



## ■ HIP

# The metabolic equivalent of task score

## A USEFUL METRIC FOR COMPARING HIGH-FUNCTIONING HIP ARTHROPLASTY PATIENTS

**T. C. Edwards,  
B. Guest,  
A. Garner,  
K. Logishetty,  
A. D. Liddle,  
J. P. Cobb**

From Imperial College  
London, London, UK

### Aims

This study investigates the use of the metabolic equivalent of task (MET) score in a young hip arthroplasty population, and its ability to capture additional benefit beyond the ceiling effect of conventional patient-reported outcome measures.

### Methods

From our electronic database of 751 hip arthroplasty procedures, 221 patients were included. Patients were excluded if they had revision surgery, an alternative hip procedure, or incomplete data either preoperatively or at one-year follow-up. Included patients had a mean age of 59.4 years (SD 11.3) and 54.3% were male, incorporating 117 primary total hip and 104 hip resurfacing arthroplasty operations. Oxford Hip Score (OHS), EuroQol five-dimension questionnaire (EQ-5D), and the MET were recorded preoperatively and at one-year follow-up. The distribution was examined reporting the presence of ceiling and floor effects. Validity was assessed correlating the MET with the other scores using Spearman's rank correlation coefficient and determining responsiveness. A subgroup of 93 patients scoring 48/48 on the OHS were analyzed by age, sex, BMI, and preoperative MET using the other metrics to determine if differences could be established despite scoring identically on the OHS.

### Results

Postoperatively the OHS and EQ-5D demonstrate considerable negatively skewed distributions with ceiling effects of 41.6% and 53.8%, respectively. The MET was normally distributed postoperatively with no relevant ceiling effect. Weak-to-moderate significant correlations were found between the MET and the other two metrics. In the 48/48 subgroup, no differences were found comparing groups with the EQ-5D, however significantly higher mean MET scores were demonstrated for patients aged < 60 years (12.7 (SD 4.7) vs 10.6 (SD 2.4),  $p = 0.008$ ), male patients (12.5 (SD 4.5) vs 10.8 (SD 2.8),  $p = 0.024$ ), and those with preoperative MET scores > 6 (12.6 (SD 4.2) vs 11.0 (SD 3.3),  $p = 0.040$ ).

### Conclusion

The MET is normally distributed in patients following hip arthroplasty, recording levels of activity which are undetectable using the OHS.

**Cite this article:** *Bone Joint Res* 2022;11(5):317–326.

**Keywords:** PROMs, Ceiling effect, Hip arthroplasty

### Article focus

- Substantial ceiling effects of some conventional patient-reported outcome measure (PROM) scores limit their ability to discriminate between high-functioning postoperative hip arthroplasty patients.
- This article introduces the concept of using metabolic equivalent of task (MET) values in a young hip arthroplasty cohort

to determine whether this approach can capture additional benefit beyond the ceiling effect of conventional PROMs.

### Key messages

- Postoperative ceiling effects of the Oxford Hip Score (OHS) and EuroQol five-dimension questionnaire (EQ-5D)

Correspondence should be sent to  
Thomas C Edwards; email:  
thomas.edwards@imperial.ac.uk

doi: 10.1302/2046-3758.115.BJR-  
2021-0445.R1

*Bone Joint Res* 2022;11(5):317–  
326.

miss clinically substantial gains in higher-level health-related quality of life.

- MET is a simple, versatile measure of physical activity, with no postoperative ceiling effect.
- Clinicians, researchers, and health economists who wish to capture the full benefit from hip arthroplasty surgery should consider using MET in addition to conventional PROMs.

### Strengths and limitations

- This multicentre, multiple-surgeon study addresses an important concept, which will inform future hip arthroplasty researchers designing studies assessing new technologies or techniques.
- The large proportion of non-responders and included cohort being substantially younger than the national average for hip arthroplasty limits its generalizability to the wider population.

### Introduction

Primary hip arthroplasty is an effective intervention for improving pain and restoring function.<sup>1</sup> Patient-reported outcome measures (PROMs) such as the Oxford Hip Score (OHS),<sup>2,3</sup> which is routinely collected before and after hip arthroplasty in the UK, reliably and predictably report considerable, cost-effective improvements in pain and function.<sup>4,5</sup> However, as a consequence of the efficacy of hip arthroplasty, the distribution of postoperative scores is highly negatively skewed. In two national registries, the modal score on the postoperative OHS was 100% (48/48 points) with up to 20% of patients recording this score.<sup>6,7</sup>

The population of patients presenting for hip arthroplasty has evolved and their expectations are different from those undergoing surgery when the OHS was introduced.<sup>8</sup> In a study by Scott et al,<sup>9</sup> 40% of hip arthroplasty patients considered returning to sporting activity a 'very important' preoperative expectation, but such activities are not captured by the OHS. By relying solely on this skewed metric, potential health gain in return to sporting activity from innovative techniques will remain undetected, while other patient groups may be inappropriately told that their clinical results are as good as they can get, despite dissatisfaction with the level of activity that they have achieved.<sup>7,10,11</sup> In the UK, the pre- to postoperative change in OHS is used by government-backed initiatives including 'getting it right first time' (GIRFT) and the NHS best practice tariff (BPT),<sup>12,13</sup> to measure success at the institutional level. With the skewed OHS as the metric, the only way to improve the health gain obtained is by refusing care until the preoperative scores are low enough. This approach may have a role in healthcare rationing but should not impede the scientific desire to measure higher-level function. As degree of health gain is so closely related to preoperative score, and as preoperative score may vary, there is a need for a metric which

measures outcome equally independently of preoperative health state.

Alternative scores have been developed with the aim of being more discriminative in high-functioning patients. Several constructs have been suggested, such as joint perception as measured by the Forgotten Joint Score (FJS) and physical activity scores. The FJS assesses patients' awareness of their joint arthroplasty performing different tasks, with the optimal outcome being a 'forgotten' artificial joint. In a study comparing the outcomes of robotic and manual THA, the authors found no clinically relevant difference using the OHS, however the robotic group did substantially better using the FJS.<sup>14</sup> The authors support the idea that the ceiling effect of the OHS limits its use for comparing high-functioning postoperative patients, with their results indicating that the FJS may be more discriminative. While this sounds encouraging, other authors have reported ceiling effects of 20% to 30% using the FJS in a postoperative hip arthroplasty cohort, and so problems may still exist with skewed distributions using this score.<sup>15,16</sup>

Physical activity metrics may be another solution, with a number of valid and reliable metrics such as the University of California, Los Angeles (UCLA) activity scale available.<sup>17</sup> This score appears to have no ceiling effect and is simple to use, however it only includes a small number of activities and does not account for the individual activity intensity.<sup>17</sup> One potential solution to this problem is the use of metabolic equivalent of task (MET) values, which numerically quantify the energy expenditure of over 800 activities comparing them to energy expenditure at rest.<sup>18</sup> This sophisticated, personalized approach to quantifying activity energy expenditure has been validated as a surrogate for general cardiovascular fitness, correlating well with both objective activity measures, such as pedometers, as well as the development of cardiovascular disease and mortality.<sup>19,20</sup> Exercise at an intensity that raises the heart rate is now well established as being an effective health maintenance intervention. MET values have been used to confirm this beneficial effect: in a twin study, 'conditioning exercise' offers substantial protection against risk of death when compared with sedentary or occasional exercising.<sup>20</sup> Although not yet commonly used following arthroplasty surgery, with a simplification to measure activity intensity without the performance time or frequency, the MET may be a robust way of comparing activity in postoperative hip arthroplasty patients, demonstrating activity levels that have real relevance to health and life expectancy.

This study therefore aims to answer two important questions: 1) does the MET have a postoperative ceiling effect that may limit its ability to discriminate between high-performing postoperative patients; and 2) can the MET demonstrate continued improvement and health gains beyond the maximal OHS, establishing differences between postoperative patients who score 48/48?

## Methods

**Study design.** This study was a retrospective analysis of anonymous data, collected prospectively from consenting primary hip arthroplasty patients as part of an ongoing, longitudinal study of gait analysis in lower limb arthroplasty (REC reference: 14/SC/1243). Patients from this study were eligible for inclusion if they underwent a primary hip arthroplasty under one of 13 surgeons at 12 sites between 2014 and 2018. Patients were excluded if they had revision surgery, an alternative hip procedure, or incomplete data either preoperatively or at one-year follow-up. Demographic data and the patient-reported answers to three PROMs questionnaires (EuroQol five-dimension questionnaire (EQ-5D), OHS, and the MET) were recorded preoperatively and at one year postoperatively.

**Demographic details.** A total of 751 patients were initially identified on our electronic database; 73 were excluded having had an alternative or revision procedure, and a further 457 patients were excluded due to lack of preoperative or one-year PROM scores. Overall 221 patients, including 117 THAs (53%) and 104 HRAs (47%), with a mean age of 59 years (SD 11), were analyzed in this study. Demographic data are detailed in Table I. The 221 responding patients with full datasets were a mean four years younger than the 457 non-responders (59 years (SD 11) vs 63 years (SD 12),  $p < 0.001$ , independent-samples  $t$ -test).

**MET score.** Using a similar methodology to Amstutz and Le Duff,<sup>21</sup> the MET asks patients to choose three physical activities that are important to them, and that are affected by their joint problem. These initially selected activities remain the same at all follow-up timepoints. Patients then rate the intensity at which they currently perform the activity on a visual scale from 0 to 100. METs are numerical values assigned to demonstrate the energy expenditure used performing different tasks. One MET is equivalent to energy expenditure during rest and is approximately equal to  $3.5 \text{ ml O}_2 \text{ kg}^{-1} \text{ min}^{-1}$  in adults.<sup>19</sup> Using Arizona State Universities compendium of activities,<sup>18</sup> the MET values associated with each activity are recorded. An example is running, which has a range of values between 4.5 (jogging on a mini-tramp) and 23 METs (running a 4.3 minute mile). Based on this reference range, the patient's self-reported intensity score is then used to work out a value for the METs they are currently doing. This is done by subtracting the lower value of the MET reference range from the higher value, multiplying this by the percentage intensity expressed as a decimal, and then adding back on the lower reference value. Using the above as an example, if a patient rated their intensity as 50% in running (range 4.5 to 23 METs) their MET score would be worked out as  $((23 \text{ to } 4.5) * 0.5) + 4.5 = 13.75$  METs. The MET is the maximum value scored from the three chosen activities. In the full MET, the frequency and duration of physical activity are recorded; we omitted these aspects from the score in favour of intensity,

**Table I.** Demographics and patient-reported outcome measures.

Variable	Data (n = 221)
Mean age, yrs (SD, range)	59.4 (11.3, 31 to 84)
<b>Sex, n (%)</b>	
Male	120 (54.3)
Female	101 (45.7)
<b>Laterality, n (%)</b>	
Right	103 (46.6)
Left	110 (49.8)
Bilateral	8 (3.6)
Mean BMI, $\text{kg/m}^2$ (SD, range)*	26.4 (4.9, 16.0 to 57.5)
<b>Operation, n (%)</b>	
THA	117 (52.9)
HRA	104 (47.1)
<b>OHS</b>	
Mean preoperative (SD, range)	25.6 (8.3, 6 to 48)
Median 1 yr (IQR)†	47 (44 to 48)
<b>EQ-5D</b>	
Median preoperative (IQR)†	0.57 (0.30 to 0.65)
Median 1 yr (IQR)†	1 (0.77 to 1)
<b>MET</b>	
Median preoperative (IQR)†	4.9 (0 to 7.5)
Mean 1 yr (SD, range)	10.7 (3.8, 0 to 23)

\*Missing data, n = 163.

†Non-parametric variable, data presented as median (interquartile range), otherwise data reported as mean (standard deviation, range) as indicated.

EQ-5D, EuroQol five-dimension questionnaire; HRA, hip resurfacing arthroplasty; IQR, interquartile range; MET, metabolic equivalent of task; OHS, Oxford Hip Score; SD, standard deviation; THA, total hip arthroplasty.

to avoid measuring cardiorespiratory fitness and also to avoid under-representing performance through measuring high-intensity but infrequently performed activities, such as skiing.<sup>21,22</sup>

**Distribution of scores.** Data were analyzed to demonstrate the distribution, presence of ceiling or floor effects, concurrent validity of the MET in terms of its responsiveness, and correlations between the MET and the two conventional PROMs. Other authors have suggested when validating physical activity metrics that a weak-to-moderate correlation would be expected between the activity metric and conventional PROMs.<sup>8</sup>

**Health gains.** Recent literature has established that as preoperative OHS increases, the improvement in score decreases.<sup>23</sup> To investigate whether health gains in the MET and EQ-5D are also limited by the level of preoperative joint symptoms, the relationship between preoperative OHS and patient improvement at one year using the three metrics was plotted. Fractional polynomial regression plots were used to demonstrate the likely increase in each metric for a given preoperative OHS score.

**48/48 sub-cohort analysis.** A subgroup analysis was performed on a cohort of patients with the maximum postoperative OHS. Previous studies have highlighted that postoperative physical activity (as measured on the UCLA activity score) in hip arthroplasty patients may be

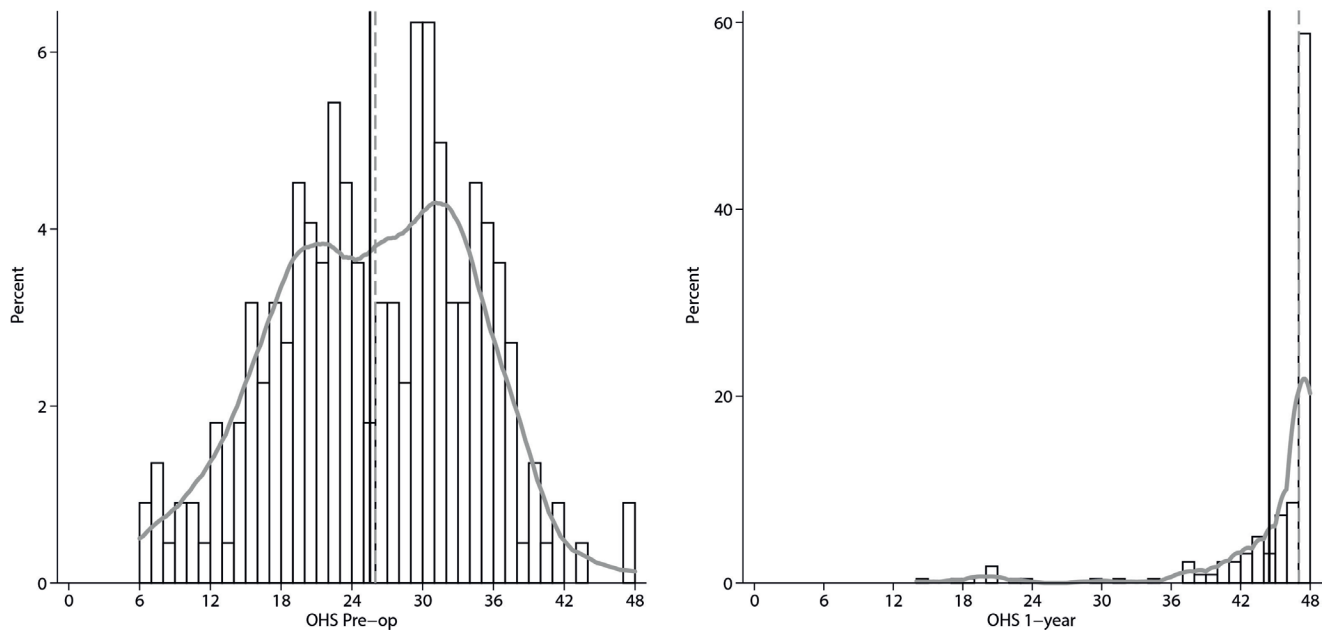


Fig. 1

Histograms with kernel (Epanechnikov) density plots demonstrating distribution of Oxford Hip Scores (OHS) preoperatively and at one-year follow-up. Solid vertical lines represent mean values, dashed vertical lines represent the median.

higher in younger patients, male patients, those with higher preoperative activity levels, and those with lower BMI.<sup>24,25</sup> Based on this, the 48/48 scoring patient cohort was divided into categories (age < or > 60 years, male or female, preoperative MET < or > 6, and BMI < or > 25 kg/m<sup>2</sup>) and compared using the MET and EQ-5D at one year postoperatively. Previous literature has classified activities as light (< 3), moderate (3 to 6), or vigorous (> 6) according to their MET values.<sup>26,27</sup> Therefore, in this analysis, the threshold for high preoperative activity was set at 6 METs.

**Statistical analysis.** Statistical analysis was performed using Stata/IC 10.1 (StataCorp, USA). Data were first tested for normality visually using histograms and normal Q-Q plots. To quantify the shape and symmetry of the distribution about the mean, kurtosis and skewness values were calculated. A standard normal distribution is generally considered to have a kurtosis value of 3 and a skew value of 0.<sup>28</sup> For independent data, parametric variables were compared using the independent-samples *t*-test, and non-parametric data were compared using the Mann-Whitney U test. Paired data were compared using the paired *t*-test. Ceiling and floor effects were calculated as the percentage of patients scoring the maximum or minimum scores, respectively. As previously indicated in the literature, ceiling or floor effects of > 15% were considered relevant.<sup>7</sup> Construct validity of the MET was assessed by examining the responsiveness of the score to change using the standardized response mean (SRM; calculated by dividing the mean change in score over the one-year time period by the standard deviation (SD) of that change), and concurrent validity was assessed by measuring correlations between scores as calculated using Spearman's rank correlation coefficient.  $R_s$  values of

0.3 to 0.5 were considered as weak-to-moderate correlations.<sup>8</sup> Statistical significance was set at  $p < 0.05$ .

## Results

**Distribution.** Preoperatively, the distribution of the OHS was normal (skew -0.14, kurtosis 2.57) and that of the EQ-5D was bimodal with a skewness value of -0.99 and a kurtosis of 4.03 (Figures 1 and 2). The MET demonstrates a slight positive skew of 0.59 representing a floor effect, with the commonest score being zero, on account of hip pain and near-normal kurtosis of 3.18 (Figure 3).

Postoperatively, both the OHS and EQ-5D scores demonstrate a substantial negative skew (Figures 1 and 2). The kurtosis value of the OHS was 13.07 with a skewness value of -3.12, while the EQ-5D demonstrated a kurtosis value of 6.23 and a skewness value of -1.96. The MET on the other hand exhibited a normal distribution postoperatively, centred around a mean of 10.7 (SD 3.8) with very little skew (0.40) and near normal kurtosis (4.46; Figure 3 and Table II).

**Ceiling and floor effects.** No floor effects were seen for the OHS or EQ-5D, but substantial ceiling effects of 41.6% and 53.8% were seen at one year follow-up in the OHS and EQ-5D, respectively. Preoperatively the MET had a moderate floor effect of 25.3%, while no relevant ceiling effect was noted postoperatively (Table II).

**Validity.** Spearman's rank correlation coefficient was weak-to-moderate, but there were statistically significant correlations between the MET and EQ-5D both preoperatively ( $r_s = 0.46$ ,  $p < 0.001$ ) and to a lesser extent at one-year follow-up ( $r_s = 0.32$ ,  $p < 0.001$ , Spearman's rank correlation). Similarly, weak-to-moderate correlations were demonstrated between the MET and the OHS both

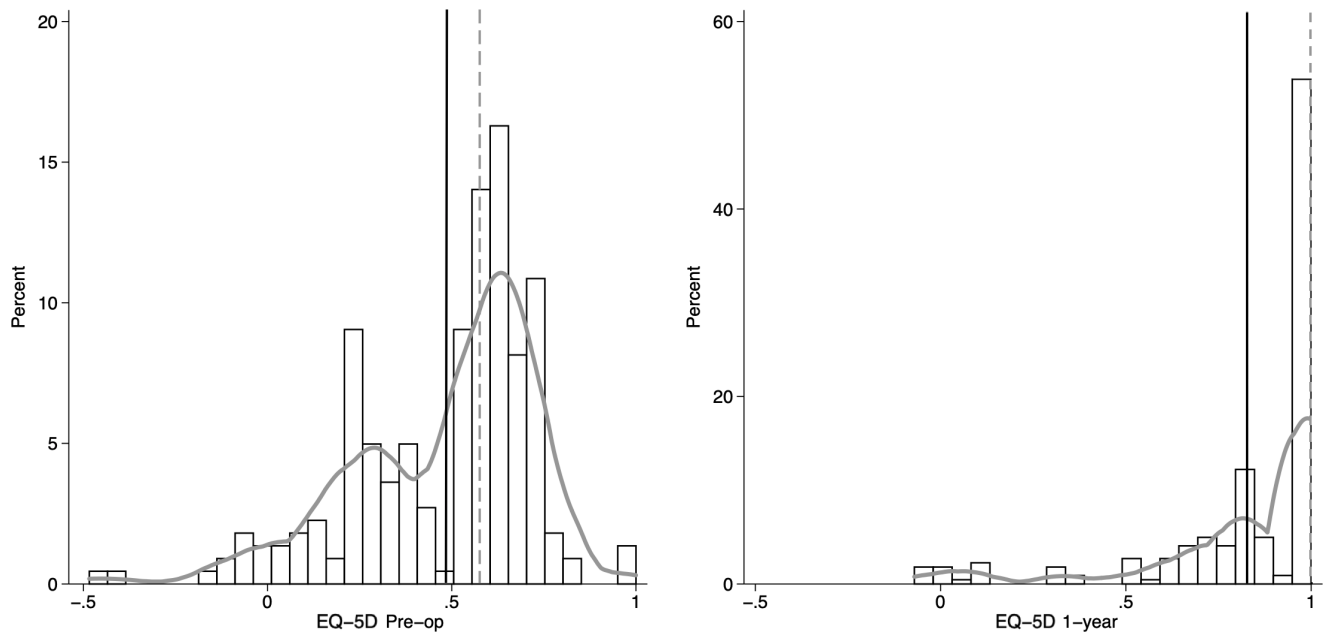


Fig. 2

Histograms with kernel (Epanechnikov) density plots demonstrating distribution of EuroQol-5D (EQ-5D) index scores preoperatively and at one-year follow-up. Solid vertical lines represent mean values, dashed vertical lines represent the median.

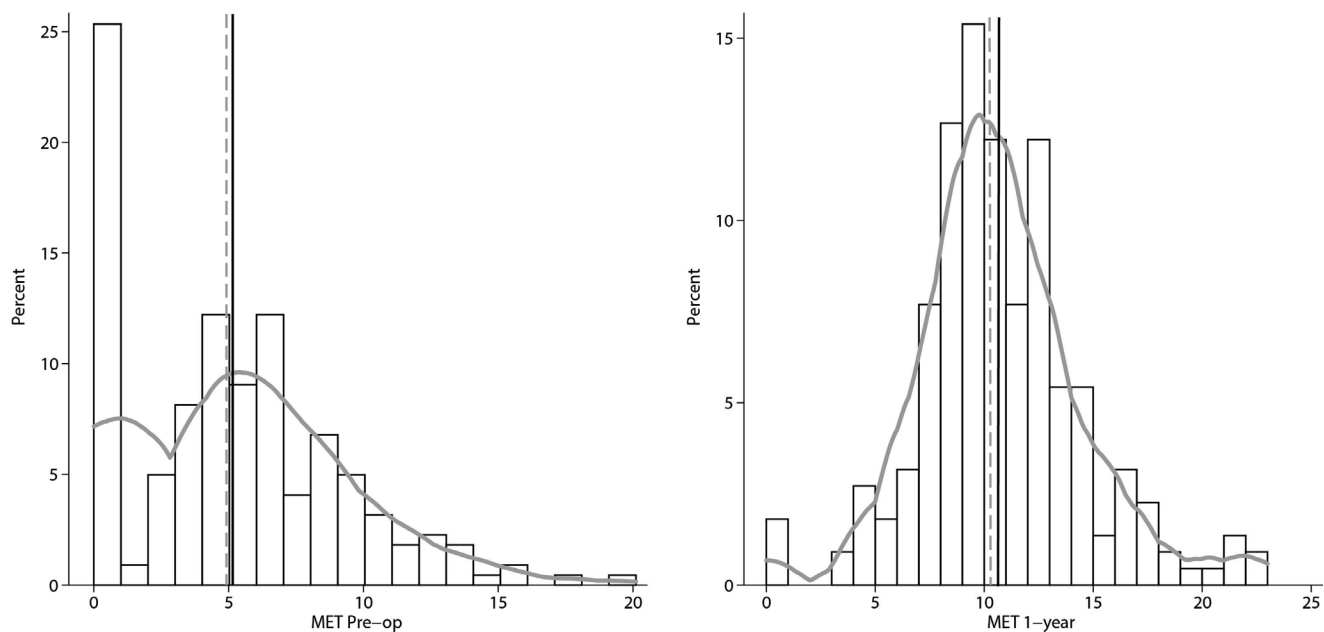


Fig. 3

Histograms with kernel (Epanechnikov) density plots demonstrating distribution of metabolic equivalent of task (MET) scores preoperatively and at one-year follow-up. Solid vertical lines represent mean values, dashed vertical lines represent the median.

preoperatively ( $r_s = 0.46$ ,  $p < 0.001$ , Spearman's rank correlation) and at one-year follow-up ( $r_s = 0.30$ ,  $p < 0.001$ , Spearman's rank correlation).

**Improvement in score and responsiveness.** All three metrics demonstrated excellent responsiveness with effect sizes as determined by the SRMs of  $> 1$  (Table III). The fractional polynomial predictive plots in Figure 4 demonstrate a strong negative relationship between preoperative OHS

and improvement in score for both EQ-5D (Figure 4a) and OHS (Figure 4b). For the MET score, this relationship is far less clear, with an initial decrease in improvement seen in patients who have lower preoperative OHS, while in patients with higher preoperative OHS the MET is progressively more responsive (Figure 4c).

**48/48 OHS.** A total of 92 postoperative patients scored 48/48 on the OHS following surgery. The histograms in

**Table II.** Distribution, ceiling and floor effects, and responsiveness of the three metrics before and after surgery.

Measure	n*	Preoperative						1-yr postoperative						p-value†	Mean difference (SD)	SRM
		Mean (SD)	Median (IQR)	Ceiling effect, %	Floor effect, %	Skew	Kurtosis	Mean (SD)	Median (IQR)	Ceiling effect, %	Floor effect, %	Skew	Kurtosis			
OHS	221	25.6 (8.3)	26 (20 to 32)	0.5	0.0	-0.14	2.57	44.7 (6.2)	47 (44 to 48)	41.6	0.0	-3.12	13.07	< 0.001	19.1 (9.5)	2.01
EQ-5D	221	0.48 (0.25)	0.57 (0.3 to 0.6)	1.4	0.0	-0.99	4.03	0.83 (0.26)	1 (0.8 to 1)	53.8	0.0	-1.96	6.23	< 0.001	0.34 (0.32)	1.06
MET	221	5.1 (4.1)	4.9 (0 to 7.5)	0.0	25.3	0.59	3.18	10.7 (3.8)	10.3 (8.6 to 12.5)	0.9	1.8	0.40	4.46	< 0.001	5.6 (4.8)	1.17

Ceiling and floor effects calculated as the percentage of patients scoring the maximum and minimum possible scores, respectively. Skewness and kurtosis values numerically represent the distribution of scores. A normal distribution has a skew of 0 and a kurtosis of 3. Standardized response mean calculated as the mean difference in scores divided by the standard deviation of that difference.

\*Indicates number of included patients.

†Paired *t*-test comparing preoperative and one-year postoperative scores.

EQ-5D, EuroQol five-dimension questionnaire; IQR, interquartile range; MET, metabolic equivalent of task; OHS, Oxford Hip Score; SD, standard deviation; SRM, standardized response mean.

**Table III.** Outcome scores by category for 48/48 Oxford Hip Score sub-cohort.

Variable	n*	MET score		EQ-5D	
		1 yr mean (SD, range)	p-value	1 yr median (IQR)	p-value†
<b>Age (yrs)</b>			0.008		0.371
< 60	44	12.7 (4.7, 0 to 23)		1 (1 to 1)	
> 60	48	10.6 (2.4, 6.3 to 17.1)		1 (1 to 1)	
<b>Sex</b>			0.024		0.280
Male	46	12.5 (4.5, 0 to 23)		1 (1 to 1)	
Female	46	10.8 (2.8, 6.3 to 19.9)		1 (1 to 1)	
<b>Preop MET</b>			0.040		0.248
< 6	54	11.0 (3.3, 0 to 23)		1 (1 to 1)	
> 6	38	12.6 (4.2, 6.2 to 23)		1 (1 to 1)	
<b>BMI (kg/m<sup>2</sup>)</b>			0.061		0.122
< 25	32	12.2 (3.5, 7.2 to 23)		1 (1 to 1)	
> 25	32	10.8 (2.5, 6.3 to 16)		1 (0.9 to 1)	

Except for p-values marked with the † symbol (Mann-Whitney U test), all p-values in this table were calculated using the independent-samples *t*-test.

\*Number of patients.

†Mann-Whitney U test.

EQ-5D, EuroQol five-dimension questionnaire; IQR, interquartile range; MET, metabolic equivalent of task; SD, standard deviation.

Figure 5 demonstrate the distribution of EQ-5D in this group, which has a strong negative skew, with the majority of patients scoring the maximal score of 1 (Figure 5a). The MET on the other hand exhibits a near normal distribution of scores, despite all patients scoring the same on the OHS (Figure 5b). When subdivided into groups, patients aged under 60 years scored significantly higher on the MET than patients over 60 years of age (mean 12.7 (SD 4.7) vs 10.6 (SD 2.4),  $p = 0.008$ , independent-samples *t*-test), as did male patients (mean 12.5 (SD 4.5) vs 10.8 (SD 2.8),  $p = 0.024$ , independent-samples *t*-test) and patients with higher activity levels on their preoperative MET scores (mean 12.6 (SD 4.2) vs 11.0 (SD 3.3),  $p = 0.040$ , independent-samples *t*-test) (Figure 6a). No significant differences were found comparing patients by BMI using the MET or comparing any of the groups using the EQ-5D (Figure 6b, Table III).

## Discussion

This retrospective study set out to determine whether the MET score could capture differences in function that were not detectable by the OHS or EQ-5D in an active hip arthroplasty population. This question was answered: the MET score does deliver a symmetrical metric with a normal distribution in a postoperative population, capturing differences in activity levels that were not detectable using the OHS. This question is relevant for health economists, policy makers, and those designing clinical trials. Health benefits that can be captured simply, without the need for expensive equipment or licences, should help drive commissioning choices. By demonstrating that patients can improve past the OHS maximum score, we have revealed an opportunity that was otherwise denied: using this metric, surgeons who are currently penalized for failing to deliver adequate health gains may now be

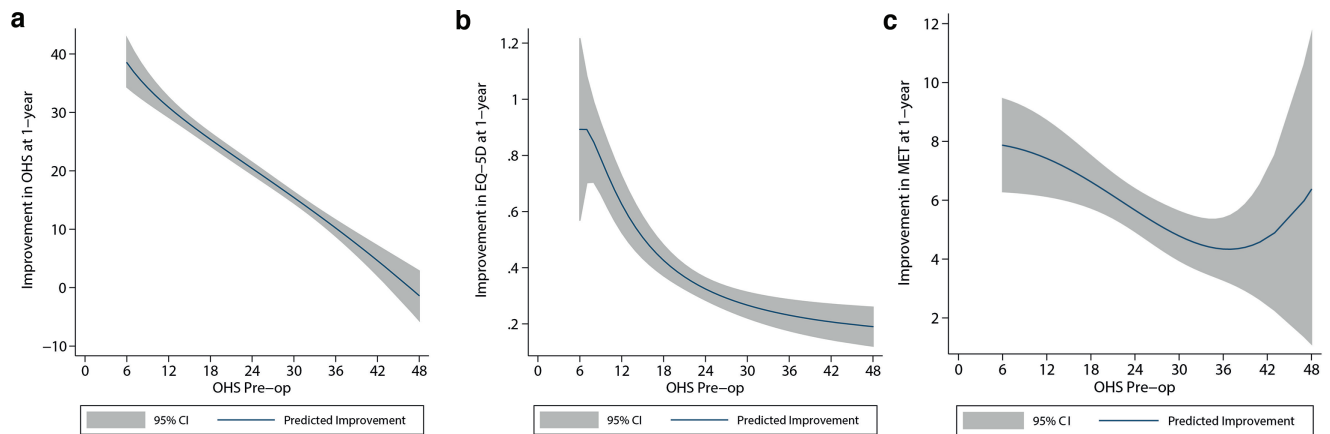


Fig. 4

Fractional polynomial regression plots demonstrating predicted improvement in score at one-year follow-up for a given preoperative Oxford Hip Score (OHS) for a) OHS, b) EuroQol five-dimension questionnaire (EQ-5D), and c) metabolic equivalent of task (MET). Fractional polynomial fit line with 95% confidence intervals (CIs) demonstrated in grey.

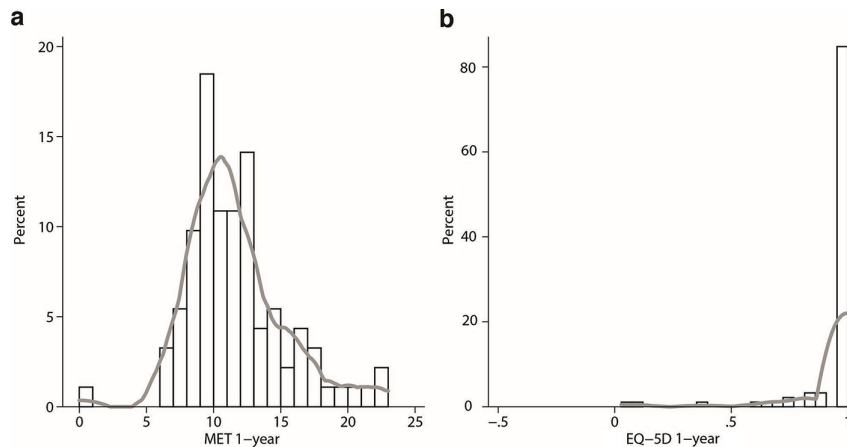


Fig. 5

Histograms with kernel (Epanechnikov) density plots demonstrating distribution of a) metabolic equivalent of task score (MET) and b) EuroQol five-dimension questionnaire (EQ-5D), at one-year follow-up for the subgroup of patients who all scored 48/48 on the Oxford Hip Score ( $n = 92$ ).

able to justify their offering of arthroplasty in younger and more active patients. By restricting health gains to the OHS, commissioners may unfairly restrict access to arthroplasty surgery, or unfairly penalize hospitals for not achieving satisfactory results if these decisions are based solely on health gains as measured by the OHS.

Although there is more work to be done in this area, the aspects of validity measured in the present study support its use as a metric for the outcome of hip arthroplasty surgery. The MET demonstrated evidence of concurrent validity with weak-to-moderate correlations found with both the OHS and EQ-5D. Naal et al<sup>17</sup> used a similar approach, establishing weak-to-moderate correlations with three different physical activity scores and the OHS. One potential limitation was that the present study did not validate the MET against another validated physical activity metric or objective physical activity measures such

as a pedometer or exercise log. However, the authors note that the face validity of using MET values has already been well established by other similar MET-based scores.<sup>18,19</sup> Although not validated specifically for use in arthroplasty, the International Physical Activity Questionnaire (IPAQ) score is a MET-based score, shown to be valid and reliable for use in the general population measuring activity levels.<sup>19</sup> It differs from the MET, being a better measure of cardiorespiratory fitness, whereas the MET is personalized to patients' sporting aspirations. The major advantage of the MET is that no matter what activities patients choose, the scores are comparable and relevant to their joint disease. Furthermore, the numeric MET values assigned by the University of Arizona are objective, being based upon oxygen consumption.<sup>18,19</sup> Therefore, the authors considered concurrent validity with other hip-specific and generic PROMs, alongside responsiveness, encouraging

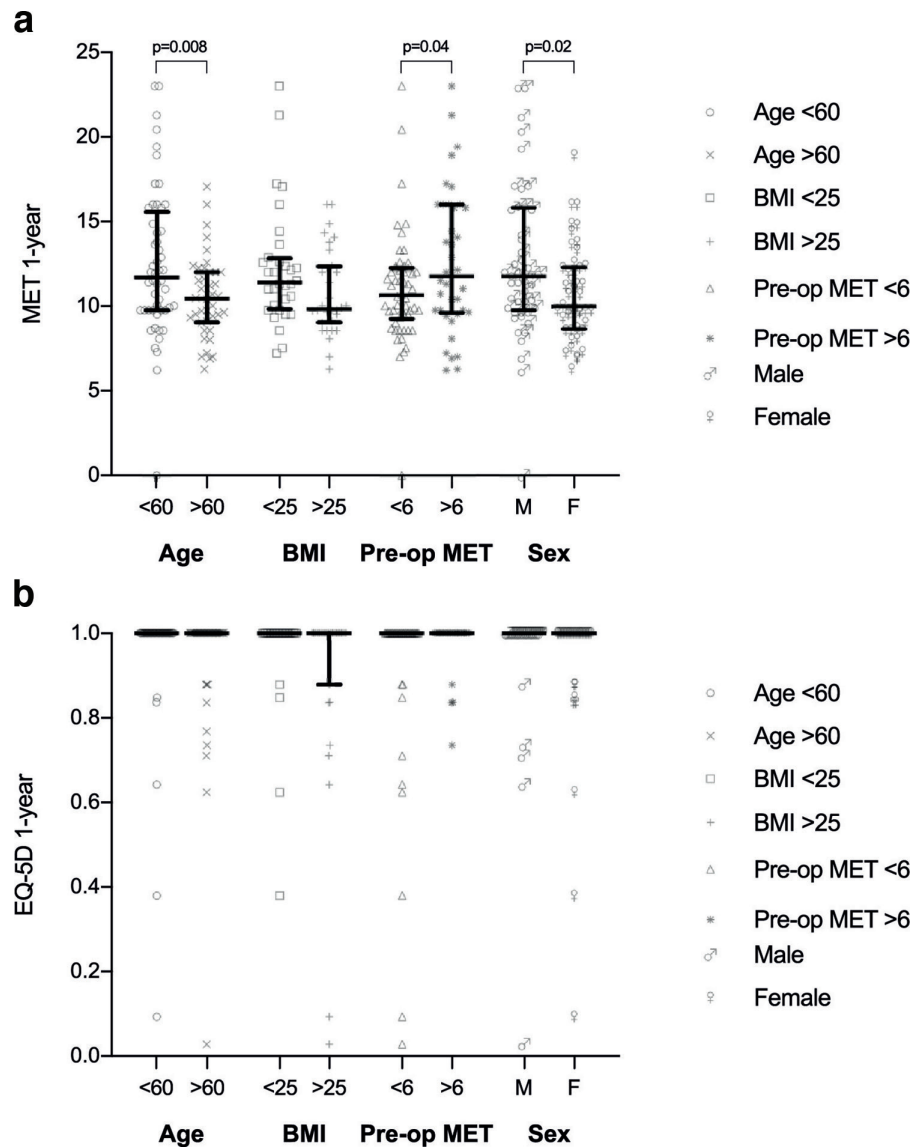


Fig. 6

Column scatter for the subgroup of 92 patients scoring 48/48 on the Oxford Hip Score, compared by age group, BMI, preoperative metabolic equivalent of task (MET) score, and sex using: a) MET scores; and b) EuroQol five-dimension questionnaire (EQ-5D) scores. The solid horizontal line represents the median and the whiskers represent the interquartile range. Statistically significant p-values have been indicated.

validation data for using the score in this cohort, however further work in this area would be an interesting avenue for future research.

Responsiveness is considered another aspect of construct validity.<sup>29</sup> The greater the responsiveness, the more accurate a metric is in detecting change when it has occurred. The MET had a SRM of 1.17, which indicates a large effect size or an excellent response to change over time.<sup>8</sup> The calculated SRMs for EQ-5D and OHS in this cohort were found to be similar to previously published literature, further validating our findings.<sup>30</sup>

Unlike the OHS and EQ-5D, the postoperative MET had a normal distribution and exhibited no ceiling effect. Substantial postoperative ceiling effects were found for the OHS (41.6%) and EQ-5D (53.8%). In general,

ceiling effects or floor effects are considered problematic when 15% or more of the cohort score the best or worst scores.<sup>7,10</sup> By having large numbers of patients scoring the best or worst scores, the metric is rendered insensitive to detecting differences at the extremes of the scale.<sup>7,10</sup> Other studies have demonstrated strong ceiling effects in the OHS of 19.9%,<sup>6</sup> and even more pronounced ceiling effects for the EQ-5D of 39.8%.<sup>30</sup> While the pattern of these findings support our results, our population demonstrated a much higher percentage ceiling effect for both metrics. This may be related to the studied population which included a younger, more active cohort than that used in other studies. While other scores have been developed with the aim of reducing the impact of ceiling effect, unfortunately problematic ceiling effects may still



exist. In a recent study, the FJS was reported to have a ceiling effect of 31.9%, similar to those reported for more conventional PROMs.<sup>15</sup> In addition, the FJS has reported a substantial floor effect of 22.4%, suggesting that there may be problems discriminating at both ends of the score.<sup>31</sup>

While the MET showed no postoperative ceiling effect, it did show a preoperative floor effect, similarly to the FJS. This is not surprising given that the formulation of the question specifies the selection of tasks that have been negatively affected by the respondent's hip pain. A similar preoperative floor effect has been observed in validation studies looking at other physical activity-based outcome measures such as the Tegner score.<sup>17</sup> When using MET solely as an assessment of postoperative outcome rather than of preoperative disease state, this floor effect is unimportant. If it were to be used for the former, the question may have to be re-formulated.

Both the OHS and EQ-5D demonstrated very little predicted improvement towards the upper end of the preoperative OHS scale. The MET on the other hand shows continued predicted improvements, with a 6 MET improvement predicted for patients who score 48/48 on the OHS. A large registry study by Price et al<sup>23</sup> demonstrated a similar effect using the OHS, with the likelihood of seeing a meaningful clinical improvement decreasing with higher preoperative scores. The authors conclude that at a preoperative score of 40 or above, there was a 0% chance of meaningful improvement, suggesting this as a threshold for referral.<sup>23</sup> The present study suggests that even though these higher-scoring preoperative patients do not show improvement using the OHS, they do show considerable improvement using the MET. Setting a referral threshold at 40 may restrict access to high-functioning patients who may want to return to a preferred sporting activity.

While it is certainly important to use conventional PROMs to record health gains, the assumption that no further benefit can be achieved past the maximal score may mean that these overall health gains are under-represented. In doing so, one may unfairly restrict access to our highly effective surgical interventions for higher-functioning patients who are unable to perform their desired sporting activity. Without an additional activity metric, the considerable improvement in quality of life delivered by returning them to their preferred sporting activity may be reported as a failure, since the improvement in function captured by change in OHS may be smaller than average.

The subgroup analysis further emphasizes the point that the patients who score 48/48 are not necessarily performing at a similar level to one another. Despite identical OHS scores, patients > 60 years old had a mean MET of 10.6 METs compared to the 12.7 METs scored by the under 60s. A similar effect was noted for the male sex and those with higher preoperative MET scores. To put those scores into perspective, an activity such as Nordic walking at a fast pace scores 9.5 METs.<sup>18</sup> A fast run at 9 mph scores 12.8 METs,<sup>18</sup> so a difference of 2 to 3 METs translates into

the difference between patients performing a fast walk or a fast run. Clinically this would likely be a noticeable benefit. Other studies have shown the effect of age, sex, and preoperative activity levels on postoperative physical activity. Williams et al,<sup>24</sup> in a study of 736 primary joint arthroplasty operations, found male sex, younger age, preoperative UCLA scores, and lower BMI to be overall predictors for achieving higher postoperative activity levels. The authors report that males are nearly five times more likely to achieve a UCLA activity score > 7 post-hip arthroplasty when compared to females (odds ratio 4.84, 95% confidence interval 2.93 to 7.99).<sup>24</sup> These findings have been corroborated by a number of other studies, concurring with the findings of the present study.<sup>25,32,33</sup>

There are a number of limitations to this study. First, a large proportion of patients (61%) did not have preoperative or one-year postoperative scores, and the included patients were younger than those with missing data. It is possible that this younger cohort who completed the online questionnaire were more physically active and motivated than those who did not respond. Furthermore, our studied population was considerably younger than the national average for hip arthroplasty. While the authors believe this young population to be ideal for investigating the MET, it is worth noting that our findings may not be generalizable to the wider population of hip arthroplasty patients. Second, the MET does not factor in frequency of the activity, only intensity, so it cannot be used as a metric of fitness. Additionally, a high MET value may not correlate with impact on the hip joint, nor on the number of hip cycles. For instance, canoeing with vigorous effort scores a MET of 12.5.<sup>18</sup> This scores similarly to running at 9 mph (12.8 METs),<sup>18</sup> however running has greater impact on the hip joint and may not be attempted following hip arthroplasty in an effort to protect the longevity of the implant. Although our score did not take this into account, patients were asked to pick activities that were of importance to them and that their joint trouble affected, thus directing them to choose activities specific to the hip. Finally, as data in this study were retrospectively analyzed, there remains a risk of selection bias.

In conclusion, this study demonstrates that a simple, patient-centred activity metric (MET) can pick up important health gains in return to higher-level sporting activity, which are missed by the OHS in a younger, active population. The MET showed evidence of construct validity, good responsiveness to change, and no post-operative ceiling effect, with health gains not limited by preoperative OHS. A patient-centred physical activity metric may have a useful role in addition to conventional function-based PROMs scores where the functional outcome of hip arthroplasty is relevant.

### Twitter

Follow T. C. Edwards @edwards\_tomc  
Follow A. Garner @DrAmyGarner  
Follow K. Logishetty @klogishetty  
Follow J. P. Cobb @orthorobodoc

## References

- Mei XY, Gong YJ, Safir O, Gross A, Kuzyk P. Long-term outcomes of total hip arthroplasty in patients younger than 55 years: a systematic review of the contemporary literature. *Can J Surg*. 2019;62(4):249–258.
- Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br*. 1996;78-B(2):185–190.
- Murray DW, Fitzpatrick R, Rogers K, et al. The use of the Oxford hip and knee scores. *J Bone Joint Surg Br*. 2007;89-B(8):1010–1014.
- Harris K, Dawson J, Gibbons E, et al. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Relat Outcome Meas*. 2016;7:101–108.
- Dakin H, Eibich P, Beard D, Gray A, Price A. The use of patient-reported outcome measures to guide referral for hip and knee arthroplasty. *Bone Joint J*. 2020;102-B(7):950–958.
- Paulsen A, Odgaard A, Overgaard S. Translation, cross-cultural adaptation and validation of the Danish version of the Oxford hip score: Assessed against generic and disease-specific questionnaires. *Bone Joint Res*. 2012;1(9):225–233.
- Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. *BMJ Open*. 2015;5(7):e007765.
- Dawson J, Beard DJ, McKibbin H, Harris K, Jenkinson C, Price AJ. Development of a patient-reported outcome measure of activity and participation (the OKS-APQ) to supplement the Oxford knee score. *Bone Joint J*. 2014;96-B(3):332–338.
- Scott CEH, Bugler KE, Clement ND, MacDonald D, Howie CR, Biant LC. Patient expectations of arthroplasty of the hip and knee. *J Bone Joint Surg Br*. 2012;94-B(7):974–981.
- Gulledge CM, Lizzio VA, Smith DG, Guo E, Makhni EC. What are the floor and ceiling effects of patient-reported outcomes measurement information system computer adaptive test domains in orthopaedic patients? A systematic review. *Arthroscopy*. 2020;36(3):901–912.
- Edwards TC, Logishetty K, Cobb JP. Letter to the Editor on “Patient-reported outcomes following total hip arthroplasty: a multicenter comparison based on surgical approaches.” *J Arthroplasty*. 2020;35(9):2686–2687.
- Briggs T. A national review of adult elective orthopaedic services in England: Getting It Right First Time. British Orthopaedic Association. 2015. <https://gettingitrightfirsttime.co.uk/wp-content/uploads/2018/07/GIRFT-National-Report-Mar15-Web.pdf> (date last accessed 5 April 2022).
- No authors listed. 2022/23 National Tariff Payment System – a consultation notice. 2022. <https://www.england.nhs.uk/wp-content/uploads/2021/12/22-23-National-Tariff-Payment-System-a-consultation-notice-v2.pdf> (date last accessed 5 May 2022).
- Clement ND, Gaston P, Bell A, et al. Robotic arm-assisted versus manual total hip arthroplasty. *Bone Joint Res*. 2021;10(1):22–30.
- Puliero B, Blakeney WG, Beaulieu Y, Vendittoli PA. Joint perception after total hip arthroplasty and the forgotten joint. *J Arthroplasty*. 2019;34(1):65–70.
- Larsson A, Rolsson O, Kärrholm J. Evaluation of Forgotten Joint Score in total hip arthroplasty with Oxford Hip Score as reference standard. *Acta Orthop*. 2019;90(3):253–257.
- Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for patients undergoing total joint arthroplasty? *Clin Orthop Relat Res*. 2009;467(4):958–965.
- Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc*. 2011;43(8):1575–1581.
- Hagströmer M, Oja P, Sjöström M. The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutr*. 2006;9(6):755–762.
- Kujala UM, Kaprio J, Sarna S, Koskenvuo M. Relationship of leisure-time physical activity and mortality: the Finnish twin cohort. *JAMA*. 1998;279(6):440–444.
- Amstutz HC, Le Duff MJ. Effects of physical activity on long-term survivorship after metal-on-metal hip resurfacing arthroplasty: is it safe to return to sports? *Bone Joint J*. 2019;101-B(10):1186–1191.
- Le Duff MJ, Amstutz HC. Sporting activity after hip resurfacing: changes over time. *Orthop Clin North Am*. 2011;42(2):161–167.
- Price AJ, Kang S, Cook JA, et al. The use of patient-reported outcome measures to guide referral for hip and knee arthroplasty. *Bone Joint J*. 2020;102-B(7):941–949.
- Williams DH, Greidanus NV, Masri BA, Duncan CP, Garbus DS. Predictors of participation in sports after hip and knee arthroplasty. *Clin Orthop Relat Res*. 2012;470(2):555–561.
- Lübbecke A, Zimmermann-Sloutskis D, Stern R, et al. Physical activity before and after primary total hip arthroplasty: a registry-based study. *Arthritis Care Res (Hoboken)*. 2014;66(2):277–284.
- Vähä-Yppä H, Vasankari T, Husu P, et al. Validation of cut-points for evaluating the intensity of physical activity with accelerometry-based mean amplitude deviation (MAD). *PLoS One*. 2015;10(8):e0134813.
- Jetté M, Sidney K, Blümchen G. Metabolic equivalents (METs) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clin Cardiol*. 1990;13(8):555–565.
- Hoffman JIE. Chapter 6 - Normal Distribution. In: Hoffman JIE, ed. *Biostatistics for Medical and Biomedical Practitioners*. Cambridge, Massachusetts: Academic Press, 2015: 101–119.
- Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102–109.
- Ostendorf M, van Stel HF, Buskens E, et al. Patient-reported outcome in total hip replacement. A comparison of five instruments of health status. *J Bone Joint Surg Br*. 2004;86-B(6):801–808.
- Hamilton DF, Loth FL, Giesinger JM, et al. Validation of the English language Forgotten Joint Score-12 as an outcome measure for total hip and knee arthroplasty in a British population. *Bone Joint J*. 2017;99-B(2):218–224.
- Dahm DL, Barnes SA, Harrington JR, Sayeed SA, Berry DJ. Patient-reported activity level after total knee arthroplasty. *J Arthroplasty*. 2008;23(3):401–407.
- Wylde V, Blom A, Dieppe P, Hewlett S, Learmonth I. Return to sport after joint replacement. *J Bone Joint Surg Br*. 2008;90-B(7):920–923.

### Author information:

- T. C. Edwards, BSc(Hons), MBBS, MRCS, Clinical Research Fellow
- B. Guest, BSc, Physician Associate
- A. Garner, BA(Hons) (Oxon), BMBCh (Oxon), MRCS, Clinical Research Fellow
- K. Logishetty, MSc, MRCS, NIHR Clinical Lecturer
- A. D. Liddle, DPhil, FRCS (T&O), Senior Clinical Lecturer
- J. P. Cobb, MCh, FRCS (T&O), Chair in Orthopaedics MSk Lab, Imperial College London, Sir Michael Uren Biomedical Engineering Research Hub, London, UK.

### Author contributions:

- T. C. Edwards: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft.
- B. Guest: Data curation, Formal analysis, Writing – review & editing.
- A. Garner: Data curation, Formal analysis, Writing – review & editing.
- K. Logishetty: Data curation, Formal analysis, Writing – review & editing.
- A. D. Liddle: Conceptualization, Methodology, Formal analysis, Supervision, Writing – review & editing.
- J. P. Cobb: Conceptualization, Methodology, Formal analysis, Supervision, Writing – review & editing.

### Funding statement:

- The authors disclose receipt of the following financial or material support for the research, authorship, and/or publication of this article: an institutional research support grant from the Sir Michael Uren Foundation (as reported by J. P. Cobb). Infrastructure support was provided by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC).

### Acknowledgements:

- The authors would like to acknowledge the support of the editors and reviewers of this manuscript, for their invaluable contributions.

### Ethical review statement:

- Ethical approval was granted for data collected and used as part of this study (REC Reference: 14/SC/1243, IRAS ID: 136430).

### Open access funding:

- The authors report that they received open access funding for their manuscript from the Imperial College London open access fund.

© 2022 Author(s) et al. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (CC BY-NC-ND 4.0) licence, which permits the copying and redistribution of the work only, and provided the original author and source are credited. See <https://creativecommons.org/licenses/by-nc-nd/4.0/>