■ HIP

Does performance-based remuneration improve outcomes in the treatment of hip fracture?

RESULTS FROM THE WHITE MULTICENTRE HIP FRACTURE COHORT

X. L. Griffin,
J. Achten,
N. Parsons,
M. L. Costa,
On behalf of the WHiTE collaborators

From University of Oxford, Oxford, UK

Aims

The aim of this study was to determine whether national standards of best practice are associated with improved health-related quality of life (HRQoL) outcomes in hip fracture patients.

Methods

This was a multicentre cohort study conducted in 20 acute UK NHS hospitals treating hip fracture patients. Patients aged ≥ 60 years treated operatively for a hip fracture were eligible for inclusion. Regression models were fitted to each of the "Best Practice Tariff" indicators and overall attainment. The impact of attainment on HRQoL was assessed by quantifying improvement in EuroQol five-dimension five-level questionnaire (EQ-5D-5L) from estimated regression model coefficients.

Results

A total of 6,532 patients provided both baseline and four-month EQ-5D-5L, of whom 1,060 participants had died at follow-up. Best practice was achieved in the care of 57% of participants; there was no difference in age, cognitive ability, and mobility at baseline for the overall attainment and non-attainment groups. Attaining at least 'joint care by surgeon and orthogeriatrician', 'delirium assessment', and 'falls assessment' was associated with a large, clinically relevant increase in four months EQ-5D-5L of 0.094 (bootstrapped 95% confidence interval (CI) 0.046 to 0.146).

Conclusion

National standards with enhanced remuneration in hip fracture care results in improvement in individual patients' HRQoL.

Cite this article: *Bone Joint J* 2021;103-B(5):881–887.

Introduction

Pay-for-performance initiatives have been used in several healthcare systems in an attempt to generate and drive improvements in patient outcomes.[1,2] In the UK National Health Service (NHS), Best Practice Tariffs (BPTs) were introduced with the explicit intention to 'incentivise and adequately reimburse providers' to deliver improved clinical and cost-effective healthcare.[3]

Fragility hip fracture is a leading cause of death and disability worldwide. As populations age, the number of hip fractures is increasing rapidly with an annual worldwide incidence of 1.6 million.[4] In the NHS, hip fracture care has been part of the BPT system of payment since April 2010. Reimbursement to providers is linked to specific performance indicators based on the care delivered in the early period following injury (Table I).[5] Similar national audit programmes have been developed across the world.[6]

Overall, mortality for patients with hip fracture have improved since the introduction of BPT; 30-day mortality has fallen from 8.4% in 2012 to 7.1% in 2017.[7] However, it is not clear whether BPT is associated directly with improved patient-centred outcomes and whether any such improvements justify the cost and administrative burden of implementing the tariff structure.[1,2]

The World Hip Trauma Evaluation (WHiTE) cohort was established in 2014, recruiting patients

Correspondence should be sent to X. L. Griffin; email: x.griffin@qmul.ac.uk

doi:10.1302/0301-620X.103B5.BJJ-2020-1839.R1 $2.00

Bone Joint J 2021;103-B(5):881–887.

Table I. Best practice tariff criteria and their determination from the World Hip Trauma Evaluation (WHiTE) dataset.

Item	Name	Description	Categories	Coding
BPT 1	Surgery < 36	Surgery within 36 hours of admission to an emergency department	No delay, surgery in < 36 hours	Coded directly
			Medical delay: awaiting diagnosis, investigation or medical stablization	
			Administrative delay: lack of available resources	
			Other delay	
BPT 2	Joint care	Admitted under joint care of a consultant orthopaedic surgeon and geriatrician	Yes/no	Attained if GMC numbers were recorded for two appropriate clinicians
BPT 3	MDT protocol	Admitted using an agreed MDT protocol	Yes/no	Coded directly
BPT 4	Geriatrician < 72	Assessed by a geriatrician within 72 hours of admission	Yes/no	Coded directly
BPT 5	MDT rehab	Postoperative geriatrician-directed MDT rehabilitation	Yes/no	Coded directly
BPT 6	Bone health	Assessment of falls risk and bone health	Yes/no	Coded directly
BPT 7	Cognitive assessment	Preoperative and postoperative delirium assessments	Yes/no	Attained if two valid AMTS scores were reported

BPT criteria described are those in place at the time of the study and have subsequently been modified.
AMTS, Abbreviated Mental Test Score; BPT, best practice tariff; GMC, General Medical Council; MDT, multidisciplinary team.

with fragility hip fracture from a representative sample of hospitals in the UK.[8,9] WHiTE provides follow-up of all participants reporting data that mirror the UK national audit, augmented with health-related quality of life (HRQoL) outcomes.[8] Recovery of HRQoL is reported by patients and carers as the most important outcome following this injury.[10] This unique cohort therefore provides a more comprehensive insight into the effectiveness of clinical care provided, measured against the key domain of "interest to patients". The aim of this study was to determine whether best practice is associated with improved HRQoL.

Methods

Study design summary. WHiTE is an observational cohort study that collects information on assessment, treatment, and recovery of patients admitted to participating UK NHS hospitals after hip fracture.[8] Full details of the protocol have been published previously. The WHiTE study was approved by research ethics committee (REC; WHiTE Cohort approved by Camberwell St Giles Research Ethics Committee with reference 11/LO/0927). The study is registered at Current Controlled Trials: ISRCTN63982700.

Eligibility. Patients were eligible to participate in WHiTE if they were aged 60 years or older and were treated operatively for a hip fracture. The cohort is both representative of the patients treated in the participating hospitals and the wider UK population with hip fracture.[9]

Consent. Patients gave their consent to participate in the WHiTE study or, for those without capacity, agreement was provided by an appropriate consultee in line with the Mental Capacity Act 2005.[11]

Procedures. Participants enrolled in the WHiTE Cohort study were treated in accordance with local standard care pathways. The National Institute for Clinical Excellence (NICE) has issued standardized care guidelines that are used in the majority of hospitals, and are summarized elsewhere.[12] Adherence to this NICE guideline is 85% to 98% across all recommendations (with the exception of the use of total hip arthroplasty),[13] and is reported in full, annually in the national audit.[14]

Outcome data collection. Data were transcribed from clinical reporting forms, completed by recruitment centre research teams at baseline or entered directly by the central study team into the WHiTE database during follow-up telephone calls at four months (OpenClinica v. 3.7, OpenClinica, USA). Data were extracted from the database and saved to a comma-separated format for analyses.

Health-related quality of life. HRQoL was assessed using the EuroQol five-dimension five-level questionnaire (EQ-5D-5L);[15] a generic, validated, cross-disciplinary standardized health utility instrument widely used to assess HRQoL after hip fracture.[16] EQ-5D has two parts, a visual analogue scale (VAS), which measures self-rated health, and a health status instrument, which is the focus of the analysis reported here, consisting of a five-level response (no problems, slight problems, moderate problems, severe problems, and extreme problems) for five health domains related to daily activities: 1) mobility, 2) self-care, 3) usual activities, 4) pain and discomfort, and 5) anxiety and depression. Each WHiTE participant was asked to indicate their health state by selecting the most appropriate statement in each of the five dimensions; combining these provides a five-digit number that describes the participant's health state. The five-digit responses, from the EQ-5D health classifications, were converted into an overall score using a published utility algorithm for the UK population.[17,18] The clinically important difference in EQ-5D derived utility is estimated to be between 0.05 and 0.075.[19] For context, the estimated adjusted disutility associated with asthma, acute myocardial infarction, and diabetes melitus are 0.05, 0.06, and 0.07 respectively.[20] Participants in WHiTE were asked to provide (retrospective) pre-fracture assessments of HRQoL, using EQ-5D, at enrolment into the study. While there are limitations to this approach, the recall values reported are similar to an age- and sex-matched UK population.[16]

Table II. Best practice tariff attainment by indicator.

Name	Description	Indicator attained, n (%)	
		No	Yes
Surgery < 36	Surgery within 36 hours of admission to an emergency department	1,768 (20.8)	6,722 (79.2)
Joint care	Admitted under joint care of an consultant orthopaedic surgeon and geriatrician	506 (5.8)	8,167 (94.2)
MDT protocol	Admitted using an agreed MDT protocol	93 (1.1)	8,436 (98.9)
Geriatrician < 72	Assessed by a geriatrician within 72 hours of admission	561 (6.7)	7,773 (93.3)
MDT rehab	Postoperative geriatrician-directed MDT rehabilitation	126 (1.5)	8,222 (98.5)
Bone health	Assessment of falls risk and bone health	496 (5.8)	8,057 (94.2)
Cognitive assessment	Preoperative and postoperative delirium assessments	1,161 (13.4)	7,512 (86.6)
All		3,771 (43.5)	4,902 (56.5)

MDT, multidisciplinary team.

Some participants were unable to complete the EQ-5D themselves, so this was completed by either the next of kin, a relative, carer, or other proxy. Proxy-reported EQ-5D index scores have previously been shown to be an acceptable source of data in a similar population to the WHiTE cohort,[21] although evidence from more recent studies is mixed.[22,23]

Best practice tariff indicators. The performance indicators assessed as part of BPT during the study period, and the means by which attainment or otherwise of these indicators were derived from the database, and shown in Table I.

If data were not available for an individual participant to make a valid assessment of whether the indicator had been attained (e.g. a missing datum), then the individual criterion was marked as missing. Overall BPT attainment was assessed as all seven criteria being attained for any individual participant.

Other data. Participant demographic variables including age, sex, alcohol consumption, smoking status, and reported diabetes or renal failure were collected at baseline. Participants residence and mobility prior to the fracture; fracture type, surgical treatment, American Society of Anesthesiologists grade (ASA; I, II, III, IV, V)[24] and Abbreviated Mental Test scores (AMTS; 1 to 10) were also recorded. At four months post-injury, further data on residency and mobility were collected.

Sample size. The data reported here are based upon the data extract for the pre-specified analysis of the WHiTE cohort of the first 6,000 complete outcome datasets. Full details are reported in the published protocol.[8] In summary, an initial sample size of 6,000 patients would provide considerable power to estimate outcomes of interest with a high level of precision. Assuming that we are interested in patient subgroups no smaller than 5% of this total (n = 300) and a clinically important difference in utility of between 0.05 and 0.075,[19] we had power to detect small to moderate effects sizes in our planned multiple regression analyses.[25]

Statistical analysis. This is an elderly and frail population, with a high mortality at four months post-injury.[26] In line with EuroQoL recommendations,[15] for our primary analysis, we 'imputed' EQ-5D values for those participants who died before the four months post-injury EQ-5D assessment with a value of zero; we call this 'death-adjusted' EQ-5D and simply EQ-5D hereafter.[27] We tested the impact of this approach through sensitivity analyses reported in Supplementary Table ii and Figure a.

Baseline HRQoL is strongly associated with HRQoL at four months post-injury[26,27] and approximately 35% of the variation in baseline HRQoL is explained by routinely reported participant characteristics (sex, age, pre-fracture mobility, and pre-fracture residency).[28] Therefore, to test the hypothesis that EQ-5D, at four months after injury, was higher if a BPT indicator had been attained, we fitted regression models to adjust for baseline EQ-5D and these characteristics. Regression models were fitted to each of the seven individual BPT indicators and overall BPT attainment. The overall impact of the BPT indicators on HRQoL was assessed by quantifying the improvement

Table III. Participant characteristics for those who attained all seven best practice tariff criteria (Yes) and those who did not (No).

Participant characteristic	No (n = 3,771)	Yes (n = 4,902)
Age, n (%)		
< 80 yrs	1,407 (37.3)	1,688 (34.4)
80+ yrs	2,364 (62.7)	3,214 (65.6)
Sex, n (%)		
Female	2,675 (70.9)	3,615 (73.8)
Male	1,096 (29.1)	1,287 (26.2)
AMTS, n (%)		
0 to 3: Severe impairment	559 (16.5)	887 (18.1)
4 to 6: Moderate impairment	295 (8.7)	433 (8.8)
7 to 10: Normal	2,537 (74.8)	3,582 (73.1)
ASA, n (%)		
I	84 (2.4)	104 (2.2)
II	866 (25.0)	1,441 (30.6)
III	1,966 (56.8)	2,620 (55.7)
IV	539 (15.6)	532 (11.3)
V	7 (0.2)	6 (0.1)
Mobility (pre-fracture), n (%)		
No functional mobility	85 (2.3)	106 (2.2)
Freely mobile without aids	1,530 (41.6)	1,968 (40.3)
Mobile outdoors with one aid	903 (24.5)	1,160 (23.7)
Mobile outdoors with two aids or frame	644 (17.5)	775 (15.9)
Some indoor mobility but never outside without help	492 (13.4)	833 (17.0)
Unknown	28 (0.8)	46 (0.9)
Residency (pre-fracture), n (%)		
Own home/sheltered housing	3,107 (84.3)	4,052 (82.7)
Residential care	303 (8.2)	475 (9.7)
Nursing care	210 (5.7)	315 (6.4)
Rehab unit	3 (0.1)	7 (0.1)
Index hospital	44 (1.2)	32 (0.7)
Other hospital in Trust	10 (0.3)	15 (0.3)
Other	8 (0.2)	6 (0.1)

AMTS, Abbreviated Mental Test Score; ASA, American Society of Anesthesiologists.

Table IV. Means and estimates of differences in four-month EuroQol five-dimension questionnaire score between those attaining (Yes) and those not attaining best practice tariff (No) for each best practice tariff indicator.

BPT	Mean EQ-5D 4 months			Adjusted analysis p-value*
	No	Yes	Difference (95% CI)	
Surgery < 36	0.429	0.430	0.010 (-0.007 to 0.027)	0.254
Joint care	0.387	0.433	0.034 (0.002 to 0.066)	0.035
MDT protocol	0.422	0.431	0.014 (-0.053 to 0.082)	0.673
Geriatrician < 72	0.397	0.433	0.026 (-0.003 to 0.054)	0.077
MDT rehab	0.392	0.431	0.015 (-0.047 to 0.076)	0.644
Bone health	0.377	0.433	0.051 (0.019 to 0.082)	0.002
Cognitive assessment	0.403	0.434	0.025 (0.004 to 0.047)	0.021
All	0.419	0.438	0.016 (0.002 to 0.030)	0.026

*Regression analysis adjusting for baseline EQ-5D, age, sex, pre-fracture mobility, and pre-fracture residence.
BPT, best practice tariff; CI, confidence interval; EQ-5D, EuroQol five-dimension questionnaire; MDT, multidisciplinary team.

in EQ-5D from the estimated regression model coefficients. The distribution of the residuals of both 'death-adjusted' and 'death-as-missing' models was confirmed to be approximately normal prior to selecting our parametric approach and are in Supplementary Figure a. Similar logistic regression models were also fitted for a binary response variable of death at four months post-injury, to test the hypothesis that mortality at four months post-injury was lower if BPT had been attained than not.

As a further step in understanding the role of the participant characteristics on BPT attainment, propensity score matching (PSM)[29] was used to reduce bias due to these variables in the regression coefficient estimates obtained from the main analysis of HRQoL at four months post-injury. Two-to-one nearest-neighbour matching was used to obtain a reduced dataset consisting of data from those participants who did not attain BPT matched as nearly as possible to two participants who did attain BPT.

All analyses were undertaken in R Project for Statistical Computing (Austria), using the MatchIt package[30] for the PSM analysis, with statistical significance assessed at the 5% level.

Results

Participants. This cohort study comprises data from 8,673 participants recruited between May 2014 and April 2017, of whom 7,391 provided a baseline EQ-5D and 6,532 both baseline and four months EQ-5D. The mean age of WHiTE cohort participants at recruitment was 83 years (SD 8.5), and the percentage female to male split was 73:27.[28] A total of 1,060 participants died prior to four months follow-up; 518 and 542 respectively in the groups for whom all BPT criteria were met or not. The sample is a nested subset of patients treated at the participating hospitals and the wider NHS. We have published elsewhere the generalizability of this cohort to both the treated population in the participating centres and the wider UK population of hip fractures and found no evidence of selection bias.[31]

BPT attainment. The overall attainment of each BPT indicator is shown in Table II; taken together, all indicators and consequently the additional tariff to the provider was only attained for 57% of participants. The poorest attainment (79.2%) was for BPT 1, 'Surgery within 36 hours from arrival in an emergency department'. Two indicators were almost universally attained: BPT 3, 'Admitted using an agreed multidisciplinary team (MDT) protocol' and BPT 5, 'Postoperative geriatrician-directed MDT rehabilitation' were not attained in fewer than 2% of cases.

Table III shows the participant characteristics for those who attained all seven BPT criteria and those who did not. The mean age for those participants attaining all seven BPT criteria was 83.0 years (standard deviation (SD) 8.5) and for those participants not attaining all seven BPT criteria was 82.5 years (SD 8.6); the mean AMTS was 7.46 in those who did attain BPT (SD 3.46) compared with 7.60 (SD 3.38) for those who did not attain BPT. Overall the participant populations who attained all BPT and those who did not were well balanced.

Health-related quality of life. Table IV shows the mean four-month post-injury EQ-5D by BPT indicator attainment; EQ-5D was always lower when BPT characteristics had not been attained (No), than when it had been attained (Yes). The proportion of participants for whom the EQ-5D was proxy reported was approximately 30% and similar between the groups that had and had not attained BPT.

Attaining indicators BPT 2, BPT 6, and BPT 7 ('Admitted under joint care of an consultant orthopaedic surgeon and geriatrician', 'Assessment of falls risk and bone health', and 'Preoperative and postoperative cognitive assessments') was associated with statistically significant and small to moderate clinical improvements in four months post-injury EQ-5D. Including all three of these BPT indicators in a single regression model, with adjustment as previously described, showed that attaining at least this subset of indicators was associated with the large, highly clinically relevant increase in four months post-injury EQ-5D of 0.094 (bootstrapped 95% CI 0.046 to 0.146). A similar magnitude of change in utility is associated with major affective psychoses.[20]

A summary of the results of the PSM and repetition of the analysis reported in Table IV are reported in Supplementary Table iv and Figures b to h. Adjusted estimates of differences in four months post-injury EQ-5D from the PSM analysis were consistent with the estimated effects from the full dataset, as were estimated effects for the other BPT indicators; increases in BPT indicators 2, 6, and 7 were 0.053 (95% CI 0.016 to 0.089), 0.054 (95% CI 0.016 to 0.093), and 0.031 (95% CI 0.007 to 0.056)).

Mortality. Mortality at four months post-injury was always lower when each BPT indicator had been attained than not.

However, these differences were not maintained after adjustment for important baseline variables except for BPT 1, 'surgery within 36 hours of admission to an emergency department' (odds ratio indicating lower mortality for Yes = 0.71; 95% CI 0.59 to 0.85, $p < 0.001$, logistic regression) (Supplementary Table i).

Further exploration of this association by subgroups of 'cause of delay' showed that the risk of death at four months post-injury for those participants with a medical delay (20.6%), namely patients for whom a clinical decision was taken to delay surgery to correct a modifiable preoperative risk, was much higher than risks for any other reason. Baseline and four-month post-injury EQ-5D scores were also statistically significantly lower in the medical delay group than the administrative delay group, namely patients who were delayed only due to non-clinical reasons such as lack of operating time or theatre space ($p = 0.001$ and $p = 0.003$ respectively, both linear regression) (Supplementary Table iii).

Discussion

In this large, multicentre, cohort study we have reported the first evidence, from the NHS, that attaining best practice tariff in the treatment of patients with hip fracture is associated not only with reduced mortality but also with improved HRQoL four months after their hip fracture. The three indicators, 'admitted under the joint care of a consultant geriatrician and orthopaedic surgeon' (BPT 2), 'assessment of falls risk and bone health' (BPT 6) and 'preoperative and postoperative delirium assessment' (BPT 7) were statistically significantly associated with better HRQoL at four months post-injury; attaining at least this subset of the indicators was associated with a large, highly clinically relevant benefit in HRQoL.

There is extremely strong evidence that mortality was higher among those participants that were delayed in receiving surgery beyond 36 hours. However, we found important differences in baseline characteristics between participants with different causes of delay, indicating that residual confounding factors may explain at least some of this variation in mortality. It seems possible that clinical judgement and pre-existing ill health explains much of the variation in mortality risks rather than delay per se.

Participants delayed for medical reasons have lower baseline EQ-5D. The outlook for these participants is poor as mortality is high (20.6% at four months); the low baseline EQ-5D may explain this, irrespective of the delay in operation. Participants delayed for administrative reasons have relatively high baseline EQ-5D, significantly higher than those with no delay. Clinicians, determining clinical priorities, may be choosing to delay operations for those patients who are generally healthier. The outlook for these patients is very good, because they have better health before hip fracture; they have the lowest four-month post-injury mortality and highest HRQoL of any participants. This may reflect decision-making, such as delay for specialist hip surgeons to perform total hip arthroplasty.

Alternative statistical approaches gave similar estimates, confidence intervals, and consistent inferences for matched subsets of the full data, providing support for the conclusions of the primary analyses (see Supplementary Material).

The principal limitations of this study are those inherent to any observational design when attempting to infer causality. However, we have sought to address these limitations through a rigorous, prospective design with a pre-published protocol and analysis plan.[8] Previous analyses have shown that there is a strong correlation between baseline and four-month post-injury EQ-5D,[16,26,28] so by including this and the other main patient characteristics (age, sex, pre-fracture mobility, and residency) in the analysis we can adjust for the effects of these factors on the main outcomes.

This study was designed to explore associations between BPT indicators and patient-level improvements in HRQoL. Changes in HRQoL were deemed to be clinically relevant or not against a prespecified difference of 0.075. It is plausible that population-level effects of BPT, or assessments of cost-effectiveness of a policy of implementing payment for performance, may be different from the inferences presented here.

There is little evidence in the data to suggest that selection bias was likely to affect the four-month post-injury EQ-5D analysis. While it is plausible that higher baseline HRQoL might be associated with attaining BPT, and therefore the observation of better four-month HRQoL outcome may be due to selection bias, the data suggest that bias in this direction was not likely; as those attaining BPT were marginally older, had a lower AMTS, and poorer mobility.

Smaller, often single centre studies have sought to address similar research questions but have been limited to using mortality alone as an outcome.[32–34] These studies have reported conflicting findings, although the largest found overall that outcomes were better in patients for whom BPT was attained than not.[33] These studies were all limited in two crucial ways; firstly by size (each was underpowered and prone to type II error) and secondly none reported outcomes that patients prioritize.[10]

We recognize that the costs associated with provision of the care required to achieve BPT criteria, the assessment of those criteria, and administration of tariff payments is a key consideration in policy-making decisions. Capturing these costs was outside the scope of this study and further research on the cost-effectiveness of the BPT system is needed.

Our findings are generally concordant with what might be expected clinically. While accepting the limitations of the design, we are confident that these data make a strong case for the benefit of the BPT in hip fracture care, in terms of HRQoL. Best practice was only delivered for 57% (4,902/8,673) of participants in this study—we expect and hope that improving this can yield further improvements in patient outcomes.

Take home message

- This study is the first to report the association between best-practice, defined as the delivery of key performance indicators, and health-related quality of life in patients recovering from a hip fracture.

Twitter

Follow Queen Mary University of London @QMULBartsTheLon
Follow Barts Health NHS Trust @NHSBartsHealth
Follow X. L. Griffin @xlgriffin

Follow M. L. Costa @oxford_trauma

Podcast

Listen to a podcast related to this article online at https://online.boneandjoint.org.uk/bjj/podcasts

Supplementary material

The supplementary material includes the results of the following analysis: Death-as-missing adjusted Euro-Qol five-dimension questionnaire (EQ-5D), Model distributions of death-adjusted and death-as-missing EQ-5D, Effects of proxy reporting on EQ-5D, Mortality, EQ-5D and attainment of "surgery within 36 hours of admission to an emergency department", and propensity score matching.

References

1. **Maynard A**. The powers and pitfalls of payment for performance. *Health Econ*. 2012;21(1):3–12.
2. **Ogundeji YK**, **Bland JM**, **Sheldon TA**. The effectiveness of payment for performance in health care: a meta-analysis and exploration of variation in outcomes. *Health Policy*. 2016;120(10):1141–1150.
3. **Department of Health and Social Care**. Update to best practice tariffs. GOV.UK. 2020. https://www.gov.uk/government/news/update-to-best-practice-tariffs
4. **Johnell O**, **Kanis JA**. An estimate of the worldwide prevalence, mortality and disability associated with hip fracture. *Osteoporos Int*. 2004;15(11):897–902.
5. **NHS England & NHS Improvement**. 2019/20 national tariff payment system: Annex D, guidance on best practice tariffs. 2019. https://improvement.nhs.uk/documents/4986/AnnexD_1920_Guidance_on_BPTs.pdf
6. **Johansen A**, **Golding D**, **Brent L**, **et al**. Using national hip fracture registries and audit databases to develop an international perspective. *Injury*. 2017;48(10):2174–2179.
7. **Metcalfe D**, **Zogg CK**, **Judge A**, **et al**. Pay for performance and hip fracture outcomes: an interrupted time series and difference-in-differences analysis in England and Scotland. *Bone Joint J*. 2019;101-B(8):1015–1023.
8. **Costa ML**, **Griffin XL**, **Achten J**, **et al**. World hip trauma evaluation (white): framework for embedded comprehensive cohort studies. *BMJ Open*. 2016;6(10):e011679.
9. **Metcalfe D**, **Costa ML**, **Parsons NR**, **et al**. Validation of a prospective cohort study of older adults with hip fractures. *Bone Joint J*. 2019;101-B(6):708–714.
10. **Haywood KL**, **Griffin XL**, **Achten J**, **Costa ML**. Developing a core outcome set for hip fracture trials. *Bone Joint J*. 2014;96-B(8):1016–1023.
11. **No authors listed**. Mental Capacity Act 2005. 2005. https://www.legislation.gov.uk/ukpga/2005/9/contents
12. **National Clinical Guideline Centre**. Hip fracture: management. 2017. National Clinical Guideline Centre. https://www.nice.org.uk/guidance/cg124 (date last accessed 3 February 2021).
13. **Perry DC**, **Metcalfe D**, **Griffin XL**, **Costa ML**. Inequalities in use of total hip arthroplasty for hip fracture: population based study. *BMJ*. 2016;353.
14. **Royal college of Physicians**. National Hip Fracture Database. 2012. http://www.nhfd.co.uk (date last accessed 13 November 2012).
15. **EuroQol**. EQ-5D. 2011. http://euroqol.org (date last accessed 31 March 2011).
16. **Parsons N**, **Griffin XL**, **Achten J**, **Costa ML**. Outcome assessment after hip fracture: is EQ-5D the answer? *Bone Joint Res*. 2014;3(3):69–75.
17. **National Institue of Health and Care Excellence**. Position statement on use of the EQ-5D-5L valuation set for England (updated November 2018). 2018. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l (date last accessed 1 December 2018).
18. **Dolan P**. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108.
19. **Walters SJ**, **Brazier JE**. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res*. 2005;14(6):1523–1532.
20. **Sullivan PW**, **Slejko JF**, **Sculpher MJ**, **Ghushchyan V**. Catalogue of EQ-5D scores for the United Kingdom. *Med Decis Making*. 2011;31(6):800–804.
21. **Devine A**, **Taylor SJC**, **Spencer A**, **Diaz-Ordaz K**, **Eldridge S**, **Underwood M**. The agreement between proxy and self-completed EQ-5D for care home residents was better for index scores than individual domains. *J Clin Epidemiol*. 2014;67(9):1035–1043.
22. **Usman A**, **Lewis S**, **Hinsliff-Smith K**, **et al**. Measuring health-related quality of life of care home residents: comparison of self-report with staff proxy responses. *Age Ageing*. 2019;48(3):407–413.
23. **Parker B**, **Petrou S**, **Underwood M**, **Madan J**. Can care staff accurately assess health-related quality of life of care home residents? a secondary analysis of data from the opera trial. *BMJ Open*. 2017;7(4):e012779.
24. **Saklad M**. Grading of patients for surgical procedures. *Anesthesiology*. 1941;2(3):281–284.
25. **Cohen J**. *Statistical power analysis for the behavioral sciences*: Lawrence Erlbaum Associates, 1988.
26. **Griffin XL**, **Parsons N**, **Achten J**, **Fernandez M**, **Costa ML**. Recovery of health-related quality of life in a United Kingdom hip fracture population: the Warwick Hip Trauma Evaluation - a prospective cohort study. *Bone Joint J*. 2015;97-B(3):372–382.
27. **Parsons N**, **Griffin XL**, **Achten J**, **Chesser TJ**, **Lamb SE**, **Costa ML**. Modelling and estimation of health-related quality of life after hip fracture: a Re-analysis of data from a prospective cohort study. *Bone Joint Res*. 2018;7(1):1–5.
28. **Parsons NR**, **Costa ML**, **Achten J**, **Griffin XL**. Baseline quality of life in people with hip fracture: results from the multicentre white cohort study. *Bone Joint Res*. 2020;9(8):468–476.
29. **Austin PC**. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011;46(3):399–424.
30. **Stuart EA**, **King G**, **Imai K**, **Ho D**. *MatchIt: nonparametric preprocessing for parametric causal inference. Journal of statistical software*. 42. Los Angeles: University of California, 2011.
31. **Metcalfe D**, **Costa ML**, **Parsons NR**, **et al**. Validation of a prospective cohort study of older adults with hip fractures. *Bone Joint J*. 2019;101-B(6):708–714.
32. **Khan SK**, **Shirley MDF**, **Glennie C**, **Fearon PV**, **Deehan DJ**. Achieving best practice tariff may not reflect improved survival after hip fracture treatment. *Clin Interv Aging*. 2014;9:2097–2102.
33. **Oakley B**, **Nightingale J**, **Moran CG**, **Moppett IK**. Does achieving the best practice tariff improve outcomes in hip fracture patients? an observational cohort study. *BMJ Open*. 2017;7(2):e014190.
34. **Khan SK**, **Weusten A**, **Bonczek S**, **Tate A**, **Port A**. The best practice tariff helps improve management of neck of femur fractures: a completed audit loop. *Br J Hosp Med*. 2013;74(11):644–647.

Author information:
X. L. Griffin, PhD, Professor of Trauma and Orthopaedic Surgery, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; Department of Trauma and Orthopaedic Surgery, Barts Health NHS Trust, London, UK.

J. Achten, PhD, Research Manager of Oxford Trauma and Emergency Care
M. L. Costa, PhD, Professor of Orthopaedic Trauma Surgery
Oxford Trauma and Emergency Care, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, UK.

N. Parsons, PhD, Associate Professor of Medical Statistics, Statistics and Epidemiology Unit, Warwick Medical School, University of Warwick, Coventry, UK.

Author contributions:
X. L. Griffin: Conceptualized the study, Designed and managed the trial, Obtained the grant funding, Acquired and interpreted the data, Drafted and approved the final submitted manuscript.
J. Achten: Conceptualized the study, Designed and managed the trial, Obtained grant funding, Acquired and interpreted the data, Drafted and approved the final submitted manuscript.
N. Parsons: Conceptualized the study, Designed and managed the trial, Obtained grant funding, Interpreted the data, Conducted the statistical analysis, Drafted and approved the final submitted manuscript.
M. L. Costa: Conceptualized the study funding, Designed and managed the trial, Obtained grant funding, Acquired and interpreted the data, Drafted and approved the final submitted manuscript.

Funding statement:
The study was sponsored by the University of Oxford. The study was funded by a NIHR Programme Development Grant and the NIHR Oxford Biomedical Research Centre. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication. No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

ICMJE COI statement:
M. L. Costa is a National Institute for Health Research (NIHR) Senior Investigator and member of the NIHR HTA General Board. X. L. Griffin is a NIHR Clinician Scientist. M. L. Costa, J. Achten, N. R. Parsons and X. L. Griffin report a NIHR grant for this study.

Data sharing:
All data requests should be submitted to the corresponding author for consideration. Access to anonymized data may be granted following review.

Acknowledgements:
Lead author, Xavier Griffin, hereby affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned have been explained. We would like to thank the patients and the research associates at all the recruitment centres, and, in particular, Stephanie Wallis and Robin Lerner (University of Oxford, Oxford, UK) for their input into trial coordination and management. In addition, we thank the members of the WHiTE Oversight Committee (Tim Chesser – Independent Chair, Iain Moppett, Antony Johansen Alwin McGibbon, Karen May and Richard Grant) and the WHiTE scientific committee (Stu White, Tim Chesser, Jenny Gould, Josephine Rowling, Mark Baxter, Philip Bell, Rafa Pinedo-Villanueva, Sallie Lamb, Andy Judge and Chris Boulton).
* Peter Hull (Addenbrookes Hospital), Mark Dunbar (Heartlands Hospital), Graham Smith (Frimley Park Hospital), John Davison (Leicester Royal Infirmary), Iain McNamara (Norfolk & Norwich Hospital), Mike Reed (Northumbria Specialist Emergency Care Hospital), Mark Farrar (Poole Hospital), Ad Gandhe (Queen Alexandra Hospital), Ansar Mahmood (Queen Elizabeth Hospital, Birmingham), Robert Handley (John Radcliffe Hospital), Andrew McAndrew (Royal Berkshire Hospital), Benedict Rogers (Royal Sussex County Hospital), Paul Fearon (Royal Victoria Infirmary), William Eardley (James Cook University Hospital), Mehool Acharya (Southmead Hospital, Bristol), Damian McClelland (Royal Stoke Hospital), Paul Dixon (Sunderland Royal Hospital), Jonathan Young (University Hospital Coventry), Ben Ollivere (Queens Medical Centre, Nottingham), Callum Clark (Wexham Park Hospital).

Ethical review statement:
The study was supported by a patient advisory group which provided input to the programme of research. Patients partnered with us for the design of the study, the informational material to support the participants, and provided insight into the burden of participation from the patient's perspective. Throughout the study, patients were involved in the management and oversight of the study. At the end of the study, the patient advisory group commented on the findings and contributed to the dissemination plan.

This article was primary edited by D. Johnstone.

This paper was presented at the 8th Fragility Fracture Network Global Congress, Oxford, 2019.