & traumatology, surgery & research : OTSR | 2019 | 105 | 1 | S75-S87 |
| Evans, P. J. and Regal, S. | Coronoid Impingement Causing Early Failure of Total Elbow Arthroplasty | Journal of Hand Surgery Global Online | 2020 | 2 | 5 | 312-315 |
| Fang, Christian and Yan, Chun-Hoi and Yee, Dennis and Lau, Tak-Wing and Wong, Tak-Man and Leung, Frankie | of a Periprosthetic Fracture with a Loose Stem: A Report of Two Cases | JBJS case connector | 2017 | 7 | 1 | e17 |
| Fevang, Bjorg-Tilde S. and Lie, Stein A. and Havelin, Leif I. and Skredderstuen, Arne and Furnes, Ove | Results after 562 total elbow replacements: a report from the Norwegian Arthroplasty Register | Journal of shoulder and elbow surgery | 2009 | 18 | 3 | 449-56 |
| Frankle, Mark A. and Herscovici, Dolfi, Jr. and DiPasquale, Thomas G. and Vasey, Matthew B. and S and ers, Roy W. | total elbow arthroplasty in the treatment of intraarticular distal humerus fractures in women older than age 65 | Journal of orthopaedic trauma | 2003 | 17 | 7 | 473-80 |
| Fritsche, C. C. and Deml, O. and Grosstuck, R. and Hofmann, G. O. | [Short- and Medium-Term Results of Total Elbow Arthroplasty after Trauma] | Kurz- und mittelfristige Ergebnisse der Ellenbogenarthroplastik nach Trauma. | 2015 | 153 | 3 | 267-76 |
| Fu, Q. Z. and Li, F. B. and Xie, J. H. and Huang, G. | Treatment of elbow joint ankylosis by repair of articular surface with periosteal autograft | Chinese journal of reparative and reconstructive surgery | 2002 | 16 | 4 | 235-236 |
| Gallucci, G. L. and Rellan, I. and Boretto, J. G. and Donndorff, A. G. and De Carli, P. | Total Elbow Arthroplasty in the Context of an Olecranon Nonunion; Surgical Technique and Report of Three Cases | Archives of Bone and Joint Surgery | 2022 | 10 | 6 | 525-529 |
| Gambirasio, R. and Ri and , N. and Stern, R. and Hoffmeyer, P. | Total elbow replacement for complex fractures of the distal humerus. An option for the elderly patient | The Journal of bone and joint surgery. British volume | 2001 | 83 | 7 | 974-8 |
| Gao, Yu and Zhang, Wei and Duan, Xin and Yang, Jing and Al-Qwbani, Mohammed and Lv, Jingtong and Xiang, Zhou | Surgical interventions for treating radial head fractures in adults | The Cochrane database of systematic reviews | 2013 | | 5 | CD008987 |
| Giannicola, G. and Calella, P. and Spinello, P. and Cinotti, G. and Bigazzi, P. and Mantovani, A. | Midterm results of radiocapitellar arthroplasty of the elbow: A multicentre prospective study on two different implants | Bone and Joint Journal | 2019 | 101 | 11 | 1362-1369 |
| Hastings, Hill, 2nd and Theng, Chee S. | Total elbow replacement for distal humerus fractures and traumatic deformity: results and complications of semiconstrained implants and design rationale for the Discovery Elbow System | American journal of orthopedics (Belle Mead, N.J.) | 2003 | 32 | 9 | 20-Aug |
| Hildebr and , K. A. and Patterson, S. D. and Regan, W. D. and MacDermid, J. C. and King, G. J. W. | Functional outcome of semiconstrained total elbow arthroplasty | Journal of Bone and Joint Surgery | 2000 | 82 | 10 | 1379-1386 |
| Hu, B. and Liu, X. W. and Huang, J. J. | Surgical treatment for distal humerus type C fractures | China journal of orthopaedics and traumatology | 2018 | 31 | 10 | 976-982 |
| Hua, K. H. and Zha, Y. J. and Chen, C. and Lu, S. and Sun, W. T. and Gong, M. Q. and Jiang, X. Y. | Progress on diagnosis and treatment of low transcondylar fractures of distal humerus | China journal of orthopaedics and traumatology | 2019 | 32 | 8 | 774-789 |
| Jeon, In-Ho and Morrey, Bernard F. and Sanchez-Sotelo, Joaquin | Coonrad-Morrey total elbow arthroplasty | Journal of shoulder and elbow surgery | 2012 | 21 | 9 | 1229-35 |
| Jiang, Bao-guo and Chen, Jian-hai and Zhang, Pei-xun and Zhang, Dian-ying and Fu, Zhong-guo | [Clinical outcomes of total elbow replacement in the treatment of complex distal humeral fractures] | [Chinese journal of surgery] | 2010 | 48 | 3 | 213-6 |
| Jiang, Xie-Yuan and Gong, Mao-Qi and Liu, Xing-Hua and He, Liang and Zhang, Li-Dan and Wang, Man-Yi and Rong, Guo-Wei and Zhang, Hong | [Semi-constrained total elbow arthroplasty for the treatment of the elbow disorders] | [Chinese journal of surgery] | 2009 | 47 | 12 | 884-7 |
| Jile, J. and Gong, M. and Zha, Y. and Liu, X. and Ting, L. and Zhang, L. and Jiang, X. | Evaluation of total elbow arthroplasty in treatment of distal humeral fracture in the elderly | National Medical Journal of China | 2015 | 95 | 47 | 3848-3851 |
| Kamineni, S. and Morrey, Bernard F. | Distal humeral fractures treated with noncustom total elbow replacement | The Journal of bone and joint surgery. American volume | 2004 | 86 | 5 | 940-7 |
| Kholinne, E. and Altamimi, L. A. and Aldayel, A. and Alsabti, R. and Kim, H. and Park, D. and Koh, K. H. and Jeon, I. H. | Primary linked total elbow arthroplasty for acute distal humerus fracture management: A systematic review of clinical outcome | CiOS Clinics in Orthopedic Surgery | 2020 | 12 | 4 | 503-513 |
| Kholinne, Erica and Arya, An and Jeon, In-Ho | Complications of modern design total elbow replacement | trauma | 2021 | 19 | | 42-49 |
| Kieser, David Christopher and Taylor, Fraser and Ball, Craig M. | interprosthetic fracture between a shoulder and elbow joint replacement | Journal of shoulder and elbow surgery | 2011 | 20 | 3 | e4-9 |
| Kincaid, Brian L. and An, Kai-Nan | prostheses | Journal of biomechanics | 2013 | 46 | 14 | 2331-41 |
| Kozak, T. K. W. and Adams, R. A. and Morrey, B. F. | Total elbow arthroplasty in primary osteoarthritis of the elbow | Journal of Arthroplasty | 1998 | 13 | 7 | 837-842 |
| Kumar, Suresh and Mahanta, Sunayan | Primary total elbow arthroplasty | Indian journal of orthopaedics | 2013 | 47 | 6 | 608-14 |
| Kwak, Jae-Man and Kholinne, Erica and Sun, Yucheng and Kim, Myung-Sun and Koh, Kyoung-Hwan and Jeon, In-Ho | Clinical results of revision total elbow arthroplasty: comparison of infected and non-infected total elbow arthroplasty | International orthopaedics | 2019 | 43 | 6 | 1421-1427 |
| Kwak, Jae-Man and Koh, Kyoung-Hwan and Jeon, In-Ho | Revision Surgery | Clinics in orthopedic surgery | 2019 | 11 | 4 | 369-379 |
| L and or, I. and Vavrik, P. and Jahoda, D. and Guttler, K. and Sosna, A. | Total elbow replacement with the Souter-Strathclyde prosthesis in rheumatoid arthritis. Long-term follow-up | The Journal of bone and joint surgery. British volume | 2006 | 88 | 11 | 1460-3 |
| Lapner, M. and King, G. J. | Elbow arthroplasty for distal humeral fractures | Instructional course lectures | 2014 | 63 | | 15-26 |
| Lapner, Michael and King, Graham J. W. | Radial head fractures | The Journal of bone and joint surgery. American volume | 2013 | 95 | 12 | 1136-43 |
| Lauder, Alex and er and Richard, Marc J. | Management of distal humerus fractures | & traumatology : orthopedie traumatologie | 2020 | 30 | 5 | 745-762 |
| Lausten, Gunnar Schwarz and Jensen, Claus Munk and Olsen, Bo S and erhoff | [Elbow arthroplasty] | Albueartroplastik. | 2006 | 168 | 19 | 1844-7 |
| Lechasseur, Benoit and Laflamme, Melissa and Leclerc, Alex and re and Bedard, Anne-Marie | Incipient malunion of an isolated humeral trochlea fracture treated with an elbow hemiarthroplasty: case report | The Journal of hand surgery | 2015 | 40 | 2 | 271-5 |
| Lee, John J. and Lawton, Jeffrey N. | Coronal shear fractures of the distal humerus | The Journal of hand surgery | 2012 | 37 | 11 | 2412-7 |
| Ma, Ching-Hou and Hsueh, Yu-Huan and Wu, Chin-Hsien and Yen, Cheng-Yo and Tu, Yuan-Kun | Does an Internal Joint Stabilizer and Standardized Protocol Prevent Recurrent Instability in Complex Persistent Elbow Instability? | Clinical orthopaedics and related research | 2022 | 480 | 7 | 1354-1370 |
| Maniscalco, P. and Pizzoli, A. L. and Brivio, L. R. and Caforio, M. | Hinged external fixation for complex fracture-dislocation of the elbow in elderly people | Injury | 2014 | 45 | | S53-S57 |
| Mannan, Syed and Ali, Mohammed and Mazur, Lukasz and Chin, Mei and Fadulelmola, Ahmed | The use of Tranexamic Acid in Total Elbow Replacement to Reduce Post-Operative Wound Infection | Journal of bone and joint infection | 2018 | 3 | 2 | 104-107 |
| and Schemitsch, Emil H. and Wild, Lisa M. and McCormack, Robert and Perey, Bertr and Goetz, Thomas and Zomar, Mauri and Moon, Karyn and M and el, Scott and Petit, Shirlet and Guy, Pierre and Leung, Irene | A multicenter, prospective, randomized, controlled trial of open reduction--internal fixation versus total elbow arthroplasty for displaced intra-articular distal humeral fractures in elderly patients | Journal of shoulder and elbow surgery | 2009 | 18 | 1 | 03-Dec |

| | | | | | | |
|---|---|---|---|---|---|---|
| Moon, Jun-Gyu and Kim, Jung-Hoon and Jung, Young-Jin and Lim, Moo-Joon and Lee, Hee-Dong | Relationship between the tilt angle of bipolar radial head prostheses and radiological radiocapitellar instability | Acta orthopaedica et traumatologica turcica | 2020 | 54 | 4 | 372-377 |
| Moro, J. K. and King, G. J. | Total elbow arthroplasty in the treatment of posttraumatic conditions of the elbow | Clinical orthopaedics and related research | 2000 | | 370 | 102-14 |
| Na, Ki-Tae and Song, Seok-Whan and Lee, Yoon-Min and Choi, Jae-Hoon | Modified triceps fascial tongue approach for primary total elbow arthroplasty | Journal of shoulder and elbow surgery | 2018 | 27 | 5 | 887-893 |
| O'Driscoll, S. W. and King, G. J. | Treatment of instability after total elbow arthroplasty | The Orthopedic clinics of North America | 2001 | 32 | 4 | 679-ix |
| O'Driscoll, S. W. and Spinner, R. J. and McKee, M. D. and Ben Kibler, W. and Hastings, Ii H. and Morrey, B. F. and Kato, H. and Takayama, S. and Imatani, J. and Toh, S. and Kerr Graham, H. | Tardy posterolateral rotatory instability of the elbow due to cubitus varus | Journal of Bone and Joint Surgery | 2001 | 83 | 9 | 1358-1369 |
| Page, Brian Joseph and Lee Brennan, Michael | Total Elbow Arthroplasty for Distal Humerus Fractures | Journal of orthopaedic trauma | 2020 | 34 | | S7-S8 |
| Pogliacomi, F. and Galavotti, C. and Cavaciocchi, M. and Corradi, M. and Rotini, R. and Ceccarelli, F. | Total elbow arthroplasty following traumas: Mid-term results | Acta Biomedica | 2013 | 84 | 3 | 212-218 |
| Pogliacomi, F. and Schiavi, P. and Defilippo, M. and Calderazzi, F. and Corradi, M. and Vaienti, E. and Ceccarelli, F. and Rotini, R. | Total elbow arthroplasty following complex fractures of the distal humerus: Results in patients over 65 years of age | Acta Biomedica | 2016 | 87 | 2 | 148-155 |
| Pooley, Joseph and Salvador Carreno, Jordi | Total elbow joint replacement for fractures in the elderly--Functional and radiological outcomes | Injury | 2015 | 46 | | S37-42 |
| Prasad, N. and Ali, A. and Stanley, D. | Total elbow arthroplasty for non-rheumatoid patients with a fracture of the distal humerus: a minimum ten-year follow-up | The bone & joint journal | 2016 | 98 | 3 | 381-6 |
| Proust, J. and Oksman, A. and Charissoux, J. L. and Mabit, C. and Arnaud, J. P. | [Intra-articular fracture of the distal humerus: outcome after osteosynthesis in patients over 60] | Resultats de l'osteosynthese des fractures articulaires de la palette humerale chez le sujet age | 2007 | 93 | 8 | 798-806 |
| Pugh, D. M. and McKee, M. D. | Advances in the management of humeral nonunion | The Journal of the American Academy of Orthopaedic Surgeons | 2003 | 11 | 1 | 48-59 |
| Ramesh, Muthu and Foead, Agus Iwan and Ali, Anuar Bin and Devadasan, Benard | Salvage of elbow function in chronic complex elbow fracture dislocation with total elbow arthroplasty: a case report | The Medical journal of Malaysia | 2013 | 68 | 4 | 353-5 |
| Ray, P. S. and Kakarlapudi, K. and Rajsekhar, C. and Bhamra, M. S. | Total elbow arthroplasty as primary treatment for distal humeral fractures in elderly patients | Injury | 2000 | 31 | 9 | 687-92 |
| Sanchez-Sotelo, Joaquin and Morrey, Bernard F. | Linked elbow replacement: a salvage procedure for distal humeral nonunion. Surgical technique | The Journal of bone and joint surgery. American volume | 2009 | 91 | | 200-12 |
| Schiavi, P. and Pogliacomi, F. and Garzia, A. and Valenti, P. and Ceccarelli, F. and Calderazzi, F. | Survival and outcome of total elbow arthroplasty for distal humeral fracture at long-term follow-up | Acta Biomedica | 2020 | 91 | | 01-Sep |
| Schmidt-Horlohe, K. and Buschbeck, S. and Wincheringer, D. and Weissenberger, M. and Hoffmann, R. | Primary radial head arthroplasty in trauma : Complications | Der Orthopade | 2016 | 45 | 10 | 853-860 |
| Seok, Hyun-Gyu and Park, Jeong-Jin and Park, Sam-Guk | Comparison of the Complications, Reoperations, and Clinical Outcomes between Open Reduction and Internal Fixation and Total Elbow Arthroplasty for Distal Humeral Fractures in the Elderly: A | Journal of clinical medicine | 2022 | 11 | 19 | |
| Shafer, Brian L. and Fehringer, Edward V. and Boorman, Richard S. and Churchill, R. Sean and Matsen, Frederick A., 3rd | Ulnar component fracture after revision total elbow arthroplasty with proximal ulnar bone loss: a report of 2 cases | Journal of shoulder and elbow surgery | 2003 | 12 | 3 | 297-301 |
| Staritz, P. and Zimmermann, R. and Lages, P. and Huth-Kuehne, A. | Long term outcome of total elbow endoprosthesis in hemophilia arthropathy may be superior to radial head prosthesis: Some case reports | Journal of Thrombosis and Haemostasis | 2011 | 9 | | 920-921 |
| Taylor, J. Ryan and Shea, Kelsey E. and Clark, Charles F. and Kelly, James D., 2nd and Schrumpf, Mark A. | Elbow hemiarthroplasty for intra-articular distal humerus fractures: results and technique | JSES reviews, reports, and techniques | 2021 | 1 | 4 | 408-413 |
| Scott P. | Distal humerus fractures | Hand clinics | 2007 | 23 | 4 | 457-vi |
| Tiusanen, Roosa E. and Tiusanen, Hannu T. and Saltychev, Mikhail and Sarantsin, Pjotr M. | Discovery R elbow system arthroplasty: results of 10-year follow-up | & traumatology : orthopedie traumatologie | 2021 | 31 | 6 | 1207-1213 |
| van der Lugt, J. C. T. and Geskus, R. B. and Rozing, P. M. | Primary Souter-Strathclyde total elbow prosthesis in rheumatoid arthritis | The Journal of bone and joint surgery. American volume | 2004 | 86 | 3 | 465-73 |
| Varecka, Thomas F. and Myeroff, Chad | Distal Humerus Fractures in the Elderly Population | The Journal of the American Academy of Orthopaedic Surgeons | 2017 | 25 | 10 | 673-683 |
| Vieira Cardoso, Diogo and Cunningham, Gregory | [Fractures of the distal humerus in the elderly : osteosynthesis or elbow arthoplasty ?] | personne agee : osteosynthese ou prothese de coude ? | 2017 | 13 | 587 | 2177-2183 |
| Tsai, Shang-Wen and Chen, Cheng-Fong and Wu, Po-Kuei and Chen, Wei-Ming | versus post-traumatic conditions: a systematic review and meta-analysis | The bone & joint journal | 2019 | 101 | 12 | 1489-1497 |
| Wang, Zheng-Hong and Xiang, Ming and Xie, Jie and Tang, Hao-Chen and Chen, Hang and Liu, Xin | [Treatment of humerus shaft fractures using minimally invasive percutaneous plate osteosynthesis through anterior approach] | Zhongguo gu shang = China journal of orthopaedics and traumatology | 2009 | 22 | 9 | 681-3 |
| Zha, Ye-Jun and Jiang, Xie-Yuan and Gong, Mao-Qi | [Choices to treat fracture of distal humerus in the elderly patients] | Zhongguo gu shang = China journal of orthopaedics and traumatology | 2012 | 25 | 9 | 773-8 |
| Zhang, Dafang and Chen, Neal | Total Elbow Arthroplasty | The Journal of hand surgery | 2019 | 44 | 6 | 487-495 |
| Zhang, Qing and Xiang, Ming and Yang, Jin-Song and Dai, Fei | Using a Semi-constrained Prosthesis with a Triceps-preserving Approach over a Minimum Follow-up Period of 4 Years | Orthopaedic surgery | 2023 | 15 | 8 | 2091-2101 |
| Zook, J. and Ward Sr, W. G. | revision | Journal of Arthroplasty | 2001 | 16 | 7 | 919-922 |

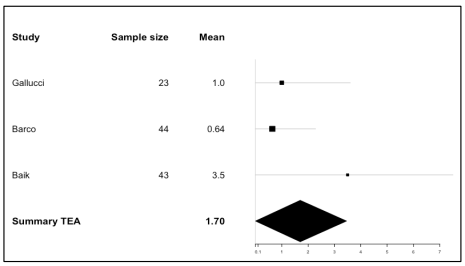
# Appendix I Assessment of Jonsson et al. 2023 against ROB-2 criteria.

(Sterne et al., 2019; Jonsson et al., 2023)



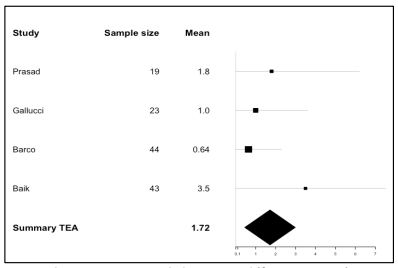
# Appendix J Sensitivity analysis for paper 3

Appendix 4: Analysis excluding imputed data (left side). Analysis including imputed data provided for comparison (right side). NRS pain DHH not included as there was no imputed data.



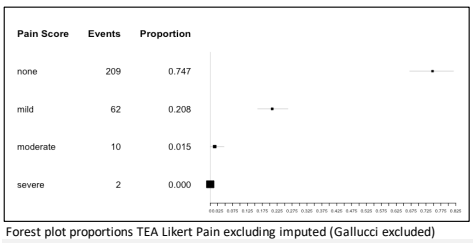

Forest plot NRS pain TEA excluding imputed (Prasad excluded)

| | | | |
|---|---|---|---|
| Studies | 3 | Mean | 1.7009 |
| Observations | 110 | LL | -0.0535 |
| | | UL | 3.4554 |
| | | I2(%) | 97.2 |
| | | tau2 | 2.3412 |
| | | H | 5.96 |
| | | Q | 71.11 |
| | | DF | 2 |
| | | P | <0.0001 |



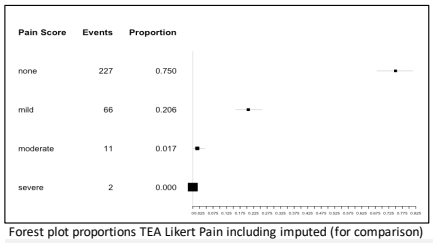

Forest plot NRS pain TEA including imputed (for comparison)

| | | | |
|---|---|---|---|
| Studies | 4 | Mean | 1.7193 |
| Observations | 129 | LL | 0.4464 |
| | | UL | 2.9921 |
| | | I2(%) | 95.9 |
| | | tau2 | 1.5803 |
| | | H | 4.92 |
| | | Q | 72.62 |
| | | DF | 3 |
| | | P | <0.0001 |



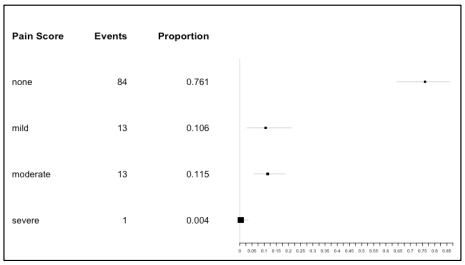

Forest plot proportions TEA Likert Pain excluding imputed (Gallucci excluded)

| | | | None | Mild | Moderate | Severe |
|---|---|---|---|---|---|---|
| Studies | 14 | | | | | |
| Observations | 283 | Events | 209 | 62 | 10 | 2 |
| | | Proportion | 0.747 | 0.2083 | 0.0152 | 0.0001 |
| | | LL | 0.6705 | 0.1597 | 0.0001 | 0 |
| | | UL | 0.8172 | 0.2609 | 0.046 | 0.0114 |
| | | I2(%) | 45.4 | 4.8 | 30.9 | 0 |
| | | tau2 | 0.0094 | 0.001 | 0.0054 | 0 |
| | | H | 1.35 | 1.03 | 1.2 | 1 |
| | | Q | 23.8 | 13.66 | 18.82 | 5.14 |
| | | DF | 13 | 13 | 13 | 13 |
| | | P | 0.033 | 0.3984 | 0.1288 | 0.9719 |



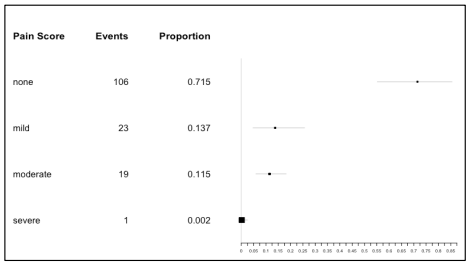

Forest plot proportions TEA Likert Pain including imputed (for comparison)

| | | | None | Mild | Moderate | Severe |
|---|---|---|---|---|---|---|
| Studies | 15 | | | | | |
| Observations | 306 | Events | 227 | 66 | 11 | 2 |
| | | Proportion | 0.7503 | 0.2055 | 0.0167 | 0.0001 |
| | | LL | 0.6813 | 0.159 | 0.0007 | 0 |
| | | UL | 0.8412 | 0.2558 | 0.0455 | 0.0105 |
| | | I2(%) | 41.5 | 0 | 26.8 | 0 |
| | | tau2 | 0.0072 | 0.001 | 0.0043 | 0 |
| | | H | 1.31 | 1 | 1.17 | 1 |
| | | Q | 23.93 | 13.84 | 19.14 | 5.19 |
| | | DF | 14 | 14 | 14 | 14 |
| | | P | 0.0467 | 0.4615 | 0.1598 | 0.983 |




Forest plot proportions DHH Likert Pain excluding imputed (Al-Hamdani and Dirckx excluded)

| | | | None | Mild | Moderate | Severe |
|---|---|---|---|---|---|---|
| Studies | 4 | | | | | |
| Observations | 111 | Events | 84 | 13 | 13 | 1 |
| | | Proportion | 0.7606 | 0.1060 | 0.1146 | 0.0039 |
| | | LL | 0.6435 | 0.0291 | 0.0579 | 0 |
| | | UL | 0.8618 | 0.2147 | 0.1847 | 0.0341 |
| | | I2(%) | 44.3 | 54.5 | 0 | 0 |
| | | tau2 | 0.0065 | 0.0112 | 0 | 0 |
| | | H | 1.34 | 1.48 | 1 | 1 |
| | | Q | 5.39 | 6.59 | 0.7 | 0.75 |
| | | DF | 3 | 3 | 3 | 3 |
| | | P | 0.1455 | 0.0862 | 0.8736 | 0.8622 |




Forest plot proportions DHH Likert Pain including imputed (for comparison)

| | | | None | Mild | Moderate | Severe |
|---|---|---|---|---|---|---|
| Studies | 6 | | | | | |
| Observations | 149 | Events | 106 | 23 | 19 | 1 |
| | | Proportion | 0.7150 | 0.1374 | 0.1147 | 0.0019 |
| | | LL | 0.5522 | 0.0479 | 0.0597 | 0 |
| | | UL | 0.8551 | 0.2567 | 0.182 | 0.0245 |
| | | tau2 | 0.0305 | 0.021 | 0.0021 | 0 |
| | | H | 1.97 | 1.73 | 1.19 | 1 |
| | | I2(%) | 74.2 | 66.7 | 29 | |
| | | Q | 19.35 | 15 | 7.04 | 0.83 |
| | | DF | 5 | 5 | 5 | 5 |
| | | P | 0.0017 | 0.0104 | 0.2174 | 0.9748 |

# Appendix K Standardised SHELF Expert briefing document.



## SHELF Expert Briefing

Thank you for agreeing to take part in an expert knowledge elicitation workshop. You will be one of a group of experts who will be asked to make judgements regarding one or more Quantities of Interest (QoIs). The QoIs for your workshop, and the importance of expert knowledge about these quantities, have been set out in your invitation letter. The purpose of this document is to explain what kind of judgements you will be asked to make.

### SHELF

The elicitation will be conducted following the Sheffield Elicitation Framework (SHELF), based on elicitation methodologies originally recommended in O'Hagan et al (2006) and subsequently refined through extensive practical experience with the SHELF approach.

The SHELF process begins with inviting experts and agreeing with them a date for the elicitation workshop. In preparation for the workshop, experts are sent this briefing document and details of the uncertain quantities of interest that will be considered in the workshop. Another important step is the gathering of evidence; you may already have received a draft 'evidence dossier' and been asked if you can contribute additional evidence.

The workshop will be conducted by an experienced facilitator who will begin by reviewing the QoIs and the final dossier of evidence relating to them. For each QoI, the facilitator will then ask all the experts to make a number of judgements which will reveal the range of opinion regarding the quantity. The facilitator will then lead a discussion aimed at exploring and understanding differences, with a view to reaching a set of 'consensus' judgements to represent the combined knowledge of the experts present.

It is important to note that you will *not* be asked to provide single estimates of any of these quantities. The elicitation process will instead involve considerations such as what a plausible range of values would be for each quantity of interest, and which values, in your opinion, are more likely than others. You may have considerable uncertainty about some of these quantities (though less than that of a lay person). This will not be of concern during the elicitation itself, as the outputs from the elicitation will reflect large uncertainty when it is present.

To illustrate the kind of judgements you will be asked to make in the workshop we imagine a freshwater biologist who is participating as an expert in a workshop where one of the quantities of interest is the average lifetime of a particular species of fish in ideal water conditions (ideal temperature, oxygenation, acidity, etc.). We will call this quantity L.

**Probabilities**

The formal way to describe and quantify uncertainty is by using probabilities, and we begin by considering how the biologist's knowledge and uncertainty about L are described by probabilities.

The concept of probability is invariably introduced to students using the *frequency* definition, according to which the probability of something happening is the frequency with which is happens in the long run, over a very long sequence of repetitions. For instance, the probability of throwing a 'Six' with a regular six-sided die is one-sixth, because in the long run, over very many tosses of the die, 'Six' will come up on one throw in six.

The frequency definition of probability only applies to events, such as tossing a die, which are *repeatable* so that we can observe how often the event occurs. However, the quantities of interest that you will be asked about in the workshop will only ever have one value. The fish average lifetime L is a typical example. The average lifetime in ideal conditions is a quantity that can only have one value. (Lifetimes of individual fish obviously vary but the average has a unique fixed value.) We cannot define the probability of the fish living on average more than 3 years by the frequency with which the average is more than 3 years in a long sequence of repetitions because it is not repeatable. The average lifetime *either is or is not* more than 3 years.

Every QoI that we ask about in expert elicitation has a single fixed true value, although that value is unknown. Probabilities in elicitation are therefore not frequency probabilities but are instead judgements. If the biologist gives a probability, say 0.4 or 40%, that the average lifetime of this species of fish is more than 3 years, this is a judgement that the biologist makes, expressing his or her degree of certainty. A probability close to 1 will mean that the biologist is almost sure that L will exceed 3, while a probability near 0 will mean that he or she is almost sure that L will *not* exceed 3. A value of 0.4 would mean that this expert judges it to be slightly less likely that L will be more than 3 than that it will be less than 3.

Notice that different experts will generally have different beliefs and will therefore assign different probabilities to statements such as "L is more than 3". In elicitation, we use a definition of probability called *personal probability* or *subjective probability*.

**Subjectivity**

You may feel uncomfortable about this element of subjectivity in the elicitation process, because subjectivity has negative connotations in everyday usage, including bias, prejudice, wishful thinking or sloppy thinking. So it is important to understand how these risks are addressed in a well-conducted elicitation exercise.

- It is inevitably true that different experts will give different probabilities, because they evaluate the evidence differently and because they bring different experiences and expertise to bear on the QoI. Some experts' judgements will be better than others. You have been invited to take part in this elicitation exercise because the organisers believe that your expertise is particularly valuable and that your judgements will be good.

- Judgements are usually elicited from several experts, thereby achieving greater objectivity by accessing a broader spectrum of opinion.

- The available evidence regarding the QoIs will be carefully reviewed, in order that each expert is fully aware of any relevant data.

- The SHELF elicitation process is designed to avoid the kinds of psychological bias that are known to be most common in eliciting probability judgements.

- The elicitation will be conducted by a facilitator who is skilled in managing the process to help experts to make the best possible judgements.

**Judgements**

The facilitator will choose what judgements to ask from the experts, within the SHELF guidance. You may be asked to make direct judgements of probability, such as, "What is your probability that L is greater than 3?" However, you will generally also be asked more indirect questions.

The facilitator will often ask experts for their *plausible range*, comprising a lower plausible limit L and an upper plausible limit U. The idea is that you should be almost certain that the QoI will lie in this range.

Another judgement that the facilitator may request is your median. This is a value M such that you think the QoI is equally likely to be above or below M.

Whatever judgements the facilitator requests will be fully explained, with clear guidance on how to make those judgements accurately. You will also do a practice elicitation before having to make your judgements about the quantities of interest, in order to familiarise you with the task.

*We hope that you will find the elicitation exercise interesting. Thank you again for agreeing to take part!*

Reference:

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. E., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006). *Uncertain Judgements: Eliciting Expert Probabilities*. Chichester: Wiley.

# Appendix L Evidence dossier pain scores after TER and DHH for acute trauma.

**Evidence dossier for Pain score distribution Expert elicitation study for Distal humerus hemiarthroplasty and total elbow Arthroplasty using a validated Numerical patient rated outcome-measure in Trauma (PEDANT)**

Adam C Watts
Catriona McDaid
Catherine Hewitt

Version 1.5, 7th Jan 2024

**Quantities of interest**
A: You will be asked to consider your median of Patient Rated Elbow Evaluation (PREE) pain at 12 months after a total elbow replacement.
B: You are being asked to consider your difference in the median PREE pain scores at 12 months in patients having a total elbow replacement or distal humerus hemiarthroplasty for acute trauma (treatment effect).
C: You will be asked to consider your standard deviation of the PREE pain score at 12 months after a total elbow replacement and Distal Humerus Hemiarthroplasty.

**The PREE score**
The PREE is a validated 20 item pain and function score used to measure outcomes of elbow interventions. It is divided into 5 questions on pain, and 15 questions on function. The two are reported individually on a scale from 0 to 50, and the two can be combined to give a total score on a scale from 0 to 100. For this elicitation exercise you are only asked to consider the difference in the median pain scores. In the PREE the five items below are scored on a numerical rating scale from 0 to 10 and the scores for each are summed to give a total pain score out of 50.

1. Rate your pain when you are at rest
2. Rate your pain when doing a task with repeated arm movement
3. Rate your pain when lifting a heavy object
4. Rate your pain at its worst
5. How often do you have pain? (0 = never, 10 = always)

**Background**
The number of fractures of the distal humerus is increasing due to the ageing population.(1) Whilst fixation of the fracture is desirable, in the older population there is a greater risk of osteoporosis and more complex fractures that mean fixation is not possible, and a randomised controlled trial has indicated that joint replacement (arthroplasty) may be a better option for these patients.(2) The increasing demand is reflected by a 3-fold increase in the number of elbow arthroplasty procedures for acute trauma recorded by the National Joint Registry (NJR) over 10 years.(3) There are two available options for prosthetic replacement for trauma, total elbow arthroplasty (TEA) where both sides of the elbow joint are replaced and distal humerus hemiarthroplasty (DHH) where only the broken distal humerus part is replaced. There are some situations where there is a clear indication for TEA, including fractures on both sides of the joint (distal humerus and coronoid fracture) and pre-existing symptomatic arthritis of the elbow. There are also areas where it might be

considered reasonable to favour TEA over DHH such as in a patient with known polyarticular inflammatory joint disease. However, for the majority of patients with an acute unreconstructable distal humerus fracture the choice between TEA and DHH is unclear. Currently approximately half of the acute trauma patients recorded in the NJR are having a TEA and half a DHH, suggesting equipoise amongst the orthopaedic community.(3)

Current evidence suggests that there are no significant differences in the outcome for patients from TEA or DHH for acute trauma as measured by functional scores and range of movement, however the quality if the evidence is low.(4)(5) Data from the NJR indicates that there may be differences in revision rate for TEA and DHH but given the small number of cases available this difference is uncertain.(3) A recent Delphi study has determined core outcome domains for elbow trauma and the Patient and Public Involvement group were clear that, of the chosen domains, pain was the most important outcome from their perspective.

**Evidence regarding pain outcomes after elbow arthroplasty for fracture.**

A systematic review of the literature from 2000 to 2023 undertaken for the purposes of this elicitation to explore the pain outcomes of TEA and DHH for acute fractures identified 1 published randomised controlled trial which reported aggregate Likert pain scores for TEA and DHH as measured using the Mayo elbow performance score, the authors were contacted to try to obtain individual pain data but the data is not available.(6) A further trial protocol was identified and the authors contacted to obtain any available data. The authors kindly supplied their data which reported pain as measured by the Oxford Elbow Score and Disabilities of the Arm Shoulder and Hand score but the numbers were below the inclusion criteria for the systematic review (7 patients) (unpublished data) The remaining evidence consisted of case series. There were 17 case series reporting pain outcomes for TEA; 4 reported pain using numerical rating scales (NRSpain) and 14 used a Likert scale for pain. In two of the 5 studies using NRSpain the mean or median was provided but with no estimate of variance so this was imputed from the minimum-maximum values as described by Hozo et al.(7) There were 6 case series reporting pain outcomes for DHH; 2 using NRSpain and 4 using Likert pain scales. The included studies are shown below.

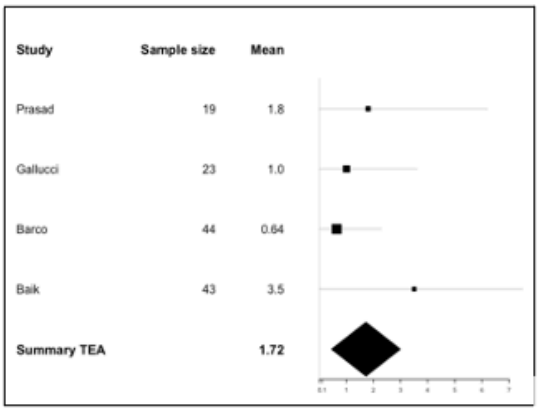
| Author | DHH or TEA | Sample size | AO Classification | Implant | Mean Age (years) | % Female | Mean Follow-up (months) |
|---|---|---|---|---|---|---|---|
| Strelzow 2021 | TEA | 40 | NA | Latitude | 79 | 88 | 70 |
| Sorensen 2014 | TEA | 20 | 13B3/13C3 | Coonrad-Morrey | 77 | 90 | 21 |
| Prasad 2008 | TEA | 15 | 13A3/13B3/13C3 | Coonrad-Morrey | 78 | 73 | 52.4 |
| Prasad 2016 | TEA | 19 | 13A3/13B/13C | Coonrad-Morrey | 68 | 63 | 156 |
| Pogliacomi 2016 | TEA | 20 | 13C | Coonrad-Morrey/ Latitude | 74 | 95 | 60 |
| Kamineni 2004 | TEA | 43 | 13A/13B/13C | Coonrad-Morrey | 69 | 72 | 84 |
| Garcia 2002 | TEA | 16 | 13A3/13B3/13C3 | Coonrad-Morrey | 73 | 75 | 36 |
| Gambirasio 2001 | TEA | 10 | 13B1/13C | Coonrad-Morrey | 85 | 100 | 17.8 |
| Gallucci 2016 | TEA | 23 | 13A2/13C | Coonrad-Morrey | 76 | 91 | 39.7 |
| Frankle 2003 | TEA | 12 | 13C | Coonrad-Morrey | 72 | 100 | 44.5 |
| Barco 2017 | TEA | 44 | NA | Coonrad-Morrey | 71 | 75 | NA |
| Antuna 2012 | TEA | 14 | 13B3/13C2/13C3 | Coonrad-Morrey | 77 | NA | 54 |
| Ali 2010 | TEA | 20 | 13A3/13B3/13C3 | Coonrad-Morrey | 72 | 80 | 63.2 |
| Jonsson 2023 | TEA | 17 | 13B3/13C1/13C3 | Latitude / Discovery | 77 | 88 | 26.4 |
| Baik 2020 | TEA | 43 | 13C | Coonrad-Morrey | 79 | 79 | 34 |
| Jost 2008 | TEA | 10 | 13A/13B/13C | Coonrad-Morrey | 65 | NA | 66 |
| Celli 2021 | TEA | 13 | 13C2/13C3 | Coonrad-Morrey | 64 | 85 | 153 |
| Mean | | 22 | | | 74 | 84 | 61 |

Table 1: Included studies TEA. (8) (9)(10)(11)(12)(13)(14)(15)(16)(17)(18)(19)(6)(20)

| Author | DHH or TEA | Sample size | AO Classification | Implant | Mean Age (years) | % Female | Mean Follow-up (months) |
|---|---|---|---|---|---|---|---|
| Al-hamdani 2019 | DHH | 24 | 13C2,13C3 | Latitude Hemi | 65 | 88 | 25 |
| Nestorson 2015 | DHH | 42 | 13C3 | Latitude Hemi | 72 | 93 | 34 |
| Rotini 2023 | DHH | 27 | 13C3/13C2/13B3 | Latitude Hemi | 64 | 78 | 82 |
| Celli 2022 | DHH | 24 | 13C2/13C3/13B3 | Latitude Hemi | 64 | 75 | 91 |
| Dirckx 2023 | DHH | 15 | 13C3/13B3 | Latitude Hemi | 75 | 93 | 31 |
| Jonsson 2023 | DHH | 18 | 13B3/13C1/13C3 | Latitude Hemi | 74 | 89 | 32 |
| Mean | | 25 | | | 69 | 86 | 49 |

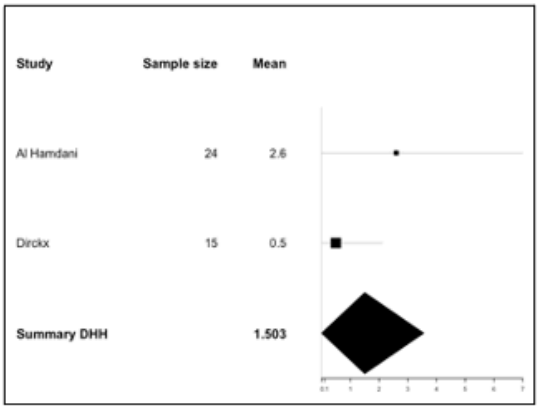
Table 2: Included studies DHH.(21)(22)(23)(24)(25)(6)

No direct comparison can be drawn between the data for TEA and DHH due to the risk of bias of included studies but a meta-analysis of the average pain outcome using a NRSpain from 0-10 is shown below with an estimate of the mean for TEA and DHH.




| | | | |
|---|---|---|---|
| Studies | 4 | Mean | 1.7193 |
| Observations | 129 | LL | 0.4464 |
| | | UL | 2.9921 |
| | | I2(%) | 95.9 |
| | | Q | 72.62 |
| | | DF | 3 |
| | | P | <0.0001 |

Figure 1. Results of meta-analysis of NRS pain scores for TEA



| | | | |
|---|---|---|---|
| Studies | 2 | Mean Pain | 1.503 |
| Observations | 39 | LL | 0.001 |
| | | UL | 3.5589 |
| | | I2(%) | 93.6 |
| | | Q | 15.6 |
| | | DF | 1 |
| | | P | <0.0001 |

Figure 2. Results of meta-analysis of NRS pain scores for DHH

A sensitivity analysis was undertaken including only those studies in which the mean and standard deviation was reported. This did not change to results for DHH but for TEA the summary NRS Pain value was reduced to 1.00.

A meta-analysis of results from the published Likert scales for pain are reported below with the probabilities and 95% confidence intervals for outcomes (no pain, mild pain, moderate pain, severe pain). Reported differences in proportions for TEA and DHH may be due to heterogeneity of inclusion criteria and the risk of bias of evidence.

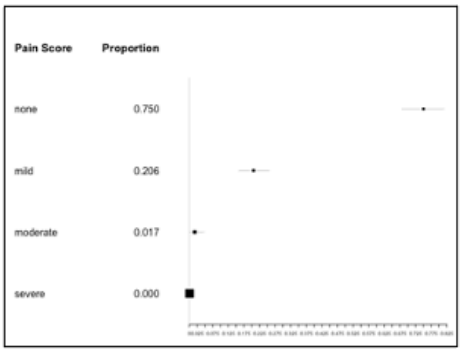



Fig 3. Summary results Likert pain scale TEA.

| | | | | | | |
|---|---|---|---|---|---|---|
| Studies | 15 | | None | Mild | Moderate | Severe |
| Observations | 306 | Events | 227 | 66 | 11 | 2 |
| | | Proportion | 0.7503 | 0.2055 | 0.0167 | 0.0001 |
| | | LL | 0.6813 | 0.159 | 0.0007 | 0.0001 |
| | | UL | 0.8142 | 0.2558 | 0.0455 | 0.0105 |
| | | I2(%) | 41.5 | 0 | 26.8 | 0 |
| | | Q | 23.93 | 13.84 | 19.14 | 5.19 |
| | | DF | 14 | 14 | 14 | 14 |
| | | P | 0.0467 | 0.4615 | 0.1598 | 0.983 |

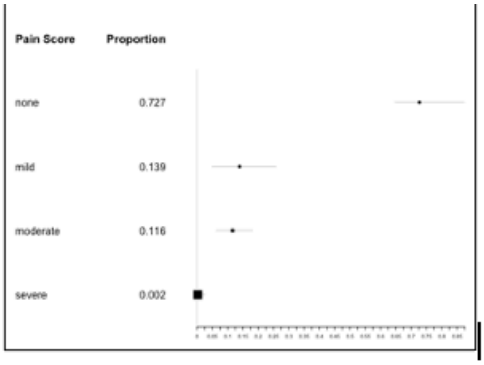



Fig 4. Summary results Likert pain scale DHH.

| | | | | | | |
|---|---|---|---|---|---|---|
| Studies | 6 | | None | Mild | Moderate | Severe |
| Observations | 149 | Events | 106 | 23 | 19 | 1 |
| | | Proportion | 0.7267 | 0.1385 | 0.1159 | 0.0019 |
| | | LL | 0.6476 | 0.0486 | 0.0625 | 0 |
| | | UL | 0.8726 | 0.2581 | 0.1806 | 0.0245 |
| | | I2(%) | 76.5 | 66.4 | 25.2 | 0 |
| | | Q | 21.32 | 14.86 | 6.69 | 0.83 |
| | | DF | 2 | 5 | 5 | 5 |
| | | P | 0.0007 | 0.011 | 0.245 | 0.9748 |

**Pain instruments**

The choice of instruments to measure pain after total elbow arthroplasty is limited. Many instruments combine pain assessment with function assessment (Oxford elbow score, DASH score, Liverpool elbow score) and the items are rarely reported separately by investigators. Pain scales such as the visual analogue scale, numerical rating scale or Likert scale have been validated as tools for measuring pain but we are not aware of any investigations of their psychometric properties in elbow arthroplasty studies. The Patient Rated Elbow Evaluation (PREE) and American Shoulder and Elbow Society (ASES) scores measure and report pain independently and have been validated in general elbow conditions and in total elbow arthroplasty studies.(26)(27) The PREE is less resource intensive as it does not require physician assessment.

Angst have reported that in patients undergoing total elbow replacement the mean PREE improved from a score of 30/50 (s.d. 9.2) pre-operation to 15/50 (s.d. 13.7) at 6 months after surgery. 54% of the 79 patients included had a diagnosis of rheumatoid arthritis.(28) Acute fracture patients were excluded from the analysis.

In a separate study Angst has measured pain six months after TER using PREE pain and a Likert pain scale from the SF-36.(27) The mean PREE pain score was 14/50 (s.d. 13.3) and the mean SF-36 pain score was 60 which equates to "mild pain" on the 6-item SF-36 Likert scale.

**References**


1. Palvanen M, Kannus P, Niemi S, Parkkari J. Secular trends in distal humeral fractures of elderly women. Bone. 2010 May;46(5):1355–8.

2. McKee MD, Veillette CJH, Hall JA, Schemitsch EH, Wild LM, McCormack R, et al. A multicenter, prospective, randomized, controlled trial of open reduction—internal fixation versus total elbow arthroplasty for displaced intra-articular distal humeral fractures in elderly patients. J Shoulder Elbow Surg. 2009 Jan;18(1):3–12.

3. 19th Annual Report of the National Joint Registry [Internet]. NJR; 2022. Available from: https://reports.njrcentre.org.uk/Portals/0/PDFdownloads/NJR%2019th%20Annual%20Report%202022.pdf

4. Stoddart MT, Panagopoulos GN, Craig RS, Falworth M, Butt D, Rudge W, et al. A systematic review of the treatment of distal humerus fractures in older adults: A comparison of surgical and non-surgical options. Shoulder Elb. 2022 May 9;175857322210998.

5. Burden EG, Batten T, Smith C, Evans JP. Hemiarthroplasty or total elbow arthroplasty for unreconstructable distal humeral fractures in patients aged over 65 years: a systematic review and meta-analysis of patient outcomes and complications. Bone Jt J. 2022 May 1;104-B(5):559–66.

6. Jonsson EÖ, Ekholm C, Hallgren HB, Nestorson J, Etzner M, Adolfsson L. Elbow hemiarthroplasty and total elbow arthroplasty provided a similar functional outcome for unreconstructable distal humeral fractures in patients aged 60 years or older: a multicenter randomized controlled trial. J Shoulder Elbow Surg. 2023 Sep;S105827462300705X.

7. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol. 2005 Dec;5(1):13.

8. Sørensen BW. Primary total elbow arthroplasty in complex fractures of the distal humerus. World J Orthop. 2014;5(3):368.

9. Prasad N, Dent C. Outcome of total elbow replacement for distal humeral fractures in the elderly: A COMPARISON OF PRIMARY SURGERY AND SURGERY AFTER FAILED INTERNAL FIXATION OR CONSERVATIVE TREATMENT. J Bone Joint Surg Br. 2008 Mar;90-B(3):343–8.

10. Prasad N, Ali A, Stanley D. Total elbow arthroplasty for non-rheumatoid patients with a fracture of the distal humerus: a minimum ten-year follow-up. Bone Jt J. 2016 Mar;98-B(3):381–6.

11. Pogliacomi F, Schiavi P, Defilippo M, Corradi M, Vaienti E, Ceccarelli F, et al. Total elbow arthroplasty following complex fractures of the distal humerus: results in patients over 65 years of age. Acta Bio-Medica Atenei Parm. 2016 Sep 13;87(2):148–55.

12. Kamineni S, Morrey BF. Distal Humeral Fractures Treated with Noncustom Total Elbow Replacement: J Bone Jt Surg. 2004 May;86(5):940–7.

13. Garcia JA, Mykula R, Stanley D. Complex fractures of the distal humerus in the elderly. The role of total elbow replacement as primary treatment. J Bone Joint Surg Br. 2002 Aug;84(6):812–6.

14. Gambirasio R, Riand N, Stern R, Hoffmeyer P. Total elbow replacement for complex fractures of the distal humerus. An option for the elderly patient. J Bone Joint Surg Br. 2001 Sep;83(7):974–8.

15. Gallucci GL, Larrondo Calderón W, Boretto JG, Castellaro Lantermo JA, Terán J, De Carli P. Total elbow arthroplasty for the treatment of distal humeral fractures. Rev Esp Cir Ortopédica Traumatol Engl Ed. 2016 May;60(3):167–74.

16. Frankle MA, Herscovici, D, DiPasquale TG, Vasey MB, Sanders RW. A Comparison of Open Reduction and Internal Fixation and Primary Total Elbow Arthroplasty in the Treatment of Intraarticular Distal Humerus Fractures in Women Older Than Age 65: J Orthop Trauma. 2003 Aug;17(7):473–80.

17. Barco R, Streubel PN, Morrey BF, Sanchez-Sotelo J. Total Elbow Arthroplasty for Distal Humeral Fractures: A Ten-Year-Minimum Follow-up Study. J Bone Jt Surg. 2017 Sep 20;99(18):1524–31.

18. Antuña SA, Laakso RB, Barrera JL, Espiga X, Ferreres A. Linked total elbow arthroplasty as treatment of distal humerus fractures. Acta Orthop Belg. 2012 Aug;78(4):465–72.

19. Ali A, Shahane S, Stanley D. Total elbow arthroplasty for distal humeral fractures: indications, surgical approach, technical tips, and outcome. J Shoulder Elbow Surg. 2010 Mar;19(2 Suppl):53–8.

20. Celli A, Paroni C, Bonucci P, Celli L. Total elbow arthroplasty for acute distal humeral fractures with humeral condyle resection or retention: a long-term follow-up study. JSES Int. 2021 Jul;5(4):797–803.

21. Al-Hamdani A, Rasmussen JV, Sørensen AKB, Ovesen J, Holtz K, Brorson S, et al. Good outcome after elbow hemiarthroplasty in active patients with an acute intra-articular distal humeral fracture. J Shoulder Elbow Surg. 2019 May;28(5):925–30.

22. Nestorson J, Ekholm C, Etzner M, Adolfsson L. Hemiarthroplasty for irreparable distal humeral fractures: medium-term follow-up of 42 patients. Bone Jt J. 2015 Oct;97-B(10):1377–84.

23. Rotini R, Ricciarelli M, Guerra E, Marinelli A, Celli A. Elbow hemiarthroplasty in distal humeral fractures: Indication, surgical technique and results. Injury. 2020 Nov;S0020138320309463.

24. Celli A, Ricciarelli M, Guerra E, Bonucci P, Ritali A, Cavallo M, et al. Elbow hemiarthroplasty for acute distal humeral fractures and their sequelae: medium- and long-term follow-up of 41 cases. J Shoulder Elbow Surg. 2022 May;31(5):1015–25.

25. Dirckx M, Tathgar A, Bellringer S, Phadnis J. Hemiarthroplasty versus open reduction internal fixation for intra-articular distal humerus fractures in older patients. Shoulder Elb. 2023 Feb;15(1):83–92.

26. Vincent JI, MacDermid JC, King GJW, Grewal R. The Patient-Rated Elbow Evaluation and the American Shoulder and Elbow Surgeons—Elbow form capture aspects of functioning that are important to patients with elbow injuries. J Hand Ther. 2021 Jul;34(3):415–22.

27. Angst F, John M, Pap G, Mannion AF, Herren DB, Flury M, et al. Comprehensive assessment of clinical outcome and quality of life after total elbow arthroplasty. Arthritis Care Res. 2005 Feb 15;53(1):73–82.

28. Angst F, Goldhahn J, Drerup S, Kolling C, Aeschlimann A, Simmen BR, et al. Responsiveness of five outcome measurement instruments in total elbow arthroplasty. Arthritis Care Res. 2012 Nov;64(11):1749–55.

# Appendix M Expert enquiry form and responses.

| Name and Title | Background and Expertise | Declaration of Interests | Additional Evidence | Clarifications and Corrections |
|---|---|---|---|---|
| MANSAT Pierre, Prof. | Master in Biomechanics 1992; Fellowship Mayo Clinic - Adult Reconstruction Upper Limb (Shoulder-Elbow) 1996/1997; PhD Thesis on Shoulder arthroplasty using FEA 2002; Professor of Orthopedic and Traumatology, University Hospital Toulouse FRANCE 2004; Main activities: Shoulder, Elbow and Hand Surgeries; 180 per-reviewed articles indexed in Pubmed; Chief-editor JSES International (Elsevier) | Chief-editor JSES International (Elsevier); Payed consultant: Medartis, Stryker, Zimmer-Biomet | None | None |
| Lars Adolfsson, Professor, MD | Specialist in orthopedic and hand surgery. 35 years experience. Over 100 peer reviewed publications on upper extremity trauma and surgery. Several on elbow arthroplasty | none | nothing very important missing | none |
| Toni Luokkala, MD PhD, Consultant Elbow and Hand Surgeon | Current employer: Wellbeing services county of Central Finland, Hospital Nova, Jyväskylä, Finland. Background: 12 years as a consultant elbow and hand surgeon including upper limb fellowship position in Wrightington hospital 2017. | Consultant, National patient insurance center, Finland | - | - |
| Joideep Phadnis | Elbow specialist | Consultant for stryker (arthroplasty implants) | none additional | none |
| Jonathan Evans | I am a Consultant Shoulder and Elbow surgeon at the Princess Elizabeth Orthopaedic Centre, Royal Devon University Hospital, Exeter. I also hold a position as a Senior Clinical Lecturer in Orthopaedics at the University of Exeter. As a clinical academic I split my time 50/50 between clinical work and research activity. Over the last few years I have published systematic reviews and meta-analysis on the outcome of total elbow replacements and distal humeral hemiarthroplasties in the context of trauma and have conducted a feasibility randomised controlled trial assessing these two interventions in unreconstructable distal humeral fractures. In | I have no relevant disclosures | I have read the dossier and there is no relevant further published evidence I would offer. I have discussed some pre-published data with Prof Watts | No further clarifications and corrections. |

| | | | | |
|---|---|---|---|---|
| | 2017 I completed my doctorate assessing outcomes in lateral elbow tendinopathy, a major component of which involved the assessment of the psychometric properties of elbow specific Patient Reported Outcome Measures (PROMs). I have maintained an interest in PROMs and have published on core outcome sets in elbow pathology and the use of advanced psychometric techniques including item response theory and computerised adaptive tests. | | from our feasibility RCT. | |
| Andrew Wright. Consultant Elbow & Shoulder Surgeon, Wrightington Hospital. | Consultant Elbow Surgeon at Wrightington Hospital UK since 2020. My practice is predominantly elbow orientated with tertiary referrals from regional colleagues. I have collaborated in a multidisciplinary working group to produce national guidelines for lateral elbow tendinopathy. | No interests to declare | No additional | N/A |
| Alexander Van Tongel MD PhD | Shoulder - Elbow surgeon University Hospital Ghent | To evaluate the average pain score and to help to lower the pain score with new techniques | - | - |

# Appendix N Linear pooling and assurance assessment for expert elicitation.

During the group elicitation stage the moderator had access to the linear pool, or mathematic aggregation, of the individually elicited expert beliefs. These were not shared with the experts to avoid influencing the group stage discussion. They do however indicate the level of agreement between experts. Linear pooling can be used instead of the RIO consensus process but the latter was chosen for this study.

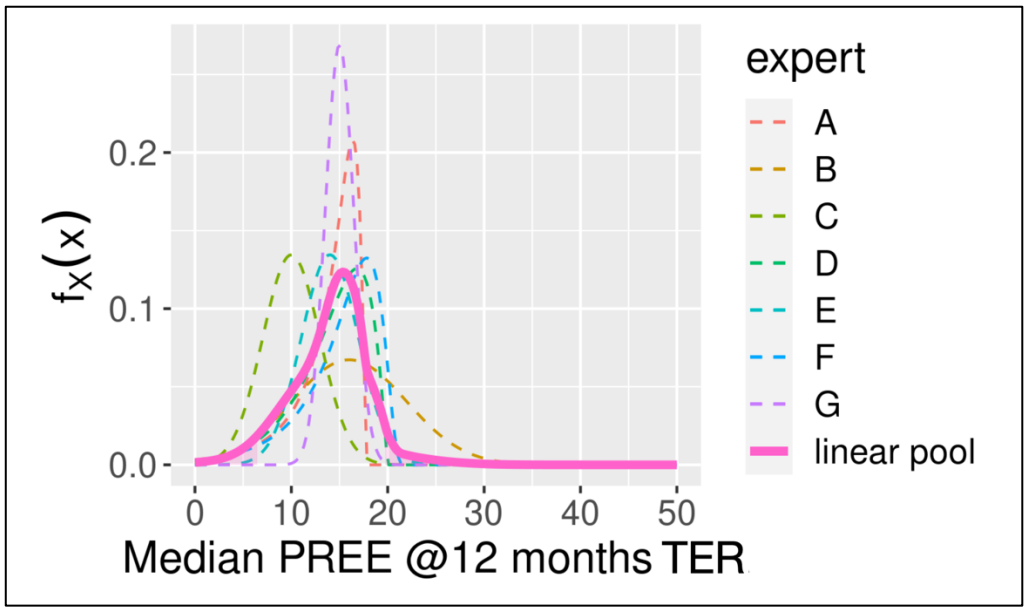



**Figure N.1** Linear pool for median PREE at 12 months after TER

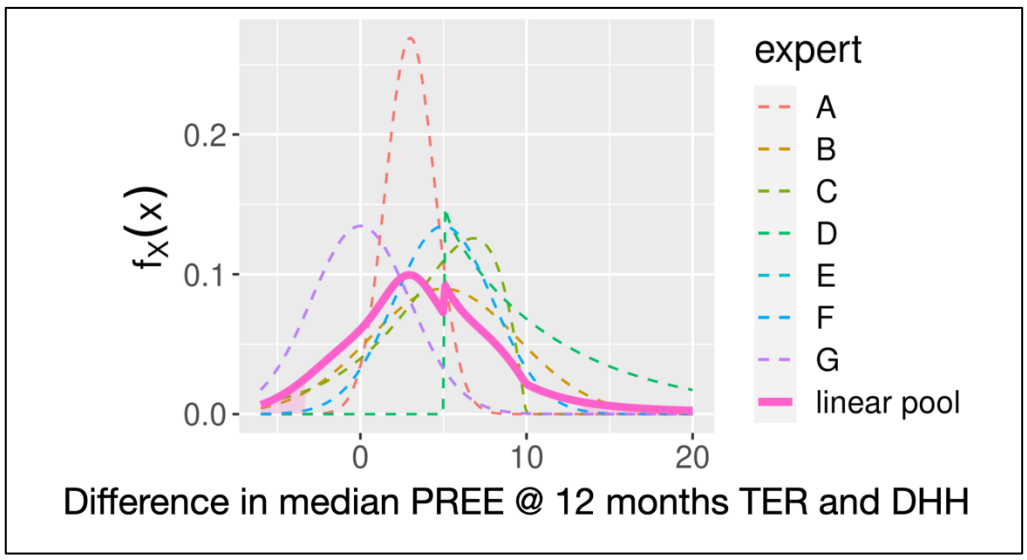



**Figure N.2** Linear pool for difference in median PREE at 12 months after TER and DHH

Assurance assessment was undertaken using the method described by Alhussain and Oakley using their online shiny app.(Alhussain and Oakley, 2020)

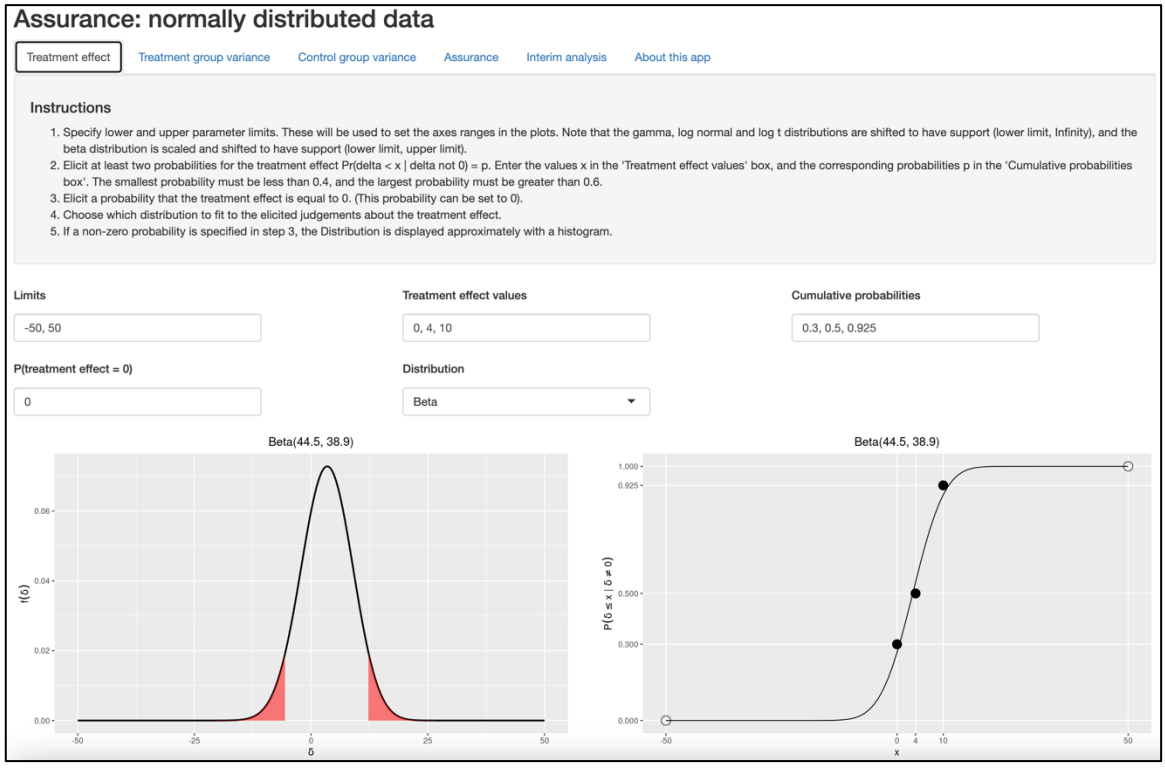



**Figure N.3** Assurance assessment: stage 1 - treatment effect.

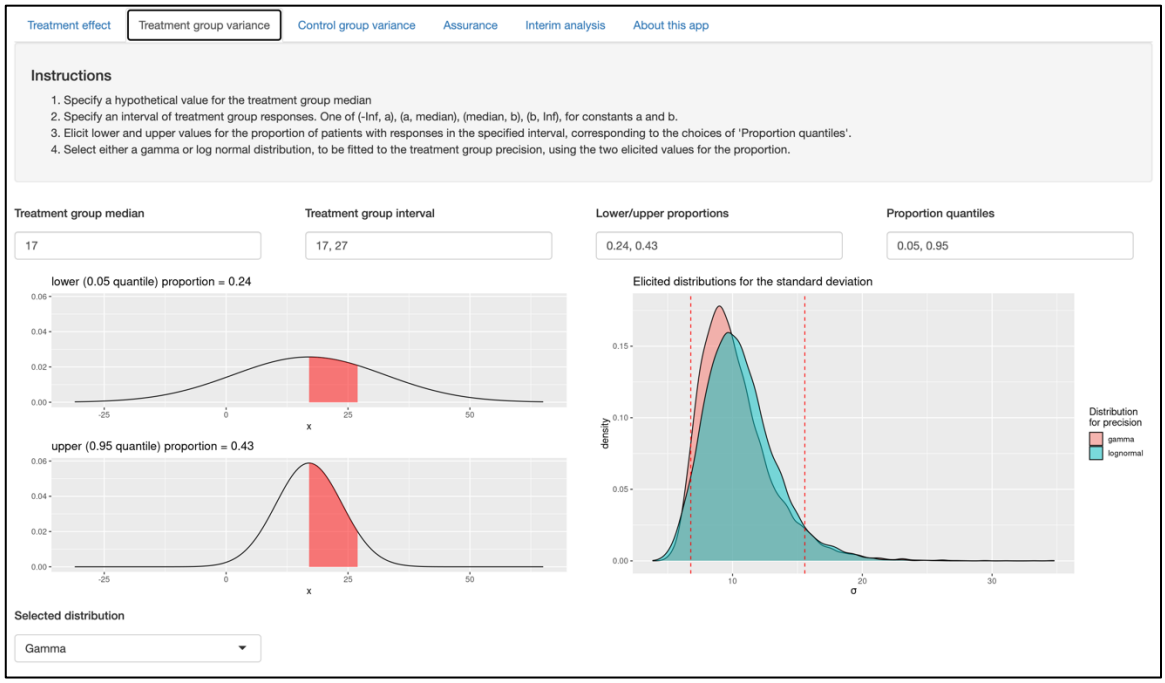



**Figure N.4** Assurance assessment: stage 2 - treatment group variance.

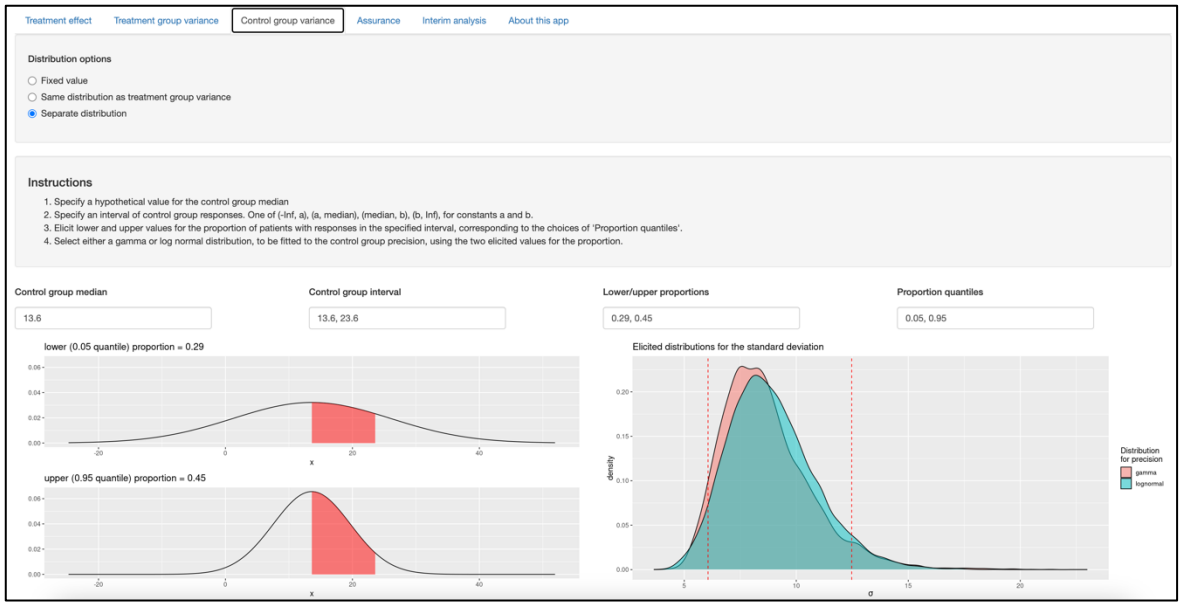



**Figure N.5** Assurance assessment: stage 3 - control group variance.

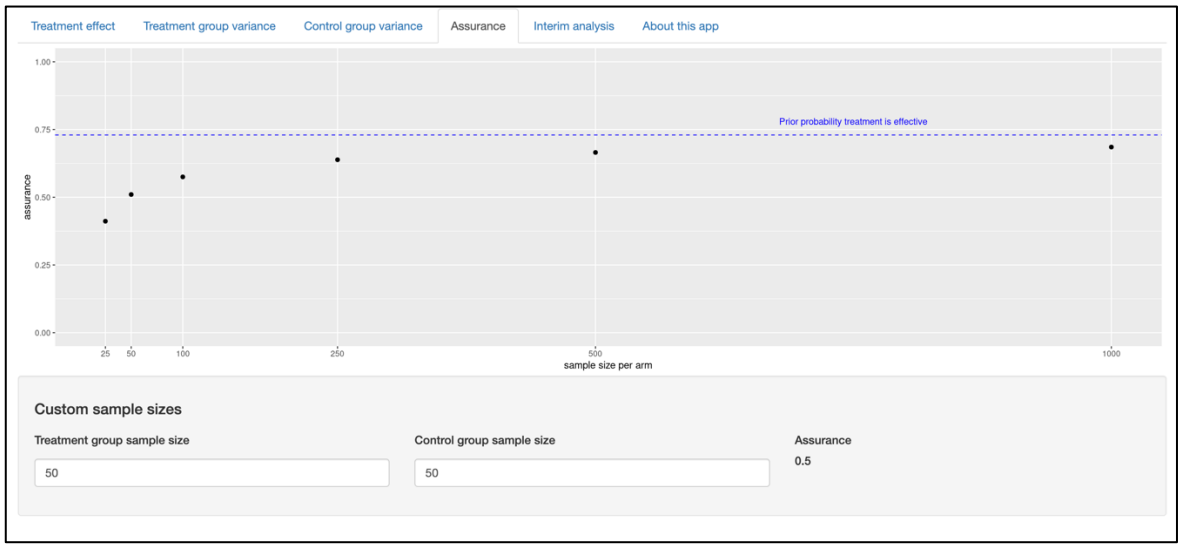



**Figure N.6** Assurance assessment: stage 4 - assurance.

The results given indicate that with the elicited median treatment effect and standard deviation for PREE pain outcomes following TER and DHH, it is not assured that a randomised trial would demonstrate a significant difference in outcomes between the two interventions when assessed using this outcome instrument.

## Appendix O G*Power output for classical power analysis based on parameters elicited in PEDANT study (Paper 4)

**t tests** - Means: Difference between two independent means (two groups)

| | | | |
|---|---|---|---|
| **Analysis:** | A priori: Compute required sample size | | |
| **Input:** | Tail(s) | = | Two |
| | Effect size d | = | 0.2828427 |
| | α err prob | = | 0.05 |
| | Power (1-β err prob) | = | 0.8 |
| | Allocation ratio N2/N1 | = | 1 |
| **Output:** | Noncentrality parameter δ | = | 2.8142493 |
| | Critical t | = | 1.9660032 |
| | Df | = | 394 |
| | Sample size group 1 | = | 198 |
| | Sample size group 2 | = | 198 |
| | Total sample size | = | 396 |
| | Actual power | = | 0.8016202 |