Received: 17 August 2023 | Revised: 29 February 2024 | Accepted: 24 March 2024

DOI: 10.1002/ejp.2272

EJP European Journal of Pain

REVIEW ARTICLE

Systematic reviews and quality assessment of patient-reported outcome measures for physical function in comparative effectiveness studies regarding acute postoperative pain after total knee arthroplasty—Do we need to start all over again?

H. Heitkamp[1] | **D. Heußner**[1] | **D. C. Rosenberger**[1] | **K. Schnabel**[1] | **D. Rosenthal**[1] | **S. Bigalke**[1] | **T. V. Maeßen**[1] | **D. Hohenschurz-Schmidt**[2,3] | **H. Liedgens**[4] | **U. Kaiser**[5] | **E. M. Pogatzki-Zahn**[1]

[1]Clinic for Anaesthesiology, Intensive Care Medicine and Pain Therapy, University Hospital Münster, Münster, Germany

[2]Pain Research Group, Department of Surgery and Cancer, Imperial College London, London, UK

[3]Research Department, University College of Osteopathy, London, UK

[4]Gruenenthal GmbH, Aachen, Germany

[5]Clinic for Anaesthesiology and Intensive Care Campus Lübeck, University Hospital Schleswig-Holstein, Lübeck, Germany

Correspondence
E. M. Pogatzki-Zahn, Department of Anesthesiology, Intensive Care and Pain Medicine, University Hospital Muenster, Albert-Schweitzer-Campus 1, A1, 48149 Muenster, Germany.
Email: pogatzki@anit.uni-muenster.de

Abstract

Background and Objective: Recently, a consensus process specified a core outcome set (COS) of domains to be assessed in each comparative effectiveness research and clinical practice related to acute postoperative pain. Physical function (PF) was one of these domains. The aim of this review was to investigate which patient-reported outcome measures (PROMs) are used to assess PF after total knee arthroplasty (TKA) in clinical trials and if they fulfil basic requirements for a COS of PROMs based on their psychometric properties.

Methods: A systematic review of randomized controlled trials and observational studies based on a search in MEDLINE, EMBASE and CENTRAL was undertaken. PROMs and performance measures were extracted and investigated, including evaluation of psychometric properties of PROMs based on COSMIN recommendations.

Results: From initially 2896 identified records, 479 studies were included in the qualitative synthesis. Only 87 of these trials (18%) assessed PF using PROMs, whereas especially performance outcome measures were used in 470 studies (98%). Application of the 'COSMIN Risk-of-Bias-Box 1' to 13 of the 14 identified PROMs resulted in insufficient content validity of the included PROMs regarding the target population based on the inauguration or development articles.

Conclusion: Our data indicate that a patient-centred postoperative assessment of PF in pain-related clinical trials early after TKA is not common, even though patient-reported assessment is widely recommended. In addition, none of the applied PROMs shows content validity based on their inauguration or development articles for the assessment of postoperative pain-related PF after TKA.

Significance: A systematic search for patient-reported outcome measures assessing postoperative, pain-related physical function after total knee arthroplasty in clinical trials and assessment of their content validity revealed none that fulfilled requirements based on COSMIN recommendations.

1 | INTRODUCTION

Postsurgical pain is often insufficiently managed (Vollert et al., 2024), leading to acute and long-term complications including impaired recovery, chronic postsurgical pain and persistent opioid use—the latter exacerbating the opioid crisis (Brat et al., 2018; Lawal et al., 2020; Rosenberger & Pogatzki-Zahn, 2022). Orthopaedic surgeries, notably total knee arthroplasty (TKA), have particularly high rates of severe acute postsurgical pain (Vollert et al., 2024) and persistent opioid use (Kent et al., 2019; Sitter & Forget, 2021). Optimizing pain management for surgeries like TKA is essential to reduce suffering and prevent long-term repercussions.

Clinical practice guidelines and evidence-based recommendations for perioperative pain management are based on Randomized Controlled Trials (RCTs) with considerable limitations (Terkawi et al., 2017). The quality of outcome assessment in trials is crucial for interpreting study results, synthesizing data, drawing meta-analyses and future allocations of patients (Kersting et al., 2020; Pogatzki-Zahn et al., 2019). However, the observable lack of standardization, and consequently, the heterogeneity in outcome assessment across clinical trials hampers comparability of results. Inconsistent use of outcome *measures*, especially of patient-reported outcome measures (PROMs), and the lack of adequate development and content validity for their corresponding indication, further complicate matters (Chiarotto et al., 2018). Establishing core outcome sets (COSs) aims to harmonize outcome assessment (Kirkham et al., 2017; Williamson et al., 2017). The process of defining COS of outcome domains (*what* to measure) and subsequently of outcome measures for the respective domains (*how* to measure) is guided by systematic literature reviews, evidence synthesis and finally, a multi-stakeholder consensus process. Appropriate psychometric properties of PROMs are required for inclusion in future COS (Prinsen et al., 2016; Williamson et al., 2017). Following the consensus-based standards for the selection of health measurement instruments (COSMIN) initiative, these are validity, reliability, sensitivity and, essentially, content validity (sufficient assessment of the construct of interest). If content validity cannot be confirmed, other psychometric properties should not be evaluated (Mokkink et al., 2010; Patrick et al., 2011; Prinsen et al., 2018; Terwee et al., 2018).

We recently started a core outcome measures in effectiveness trials (COMET) initiative-guided development process of a COS of PROMs for perioperative management of acute postsurgical pain for several types of surgery including TKA within the Innovative Medicines Initiative's (IMI) PainCare subproject PROMPT (PROMs to improve management of acute and chronic pain) (Kaiser et al., 2020). Facing the heterogeneous and inconsistent outcome assessment in perioperative pain trials related to TKA among others (Bigalke et al., 2021), an international and interdisciplinary consensus panel recommended five core outcome domains, including *physical function* (PF; with complete agreement by all stakeholder groups and suggestions to consider surgery-specific PROMs), *pain intensity*, *self-efficacy* and *adverse events* (Pogatzki-Zahn et al., 2021).

The *primary aim* of this investigation was to systematically identify and analyse PROMs for the domain of *PF* used in clinical studies evaluating effectiveness of pain management post-TKA. *The second step* was to evaluate psychometric properties of the retrieved PROMs and the quality of their developmental process, including content validity, based on the respective inauguration articles following COSMIN guidelines.

2 | METHODS

The study protocol was designed and registered in advance (PROSPERO: CRD42020148012). For reporting, we followed the preferred reporting items for systematic reviews (PRISMA) statement (Liberati et al., 2009).

The IMI PainCare PROMPT COS Initiative is registered via COMET database (https://www.comet-initiative.org/Studies/Details/1731).

2.1 | Inclusion and exclusion criteria

Eligibility criteria were predefined according to the PICOS scheme (Table S1) and documented in the protocol. We considered prospective randomized controlled or observational trials, including at least 20 adult participants who had undergone TKA surgery on one or both knees. In the case of mixed surgery samples (e.g. TKA and total hip replacement), studies were only included if

a subgroup analysis of the TKA data was performed. The studied interventions had to be pain related as indicated by postsurgical pain intensity used as a primary or secondary outcome. Furthermore, the intervention had to be part of the surgical procedure or an additional medicinal, physiotherapeutic, psychological or any other treatment. Besides *pain intensity*, *PF* had to be measured as an outcome within the first 2 weeks after surgery.

Because the definition of *PF* includes the 'ability' to carry out activities, we searched not only for PROMs but also for 'performance outcome measures' (PerfOMs) and 'clinician reported outcome measures' (ClinROMs).

2.2 | Information source and search strategy

Studies were identified by searching EMBASE (via Ovid), MEDLINE (via PubMed) and CENTRAL databases. The search was conducted without date limitation and run in April 2022 (an updated search from December 2023 is shown in Figure S1; Appendix S1). Initially, we created a search string for application on PubMed (Methods S1) and adapted it for use in EMBASE and CENTRAL. Search terms included MeSH terms (if applicable) and free-text terms regarding pain, TKA, *PF*, study design and language.

2.3 | Study selection

After removing duplicates, predefined eligibility criteria (PICOS; Table S1) were applied independently by four reviewers (DCR, HH, KS and SB) to all potentially relevant titles and abstracts. After title–abstract screening, the remaining studies were screened as full texts by three independent reviewers (DCR, DH and HH). Any conflicts were adjudicated by all reviewers or under supervision of EPZ in case of remaining disagreement. We did not appraise the methodological quality of trials because our focus was on identifying common instruments measuring *PF* in clinical trials and observational studies after TKA, rather than assessing the efficacy of interventions.

2.4 | Data extraction

Data extraction was conducted by the same reviewers (DCR, DH and HH). The employed data extraction sheet was pilot tested on 10% of the included studies and adapted accordingly. Data extracted referred to (i) general study characteristics, (ii) information about the pain-related intervention and (iii) the outcome assessment concerning *PF* (Table S2). To complete the list of potential instruments for the IMI COS, we additionally extracted the time period of PROM assessment distinguishing between 'within 2 weeks after surgery' or later.

2.5 | Data synthesis

2.5.1 | Patient-reported outcome measures

For each PROM identified in the reviewed trials, we subsequently examined their relevance using the following three steps (Figure S2):

Step 1: Identification of relevant PROMs

All articles were hand searched as full texts and the following eligibility criteria were applied: First, the instrument had to contain at least one item which *corresponded to the FDA definition of 'patient-reported'* (U.S. Department of Health et al., 2006). Instruments using only objective or clinician-reported measurement methods were excluded. Second, the instrument had to contain at least one item which assessed *PF*. Instruments assessing related constructs (e.g. *pain interference* and *fatigue*) were excluded. Third, all items assessing *PF* had to do this site specifically for the lower extremities. In case there was any ambiguity regarding these criteria, instruments were initially included. Eligibility was further discussed during the next steps.

Step 2: Identification of development or inauguration articles for identified PROMs

For all PROMs passing step 1, we hand-searched articles about the PROM's development process for further data extraction and in order to evaluate content validity (see Step 3 and Figure S2). For each article, a further extraction sheet was developed (incl. pilot testing and adaption) (Table S3). The relevant data for extraction included general study characteristics, as well as detailed information regarding the assessed construct and population in which PROMs were validated.

Step 3: Evaluation of psychometric properties of the included PROMs

First, we evaluated if the population in which each PROM had been developed matched our population of interest. The aim was to examine content validity for patients in the acute postoperative stage after TKA of each PROM based on the inauguration or development article. Second, to address the aspect of construct, we compared the description of *PF* used in the PROMs' inauguration studies with the definition agreed upon by the IMI PainCare PROMPT COS steering committee. Finally, we investigated, according to

the COSMIN checklists, the quality of the development process by using COSMIN Risk of Bias Checklist—Box 1 and, if applicable due to a good or at least decent quality of the development process of the PROM, content validity of relevant PROMs using the user manual (v1.0) 'COSMIN methodology for assessing the content validity of PROMS' (accessible: https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf) (Terwee et al., 2018).

By answering the following questions, we identified indications of the development process of the PROM and content validity related to the indication of the present aim (acute pain after TKA) (Reeve et al., 2013; Terwee et al., 2018), when development or inauguration articles were available for a single PROM. First, the target population of the PROM was assessed (questions were as follows: What kind of patients are addressed by the PROM? Do they match with our population of interest regarding diagnosis (TKA) and clinical timeframe of 2 weeks after surgery?). Second, data on participants of PROM development studies were extracted and compared with the future target population of the COS application, which are patients with acute pain, receiving perioperative pain management within the first 2 weeks after surgery (TKA). Here, a construct of the PROM (conceptual framework and/or conceptual model) was evaluated (questions were as follows: Is the underlying concept and/or definition of the PROM explained? Does the description of *PF* for the PROM development (if applicable) conceptually match with the PROMIS definition chosen by the IMI PainCare PROMPT COS steering committee?). The reporting of an explicitly described conceptual model, framework or definition was extracted and, if available, compared with the definition of IMI PROMPT for *PF* (see above). Third, the study quality of the PROM development or inauguration article was evaluated by using the COSMIN Risk of Bias Checklist—Box 1 guided the evaluation of the developmental process if sufficiently described (Terwee et al., 2018). Evaluation was performed by the 4-point Likert scale (*very good*, *adequate*, *doubtful*, *inadequate* and *N/A*), considering 35 different items in 6 sub-scores. The overall study quality was summarized by the lowest rating of any of the items in the entire box ('worst score counts' method). Rating was provided independently by two raters (DH and HH) and adjudicated afterwards. Only if the development process for a PROM was rated as very good or adequate, the content validity was assessed by using the COSMIN Checklist Box 2 and further psychometric property assessment (Figure S2).

Ultimately, only PROMs were considered eligible for the IMI COS to assess *PF* after TKA that corresponded to general eligibility criteria (step 1) and matched quality criteria for PROM development and content validity.

2.5.2 | ClinROMs and PerfOMs

For further information about clinician-reported and performance outcomes, we extracted additional data in terms of indicators of *PF* (i.e. active/passive range of motion, muscle strength and self-care), measured by a ClinROM or PerfOMs. These findings in combination provide a more comprehensive overview of the general way of assessing *PF* in a postoperative setting. Thus, we were able to compare the use of PROMs and ClinROMs/PerfOMs in relation to the year of publication, the country of study conduct, the type of pain-related intervention and the time point of outcome assessment after surgery.

3 | RESULTS

3.1 | Study selection

The database search yielded 5408 potentially eligible records, of which 2896 remained after duplicate removal. After screening of titles and abstracts, 736 full-text articles were assessed for eligibility. The main reason for article exclusion was missing *PF* assessment within 14 days after surgery ($n = 73$ of 257 excluded articles). A list of excluded studies based on full-text screening is available from the corresponding author. Finally, 479 full-text articles fulfilled predefined eligibility criteria and were included for extraction and qualitative synthesis (Figure 1).

3.2 | Study characteristics

Of these 479 trials, 455 (95%) employed a randomized controlled design, whereas 24/479 (5%) were prospective observational studies. A total number of 44,440 patients participated in these studies (range: 20–600). Most of the studies were conducted in Asia (219/479; 46%), followed by Europe (130/479; 27%), North America (110/479; 23%), Australia (12/479; 3%), Africa (4/479; <1%) and South America (1/479; <1%). Three of 479 trials (<1%) were performed across different continents ('worldwide'). The date of publication ranged from 1983 to 2022: 396/479 trials (82%) were published after 2010, 67/479 studies (14%) were published between 2000 and 2009, 19/479 studies (4%) between 1990 and 1999 and 2/479 studies (<1%) between 1980 and 1989. In most cases, the pain-related intervention was pharmacological, including regional analgesia techniques (301/479; 63%), whereas surgical and physiotherapeutic interventions were studied in 82/479 (17%) and 55/479 (11%) studies respectively (Table 1; for more detailed information about the extracted studies' individual characteristics, see Table S4).

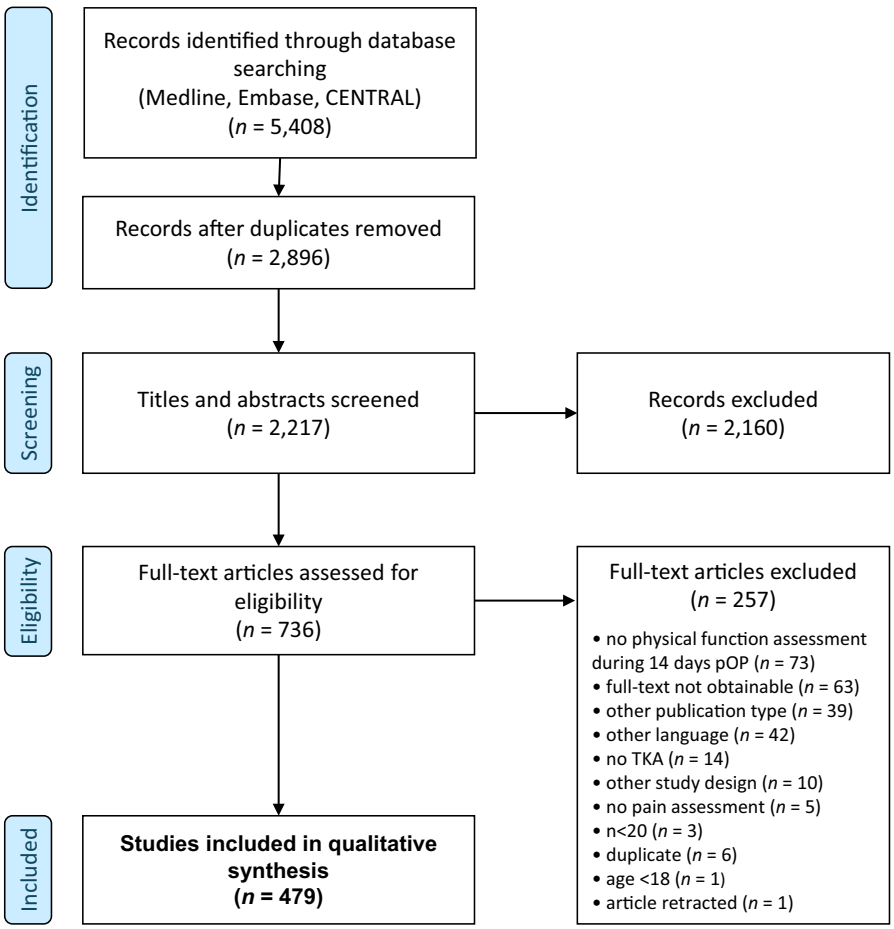

FIGURE 1 PRISMA flow chart. Abbreviations: pOP, postoperative; PRISMA, preferred reporting items for systematic reviews; TKA, total knee arthroplasty.

In 87/479 studies (18% of all studies), *PF* was assessed by PROMs. However, most of these studies additionally used other outcome assessment methods. In 9/479 studies (2%), *PF* was assessed exclusively by PROMs, compared to 78/479 studies (16%) applying both PROMs and PerfOMs/ClinROMs. In 392/479 studies (82%), the authors assessed the domain only by PerfOM and/or CLinROMs but not by PROMs (Table 2). For an overview of the used instruments in all included studies individually, see Table S5.

3.3 | Patient-reported outcome measures (PROMs)

3.3.1 | Step 1: Identification of relevant PROMs

In total, 11 different PROMs fulfilled the criteria of being at least partly patient reported and assessing site-specific *PF* for the lower extremities within 2 weeks after surgery (Figure 2a). These PROMs were used 97 times in 87 studies. The most frequently applied PROM was the 'Knee Scoring System' (KSS, 29/87; 33%) followed by the 'Western Ontario and McMaster Universities Osteoarthritis Index' (WOMAC, 23/87; 26%), the 'Hospital for Special Surgery Scoring System' (HSS, 17/87; 20%), the 'Knee Injury and Osteoarthritis Outcome Score' (KOOS, 11/87; 13%), the 'Oxford Knee Score' (OKS, 7/87; 8%), the 'Knee Society Clinical Rating System' (KSCRS, 4/87; 5%), the 'Knee Injury and Osteoarthritis Outcome Score—Joint Replacement (KOOS-JR, 2/87; 2%)', the 'Knee Injury and Osteoarthritis Outcome Score—Physical Function Short Form' (KOOS-PS, 1/87; 1%), the Lysholm Score (LS, 1/87; 1%), the Knee Outcome Survey (KOS, 1/87; 1%) and the 'Self-Reported Barthel Index' (1/87; 1%). We identified three additional PROMs which were used later than 14 days after surgery ('Lequesne Index' [LI], 'Lower Extremity Function Scale' [LEFS] and 'British Orthopedic Association Knee Function Assessment Chart' [BOA]). Six of the included 14 outcome measures were a combination of patient and clinician reported/performance based (BOA, HSS, KSCRS, KSS, LI and LS).

3.3.2 | Step 2: Identification of development or inauguration articles for identified PROMs

For 13 of the 14 identified PROMs, we found corresponding inauguration or development articles (Table 3). We

TABLE 1 Study characteristics.

Variable	*n* (%)
Included studies	479 (100)
Study design	
Randomized controlled trials	455 (95)
Prospective observational studies	24 (5)
Total number of participants	44,440 (range: 20 to 600)
Continent of study implementation	
Asia	219 (46)
Europe	130 (27)
North America	110 (23)
Australia	12 (3)
Afrika	4 (<1)
South America	1 (<1)
Worldwide	3 (<1)
Year of publication	
2020–2022	98 (20)
2010–2019	298 (62)
2000–2009	67 (14)
1990–1999	19 (4)
1980–1989	2 (<1)
Pain-related intervention	
Pharmacological/regional analgesia	301 (63)
Surgical	82 (17)
Physiotherapy	55 (11)
Psychological	3 (<1)
Others	38 (8)

did not find any appropriate article for the 'Self-Reporting Barthel Index'.

3.3.3 | Step 3: Evaluation of psychometric properties of the included PROMs

Target population of the PROM

The hand-searched inauguration and development articles showed that, based on the initial development, none of these PROMs corresponded with the exact scope of our interest. Six PROMs were developed on patients after TKA (Aichroth et al., 1978; Dawson et al., 1998; Insall et al., 1989; Lyman et al., 2016; Noble et al., 2012; Ranawatt et al., 1976; Ranawatt & Shine, 1973), whereas all other studies included patients after different surgeries (Lysholm & Gillquist, 1982; Roos et al., 1998) or with none-surgical disorders (Bellamy & Buchanan, 1986; Binkley et al., 1999; Irrgang et al., 1998; Lequesne, 1997; Lequesne et al., 1987; Perruccio et al., 2008). In case of included postsurgical patients, the assessments during development studies did not take place within 2 weeks after surgery (Noble et al., 2012; Roos et al., 1998) or the authors did not provide information about the time point of assessment (Aichroth et al., 1978; Dawson et al., 1998; Insall et al., 1989; Irrgang et al., 1998; Lyman et al., 2016; Lysholm & Gillquist, 1982; Ranawatt et al., 1976; Ranawatt & Shine, 1973) (Table 3).

Construct of the PROM (conceptual framework and/or conceptual model)

In all but one case, no information about the definition of *PF* underlying the PROM development was provided. Only Binkley et al. (Binkley et al., 1999) reported the WHO definition as basis for the development of the LEFS. It is, however, unclear to which extent this corresponds with the definition of *PF* recommended by PROMIS (see Discussion).

Study quality of PROM developmental or inauguration article

The risk of bias evaluation of the PROM development process using the COSMIN risk of bias Checklist—Box 1 showed, also based on the initial development, a lack of high-quality methodology. All PROM developments were rated as inadequate related to their general design requirements, especially—as mentioned above (Step 3b)—because almost none of the authors provided information about the construct to be measured. And even if a development study was performed, which was the case for six PROMs (Bellamy & Buchanan, 1986; Binkley et al., 1999; Dawson et al., 1998; Lyman et al., 2016; Noble et al., 2012; Perruccio et al., 2008), the applied method was rated at least 'doubtful', implying a lack in concept elicitation. Finally, the authors of only two studies conducted pilot testing or cognitive interviews to evaluate comprehensiveness and/or comprehensibility of the PROM (Bellamy & Buchanan, 1986; Dawson et al., 1998). Detailed information is shown in Table 4. Since basic criteria for content validity and general PROM design requirements were not fulfilled, further evaluation of content validity and additional psychometric properties (Steps 3d and 3e in the COSMIN methodology) were not indicated, as per COSMIN instructions (Terwee et al., 2018).

3.4 | PerfOM/ClinROM

We identified the following eight indicators measuring *PF* by a PerfOM or ClinROM (Figure 2b): *knee range of motion* (ROM, 371/470; 79%), *walking and ambulation* (W&A, 211/470; 45%), *muscle strength* (MS, 206/475; 44%),

TABLE 2 Study characteristics of clinical trials in regard to PROMs and PerfOMs/ClinROMs used for comparative effectiveness research of postoperative pain management in patients after TKA.

		Trials using PROMs $n = 87$ (18)		Trials NOT using PROMs
	Overall (n (%)) $n = 479$	PROMs only (n (%)) $n = 9$ (2)	PROMs + ClinROMs/ PerfOMs (n (%)) $n = 78$ (16)	ClinROMs/PerfOMs only (n (%)) $n = 392$ (82)
Continent				
Europe	130 (27)	6 (67)	23 (29)	101 (26)
North America	110 (23)	1 (11)	19 (24)	90 (23)
South America	1 (<1)	0	0	1 (<1)
Asia	219 (46)	1 (11)	34 (44)	183 (47)
Australia	12 (3)	0	2 (3)	10 (3)
Africa	4 (<1)	1 (11)	0	3 (<1)
Worldwide	3 (1)	0	0	3 (<1)
Year of publication				
1980–1989	2 (<1)	0	0	2 (<1)
1990–1999	19 (4)	0	2 (3)	17 (4)
2000–2009	67 (14)	0	6 (8)	61 (16)
2010–2019	298 (62)	6 (67)	48 (62)	244 (62)
2020–2022	98 (20)	3 (33)	22 (28)	73 (19)
Pain-related intervention				
Pharmacological/regional analgesia	301 (63)	3 (33)	29 (37)	271 (69)
Surgical	82 (17)	2 (22)	21 (27)	61 (16)
Physiotherapy	55 (11)	1 (11)	19 (24)	35 (9)
Psychological	3 (<1)	0	1 (1)	2 (<1)
Other	38 (8)	3 (33)	8 (10)	28 (7)

Abbreviations: ClinROMs, clinician reported outcome measures; PerfOMs, performance outcome measures; PROMs, patient reported outcome measures; TKA, total knee arthroplasty.

transfer (e.g. from sitting to standing) (Tr, 96/470; 20%), walking *stairs* (St, 82/470; 17%), *use walking aids* (WA, 53/470; 11%), *performing activities of daily living* (ADL, 20/470; 4%) and *balance issues* (BAL, 9/470; 2%). Due to our focus on PROMs in this article, a detailed analysis of the PerfOMs and ClinROMs will be reported separately.

4 | DISCUSSION

We aimed to investigate PROMs assessing *PF* in effectiveness studies on managing acute pain post-TKA and to investigate their psychometric properties guided by COSMIN. Our main result, based on 479 included studies, shows an unexpectedly rare implementation of PROMs to assess *PF* (only 18% of studies; Figure 2) and a lack of content validity in relation to our patient population of interest.

Although the necessity of PROMs in clinical trials is widely recognized (Mercieca-Bebber et al., 2018; Mokkink et al., 2010), our results indicate that *PF* in clinical studies on acute postoperative pain improvement is primarily assessed by PerfOMs and ClinROMs but not by PROMs. While only 87 trials assessed *PF* using PROMs, 392 trials used only PerfOMs and/or ClinROMs. Applying different forms of measurement can make sense as they evaluate different aspects of *PF* (Bean et al., 2011; Luna et al., 2017; Nielsen et al., 2016). Yet, neglecting PROMs means a lack of patient perspectives in treatment evaluations. While we have observed a clear increase in PROM implementation in this field since 2010, it still needs acceleration.

Our review underscores the need for identification and use of appropriate PROMs, including their harmonization, particularly after recognizing *PF* assessed by PROMs as a core domain for perioperative pain trials (Pogatzki-Zahn et al., 2021). After COSMIN, for content validity, qualitative research methods need to ensure *comprehensiveness* and *comprehensibility* of the pilot instrument, involving representatives of the target population and considering the specific research and/or clinical context (Patrick et al., 2011; Prinsen et al., 2016; Staniszewska & Haywood, 2012; Terwee et al., 2018; Weldring &

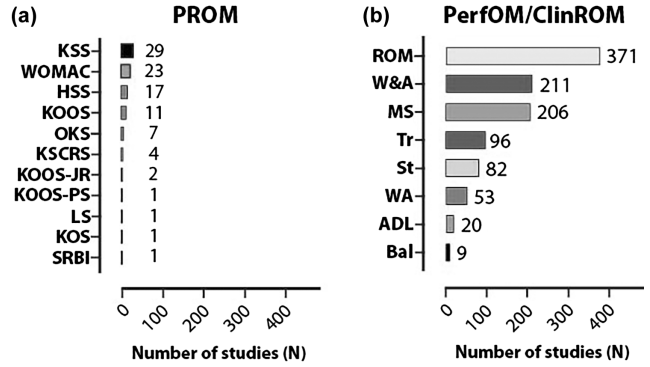

FIGURE 2 Frequency of applied PROMs (a) versus PerfO/ClinRO indicators measured by PerfOM/ClinROM (b) within 2 weeks after surgery. Abbreviations: ADL, activities of daily living; Bal, balance; ClinROM, clinician reported outcome measure; HSS, Hospital for Special Surgery Scoring System; KOOS, Knee Injury and Osteoarthritis Outcome Score; KOOS-JR, Knee Injury and Osteoarthritis Outcome Score—Joint Replacement; KOOS-PS, Knee Injury and Osteoarthritis Outcome Score—Physical Function Short Form; KOS, Knee Outcome Survey; KSCRS, Knee Society Clinical Rating System; KSS, Knee Scoring System; LS, Lysholm Score; MS, muscle strength; OKS, Oxford Knee Score; PerfOM, performance outcome measure; PROM, patient reported outcome measure; ROM, range of motion; SRBI, Self-Reporting Barthel Index; St, stairs; Tr, transfer; W&A, walking and ambulation; WA, walking aids; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

Smith, 2013). Our stepwise approach reveals that none of the identified PROMs was developed for and in our specific target population. Application of the COSMIN risk of bias—Box 1 for each identified PROM revealed insufficient development processes for all PROMs, for reasons such as lack of a conceptual model or lack of patient involvement. Per COSMIN, evaluating other psychometric properties only makes sense if good content validity is evident for the specific indication in the development process (Prinsen et al., 2018). PROMs with unclear content validity should not be considered for COS (Terwee et al., 2018), thus, further evaluation of the PROMs' quality was obsolete.

PROM development guidelines demand a detailed presentation of the underlying construct (PRO) through a conceptual model or framework (Terwee et al., 2018). In our review, only one development paper (LEFS; Binkley et al., 1999) provided a conceptual framework, based on the World Health Organization's *model of disability and handicap*. However, as the shift from ICIDH (*International Classification of Impairment, Disabilities and Handicaps*) to ICF (*International Classification of Functioning, Disability and Health*) took place in 2001 (WHO, 2002), it can be assumed that the conceptual framework of LEFS corresponds to *dysfunction* (or *impairment*), rather than *function*. All other PROMs lack information about the measured construct's origin or a conceptual framework. However, their individual items suggest they also address *disability* rather than *capability* (e.g. using questions focusing on what someone 'can NOT do'). As none of the developers provide sufficient information about their construct, it remains unclear whether any of these 10 PROMs properly assess *PF* or different constructs like *physical dysfunction* or *pain interference*.

The methodological quality of the PROM developmental studies was mainly limited by the absence of appropriate pilot studies. *Comprehensibility* was only evaluated for the WOMAC's development (Bellamy & Buchanan, 1986), and *comprehensiveness* for the OKS (Dawson et al., 1998). No other development process described the evaluation of any quality criteria. In none of the development studies, both core elements to ensure content validity, *comprehensibility* (i.e. items reflect patients' perspectives) and *comprehensiveness* (i.e. easy to understand), were assessed (Patrick et al., 2011; Terwee et al., 2018), thus measurement and interpretation of results obtained by all PROMs are insufficient.

Beyond acute postoperative pain, prior systematic reviews investigated psychometric properties of some identified PROMs (Alviar et al., 2011; Gagnier et al., 2017; Harris et al., 2016). All these authors concluded that there is no PROM for assessing *PF* with good psychometric properties, including content validity. The WOMAC and the OKS seem the most investigated PROMs for site-specific *PF*. Two reviews evaluated the OKS, attesting to *fair* content validity for long-term *PF* after TKA (Alviar et al., 2011; Gagnier et al., 2017). They cited Xie et al., who investigated content validity during cross-cultural adaption of the Singapore English and Chinese versions of the OKS (Xie et al., 2006). However, this study's process focusing on item *comprehensiveness* and *comprehensibility* did not fully meet current guidelines (Terwee et al., 2018). A similar approach indicated the KOOS's good content validity (Alviar et al., 2011). Another review, including the OKS developers, reported good evidence for content validity for the OKS and limited evidence for the KOOS, LEFS and WOMAC (Harris et al., 2016). However, the authors did not provide information on the criteria used for evaluation, making these ratings difficult to replicate. The new guidelines for rating content validity (Terwee et al., 2018) disqualify both, OKS and KOOS, possibly explaining discrepancies between our and earlier assessments using older tools (Terwee et al., 2007, 2012).

Our review's limitations mainly refer to our search strategy, where we deviated from recommendations by focusing on clinical trials to identify potential PROMs for *PF* (Prinsen et al., 2016). We may have missed PROMs specifically developed for the assessment of *PF* in the

TABLE 3 Summary of developmental or inauguration references of the identified 'patient reported outcome measures' (PROMs).

Outcome measures	Inauguration or development article	Construct being measured[a]	Study population (condition/diagnosis)	Number of subscales[b] (number of items)	Range total score; Response range	Time lag after surgery //Recall period
PROMs						
LEFS—Lower Extremity Function Scale	Binkley et al. (1999)	'The World Health Organization's model of disability served as the basis for the item generation phase of the scale development'	Lower extremity orthopaedic conditions	Single-dimension/20 items for lower-extremity function	0–80 (80 = high functional level); 0–4 (extreme difficulty to perform activity—no difficulty to perform activity)	n/a (no surgery) //'today'
KOOS—Knee Injury and Osteoarthritis Outcome Score	Roos et al. (1998)	Not reported (pain; symptoms, e.g. swelling and restricted ROM; ADL; sport and recreation function; KRQOL)	Posttraumatic/ postoperative knee OA ('young and middle-aged subjects with ACL injury, meniscus injury, or posttraumatic OA')	5 subscales/42 items: Pain (9); Symptoms (7); ADL function (17); Sports/recreation (5); KRQOL (4)	0–100 (0 = extreme knee problems − 100 = no knee problems); 0–4 in levels of severity or frequency (converted in per cent scale)	>6 weeks (POM 3, 6 and 12) //1 week
KOOS-JR—'Knee Injury and Osteoarthritis Outcome Score'—Joint Replacement	Lyman et al. (2016)	Not reported ('knee health' with aspects of pain, symptom severity and ADL)	End-stage knee OA, undergoing TKA	Single-dimension/7 items for 'knee health'	0–100 (0 = total knee disability − 100 = perfect knee health); 0–4 in levels of severity or frequency (converted to an interval score)	n/a (no surgery) //not reported
KOOS-PS—'Knee Injury and Osteoarthritis Outcome Score'—Physical Function Short-Form	Perruccio et al. (2008)	Not reported ('to develop physical function states that represent the progression of physical disability from early to late disease for individuals with OA of the knee')	Knee OA	Single-dimension/7 items for function DAL and 'higher level activities'	0–28; 0–4 in levels of severity (0 = no difficulty with activity, 4 = extreme difficulty)	n/a (no surgery) //1 week
OS—Knee Outcome Survey	Irrgang et al. (1998)	Not reported ('functional limitations that result from a wide variety of pathological disorders and impairments of the knee')	Various knee disorders	Two separate (sub)scales/17 items: ADL Scale/Symptoms(7), Sports Activity Scale/Functional disability (10)	0–80 (final score expressed as percentage); 0–5 (5 = being not affected at all)	n/a //1–2 days
OKS—Oxford Knee Score	Dawson et al. (1998)	Not reported	TKA due to OA	Single-dimension/12 items	12–60; 1–5 (least to most difficulty/severity)	Not reported //4 weeks
WOMAC—Western Ontario and McMaster Universities Osteoarthritis Index	Bellamy and Buchanan (1986)	Not reported	Knee and hip OA	3 subscales/24 items: pain (5), stiffness (2) and difficulty DAL (17)	0–96; 0–4 (discomfort or disability from 0 = none to 4 = extremely)	n/a (no surgery) //not reported

(Continues)

TABLE 3 (Continued)

Outcome measures	Inauguration or development article	Construct being measured[a]	Study population (condition/diagnosis)	Number of subscales[b] (number of items)	Range total score; Response range	Time lag after surgery //Recall period
SRBI—Self-Report Barthel Index	None found	-	-	-	-	-
Hybrids (additional PerfOM/ClinROM Items)						
BOA—British Orthopaedic Association Knee Function Assessment Chart	Aichroth et al. (1978)	Not reported	TKA/knee OA	11 subscales/12 items: Satisfaction, pain, ability to walk, walking aids, gait, flexion deformity, maximal flexion, extension lag, valgus angle, varus angle, arising from chair and stair climbing	12–55; 0–4, 1–4, 0–5 or 1–5 (higher scores for more function/less pain/less severity)	Not reported ('follow up months') //not reported
HSS—Hospital for Special Surgery Score	Ranawatt and Shine (1973); Ranawatt et al. (1976)	Not reported	TKA	7 subscales/12 items: pain (30 pts), function (22 pts) range of motion (18 pts), muscle strength (10 pts), deformity (10 pts), instability (10 pts) and subtractions (assistive device, extension lag, valgus/varus deformity)	0–100 (excellent: 85–100, good: 70–84, fair: 60–69 and poor <60); 0 to 5–15 pts per item scale, subtractions of 1–5 pts for specific items	>6 weeks (POM 6, 12 and up to 4 years postop) //not reported
KSCRS—Knee Society Clinical Rating System	Noble et al. (2012)	Not reported	TKA	4 Subscales/34 items: Objective Knee Score/7 items: 4 items physician assessed: alignment, instability and joint motion; 3 items patient assessed: symptoms; satisfaction score/5 items (patient assessed): pain and knee function expectation score/3 items (patient assessed): pain relief, ADL and sports/recreation; functional activity score/19 items (patient assessed): walking and standing (5), standard activities (6), advances activities (5) and discretionary activities (3)	Objective Knee Score 100 points (10–25 pts with optional subtraction of 2–15 pts), Satisfaction Score 40 points (each 8 pts), expectation score 15 points (5 pts each), functional activity score 100 points (5–6 pts each) (the higher the better the outcome)	>6 weeks (POM 12) //not reported
KSS—Knee Society Score	Insall et al. (1989)	Not reported ('knee rating and functional assessment')	TKA	2 subscales/10 items: Knee rating/8 items (physician assessed): pain (50 pts), ROM (25 pts) and stability (25 pts); deductions for flexion contractures, extension lag and misalignment Function score/2 items (patient assessed): mobility walking distance (50 pts) and stair climbing (50 pts); deduction for walking aids[c]	0–100 for both scores (higher scores better the outcome: 100 = well-aligned knee with no pain and instability, sufficient ROM/100 = pat. can walk unlimited distance and climb stairs normally); 0 to 10–50 pts per item scale, deductions of 2–20 pts for specific items	Not reported //not reported

TABLE 3 (Continued)

Outcome measures	Inauguration or development article	Construct being measured[a]	Study population (condition/diagnosis)	Number of subscales[b] (number of items)	Range total score; Response range	Time lag after surgery //Recall period
LI—Lequesne Index	Lequesne et al. (1987)	Not reported (to evaluate drug efficacy 'in the short term and to follow up the course of the disease in the long term'; 'index of severity for OA of the knee')	Hip and knee OA	3 subscales/10 items: Pain/discomfort (5), maximum distance walked (1) with additions for walking aids[c] and ADL (4)	0–24 (sum score for severity of handicap: 1–4 = mild, 5–7 = moderate, 8–10 = severe, 11–13 = very severe and >14 = extremely severe); Pain 0–2; ADL 0–2 (0 = easy, 0.5–1.5 = with difficulty, 2 = impossible); walking impairment 0–6, plus possible additions (1–2)	n/a (no surgery) //not reported
LS—Lysholm Score	Lysholm and Gillquist (1982)	Not reported ('a scoring scale for knee ligament surgery follow-up emphasizing evaluation of symptoms of instability. Instability is defined as "giving way" during activity')	Knee ligament surgeries	8 subscales/8 items ('filled with the patients' collaboration'): Limp (5 pts), need for support (5 pts), stair climbing (10 pts) and squatting (5 pts); 'walking, running and jumping (70 pts)'/3 items: instability (30 pts), pain (30 pts) and swelling (10 pts); Atrophy of thigh (5 pts)	0–100 (the higher the better); 0 to 5–30 pts per item scale	>6 weeks ('1–8 years after treatment') //not reported

Abbreviations: ADL, activities of daily living; KRQOL, knee-related quality of life; n/a, not applicable; OA, osteoarthritis; POM, postoperative month; pts, points; ROM, range of motion/movement.

[a]As described in the paper, if no clear construct was described, we are citing in () the closest description of the aim.

[b]Described in the papers as subscales, domains and dimensions.

[c]In some reviews, the use of walking aids is counted as an individual item.

TABLE 4 COSMIN Risk of Bias Checklist—Box 1 applied for each PROM.

Items – COSMIN risk of bias checklist for PROM development	KOOS	KOOS-JR	KOOS-PS	KOS	KSCRS	LEFS	Lequesne	LS	OKS	WOMAC	KSS	HSS	BOA
1a. PROM design													
General Design requirements													
1. clear construct description	🔴	🔴	🔴	🔴	🔴	🟢	🔴	🔴	🔴	🔴	🔴	🔴	🔴
2. origin of construct clear	🟡	🟡	🟡	🟡	🟡	🟢	🔴	🟡	🔴	🔴	🟡	🟡	🟡
3. clear description of target population	🟢	🟢	🟢	🟢	🟢	🟢	🔴	🟢	🟢	🔴	🔴	🔴	🔴
4. description of context of use	🟢	🟡	🟢	🟡	🟢	🟢	🟢	🟡	🟡	🟢	🟡	🔴	🟢
5. PROM development study performed	🟡	🟢	🟢	🔴	🟢	🟢	🔴	🔴	🟢	🟢	🔴	🔴	🔴
Concept elicitation (relevance and comprehensiveness)													
6. appropriate qualitative data collection	🟡	🟡	🟡	-	🟡	🟡	🔴	🟡	🟢	🟡	-	-	-
7. skilled moderators/interviewers	🟡	n/a	n/a	-	n/a	n/a	n/a	n/a	🟡	🟢	-	-	-
8. appropriate interview guide	🟡	n/a	n/a	-	n/a	n/a	n/a	n/a	🟡	🟡	-	-	-
9. recording and transcribed verbatim	🟡	n/a	n/a	-	n/a	n/a	n/a	n/a	🟡	🟡	-	-	-
10. appropriate data analysis	🟡	n/a	🟢	-	🟡	🟡	🔴	n/a	🟡	🟡	-	-	-
11. at least part of data independently coding	🔴	n/a	n/a	-	n/a	🟡	🔴	n/a	🟡	🟡	-	-	-
12. data collection continued until saturation	🟡	n/a	n/a	-	n/a	🔴	🔴	n/a	🟢	🟡	-	-	-
13. sample size appropriate	🟢	🟡	🟢	-	🟢	🟡	n/a	🟢	n/a	🟢	-	-	-
SUBTOTAL QUALITY ELICITATION STUDY	🔴	🟡	🟡	n/a	🟡	🔴	🔴	🟡	🟡	🟡	-	-	-
TOTAL QUALITY OF THE PROM DESIGN	🔴	🔴	🔴	n/a	🔴	🔴	🔴	🔴	🔴	🔴	🔴	🔴	🔴
1b. Cognitive interview study or other pilot test													
14. cognitive interview or pilot testing conducted	🔴	🔴	🔴	🔴	🟢	🟢	🔴	🔴	🟢	🟢	🔴	🔴	🔴
General design requirements													
15. relevant target population	-	-	-	-	🟢	🟢	-	-	🟢	🟢	-	-	-
Comprehensibility													
16. patients asked about comprehensibility	-	-	-	-	🔴	🔴	-	-	🟡	🟢	-	-	-
17. all items tested in final form	-	-	-	-	-	-	-	-	-	🟡	-	-	-
18. appropriate qualitative method	-	-	-	-	-	-	-	-	-	🟡	-	-	-
19. appropriate number of patients	-	-	-	-	-	-	-	-	-	🟢	-	-	-
20. skilled interviewers	-	-	-	-	-	-	-	-	-	🟡	-	-	-
21. appropriate interview guide	-	-	-	-	-	-	-	-	-	🟡	-	-	-
22. recording and transcribed verbatim	-	-	-	-	-	-	-	-	-	🟡	-	-	-
23. appropriate data analysis	-	-	-	-	-	-	-	-	-	🟡	-	-	-
24. analysis by at least two researchers	-	-	-	-	-	-	-	-	-	🟡	-	-	-
25. PROM comprehensibility adaption	-	-	-	-	-	-	-	-	-	🟡	-	-	-
SUBTOTAL QUALITY OF COMPREHENSIBILITY STUDY	-	-	-	-	-	-	-	-	🟡	🟡	-	-	-
Comprehensiveness													
26. patients asked about comprehensiveness	-	-	-	-	🔴	🔴	-	-	🟢	🟡	-	-	-
27. final set of items tested	-	-	-	-	-	-	-	-	🟢	-	-	-	-
28. appropriate method	-	-	-	-	-	-	-	-	🟢	-	-	-	-
29. appropriate number of patients	-	-	-	-	-	-	-	-	🟢	-	-	-	-
30. skilled interviewers	-	-	-	-	-	-	-	-	n/a	-	-	-	-
31. appropriate interview guide	-	-	-	-	-	-	-	-	n/a	-	-	-	-
32. recording and transcribed verbatim	-	-	-	-	-	-	-	-	n/a	-	-	-	-
33. appropriate data analysis	-	-	-	-	-	-	-	-	🟡	-	-	-	-
34. analysis by at least two researchers	-	-	-	-	-	-	-	-	🟡	-	-	-	-
35. PROM comprehensiveness adaption	-	-	-	-	-	-	-	-	🟢	-	-	-	-
SUBTOTAL QUALITY OF COMPREHENSIVENESS STUDY	-	-	-	-	🔴	🔴	🔴	-	🟡	🟡	-	-	-
TOTAL QUALITY OF THE PILOT STUDY	🔴	🔴	🔴	🔴	🔴	🔴	🔴	🔴	🟡	🟡	🔴	🔴	🔴
TOTAL QUALITY OF THE PROM DEVELOPMENT STUDY	🔴	🔴	🔴	🔴	🔴	🔴	🔴	🔴	🔴	🔴	🔴	🔴	🔴

Note: 🟢, very good; 🟢, adequate; 🟡, doubtful; 🔴, inadequate; -, not investigated regarding COSMIN manual.

Abbreviations: BOA, British Orthopaedic Association Knee Function Assessment Chart; HSS, Hospital for Special Surgery Scoring System; KOOS, Knee Injury and Osteoarthritis Outcome Score; KOOS-JR, Knee Injury and Osteoarthritis Outcome Score—Joint Replacement; KOOS-PS, Knee Injury and Osteoarthritis Outcome Score—Physical Function Short-Form; KOS, Knee Outcome Survey; KSCRS, Knee Society Clinical Rating System; KSS, Knee Scoring System; LEFS, Lower Extremity Function Scale; LS, Lysholm Score; n/a, not applicable; OKS, Oxford Knee Score; POM, postoperative month; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

acute postoperative phase, especially those not yet used in clinical trials. Nonetheless, our main aim and search strategy demonstrate the current status of PROM usage in assessing *PF* in acute postoperative pain research, and we found no additional relevant instruments by hand search (e.g. in identified systematic reviews or an updated search). Another limitation is that inauguration or development publications were searched by hand for each PROM and not systematically. Finally, three PROMs (i.e. LI, LEFS and BOA) were included for the second step (rating of content validity), although used later than 14 days postoperatively, as we could not exclude that their development processes included the acute postsurgical phase.

Reflecting the limitations of PROMs for *PF*, we face the lack of an adequate PROM for *PF* to be included in the future IMI PROMPT COS of measures. Thus, the next step should be an ISPOR- and COSMIN-guided development of a suitable PROM for assessing self-reported *PF* in patients early after TKA (Basch et al., 2011; Patrick et al., 2011; Prinsen et al., 2016). During this process, constructs other than *PF*, such as *pain interference*, should be considered as possible alternatives to be implemented into a COS, especially because *pain interference* was recently included in a COS for acute pain in general by a large international consensus process (Bova et al., 2023). Both are possibly based on different constructs, thus, a thorough differentiation between supposedly related constructs such as *function*, *dysfunction* and *interference* is important. Another challenge will be promoting widespread use of future COS in clinical trials. *PF* was only rarely measured from patients' perspective in pain management effectiveness trials post-TKA, and applied PROMs date back up to five decades. Manifest barriers seem to hinder implementing new PROMs in comparative effectiveness research, possibly due to trial authors unaware of current outcome research and advances in this field. Furthermore, trial designs often default to employ frequently used tools for outcome assessment, regardless of their validity or appropriateness for specific scientific purposes.

Moreover, *PF* as a domain is also commonly unattended in half of the studies on perioperative pain management for patients after TKA (Bigalke et al., 2021). In contrast, the IMI PainCare COS Panel has recommended *PF* as a core domain for the future IMI PROMPT COS domains for perioperative management of acute pain after surgery (Pogatzki-Zahn et al., 2021). This implies that current evidence for best pain management after TKA is still hampered by missing *PF* as one of the major outcome domains. Thus, and because of the unsatisfactory situation of PROMs for *PF* illustrated by this review, the development of a PROM with profound psychometric properties is urgently needed, also for further establishment of transculturally adapted and translated PROMs (Beaton et al., 2000; Rupareliya & Shukla, 2020).

5 | CONCLUSION

Despite widespread recommendations, our data demonstrate the prevailing lack of patient-reported assessment of *PF* in context of clinical trials on acute pain management post-TKA. Notably, none of the currently applied PROMs was specifically developed for this target population, and methodological quality of PROM development, based on the inauguration or development articles, lacks comprehensibility and comprehensiveness. The overall limited quality of the PROMs for evaluating early postoperative pain-related *PF* after TKA does not allow to recommend any of the PROMS for a COS. Rather, the lack of conceptual framework for postsurgical *PF* leads to the urgent need for its establishment as a foundation for the development of a new PROM tailored to this specific target population to ensure a more comprehensive and patient-focused assessment of *PF* in future research and clinical practice.

CONFLICT OF INTEREST STATEMENT

HH, DH, KS, DCR, SB, TM, DH-S and UK: none. HL: HL is employee of Grünenthal GmbH. EMP-Z: received financial support from Grünenthal for research activities and advisory board and lecture fees from Grünenthal, Novartis and Medtronic. In addition, she receives scientific support from the German Research Foundation (DFG), the Federal Ministry of Education and Research (BMBF), the Federal Joint Committee (G-BA) and the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777500. This joint undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. All money went to the institution EMP-Z is working for.

IMI STATEMENT

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No (777500). This joint undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. Further information is found under: www.imi-paincare.eu and www.imi.europa.eu. The statements and opinions presented here reflect the author's view and neither IMI nor the European Union, EFPIA or any Associated Partners are responsible for any use that may be made of the information contained therein.

ACKNOWLEDGEMENTS

Open Access funding enabled and organized by Projekt DEAL.

ORCID

H. Heitkamp https://orcid.org/0000-0002-6336-6697
D. Heußner https://orcid.org/0000-0002-5540-8973
D. C. Rosenberger https://orcid.org/0000-0002-7079-6476
D. Hohenschurz-Schmidt https://orcid.org/0000-0002-1964-6069
E. M. Pogatzki-Zahn https://orcid.org/0000-0003-0981-3940

REFERENCES

Aichroth, P., Freeman, M. A. R., Smillie, I. S., & Souter, W. A. (1978). A knee function assessment chart. *The Journal of Bone and Joint Surgery*, *60*, 308.

Alviar, M. J., Olver, J., Brand, C., Tropea, J., Hale, T., Pirpiris, M., & Khan, F. (2011). Do patient-reported outcome measures in hip and knee arthroplasty rehabilitation have robust measurement attributes? A systematic review. *Journal of Rehabilitation Medicine*, *43*(7), 572–583. https://doi.org/10.2340/16501977-0828

Basch, E., Abernethy, A. P., & Reeve, B. B. (2011). Assuring the patient centeredness of patient-reported outcomes: Content validity in medical product development and comparative effectiveness research. *Value in Health*, *14*(8), 965–966. https://doi.org/10.1016/j.jval.2011.10.002

Bean, J. F., Olveczky, D. D., Kiely, D. K., LaRose, S. I., & Jette, A. M. (2011). Performance-based versus patient-reported physical function: What are the underlying predictors? *Physical Therapy*, *91*(12), 1804–1811. https://doi.org/10.2522/ptj.20100417

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Bellamy, N., & Buchanan, W. W. (1986). A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. *Clinical Rheumatology*, *5*(2), 231–241.

Bigalke, S., Maeßen, T. V., Schnabel, K., Kaiser, U., Segelcke, D., Meyer-Frießem, C. H., Liedgens, H., Macháček, P. A., Zahn, P. K., & Pogatzki-Zahn, E. M. (2021). Assessing outcome in postoperative pain trials: Are we missing the point? A systematic review of pain-related outcome domains reported in studies early after total knee arthroplasty. *Pain*, *162*(7), 1914–1934. https://doi.org/10.1097/j.pain.0000000000002209

Binkley, J. M., Stratford, P. W., Lott, S. A., & Riddle, D. L. (1999). The Lower Extremity Functional Scale (LEFS): Scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. *Physical Therapy*, *79*(4), 371–383. https://doi.org/10.1093/ptj/79.4.371

Bova, G., Domenichiello, A., Letzen, J. E., Rosenberger, D. C., Siddons, A., Kaiser, U., Anicich, A., Baron, R., Birch, J., Bouhassira, D., Casey, G., Golden, K., Iyengar, S., Karp, B. I., Liedgens, H., Meissner, W., Nicholson, K., Pogorzala, L., Ryan, D., … Pogatzki-Zahn, E. M. (2023). Developing consensus on core outcome sets of domains for acute, the transition from acute to chronic, recurrent/episodic, and chronic pain: Results of the INTEGRATE-pain Delphi process. *EClinicalMedicine*, *66*, 102340. https://doi.org/10.1016/j.eclinm.2023.102340

Brat, G. A., Agniel, D., Beam, A., Yorkgitis, B., Bicket, M., Homer, M., Fox, K. P., Knecht, D. B., McMahill-Walraven, C. N., Palmer, N., & Kohane, I. (2018). Postsurgical prescriptions for opioid naive patients and association with overdose and misuse: Retrospective cohort study. *British Medical Journal*, *360*, j5790. https://doi.org/10.1136/bmj.j5790

Chiarotto, A., Boers, M., Deyo, R. A., Buchbinder, R., Corbin, T. P., Costa, L. O. P., Foster, N. E., Grotle, M., Koes, B. W., Kovacs, F. M., Lin, C.-W. C., Maher, C. G., Pearson, A. M., Peul, W. C., Schoene, M. L., Turk, D. C., van Tulder Maurits, W., Terwee, C. B., & Ostelo, R. W. (2018). Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain*, *159*(3), 481–495. https://doi.org/10.1097/j.pain.0000000000001117

Dawson, J., Fitzpatrick, R., Murray, D., & Carr, A. (1998). Questionnaire on the perceptions of patients about total knee replacement. *The Journal of Bone & Joint Surgery British*, *80*(1), 63–69. https://doi.org/10.1302/0301-620X.80B1.7859

Gagnier, J. J., 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. *The Journal of Arthroplasty*, *32*(5), 1688–1697. e7. https://doi.org/10.1016/j.arth.2016.12.052

Harris, K., Dawson, J., Gibbons, E., Lim, C. R., Beard, D. J., Fitzpatrick, R., & Price, A. J. (2016). Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Related Outcome Measures*, *7*, 101–108. https://doi.org/10.2147/PROM.S97774

Insall, J. N., Dorr, L. D., Scott, R. D., & Scott, W. N. (1989). Rationale of the knee society clinical rating system. *Clinical Orthopaedics and Related Research*, *248*, 13–14.

Irrgang, J. J., Snyder-Mackler, L., Wainner, R. S., Fu, F. H., & Harner, C. D. (1998). Development of a patient-reported measure of function of the knee. *The Journal of Bone and Joint Surgery*, *80*(8), 1132–1145. https://doi.org/10.2106/00004623-199808000-00006

Kaiser, U., Liedgens, H., Meissner, W., Weinmann, C., Zahn, P., & Pogatzki-Zahn, E. (2020). Developing consensus on core outcome domains and measurement instruments for assessing effectiveness in perioperative pain management after sternotomy, breast cancer surgery, total knee arthroplasty, and surgery related to endometriosis. *Trials*, *21*(1), 773. https://doi.org/10.1186/s13063-020-04665-9

Kent, M. L., Hurley, R. W., Oderda, G. M., Gordon, D. B., Sun, E., Mythen, M., Miller, T. E., Shaw, A. D., Gan, T. J., Thacker, J. K. M., McEvoy, M. D., & POQI-4 Working Group. (2019). American Society for Enhanced Recovery and Perioperative Quality Initiative-4 joint consensus statement on persistent postoperative opioid use: Definition, incidence, risk factors, and health care system initiatives. *Anesthesia and Analgesia*, *129*(2), 543–552. https://doi.org/10.1213/ANE.0000000000003941

Kersting, C., Kneer, M., & Barzel, A. (2020). Patient-relevant outcomes: What are we talking about? A scoping review to improve conceptual clarity. *BMC Health Services Research*, *20*(1), 596. https://doi.org/10.1186/s12913-020-05442-9

Kirkham, J. J., Davis, K., Altman, D. G., Blazeby, J. M., Clarke, M., Tunis, S., & Williamson, P. R. (2017). Core outcome set-STAndards for development: The COS-STAD recommendations. *PLoS Medicine*, *14*(11), e1002447. https://doi.org/10.1371/journal.pmed.1002447

Lawal, O. D., Gold, J., Murthy, A., Ruchi, R., Bavry, E., Hume, A. L., Lewkowitz, A. K., Brothers, T., & Wen, X. (2020). Rate and risk factors associated with prolonged opioid use after surgery. *JAMA Network Open*, *3*(6), e207367. https://doi.org/10.1001/jamanetworkopen.2020.7367

Lequesne, M. G. (1997). The algofunctional indices for hip and knee osteoarthritis. *The Journal of Rheumatology*, *24*(4), 779–781.

Lequesne, M. G., Mery, C., Samson, M., & Gerard, P. (1987). Indexes of severity for osteoarthritis of the hip and knee. Validation—Value in comparison with other assessment tests. *Scandinavian Journal of Rheumatology. Supplement*, *65*, 85–89. https://doi.org/10.3109/03009748709102182

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, *6*(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

Luna, I. E., Kehlet, H., Peterson, B., Wede, H. R., Hoevsgaard, S. J., & Aasvang, E. K. (2017). Early patient-reported outcomes versus objective function after total hip and knee arthroplasty: A prospective cohort study. *The Bone & Joint Journal*, *99*(9), 1167–1175. https://doi.org/10.1302/0301-620X.99B9.BJJ-2016-1343.R1

Lyman, S., Lee, Y.-Y., Franklin, P. D., Li, W., Cross, M. B., & Padgett, D. E. (2016). Validation of the KOOS, JR: A short-form knee arthroplasty outcomes survey. *Clinical Orthopaedics and Related Research*, *474*(6), 1461–1471. https://doi.org/10.1007//s11999-016-4719-1

Lysholm, J., & Gillquist, J. (1982). Evaluation of knee ligament surgery results with special emphasis on use of a scoring scale. *The American Journal of Sports Medicine*, *10*(3), 150–154. https://doi.org/10.1177/036354658201000306

Mercieca-Bebber, R., King, M. T., Calvert, M. J., Stockler, M. R., & Friedlander, M. (2018). The importance of patient-reported outcomes in clinical trials and strategies for future optimization. *Patient Related Outcome Measures*, *9*, 353–367. https://doi.org/10.2147/PROM.S156279

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology*, *10*, 22. https://doi.org/10.1186/1471-2288-10-22

Nielsen, L. M., Kirkegaard, H., Østergaard, L. G., Bovbjerg, K., Breinholt, K., & Maribo, T. (2016). Comparison of self-reported and performance-based measures of functional ability in elderly patients in an emergency department: Implications for selection of clinical outcome measures. *BMC Geriatrics*, *16*(1), 199. https://doi.org/10.1186/s12877-016-0376-1

Noble, P. C., Scuderi, G. R., Brekke, A. C., Sikorskii, A., Benjamin, J. B., Lonner, J. H., Chadha, P., Daylamani, D. A., Scott, N. W., & Bourne, R. B. (2012). Development of a new Knee Society scoring system. *Clinical Orthopaedics and Related Research*, *470*(1), 20–32. https://doi.org/10.1007/s11999-011-2152-z

Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—Eliciting concepts for a new PRO instrument. *Value in Health*, *14*(8), 967–977. https://doi.org/10.1016/j.jval.2011.06.014

Perruccio, A. V., Stefan Lohmander, L., Canizares, M., Tennant, A., Hawker, G. A., Conaghan, P. G., Roos, E. M., Jordan, J. M., Maillefert, J. F., Dougados, M., & Davis, A. M. (2008). The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS)—An OARSI/OMERACT initiative. *Osteoarthritis and Cartilage*, *16*(5), 542–550. https://doi.org/10.1016/j.joca.2007.12.014

Pogatzki-Zahn, E. M., Liedgens, H., Hummelshoj, L., Meissner, W., Weinmann, C., Treede, R.-D., Vincent, K., Zahn, P., IMI-PainCare PROMPT Consensus Panel, & Kaiser, U. (2021). Developing consensus on core outcome domains for assessing effectiveness in perioperative pain management: Results of the PROMPT/IMI-PainCare Delphi Meeting. *Pain*, *162*(11), 2717–2736. https://doi.org/10.1097/j.pain.0000000000002254

Pogatzki-Zahn, E., Schnabel, K., & Kaiser, U. (2019). Patient-reported outcome measures for acute and chronic pain: Current knowledge and future directions. *Current Opinion in Anaesthesiology*, *32*(5), 616–622. https://doi.org/10.1097/ACO.0000000000000780

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1147–1157. https://doi.org/10.1007/s11136-018-1798-3

Prinsen, C. A. C., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., Williamson, P. R., & Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a "Core outcome set"—A practical guideline. *Trials*, *17*(1), 449. https://doi.org/10.1186/s13063-016-1555-2

Ranawatt, C. S., Insall, J., & Shine, J. (1976). Duo-condylar knee arthroplasty. Hospital for special surgery design. *Clinical Orthopaedics and Related Research*, *120*, 76–82.

Ranawatt, C. S., & Shine, J. J. (1973). Duo condylar total knee arthroplasty. *Clinical Orthopaedics and Related Research*, *94*, 185–195.

Reeve, B. B., Wyrwich, K. W., Wu, A. W., Velikova, G., Terwee, C. B., Snyder, C. F., Schwartz, C., Revicki, D. A., Moinpour, C. M., McLeod, L. D., Lyons, J. C., Lenderking, W. R., Hinds, P. S., Hays, R. D., Greenhalgh, J., Gershon, R., Feeny, D., Fayers, P. M., Cella, D., … Butt, Z. (2013). ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Quality of Life Research*, *22*(8), 1889–1905. https://doi.org/10.1007/s11136-012-0344-y

Roos, E. M., Roos, H. P., Lohmander, L. S., Ekdahl, C., & Beynnon, B. D. (1998). Knee Injury and Osteoarthritis Outcome Score (KOOS)—Development of a self-administered outcome measure.

The Journal of Orthopaedic and Sports Physical Therapy, *28*(2), 88–96. https://doi.org/10.2519/jospt.1998.28.2.88

Rosenberger, D. C., & Pogatzki-Zahn, E. M. (2022). Chronic postsurgical pain—Update on incidence, risk factors and preventive treatment options. *BJA Education*, *22*(5), 190–196. https://doi.org/10.1016/j.bjae.2021.11.008

Rupareliya, D. A., & Shukla, Y. (2020). Need for cross-cultural adaptation of self-reported health measures: Review study. *Indian Journal of Physiotherapy & Occupational Therapy*, *14*, 34–37. https://doi.org/10.37506/ijpot.v14i2.2597

Sitter, T., & Forget, P. (2021). Persistent postoperative opioid use in Europe: A systematic review. *European Journal of Anaesthesiology*, *38*(5), 505–511. https://doi.org/10.1097/EJA.0000000000001346

Staniszewska, S., Haywood, K. L., Brett, J., & Tutton, L. (2012). Patient an public involvement in patient-reported outcome measures. *The Patient-Patient-Centered Outcomes Research*, *5*(2), 79–87. https://doi.org/10.2165/11597150-000000000-00000

Terkawi, A. S., Mavridis, D., Sessler, D. I., Nunemaker, M. S., Doais, K. S., Terkawi, R. S., Terkawi, Y. S., Petropoulou, M., & Nemergut, E. C. (2017). Pain management modalities after total knee arthroplasty: A network meta-analysis of 170 randomized controlled trials. *Anesthesiology*, *126*(5), 923–937. https://doi.org/10.1097/ALN.0000000000001607

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, *21*(4), 651–657. https://doi.org/10.1007/s11136-011-9960-1

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., De Vet, H. C., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, *27*, 1159–1170. https://doi.org/10.1007/s11136-018-1829-0

U.S. Department of Health; Human Services FDA Center for Drug Evaluation; Research; Human Services FDA Center for Biologics Evaluation; Human Services FDA Center for Devices; Radiological Health. (2006). Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims: Draft guidance. *Health and Quality of Life Outcomes*, *4*, 79. https://doi.org/10.1186/1477-7525-4-79

Vollert, J., Segelcke, D., Weinmann, C., Schnabel, K., Fuchtmann, F., Rosenberger, D. C., Komann, M., Maessen, T., Sauer, L., Kalso, E., Fletcher, D., Lavand'homme, P., Kaiser, U., Liedgens, H., Meissner, W., & Pogatzki-Zahn, E. M. (2024). Responsiveness of multiple patient-reported outcome measures for acute postsurgical pain: Primary results from the international multicentre PROMPT NIT-1 study. *British Journal of Anaesthesia*, *132*(1), 96–106. https://doi.org/10.1016/j.bja.2023.10.020

Weldring, T., & Smith, S. M. S. (2013). Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). *Health Services Insights*, *6*, 61–68. https://doi.org/10.4137/HSI.S11093

Williamson, P. R., Altman, D. G., Bagley, H., Barnes, K. L., Blazeby, J. M., Brookes, S. T., Clarke, M., Gargon, E., Gorst, S., Harman, N., Kirkham, J. J., McNair, A., Prinsen, C. A. C., Schmitt, J., Terwee, C. B., & Young, B. (2017). The COMET handbook: Version 1.0. *Trials*, *18*(S3), 280. https://doi.org/10.1186/s13063-017-1978-4

World Health Organization (WHO). (2002). *Towards a common language for functioning, disability and health—ICF*.

Xie, F., Li, S.-C., Roos, E. M., Fong, K.-Y., Lo, N.-N., Yeo, S.-J., Yang, K. Y., Yeo, W., Chong, H. C., & Thumboo, J. (2006). Cross-cultural adaptation and validation of Singapore English and Chinese versions of the Knee Injury and Osteoarthritis Outcome Score (KOOS) in Asians with knee osteoarthritis in Singapore. *Osteoarthritis and Cartilage*, *14*(11), 1098–1103. https://doi.org/10.1016/j.joca.2006.05.005

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Heitkamp, H., Heußner, D., Rosenberger, D. C., Schnabel, K., Rosenthal, D., Bigalke, S., Maeßen, T. V., Hohenschurz-Schmidt, D., Liedgens, H., Kaiser, U., & Pogatzki-Zahn, E. M. (2024). Systematic reviews and quality assessment of patient-reported outcome measures for physical function in comparative effectiveness studies regarding acute postoperative pain after total knee arthroplasty—Do we need to start all over again? *European Journal of Pain*, *28*, 1415–1430. https://doi.org/10.1002/ejp.2272