



■ TRAUMA

Five-year follow-up results of the PROFHER trial comparing operative and non-operative treatment of adults with a displaced fracture of the proximal humerus

H. H. Handoll,
A. Keding,
B. Corbacho,
S. D. Brealey,
C. Hewitt,
A. Rangan

From University of
York, York, United
Kingdom

■ H. H. Handoll, DPhil, Reader in Orthopaedics, School of Health and Social Care, Health and Social Care Institute, Teesside University, Middlesbrough, Tees Valley TS1 3BA, UK.

■ A. Keding, BSc, MSc, Statistician
■ B. Corbacho, BSc, MSc, Research Fellow in Health Economics
■ S. D. Brealey, BSc, PhD, Senior Trial Manager/Research Fellow in Health Sciences
■ C. Hewitt, BSc, MSc, PhD, Professor and Deputy Director, York Trials Unit, Department of Health Sciences, University of York, Lower Ground Floor, ARRC Building, Heslington, York YO10 5DD, UK.

■ A. Rangan, ChM, FRCS (Tr&Orth), Professor of Orthopaedic Surgery, Department of Trauma and Orthopaedics, James Cook University Hospital, South Tees Hospitals NHS Trust, Marton Road, Middlesbrough TS4 3BW, UK.

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

©2017 Handoll et al
doi:10.1302/0301-620X.99B3.
BJJ-2016-1028 \$2.00

Bone Joint J
2017;99-B:383–92.
Received 15 October 2016;
Accepted after revision 28
November 2016

Aims

The PROximal Fracture of the Humerus Evaluation by Randomisation (PROFHER) randomised clinical trial compared the operative and non-operative treatment of adults with a displaced fracture of the proximal humerus involving the surgical neck. The aim of this study was to determine the long-term treatment effects beyond the two-year follow-up.

Patients and Methods

Of the original 250 trial participants, 176 consented to extended follow-up and were sent postal questionnaires at three, four and five years after recruitment to the trial. The Oxford Shoulder Score (OSS; the primary outcome), EuroQol 5D-3L (EQ-5D-3L), and any recent shoulder operations and fracture data were collected. Statistical and economic analyses, consistent with those of the main trial were applied.

Results

OSS data were available for 164, 155 and 149 participants at three, four and five years, respectively. There were no statistically or clinically significant differences between operative and non-operative treatment at each follow-up point. No participant had secondary shoulder surgery for a new complication. Analyses of EQ-5D-3L data showed no significant between-group differences in quality of life over time.

Conclusion

These results confirm that the main findings of the PROFHER trial over two years are unchanged at five years.

Cite this article: *Bone Joint J* 2017;99-B:383–92.

We report the five-year follow-up of the PROximal Fracture of the Humerus Evaluation by Randomisation (PROFHER) trial (trial registration identifier: ISRCTN50850043).

PROFHER was a pragmatic, multi-centre randomised controlled trial (RCT), funded by the United Kingdom National Institute for Health Research (NIHR), which compared operative and non-operative treatment of adults with a displaced fracture of the proximal humerus involving the surgical neck.¹

Between September 2008 and April 2011, 250 adults were recruited into the trial. At two-year follow-up, the primary outcome and the Oxford Shoulder Score (OSS)^{2,3} were available for 215 participants.⁴ The results showed no significant difference between operative and non-operative treatment by OSS over two years ($p = 0.479$) or other patient-reported clinical outcomes in the two years following fracture;^{4,5} and the cost of surgery was considerably greater.⁶

The initial choice of a two-year follow-up for PROFHER was a pragmatic one which balanced feasibility and the expectation that any differences in the OSS between the two treatment groups at two years would represent a true and enduring effect. However, there is insufficient evidence from other RCTs to confirm this assumption.⁷ Recovery from serious injuries such as a fracture of the proximal humerus is a long and often incomplete process that can be hindered by complications. A substantial proportion (15/74, 20%) of participants in a trial with less severe ('minimally displaced two-part') fractures than in PROFHER had continuing 'severe' disability after two years, although less than that at one year (30/84, 37%).⁸

We reasoned that a five-year follow-up would allow for delays in recovery, potential functional deterioration, and subsequent operations resulting from complications, such as avascular necrosis and complications of surgical fixation or humeral head replacement,

Table I. Inclusion and exclusion criteria of the PROFHER trial

Criteria
Inclusion criteria:
Adults (aged ≥ 16 years) presenting within 3 wks of their injury with a radiologically confirmed displaced fracture of the humerus involving the surgical neck. This should include all two-part surgical neck fractures; three-part (including surgical neck) and four-part fractures of proximal humerus (Neer Classification). ¹⁰ It may also include displaced surgical neck fractures that do not meet the exact displacement criteria of the Neer Classification (1 cm or/and 45° angulation of displaced parts) where this reflects an individual surgeon's uncertainty (e.g. whether, or not, the surgical neck fracture should be treated surgically).
Exclusion criteria:
Associated dislocation of the injured joint of the shoulder.
Open fracture.
Mentally incompetent patient: unable to understand trial procedure or instructions for rehabilitation; significant mental impairment that would preclude compliance with rehabilitation and treatment advice.
Comorbidities precluding surgery/anaesthesia.
A clear indication for surgery such as severe soft-tissue compromise requiring surgery/emergency treatment (nerve injury/dysfunction).
Multiple injuries: same limb fractures; other upper limb fractures.
Pathological fractures (other than osteoporotic) and terminal illness.
Participant not resident in catchment area of trauma centre.

which could arise or become symptomatic later on. The extension made practical sense as the infrastructure was already in place and the potential availability of a large group of patients presented an opportunity to gain reliable evidence about patient-reported longer term outcome, as well as insight into the feasibility of future research.

We set up the extended follow-up study at the Yorks Trials Unit, securing ethical approval in September 2010 from the institution,⁵ before the end of recruitment to, and without knowing the results of, the first study.

Our primary aim was to obtain three, four and five-year data on the key outcomes (OSS, EuroQol 5D-3L (EQ-5D-3L),⁹ and subsequent surgery) to determine whether the effect of treatment detected at two-year follow-up had persisted or changed. A further aim, linked to the collection of EQ-5D-3L data and information about any further surgery, was to examine the potential effect on our economic analysis⁶ of any change in health related quality of life (HRQoL) and the costs related to this.

Our secondary aims were to generate longer term condition-specific data on shoulder function that would provide reference data for the interpretation of the findings of PROFHER and future studies of proximal humeral fractures and to inform future research in this area on the appropriate duration of follow-up.

Patients and Methods

The methodology of the main trial is reported elsewhere.^{1,5} The inclusion and exclusion criteria are shown in Table I.¹⁰ The final version of the extended study protocol (version 3.0; 09 August 2012) is published on the NIHR website.¹¹ All related amendments were reviewed and approved by the Leeds West Research Ethics Committee (08/H1311/12).

Data collection. Postal questionnaires were sent at three, four and five years after the start of the original trial to the 176 participants who had completed and returned a con-

sent form sent on receipt of their 24-month questionnaire. A pre-notification letter was sent before this and, when necessary, reminders were sent after two and four weeks, with the option to complete questionnaires by telephone after six weeks. To maximise collection of data at the three time points, patients were asked to complete a short questionnaire restricted to the OSS, EQ-5D-3L, recent operations on their shoulder, and recent fractures. Patients were also sent an unconditional £5.00 incentive payment with each questionnaire. We also collected data from NHS Digital, using the NHS Summary Care Records available electronically for authorised staff, on patient mortality at regular intervals before sending the questionnaires to avoid distressing bereaved families or friends.

Outcomes. The primary outcome measure was the OSS, which assesses pain, function and activities of daily living.^{2,3} It contains 12 items, each with five categories of response, and a range of total scores from 0 (worst outcome) to 48 (best outcome).³ Secondary outcomes were the EQ-5D-3L, used to estimate utilities (HRQoL weights),⁹ further shoulder surgery and further fractures. While mortality was a secondary outcome in the main follow-up, it was reported solely as a reason for loss to follow-up in the extended follow-up: mortality and definitive treatment of the fracture after two years could not reasonably be expected to be linked and would not anyway be listed as a cause of death. Overall, the OSS and EQ-5D-3L were collected at six, 12, 24, 36, 48 and 60 months; EQ-5D-3L data were also collected at baseline and three months. Secondary shoulder surgery and further fractures were collected from hospital forms at one and two years and from patient questionnaires at three, four and five years' follow-up.

Sample size. The main study was designed to detect a standard effect size of 0.4 (approximating to five OSS points) with 80% power using 5% significance level, and needed approximately 200 participants at two years.¹ We

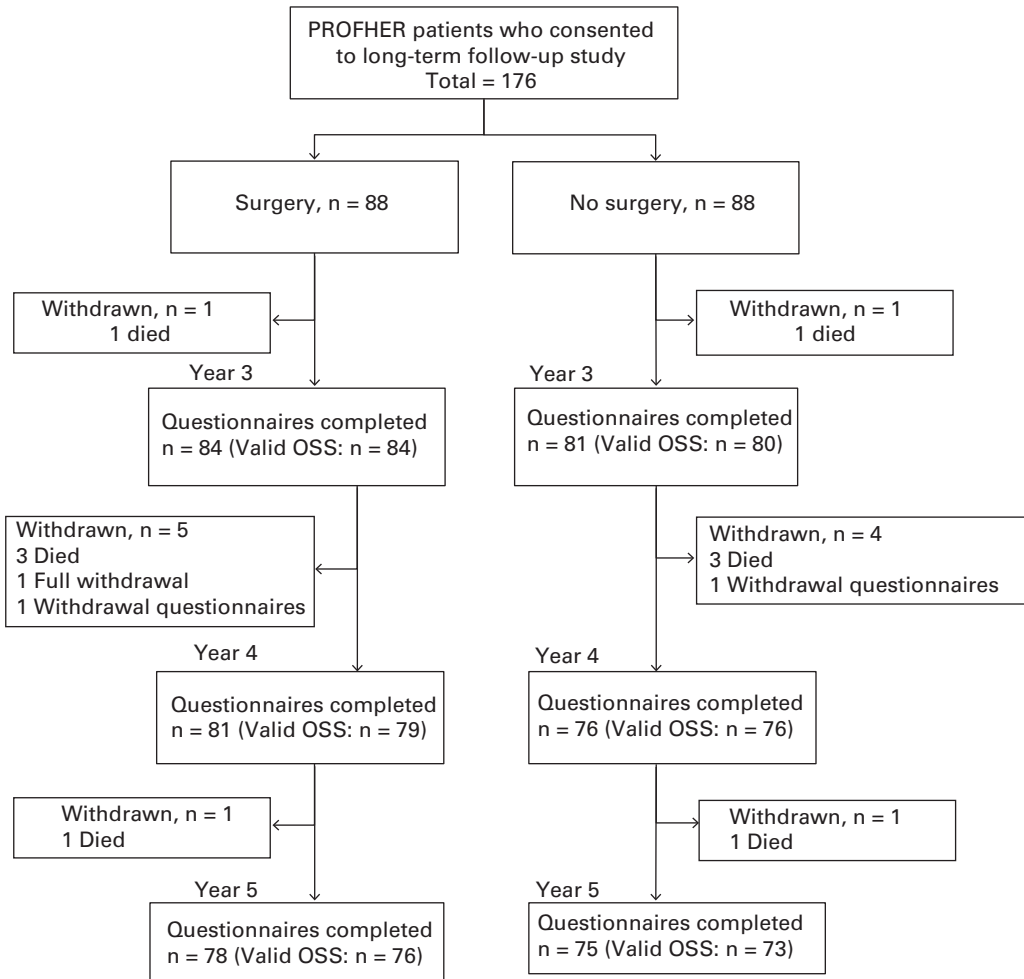


Fig. 1

Participant flow diagram.

assumed a 20% attrition rate at five years and based our proposal on a final sample size of 160 which would provide 71% power to detect a standard effect size of 0.4 using 5% significance level. Given the reduced statistical power for the extended follow-up, significance testing was limited to the primary outcome alone.

Statistical and economic analyses. All analyses were performed using Stata version 13.1 (StataCorp, College Station, Texas) and were on an intention-to-treat basis, participants being analysed in the groups to which they were randomised. Significance tests were two-sided at the 5% significance level.

Primary analysis. OSS data from the extended follow-up time points were added to the primary analysis model of the PROFHER trial.⁴ The analysis compared OSS data from the two treatment groups over all follow-up assessments using a multilevel regression model. In order to account for the correlation of outcomes over time from the same patients, time points were nested within patients. The model adjusted for the fixed effects of treatment group;

time (six months, one, two, three, four and five years); interaction between treatment group and time; tuberosity involvement at baseline (yes or no); age (< 65 years, ≥ 65 years), and gender and health status at baseline (EQ-5D-3L). The unstructured covariance pattern was retained from the primary analysis model. Patients with valid OSS data at one or more follow-up points for the standard or extended follow-up as well as complete covariate data were included in the analysis. Estimates of the difference in OSS between treatment groups, 95% confidence intervals (CI) and p-values were obtained for the extended follow-up at three, four and five years.

In a sensitivity analysis, the multilevel model was repeated substituting missing data with data derived by multiple imputation by chained equations. Missing outcome and covariate data were imputed from age, gender, tuberosity involvement, EQ-5D-3L index at baseline and available OSS data at other follow-up points.

Subgroup analyses. As with the main trial, the possibility of differential long-term treatment responses for older

Table II. Baseline characteristics (demographics) at randomisation and five years' follow-up

Characteristic	All randomised PROFHER patients		Patients with OSS data at 5 yrs	
	Operative (n = 125)	Non-operative (n = 125)	Operative (n = 76)	Non-operative (n = 73)
Gender				
Male, n (%)	28 (22.4)	30 (24.0)	19 (25.0)	15 (20.6)
Female, n (%)	97 (77.6)	95 (76.0)	57 (75.0)	58 (79.5)
Age (yrs)				
Mean (SD; range)	66.60 (11.80; 27.04 to 92.04)	65.43 (12.09; 24.63 to 89.02)	65.80 (10.12; 37.09 to 87.76)	65.33 (11.35; 31.33 to 84.56)
Median (IQR)	67.42 (61.73 to 75.48)	66.12 (58.09 to 74.34)	65.69 (61.98 to 73.47)	65.37 (57.60 to 74.41)
Age (group)				
< 65 yrs, n (%)	51 (40.8)	57 (45.6)	34 (44.7)	36 (49.3)
≥ 65 yrs, n (%)	74 (59.2)	68 (54.4)	42 (55.3)	37 (50.7)
Ethnicity				
White, n (%)	124 (99.2)	125 (100.0)	75 (98.7)	73 (100.0)
Black, n (%)	1 (0.8)	0 (0.0)	1 (1.3)	0 (0.0)
Education				
No formal qualifications, n (%)	66 (52.8)	68 (54.4)	35 (46.1)	35 (48.0)
Some qualifications but no degree, n (%)	47 (37.6)	43 (34.4)	34 (44.7)	25 (34.3)
Degree or higher, n (%)	12 (9.6)	14 (11.2)	7 (9.2)	13 (17.8)
Employment				
Part-time, n (%)	12 (9.6)	7 (5.6)	10 (13.2)	5 (6.9)
Full-time, n (%)	17 (13.6)	22 (17.6)	12 (15.8)	15 (20.6)
Self-employed, n (%)	1 (0.8)	3 (2.4)	0 (0.0)	3 (4.1)
Retired, n (%)	78 (62.4)	82 (65.6)	43 (56.6)	45 (61.6)
Not employed but seeking work, n (%)	3 (2.4)	1 (0.8)	2 (2.6)	1 (1.4)
Other, n (%)	12 (9.6)	9 (7.2)	9 (11.8)	3 (4.1)
Missing, n (%)	2 (1.6)	1 (0.8)	0 (0.0)	1 (0.7)
Diabetes				
Yes, n (%)	18 (14.4)	13 (10.4)	8 (10.5)	8 (11.0)
No, n (%)	106 (84.8)	111 (88.8)	67 (88.2)	64 (87.7)
Missing, n (%)	1 (0.8)	1 (0.8)	1 (1.3)	1 (1.4)
Smoking status				
Yes, n (%)	24 (19.2)	40 (32.0)	13 (17.1)	16 (21.9)
No, n (%)	96 (76.8)	81 (64.8)	61 (80.3)	55 (75.3)
Missing, n (%)	5 (4.0)	4 (3.2)	2 (2.6)	2 (2.7)
Steroid use				
Yes, n (%)	6 (4.8)	7 (5.6)	4 (5.3)	6 (6.9)
No, n (%)	118 (94.4)	116 (92.8)	72 (94.7)	67 (91.8)
Missing, n (%)	1 (0.8)	2 (1.6)	0 (0.0)	1 (1.4)
Health status (EQ-5D-3L Index)				
n	123	121	75	70
Mean (SD, range)	0.43 (0.37, -0.36 to 1)	0.38 (0.37, -0.35 to 1)	0.43 (0.36, -0.35 to 1)	0.34 (0.36, -0.35 to 1)
Median (IQR)	0.59 (0.09 to 0.73)	0.26 (0.07 to 0.66)	0.59 (0.08 to 0.69)	0.24 (0.07 to 0.66)

PROFHER, PROximal Fracture of the Humerus Evaluation by Randomisation trial; SD, standard deviation; IQR, interquartile range; EQ-5D-3L, EuroQol 5D-3L

patients (subgroups: < 65 years *versus* ≥ 65 years) and more complex fractures (subgroups: involvement of no tuberosities/one or both tuberosities) was explored. Expectations of the benefit of surgery over conservative treatment, established before the main trial results were known, were that this was greater in patients < 65 years and in patients with fractures involving one or both tuberosities,¹ and that these benefits might only emerge in the longer term.¹¹ Unadjusted mean OSSs by subgroups and treatment arm were therefore explored. Due to the substantially reduced statistical power for the subgroups, no statistical testing was performed.

Secondary outcomes. We calculated the annual and overall frequencies of shoulder surgery and fractures in each

treatment group that had occurred within the previous year. Extended follow-up data were combined with those of the main trial to establish the number of participants in each treatment group who had secondary shoulder surgery or a further fracture over five years. Free text providing details of further surgery and non-pre-specified fractures was categorised by two independent observers (HH and AR), who were blinded to the treatment group.

Economic analysis. The economic analysis aimed to explore whether the results from the PROFHER trial were sustained over a five-year time period by determining the between-group differences in HRQoL (measured via the EQ-5D-3L) at set times (three, four and five years) and

Table III. Baseline characteristics (fracture data) at randomisation and five years' follow-up

Characteristic	All randomised PROFHER patients		Patients with OSS data at 5 yrs	
	Operative (n = 125)	Non-operative (n = 125)	Operative (n = 76)	Non-operative (n = 73)
Time since injury (days)				
Mean (SD)	5.78 (4.90)	5.69 (4.89)	5.93 (5.17)	5.82 (4.59)
Median (IQR)	4 (0 to 19)	4 (0 to 21)	4.5 (0 to 19)	4 (0 to 18)
Affected shoulder				
Left, n (%)	57 (45.6)	68 (54.4)	32 (42.1)	40 (54.8)
Right, n (%)	68 (54.4)	57 (45.6)	44 (57.9)	33 (45.2)
Tuberosity involvement				
Yes, n (%)	99 (79.2)	94 (75.2)	58 (76.3)	58 (79.5)
No, n (%)	26 (20.8)	31 (24.8)	18 (23.7)	15 (20.6)
Tuberosity involvement (detail)				
Tuberosity not involved or missing, n (%)	26 (20.8)	31 (24.8)	18 (23.7)	15 (20.6)
Greater tuberosity, n (%)	58 (46.4)	61 (48.8)	34 (44.7)	36 (49.3)
Lesser tuberosity, n (%)	7 (5.6)	3 (2.4)	4 (5.3)	1 (1.4)
Greater and lesser tuberosity, n (%)	34 (20.8)	30 (24.0)	20 (26.3)	21 (28.8)
Fractures in the past 10 yrs				
Yes, n (%)	33 (26.4)	33 (26.4)	19 (25.0)	19 (26.0)
No, n (%)	92 (73.6)	90 (72.0)	57 (75.0)	53 (72.6)
Missing, n (%)	0 (0.0)	2 (1.6)	0 (0.0)	1 (1.4)
Previous surgery for fractures				
Yes, n (%)	8 (6.4)	12 (9.6)	3 (4.0)	9 (12.3)
No, n (%)	23 (18.4)	21 (16.8)	14 (18.4)	10 (13.7)
Missing, n (%)	2 (1.6)	0 (0.0)	2 (2.6)	0 (0.0)
No previous fractures, n (%)	92 (73.6)	92 (73.6)	57 (75.0)	54 (74.0)
Shoulder on dominant side				
Yes, n (%)	67 (53.6)	61 (48.8)	40 (52.6)	36 (49.3)
No, n (%)	56 (44.8)	62 (49.6)	34 (44.7)	35 (48.0)
Missing, n (%)	2 (1.6)	2 (1.6)	2 (2.6)	2 (2.7)
Injury mechanism				
Fall or trip from standing height or less, n (%)	90 (72.0)	96 (76.8)	55 (72.4)	58 (79.5)
Fall downstairs/steps or from a height, n (%)	18 (14.4)	17 (13.6)	11 (14.5)	9 (12.3)
Other, n (%)	15 (12.2)	9 (7.2)	8 (10.5)	5 (6.9)
Missing, n (%)	2 (1.6)	3 (2.4)	2 (2.6)	1 (1.4)

PROFHER, PROximal Fracture of the Humerus Evaluation by Randomisation; OSS, Oxford Shoulder Score; SD, standard deviation; IQR, interquartile range

examining how this difference evolved over time. We also planned to estimate costs of any further shoulder surgery and report these descriptively.

The methods used to process the EQ-5D-3L data and calculate quality-adjusted life years (QALY) scores were the same as those described in our previous cost-effectiveness report.⁶ Briefly, the EQ-5D-3L data were transformed into 'health-related quality of life weights' (utilities) using the United Kingdom general population tariff which assigns societal values to each health state.¹² QALYs were calculated by combining the utility estimates by the duration of time in each health state using the area under the curve method following the trapezium rule which assumed linear interpolation between follow-up points.¹³ A discount rate was applied to QALYs after 12 months, at an annual rate of 3.5%.¹⁴

In the main trial, the base-case analysis was conducted for the imputed dataset by means of multiple imputation with chained equations, using seemingly unrelated regression

analysis.⁶ This method accounts for the correlation between costs and effects from the same individuals and imputes the missing data. However, other regression-based methods are available for handling missing data in longitudinal studies, principally mixed models, and results may be sensitive to the methods used.¹⁵ A multilevel model similar to the primary OSS analysis was therefore conducted to investigate whether the results obtained in the main trial were robust to this alternative method of analysis. Therefore, the mean difference in utilities and QALYs (with 95% CIs) between the two groups was estimated using a multilevel model that adjusted for the fixed effects of treatment group, time (three and six months, one, two, three, four and five years), interaction between treatment group and time, tuberosity involvement at baseline, age, gender and baseline utility.

Uncertainty around the results was explored by means of sensitivity analysis that used multiple imputation by chained equations to replace missing data on QALYs in the

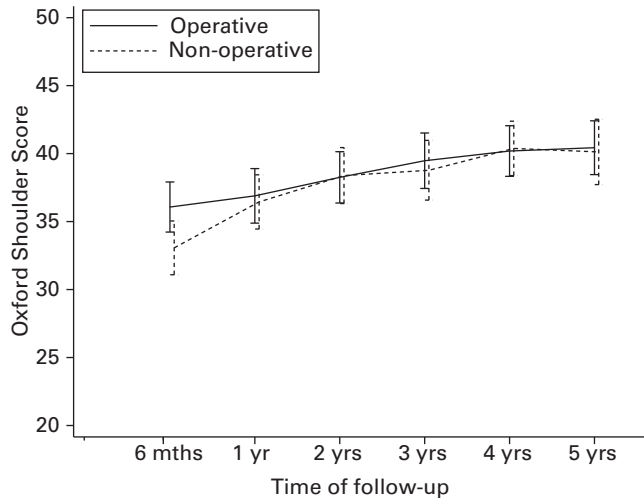


Fig. 2

Unadjusted mean Oxford Shoulder Scores (OSS) by allocated treatment (patients with available OSS only). Errors bars represent 95% confidence intervals.

multilevel model where missing outcome and covariate data were imputed from age, gender, tuberosity involvement, and baseline utility.

Results

Of 176 patients (81% of the 218 who returned questionnaires at two years; 70% of 250 randomised trial participants) who consented to long-term follow-up at two years after randomisation, valid OSSs were received for 164 (93%) at three years, 155 (88%) at four years, and 149 (85%) at five years' follow-up (Fig. 1). Retention was therefore slightly lower than anticipated in the extended follow-up. However, additional power was gained by the multilevel analysis. A total of ten patients died during the long-term follow-up, five in each trial arm.

As found at baseline (except for smoking status, which did not affect the OSS results) and two-year follow-up,⁴ patient characteristics were balanced between groups at five-year follow-up in the 149 patients with complete OSS data (Tables II and III). Furthermore, the characteristics of the RCT population remained representative, as none of the baseline characteristics differed meaningfully between participants at the start of the trial and those remaining at the end.

Primary outcome (OSS). Unadjusted OSS outcomes for patients with valid data were very similar in both groups for the extended follow-up period (Fig. 2). This featured a trend of small score increases between two and four years, with little difference in the fifth year. OSS scores were skewed towards maximum OSS shoulder function: over half the population had stable and satisfactory shoulder function³ at all three follow-up points: three years (median 42, interquartile range (IQR) 35 to 47.5); four years (median 43, IQR 37 to 48); five years (median 44, IQR 36 to 48).

When adding the long-term OSS follow-up data to the existing multilevel analysis, group differences were not statistically significant at any of the long-term follow-up time points. This was true for the primary analysis model including all patients with available outcome data at any time point as well as the sensitivity analysis including all patients using data derived by multiple imputation (Tables IV and V). None of the estimated mean differences was clinically meaningful; almost all were smaller than one OSS score point in magnitude with no consistent trend for the direction of the treatment effect.

The substantial overlap of the confidence intervals for the unadjusted OSS scores indicate that there were no marked differences between the treatment groups for the subgroups based on age (Fig. 3) or tuberosity involvement (Fig. 4). In both subgroups, the patterns of OSS score differences were not consistent with prior expectations.

Secondary outcomes. Only one patient reported further shoulder surgery during the extended follow-up period. This was a reverse shoulder replacement in year three in a non-operative group patient who had already undergone surgery (arthroscopic capsular release and subacromial decompression) during the main follow-up. Consequently, the number of patients who needed secondary surgery remained at 11 in each treatment group.⁴

A total of 81 further fractures were reported by 52 patients over the five-year follow-up period. A small number of fractures are likely to be duplicated from one year to the next but as this could not be known definitively, patient data were accepted as submitted, with the exception of one participant who provided the date of their fracture. There were more fractures in the non-operative group (50 fractures, 33 patients) than the operative group (31 fractures, 19 patients), especially of the spine and hip (Table VI).

Economic analyses. Inevitably, when compared with the 125 randomised into each treatment group, the extent of missing EQ-5D-3L data increased considerably in the extended follow-up period. For the 176 participants who consented to long-term follow-up, complete EQ-5D-3L scores were available for 159 (90%) at three years, 153 (86%) at four years and 151 (86%) at five years.

Figure 5 shows the distribution of mean utilities (EQ-5D-3L scores) for all the available patients across the five years for the two groups. Patients in the operative group started from a higher mean baseline utility (0.43; -0.36 to 1, operative *versus* 0.38; -0.35 to 1, non-operative). However, at the end of the second year there was little difference in EQ-5D-3L scores between treatment groups. This finding was consistent at three, four and five years with the 95% CIs overlapping at each assessment point. The same pattern applied for the analysis of utilities when adjusted for baseline utility or for all covariates (Table VII).

Between-group mean difference in QALYs based on individual patients' utilities are shown in Table VIII. At the end of the five years, patients allocated to the non-operative group generally had a marginally higher QALY gain than

Table IV. Extended primary analysis multilevel regression model of Oxford Shoulder Score (OSS).^a Mean OSS estimates, with 95% confidence intervals (CI), over time by treatment group and statistical significance of group differences

	Operative, mean (95% CI)	Non-operative, mean (95% CI)	Difference (95% CI)	p-value
Patients (n) ^{†‡}	114	117	231	
6 mths [‡]	37.84 (35.93 to 39.65)	35.59 (33.62 to 37.45)	2.25 (-0.07 to 4.57)	0.058
1 yr [‡]	39.23 (37.38 to 40.99)	38.80 (36.99 to 40.53)	0.42 (-1.78 to 2.63)	0.706
2 yrs [‡]	40.11 (38.24 to 41.90)	40.40 (38.59 to 42.13)	-0.29 (-2.53 to 1.95)	0.800
Patients (n) [†]	114	117	231	
3 yrs	40.53 (38.73 to 42.25)	40.36 (38.58 to 42.06)	0.17 (-2.02 to 2.35)	0.880
4 yrs	40.87 (39.04 to 42.62)	41.45 (39.67 to 43.16)	-0.58 (-2.81 to 1.64)	0.607
5 yrs	40.89 (39.99 to 42.70)	41.98 (40.14 to 43.74)	-1.09 (-3.41 to 1.23)	0.356

^a multilevel model of OSS (score range 0 to 48, higher scores indicate better outcomes) adjusted for treatment group, time (six, 12, 24, 36, 48 and 60 months), group×time interaction, baseline EuroQoL-5D-3L index, gender, age group (< 65 years/≥ 65 years) and tuberosity involvement at baseline (yes/no)

[†] number of patients included in the analyses (complete baseline characteristics and valid OSS score for at least one follow-up, same for primary and long-term analyses)

[‡] rows obtained from original PROximal Fracture of the Humerus Evaluation by Randomisation trial analysis

Table V. Multilevel regression model of Oxford Shoulder Score (OSS); data derived by multiple imputation:^a mean OSS estimates, with 95% confidence intervals (CI), over time by treatment group and statistical significance of group differences

	Operative, mean (95% CI)	Non-operative, mean (95% CI)	Difference (95% CI)	p-value
Patients (n) [†]	125	125	250	
6 mths [†]	37.96 (36.07 to 39.76)	35.67 (33.71 to 37.54)	2.28 (-0.04 to 4.61)	0.054
1 yr [†]	39.29 (37.48 to 41.03)	38.84 (37.03 to 40.56)	0.46 (-1.72 to 2.64)	0.680
2 yrs [†]	40.18 (38.36 to 41.93)	40.54 (38.72 to 42.28)	-0.36 (-2.58 to 1.87)	0.752
Patients (n)	125	125	250	
3 yrs	40.59 (38.79 to 42.31)	40.22 (38.46 to 41.91)	0.36 (-1.86 to 2.58)	0.748
4 yrs	40.97 (39.14 to 42.71)	41.52 (39.84 to 43.13)	-0.55 (-5.64 to 1.53)	0.602
5 yrs	40.96 (39.10 to 42.75)	41.90 (40.13 to 43.59)	-0.93 (-3.19 to 1.32)	0.416

^amissing OSS and covariate data derived by multiple imputation. Multilevel model adjusted for treatment group, time (six, 12, 24, 36, 48 and 60 months), group×time interaction, baseline EuroQoL 5D-3L index, gender, age group (< 65 years/ ≥ 65 years) and tuberosity involvement at baseline (yes/no)

[†] rows obtained from original PROximal Fracture of the Humerus Evaluation by Randomisation trial analysis

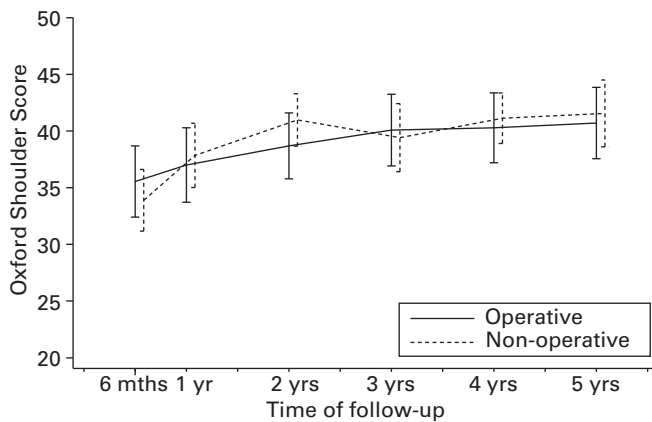


Fig. 3a

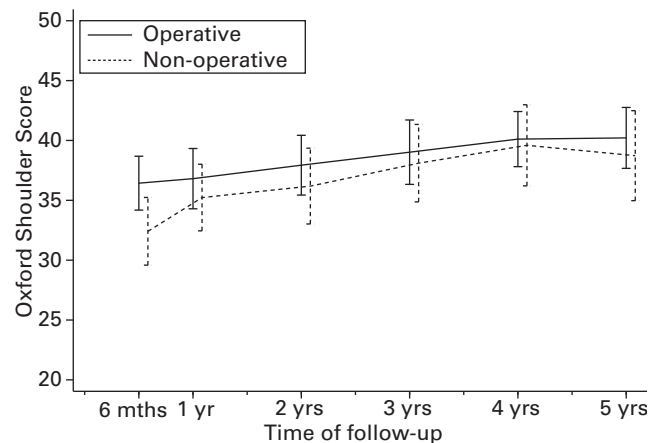


Fig. 3b

Unadjusted mean Oxford Shoulder Scores (OSS) by allocation and age group (patients with available OSS only): a) age < 65 years; b) age ≥ 65 years. Errors bars represent 95% confidence intervals.

patients allocated to the operative group. Hence the QALY gain for non-operative patients is maintained over time whether data are adjusted for baseline utility or for all covariates. The mixed model was repeated substituting missing data with data derived by multiple imputation by chained equations. For both analyses, there were negligible differences in the QALYs between the two groups at the different follow-up times (Table VIII).

Discussion

The extended follow-up found no statistically or clinically significant differences between operative and non-operative treatment of displaced fractures of the proximal humerus involving the surgical neck at three, four or five years in the OSS, our primary outcome. Nor was there any trend for group differences relating to age or fracture type.

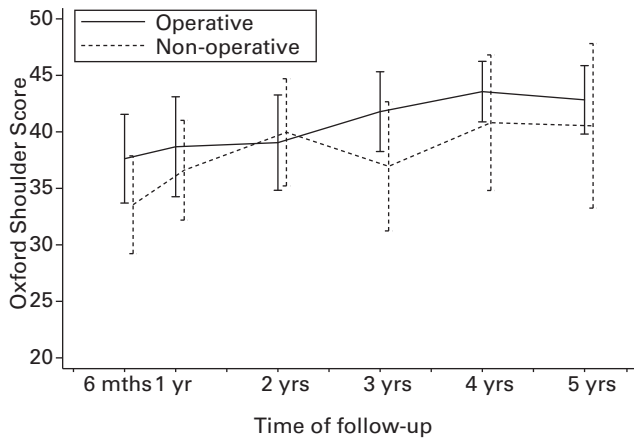


Fig. 4a

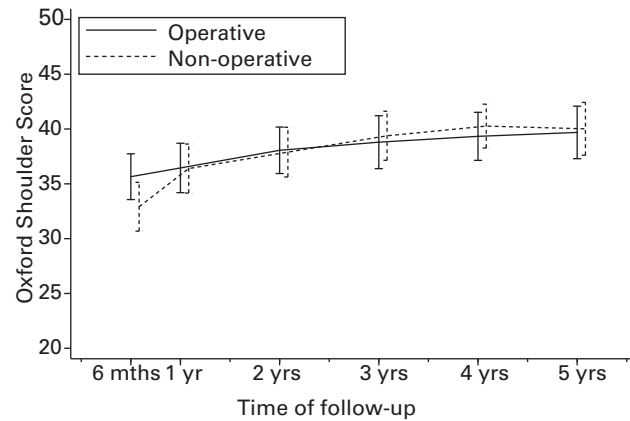


Fig. 4b

Unadjusted mean Oxford shoulder Scores (OSS) by allocation and tuberosity involvement group (patients with available OSS only): a) neither tuberosities involved; b) one or both tuberosities involved. Errors bars represent 95% confidence intervals.

Table VI. Further fractures by treatment arm

	Operative (n)			Non-operative (n)			Total (n)		
	M0 to 24	M24 to 60	Total	M0 to 24	M24 to 60	Total	M0 to 24	M24 to 60	Total
Shoulder/upper arm	1	5	6	2	4	6	3	9	12
Wrist	3	6	9	5	7	12	8	13	21
Hip	3	1	4	7	2	9	10	3	13
Spine	1	0	1	1	10	11	2	10	12
Elbow	0	0	0	1	2	3	1	2	3
Ankle	2	1	3	0	1	1	2	2	4
Other	0	8	8	2	6	8	2	14	16
Total fractures	10	21	31	18	32	50	28	53	81
Total patients	10	12	19	15	21	33	25	33	52

M0 to 24, follow-up up to two years; M24 to 60, extended follow-up from two to five years

These findings mirror those of the main trial.⁴ No trial participant had secondary shoulder surgery for a new complication during the extended follow-up period. The between-group differences in utilities, based on EQ-5D-3L data, at three, four or five years were very small: the 95% CIs overlapped at each assessment. The same lack of statistically significant between-group differences applied to the HRQoL analysis that showed the trend for a QALY gain for participants in the non-operative group was maintained over time. Sensitivity analyses indicated minimal differences between the two groups at each follow-up time.

By exceeding the original target of 200 participants at two-year follow-up, PROFHER was sufficiently powered at final follow-up. By contrast, we were 11 short of the 160 participants with OSS data at five years, and therefore did not meet the revised statistical power criteria for the extended follow-up. However, we believe this is unlikely to affect the validity of the results. First, loss to follow-up, including identical mortality (five in each group), was balanced in the two groups. Secondly, baseline characteristics

at five years were comparable between groups as well as being representative of the original population. Thirdly, much of the missing data were accounted for in the multi-level analysis, which included 231 patients. Fourthly, the between-group differences were small: the 95% CIs at each follow-up time were less than the minimal clinically important difference of five points. Fifthly, the between-group differences in the EQ-5D-3L were also very small, again reflecting comparability of the groups. Finally, there were no new complications warranting surgery.

Although there were no cost data to replicate the incremental cost-effectiveness analysis conducted for the PROFHER trial, the analyses of the health utility data for the five-year period produced results that are consistent with the main trial analysis:⁶ in general, patients allocated to surgery reported lower HRQoL. The very small differences in HRQoL between the two groups found for the mixed model and multiple imputation analyses indicate negligible differences in quality of life between the treatment groups. The costs of the only shoulder operation reported for the

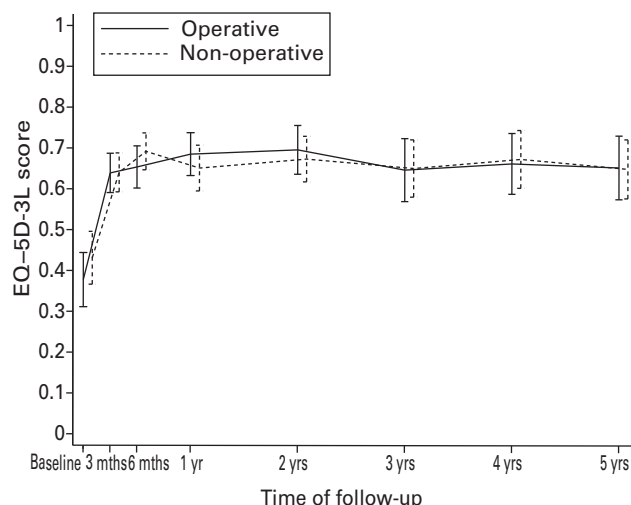


Fig. 5

Mean EuroQol-5D-3L (EQ-5D-3L) scores at baseline and follow-up points to five years. Error bars represent 95% confidence intervals.

Table VII. Multilevel regression model of EuroQol-5D-3L (EQ-5D-3L):* mean EQ-5D-3L estimates, and standard error of the mean (SEM) over time by treatment group and group differences, with 95% confidence intervals (CI)

	Operative mean (SEM)	Non-operative mean (SEM)	Difference (95% CI) (operative – non-operative)
Patients (n)	123	121	244
3 mths	0.61 (0.03)	0.60 (0.03)	0.01 (-0.06 to 0.08)
6 mths	0.66 (0.03)	0.63 (0.03)	0.03 (-0.04 to 0.10)
12 mths	0.63 (0.03)	0.66 (0.03)	-0.02 (-0.09 to 0.05)
2 yrs	0.66 (0.03)	0.66 (0.03)	-0.00 (-0.08 to 0.07)
3 yrs	0.65 (0.03)	0.63 (0.03)	0.02 (-0.06 to 0.10)
4 yrs	0.67 (0.03)	0.62 (0.04)	0.05 (-0.04 to 0.14)
5 yrs	0.65 (0.04)	0.62 (0.04)	0.03 (-0.07 to 0.13)

*multilevel model for EQ-5D-3L (score range 0 to 1, higher scores indicate better health related quality of life) adjusted for treatment allocation, time (three, six, 12, 24, 36, 48 and 60 months), group-time interaction, baseline EQ-5D-3L index, gender, age group and tuberosity involvement at baseline (yes/no). Number of patients included in the analyses (complete baselines characteristics and EQ-5D-3L score for at least one follow-up)

Table VIII. Health related quality of life. Mixed model and multiple imputation sensitivity analyses at each follow-up time up to five years

Follow-up	Mixed model* difference QALYs (adjusted for covariates) (operative – non-operative) (95% CI) (n = 200) †	Multiple imputation ‡ difference QALYs (adjusted for covariates) (operative – non-operative) (95% CI) (n = 250)
3 mths	-0.001 (-0.02 to 0.02)	-0.002 (-0.03 to 0.02)
6 mths	0.028 (-0.03 to 0.04)	-0.000 (-0.03 to 0.03)
1 yr	-0.004 (-0.06 to 0.05)	-0.004 (-0.06 to 0.05)
2 yrs	-0.031 (-0.15 to 0.09)	-0.024 (-0.15 to 0.10)
3 yrs	-0.061 (-0.25 to 0.12)	-0.034 (-0.23 to 0.16)
4 yrs	-0.063 (-0.32 to 0.19)	-0.027 (-0.29 to 0.24)
5 yrs	-0.042 (-0.36 to 0.28)	-0.013 (-0.35 to 0.32)

* multilevel model for quality-adjusted life years (QALYs) adjusted for treatment allocation, time (three, six, 12, 24, 36, 48 and 60 months), group-time interaction, baseline utility, gender, age group and tuberosity involvement at baseline (yes/no)

† number of patients included in the analyses (complete baselines characteristics and QALYs score for at least one follow-up): 106 operative; 94 non-operative

‡ missing Euroqol-5D-3L and covariate data derived by multiple imputation. Multilevel model adjusted for treatment group, time (six, 12, 24, 36, 48 and 60 months), group×time interaction, baseline utility, gender, age group (< 65 years/≥65 years) and tuberosity involvement at baseline (yes/no) CI, confidence interval

extended follow-up would not have affected the findings of the main trial.

We consider that it is unsafe to draw any conclusions from the observed differences in participants incurring further fractures between the two groups on the basis of treat-

ment group. We suggest that this is primarily a chance effect. In terms of known risk factors for fractures (such as higher age, female gender, previous fracture and smoking), the two groups were at similar risk of further fracture at baseline except for smoking status, where there was a

higher incidence of smokers in the non-operative group. This may partly explain a higher number of fractures in that group. Known inaccuracies, relating to both under- and over-reporting, of self-reported fractures¹⁶ are of some concern and indeed, based on additional participant commentary, we have confirmed one instance of duplicate reporting over time. We also have no information about whether there was any difference in the advice offered and medication provided for preventing further fractures in the two groups.

Our findings of an absence of treatment differences on the OSS in the extended follow-up underpin the main findings for the two-year follow-up. The only case of further surgery over the extended follow-up was further surgery for a patient who had already had surgery for a complication that occurred within the two-year follow-up.⁵ Given that most (15 of 22) secondary surgery occurred in the first year, this finding and the lack of difference in the OSS provide reassurance that late symptomatic complications are rare. The HRQoL results show that the PROFHER economic analysis was applicable over a five-year period. The overall OSS results show that most patients had attained satisfactory shoulder function by two years: this was subsequently sustained. Therefore, the two-year follow-up would have been sufficient for the PROFHER trial, and this finding could inform the length of follow-up for future RCTs on these fractures.



Take home message:

- The results of the extended follow-up underpin the main findings of the PROFHER trial.

- There was no significant difference in patient-reported outcome between operative and non-operative treatment for the majority of adults with proximal humeral fractures involving the surgical neck.

Author contributions:

H. Handoll: Advised on and contributed to methods and reporting throughout the trial, Wrote the first and revised drafts of the paper incorporating separate reports from AK and BC.

A. Keding: Trial statistician, Provided advice on methods, Produced and implemented statistical analysis plan, Contributed to the preparation of the paper.

B. Corbacho: Trial health economist, Produced the health economics analysis plan, Conducted the economics analysis, Contributed to the preparation of the paper.

S. Brealey: Trial manager, Advised on the design and coordinated the implementation of the extended follow-up including data collection, Contributed to the writing.

C. Hewitt: Statistician, Independently repeated primary analysis, Commented on paper.

A. Rangan: Chief investigator, Advised on the study design and on the clinical aspects of the analysis, Contributed to the writing.

We are grateful to the patients who generously completed questionnaires for the extended follow-up.

We thank R. Clarkson and L. Kottam (both at James Cook University Hospital, Middlesbrough, United Kingdom) for checking the Summary Care Records of patients for mortality, and staff at various participating sites for their help in locating missing patients.

This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment Programme (Project number: 06/404/53).

This paper presents independent research commissioned by the NIHR. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the UK National Institute for Health Research, the UK National Health Service, or the UK Department of Health.

The sponsor (Teesside University) managed the grant application process and monitored the study but it had no direct role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Although none of the authors has received or will receive benefits for personal or professional use from a commercial party related directly or indirectly to the subject of this article, benefits have been or will be received but will be directed solely to a research fund, foundation, educational institution, or other non-profit organization with which one or more of the authors are associated.

This is an open-access article distributed under the terms of the Creative Commons Attribution licence (CC-BY-NC), which permits unrestricted use, distribution, and reproduction in any medium, but not for commercial gain, provided the original author and source are credited.

This article was primary edited by G. Scott and first proof edited by A. C. Ross.

References

1. Handoll H, Brealey S, Rangan A, et al. Protocol for the PROFHER (PROximal Fracture of the Humerus: Evaluation by Randomisation) trial: a pragmatic multi-centre randomised controlled trial of surgical versus non-surgical treatment for proximal fracture of the humerus in adults. *BMC Musculoskelet Disord* 2009;10:140.
2. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg [Br]* 1996;78-B:593–600.
3. Dawson J, Rogers K, Fitzpatrick R, Carr A. The Oxford shoulder score revisited. *Arch Orthop Trauma Surg* 2009;129:119–123.
4. Rangan A, Handoll H, Brealey S, et al. Surgical vs nonsurgical treatment of adults with displaced fractures of the proximal humerus: the PROFHER randomized clinical trial. *JAMA* 2015;313:1037–1047.
5. Handoll H, Brealey S, Rangan A, et al. The PROFHER (PROximal Fracture of the Humerus: evaluation by Randomisation) trial - a pragmatic multicentre randomised controlled trial evaluating the clinical effectiveness and cost-effectiveness of surgical compared with non-surgical treatment for proximal fracture of the humerus in adults. *Health Technol Assess* 2015;19:1–280.
6. Corbacho B, Duarte A, Keding A, et al. Cost effectiveness of surgical versus non-surgical treatment of adults with displaced fractures of the proximal humerus: economic evaluation alongside the PROFHER trial. *Bone Joint J* 2016;98-B:152–159.
7. Handoll HH, Brorson S. Interventions for treating proximal humeral fractures in adults. *Cochrane Database Sys Rev* 2015;11:CD000434.
8. Hodgson SA, Mawson SJ, Saxton JM, Stanley D. Rehabilitation of two-part fractures of the neck of the humerus (two-year follow-up). *J Shoulder Elbow Surg* 2007;16:143–145.
9. van Reenen M, Oppe M. EQ-5D-3L User Guide. http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-3L_UserGuide_2015.pdf (date last accessed 29 November 2016).
10. Neer CS II. Displaced proximal humeral fractures. I. Classification and evaluation. *J Bone Joint Surg [Am]* 1970;52-A:1077–1089.
11. No authors listed. Protocol for an extended follow-up of patients for the PROFHER trial. <http://www.nets.nihr.ac.uk/projects/hta/0640453> (date last accessed 07 February 2017).
12. Dolan P, Gudex C, Kind P, Williams A. A social tariff for Euroqol: Results from a UK general population survey. Centre for Health Economics Discussion Paper Sep; 138, 1995. <http://www.york.ac.uk/che/pdf/DP138.pdf> (date last accessed 29 November 2016).
13. Billingham LJ, Abrams KR, Jones DR. Methods for the analysis of quality-of-life and survival data in health technology assessment. *Health Technol Assess* 1999;3:1–152.
14. No authors listed. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013 <http://www.nice.org.uk/article/pmg9/chapter/Foreword> (date last accessed 29 November 2016).
15. Peters SA, Bots ML, den Ruijter HM, et al. Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models. *J Clin Epidemiol* 2012;65:686–695.
16. Chen Z, Kooperberg C, Pettinger MB, et al. Validity of self-report for fractures among a multiethnic cohort of postmenopausal women: results from the Women's Health Initiative observational study and clinical trials. *Menopause* 2004;11:264–274.