Does pay for performance improve patient outcomes in a national health service? Results from the WHiTE multicentre hip fracture cohort.

Abstract

Aim: The aim of this study was to determine whether best practice is associated with improved Health-Related Quality of Life (HRQoL) in hip fracture patients.

Patients and methods: This was a multi-centre comprehensive cohort study conducted in 20 acute UK National Health Service hospitals treating hip fracture patients. Patients aged ≥60 years treated operatively for a hip fracture were eligible for inclusion.

HRQoL was assessed using the Euro-Qol 5 Dimensions 5 Levels Score (EQ-5D-5L) at baseline and 4-months post-injury by patient or proxy (if the patient lacked capacity). Hospital reimbursement is linked to attaining all best practice indicators for any individual participant. Regression models were fitted to each of the indicators and overall attainment. The impact of attainment on HRQoL was assessed by quantifying improvement in EQ-5D-5L from estimated regression model coefficients.

Results and conclusion: 6532 patients provided both baseline and 4-month EQ-5D-5L amongst whom 1060 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 4-months EQ-5D-5L of 0.094 (bootstrapped 95% CI; 0.046 - 0.146). There is a strong case for the benefit of pay for performance in hip fracture care in terms of improving individual patients' HRQoL.

Clinical relevance: This is the only large prospective cohort study nested within a registry, collecting patient-reported data in addition to those collected by national registries. 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.

Word count 293

Trial Registration: ISRCTN63982700

Funding: The trial was funded by the UK National Institute for Health Research and supported by the NIHR Oxford Biomedical Research Centre.

Introduction

Pay-for-performance initiatives have been used in several healthcare systems in an attempt to drive improvements in patient outcomes. (1,2) In the UK National Health Service (NHS), best practice tariffs (BPT) were introduced with the explicit intention to 'incentivise and adequately reimburse providers' to deliver 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 rising 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 1).(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 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 mirrors the UK national audit augmented with health-related quality-of-life (HRQoL). Recovery of health-related quality-of-life 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.

Item	Name	Description	Categories	Coding
BPT 1	Surgery<36	Surgery within 36 hours of admission to an emergency department	a. No delay, surgery in <36 hours b. Medical delay: awaiting diagnosis, investigation or medical stabilisation c. Administrative delay: lack of available resources d. Other delay	Coded directly
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	Post-operative 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	Pre-operative and post-operative delirium assessments	Yes/no	Attained if two valid AMTS scores were reported

Table 1: Best practice tariff criteria and their determination form the WHiTE dataset

GMC: General Medical Council; MDT: Multi-disciplinary team

Patients and Methods

Study Design Summary

The World Hip Trauma Evaluation (WHiTE) (8) is an observational cohort study that collects information on assessment, treatment and recovery of patients admitted to participating UK NHS hospitals after hip fracture. Full details of the protocol have been published previously. The WHiTE study was approved by research ethics committees (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 2015.

Procedures

Participants enrolled in the WHiTE Cohort study were treated in accordance with local standard care pathways. The National Institute for Clinical Excellence (NICE) have issued standardised care guidelines that are used in the majority of hospitals, and are summarised elsewhere. (11) Adherence to this NICE guideline is 85-98% across all recommendations (with the exception of the use of total hip replacement), (12) and is reported in full annually in the national audit.(13)

Outcomes

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, V3.7, OpenClinica LLC, Waltham, MA, USA). Data were extracted from the database and saved to a comma separated format for analyses.

Health-related quality-of-life

Health-related quality-of-life (HRQoL) was assessed using the Euro-Qol 5 Dimensions 5 Levels Score (EQ-5D-5L); (14) a generic, validated, cross-disciplinary standardised health

utility instrument widely used to assess HRQoL after hip fracture. (15) 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, (i) mobility, (ii) self-care, (iii) usual activities, (iv) pain and discomfort and (v) 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 together provides a 5-digit number that describes the participant's health state. The 5-digit responses, from the EQ-5D health classifications, were converted into an overall score using a published utility algorithm for the United Kingdom population.(16,17) The clinically important difference in EQ-5D derived utility is estimated to be between 0.05 and 0.075.(18) For context, the estimated adjusted disutility associated with asthma, acute myocardial infarction and diabetes melitus are 0.05, 0.06 and 0.07 respectively. (19) Participants in WHiTE were asked to provide (retrospective) pre-fracture assessments of HRQoL, using EQ-5D, at enrolment into the study; whilst there are limitations to this approach, the recall values reported are similar to an age and sex matched UK population.(15)

Some participants were unable to complete the EQ-5D themselves, so this was completed by either the next of kin (NOK), 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,(20) although evidence from more recent studies is mixed.(21,22)

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 determined from the database, are in Table 1.

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 Anaesthesiologists (ASA; I, II, III, IV, V) 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 6000 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 utility of between 0.05 and 0.075 (18), we had power to detect small to moderate effects (23) sizes in our planned multiple regression analyses.

Statistical Analysis

This is an elderly and frail population, with a high mortality at four months post-injury. (24) In line with EuroQoL recommendations, (14) 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(25) and simply EQ-5D hereafter. We tested the impact of this approach through sensitivity analyses reported in Supplementary Table 6 and Figure 1.

Baseline HRQoL is strongly associated with HRQoL at four months post-injury(24,25) 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). (26) 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 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 1. 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) (27) 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 (Vienna, Austria), using the MatchIt package(28) for the PSM analysis, with statistical significance assessed at the 5% level.

Results

Participants

This cohort study comprises data from 8673 participants recruited between May 2014 and April 2017, of whom 7391 provided a baseline EQ-5D and 6532 both baseline and four months EQ-5D. The mean age of WHiTE cohort participants at recruitment was 83 years (standard deviation (sd) = 8.5), and the percentage female to male split was 73:27. (26) 1060 participants died prior to 4 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 generalisability 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. (29)

BPT attainment

The overall attainment of each BPT indicator is shown in Table 2; 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, S*urgery within 36 hours from arrival in an emergency department*. Two indicators were almost universally attained; BPT 3 *Admitted using an agreed multi-disciplinary team (MDT) protocol* and BPT 5 *Post-operative geriatrician-directed MDT rehabilitation* were not attained in fewer than 2% of cases.

Name	Description	Indicator attained	
		No*; n, (%)*	*Yes; n, (%)*
Surgery<36	Surgery within 36 hours of admission to an emergency department	1768 (20.8)	6722 (79.2)
Joint care	Admitted under joint care of an consultant orthopaedic surgeon and geriatrician	506 (5.8)	8167 (94.2)
MDT protocol	Admitted using an agreed MDT protocol	93 (1.1)	8436 (98.9)
Geriatrician<72	Assessed by a geriatrician within 72 hours of admission	561 (6.7)	7773 (93.3)
MDT rehab	Post-operative geriatrician-directed MDT rehabilitation	126 (1.5)	8222 (98.5)
Bone health	Assessment of falls risk and bone health	496 (5.8)	8057 (94.2)
Cognitive assessment	Pre-operative and post-operative delirium assessments	1161 (13.4)	7512 (86.6)
All	-	3771 (43.5)	4902 (56.5)

Table 2: Best practice tariff attainment by indicator

Table 3 shows the participant characteristics for those who attained all seven BPT (*All*) criteria and those that did not. The mean age for those participants attaining all seven BPT criteria was 83.0 years (sd = 8.5) and for those participants not attaining all seven BPT criteria was 82.5 years (sd = 8.6); the mean Abbreviated Mental Test Score (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. There was no evidence that those who did not attain *All* BPT were older, had a lower AMTS or had poorer mobility.

Participant characteristic		***No; n,* (%) (n=3771)**	***Yes; n,* (%) (n=4902)**
Age (years)	<80	1407 (37.3)	1688 (34.4)
	80+	2364 (62.7)	3214 (65.6)
Sex	F	2675 (70.9)	3615 (73.8)
	M	1096 (29.1)	1287 (26.2)
AMTS	0-3: Severe impairment	559 (16.5)	887 (18.1)
	4-6: Moderate impairment	295 (8.7)	433 (8.8)
	7-10: Normal	2537 (74.8)	3582 (73.1)
ASA	I	84 (2.4)	104 (2.2)
	II	866 (25.0)	1441 (30.6)
	III	1966 (56.8)	2620 (55.7)
	IV	539 (15.6)	532 (11.3)
	V	7 (0.2)	6 (0.1)
Mobility (pre-fracture)	No functional mobility	85 (2.3)	106 (2.2)
	Freely mobile without aids	1530 (41.6)	1968 (40.3)
	Mobile outdoors with one aid	903 (24.5)	1160 (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)	Own home/Sheltered housing	3107 (84.3)	4052 (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)

Table 3: Participant characteristics for those who attained all seven BPT criteria (Yes) and those who did not (No)

Health-related quality-of-life

Table 4 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 in 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, Pre-operative and post-operative cognitive assessments*) was associated with statistically significantly 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 - 0.146). A similar magnitude of change in utility is associated with major affective psychoses. (19)

A summary of the results of the PSM and repetition of the analysis reported in Table 4 are reported in the supplementary material Table 8 and Figure 2. Adjusted estimates of differences in four months post-injury EQ-5D were consistent with the estimated effects from the full dataset, as were estimated effects for the other BPT indicators; increases (95% CI) in BPT indicators 2, 6 and 7 were 0.053 (0.016 - 0.089), 0.054 (0.016 - 0.093) and 0.031 (0.007 - 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 0.59-0.85, $p<0.001$) (Table 5, Supplementary material).

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%), that is patients for whom a clinical decision was taken to delay surgery to correct a modifiable

preoperative risk, were much higher than risks for any other reason. Baseline and four-month post-injury EQ-5D were also statistically significantly lower in the *medical delay* group than the *administrative delay* group, that is patients who were delayed only due to non-clinical reasons such as lack of operating time or theatre space (p=0.001 and 0.003 respectively). (Table 7, Supplementary material).

BPT	Mean EQ-5D 4 months			Adjusted Analysis [1]
	No	*Yes*	Difference (95% CI)	p
Surgery<36	0.429	0.430	0.010 (-0.007 - 0.027)	0.254
Joint care	0.387	0.433	0.034 (0.002 - 0.066)	0.035
MDT protocol	0.422	0.431	0.014 (-0.053 - 0.082)	0.673
Geriatrician<72	0.397	0.433	0.026 (-0.003 - 0.054)	0.077
MDT rehab	0.392	0.431	0.015 (-0.047 - 0.076)	0.644
Bone health	0.377	0.433	0.051 (0.019 - 0.082)	0.002
Cognitive assessment	0.403	0.434	0.025 (0.004 - 0.047)	0.021
All	0.419	0.438	0.016 (0.002 - 0.030)	0.026

(1) Regression analysis adjusting for baseline EQ-5D, age, sex, pre-fracture mobility and pre-fracture residence

Table 4: Means and estimates of differences in 4 months EQ-5D between those attaining (Yes) and those not attaining BPT (No) for each BPT indicator

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 people with hip fracture is associated not only with reduced mortality but also with improved health-related quality-of-life 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 *pre-operative and post-operative 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 amongst 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 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 4 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 healthy. The outlook for these patients is very good, because they are in good 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 for example.

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

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, (15,24,26) 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 be important for 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 months HRQoL outcome due to selection bias, the data suggests 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. (30-32) These studies have reported conflicting findings, although the largest found overall that outcomes were better in patients for whom BPT was attained than not. (30) 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 prioritise. (10)

We recognise 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 cost 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. Whilst 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% of participants in this study – we expect and hope that improving this can yield further improvements in patient outcomes.

References

1. Maynard A. The powers and pitfalls of payment for performance. Health Econ. 2012 Jan;21(1):3-12. doi: 10.1002/hec.1810.

2. Ogundeji YK, Bland JM, Policy TSH, 2016. The effectiveness of payment for performance in health care: a meta-analysis and exploration of variation in outcomes. Elsevier

3. Update to best practice tariffs. Available from: https://www.gov.uk/government/news/update-to-best-practice-tariffs

4. Johnell O, Kanis J. An estimate of the worldwide prevalence, mortality and disability associated with hip fracture. Osteoporosis International. 2004;15(11):897–902.

5. 2019/20 National Tariff Payment System: Annex D [Internet]. 1st ed. London: NHS England & NHS Improvement; 2019. Available from: https://improvement.nhs.uk/documents/4986/AnnexD_1920_Guidance_on_BPTs.pdf

6. Johansen A, Golding D, Brent L, Close J, Gjertsen J-E, Holt G, Hommel A, Pedersen AB, Dieter Röck N, Thorngren K-G. Using national hip fracture registries and audit databases to develop an international perspective. Injury. 2017 Oct;48(10):2174–9.

7. Metcalfe D, Zogg CK, Judge A, Perry DC, Gabbe B, Willett K, Costa ML. Pay for performance and hip fracture outcomes: an interrupted time series and difference-in-differences analysis in England and Scotland. Bone Joint J. 2019 Aug;101-B(8):1015–23.

8. Costa ML, Griffin XL, Achten J, Metcalfe D, Judge A, Pinedo-Villanueva R, Parsons N. World Hip Trauma Evaluation (WHiTE): framework for embedded comprehensive cohort studies. BMJ Open. 2016 Oct 21;6(10):e011679.

9. Metcalfe D, Costa ML, Parsons NR, Achten J, Masters J, Png ME, Lamb SE, Griffin XL. Validation of a prospective cohort study of older adults with hip fractures. Bone Joint J. 2019 Jun;101-B(6):708–14.

10. Haywood KL, Griffin XL, Achten J, Costa ML. Developing a core outcome set for hip fracture trials. Bone Joint J. 2014 Aug;96-B(8):1016–23.

11. Hip Fracture: management [Internet]. London: National Clinical Guideline Centre; 2017. Available from: http://www.ncgc.ac.uk

12. Perry DC, Metcalfe D, Griffin XL, Costa ML. Inequalities in use of total hip arthroplasty for hip fracture: population based study. 2016 Apr 27;353:i2021.

13. National Hip Fracture Database [Internet]. [cited 2012 Nov 13]. Available from: http://www.nhfd.co.uk

14. EuroQol [Internet]. EuroQol Group. [cited 2011 Mar 31]. Available from: http://euroqol.org

15. Parsons N, Griffin XL, Achten J, Costa ML. Outcome assessment after hip fracture: is EQ-5D the answer? Bone Joint Res. 2014 Mar 5;3(3):69–75.

16. National Institue of Health and Care Excellence, editor. Position statement on use of the EQ-5D-5L valuation set for England (updated November 2018) [Internet]. www.nice.org.uk. [cited 2018 Dec 1]. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l

17. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997 Nov;35(11):1095–108.

18. 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 Aug 1;14(6):1523–32.

19. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D Scores for the United Kingdom. Medical Decision Making. 2011 Nov 8;31(6):800–4.

20. 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 Sep;67(9):1035–43.

21. Usman A, Lewis S, Hinsliff-Smith K, Long A, Housley G, Jordan J, Gage H, Dening T, Gladman JR, Gordon AL . Measuring health-related quality of life of care home residents: comparison of self-report with staff proxy responses. Age Ageing. 2019 May;48(3):407–13.

22. 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 Apr;7(4):e012779.

23. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates; 1988. 1 p.

24. 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 Mar;97-B(3):372–82.

25. 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 Jan;7(1):1–5.

26. Parsons N, 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 Aug; 9(8): 468–476.

27. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res. 2011 May; 46(3): 399–424.

28. Stuart EA, King G, Imai K, Ho D. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software. University of California, Los Angeles; 2011;42(8).

29. Metcalfe D, Costa ML, Parsons NR, Achten J, Masters J, Png ME, Lamb SE, Griffin XL. Validation of a prospective cohort study of older adults with hip fractures. Bone Joint J. 2019 Jun;101-B(6):708–14.

30. 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 Feb 6;7(2):e014190.

31. 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;9:2097–102.

32. 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 (Lond). 2013 Nov;74(11):644–7.