
Article:
Knowlson, Catherine, Brealey, Stephen Derek orcid.org/0000-0001-9749-7014, Keding, Ada orcid.org/0000-0002-1182-887X et al. (2 more authors) (2023) Examination of early treatment effects and related biases during the conduct of two UK-wide pragmatic orthopaedic surgical trials:ProFHER and UK FROST. Bone & Joint Open. pp. 96-103. ISSN: 2633-1462

https://doi.org/10.1302/2633-1462.42.BJO-2022-0139

Follow us @BoneJointOpen

Freely available online OPEN ACCESS

BJO

GENERAL ORTHOPAEDICS

Examination of early treatment effects and related biases during the conduct of two UK-wide pragmatic orthopaedic surgical trials: ProFHER and UK FROST

C. N. Knowlson,
S. Brealey,
A. Keding,
D. Torgerson,
A. Rangan

From University of York Trials Unit, York, UK

Aims

Early large treatment effects can arise in small studies, which lessen as more data accumulate. This study aimed to retrospectively examine whether early treatment effects occurred for two multicentre orthopaedic randomized controlled trials (RCTs) and explore biases related to this.

Methods

Included RCTs were ProFHER (PROximal Fracture of the Humerus: Evaluation by Randomisation), a two-arm study of surgery versus non-surgical treatment for proximal humerus fractures, and UK FROST (United Kingdom Frozen Shoulder Trial), a three-arm study of two surgical and one non-surgical treatment for frozen shoulder. To determine whether early treatment effects were present, the primary outcome of Oxford Shoulder Score (OSS) was compared on forest plots for: the chief investigator's (CI) site to the remaining sites, the first five sites opened to the other sites, and patients grouped in quintiles by randomization date. Potential for bias was assessed by comparing mean age and proportion of patients with indicators of poor outcome between included and excluded/non-consenting participants.

Results

No bias in treatment effect was observed overall for the CI site, or the first five sites, compared with the remaining sites in either trial. An early treatment effect on the OSS was observed for the first quintile of participants recruited to ProFHER only (clinically relevant difference of seven points). Selection bias for age was observed in the ProFHER trial only, with slightly younger patients being recruited into the study. Both trials showed some selection bias for markers of poor prognosis, although these did not appear to change over time.

Conclusion

No bias in treatment effects overall were found at the CI or early sites set-up. An early treatment effect was found in one of the two trials, which was likely a chance effect as this did not continue during the study. Selection bias was observed in both RCTs, however this was minimal and did not impact on outcome.

Cite this article: *Bone Jt Open* 2023;4-2:96–103.

Keywords: Randomized controlled trials, Orthopaedic surgery, Selection bias, Performance bias

Introduction

Randomized controlled trials (RCTs) are the gold standard for evaluating the effectiveness of an intervention.[1] Their validity depends on a number of factors including an adequate sample size to achieve an appropriate level of statistical power and an appropriate methodology to minimize bias.[1] Sample size is important because there is evidence that small studies can find large treatment effects, but when additional trials are performed and combined in a meta-analysis, the effect sizes usually become smaller.[2] This may be because of spurious findings in small studies so that there is regression to the mean as more studies are conducted.[3] Alternatively,

Correspondence should be sent to Stephen Brealey; email: stephen.brealey@york.ac.uk

doi: 10.1302/2633-1462.42.BJO-2022-0139

Bone Jt Open 2023;4-2:96–103.

these effects could be due to bias at the few participating sites, such as: preference of clinicians for one treatment over the other; bias in the selection of which patients to recruit; awareness of treatment allocation by the patient or clinician; or the rate of participant withdrawal.[4,5] Single-centre studies, or the site at which the Chief Investigator (CI) is based in a multicentre study, may be at increased risk of bias as for various reasons. They are usually more engaged, as shown by the higher recruitment rate compared to the other sites.[6]

Surgical trials can be particularly at risk of bias, due to difficulties with surgeon equipoise.[7-9] Surgeons may have a preference for what they believe is the superior treatment, or they may feel conflicted over the safety/eligibility of individuals.[7,8,10] For example, in trials involving fractures, surgeons have been reported to be less willing to randomize patients with more complex fractures.[7] Additionally, some surgeons may feel uncomfortable describing to patients that the best treatment is not known,[8] as this conflicts with having confidence in the right option for each patient in routine clinical practice.[9] Consequently, this can introduce selection bias as to who is enrolled into the study,[9] and performance bias because of one group getting more attention than another.[11] Pragmatic surgical trials are particularly susceptible to these biases as blinding is often impractical.[12]

Risk of bias is routinely assessed in RCTs using standardized tools such as the Cochrane Risk of Bias assessment tool (RoB2) during systematic review.[13] This type of assessment is done after the trial results are reported and is based on the overall conduct of the study. We considered it could be useful to examine whether there was any evidence of unconscious bias during the conduct of a trial using data that would not routinely be included in a trial report; to our knowledge, this has not been considered before. This could be useful for trial teams to detect whether there are early treatment effects due to bias that could affect the overall treatment effect at the end of the study. This could have implications for decisions about stopping a study as trial data accumulate, which will affect the future evidence base to guide policy and practice in that clinical setting.[14]

We hypothesized that if the adequately powered pragmatic orthopaedic surgical trials ProFHER (PROximal Fracture of the Humerus: Evaluation by Randomisation)[15] and UK FROST ((United Kingdom Frozen Shoulder Trial)[16] had been conducted as small single site studies with fewer patients, there may have been an exaggerated treatment effect, as opposed to the actual results of no difference in outcome between treatments. Furthermore, if an exaggerated treatment effect was present, we hypothesized that this could either be due to spurious findings that are diminished by regression to the mean as more sites/patients get involved, or due to bias at the start of the trial. This was explored in the context of the early sites set-up compared with the remaining sites, namely the CI's site which is the lead site to host the research, and the first five sites to open, which represent the early adopters of the trial. This would also allow for an exploration of a biased treatment effect in smaller samples. The presence of early bias in treatment effect as patients were recruited into the trial was also explored.

Methods

Summary of UK FROST and ProFHER. UK FROST was a three-arm RCT which randomized adult participants with a frozen shoulder to manipulation under anaesthesia, arthroscopic capsular release, or early structured physiotherapy in a 2:2:1 ratio.[16] ProFHER was a two-arm RCT which randomized adult participants with displaced fractures of the proximal humerus to surgery or sling immobilization in a 1:1 ratio.[15] These two RCTs had a large number of participating sites and included interventions that compared different types of surgery, or surgery with non-surgical treatment. UK FROST recruited from 35 NHS hospitals between April 2015 and December 2018 (45 months), and ProFHER recruited from 33 NHS hospitals between September 2008 and April 2011 (32 months).

Statistical analysis. To determine if there was an exaggerated treatment effect at the CI's site, the outcome data for participants recruited at this site for each trial were averaged, as were the data for participants from the remaining sites. For both trials, the primary outcome was the Oxford Shoulder Score (OSS),[17] a shoulder-specific patient-reported outcome measure (total scores of 0 (worst outcome) to 48 (best outcome)), and therefore was the outcome used for analysis in this study at the one-year follow-up. The target difference between treatment groups for the two trials was set at a threshold of five or four points when testing for differences between surgery and non-surgical options or between surgical options, respectively. Data were analyzed by forest plot. To test for the presence of an early bias in treatment effect over time, the mean outcome for the quintiles of randomized patients was calculated and analyzed by forest plot. For both, a fixed effects model was used and the I^2 value to determine heterogeneity. As UK FROST had three arms, separate analyses were carried out to compare all treatments. Review Manager (RevMan) 5 was used to undertake these analyses. This was repeated for the first five sites open compared with the remaining sites.

To examine the presence of selection bias, we explored whether there were differences in age or predictors of poor outcome between the patients who were randomized and those who either did not consent or were ineligible to take part. For age, the mean and standard deviation (SD) were calculated for the trial participants and for patients who were ineligible, eligible but did not consent, and the latter groups combined. For

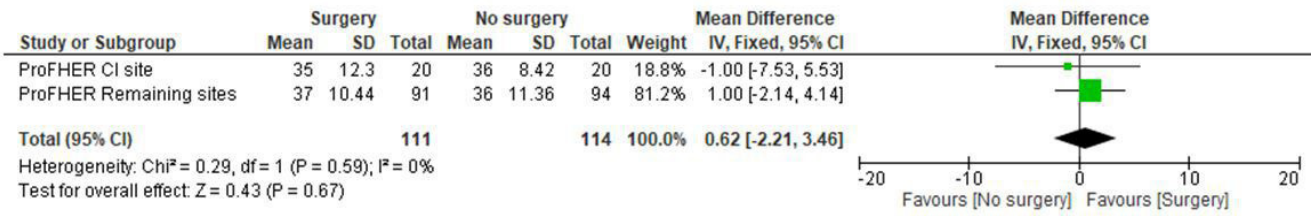

Fig. 1

Forest plot of mean Oxford Shoulder Score, comparing the chief investigator's site to the remaining sites of the ProFHER trial. CI, confidence interval; IV, inverse variance; SD, standard deviation.

Table I. Summary of heterogeneity in outcome comparisons.

Study name	Comparison by forest plot	Heterogeneity (I^2 value)
ProFHER	Patients from CI site vs the remaining sites	0%
	Patients from first five sites vs the remaining sites	0%
	Quintiles of patients recruited	38%
UK FROST	Patients from CI site vs the remaining sites – ACR vs MUA	0%
	Patients from CI site vs the remaining sites – ESP vs ACR	50%
	Patients from CI site vs the remaining sites - ESP vs MUA	0%
	Patients from first five sites vs the remaining sites – ACR vs MUA	44%
	Patients from first five sites vs the remaining sites – ESP vs ACR	0%
	Patients from first five sites vs the remaining sites - ESP vs MUA	0%
	Quintiles of patients recruited – ACR vs MUA	0%
	Quintiles of patients recruited – ESP vs ACR	22%
	Quintiles of patients recruited – ESP vs MUA	24%

ACR, arthroscopic capsular release; CI, chief investigator; ESP, early structured physiotherapy; MUA, manipulation under anaesthesia.

the ProFHER trial, the predictor of poor outcome was whether either tuberosity (a rounded prominence) of the humeral bone was involved in the fracture;[15] for UK FROST it was diabetic status.[16] The percentage of individuals who had tuberosity involved or were diabetic, for the respective trials, was calculated for the following groups: trial participants, ineligible patients, eligible but non-consenting patients, and the latter groups combined. To assess whether these changed over time, the participants were ordered by randomization date and split into quintiles (i.e. five equal groups). Each group was analyzed as above. The non-consenting and ineligible patients were combined and ordered by date of eligibility so that the quintiles matched the same time periods as the recruited group.

Results

For the ProFHER trial, the OSS did not differ significantly between treatment groups for participants recruited by the CI's site or for the remaining sites (Figure 1, Table I). There was no difference in OSS between treatment groups for the analyses of the first five sites or at the remaining sites (Supplementary Figure a). Neither were there significant differences in treatment outcomes when comparing the CI site, or first five sites, with the remaining sites.

In UK FROST, participants from the CI sites showed no significant difference in OSS between any of the three treatments compared to participants from the other sites (Figure 2, Table I). The same applies to the analyses of the first five sites opened to recruitment compared to the remaining sites (Supplementary Figure b). When comparing treatment differences between the CI site, or the first five sites, with the remaining sites there was a four- and three-point favourability of early structured physiotherapy (ESP) compared with arthroscopic capsular release (ACR) at the CI site, respectively. Data, however, were only available for a small number of participants at this site who received ESP (three participants). No differences in the OSS in any of the comparisons exceeded the targets set for being clinically important.

The mean OSS in the quintiles of participants recruited to ProFHER grouped in the order of randomization showed that there was heterogeneity beyond chance (Figure 3, Table I). The largest difference was the first quintile of participants recruited, who had favourable outcome with surgery of seven OSS points (Figure 3). The second largest difference was of four OSS points for the fourth quintile and favoured no surgery (Figure 3).

In UK FROST, there was no heterogeneity between the quintiles of participants when comparing the surgical options to each other (Figure 4a). There was, however, some heterogeneity when comparing the non-surgical arm (ESP) to either of the surgical options (Figures 4b and 4c). Over time (from the first to fifth quintile) there

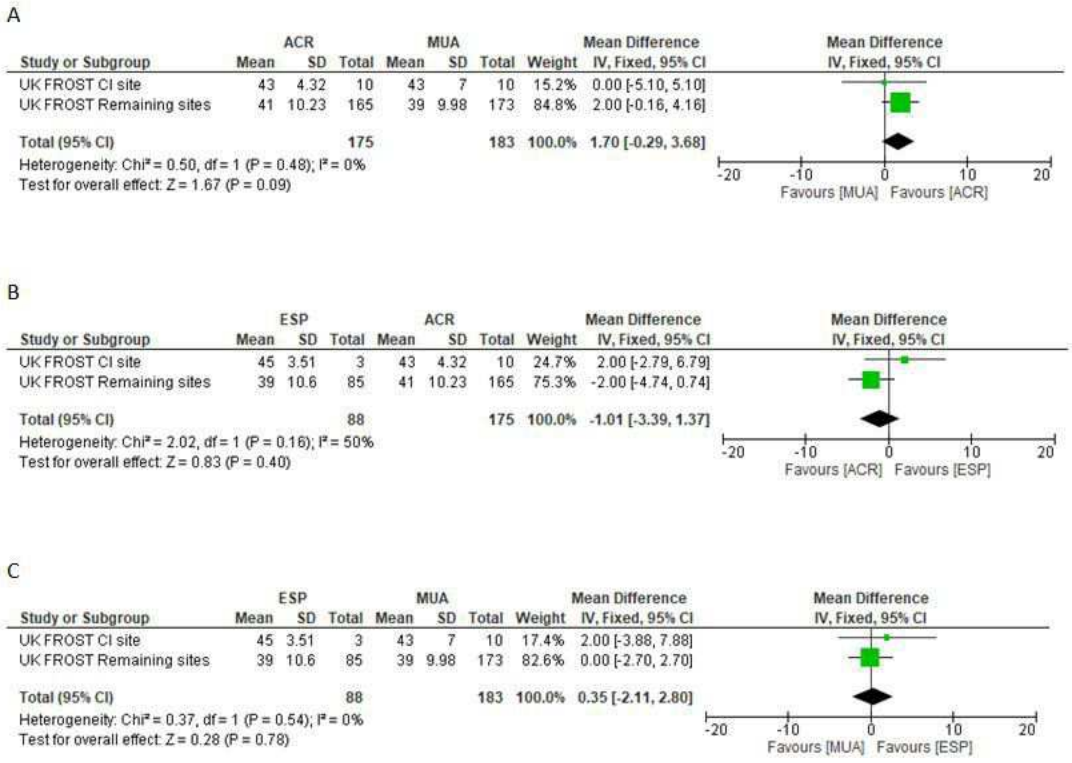

Fig. 2

Forest plot of mean Oxford Shoulder Score, comparing the chief investigator's site to the remaining sites of the UK FROST trial. ACR, arthroscopic capsular release; CI, confidence interval; ESP, early structured physiotherapy; IV, inverse variance; MUA, manipulation under anaesthesia; SD, standard deviation.

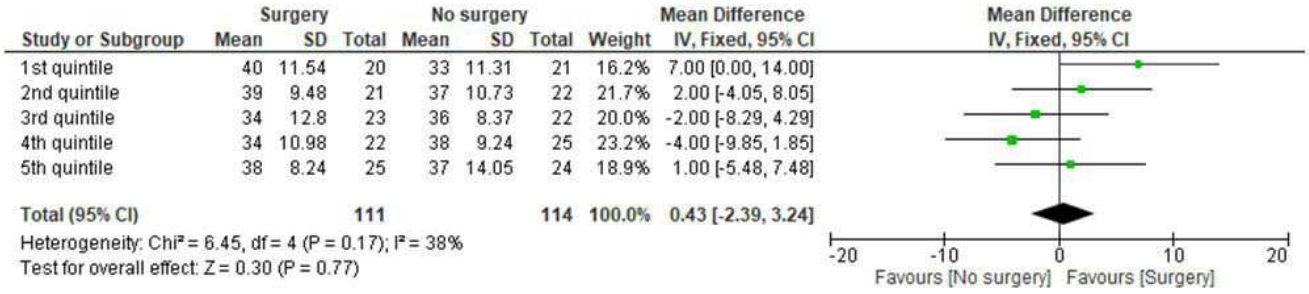

Fig. 3

Forest plot of mean Oxford Shoulder Score, comparing quintiles of participants in order of recruitment to the ProFHER trial. CI, confidence interval; IV, inverse variance; SD, standard deviation.

was a slight trend observed from favouring of non-surgical treatment (ESP) to surgical treatment (either MUA or ACR; Figures 4b and 4c).

For UK FROST, ages were similar between patients who were and were not randomized, and when categorized as non-consenting and ineligible patients, differences were still negligible (Table I). For ProFHER, patients who were not randomized were slightly older than those who were randomized. Patients considered for ProFHER but who were not randomized were less likely to have tuberosity involved in the fracture (a predictor of poor outcome)[15] than those who were randomized, especially for patients not randomized due to ineligibility (Table II). For UK FROST, patients who were not randomized, either due to ineligibility or because they did not consent, were more likely to be diabetic (a predictor of poor outcome)[16] than patients who were randomized (Table II). For ProFHER, 55% (313/563) of eligible patients declined to consent, whereas for UK FROST this was 37% (298/801) (Table II).

The mean age of participants (with error bars for SD) considered for ProFHER and UK FROST did not change over time for the randomized compared with not randomized groups, i.e. those patients who were ineligible or did not consent (Figures 5a and 5b). The percentage

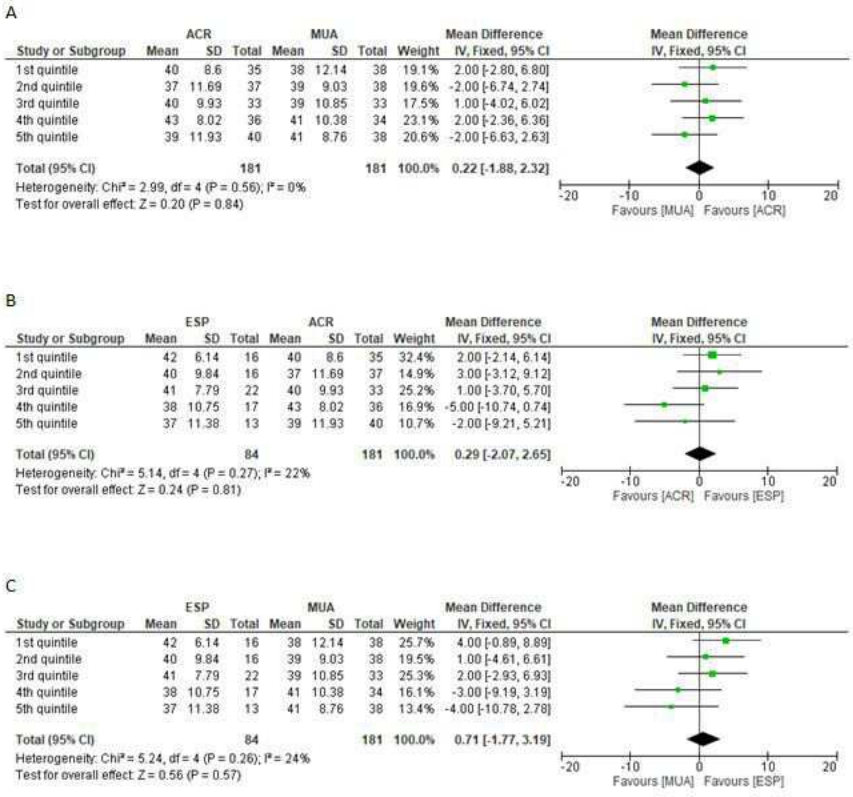

Fig. 4

Forest plot of mean Oxford Shoulder Score, comparing quintiles of participants in order of recruitment to the UK FROST trial. ACR, arthroscopic capsular release; CI, confidence interval; ESP, early structured physiotherapy; IV, inverse variance; MUA, manipulation under anaesthesia; SD, standard deviation.

of patients with tuberosity involvement considered for ProFHER varied between quintiles of patients ordered by date randomized, but there was no trend observed over time. (Figure 5a). The same applies for the percentage of patients with diabetes who were considered for UK FROST (Figure 5).

Discussion

This study explored whether biased treatment effects were observed in the CI's site or first five sites that opened to recruitment, namely the early sites set-up (and with smaller sample size compared with overall), or in the quintiles of patients recruited over time to either the ProFHER[15] or UK FROST[16] trial. No clinically relevant differences in treatment outcome (the OSS) were found for the CI site, or first five sites, compared to the remaining sites for either ProFHER or UK FROST. For UK FROST there was also no difference in outcome when comparing the first quintile of participants randomized to the rest. However, an early treatment effect was observed in favour of surgery for ProFHER in the first quintile of patients. There was more heterogeneity in treatment effect over time when comparing non-surgical treatment with surgery in both trials than when comparing two different types of surgery in UK FROST. Overall, patients enrolled to ProFHER were slightly younger. In both trials, patients with a key marker for a poorer prognosis were more likely to be recruited. The difference in age and prognostic markers between the two groups did not change over time.

Regression to the mean may explain why trials with only a small number of participants are more likely to show large treatment effect size.[2,18] When larger trials are included in meta-analyses of trials evaluating the same intervention, the overall effect size lessens compared with that of the smaller trials.[2] Furthermore, it has been observed that for 43% of trials with a large effect size, subsequent trials with larger sample sizes either found no statistically significant benefit or had smaller effect sizes.[18] These effects could be due to the small sample size causing imprecise or chance results so that regression to the mean is observed in subsequent trials.[2] This could explain the clinically significant treatment effect early on in the ProFHER trial in favour of surgery, as differences in either age or tuberosity involvement did not appear to change over time in those who were or were not randomized. Overall, however, there was greater heterogeneity when comparing surgical to non-surgical approaches than when comparing different types of surgery in this study. This could be explained by smaller sample sizes in the quintiles when comparing surgery with non-surgical

Table II. Mean age of randomized and non-randomized patients and the proportion with predictors of poor outcome.

Trial	Mean age, yrs (SD)	Tuberosity involved, n (%)
ProFHER		
Randomized (n = 250)	65.5 (12.0)	193/250 (77.2)
Not randomized (n = 1,000)	69.3 (15.5)	640/1,000 (64.0)
Non-consenting (n = 313)	67.9 (13.1)*	228/313 (72.8)
Ineligible (n = 687)	69.9 (16.4)†	411/687 (59.8)
UK FROST	**Mean age, yrs (SD)**	**Diabetic (%)**
Randomized (n = 503)	54.3 (7.7)	150/503 (29.8)
Not randomized (n = 411)	53.4 (8.5)	103/411 (25.1)
Non-consenting (n = 298)	53.5 (7.9)	74/298 (24.8)
Ineligible (n = 113)	53.3 (9.9)	29/113 (25.7)

*Data available for 297 patients.
†Data available for 643 patients.
SD, standard deviation.

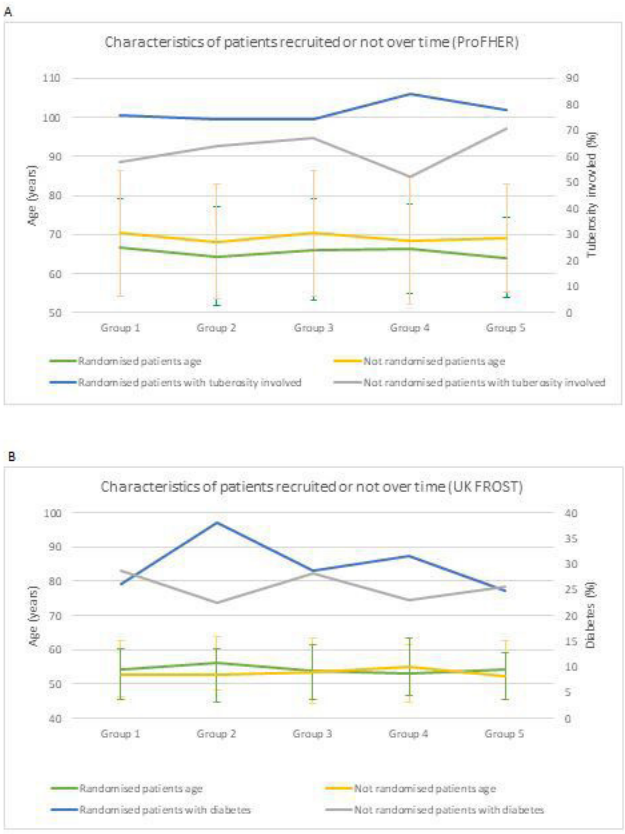

Fig. 5
Mean age and prevalence of predictors of poor outcome in participants considered for the a) ProFHER trial and b) UK FROST trial over time, comparing those who were randomized to those who were not.

approaches, leading to more evidence of sampling error than necessarily bias in the treatment effect. The trend for UK FROST was a move to favouring both surgical arms over time. This could suggest that the ESP may not have continued to be implemented to the same extent over time due to possible resource constraints.

We hypothesized that sites may inadvertently avoid recruiting participants who were older or had indicators of poor prognosis, in favour of not leaving the treatment decision to chance. Regarding participant age, this only appeared to occur for ProFHER, although the age difference was minimal. This could be because the participants were generally older than UK FROST participants, and therefore more likely to have comorbidities that are associated with a higher risk of complications from surgery.[19,20] However, the opposite was observed for indicators of poor prognosis, as participants in each trial were marginally more likely to have predictors of poor outcome (fractures with tuberosity involved for ProFHER and diabetes for UK FROST) than the patients who were not randomized. For ProFHER, the increased recruitment of patients with more complex fractures (i.e. with tuberosity involved) contrasts with the findings of Phelps et al[7] for a surgical trial on fractures of the distal femur. The reason given by surgeons in the Phelps et al[7] study was that surgeons felt one of the arms was not appropriate for the more complex fractures. The same could apply here, if the ProFHER surgeons felt that the lack of tuberosity involved in the proximal humerus fracture made surgery an unnecessary risk, as simpler fractures may be considered easier to treat with non-surgical care. As for UK FROST, a possible reason for being more likely to recruit patients with diabetes is that they have a more resistant frozen shoulder and are therefore more suitable for the trial.[21] It is reassuring, however, that there was no trend over time during trial recruitment for either tuberosity involvement or diabetes that suggests an increase or decrease in trust of the trial treatments as the trial progressed. Selection bias is also not limited to the surgeon, as patients have preferences too. This was previously analyzed for the ProFHER trial, for which non-consent could be attributed to patient preference more frequently than surgeon preference. More than half of the eligible patients who did not consent preferred surgery, whereas 17% were not enrolled due to lack of surgeon equipoise.[22]

This study explored evidence of bias in early treatment effects in terms of the early sites set-up and recruitment of patients over time during the conduct of two trials after they were completed. This could be due to unconscious bias, such as a change in surgeon preference as they become more experienced with an intervention and the protocol requirements of the study. Exploring the potential for bias is important for confirming the validity of a trial's findings and is usually done at the end of a study or systematic review. It could, however, be useful to monitor evidence of bias during a trial to give oversight committees the opportunity to find ways to reduce it. For example, collecting a minimal amount of data from screened patients, including age and important prognostic factors, allows checks to be made as to whether

patients who are not randomized differ significantly from those who are. Comparing outcomes between sites or groups of patients as the trial progresses could be more difficult, especially if outcome data are collected a long time after randomization, as by the time a trend is identified there is little time to act. It is important to note that small sample sizes may also be misleading, and as more data are collected the treatment effect could regress to the mean, but being vigilant about any emerging trends will provide time to offer site training or potentially close or open new sites if a clear problem is observed. This could be added to the responsibilities of an independent Data Monitoring and Ethics Committee.

Overall, this study observed an early treatment effect in ProFHER but not in UK FROST, which was likely a chance finding rather than a bias at the start of the study. Interestingly, there was more evidence of heterogeneity when comparing outcomes between non-surgical treatment with surgery than with two surgical procedures. Both trials had evidence of selection bias in terms of recruiting patients with markers of poor prognosis, but this was minimal. Importantly, despite evidence of some marginal bias during the conduct of both trials, this had little effect on the primary outcome at one year. This study was an initial exploration into the possibility of early treatment effects and related biases during the conduct of two key orthopaedic surgical trials of the shoulder and, reassuringly, there was no evidence for concern. Examining this across a wider range of such trials could be useful, to be undertaken in the future.

Take home message

- Early large and biased treatment effects can arise in small studies. This study retrospectively examined whether early treatment effects occurred for two multi-centre orthopaedic randomized controlled trials and explored related biases.
- Overall, there was a possible early treatment effect in ProFHER but not in UK FROST. Importantly, despite evidence of some marginal bias during the conduct of both trials, this had little effect on the primary outcome at one year.
- Exploring early treatment effects with small sample sizes may be misleading, but by being vigilant about any emerging trends will provide time to offer site training or potentially close or open new sites if a clear problem is observed. This could be added to the responsibilities of an independent Data Monitoring and Ethics Committee.

Twitter

Follow the authors @YorkTrialsUnit

Supplementary material

Forest plots of mean patient outcome, comparing the first five sites opened to the remaining sites for the ProFHER and UK FROST trials.

References

1. **Akobeng AK**. Understanding randomised controlled trials. *Arch Dis Child*. 2005;90(8):840–844.
2. **Pereira TV**, **Horwitz RI**, **Ioannidis JPA**. Empirical evaluation of very large treatment effects of medical interventions. *JAMA*. 2012;308(16):1676–1684.
3. **Ioannidis JPA**. Why most discovered true associations are inflated. *Epidemiology*. 2008;19(5):640–648.
4. **Gurusamy KS**, **Gluud C**, **Nikolova D**, **Davidson BR**. Assessment of risk of bias in randomized clinical trials in surgery. *Br J Surg*. 2009;96(4):342–349.
5. **Schulz KF**, **Chalmers I**, **Hayes RJ**, **Altman DG**. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273(5):408–412.
6. **Knowlson C**, **Dean A**, **Doherty L**, **Fairhurst C**, **Brealey S**, **Torgerson DJ**. Recruitment patterns in multicentre randomised trials fit more closely to Price's Law than the Pareto Principle: A review of trials funded and published by the United Kingdom Health Technology Assessment Programme. *Contemp Clin Trials*. 2022;113:106665.
7. **Phelps EE**, **Tutton E**, **Griffin X**, **Baird J**, **TrAFFix study co-applicants**. Facilitating trial recruitment: A qualitative study of patient and staff experiences of an orthopaedic trauma trial. *Trials*. 2019;20(1):492.
8. **Donovan JL**, **de Salis I**, **Toerien M**, **Paramasivan S**, **Hamdy FC**, **Blazeby JM**. The intellectual challenges and emotional consequences of equipoise contributed to the fragility of recruitment in six randomized controlled trials. *J Clin Epidemiol*. 2014;67(8):912–920.
9. **Wright JG**, **Katz JN**, **Losina E**. Clinical trials in orthopaedics research. Part I. Cultural and practical barriers to randomized trials in orthopaedics. *J Bone Joint Surg Am*. 2011;93-A(5):e15.
10. **Boutron I**, **Ravaud P**, **Nizard R**. The design and assessment of prospective randomised, controlled trials in orthopaedic surgery. *J Bone Joint Surg Br*. 2007;89-B(7):858–863.
11. **Kahan BC**, **Rehal S**, **Cro S**. Risk of selection bias in randomised trials. *Trials*. 2015;16:405.
12. **Probst P**, **Grummich K**, **Heger P**, **et al**. Blinding in randomized controlled trials in general and abdominal surgery: protocol for a systematic review and empirical study. *Syst Rev*. 2016;5:48.
13. **Higgins JPT**, **Savović JP**. Chapter 8: assessing risk of bias in a randomized trial cochrane handbook for systematic reviews of interventions version. In: *(Updated February 2021) Cochrane*. Vol 62. 2021.
14. **DAMOCLES Study Group, NHS Health Technology Assessment Programme**. A proposed charter for clinical trial data monitoring committees: helping them to do their job well. *Lancet*. 2005;365(9460):711–722.
15. **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(10):1037–1047.
16. **Rangan A**, **Brealey SD**, **Keding A**, **et al**. Management of adults with primary frozen shoulder in secondary care (UK FROST): a multicentre, pragmatic, three-arm, superiority randomised clinical trial. *Lancet*. 2020;396(10256):977–989.
17. **Dawson J**, **Fitzpatrick R**, **Carr A**. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br*. 1996;78-B(4):593–600.
18. **Nagendran M**, **Pereira TV**, **Kiew G**, **et al**. Very large treatment effects in randomised trials as an empirical marker to indicate whether subsequent trials are necessary: meta-epidemiological assessment. *BMJ*. 2016;355:i5432.
19. **Barnett K**, **Mercer SW**, **Norbury M**, **Watt G**, **Wyke S**, **Guthrie B**. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet*. 2012;380(9836):37–43.
20. **Ho VP**, **Schiltz NK**, **Reimer AP**, **Madigan EA**, **Koroukian SM**. High-risk comorbidity combinations in older patients undergoing emergency general surgery. *J Am Geriatr Soc*. 2019;67(3):503–510.
21. **Dyer BP**, **Burton C**, **Rathod-Mistry T**, **Blagojevic-Bucknall M**, **van der Windt DA**. Diabetes as a prognostic factor in frozen shoulder: A systematic review. *Arch Rehabil Res Clin Transl*. 2021;3(3):100141.
22. **Keding A**, **Handoll H**, **Brealey S**, **et al**. The impact of surgeon and patient treatment preferences in an orthopaedic trauma surgery trial. *Trials*. 2019;20(1):570.

Author information:
- C. N. Knowlson, MSci, PhD, Research Fellow
- S. Brealey, BSc, PhD, Research Fellow
- A. Keding, BSc, MSc, Statistician
- D. Torgerson, MSc, PhD, Co-Director
York Trials Unit, Department of Health Sciences, University of York, York, UK.
- A. Rangan, MBBS, FRCS, Professor of Orthopaedic Surgery, James Cook University Hospital, Middlesbrough, UK.

Author contributions:
- C. N. Knowlson: Data curation, Investigation, Formal analysis, Visualization, Writing – original draft.
- S. Brealey: Conceptualization, Methodology, Supervision, Writing – review & editing.
- A. Keding: Methodology, Supervision, Writing – review & editing.
- D. Torgerson: Writing – review & editing.
- A. Rangan: Writing – review & editing.

Funding statement:
- The authors disclose receipt of the following financial or material support for the research, authorship, and/or publication of this article: this study is supported by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (Grant Reference Number 13/26/01). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. ProFHER and UK FROST were studies funded by the NIHR (Health Technology Assessment programme (project No. 06/404/53 and 13/26/01)).

ICMJE COI statement:
- A. Rangan reports grants from NIHR, AOUK, CIHR, DePuy Johnson & Johnson, and Wishbone Trust, NZ; and speaker payments and expenses from DePuy Johnson & Johnson, unrelated to this study. A. Rangan is also a trustee of the British Orthopaedic Association and the British Elbow & Shoulder Society, and a surgeon member of the NJR Steering Committee.

Acknowledgements:
- We thank the patients participating in this trial, and our collaborators at the participating hospitals, who provided the data to make this analysis possible.

Ethical review statement:
- No ethical review was required for this secondary analyses of data.

Open access funding
- The open access fee for this study was funded by the NIHR HTA programme (project 13/26/01).