Heliyon 10 (2024) e36645

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Review article

Reliability and validity of patient-reported outcome measures in assessing knee osteoarthritis in the Chinese population: A systematic review

Jiayi Ren [a,1], Hongyuan Lu [b,1], Hang Gao [a,1], Xinglai Zhang [a,1], Yongni Zhang [c], Jin Li [a], Haoxiang He [a], Jiming Tao [a,*]

[a] Department of Rehabilitation Medicine, Shanghai Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, Shanghai, 201203, PR China
[b] Department of Spine, Guanghua Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, Shanghai, 200052, PR China
[c] Duquesne- China Health Institute, Duquesne University, United States

ARTICLE INFO

Keywords:
Reliability
Validity
Knee osteoarthritis
Patient-reported outcome measures
Systematic review

ABSTRACT

Objective: Knee osteoarthritis (KOA) is a prevalent condition in China, necessitating effective assessment tools for treatment outcomes. This study systematically reviews and analyzes the reliability, validity, and selection of patient-reported outcome measures (PROMs) for evaluating KOA.
Methods: Following PRISMA guidelines, a literature search was conducted across seven databases, including CNKI, PubMed, and Embase, covering publications from December 2012 to December 2022. The methodological quality of the studies was assessed using the COSMIN checklist.
Results: Twenty-one studies met the inclusion criteria, involving eight types of KOA PROMs. The Oxford Knee Score (OKS) and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) were the most frequently utilized, appearing in nine and four studies, respectively. OKS achieved a "strong" rating in internal consistency, test-retest reliability, content validity, responsiveness, and measurement error, while WOMAC received a "strong" rating in internal consistency, test-retest reliability, and content validity, with a "moderate" rating in structural validity.
Conclusion: Both OKS and WOMAC are effective PROMs for evaluating KOA in China. However, the choice of a specific tool should be based on the study's objectives and the practical context, considering each tool's reliability, validity, and other measurement properties.

1. Introduction

Knee osteoarthritis (KOA) is a prevalent chronic condition that significantly impacts the quality of life and functional abilities of patients [1–4]. With medical advancements, there has been an increasing focus on scientifically evaluating the efficacy of KOA treatments. Patient-reported outcome measures (PROMs), with their unique capability to capture patients' self-reported perspectives,

* Corresponding author. Department of rehabilitation medicine, Shanghai Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, No.528 Zhangheng Road, Zhangjiang Town, Pudong New Area, Shanghai 201203, PR China.
E-mail address: shuguangkf@163.com (J. Tao).
[1] These authors are regarded as co-first authors.

https://doi.org/10.1016/j.heliyon.2024.e36645
Received 6 January 2024; Received in revised form 5 June 2024; Accepted 20 August 2024
Available online 2 September 2024

focus on patient experiences and health outcomes, have emerged as an important tool for assessing the effectiveness of KOA treatments [5–7].

In recent years, PROMs have garnered widespread attention for their unique advantages [8,9]. Unlike other assessment methods, PROMs can comprehensively and accurately reflect the subjective feelings and experiences of patients. This offers clinicians and researchers a more holistic and objective set of metrics for evaluating treatment outcomes. Moreover, PROMs can measure the impact of treatments, aiding in shared decision-making between doctors and patients [6,7]. Thus, selecting appropriate PROMs is important in assessing the treatment effects on KOA patients.

Currently, research on PROMs in the assessment of KOA has several limitations. Firstly, existing studies are limited in scope, involving a narrow range of PROMs, making it challenging to provide reliable guidance on selection [10]. Secondly, the methodological quality of studies on PROMs varies greatly, lacking a unified and standardized evaluation system. This increases the difficulty for clinicians and researchers in selecting specific PROMs and limits their application in KOA assessment.

To address these issues, this study aims to conduct a comprehensive evaluation of the reliability, validity, and selection of PROMs in KOA assessment through systematic review analysis. Adopting the PRISMA standards and the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist, this research will rigorously screen and assess the quality of related literature to ensure the scientific integrity and reliability of the results [11]. By providing evaluations of PROMs in aspects such as internal consistency, test-retest reliability, content validity, responsiveness, and measurement error, this study aims to support clinical decision-making and treatment planning. With the selection of suitable PROMs, medical professionals can accurately and comprehensively assess the treatment effects on KOA patients, thereby optimizing treatment plans and enhancing patients' quality of life.

This research will systematically review the reliability, validity, and selection of PROMs in the assessment of KOA in China, identifying high-quality scales for use. Through the application of PRISMA standards and the COSMIN checklist, we aim to ensure the scientific rigor and reliability of our research methods. The outcomes of this study will provide clinicians and researchers with robust evidence on the selection of PROMs for KOA, laying the foundation for improving the quality of life of KOA patients.

2. Materials and methods

2.1. Literature search strategy

The study adhered to The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [12], guided by the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) recommendations. From December 2012 to December 2022, a systematic search was conducted across several databases, including CNKI, PubMed, Embase, Cochrane, CINAHL, Sportdiscus, and Scopus. English keywords utilized in the search comprised "Total knee replacement," "Knee replacement arthroplasty," "Unicompartmental knee arthroplasty," "Knee osteoarthritides," "Knee osteoarthritis," "Index," "Measure," "Instrument," "Scale," "Questionnaire," "Reliability," "Validity," "Responsiveness," and "Psychometric properties." The Chinese keywords included "膝关节炎 (knee osteoarthritis)," "膝关节置换术 (Knee replacement surgery)," "量表 (Scale)," "问卷 (Questionnaire)," "信度 (Reliability)," "效度 (Validity)," and "反应性 (Responsiveness)." All keywords were searched using Boolean logic to ensure a thorough retrieval process.

2.2. Inclusion and exclusion criteria

Studies were eligible for inclusion if they were published in English or Chinese in academic journals, involved KOA patients from China, and were original research utilizing PROMs. Exclusion criteria encompassed basic experimental studies such as animal or cadaver research, systematic reviews or commentary articles, case reports, and articles with incomplete data.

2.3. Data screening and extraction

A comprehensive literature search, coupled with a double-blind screening and data extraction process, was employed to minimize bias in this systematic review. Two researchers (Jiayi Ren and Hongyuan Lu) independently screened the literature, resolving any discrepancies through consultation with a third author, Jiming Tao. Following selection, key data such as study dates, clinical diagnoses, scale information, reliability, validity, and responsiveness were extracted from the chosen articles.

2.4. Methodological quality assessment

To control for literature bias, the methodological quality of the studies was assessed using the COSMIN checklist, focusing on three main measurement properties: reliability, validity, and responsiveness, encompassing eight sub-properties in total. The study referenced standards [13] that align with the COSMIN scale, applying a 2018 rating system to categorize studies into four levels: excellent, good, fair, and poor. Two researchers independently conducted each article's methodological quality rating, following COSMIN guidelines and quality standards (https://www.cosmin.nl/), to further reduce selection and researcher bias.

2.5. Measurement attribute interpretation

Reliability focuses on the stability and consistency of scale outcomes, including internal consistency, test-retest reliability, and measurement error. Validity measures the closeness between the true outcomes and those provided by the measurement tool, covering content and structural validity. Responsiveness examines the capacity of a scale to detect changes over time. To evaluate measurement properties of each PROM, the outcomes of each study's rating method were qualitatively classified as sufficient (+), insufficient (−), or indeterminate (?). Quality standards for assessing each measurement property are listed in Table 1.

2.6. *Summary of level of evidence*

Multiple assessments of the same scale may yield varying results and levels of evidence. The method [14] was utilized to achieve a broader level of evidence for each scale. Summaries of evidence for each PROM illustrated the number and methodological quality of studies (using the COSMIN checklist), the measurement properties of PROMs [15], and the consistency of results. Following the recommendations of the Cochrane Back Review Group, the overall assessment of evidence for PROMs was categorized as strong, moderate, limited, controversial, or unknown (Table 2). Two researchers independently conducted the ratings, with any disagreements resolved through consultation with a third party.

Table 1
Measurement attribute quality standards.

Characteristics		Definition	Rating	Quality criteria
Reliability	Internal consistency	Interrelationship among items within the same factor structure	(+)	(Uni)dimensional subscale with Cronbach's alpha (s) greater than or equal to 0.70
			?	Dimension unknown or Cronbach's alpha undetermined
			(−)	(Uni)dimensional subscale with Cronbach's alpha (s) < 0.70
	Test-retest reliability	Proportion of total variance in measurements caused by true differences among patients	(+)	ICC/Weighted Kappa ≥0.70 or Pearson r ≥ 0.80
			?	ICC/Weighted Kappa and Pearson r undetermined
			(−)	ICC/Weighted Kappa <0.70 or Pearson r < 0.80
	Measurement error	Systematic and random errors in patient ratings caused by true variation in non-measured constructs	(+)	MIC > SDC or MIC outside LOA
			?	MIC undefined
			(−)	MIC < SDC or MIC inside LOA
Validity	Content validity	Degree to which scale content reflects the response of target subjects	(+)	Target population perceives all items in the questionnaire as relevant and perceives the questionnaire as comprehensive
			?	No participation from the target population
			(−)	Target population perceives items in the questionnaire as irrelevant or perceives the questionnaire as incomplete
	Construct validity	Dimensions of the target construct measured by the scale	(+)	Factor explanation of ≥50 % variance
			?	Variance explanation not mentioned
			(−)	Factor explanation of <50 % variance
	Hypothesis testing	Degree of association between scale scores consistent with the theoretical derivation and scores on other scales	(+)	Results consistent with hypotheses
			?	No hypotheses defined (defined by the review team)
			(−)	Results inconsistent with hypotheses
	Cross-cultural validity	Extent to which items on translated or cross-culturally adapted scales reflect the response of the original version of the scale	(+)	In multigroup factor analysis, no significant differences between groups based on factors such as age, gender, language
			?	No multigroup factor analysis conducted
			(−)	In multiple factor analysis, there are significant differences between inter-group factors (such as age, gender, language)
Responsiveness		Measurement of longitudinal validity. Similar to construct validity, longitudinal validity should be evaluated by testing predefined hypotheses	(+)	(Using instruments to measure the correlation of the same structure, with a correlation of ≥0.50 or at least 75 % of the results in accordance with the hypothesis or AUC ≥0.70), and the correlation with the relevant structure is higher than that with the irrelevant structure
			?	Only correlated with unrelated structures
			(−)	The correlation with the structure of the scale is < 0.50 OR <75 % of the results in accordance with the hypothesis OR AUC <0.70 OR the correlation with the relevant structure is lower than that with the irrelevant structure

*ICC = correlation coefficient, MIC = minimal important change, SDC = minimal detectable change, LOA = limits of agreement, AUC = area under the curve. (Reprinted with modifications from: Huang H， Grant JA， Miller BS， Mirza FM， Gagnier JJ. A systematic review of the psychometric properties of patient reported outcome instruments for use in patients with rotator cuff disease.AmJ Sports Med.2015 Oct; 43 [10]:2572-82. Epub 2015 Jan 26).

Table 2
Grading of the level of evidence for overall quality.

Level	Grade	Evaluation Criteria
Strong	+++ or —	Consistently found to be "good" methodological quality across multiple studies, or "excellent" in one study
Moderate	++ or –	Consistently found to be "fair" methodological quality across multiple studies, or "good" in one study
Limited	+ or -	Found to be "fair" methodological quality in one study
Controversial	±	Controversial results
Unknown	?	Only studies with poor methodological quality

3. Results

3.1. Results of literature screening

As illustrated in Fig. 1, the literature search yielded a total of 4933 papers. The distribution of articles across databases was as follows: CNKI contributed 72 papers (1.4 %), PubMed 747 papers (15.2 %), Embase 713 papers (14.5 %), Cochrane 1405 papers (28.5 %), CINAHL 555 papers (11.3 %), SPORTDiscus 185 papers (3.7 %), and Scopus 1256 papers (25.4 %). After the removal of duplicates, 3268 articles remained. Titles, abstracts, keywords, types of literature, and full texts of each article were independently screened by two reviewers, leading to the consensus inclusion of 21 articles [16–36] for further analysis.

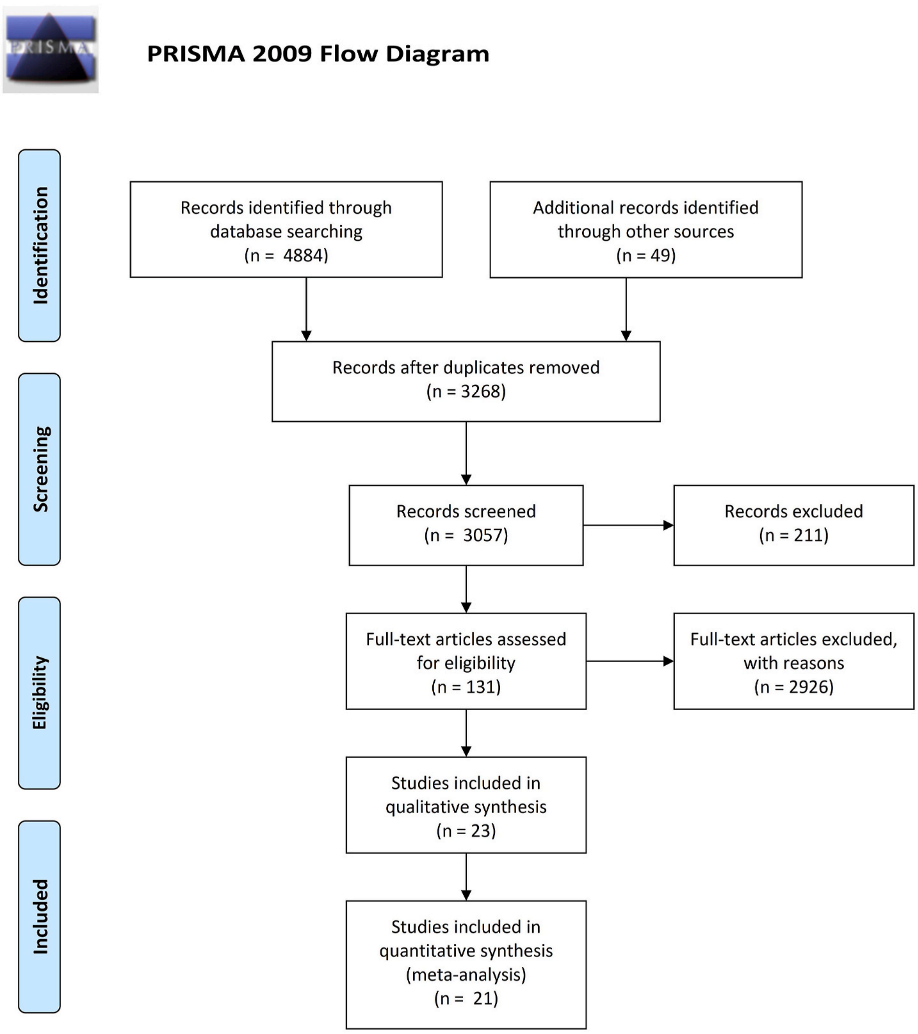

Fig. 1. PRISMA study flow diagram.

3.2. Overview of PROMs evaluated in included studies

The 21 included articles assessed eight different PROMs for KOA, as summarized in Table 3. The most frequently evaluated scales were the OKS in nine articles and the WOMAC in four articles. Other evaluated scales included the Knee injury and Osteoarthritis Outcome Score (KOOS) in two articles, the Intermittent and Constant Osteoarthritis Pain (ICOAP) in two articles, and the Lower Extremity Functional Scale (LEFS), Patient Expectations of Physician Interaction (PEPPI-10), Pain Catastrophizing Scale (PCS), and the Japanese Knee Osteoarthritis Measure (JKOM) each in one article.

3.3. Methodological quality and rating of measurement properties of scales

The methodological quality and COSMIN grading of each PROM study are detailed in Table 4.

Internal consistency: Out of the 21 articles, 20 (95.2 %) analyzed internal consistency. Except for JKOM, which was classified as uncertain due to an undetermined Cronbach's alpha, all studies were deemed sufficient. Except for three articles on PCS and OKS rated as fair, the methodological quality was generally rated as excellent.

Test-retest reliability: Fifteen out of 21 articles (71.4 %) conducted studies on test-retest reliability, all receiving sufficient ratings for this measurement property. The methodological quality of studies on OKS, KOOS, ICOAP, LEFS, and PCS was rated as good or fair.

Content validity: Seven out of 21 articles (33.3 %) conducted content validity studies, all achieving sufficient ratings. The methodological quality of WOMAC, KOOS, and ICOAP studies was rated as good or excellent.

Structural validity: Nineteen out of 21 articles (90.5 %) analyzed structural validity. However, studies on OKS, WOMAC, KOOS, ICOAP, LEFS, and PCS were rated as unknown due to unspecified explained variance, and the methodological quality was rated as fair or biased due to the absence of exploratory factor analysis or insufficient sample size.

Measurement error: Only three articles (14.3 %) studied measurement error, with OKS and ICOAP receiving positive evaluations. KOOS was rated as insufficient due to undefined Standardized Response Mean (SRM) and Minimal Detectable Change (MDC), and the methodological quality was considered moderate.

Responsiveness: Three articles (14.3 %) analyzed scale responsiveness, with OKS, ICOAP, and LEFS all receiving positive evaluations and an excellent methodological quality rating.

3.4. Overall measurement property ratings of included PROMs

Based on the quantity and quality of evidence (Table 5), although structural validity was limited, OKS remained the best-evidenced measure for assessing KOA, with its nine articles achieving a strong overall rating across seven measurement properties, including internal consistency, test-retest reliability, content validity, responsiveness, measurement error, hypothesis testing, and cross-cultural validity. WOMAC, despite lacking data on responsiveness, measurement error, and hypothesis testing, received a strong overall rating in internal consistency, test-retest reliability, and content validity, with structural validity rated as moderate.

ICOAP, while achieving a strong overall rating in four measurement properties, including internal consistency, content validity, responsiveness, and measurement error based on two articles, was rated as moderate in cross-cultural validity, limited in test-retest reliability, and controversial in hypothesis testing.

Table 3
Summary of the scale.

Abbreviation	Full Name	Target Population	Measurement Dimension	Number of Items
OKS	The Oxford Knee Score	Knee joint replacement surgery/knee arthritis	Pain, daily life functioning	12
WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index Function Score	Knee arthritis	Pain, stiffness, and difficulty in performing daily activities	24
ICOAP	intermittent and constant osteoarthritis pain	Knee arthritis	The pain experience of patients with knee osteoarthritis, including pain intensity, frequency, and duration, as well as its impact on emotions, sleep, and quality of life	11
LEFS	Lower Limb Functional Scale	Lower limb muscle and skeletal dysfunction	Physical activity	20
PEPPI-10	10-item Perceived Efficacy in Patient Physician Interaction scale	All patients	Doctor-patient interaction	10
KOOS	Knee Injury and Osteoarthritis Outcome Score	Knee joint injury	Frequency and severity of pain, symptoms, activities of daily living (ADL), exercise/leisure	42
PCS	Pain Catastrophizing Scale	Fibromyalgia/low back pain/knee arthritis	Degree of pain (divided into three sub-scales: rumination, for example, I cannot stop thinking about how much it hurts; magnification, for example, I worry that something serious might happen; and helplessness, for example, I have no way to reduce the intensity of the pain)	13
JKOM	The Japanese knee osteoarthritis evaluation scale	Knee arthritis	Pain and stiffness, daily life status, usual activity status, health status	25

Table 4
Summary of COSMIN grading assessment for the measurement properties of the scale.

Scale	Source	Sample size	Internal consistency	Test-retest reliability	Content validity	Construct validity	Reactiveness	Measurement error	Hypothesis testing	Cross-cultural validity
OKS	R. T. H. Cheung, 2017	100	E/(+)	0	0	G/?	0	0	0	0
	K. Lin, 2017	253	E/(+)	G/(+)	0	G/(+)	E/(+)	0	G/(+)	E/(+)
	Lin Kai, 2017	114	E/(+)	0	E/(+)	G/?	0	0	0	G/(+)
	Li Yangjie, 2019	96	E/(+)	0	G/(+)	F/(+)	0	0	0	0
	Wu Hao, 2020	30	E/(+)	G/(+)	F/(+)	P/(+)	0	0	0	0
	J. R. M. Ngwayi, 2020	149	F/(+)	G/(+)	0	G/?	0	0	G/(+)	0
	C. Chen, 2020	30	E/(+)	G/(+)	0	P/?	0	0	0	0
	W. Deng, 2021	267	F/(+)	E/(+)	0	G/(+)	0	E/(+)	G/(+)	0
	X. L. Liu, 2022	159	E/(+)	0	0	F/?	0	0	G/(+)	0
WOMAC	T. Symonds, 2015	287	E/(+)	E/(+)	0	F/(+)	0	0	0	G/(+)
	Xia Chuantao, 2015	177	E/(+)	E/(+)	G/(+)	G/(+)	0	0	0	F/(+)
	Shen Zhengdong, 2019	113	E/(+)	F/(+)	G/(+)	G/?	0	0	0	0
	Chen Wei, 2019	24	0	P/+	0	0	0	0	0	0
KOOS	Cheung, 2016	100	E/(+)	G/(+)	0	P/(+)	0	0	G/(+)	0
	Cheng, 2019	125	E/(+)	G/(+)	G/(+)	G/?	0	F/(−)	0	0
ICOAP	Zhang, 2017	108	E/(+)	F/(+)	0	G/(+)	E/(+)	0	F/(+)	0
	Sit, 2019	110	E/(+)	G/(+)	E/(+)	F/?	0	E/(+)	G/(+)	G/(+)
LEFS	Xu, 2020	108	E/(+)	G/(+)	0	G/?	G/(+)	0	0	G/(+)
PEPPI-10	Zhao, 2016	110	E/(+)	0	0	E/(+)	0	0	0	0
PCS	Ong, 2021	675	F/(+)	0	0	E/(?)	0	0	F/(+)	0
JKOM	Xu Shouyu, 2013	171	E/(?)	G/(+)	0	0	0	0	0	0

*Methodology quality rating: E = excellent, G = good, F = fair, P = poor, 0 = unknown; Measurement attribute assessment: sufficient(+), insufficient(−), or uncertain(?).

Table 5
Overall measurement attribute rating.

Scales	Internal consistency	Test-retest reliability	Content validity	Construct validity	Responsiveness	Measurement error	Hypothesis testing	Cross-cultural validity
OKS	+++	+++	+++	+	+++	+++	+++	+++
WOMAC	+++	+++	+++	++	0	0	0	±
KOOS	+++	++	++	±	0	?	++	0
ICOAP	+++	+	+++	±	+++	+++	±	++
LEFS	+++	++	0	–	++	0	0	++
PEPPI-10	+++	0	0	+++	0	0	0	0
PCS	+	0	0	–	0	0	+	0
JKOM	–	0	0	++	0	0	0	0

4. Discussion

This review systematically evaluates the reliability, validity, and selection of PROMs in the assessment of KOA within a Chinese population. Among the multitude of PROMs available, the OKS and the WOMAC have demonstrated excellent reliability and validity [37–40]. These findings provide robust tool selection recommendations for clinicians and researchers assessing the effects of KOA (Fig. 2).

Assessing the outcomes of KOA treatments involves a complex process that requires the consideration of various factors. PROMs offer an objective tool for clinicians to more accurately gauge treatment effectiveness and disease progression. Especially in aspects such as pain and functional recovery, PROMs provide essential feedback, assisting in the adjustment of individualized treatment plans [5]. While previous studies have extensively evaluated different PROMs, this study is the first to systematically compare the applicability of various PROMs in the assessment of KOA in China. Consistent with conclusions from some prior studies, we also find that OKS and WOMAC achieve high ratings across most measurement properties [41,42]. Compared to previous piecemeal assessments, our study offers a more comprehensive and objective perspective.

OKS is specifically designed to assess the postoperative recovery of knee replacement surgery patients but has also been widely applied to evaluate the pain and functional status of KOA patients. The widespread use of OKS reflects its high reliability and validity in measuring knee function, especially in terms of content validity and responsiveness. Additionally, the design of OKS is straightforward, making it easy to understand and complete, thus serving as an ideal tool for assessing knee function impairment and treatment outcomes [37,43,44]. WOMAC, a specific scale for assessing pain, stiffness, and functional limitations in patients with lower limb osteoarthritis, is particularly applicable to patients with knee and hip osteoarthritis. WOMAC is considered one of the gold standards for assessing osteoarthritis pain and functional impairment, with the advantage of reflecting the patients' specific difficulties and challenges in detail [38,43,45]. In this study, although WOMAC showed some shortcomings in structural validity, it still performed excellently in terms of internal consistency and test-retest reliability. Overall, both tools provide reliable data support for assessing symptoms and function in KOA.

Despite aiming for comprehensiveness and objectivity, this study has limitations. For example, the literature included mainly comes from seven databases, potentially missing valuable studies from other sources. Additionally, the number of studies on some PROMs is relatively small, possibly affecting the comprehensiveness and objectivity of the evaluations. It is important to note that this study assesses the measurement properties of individual PROMs and then uses the COSMIN checklist to summarize the evidence for each PROM. This approach is fundamentally different from traditional meta-analysis, each with its advantages and limitations. The COSMIN method focuses on assessing and summarizing the measurement properties (e.g., reliability, validity, responsiveness) of individual scales or tools across studies, emphasizing the quality and applicability of scales. In contrast, meta-analysis aims to quantitatively combine results from different studies to estimate an overall effect size, focusing on the statistical synthesis of study outcomes. The COSMIN method is particularly suitable when evaluating tool or scale quality, especially when the research goal is to select or recommend the most appropriate measurement tools. Meta-analysis is applicable for the quantitative assessment of specific interventions or factor effects. Although no meta-analysis was conducted, this study provides a powerful method to assess the quality and applicability of measurement tools.

Recent advances in the application of PROMs in osteoarthritis assessment have been significant, thanks to the integration of digital health technologies and big data analytics. These advances have not only enabled real-time monitoring of patient symptoms and quality of life but also facilitated the development of personalized treatment strategies and improved patient care [5]. Digital health technologies, such as mobile apps, wearable devices, and remote monitoring systems, have been integrated with PROMs, improving the quality and frequency of data collection and enhancing patient engagement. This allows doctors to adjust treatment plans based on real-time data, assessing patients' functional status more accurately [46]. Furthermore, disease-specific PROMs are being developed to more finely reflect the unique experiences of osteoarthritis patients, accurately capturing the impact on aspects such as pain and functional impairment and guiding the formulation of personalized treatment plans [47]. Additionally, the application of machine learning algorithms in PROM data analysis offers new perspectives for identifying key factors in treatment outcomes and predicting treatment effectiveness. This helps doctors make early and accurate predictions and provides a scientific basis for formulating effective personalized treatment strategies [48].

The current study's findings offer significant reference value for future research, especially regarding selecting and using PROMs.

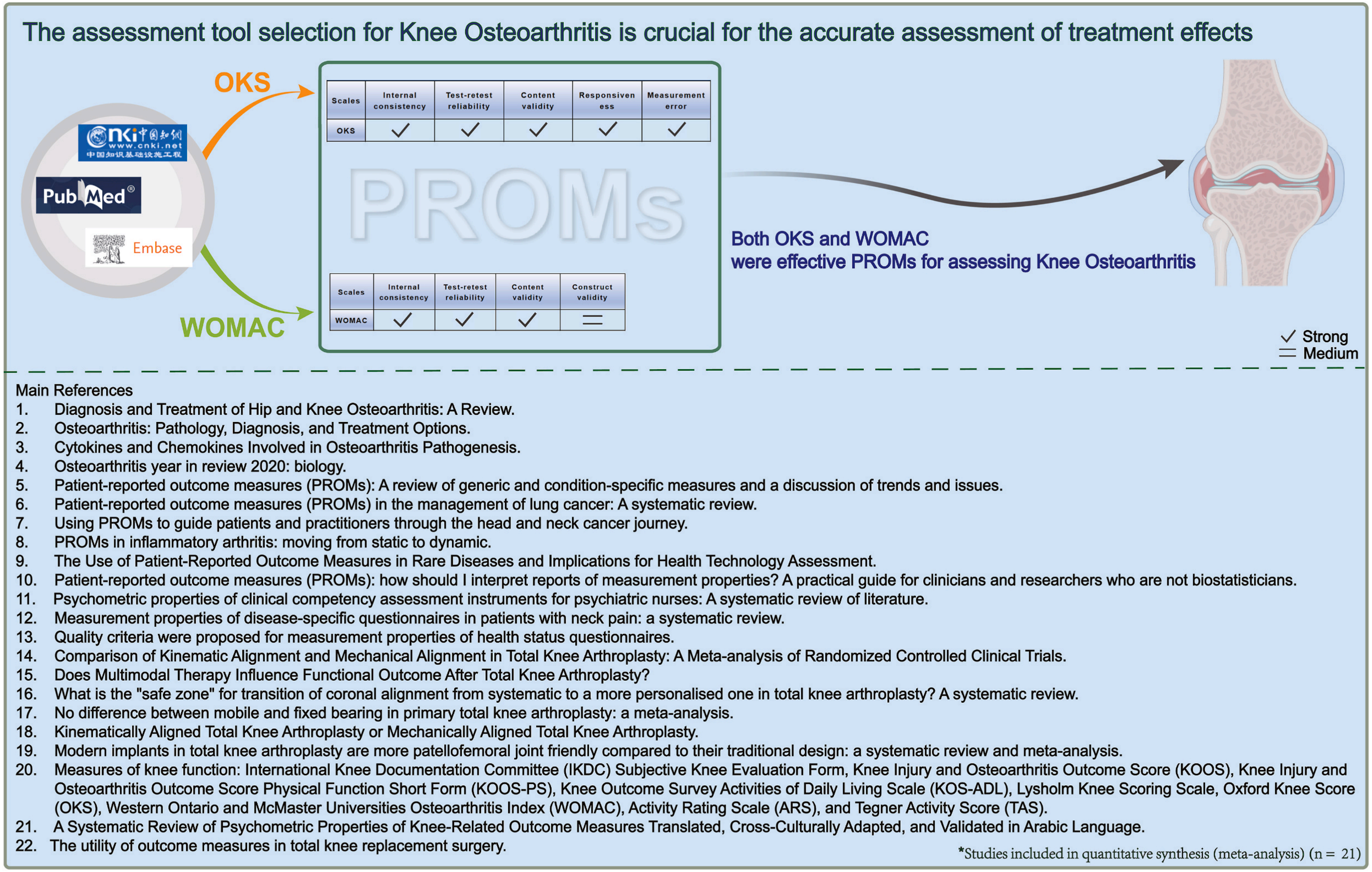

Fig. 2. Mechanism diagram for this study.

Considering the complexity of KOA, future studies need to further explore other potential assessment tools and validate them in larger samples and diverse populations. Notably, patient involvement and co-design approaches have revolutionized the development of PROMs, particularly in the field of osteoarthritis, ensuring that evaluation tools authentically reflect outcomes of concern to patients, enhancing their effectiveness and application value in clinical and research settings.

The conclusions of this study underscore the importance of OKS and WOMAC in assessing KOA. They have shown high reliability and validity across multiple measurement properties, providing valuable tools for clinical and research work. However, research is a continually advancing process, and as technology evolves and more studies accumulate, future work may identify even more precise and practical assessment tools. Hopefully, this research will offer new insights into KOA assessment and stimulate further study in related fields.

Funding

This study was supported by Key supported discipline in the Shanghai Health System - Rehabilitation Medicine (2023ZDFC0301) and Science and Technology Development Project of Shanghai University of Traditional Chinese Medicine (Natural Science category 23KFL009).

Ethical statement

Our study did not require an ethical board approval because it did not contain human or animal trials.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author/s.

CRediT authorship contribution statement

Jiayi Ren: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration. **Hongyuan Lu:** Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Hang Gao:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Conceptualization. **Xinglai Zhang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology. **Yongni Zhang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Conceptualization. **Jin Li:** Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Haoxiang He:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation. **Jiming Tao:** Writing – review & editing, Writing – original draft, Visualization, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e36645.

References

[1] J.N. Katz, K.R. Arant, R.F. Loeser, Diagnosis and treatment of hip and knee osteoarthritis: a review, JAMA 325 (6) (2021) 568–578, https://doi.org/10.1001/jama.2020.22171.
[2] B. Abramoff, F.E. Caldera, Osteoarthritis: pathology, diagnosis, and treatment options, Med Clin North Am 104 (2) (2020) 293–311, https://doi.org/10.1016/j.mcna.2019.10.007.
[3] V. Molnar, V. Matišić, I. Kodvanj, et al., Cytokines and chemokines involved in osteoarthritis pathogenesis, Int. J. Mol. Sci. 22 (17) (2021) 9208, https://doi.org/10.3390/ijms22179208. Published 2021 Aug 26.

[4] M.H.J. van den Bosch, Osteoarthritis year in review 2020: biology, Osteoarthritis Cartilage 29 (2) (2021) 143–150, https://doi.org/10.1016/j.joca.2020.10.006.
[5] K. Churruca, C. Pomare, L.A. Ellis, et al., Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues, Health Expect. 24 (4) (2021) 1015–1024, https://doi.org/10.1111/hex.13254.
[6] Y.B. Bouazza, I. Chiairi, O. El Kharbouchi, et al., Patient-reported outcome measures (PROMs) in the management of lung cancer: a systematic review, Lung Cancer 113 (2017) 140–151, https://doi.org/10.1016/j.lungcan.2017.09.011.
[7] S.N. Rogers, B. Barber, Using PROMs to guide patients and practitioners through the head and neck cancer journey, Patient Relat. Outcome Meas. 8 (2017) 133–142, https://doi.org/10.2147/PROM.S129012. Published 2017 Nov 8.
[8] Y. El Miedany, PROMs in inflammatory arthritis: moving from static to dynamic, Clin. Rheumatol. 32 (6) (2013) 735–742, https://doi.org/10.1007/s10067-013-2228-0.
[9] A. Whittal, M. Meregaglia, E. Nicod, The use of patient-reported outcome measures in rare diseases and implications for health technology assessment, Patient 14 (5) (2021) 485–503, https://doi.org/10.1007/s40271-020-00493-w.
[10] M. Davidson, J. Keating, Patient-reported outcome measures (PROMs): how should I interpret reports of measurement properties? A practical guide for clinicians and researchers who are not biostatisticians, Br. J. Sports Med. 48 (9) (2014) 792–796, https://doi.org/10.1136/bjsports-2012-091704.
[11] E. Sheikhbahaeddinzadeh, T. Ashktorab, A. Ebadi, Psychometric properties of clinical competency assessment instruments for psychiatric nurses: a systematic review of literature, J. Educ. Health Promot. 11 (2022) 396, https://doi.org/10.4103/jehp.jehp_1779_21. Published 2022 Dec 28.
[12] M.J. Page, J.E. McKenzie, P.M. Bossuyt, et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, BMJ 372 (2021) n71, https://doi.org/10.1136/bmj.n71. Published 2021 Mar 29.
[13] C.A.C. Prinsen, L.B. Mokkink, L.M. Bouter, et al., COSMIN guideline for systematic reviews of patient-reported outcome measures, Qual. Life Res. 27 (5) (2018) 1147–1157, https://doi.org/10.1007/s11136-018-1798-3.
[14] J.M. Schellingerhout, A.P. Verhagen, M.W. Heymans, B.W. Koes, H.C. de Vet, C.B. Terwee, Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review, Qual. Life Res. 21 (4) (2012) 659–670, https://doi.org/10.1007/s11136-011-9965-9.
[15] C.B. Terwee, S.D. Bot, M.R. de Boer, et al., Quality criteria were proposed for measurement properties of health status questionnaires, J. Clin. Epidemiol. 60 (1) (2007) 34–42, https://doi.org/10.1016/j.jclinepi.2006.03.012.
[16] C. Chen, W. Wang, H. Wu, et al., Cross-cultural translation and validation of the Chinese oxford knee score and the activity and participation questionnaire, J. Orthop. Surg. 28 (2) (2020) 2309499020910668, https://doi.org/10.1177/2309499020910668.
[17] A.S.K. Cheng, K.C. Chan, S.Y. Chan, et al., Cross-cultural adaptation and validation of the Hong Kong version of the knee injury and osteoarthritis outcome score (HK-KOOS) for patients with knee osteoarthritis, Occup. Ther. Int. 2019 (2019) 8270637, https://doi.org/10.1155/2019/8270637. Published 2019 Aug 14.
[18] R.T. Cheung, S.P. Ngai, K.K. Ho, Chinese adaptation and validation of the Knee Injury and Osteoarthritis Outcome Score (KOOS) in patients with knee osteoarthritis, Rheumatol. Int. 36 (10) (2016) 1449–1454, https://doi.org/10.1007/s00296-016-3539-7.
[19] R.T.H. Cheung, S.P.C. Ngai, K.K.W. Ho, Chinese translation and validation of the Oxford Knee Scale for patients with knee osteoarthritis, Hong Kong Physiother. J. 37 (2017) 46–49, https://doi.org/10.1016/j.hkpj.2017.03.002. Published 2017 May 10.
[20] J.R.M. Ngwayi, J. Tan, N. Liang, D.E. Porter, Reliability and validity of 3 different Chinese versions of the Oxford knee score (OKS), Arthroplasty 2 (1) (2020) 31, https://doi.org/10.1186/s42836-020-00049-1. Published 2020 Oct 21.
[21] K. Lin, L. Bao, J. Wang, K. Fujita, K. Makimoto, X. Liao, Validation of the Chinese (Mandarin) version of the oxford knee score in patients with knee osteoarthritis, Clin. Orthop. Relat. Res. 475 (12) (2017) 2992–3004, https://doi.org/10.1007/s11999-017-5495-2.
[22] W.J. Ong, Y.H. Kwan, Z.Y. Lim, et al., Measurement properties of Pain Catastrophizing Scale in patients with knee osteoarthritis, Clin. Rheumatol. 40 (1) (2021) 295–301, https://doi.org/10.1007/s10067-020-05163-8.
[23] R.W.S. Sit, D.C.C. Chan, W. Wong, B.H.K. Yip, L.L.Y. Chow, S.Y.S. Wong, Translation, cross-cultural adaptation and validation of the traditional Chinese intermittent and constant osteoarthritis pain (ICOAP) questionnaire for knee osteoarthritis, BMJ Open 9 (3) (2019) e026006, https://doi.org/10.1136/bmjopen-2018-026006. Published 2019 Mar 30.
[24] T. Symonds, B. Hughes, S. Liao, Q. Ang, N. Bellamy, Validation of the Chinese western Ontario and McMaster Universities osteoarthritis Index in patients from mainland China with osteoarthritis of the knee, Arthritis Care Res. 67 (11) (2015) 1553–1560, https://doi.org/10.1002/acr.22631.
[25] W. Deng, H. Shao, Y. Zhou, H. Li, Z. Wang, Y. Huang, Reliability and validity of commonly used patient-reported outcome measures (PROMs) after medial unicompartmental knee arthroplasty, Orthop Traumatol Surg Res 108 (8) (2022) 103096, https://doi.org/10.1016/j.otsr.2021.103096.
[26] X.L. Liu, Y.Y. Huang, T. Wang, et al., Psychometric assessment of the Chinese version of the Oxford Knee Score in breast cancer survivors experiencing hormone treatment-related knee dysfunction, Asia Pac J Oncol Nurs 9 (3) (2022) 135–142, https://doi.org/10.1016/j.apjon.2022.01.001. Published 2022 Jan 19.
[27] L. Xu, C. Wang, C. Zhang, X. Feng, W. Tong, Cross-cultural adaption and validation of simplified Chinese version of the lower extremity function scale in patients with knee osteoarthritis, Clin. Rheumatol. 39 (10) (2020) 3041–3048, https://doi.org/10.1007/s10067-020-05077-5.
[28] C. Zhang, D.H. Liu, Y.L. Qu, et al., Transcultural adaptation and validation of the Chinese version of the intermittent and constant osteoarthritis pain (ICOAP) measure in patients with knee osteoarthritis, Osteoarthritis Cartilage 25 (4) (2017) 506–512, https://doi.org/10.1016/j.joca.2016.11.012.
[29] H. Zhao, W. Luo, R.C. Maly, J. Liu, J. Lee, Y. Cui, Validation of the Chinese version 10-item perceived efficacy in patient-physician interactions scale in patients with osteoarthritis, Patient Prefer. Adherence 10 (2016) 2189–2195, https://doi.org/10.2147/PPA.S110883. Published 2016 Oct 28.
[30] W. Chen, Y. Guo, X. Li, et al., Test-retest reliability of the western Ontario and McMaster Universities osteoarthritis Index, Chin. J. Rehabil. Theory Pract. 16 (1) (2010) 23–24, https://doi.org/10.3969/j.issn.1006-9771.2010.01.010.
[31] Y. Li, Q. Sun, C. Deng, et al., Reliability and validity analysis of the Oxford Knee Score applied to knee osteoarthritis in the northeastern region of China, Pract J Orthop 25 (10) (2019) 936–940.
[32] K. Lin, Cross-sectional and Longitudinal Study of the Chinese Version of the Oxford Knee Score in Patients Undergoing Total Knee Arthroplasty [dissertation], Southern Medical University, Guangdong, 2018.
[33] Z. Shen, H. Yu, J. Wang, et al., Application of the modified western Ontario and McMaster Universities osteoarthritis Index in knee osteoarthritis, Chin. Med. J. 99 (7) (2019) 537–541, https://doi.org/10.3760/cma.j.issn.0376-2491.2019.07.012.
[34] H. Wu, W. Wang, M. Sun, et al., Application of the Chinese version of the oxford knee score (OKS) in the Chinese population, in: Proceedings of the 8th Forum of the Chinese Journal of Orthopedics, 2015, pp. 249–254.
[35] C. Xia, F. Yu, H. Fang, et al., Reliability and validity evaluation of the western Ontario and McMaster Universities osteoarthritis Index in kashin-beck disease, Chin. J. Endemiol. 34 (4) (2015) 274–277, https://doi.org/10.3760/cma.j.issn.2095-4255.2015.04.011.
[36] S. Xu, X. Yao, Y. Wu, et al., Reliability study of the Chinese version of the Japanese knee osteoarthritis function scale, Chin. J. Rehabil. Med. 29 (8) (2014) 723–725, https://doi.org/10.3969/j.issn.1001-1242.2014.08.006.
[37] Z.X. Gao, N.J. Long, S.Y. Zhang, W. Yu, Y.X. Dai, C. Xiao, Comparison of kinematic alignment and mechanical alignment in total knee arthroplasty: a meta-analysis of randomized controlled clinical trials, Orthop. Surg. 12 (6) (2020) 1567–1578, https://doi.org/10.1111/os.12826.
[38] A.A. Magan, S.S. Ahmed, B. Paton, S. Konan, F.S. Haddad, Does multimodal therapy influence functional outcome after total knee arthroplasty? Orthop Clin North Am. 51 (4) (2020) 453–459, https://doi.org/10.1016/j.ocl.2020.06.011.
[39] B.L. Schelker, A.M. Nowakowski, M.T. Hirschmann, What is the "safe zone" for transition of coronal alignment from systematic to a more personalised one in total knee arthroplasty? A systematic review, Knee Surg. Sports Traumatol. Arthrosc. 30 (2) (2022) 419–427, https://doi.org/10.1007/s00167-021-06811-5.
[40] F. Migliorini, N. Maffulli, F. Cuozzo, M. Pilone, K. Elsner, J. Eschweiler, No difference between mobile and fixed bearing in primary total knee arthroplasty: a meta-analysis, Knee Surg. Sports Traumatol. Arthrosc. 30 (9) (2022) 3138–3154, https://doi.org/10.1007/s00167-022-07065-5.
[41] T. Takahashi, J. Ansari, H.G. Pandit, Kinematically aligned total knee arthroplasty or mechanically aligned total knee arthroplasty, J. Knee Surg. 31 (10) (2018) 999–1006, https://doi.org/10.1055/s-0038-1632378.
[42] A.K. Choudhury, K. Gupta, S. Ansari, B.S. Raja, S. Paul, R.B. Kalia, Modern implants in total knee arthroplasty are more patellofemoral joint friendly compared to their traditional design: a systematic review and meta-analysis, Arch. Orthop. Trauma Surg. 143 (9) (2023) 5873–5889, https://doi.org/10.1007/s00402-023-04832-y.

[43] N.J. Collins, D. Misra, D.T. Felson, K.M. Crossley, E.M. Roos, Measures of knee function: international knee documentation committee (IKDC) subjective knee evaluation form, knee injury and osteoarthritis outcome score (KOOS), knee injury and osteoarthritis outcome score physical function short form (KOOS-ps), knee outcome survey activities of daily living scale (KOS-adl), lysholm knee scoring scale, oxford knee score (OKS), western Ontario and McMaster Universities osteoarthritis Index (WOMAC), activity rating scale (ARS), and tegner activity score (TAS), Arthritis Care Res. 63 (Suppl 11) (2011) S208–S228, https://doi.org/10.1002/acr.20632, 0 11.
[44] M. Ateef, M. Alqahtani, M. Alzhrani, A.A. Alkathiry, A. Alanazi, S.A. Alshewaier, A systematic review of psychometric properties of knee-related outcome measures translated, cross-culturally adapted, and validated in Arabic language, Healthcare (Basel) 10 (9) (2022) 1631, https://doi.org/10.3390/healthcare10091631. Published 2022 Aug 26.
[45] M.M. Dowsey, P.F. Choong, The utility of outcome measures in total knee replacement surgery, Int J Rheumatol 2013 (2013) 506518, https://doi.org/10.1155/2013/506518.
[46] V.V. Lee, N.Y. Lau, D.J.Y. Xi, et al., A systematic review of the development and psychometric properties of constipation-related patient-reported outcome measures: opportunities for digital health, J Neurogastroenterol Motil 28 (3) (2022) 376–389, https://doi.org/10.5056/jnm22004.
[47] S. van Pottelberghe, N. Kupper, E. Scheirlynck, et al., Are disease-specific patient-reported outcomes measures (PROMs) used in cardiogenetics? A systematic review, Eur. J. Hum. Genet. (2023), https://doi.org/10.1038/s41431-023-01510-w. Published online December 14.
[48] C. Klemt, A.C. Uzosike, J.G. Esposito, et al., The utility of machine learning algorithms for the prediction of patient-reported outcome measures following primary hip and knee total joint arthroplasty, Arch. Orthop. Trauma Surg. 143 (4) (2023) 2235–2245, https://doi.org/10.1007/s00402-022-04526-x.