



Original research

The Relationship Between Preoperative Oxford Hip and Knee Score and Change in Health-Related Quality of Life After Total Hip and Total Knee Arthroplasty: Can It Help Inform Rationing Decisions?

David P. Gwynne-Jones, MA, BM, BCh, FRCS, FRACS (Orth)^{a, b, *}, Trudy Sullivan, PhD^c,
Ross Wilson, PhD^a, J. Haxby Abbott, PhD, FNZCP^a

^a Department of Surgical Sciences, Centre for Musculoskeletal Outcomes Research, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand

^b Department of Orthopaedic Surgery, Dunedin Hospital, Southern District Health Board, Dunedin, New Zealand

^c Dunedin School of Medicine, Dunedin, New Zealand

ARTICLE INFO

Article history:

Received 2 February 2020

Received in revised form

2 March 2020

Accepted 9 April 2020

Available online 21 May 2020

Keywords:

Health-related quality of life

Total hip arthroplasty

Total knee arthroplasty

Outcomes

Prioritization

Rationing

ABSTRACT

Background: In countries with publicly funded health care, there is an increasing need for explicit rationing for total joint arthroplasty (TJA). The Oxford Hip and Knee Scores (OHS/OKS) have been used to set access thresholds for TJA despite not being developed for that purpose. The aim of this study was to determine whether preoperative OHS/OKS can aid rationing decisions by investigating the changes in general health-related quality of life after TJA.

Methods: OHS/OKS, Short Form-12, and Short Form-6D (SF-6D) scores were collected preoperatively and at 1 year postoperatively in a cohort of patients undergoing total hip arthroplasty (THA; n = 713) and total knee arthroplasty (TKA; n = 520). The association between preoperative OHS/OKS and postoperative score and the change in OHS/OKS and SF-6D was investigated, adjusting for age and gender. **Results:** The mean Oxford scores improved from 13.9 to 40.7 (OHS) and 15.6 to 37.4 (OKS). The mean SF-6D improved after THA (0.53 to 0.80) and TKA (0.56 to 0.78) (all $P < .0001$). Poorer preoperative Oxford scores were associated with poorer postoperative OHS/OKS and SF-6D but larger improvements. For every 5 points lower preoperative OHS/OKS, the postoperative SF-6D score was worse by a margin of 0.019 (THA) and 0.023 (TKA).

Conclusions: Preoperative OHS/OKS can help inform rationing decisions. A lower preoperative OHS/OKS will result in greater gains but a lower final outcome score in general health-related quality of life.

© 2020 The Authors. Published by Elsevier Inc. on behalf of The American Association of Hip and Knee Surgeons. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Total hip arthroplasty (THA) and total knee arthroplasty (TKA) are very successful interventions for end-stage osteoarthritis (OA). With an aging population and rising rates of obesity, the demand for THA and TKA is increasing [1,2]. In countries with limited publicly funded health care, there is an increasing need for explicit rationing [3,4]. Almost half of National Health Service trusts in the United Kingdom are now rationing THA and TKA [5]. There are concerns that delaying surgery until a patient has deteriorated to a threshold

score may have a deleterious effect on their final outcome [6], which could be seen as an unintended consequence of rationing.

In New Zealand, Ministry of Health policy requires District Health Boards to complete surgery within 4 months of a decision to offer publicly funded surgery. Those patients who cannot, due to capacity constraints, be operated on within 4 months are declined surgery and returned to the care of their general practitioner (GP). Since 2000, various tools or scoring systems have been used to help prioritize patients with the emphasis on offering surgery to those patients with the worst symptoms. These tools have been validated and shown to be effective but lack discrimination around the threshold score [4]. They are not designed to assess outcomes after surgery.

Condition-specific scores such as the Oxford Hip Score (OHS) and Oxford Knee Score (OKS) were originally developed to assess outcomes after joint replacement [7]. They have been used in some regions in the United Kingdom to determine eligibility for surgery

* Corresponding author. Department of Orthopaedics, Dunedin Hospital, Great King Street, Dunedin, New Zealand. Tel.: + 64 3 474 0999x58618.

E-mail address: david.gwynne-jones@otago.ac.nz

despite not having been designed for that purpose and not shown to be predictive of patient satisfaction after THA and TKA [8].

In addition to condition-specific scores, generic health-related quality of life (HRQoL) measures such as the Short Form-12 (SF-12) [9], the Short Form-6D (SF-6D) [10], and the Euroqol-5D [11] can be used to assess outcomes. The SF-12 [9] has 12 questions covering 8 domains: physical functioning, role participation (physical and emotional), social functioning, bodily pain, mental health, general health, and vitality. It is usually reported as a physical component score (PCS) and a mental component score (MCS). The SF-6D [10] is a preference-based single-index measure of health, derived from the SF-12, which can be used to calculate quality-adjusted life years for use in cost-utility analysis. The SF-6D focuses on 7 of the 8 health domains covered by the SF-12, with only the general health domain not included. The EQ-5D [11] is a preference-based HRQoL measure that describes health across 5 dimensions—mobility, self-care, usual activities, pain/discomfort, and anxiety/depression and is also commonly used in cost-utility analysis.

An advantage of generic HRQoL measures is they can be used to compare outcomes of procedures within orthopaedics or with other specialties, which may help inform resource allocation in a publicly funded health system.

The aim of this study was to determine whether preoperative OHS/OKS can help inform rationing decisions by investigating the changes in general HRQoL after THA and TKA.

Material and methods

Data Set

This cohort study comprises 1233 patients who underwent THA ($n = 713$) or TKA ($n = 520$) at our institution between 2006 and 2010. Patient demographic data including the age, sex, and joint replaced were collected. Patient-reported scores (OHS/OKS and the SF-12) were collected preoperatively and postoperatively (at approximately 1 year) using an arthroplasty audit database (OrthoWave, Stryker, Sydney, Australia). SF-12 PCS and MCS and SF-6D scores were calculated using responses to individual questions from the SF-12 [9,10]. To enable comparability with results from other studies, we also report EQ-5D-3L scores that were mapped from the SF-12 scores [EQ-5D(SF-12)] using a published crossover algorithm [12]. The SF-6D has half the range of the EQ-5D-3L and a correspondingly lower minimum important difference (MID) (0.041) than the EQ-5D (0.074) [13]. The MID for OHS and OKS was taken as 5 points [14]. The MID for SF-12 PCS and MCS was also taken as 5 points [15].

Statistical Analysis

Patients were grouped into 5 bands based on their preoperative Oxford score: <10, 10–14, 15–19, 20–24, and over 25 points. For both THA and TKA, preoperative and postoperative mean and standard deviation were calculated for each patient-reported outcome measure. The change in scores (postoperative minus preoperative) was tested using a paired *t* test.

The association between preoperative Oxford scores and postoperative Oxford, SF-12 PCS and MCS, and SF-6D scores was assessed graphically by plotting the postoperative score and the change in score, for each outcome measure, against preoperative Oxford scores. Curves were fitted using robust locally weighted regression smoothing to visualize the relationship between preoperative and postoperative outcomes [16]. Linear regression, adjusted for age and sex, was used to estimate postoperative outcome scores and change in scores, conditional on preoperative

Oxford score, for the THA and TKA cohorts. All statistical analyses were conducted using R version 3.5.1 [17].

Results

Baseline data were collected from patients who underwent a THA ($n = 713$) or a TKA ($n = 520$). Follow-up surveys were completed at a mean of 13 months postoperatively by 945 patients (THA [$n = 569$]; TKA [$n = 375$]), a completion rate of 77%. There were no significant differences in age, sex, or baseline measures between those completing and those not completing the follow-up survey. Most patients were female, with no significant difference in gender mix between THA and TKA patients ($P = .18$). Patients with hip OA were approximately 3 years younger and had poorer preoperative Oxford ($P = .0001$) and SF-6D scores ($P = .0006$) than patients with knee OA but had no significant difference in SF-12 PCS ($P = .64$) or MCS ($P = .08$).

After surgery, there was a significant improvement in all mean unadjusted outcome scores for both THA and TKA ($P < .0001$) (Table 1). Unadjusted mean scores after THA were significantly better than those after TKA for Oxford scores (3.3, $P < .0001$) and SF-12 PCS (2.2, $P = .009$). There was no significant difference in the postoperative score between THA and TKA cohorts for SF-12 MCS (0.6, $P = .4$), SF-6D (0.02, $P = .063$), or EQ5D (SF-12) (0.028, $P = .067$). The mean improvement was greater for THA than TKA on all scores including OHS/OKS (5, $P < .001$), SF-6D (0.05, $P < .001$), and EQ-5D(SF-12) (0.078, $P < .001$) (Table 1).

Patients with a lower preoperative Oxford score achieved a lower mean postoperative Oxford score, SF-12 PCS, SF-12 MCS, and SF-6D than those with higher preoperative scores, after both THA and TKA. The improvement (gain in scores) was greater in those with lower preoperative scores on all outcome scores (Fig. 1, Table A1).

In adjusted regression models, age was a significant (but nonlinear) predictor of postsurgery improvement in outcome scores for Oxford, SF-12 PCS, and SF-6D for both hips ($P < .001$) and knees ($P < .04$) but not for SF-12 MCS ($P > .15$). (Figure A1). The improvement in OHS was greatest in patients aged 60 years and in OKS for patients aged 70 years. The gain in SF-6D utility was greater in younger patients and declined with increasing age especially after 70 years. Men had significantly smaller improvements than women in Oxford scores for hips ($P = .007$) but not knees. There was

Table 1
Patient-reported outcome measures before and 1 y after joint arthroplasty surgery.

Outcome measure	Preoperative	Postoperative 1 y	Change ^d
	Mean (SD)	Mean (SD)	
Total hip arthroplasty (THA)			
Total Oxford score	13.9 (6.6) ^b	40.7 (7.3) ^b	26.8 (9.2) ^b
SF-12 PCS	28.1 (5.3)	43.5 (11.0) ^a	15.4 (10.9) ^a
SF-12 MCS	43.0 (12.0)	54.4 (9.4)	11.4 (12.8) ^c
SF-6D utility value	0.53 (0.11) ^b	0.80 (0.15)	0.27 (0.17) ^b
EQ-5D from SF-12	0.383(0.22) ^a	0.775 (0.20)	0.392 (0.25) ^b
Total knee arthroplasty (TKA)			
Total Oxford score	15.6 (6.1) ^b	37.4 (8.2) ^b	21.8 (9.3) ^b
SF-12 PCS	28.3 (5.3)	41.3 (10.4) ^a	13.0 (10.3) ^a
SF-12 MCS	44.7 (11.4)	53.8 (9.6)	9.2 (11.8) ^c
SF-6D utility value	0.56 (0.10) ^b	0.78 (0.15)	0.22 (0.15) ^b
EQ-5D from SF-12	0.433 (0.20) ^a	0.747 (0.19)	0.314 (0.25) ^b

^a Indicates difference between the THA and TKA groups is statistically significant $P < .01$.

^b Indicates difference between the THA and TKA groups is statistically significant $P < .001$.

^c Indicates difference between the THA and TKA groups is statistically significant $P = .02$.

^d All changes (preoperative to postoperative) are highly statistically significant ($P < .0001$).

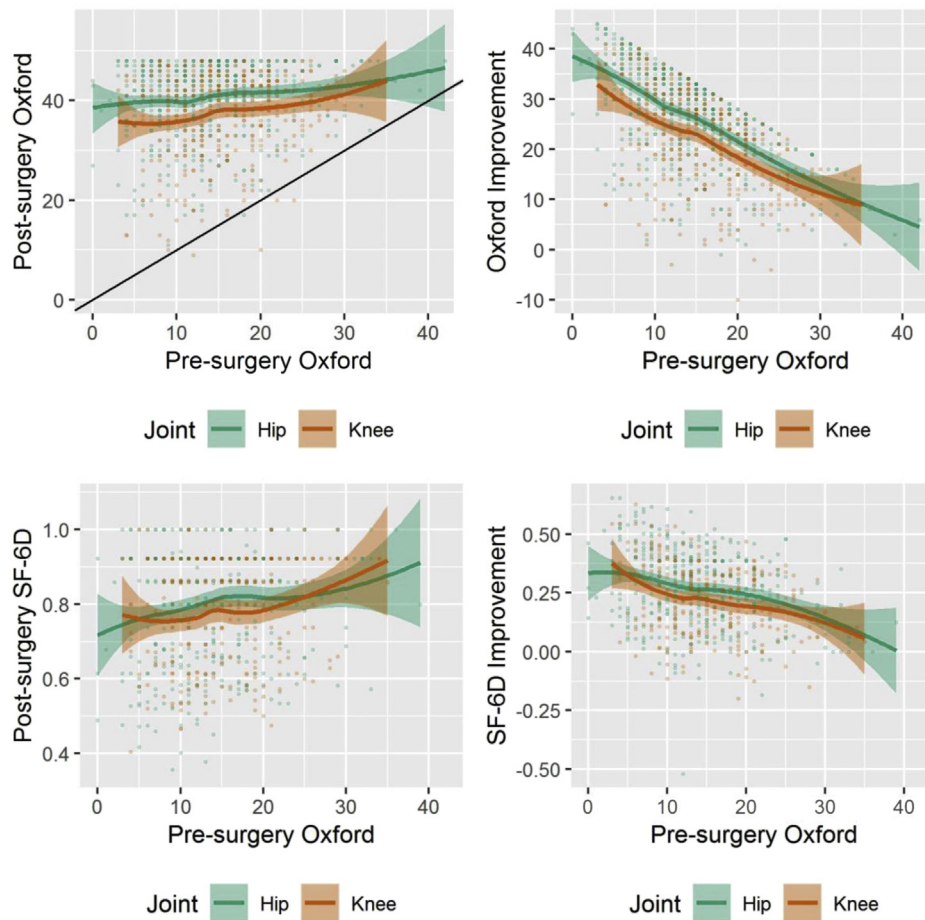


Figure 1. Postsurgery and improvement in total Oxford Hip or Knee Score and SF-6D utility 1 y after surgery, by presurgery total Oxford score (unadjusted with 95% confidence intervals).

no significant difference between men and women for SF-12 or SF-6D outcomes.

After adjusting for age and gender, the mean postoperative Oxford, SF-12 PCS, and SF-6D scores for both THA and TKA were significantly lower for those with poorer Oxford scores at baseline (Table A1). The difference in mean scores between the poorest preoperative group (OHS/OKS < 10) and the best preoperative group (OHS/OKS > 25) was OHS 3.9, OKS 4.6 points, SF-12 PCS 5.0 (THA), 5.4 (TKA), SF-6D 0.08 (THA), and 0.10 (TKA).

In the continuous regression model, the postoperative OHS was 0.8 points and OKS 1.2 points poorer for every 5-point difference in preoperative OHS/OKS (Fig. 2a). Similarly, for SF-6D, the postoperative score was worse by 0.018 (THA) and 0.023 (TKA) for every 5-point decrease in preoperative OHS/OKS (Fig. 2c). The difference in change in the Oxford score between THA and TKA was statistically significant for all preoperative OHS/OKS < 28 points (Fig. 2b). A similar pattern was seen for the gains in SF-6D score according to preoperative OHS/OKS but was not statistically significant because of wide confidence intervals (Fig. 2d).

Discussion

Our results show significant improvement in HRQoL at 1 year after THA and TKA. The postoperative SF-6D and EQ-5D(SF-12) scores after THA and TKA are comparable; however, there is a significantly larger gain after THA than TKA on all outcome scores. A poorer preoperative OHS/OKS is associated with a lower HRQoL

score at 1 year, compared with higher preoperative scores. There was a clinically important difference in postoperative SF-6D scores between the groups with the lowest and highest preoperative Oxford scores, after adjusting for age and gender, that is, more than the MID reported for SF-6D [13]. This suggests that a clinically relevant poorer HRQoL outcome may result from rationing access to arthroplasty on the basis of lower Oxford scores. However, larger HRQoL gains were associated with lower preoperative Oxford scores. These findings are relevant to health services such as ours, in which the use of explicit rationing over recent years has led to the mean preoperative OHS/OKS score falling to around 10 points in our institution [3,4].

It is well established that, for both THA and TKA, lower preoperative Oxford scores are associated with a lower postoperative Oxford score and a higher change in score [7,18–20]. Less has been published on generic HRQoL measures, with most studies reporting EQ-5D scores. Our results show that the mean improvement in SF-6D in our patients after both THA (0.27) and TKA (0.22) is well above the MID of 0.041. This is higher than that previously reported [21–23], mainly due to lower preoperative SF-6D in our cohort. The mean EQ-5D(SF-12) score also improved significantly from 0.383 to 0.775 (a gain of 0.392) after THA and from 0.433 to 0.747 (+0.314) after TKA. These scores and changes were similar or greater than results reported in other studies using the EQ-5D [18,24–26].

Dakin et al reported an improvement in the mean EQ-5D score from 0.39 to 0.71 (+0.32) after TKA, with the greatest gains in those with an OKS in the lowest quintile (OKS < 12) [18]. Subsequently,

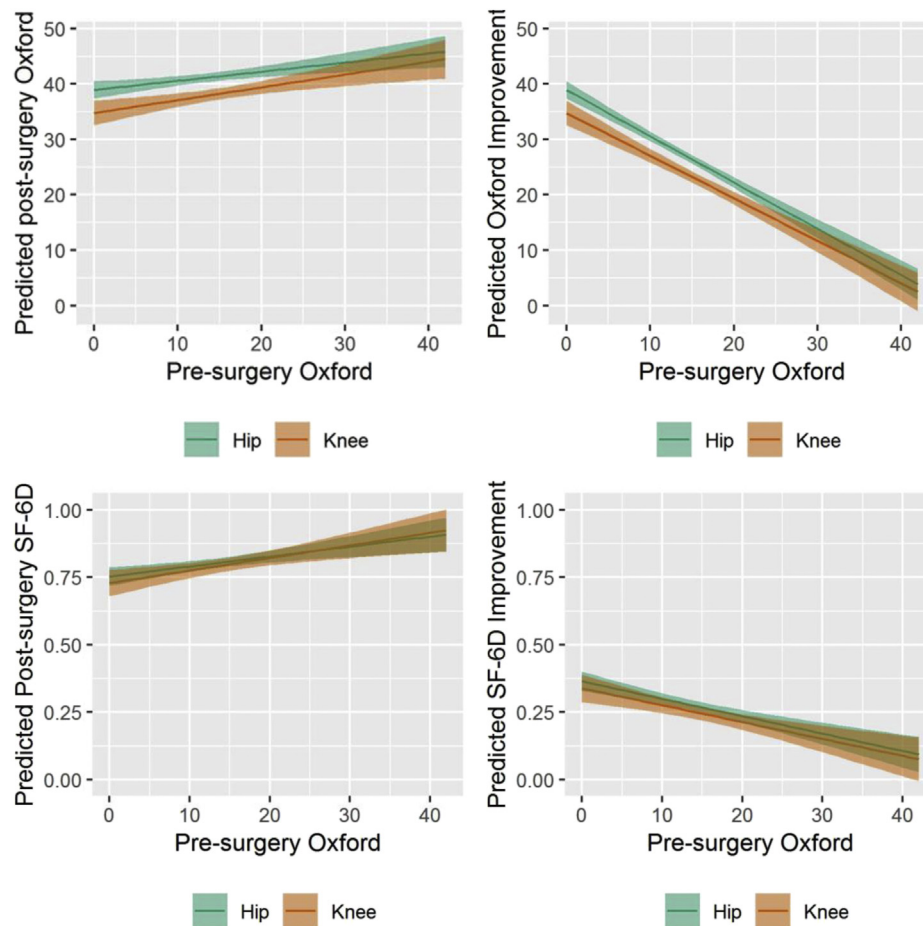


Figure 2. Postsurgery and improvement in total Oxford Hip or Knee Score and SF-6D utility 1 y after surgery, by presurgery total Oxford score (adjusted for age and gender).

Eibich et al. [19] reported similar findings after THA and TKA, also demonstrating that gains in HRQoL (as measured by the EQ-5D) were still evident in patients with preoperative OHS >46 and OKS >44 [19]. Gordon et al. [27], using data from the Swedish hip registry, also reported lower preoperative scores were associated with poorer postoperative EQ-5D scores. Our results, in a sample with poorer mean preoperative scores confirm that a larger gain in HRQoL results in patients with poorer preoperative OHS/OKS, but at a lower final EQ-5D score.

These findings have implications on the cost-effectiveness of total joint arthroplasty in relation to preoperative status. Dakin et al. [18] suggested that if TKAs were to be rationed based on the magnitude of HRQoL gains, OKS would be a reasonable tool to use to set the threshold. They suggested that the most cost-effective preoperative OKS was 12–15 points, but TKA remained cost-effective even in patients with an OKS up to 35–40 points, depending on the American Society of Anesthesiologists grade. Schilling et al concluded that TKA was likely to be cost-effective for most patients except those with an unusually high HRQoL [26]. Ferket et al. [28] found only small improvements in SF-12 PCS (+1.7) and SF-6D HRQoL scores after TKA (+0.008). They suggested that TKA would be more effective if restricted to patients with SF-12 PCS <50 and was most attractive from an economic viewpoint in patients with a score <35 points. In contrast, the mean preoperative PCS in our study was 28.3, with a clinically relevant improvement of 13 points.

Consistent with other reports [19,25], we found greater improvement in OHS/OKS after THA than TKA. The differences in change in OHS/OKS (5 points), SF-6D (0.050), and EQ-5D(SF-12) (0.078) are at or above the reported MID for these scores [13,14]. Our results suggest that to achieve a similar gain in both Oxford score and HRQoL at 1 year, a patient with knee OA should have an Oxford score 3–4 points worse than a patient with hip OA. This was statistically significant for the Oxford score but did not reach statistical significance for SF-6D utility despite showing the same pattern. Jenkins et al reported similar findings with patients undergoing TKA needing an OKS 8 points lower than for a patient undergoing THA to offer the same value for money over a patient's lifetime [25]. This gives some limited support to prioritizing THA over TKA.

Oxford scores alone should not be used to determine access to THA and TKA. Patient-reported scores such as the Oxford score may be open to “gaming” by a patient or referring GP if it becomes known that they are being used to determine whether a patient qualifies for surgery. The decision to offer surgery is complex and should involve clinical and radiological assessment of the patient by an orthopaedic surgeon. Increasingly, however, some form of gate-keeping is required, which may be by surgeons, managers, commissioners, or GPs. This study shows that preoperative Oxford score can help inform these rationing decisions. A patient with a poorer preoperative score will gain more than a patient with a higher score, which may be seen as cost-effective. However, a

system which requires a patient to wait until they have deteriorated to a poorer score may prejudice their final outcome and increase indirect costs. The results are a reflection of the New Zealand health care system where patients are prioritized by severity. We used data from 2006 to 2010 before rationing became so severe. Oxford scores were not used to determine access during this period, and therefore, we do not believe that patients inflated their scores. Despite the lack of explicit rationing, there were relatively few patients with Oxford scores over 25 points, and the mean preoperative scores are lower than those reported in the National Joint Registry of England, Wales, Northern Ireland and the Isle of Man [29] and by other authors [8,15,18,19,25]. However, the results should be generalizable to other public health systems that are under financial pressure. Although a total joint arthroplasty may be cost-effective in patients with a preoperative Oxford score of >35 points as reported by others [18,19], there is a more limited gain in HRQoL. It is likely that, increasingly, publicly funded services will not be able to routinely offer surgery to these patients.

Strengths of this study are that we used prospectively gathered data and were able to compare the SF-6D measure with the condition-specific Oxford scores that have been widely used elsewhere. Although the OHS/OKS were not designed to be used for rationing, in practice, they have been widely used to dictate thresholds. Therefore, we believe this study is relevant to the current debate on rationing and gives some support for their use.

Limitations to this study include the observational design and the absence of comorbidity data, which may influence both preoperative and postoperative HRQoL scores. As patients did not complete an EQ-5D questionnaire, the EQ-5D-3L scores used in this study were derived from the SF-12 scores. They were calculated to enable comparisons with published studies and showed similar changes. Because the algorithm was based on UK preference weights and full cost data were unavailable, quality-adjusted life years were not calculated. The 1-year response rate of 77% is not uncommon in large observational studies but may lead to some bias. However, we found no difference in baseline variables between responders and nonresponders. The maximal benefit may occur after 12 months, so these results may not be fully representative of the maximal outcome. However, the use of 1-year data is consistent with other reports and therefore allows comparison. A longer follow-up increases the chances of other conditions developing that may impact the more general HRQoL scores.

Conclusions

Rationing for joint replacement, while unpalatable to many surgeons, is inevitable in publicly funded systems where demand exceeds capacity. This study shows that, despite not being designed for the purpose, OHS/OKS can help inform these decisions. Delaying access to a patient until they have deteriorated to a level where they have a greater gain after surgery may come at the expense of not achieving the same outcome as patients with less severe symptoms.

Conflict of interest

The authors declare there are no conflicts of interest. Dunedin Hospital receives an educational grant from DePuy Synthes to support an Arthroplasty Fellow. The authors do not believe that this has any relevance to the contents of this article.

References

- [1] Hooper G, Lee AJ-J, Rothwell A, Frampton C. Current trends and projections in the utilization rates of hip and knee replacement in New Zealand from 2001 to 2026. *N Z Med J* 2014;127:82.
- [2] Culliford D, Maskell J, Judge A, Cooper C, Prieto-Alhambra D, Arden NK. Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink. *Osteoarthritis Cartilage* 2015;23:594.
- [3] Gwynne-Jones DP, Iosua E. Rationing of hip and knee replacement: effect on the severity of patient reported symptoms and the demand for surgery in Otago. *N Z Med J* 2016;129:59.
- [4] Gwynne-Jones DP, Iosua E, Stout KM. Rationing for total hip and knee arthroplasty using the New Zealand orthopaedic association score: effectiveness and comparison with patient reported scores. *J Arthroplasty* 2016;31(5):957.
- [5] Royal College of Surgeons Political. <https://updates-rcseng.co.uk/4D4N-IXUW-FA117Y9T0C/cr.aspx>; 2019. [Accessed 8 April 2019].
- [6] Garbuz DS, Xu M, Duncan CP, Masri BA, Sobolev B. Delays worsen quality of life outcome of primary total hip arthroplasty. *Clin Orthop Relat Res* 2006;447:79.
- [7] Murray DW, Fitzpatrick R, Rogers K, et al. The use of the Oxford hip and knee scores. *J Bone Joint Surg* 2007;89-B:1010.
- [8] Judge A, Arden NK, Price A, et al. Assessing patients for joint replacement. *J Bone Joint Surg* 2011;93B:1660.
- [9] Ware Jr JE, Kosinski M, Turner-Bowker DM, Gandek B. How to score version 2 of the SF-12® health survey (with a supplement documenting version 1). Lincoln, RI: QualityMetric Incorporated; 2002.
- [10] Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care* 2004;42(9):851.
- [11] EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16(3):199.
- [12] Gray AM, Rivero-Arias O, Clarke PM. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Med Decis Making* 2006;26(1):18.
- [13] 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;14:1523.
- [14] Beard DJ, Harris K, Dawson J, et al. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol* 2015:73.
- [15] Clement ND, MacDonald D, Simpson AH. The minimal clinically important difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc* 2014;22(8):1933.
- [16] Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979;74(368):829.
- [17] R Core Team. R: A language and environment for statistical computing, v3.5.1. Vienna, Austria: R Foundation for Statistical Computing; 2018. <http://www.R-project.org/>.
- [18] Dakin H, Gray A, Fitzpatrick R, MacLennan G, Murray D, KAT Trial Group. Rationing of total knee replacements: a cost effectiveness analysis on a large trial data set. *BMJ* 2012;2:e000332.
- [19] Eibich P, Dakin HA, Price AJ, Beard D, Arden NK, Gray AM. Associations between preoperative Oxford hip and knee scores and costs and quality of life of patients undergoing primary total joint replacement in the NHS England: an observational study. *BMJ Open* 2018;8:e019477.
- [20] Hajat S, Fitzpatrick R, Morris R, et al. Does waiting for total hip replacement matter? Prospective cohort study. *J Health Serv Res Policy* 2002;7(1):19.
- [21] Hamilton DF, Clement ND, Burnett R, et al. Do modern total knee replacements offer better value for money? A health economic analysis. *Int Orthop* 2013;37:2147.
- [22] Elmallah RK, Chughtai M, Khloppas A, et al. Determining cost-effectiveness of total hip and knee arthroplasty using the Short Form-6D utility measure. *J Arthroplasty* 2017;32:351.
- [23] Liebs TR, Herzberg W, Ruther W, Russlies M, Hassenpflug J. Quality-adjusted life years gained by hip and knee replacement surgery and its aftercare. *Arch Phys Med Rehabil* 2016;97:691.
- [24] Rolfson O, Karrholm J, Dahlberg LE, Garellick G. Patient-reported outcomes in the Swedish hip arthroplasty register. Results of a nationwide prospective observational study. *J Bone Joint Surg Br* 2011;93(B):867.
- [25] Jenkins PJ, Clement ND, Hamilton DF, Gaston P, Patton JT, Howie CR. Predicting the cost-effectiveness of total hip and knee replacement. A health economic analysis. *Bone Joint J* 2013;95(B):115.
- [26] Schilling CG, Dowsey MM, Petrie DJ, Clarke PM, Choong PF. Predicting the long-term gains in health-related quality of life after total knee arthroplasty. *J Arthroplasty* 2017;32:395.
- [27] Gordon M, Greene M, Frumento P, Rolfson O, Garellick G, Stark A. Age and health related quality of life after total hip replacement. Decreasing gains in patients above 70 years of age. *Acta Orthop* 2014;85(3):244.
- [28] Ferket BS, Feldman Z, Zhou J, Oei EH, Bierma-Zeinstra SMA, Mazumdar M. Impact of total knee replacement practice: cost effectiveness analysis of data from the osteoarthritis initiative. *BMJ* 2017;356:1.
- [29] National Joint Registry for England, Wales, Northern Ireland and the Isle of Man. In: 15th Annual report; 2018. <https://www.njr.org.uk/wp-content/uploads/2018/11/NJR-15th-Annual-Report-2018.pdf>. [Accessed 11 May 2020].

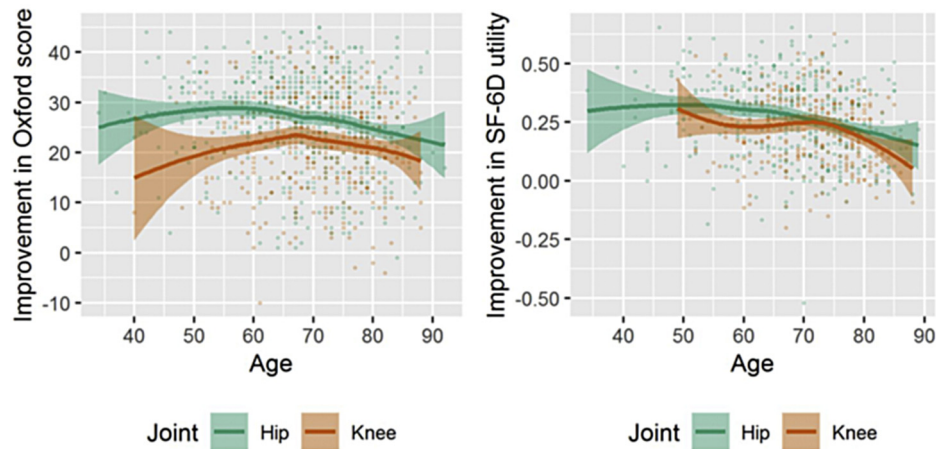


Figure A1. Improvement in Oxford score and SF-6D utility by age of the patient (unadjusted with 95% confidence intervals).

Table A1
Adjusted outcomes after joint arthroplasty surgery, by presurgery Oxford score.

Outcome measure	Presurgery Oxford score				
	<10	10-14	15-19	20-24	25+
Postoperative scores 1 y					
Oxford					
Hip	39.9 (38.5 to 41.2)	40.6 (39.4 to 41.9)	42.4 (41.0 to 43.7) ^b	42.2 (40.3 to 44.2) ^a	43.8 (41.3 to 46.4) ^b
Knee	37.1 (35.1 to 41.2)	37.3 (35.7 to 41.9)	38.7 (37.1 to 43.7)	39.5 (37.5 to 44.2)	41.7 (39.0 to 46.4) ^b
PCS					
Hip	41.7 (39.6 to 45.2)	42.5 (40.5 to 44.5)	46.5 (44.3 to 48.8) ^c	47.2 (44.3 to 50.1) ^b	46.7 (42.7 to 51.6) ^a
Knee	42.1 (38.9 to 45.2)	41.9 (39.4 to 44.5)	41.9 (39.2 to 48.8)	43.5 (40.1 to 50.1)	47.5 (43.5 to 51.6) ^a
MCS					
Hip	53.0 (51.1 to 56.8)	54.8 (52.9 to 56.6)	55.1 (53.0 to 57.5)	54.3 (51.6 to 60.1)	56.3 (52.5 to 62.0)
Knee	53.9 (51.0 to 56.8)	52.9 (50.6 to 56.6)	55.1 (52.6 to 57.5)	57.0 (53.8 to 60.1)	58.2 (54.5 to 62.0)
SF-6D					
Hip	0.77 (0.74 to 0.83)	0.80 (0.77 to 0.83)	0.83 (0.79 to 0.86) ^a	0.83 (0.79 to 0.88) ^a	0.85 (0.79 to 0.94) ^a
Knee	0.78 (0.74 to 0.83)	0.78 (0.75 to 0.83)	0.79 (0.76 to 0.86)	0.83 (0.78 to 0.88)	0.88 (0.82 to 0.94) ^b
Average improvement in outcome scores after surgery					
Oxford					
Hip	33.4 (32.0 to 34.7)	28.6 (27.3 to 29.8) ^c	25.7 (24.3 to 27.1) ^c	20.5 (18.5 to 22.5) ^c	14.4 (11.8 to 17.0) ^c
Knee	29.8 (27.7 to 34.7)	25.2 (23.6 to 29.8) ^c	21.7 (20.1 to 27.1) ^c	17.9 (16.0 to 22.5) ^c	14.2 (11.4 to 17.0) ^c
PCS					
Hip	14.9 (12.8 to 18.0)	15.3 (13.3 to 17.3)	18.6 (16.3 to 20.9) ^a	18.3 (15.4 to 21.2)	11.5 (7.4 to 19.8)
Knee	14.9 (11.7 to 18.0)	13.8 (11.3 to 17.3)	13.7 (11.1 to 20.9)	13.5 (10.1 to 21.2)	15.7 (11.6 to 19.8)
MCS					
Hip	17.1 (14.7 to 20.8)	12.3 (10.0 to 14.6) ^b	7.0 (4.4 to 11.2) ^c	5.7 (2.4 to 12.0) ^c	6.2 (1.6 to 10.9) ^c
Knee	17.2 (13.6 to 20.8)	10.9 (8.1 to 14.6) ^b	8.2 (5.1 to 11.2) ^c	8.1 (4.3 to 12.0) ^c	4.7 (0.0 to 10.9) ^c
SF-6D					
Hip	0.32 (0.29 to 0.37)	0.28 (0.25 to 0.31) ^a	0.26 (0.23 to 0.30) ^b	0.24 (0.20 to 0.29) ^b	0.15 (0.09 to 0.24) ^c
Knee	0.32 (0.27 to 0.37)	0.25 (0.21 to 0.31) ^a	0.22 (0.18 to 0.30) ^c	0.22 (0.17 to 0.29) ^b	0.18 (0.12 to 0.24) ^c

Adjusted outcome for the presurgery category significantly different to the reference (Oxford <10) category: ^a $P < .05$; ^b $P < .01$; ^c $P < .001$.

Adjusted outcome for the knee arthroplasty cohort significantly different to the hip arthroplasty cohort: ^a $P < .05$; ^b $P < .01$; ^c $P < .001$.

No patients with a presurgery Oxford score of 25+ had a postsurgery score <27.