



Machine learning algorithms do not outperform preoperative thresholds in predicting clinically meaningful improvements after total knee arthroplasty

Siyuan Zhang¹ · Bernard Puang Huh Lau² · Yau Hong Ng² · Xinyu Wang² · Weiliang Chua²

Received: 25 April 2021 / Accepted: 12 June 2021 / Published online: 10 July 2021
© European Society of Sports Traumatology, Knee Surgery, Arthroscopy (ESSKA) 2021

Abstract

Purpose Patient-reported outcome measures (PROMs) are important measures of success after total knee arthroplasty (TKA) and being able to predict their improvements could enhance preoperative decision-making. Our study aims to compare the predictive performance of machine learning (ML) algorithms and preoperative PROM thresholds in predicting minimal clinically important difference (MCID) attainment at 2 years after TKA.

Methods Prospectively collected data of 2840 primary TKA performed between 2008 and 2018 was extracted from our joint replacement registry and split into a training set (80%) and test set (20%). Using the training set, ML algorithms were developed using patient demographics, comorbidities and preoperative PROMs, whereas the optimal preoperative threshold was determined using ROC analysis. Both methods were used to predict MCID attainment for the SF-36 PCS, MCS and WOMAC at 2 years postoperatively, with predictive performance evaluated on the independent test set.

Results ML algorithms and preoperative PROM models performed similarly in predicting MCID for the SF-36 PCS (AUC: 0.77 vs 0.74), MCS (AUC: 0.95 vs 0.95) and WOMAC (AUC: 0.89 vs 0.88). For each outcome, the most important predictor of MCID attainment was the patient's preoperative PROM score. ROC analysis also identified optimal preoperative threshold values of 33.6, 54.1 and 72.7 for the SF-36 PCS, MCS and WOMAC, respectively.

Conclusion ML algorithms did not perform significantly better than preoperative PROM thresholds in predicting MCID attainment after TKA. Future research should routinely compare the predictive ability of ML algorithms with existing methods and determine the type of clinical problems which may benefit the most from it.

Level of evidence II.

Keywords Total knee arthroplasty · Machine learning · Artificial intelligence · Patient reported outcome measures · MCID

Introduction

Patient-reported outcome measures (PROMs) are used to measure improvements in pain, function, and quality of life after total knee arthroplasty (TKA) [10]. Given the trend towards more patient-centric and value-based care, PROMs, especially when interpreted in the context of the minimal

clinically important difference (MCID), are increasingly recognized as key measures of success after TKA [2]. Being able to predict a patient's likelihood of MCID attainment could facilitate more personalized preoperative counselling and enable more informed decision-making [6].

Using receiver operating characteristic (ROC) analysis, preoperative PROMs have been used to predict MCID attainment, since patients with poorer preoperative PROMs have been observed to experience greater improvement [2, 3]. Berliner et al. found that preoperative PROM thresholds (KOOS: 58, SF-12 PCS: 34) could predict MCID attainment for TKA reasonably well, achieving an AUC of 0.76 for KOOS and 0.65 for SF-12 PCS [2].

In contrast, machine learning (ML) is a relatively newer method of data analysis which uses computer algorithms capable of learning and improving through experience [22].

✉ Weiliang Chua
wei_liang_chua@nuhs.edu.sg

¹ Yong Loo Lin School of Medicine, National University of Singapore, 10 Medical Drive, Singapore 117597, Singapore

² Department of Orthopaedic Surgery, National University Hospital, Level 11, NUHS Tower Block, 1E Kent Ridge Road, Singapore 119228, Singapore

In simple terms, ML algorithms work by recognizing patterns within real-world data and then using these insights to generate predictions without being explicitly programmed [11]. Recent studies have also demonstrated the feasibility of using ML algorithms to predict postoperative MCID attainment after TKA, achieving an AUC: 0.60–0.89 for outcomes such as Knee injury and Osteoarthritis Outcome Score (KOOS), Short Form-36 (SF-36) and Oxford Knee Score (OKS) [9, 12, 14].

Although both methods are fairly accurate in predicting MCID attainment, there has been no prior study directly comparing their performance on the same dataset. As such, it is unclear whether ML algorithms offer a significant advantage compared to the use of preoperative PROMs in predicting MCID attainment.

Given the growing interest to identify accurate and accessible prediction tools that can be used to improve preoperative decision-making, this study aims to evaluate and compare the performance of ML algorithms and preoperative PROM thresholds in predicting MCID attainment for the SF-36 PCS, MCS and WOMAC at 2 years after TKA. It is hypothesized that ML algorithms will outperform preoperative PROM thresholds in predicting MCID attainment after TKA.

Materials and methods

Data

Using prospectively collected data from a single institution's joint replacement registry, 3537 adult patients who had undergone elective primary TKA between 2008 and 2018 were identified, of which 3021 (85.4%) completed their postoperative 2-year follow-up. Only patients with unilateral TKA due to primary osteoarthritis were included and patients with simultaneous bilateral TKA ($n=138$), inflammatory arthritis ($n=36$), post-traumatic arthritis ($n=3$), malignancy ($n=2$) and avascular necrosis ($n=2$) were excluded. Data were subsequently collected for the remaining 2840 patients. Ethics approval from the institutional review board (DSRB: 2020/00613) was obtained prior to study initiation, with a waiver of informed consent due to the use of deidentified data.

Outcomes

The two PROMs used in this study are the Short Form-36 (SF-36) and Western Ontario and McMaster University Osteoarthritis Index (WOMAC) at 2 years postoperatively. The SF-36 is a generic health questionnaire that measures an individual's health-related quality of life, with its 8 domains aggregated into the physical component summary (PCS) and

mental component summary (MCS) [15, 25]. In comparison, WOMAC is a disease-specific questionnaire for lower limb osteoarthritis and has 3 main dimensions of pain, stiffness and function [26, 27].

Published anchor-based MCID values were used as they take into account the patient's perspective and are more likely to represent subjective meaningful change [16, 17]. MCID values of 10.0 points for SF-36 PCS, 5.0 for SF-36 MCS and 15.0 for WOMAC were adopted based on work by Escobar et al. and Quintana et al. [8, 21].

Statistical analysis

Statistical analysis and predictive modelling were performed using Python 3.7 (Python Software Foundation, Wilmington, DE) and R software version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria; 2019). First, the 2840 TKAs were randomly split into a training set (80%) and test set (20%). Data in the training set was used to train and develop the various predictive models while evaluation of model performance was performed using the independent test set to assess its ability to generalize to unseen data.

Machine learning algorithms

For the ML approach, four of the most popular supervised ML algorithms: (1) random forest (RF), (2) extreme gradient boosting (XGB), (3) support vector machines (SVM) and (4) logistic regression with L1-regularization (LASSO) were used.

Input variables for ML models include patient demographics, comorbidities and preoperative PROMs (Table 1). Input variables were limited to preoperative data as we believe the greatest utility of such prediction models would be to aid patients and surgeons in preoperative decision-making. Missing input variables—BMI ($n=9$) and preoperative pain score ($n=5$) were imputed with median values in the training or test set correspondingly. Upsampling (random oversampling) was implemented during model training due to class imbalance which could potentially affect the predictive performance of some algorithms [1]. All ML models were trained using fivefold cross-validation in the training set to optimize their hyperparameters before the final evaluation their predictive performance on the independent test set.

Preoperative PROM thresholds

A non-parametric ROC analysis, similar to the method described by Berliner et al. was used to determine the preoperative PROM threshold that could best predict MCID attainment in the training set [2]. The optimal threshold value was determined using the Youden method, which corresponds to the point on the ROC curve that maximizes the

Table 1 Input variables for ML models

	SF-36 PCS	SF-36 MCS	WOMAC
Demographics	Age Sex: Male or Female Race: Chinese, Indian, Malay, Others BMI BMI (categorical): <18.5, 18.5 – 29.9, ≥30.0 Payment class: Private or Public		
Comorbidities	Diabetes Hypertension Ischemic Heart Disease Stroke Cancer Respiratory disease		
Preoperative Data	Preop Pain Score Surgeon Preop PCS Preop MCS	Preop PCS Preop MCS	Preop WOMAC

classifier’s combined sensitivity and specificity [28]. This optimal threshold value was then implemented in the unseen test set to evaluate its ability to predict MCID attainment.

Evaluation of predictive performance

The area under ROC curve (AUC) is a commonly used evaluation metric for predicting binary outcomes, measuring the model’s ability to discriminate between different classes. In general, an AUC of 0.7–0.8 indicates fair discriminative ability, 0.8–0.9 indicates good discriminative ability and 0.9–1.0 indicates excellent discriminative ability [24]. Other classification metrics include the F1 score, Brier score, Sensitivity and Specificity.

Results

Baseline patient characteristics of the 2840 TKAs are reported in Table 2. 70.7% (n = 2008) of patients were female. The mean age was 66.3 (SD: 8.2) years and the mean BMI was 28.0 (SD: 5.0).

Improvements in PROMs are summarized in Table 3. At 2 years postoperatively, mean improvements for SF-36 PCS, SF-36 MCS and WOMAC were +15.4 (SD: 9.2), +3.3 (SD: 8.8) and +26.5 (SD: 13.8) respectively. MCID attainment was 73.6% for SF-36 PCS, 38.5% for SF-36 MCS and 81.2% for WOMAC.

Predictive performance

The predictive performance of ML algorithms and preoperative PROM thresholds are reported in Table 4. All models were evaluated on the unseen test set. For brevity, only the top two performing ML models were presented. For all outcomes, ML models and preoperative PROM

thresholds performed similarly in terms of their discriminative ability (AUC). For the SF-36 PCS, ML models achieved a maximum AUC of 0.77 while the preoperative PROM threshold achieved an AUC of 0.74. For the SF-36 MCS, ML models achieved a maximum AUC of 0.95 while the preoperative PROM threshold achieved a similar AUC of 0.95. For the WOMAC, ML models achieved a maximum AUC of 0.89 while the preoperative PROM threshold achieved an AUC of 0.88.

Using ROC analysis on the training data, the identified optimal preoperative PROM thresholds are 33.6 for the PCS, 54.1 for the MCS and 72.7 for the WOMAC (Figs. 1, 2, and 3). These threshold values are preoperative cut-offs that can be used to predict MCID attainment for the same PROM. For example, using a preoperative threshold of 72.7 for the WOMAC, it can predict MCID attainment in the test set with 84.7% sensitivity and 71.2% specificity.

Table 2 Baseline patient characteristics

Sex	Female	
	2008 (70.7%)	
Age	Mean (SD)	Range
	66.3 (8.2)	41–88
BMI	Mean (SD)	Range
	28.0 (5.0)	13.6–57.6
Race	Chinese	
	2063 (72.6%)	
Payment class	Public	
	1988 (70.0%)	
Comorbidities		
Diabetes	716 (25.2%)	
Hypertension	1890 (66.5%)	
Ischemic heart disease	356 (12.5%)	
Stroke	93 (3.3%)	
Cancer	142 (5.0%)	
Respiratory disease	242 (8.5%)	

Table 3 PROM improvements and MCID attainment

PROM	Preop Mean (SD)	2-year Mean (SD)	Improvement Mean (SD)	MCID attainment (%)
Pain Score	6.3 (2.1)	–	–	–
SF-36 PCS	31.9 (6.9)	47.4 (7.8)	+ 15.4 (9.2)	2089 (73.6%)
SF-36 MCS	54.9 (8.1)	58.2 (4.5)	+ 3.3 (8.8)	1092 (38.5%)
WOMAC	63.1 (13.0)	89.7 (7.8)	+ 26.5 (13.8)	2305 (81.2%)

Table 4 Predictive performance of ML algorithms and preoperative PROM thresholds

	SF-36 PCS			SF-36 MCS			WOMAC		
	XGB	SVM	Threshold	XGB	SVM	Threshold	RF	LASSO	Threshold
AUC	0.77 (0.72–0.81)	0.76 (0.71–0.81)	0.74 (0.70–0.79)	0.95 (0.94–0.97)	0.95 (0.94–0.97)	0.95 (0.94–0.97)	0.89 (0.85–0.92)	0.89 (0.86–0.92)	0.88 (0.85–0.91)
Threshold	0.49	0.48	33.6	0.50	0.51	54.1	0.61	0.47	72.7
F1 Score	0.80	0.81	0.79	0.86	0.85	0.86	0.88	0.84	0.88
Brier Score	0.27	0.27	0.29	0.11	0.12	0.11	0.18	0.23	0.18
Sensitivity (%)	75.2	76.4	73.6	95.6	93.1	93.1	83.2	75.5	84.7
Specificity (%)	67.1	63.8	63.2	84.9	86.8	86.3	76.6	82.0	71.2

ML algorithms: *RF* random forest, *SVM* support vector machine, *XGB* extreme gradient boosting, *LASSO* logistic regression with L1-regularization; Threshold: preoperative PROM thresholds

Machine learning variable importance

To understand the variables which drive prediction in our ML models, variable importance was assessed using permutation importance, a model-agnostic method which has been shown to generate reliable insights correlating with clinical intuition [4]. The permutation importance scores of the top five variables for each PROM is summarized in Table 5. For each PROM, the single most important predictor of MCID attainment was consistently its corresponding preoperative score. In comparison, other variables such as patient age and operating surgeon played a relatively minor role, as evident by their minute permutation importance scores.

Discussion

The key finding of our study was that ML algorithms did not perform significantly better than preoperative PROM thresholds in predicting MCID attainment after TKA. Our study is the first to directly compare the predictive performance of ML algorithms with preoperative PROM thresholds in predicting MCID attainment after TKA. Our results have shown that simple preoperative PROM thresholds perform just as well as ML algorithms in predicting MCID attainment for SF-36 PCS, MCS and WOMAC after TKA. This is likely because the single most important predictor of MCID attainment after TKA is the patient's preoperative PROM, with other variables such as patient demographics and comorbidities shown to play a relatively minor role.

However, this is not a denunciation of ML. Rather than saying that ML is not useful, our results simply make the argument that ML algorithms may not perform significantly better than existing methods when applied to this clinical problem and trained using currently available data.

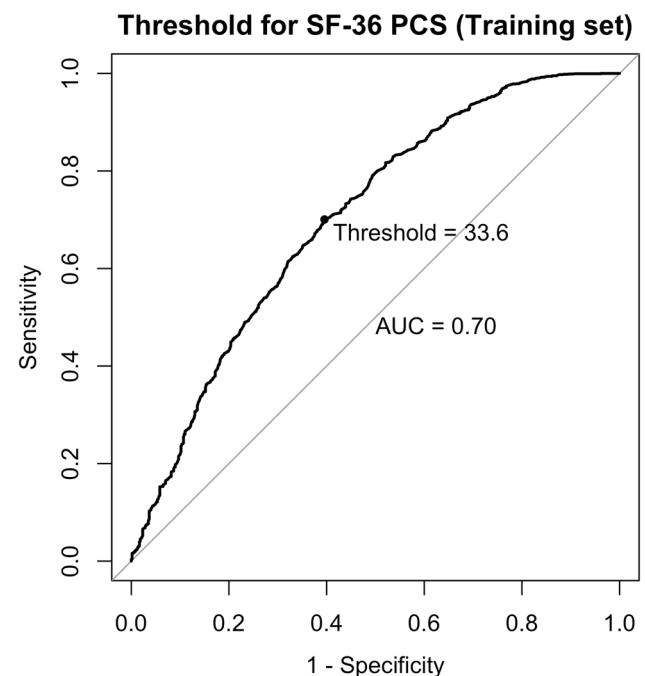


Fig. 1 Optimal preoperative threshold for SF-36 PCS in the training set

Furthermore, current registry data may lack the granularity required by ML algorithms to generate deeper insights beyond those offered by preoperative PROMs.

While artificial intelligence has the potential to transform healthcare, not all clinical problems may warrant the use of ML algorithms or even benefit significantly from it, especially if existing methods may work just as well [22]. However, it remains unclear what types of clinical problems and datasets may benefit the most from ML algorithms. Some studies also do not routinely compare their performance with conventional statistical techniques, making it difficult to critically evaluate the comparative advantage of ML algorithms. This issue has been highlighted by Christodoulou et al. in a

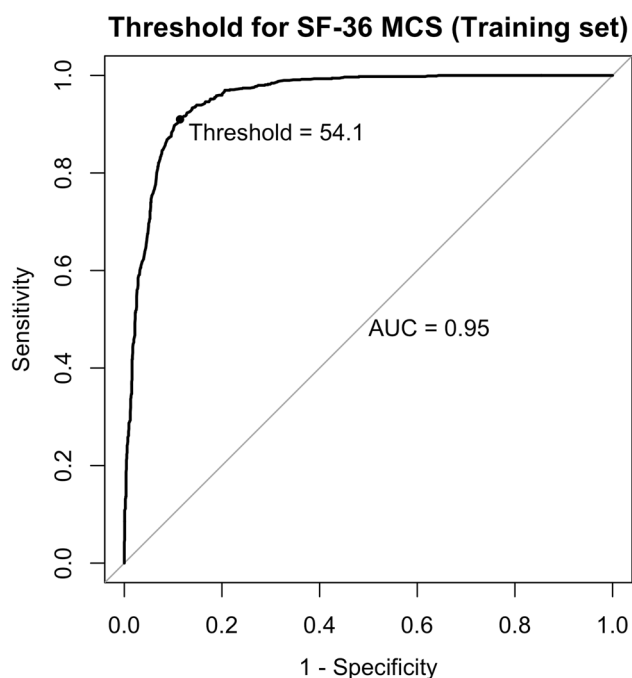


Fig. 2 Optimal preoperative threshold for SF-36 MCS in the training set

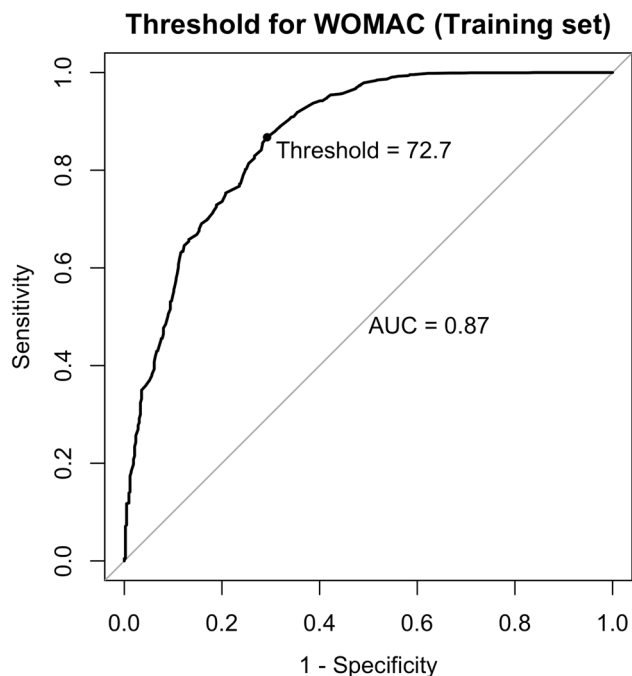


Fig. 3 Optimal preoperative threshold for WOMAC in the training set

systematic review of 71 studies, which found that ML algorithms did not outperform conventional logistic regression in predicting various clinical outcomes [5]. A recent study by Pua et al. also found that ML algorithms did not perform

better than logistic regression in predicting severe walking limitation after TKA [20]. Given that this is a vital topic that remains poorly understood, future research should routinely evaluate the comparative advantage of ML algorithms over existing methods and determine the types of clinical problems which may benefit the most from it.

In addition, this study has determined optimal preoperative PROM thresholds using ROC analysis—33.6 for the SF-36 PCS, 54.1 for the MCS and 72.7 for the WOMAC. These findings are consistent with prior work by Berliner et al., who reported a preoperative threshold of 34 points for the SF-12 PCS [2]. Moreover, this is the first study to report preoperative PROM thresholds for predicting MCID attainment in SF-36 MCS and WOMAC after TKA. These findings show that despite its simplicity, preoperative PROM thresholds are effective and remain a clinically viable tool in predicting postoperative functional improvements, with patients exceeding such thresholds less likely to attain a clinically meaningful change.

With increased ability to predict functional improvements after TKA, there is a greater need to define the utility and role of such prediction tools in clinical practice. Price et al. reported that a preoperative OKS threshold of 41 is predictive of MCID attainment and proposed using the threshold to guide specialist referral by primary care physicians [19]. While this is a step in the right direction towards better resource utilization and individualized patient care, such thresholds remain imperfect prediction tools with significant false positives and negatives. Furthermore, the MCID is derived from population estimates and using a point value to assess clinically meaningful change may risk misclassifying individuals with varying thresholds to perceive improvement [7]. As such, these prediction tools may be better suited for preoperative surgical counselling and shared decision-making—whereby patients can be individually counselled on their expected improvements so as to make more informed decisions regarding TKA [2]. Prior studies have also shown that modifying patient expectations prior to joint arthroplasty can improve postoperative satisfaction, thus highlighting the potential of such decision-making tools to encourage more realistic expectations regarding TKA [13, 18].

There are several limitations in this study. First, since data used in this study was extracted from a single institution's joint replacement registry, it is unclear the extent to which these findings are generalizable to other healthcare institutions. Next, a potential criticism is that our ML models could have performed better if given additional input data. However, input variables were selected based on current understanding of the predictors of MCID attainment after TKA and similar to datasets used by prior studies employing ML algorithms [9, 12, 14]. Furthermore, input variables for ML algorithms were limited to preoperative data as the greatest utility of such prediction tools was to enhance preoperative

Table 5 Permutation importance scores for ML models

SF-36 PCS		SF-36 MCS		SF-36 MCS	
Variable	Importance	Variable	Importance	Variable	Importance
Preop PCS	0.211	Preop MCS	0.457	Preop WOMAC	0.309
Age	0.040	BMI	0.001	Preop Pain Score	0.019
Preop MCS	0.011	-	-	Age	0.008
Surgeon	0.005	-	-	Payment Class	0.006
Diabetes	0.003	-	-	Surgeon	0.004

Importance is defined as the mean decrease in AUC when the variable is randomly shuffled (permuted)

counselling and modify patient expectations. While the inclusion of postoperative data may potentially improve its predictive performance, such algorithms have limited utility in the preoperative period. Lastly, there were other PROMs such as the Oxford Knee Score and Knee Society Score, as well as other subjective outcomes such as patient satisfaction and expectation fulfilment that were not considered in this study [23].

Conclusion

Machine learning algorithms did not perform significantly better than preoperative PROM thresholds in predicting MCID attainment after TKA. Future research should routinely evaluate the predictive advantage of ML algorithms over existing methods and determine the type of clinical problems which may benefit the most from it.

Funding No funding was received for this study.

Declarations

Conflict of interest The authors declare that there are no conflicting interests.

Ethical approval Ethics approval was obtained from the institutional review board prior to the initiation of this study (DSRB: 2020/00613).

References

- Batista GEAPA, Prati RC, Monard MC (2004) A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor Newsl* 6(1):20–29
- Berliner JL, Brodke DJ, Chan V, SooHoo NF, Bozic KJ (2017) Can preoperative patient-reported outcome measures be used to predict meaningful improvement in function after TKA? *Clin Orthop Relat Res* 475(1):149–157
- Berliner JL, Brodke DJ, Chan V, SooHoo NF, Bozic KJ (2016) John charnley award: Preoperative patient-reported outcome measures predict clinically meaningful improvement in function after THA. *Clin Orthop Relat Res* 474(2):321–329
- Cava W, Bauer C, Moore JH, Pendergrass SA (2020) Interpretation of machine learning predictions for patient outcomes in electronic health records. *AMIA Annu Symp Proc* 2019:572–581
- Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B (2019) A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 110:12–22
- Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, Cording E, Tomson D, Dodd C, Rollnick S, Edwards A, Barry M (2012) Shared decision making: a model for clinical practice. *J Gen Intern Med* 27(10):1361–1367
- Engel L, Beaton DE, Touma Z (2018) Minimal clinically important difference: a review of outcome measure score interpretation. *Rheum Dis Clin North Am* 44(2):177–188
- Escobar A, Quintana JM, Bilbao A, Aróstegui I, Lafuente I, Vidaurreta I (2007) Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage* 15(3):273–280
- Fontana MA, Lyman S, Sarker GK, Padgett DE, MacLean CH (2019) Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? *Clin Orthop Relat Res* 477(6):1267–1279
- Gagnier JJ, Mullins M, Huang H, Marinac-Dabic D, Ghambaryan A, Eloff B, Mirza F, Bayona M (2017) A systematic review of measurement properties of patient-reported outcome measures used in patients undergoing total knee arthroplasty. *J Arthroplasty* 32(5):1688–1697
- Haeberle HS, Helm JM, Navarro SM, Karnuta JM, Schaffer JL, Callaghan JJ, Mont MA, Kamath AF, Krebs VE, Ramkumar PN (2019) Artificial intelligence and machine learning in lower extremity arthroplasty: a review. *J Arthroplasty* 34(10):2201–2203
- Harris AHS, Kuo AC, Bowe TR, Manfredi L, Lalani NF, Giori NJ (2020) Can machine learning methods produce accurate and easy-to-use preoperative prediction models of one-year improvements in pain and functioning after knee arthroplasty? *J Arthroplasty* 36(1):112–117.e116
- Hawker GA, Conner-Spady BL, Bohm E, Dunbar MJ, Jones CA, Ravi B, Noseworthy T, Dick D, Powell J, Paul P, Marshall DA, BEST-Knee Study Team (2021) Patients' preoperative expectations of total knee arthroplasty and satisfaction with outcomes at one year: A prospective cohort study. *Arthritis Rheumatol* 73(2):223–231
- Huber M, Kurz C, Leidl R (2019) Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning. *BMC Med Inform Decis Mak* 19(1):3
- Laucis NC, Hays RD, Bhattacharyya T (2015) Scoring the SF-36 in orthopaedics: a brief guide. *J Bone Joint Surg AM* 97(19):1628–1634
- Maredupaka S, Meshram P, Chatte M, Kim WH, Kim TK (2020) Minimal clinically important difference of commonly used

- patient-reported outcome measures in total knee arthroplasty: review of terminologies, methods and proposed values. *Knee Surg Relat Res* 32(1):19
17. Mouelhi Y, Jouve E, Castelli C, Stephanie G (2020) How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. *Health Qual Life Outcomes* 18:136
 18. Padilla JA, Feng JE, Anoushiravani AA, Hozack WJ, Schwarzkopf R, Macaulay WB (2019) Modifying patient expectations can enhance total hip arthroplasty postoperative satisfaction. *J Arthroplasty* 34(7S):S209–S214
 19. Price AJ, Kang S, Cook JA, Dakin H, Blom A, Arden N, Fitzpatrick R, Beard DJ, ACHE Study team (2020) The use of patient-reported outcome measures to guide referral for hip and knee arthroplasty. *Bone Joint J.* 102(7):941–949
 20. Pua YH, Kang H, Thumboo J, Clark RA, Chew ES, Poon CL, Chong HC, Yeo SJ (2020) Machine learning methods are comparable to logistic regression techniques in predicting severe walking limitation following total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc* 28(10):3207–3216
 21. Quintana JM, Escobar A, Arostegui I, Bilbao A, Azkarate J, Goenaga JI, Arenaza JC (2006) Health-related quality of life and appropriateness of knee or hip joint replacement. *Arch Int Med* 166(2):220–226
 22. Rajkomar A, Dean J, Kohane I (2019) Machine learning in medicine. *N Engl J Med* 380(14):1347–1358
 23. Ramkumar HJD, Noble PC (2015) Patient-reported outcome measures after total knee arthroplasty: a systematic review. *Bone Joint Res* 4(7):120–127
 24. Safari S, Baratloo A, Elfil M, Negida A (2016) Evidence based emergency medicine; part 5 receiver operating curve and area under the curve. *Emerg* 4(2):111–113
 25. Thumboo J, Feng PH, Boey ML, Soh CH, Thio S, Fong KY (2000) Validation of the chinese SF-36 for quality of life assessment in patients with systemic lupus erythematosus. *Lupus* 9(9):708–712
 26. Walker LC, Clement ND, Deehan DJ (2019) Predicting the outcome of total knee arthroplasty using the WOMAC score: a review of the literature. *J Knee Surg* 32(8):736–741
 27. Xie F, Li SC, Goeree R, Tarride JE, O'Reilly D, Lo NN, Yeo SJ, Yang KY, Thumboo J (2008) Validation of chinese western ontario and mcmaster universities osteoarthritis index (WOMAC) in patients scheduled for total knee replacement. *Qual Life Res* 17(4):595–601
 28. Youden WJ (1950) Index for rating diagnostic tests. *Cancer* 3(1):32–35

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.