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

Helen Handoll,[1]* Stephen Brealey,[2] Amar Rangan,[3] Ada Keding,[2] Belen Corbacho,[2] Laura Jefferson,[2] Ling-Hsiang Chuang,[4] Lorna Goodchild,[5] Catherine Hewitt[2] and David Torgerson[2]

[1]Health and Social Care Institute, School of Health and Social Care, Teesside University, Middlesbrough, UK
[2]York Trials Unit, Department of Health Sciences, University of York, York, UK
[3]Department of Trauma and Orthopaedics, James Cook University Hospital, Middlesbrough, UK
[4]Pharmerit Europe, Rotterdam, the Netherlands
[5]Department of Rehabilitation, James Cook University Hospital, Middlesbrough, UK

*Corresponding author

Declared competing interests of authors: Amar Rangan has obtained grants and personal fees from De Puy Ltd and grants from JRI Ltd; both are outside the submitted work. In addition, he has a UK and European patent application pending. None of these influenced the trial or the report.

Published March 2015
DOI: 10.3310/hta19240

Scientific summary

The ProFHER trial

Health Technology Assessment 2015; Vol. 19: No. 24
DOI: 10.3310/hta19240

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Proximal humeral fractures occur mainly in older adults and account for 5–6% of all fractures in adults. It is estimated that the substantial costs associated with these fractures are about one-fifth of those of hip fracture in the first year. Approximately 40% of these fractures are displaced fractures involving the surgical neck. As reflected in the variation in practice, the role of surgery for this group of fractures is unclear but its use is increasing.

Objectives

The primary objective was to evaluate the clinical effectiveness and cost-effectiveness of surgery compared with non-surgical treatment of the majority of displaced fractures of the proximal humerus involving the surgical neck in adults.

Secondary objectives included:

- to describe the study population in terms of key characteristics and in particular to categorise the trial fractures according to the Neer classification
- to ensure that both groups of patients received comparable and good standards of care
- to explore the effect of age (< 65 years vs. ≥ 65 years), type of fracture (involving none vs. one or both tuberosities) and patients' treatment preference on the primary outcome [Oxford Shoulder Score (OSS)].

Design

The PROximal Fracture of the Humerus: Evaluation by Randomisation (ProFHER) trial was a pragmatic parallel-group multicentre randomised controlled trial (RCT) with an economic evaluation. Patients were randomised on an equal basis to receive surgical or non-surgical treatment via a remote randomisation service. Randomisation was stratified by tuberosity involvement (yes or no) with allocation using random block sizes. Follow-up was for 2 years. Data collection, which included copies of baseline radiographs for all randomised patients, was by means of hospital and physiotherapy forms, and patient questionnaires. There was no blinding of outcome assessment. However, the independent classification of the baseline radiographs and all coding of data, which involved at least two independent raters, was carried out blind to treatment allocation.

Throughout the conduct of the trial we emphasised the importance of good practice from clinical and research perspectives; standardised protocols and care pathways; and comparable and sufficient expertise of care providers in the two groups. As far as possible we adopted procedures that reflected and were compatible with usual practice and that did not delay treatment, add unnecessarily to the clinical workload or place burden on participants.

Setting

Trial recruitment was undertaken in the orthopaedic departments of 33 of the 35 participating acute NHS hospitals; 33 participating hospitals were based in England and one each was based in Scotland and Wales. Patient care pathways included outpatient and community-based rehabilitation. Patients were recruited from fracture clinics or orthopaedic wards from 17 September 2008 up to 13 April 2011.

Participants

Adults (aged ≥ 16 years) presenting to the participating hospital within 3 weeks of their injury with a radiographically confirmed displaced fracture of the proximal humerus involving the surgical neck were considered for inclusion. Fractures that were displaced enough for the treating surgeon to consider surgical intervention were considered for inclusion, even if the degree of displacement did not meet the exact displacement criteria of the Neer classification (1 cm and/or 45° angulation of displaced parts). (We refer to fractures not meeting these criteria as 'one-part' fractures.) Exclusion criteria were associated dislocation of the injured shoulder joint; open fracture; lack of mental capacity to understand the trial or instructions for rehabilitation; comorbidities precluding surgery/anaesthesia; clear indication for surgery such as severe soft-tissue compromise; multiple injuries (upper limb fractures); pathological fractures (other than osteoporotic) and terminal illness; and participant not resident in the hospital catchment area. Patients were provided with an ethically approved information sheet and written consent was obtained before randomisation.

Interventions

The choice of surgical intervention was left to the treating surgeons, who used techniques that they were fully experienced with to avoid learning curve problems. Non-surgical treatment was initial sling immobilisation followed by active rehabilitation. Measures taken to ensure comparable delivery of care and rehabilitation for both groups of patients included the development and provision of an information leaflet on personal care during initial sling immobilisation, a physiotherapy protocol to promote standard rehabilitation care and promotion of the need for patients to perform home exercises.

Main outcome measures

The primary outcome measure was the OSS (scale 0–48, with higher scores indicating a better outcome) assessed at 6, 12 and 24 months. The trial was powered to detect a clinically important difference in the OSS of 5 points (equivalent to an effect size of 0.4).

Secondary outcomes were the 12-item Short Form health survey (SF-12) [physical and mental component summary scores (PCS; MCS) scale 0–100, with higher scores indicating better health], surgical and other shoulder fracture-related complications, secondary surgery to the shoulder or increased/new shoulder-related therapy, medical complications during the inpatient stay and mortality. European Quality of Life-5 Dimensions (EQ-5D) data and treatment costs incurred in hospital and subsequently were also collected.

Statistical analysis

Both the clinical effectiveness and economic analyses were conducted to a prespecified and externally endorsed analysis plan. Intention-to-treat analysis was conducted throughout.

The primary clinical effectiveness analysis compared the results of the two groups over all three follow-up assessments. A random slope multilevel model was fitted to the data with time points nested in patients to allow for clustering of data within each patient. This model adjusted for the fixed effects of treatment group, time (6, 12 or 24 months), the interaction between treatment and time, tuberosity involvement at baseline (yes or no), age (< 65 years or ≥ 65 years), gender and health status at baseline (EQ-5D). Different covariance patterns for the repeated measurements were explored and the most appropriate pattern selected for the final model. Estimates of the difference in OSS between treatment groups were assessed overall and at individual time points. All main analyses were performed using two-sided significance tests at the 5% significance level.

Two subgroup analyses, whose expected direction of effect was specified a priori, were planned for the OSS only. These were based on age (< 65 years or ≥ 65 years), with younger patients expected to benefit more from surgery, and tuberosity involvement (yes or no) recorded by the surgeon who assessed patient eligibility at baseline, with patients expected to benefit from surgery over non-surgical treatment when one or both tuberosities were involved. A sensitivity analysis of tuberosity involvement in terms of Neer one- and two-part fractures compared with Neer three- and four-part fractures was undertaken. This was based on the classification of the fractures by two independent experts using baseline radiographs. The primary analysis model was extended for each of these subgroup analyses by including, in turn, an interaction of treatment group with age group, tuberosity involvement or Neer parts. The differences in 2-log likelihood model fit between these and the base model were assessed using the chi-squared test at $p < 0.10$.

Reflecting a chance baseline imbalance, smoking status was added to the primary model as a sensitivity analysis. To ascertain the impact of prior beliefs, an interaction of treatment group with patient treatment preference was added to the base model. The variability in the OSS across recruitment sites was analysed descriptively.

The SF-12 PCS and MCS scores were analysed by multilevel modelling using the same predictors as in the OSS primary model. Frequencies of shoulder surgery- and fracture-related complications and any treatments for these were compared between treatment groups using the chi-squared test. Mortality rates were reported and compared between treatment groups using the chi-squared test.

Economic analysis

The economic analysis was carried out from the UK NHS and Personal Social Services perspective. Data on health utilities, obtained from the EQ-5D data collected from patient questionnaires, were converted into quality-adjusted life-years (QALYs) for each patient using the area under the curve method. Costs were expressed in UK pounds sterling at a 2012 price base.

Differences in mean costs and QALYs at 2 years were used to derive an estimate of the cost-effectiveness of surgery and non-surgical treatment. Multiple imputation was used to derive the data set for the base-case analysis. An additional analysis was conducted for complete cases (in which patients with any missing data are excluded). We used sensitivity analyses to explore the impact of incorporating both types of resource use (shoulder and non-shoulder related) and using different data sources. Cost-effectiveness acceptability curves (CEACs) were used to express the probability of whether or not surgery is cost-effective at the willingness-to-pay threshold.

Clinical effectiveness results

Of 1250 patients screened, 563 were considered eligible, of whom 250 were recruited. The mean age of the trial participants was 66 years and 192 (77%) were female. Independent characterisation of radiographs confirmed that the patients included in the ProFHER trial had sustained injuries that are typically considered for surgical intervention in contemporaneous practice. The assessment, based on the Neer classification, identified 18 one-part fractures, 128 two-part fractures and 104 three- or four-part fractures.

Of the 125 patients allocated to surgery, 109 received surgery. Consultant surgeons were predominantly involved in the management of these fractures and usually performed the operations. The choice of implants, with the majority being locking plates and a minority receiving hemiarthroplasty, is compatible with expectations of current practice. Of the 125 patients allocated non-surgical treatment, 123 received conservative care. Good and comparable non-surgical care in the two groups was achieved as evidenced

by (1) the provision of the ProFHER trial sling information leaflet to nearly all trial patients; (2) the equivalent provision of physiotherapy in terms of timing, numbers of sessions and interventions applied; and (3) comprehensive recording by physiotherapists showing that the majority of patients were doing their home exercises.

The treatment groups were balanced for all baseline characteristics, including Neer categories, except for smoking status. These profiles also applied to the baseline characteristics of the 215 patients (106 surgery patients vs. 109 non-surgery patients) with OSS data at 2 years.

The primary analysis included data from 231 patients (114 surgery patients vs. 117 non-surgery patients) with valid OSS data for at least one follow-up time point and complete baseline covariates. We found that there were no statistically significant differences in OSS between the two treatment groups over the 2-year period [difference of 0.75 points in favour of the surgery group, 95% confidence interval (CI) –1.33 to 2.84; $p = 0.479$] or at individual time points. All 95% CIs excluded the prespecified difference in OSS of 5 points, thus indicating that any observed difference was not of 'clinical significance'.

There was no statistically significant effect of treatment group when including interactions with age [comparison with base model: $\chi^2(1) = 1.24$, $p = 0.265$] or fracture type [assessed by tuberosity involvement at baseline; $\chi^2(1) < 0.01$, $p = 0.954$] and Neer classification [$\chi^2(1) = 0.05$, $p = 0.823$] in the two planned subgroup analyses. Thus, our prior expectations of directions of effect (subgroup differences) were not supported by these results and strengthen the case for not differentiating treatment (use of surgery) on the basis of these characteristics. Similarly, the effect of treatment group remained not statistically significant when accounting for smoking status and patient treatment preference.

We found no statistically significant differences between treatment groups in the SF-12 PCS score. The PCS score was on average 1.8 score points higher in the surgical group than in the non-surgical group (95% CI –0.84 to 4.39, $p = 0.184$). The same lack of significance applied for the SF-12 MCS score. The MCS score was on average 1.3 points lower in the surgical group than in the non-surgical group (95% CI –3.80 to 1.23, $p = 0.317$).

All 10 early medical complications occurred in surgery group participants. Slightly more patients in the surgery group than in the non-surgery group experienced a surgical or shoulder fracture-related complication [30 (24%) vs. 23 (18%)]. Although the same number in each group ($n = 11$; 9%) had secondary surgery to the shoulder within the 2-year follow-up period, slightly more in the surgery group than in the non-surgery group had increased or new shoulder-related therapy [seven (5.6%) vs. four (3.2%)]. Neither of these differences were statistically significant (patients with complications $p = 0.279$; patients undergoing further surgery or therapy $p = 0.575$). The same finding applied to mortality, with slightly more deaths in the surgery group than in the non-surgery group [nine (7.2%) vs. five (4.0%), $p = 0.271$]. In total, 28 patients in each group experienced at least one serious adverse event (SAE). Nine patients experienced at least one non-SAE, of whom four were in the surgery group and five were in the non-surgery group.

Cost-effectiveness results

The base-case economic analysis showed that, at 2 years, the cost of surgical intervention was, on average, £1780.73 more per patient (95% CI £1152.71 to £2408.75) than the cost of non-surgical intervention. It was also slightly less beneficial in terms of utilities, although this difference was not statistically significant (difference of –0.0158, 95% CI –0.13 to 0.10 when adjusted for baseline utility; –0.0101, 95% CI –0.13 to 0.11 when adjusted for covariates). The net monetary benefit associated with surgery is negative, indicating that the resources to be displaced would be greater than the benefit to be gained if surgery were implemented in the NHS. Furthermore, the CEAC showed that surgery had only a 5% probability of achieving the criterion of costing < £20,000 to gain a QALY. Therefore, surgery appears to

be a dominated treatment option and not a cost-effective use of health-care resources. These findings were robust to the three sensitivity analyses that were undertaken. Although surgery did not result in a dominated alternative for the base-case analysis or the sensitivity analyses that included both shoulder and non-shoulder-related resource use, the incremental cost-effectiveness ratios were still above National Institute for Health and Care Excellence cost-effectiveness thresholds (£20,000–30,000 per QALY gained).

Conclusion

The ProFHER trial has provided robust clinically relevant evidence showing that current surgical practice does not result in a better outcome for most patients with a displaced fracture of the proximal humerus involving the surgical neck and that, in addition, it is not cost-effective in the UK setting.

It is important that non-surgical care should be of a good standard, including the availability of a leaflet about sling immobilisation, timely access to physiotherapy and promotion of home exercises. The potential need for remedial surgery for severe symptomatic complications in around 5–10% of these patients should be factored into forecasts for hospital budgets.

Recommendations for research

Given the above findings and the existence of five ongoing trials, initiating further RCTs on this question is not appropriate. The setting up of a national database of these fractures, with the systematic and prospective collection of data on epidemiology, management and outcome, including of patient-reported outcomes, should be considered. In addition, research is required to establish the best approach to providing patient information on the early treatment of these fractures.

Trial registration

This trial is registered as ISRCTN50850043.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 5.116

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 06/404/53. The contractual start date was in March 2008. The draft report began editorial review in December 2013 and was accepted for publication in June 2014. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).