Systematic Review

JOURNAL OF CHILDREN'S ORTHOPAEDICS

Journal of Children's Orthopaedics
2024, Vol. 18(1) 96–110

DOI: 10.1177/18632521231207081
journals.sagepub.com/home/cho

Quality of patient- and proxy-reported outcomes for children with impairment of the lower extremity: A systematic review using the COnsensus-based Standards for selection of health Measurement INstruments methodology

Tim FF Saris[1,2], JP Ruben Kalle[3], Inger N Sierevelt[4], Denise Eygendaal[1], and Christiaan JA van Bergen[1,2]

Abstract

Background: Patient-reported outcome measures have become crucial in the clinical evaluation of patients. Appropriate selection, in a young population, of the instrument is vital to providing evidence-based patient-centered healthcare. This systematic review applies the COnsensus-based Standards for selection of health Measurement INstruments methodology to provide a critically appraised overview of patient-reported outcome measures targeted at pediatric orthopedic patients with lower limb impairment.

Method: A systematic search of electronic databases was performed to identify original studies reporting the development and/or validation of patient-reported outcome measures evaluating children with impairment of the lower extremity. Data extraction, quality assessment, and risk of bias evaluation were performed following the COnsensus-based Standards for selection of health Measurement INstruments guidelines and Preferred Reporting Items for Systematic reviews and Meta-Analyses statement.

Results: A total of 6919 articles were screened. Thirty-three studies were included, reporting evidence on the measurement properties of 13 different patient-reported outcome measures and 20 translations. Four studies reported on content validity and patient-reported outcome measure development. The methodological quality of studies on structural validity, content validity, or patient-reported outcome measure development was mostly rated as "doubtful" or "very good." The quality of evidence on measurement properties varied noticeably, with most studies needing to perform improve their methodological quality to justify their results.

Conclusion: This review provides an extensive overview of all available patient-reported outcome measures for patients with lower extremity impairment within pediatric orthopedics. We cautiously advise the use of four patient-reported outcome measures. However, the scarce availability of research on content validity and patient-reported outcome measure development highlights an area for future research endeavors to improve our knowledge on the currently available patient-reported outcome measures.

Level of evidence: Diagnostic level I

Keywords: COnsensus-based Standards for selection of health Measurement INstruments, pediatric orthopedics, lower extremity, measurement properties, review (publication type)

[1]Erasmus MC, Rotterdam, The Netherlands
[2]Amphia Hospital, Breda, The Netherlands
[3]Utrecht University, Utrecht, The Netherlands
[4]Stichting SCORE, Amsterdam, The Netherlands

Corresponding Author:
Tim FF Saris, Erasmus MC, Dr. Molewaterplein 40, 3015GD Rotterdam, The Netherlands.
Email: tim.saris@me.com

Introduction

A patient-reported outcome (PRO) is an informational statement coming directly from patients describing their mental and physical well-being over time without the clinicians' interpretation.[1] Measuring these outcomes accurately requires a validated patient-reported outcome measure (PROM).[1] Clinicians and researchers can, through PROMs, evaluate a patient's health status and track changes over time.[2,3] PROMs are specifically developed to evaluate the patient within a predefined study population and require validation studies to provide adequate scientific justification for proper usage. A recent study by Arguelles et al.[4] revealed improper use of the majority of PROMs in pediatric orthopedic research. These results describe the adverse effects of using PROMs without proper validation studies, consequently influencing the findings of researchers and clinicians using these pediatric orthopedic PROMs.

The selection of a PROM is vital to its effectiveness. Factors essential to the proper selection and application of a PROM are the study population and disease for which the PROM was developed and subsequently validated. Therefore, it is imperative that clinicians and researchers are provided with validated PROMs. Traumatic injuries/impairments of the extremities are among the most common pediatric diagnoses.[5,6] Meanwhile, clinical follow-up through survival-based outcomes has transitioned to clinical follow-up using PROMs.[7,8] The need for proper use of validated PROMs has become more apparent because of this transition and increased incidence of injury.

This study is part of a research collaboration that aims to provide a comprehensive overview of PROMs used in pediatric orthopedics, evaluate the methodological quality of the respective validation studies, and formulate a recommendation on proper PROMs selection. A comprehensive overview of PROMs used for evaluating upper extremity problems has been previously published as part of this joint effort.[9] This review will provide an extensive overview of PROMs used in orthopedic patients with a lower extremity impairment.

Methods and materials

A detailed study protocol has been previously published.[10]

Design: This systematic review was performed in accordance with the 10-step procedure within the updated COnsensus-based Standards for selection of health Measurements INstruments (COSMIN) risk of bias guidelines.[11–13] This systematic review adhered to the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) statement when reporting the synthesized evidence and data.[14]

Pre-registration: This study was pre-registered in PROSPERO with registration number: CRD42021287323.

Search strategy: The electronic database of PubMed and Embase were systematically searched to identify all relevant and current studies between 1 January 2000 and 1 December 2022. A clinical librarian was consulted to design a search string for both search engines to improve sensitivity and ensure proper identification of all relevant studies. The search was restricted to English and/or Dutch studies using a language filter. The search strings for each database can be found in greater detail in Appendix 1. In addition, a validated pediatric study search filter (by Leclercq et al.[15]) and two validated data filters (by Terwee et al.[16]) were used to enhance sensitivity.

Eligibility criteria: Articles were considered eligible for inclusion if the following criteria were met: the full-text article of the original study was available and must describe the development and/or the evaluation of one or more measurement properties of a generic and/or disease-specific patient-reported and/or proxy-reported questionnaire of any language, in a population consisting of children (0–18 years old) with an impairment of the lower extremity. The exclusion criteria were as follows: a study using PROMs as an outcome measurement instrument or a PROM in a study population consisting of children with prosthetics.

Study selection: Study selection was split into two phases: a screening selection through title and abstract followed by a careful selection through full text. Two reviewers (T.F.F.S. and J.P.R.K.) independently identified eligible articles, and disagreements were resolved by a third reviewer (C.J.A.v.B.).

Data extraction and appraisal: The data of all included studies were extracted and underwent qualitative appraisal using the COSMIN methodology for qualitative evaluation of PROMs.[11] More detailed information on the taxonomy, guideline, and checklists can be found in studies by Mokkink et al.,[17,18] Terwee et al.,[19] and Prinsen et al.[11]

Evaluation of the methodological quality: The methodological quality of all included studies assessing the following measurement properties were rated with the COSMIN Risk of bias checklist;[12] validity (structural validity, hypotheses testing for construct validity and cross-cultural validity), reliability (internal consistency, reliability, and measurement error), and/or responsiveness of a PROM. To evaluate the overall methodological quality, the study was rated on a 4-point scale: "very good," "adequate," "doubtful," or "inadequate." After this "the worst score counts" principle was applied to determine the study's overall methodological quality on the PROM's psychometric properties.

Data extraction: The data on the patient characteristics of the included study populations, the characteristics of the included PROMs, and the evidence within each study on a psychometric property were extracted using the predefined tables within the COSMIN risk of bias checklist.[11]

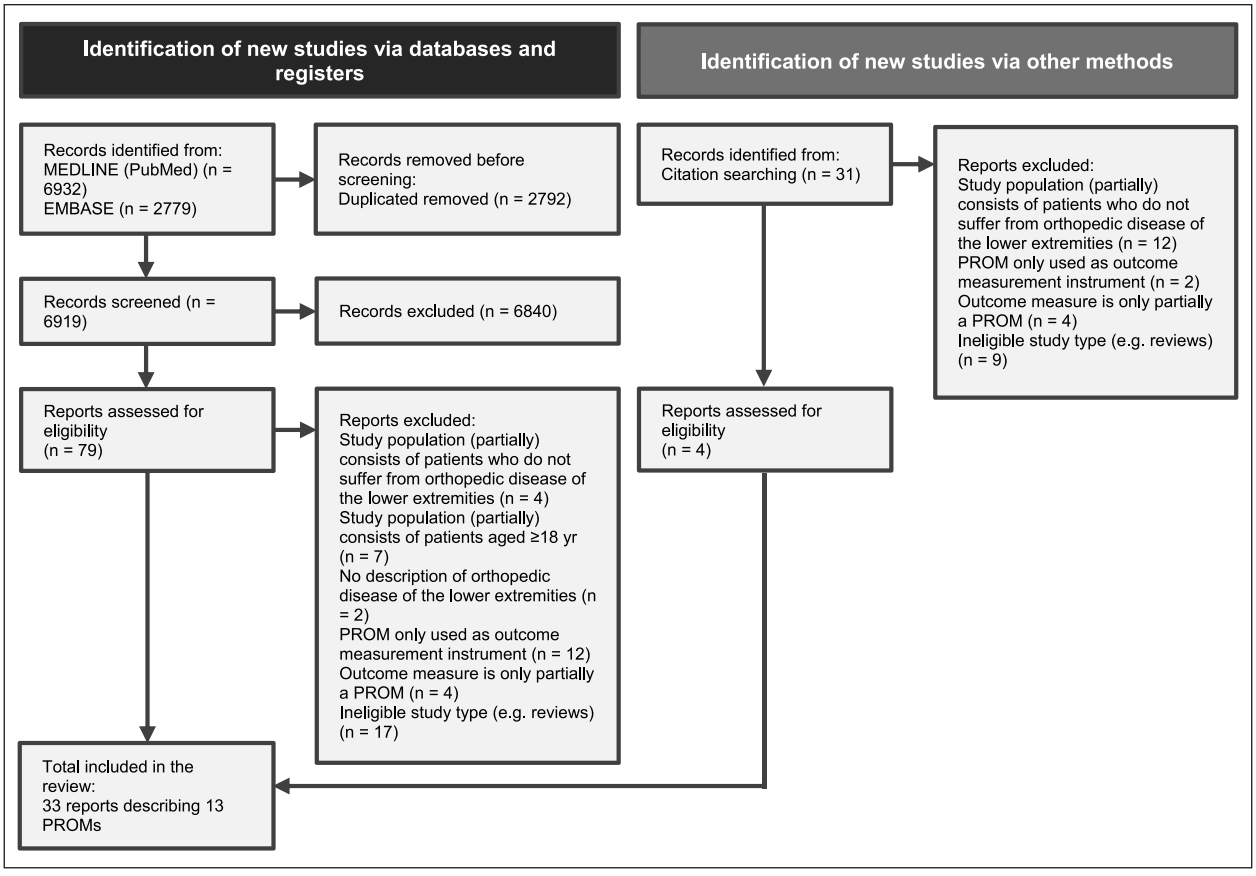

Figure 1. PRISMA flowchart detailing the article selection process with the inclusion and exclusion criteria.

Assessment of psychometric properties: A set of predefined criteria[11] was constructed to standardize the rating of the studies on measurement properties. The results gathered from each individual study were rated as "sufficient (+)," "insufficient (-)," or "indeterminate (?)" accordingly.

Evidence synthesis: The evaluation of the methodological quality and the assessment of the studies on psychometric properties were performed by two reviewers (T.F.F.S. and J.P.R.S.) independently. If consensus could not be reached, an additional reviewer (C.J.A.v.B. and/or I.N.S.) was consulted. To ensure an evaluation and assessment of high quality, both reviewers upheld the predefined criteria corresponding to the COSMIN risk of bias checklist. The inter-rater agreement was considered appropriate when reviewers reached >80% agreement, a value proposed by the COSMIN development group.[20]

Results

General characteristics of included studies and instruments: After duplicate removal (n=2792), a total of 6919 original studies were identified. After careful selection using the predefined inclusion and exclusion criteria, 33 original studies were included. The PRISMA flow diagram describing the selection process is shown in Figure 1. The inter-rater agreement was calculated to be 87.8%, which exceeded the predefined minimum of 80% and was therefore deemed appropriate.

The key characteristics of the included studies and their study populations are described in Table 1. The 33 studies reported evidence on 13 individual PROMs and 20 translations. The least evaluated psychometric property was the responsiveness of the instrument (n=1). The most evaluated psychometric property was the construct validity (n=26). The detailed characteristics of the included measurement instruments can be found in Table 2.

Synthesized evidence: The results of the methodological quality assessment of the included studies, according to the COSMIN guidelines, can be found in Table 3. In addition, a comprehensive summary of the pooled results per psychometric properties reported for each study on an original PROM, according to COSMIN guidelines and GRADE approach, can be found in Table 4.

Content validity and PROM development: Four studies described content validity and PROM development. To determine the methodological quality of PROM development, a study should report on the comprehensibility and comprehensiveness of the instrument. Two studies[25,30] reported on these parameters and were rated as "adequate."

Table 1. Characteristics of the included studies.

PROM	References	n	Age, mean (SD, range), years	Gender % female	Disease	Country	Language
ACL-QOL	Marien et al.[21]	294	16.2 (1.3, 10.9–17.9)	64	Anterior cruciate ligament injury	Canada	English
Barnhoft	Herngren et al.[22]	61	No mean (8–15)	Not reported	Slipped capital femoral epiphysis	Sweden	Swedish
BPII2.0	Lafave et al.[23]	133	14.9 (1.9, 8–18)	72	Patellafemoral instability	The United states of America	English
DSI	Masrouha et al.[24]	91	8.8 (3.3)	23	Idiopathic talipes equinovarus	The United states of America	English
GOAL-LD	Dermott et al.[25]	25	13.7 (9.0–17.9)	60	Lower limb differences	Canada	English
	Griffiths et al.[26]	137	12.5 (3.2)	50	Cerebral palsy	Australia	English
KOOS-Child	van der Velden et al.[27]	101	15.0 (1.85)	66	Acute and chronic knee injuries	The Netherlands	Dutch
	Rioux Trottier et al.[28]	99	Control group: 13.3 (8.1–16.4) Pain group: 14.1 (9.2–16.8)	Control group: 57 Pain group: 60	Various knee problems	Canada	French
	Moutzouri et al.[29]	60	11.0 (1.8)	50	Knee injuries	Greece	Greek
	Örtqvist et al.[30]	34	No mean (10–16)	50	Knee injuries	Sweden	Swedish and English
	Örtqvist et al.[31]	115	13.0 (1.85)	56	Various knee problems	Sweden	Swedish
	Neuhaus et al.[32]	47	Conservative group: 13.4 (1.8) Intervention group: 15.3 (1.9)	Conservative group: 79 Intervention group: 78	Acute knee injuries	Germany	German
	Hansen et al.[33]	153	13.7 (10–15)	52	ACL deficiency	Denmark	Danish
LIMB-QoL	Chhina et al.[34]	17	13 (8–17)	41	Lower limb deficiency	Canada	English
PEDI-IKDC	van der Velden et al.[27]	101	15.0 (1.85)	66	Acute and chronic knee injuries	The Netherlands	Dutch
	Boykin et al.[35]	135	15 (10.1–18.9)	41	Anterior cruciate ligament injuries	United states of America	English
	Macchiarola et al.[36]	89	12.6 (2.1)	48	Acute and chronic knee injuries	Italy	Italian
	Schafer et al.[37]	145	14 (7–17)	53	Knee pain	United states of America	English
	Kocher et al.[38]	589	14.6 (2.5)	51	Acute knee injuries	United states of America	English
	Jacobsen et al.[39]	99	14 (9–18)	49	Acute knee injuries	Denmark	Danish
	Schmitt et al.[40]	673	14.2	54	Various knee problems	United states of America	English
	Herrera Rodríguez et al.[41]	50	11.8 (8–15)	Not reported	Traumatic knee injuries	Colombia	Spanish
PediSKV	Marot et al.[42]	44	13.8 (1.5)	36	Knee pain	France	French
PED-QL	Kothari et al.[43]	95	Pediatric flat feet: 12.0 (2.0) Typical developing flat feet: 11.9 (2.3)	Pediatric flat feet: 60 Typical developing flat feet: 34	(Congenital) feet abnormalities	The United Kingdom	English
PODCI	Lee et al.[44]	92	10.2 (3.7, 5.1–18)	34	Cerebral palsy	South Korea	Korean
PROMIS	Adjei et al.[45]	294	13.7 (2.1)	50	Lower extremity injuries	The United states of America	English
	Schafer et al.[37]	145	14 (7–17)	53	Knee pain	The United states of America	English
	Masrouha et al.[24]	91	8.8 (3.3)	23	Idiopathic club feet	The United States of America	English
OxAFQ-c	Burger et al.[46]	64	10.8 (3.4)	42	Congenital foot/ankle disease	The Netherlands	Dutch
	Morris et al.[47]	16	9.5 (5–15)	38	Congenital foot/ankle disease	The United Kingdom	English
	Hajebrahimi et al.[48]	49	8.8 (3.3, 6–18)	27	Congenital talipes equinovarus	Turkey	Turkish
	Kothari et al.[43]	95	Pediatric flat feet: 12.0 (2.0) Typical developing flat feet: 11.9 (2.3)	Pediatric flat feet: 60 Typical developing flat feet: 34	(Congenital) feet abnormalities	The United Kingdom	English
	Martinelli et al.[49]	61	10.9 (1.6)	46	Symptomatic flexible flatfeet	Italy	Italian
	Morris et al.[50]	80	12.2 (2.5)	48	Congenital foot/ankle disease	The United Kingdom	English
	Morris et al.[51]	158	11.1 (2.9)	Not reported	Congenital foot/ankle disease	The United Kingdom	English
	Cho et al.[52]	169	10.8 (3.3)	40	Congenital foot/ankle disease	South Korea	Korean
	Alotaibi et al.[53]	87	8.3 (2.9)	44	Congenital foot/ankle disease	Saudi Arabia	Arabic

SD: standard deviation; ACL-QOL: Anterior Cruciate Ligament Quality of Life; ACL: anterior cruciate ligament; BPII2.0: Banff Patellofemoral Instability Instrument 2.0; DSI: disease-specific instrument; GOAL-LD: Gait Outcome Assessment List for Lower-limb Differences; KOOS: Knee Osteoathritis Outcome Score; LIMB-QOL: lower LIMB Quality of Life; PEDI-IKDC: Pediatric Evaluation of Disability Inventory Internal Knee Documentation Committee; PediSKV: Pediatric Simple Knee Value; PED-QL: PEDiatric Quality of Life inventory; PROMIS: Patient-Reported Outcomes Measurement Information System; PODCI: Pediatric Outcomes Data Collection Instrument; OxAFQ-c: Oxford Ankle Foot Questionnaire for children.

Table 2. Characteristics of the included PROMs.

PROM (reference to first article)	Construct(s)	Target population	Mode of administration	Recall period	(Sub)scale(s) (number of items)	Response options	Range of scores/scoring	Original language	Available tanslations[a]
ACL-QOL [21]	Quality of life	Adolescent patients with ACL injury	Self-report	6, 12 and 24 months	5 domains (32 items)	Sliding scale	0–100	English	
Barnhoft[22]	Manual ability	Children 8–15 years with hip disorder	Parent proxy	6 weeks				English	Swedish
BPII2.0[23]		Adolescent patients with patellar instability	Self-report	2 weeks	5 domains			English	
DSI[24]	Functional assessment	Children with idiopathic clubfoot						English	English
GOAL-LD[25]	Lower extremity function	Children 5–18 years (with cerebral palsy)	Self-report and parent proxy		6 domains (45 items)	6 point Likert rating scale	0–5	English	
KOOS-Child[30]	Functional assessment	Children 10–18 years with knee injury	Self-report with parental help	2 weeks and 1 year	5 domain (48 items)	Rating scale and 5 level Likert rating scale	0–100	Swedish	Dutch, French, Greek, English, German, Danish
LIMB-QOL[34]	Quality of life	Children 8–18 years with lower limb deformities	Self-report	1 week	11 scales (159 items)	5 point Likert rating scale	0–4	English	
PEDI IKDC[54]	Functional assessment	Children 9–18 years with knee injuries	Self-report and parent proxy	2 weeks and 1 year	4 domains (18 items)	5 level Likert rating scale	0–100	English	Dutch, Italian, Danish, Spanish
Pedi-SVK[42]	Functional assessment	Pediatric patient with knee pain	Self-report		1 scale	Sliding scale	0–100	French	
PED-QL[43]	Perceived limitations	Children 2–18 years	Self-report and parent proxy		4 domains (23 items)	5 point Likert rating scale	0–4	English	
PODCI[44]	Perceived limitations	Children with cerebral palsy	Self-report					English	Korean
PROMIS[45]	Lower extremity function	Children 8–17 years with musculoskeletal disorder	Self-report and parent proxy					English	
OxAFQ-c[47]	Functional assessment	Children 5–17 years with foot ankle problems	Self-report and parent proxy	2 weeks and 4–6 months	4 domains (15 items)	5 points Likert rating scale	0–4	English	Dutch, Italian, Danish, Turkish, Korean, Arabic

PROM: patient-reported outcome measurement; ACL-QOL: Anterior Cruciate Ligament Quality of Life; ACL: anterior cruciate ligament; BPII2.0: Banff Patellofemoral Instability Instrument 2.0; DSI: disease-specific instrument; GOAL-LD: Gait Outcome Assessment List for Lower-limb Differences; KOOS: Knee Osteoathritis Outcome Score; LIMB-QOL: lower LIMB Quality of Life; PEDI-IKDC: Pediatric Evaluation of Disability Inventory Internal Knee Documentation Committee; PediSKV: Pediatric Simple Knee Value; PED-QL: PEDiatric Quality of Life inventory; PROMIS: Patient-Reported Outcomes Measurement Information System; PODCI: Pediatric Outcomes Data Collection Instrument; OxAFQ-c: Oxford Ankle Foot Questionnaire for children.

[a]PROM translations that have been cross-cultural adapted and/or validated in the population of interest of this review.

Table 3. Methodological quality and ratings of measurements properties of the included PROMs.

PROM	References	Measurement property	Methodological quality	Rating
ACL-QOL	Marien et al.[21]	Internal consistency	Very good	?
		Reliability	Doubtful	+
		Measurement error	Doubtful	?
		Responsiveness: construct approach (hypothesis testing) before and after intervention	Adequate	4+
Barnhoft	Herngren et al.[22]	Structural validity	Inadequate	?
		Reliability	Doubtful	+
		Hypothesis testing for construct validity: convergent validity	Adequate	1+/1−
BPII2.0	Lafave et al.[23]	Internal consistency	Very good	?
		Reliability	Doubtful	+
		Hypothesis testing for construct validity: convergent validity	Adequate	?
DSI	Masrouha et al.[24]	Hypothesis testing for construct validity: convergent validity	Doubtful	?
GOAL-LD	Dermott et al.[25]	PROMS development quality of pilot study	Adequate	
	Griffiths et al.[26]	Internal consistency	Very good	?
		Reliability	Adequate	+
		Hypothesis testing for construct validity: convergent validity	Adequate	2+
		Hypothesis testing for construct validity: discriminative validity	Doubtful	2+
KOOS-Child	Van der Velden et al.[27]	Reliability	Adequate	+
		Measurement error	Adequate	?
		Hypothesis testing for construct validity: convergent validity	Adequate	13+/7−
		Responsiveness: construct approach (hypothesis testing) before and after intervention	Inadequate	14+/6−
	Trottier et al.[28]	Internal consistency	Very good	-
		Hypothesis testing for construct validity: convergent validity	Adequate	7+
		Hypothesis testing for construct validity: discriminative validity	Doubtful	4+/4-
	Moutzouri et al.[29]	Structural validity	Inadequate	?
		Internal consistency	Very good	+
		Reliability	Inadequate	?
		Hypothesis testing for construct validity: convergent validity	Adequate	?
		Responsiveness: construct approach (hypothesis testing) before and after intervention	Inadequate	+
	Ortqvist et al.[30]	PROMS development quality of pilot study	Adequate	
	Ortqvist et al.[31]	Internal consistency	Very good	+
		Reliability	Adequate	+
		Measurement error	Adequate	?
		Hypothesis testing for construct validity: convergent validity	Very good	7+
		Responsiveness: construct approach (hypothesis testing) comparison between subgroups	Very good	5+
	Neuhaus et al.[32]	Internal consistency	Very good	+
		Reliability	Doubtful	?
		Responsiveness: construct approach (hypothesis testing) before and after intervention	Doubftul	?
	Hansen et al.[33]	Structural validity	Adequate	−
		Measurement invariance	Inadequate	+
		Reliability	Adequate	?
		Responsiveness	Adequate	?

(Continued)

Table 3. (Continued)

PROM	References	Measurement property	Methodological quality	Rating
LIMB-Q	Chhina et al.[34]	PROMS development quality of pilot study	Very good	
PEDI IKDC	Van der Velden et al.[27]	Reliability	Adequate	+
		Measurement error	Adequate	?
		Hypothesis testing for construct validity: convergent validity	Adequate	6+/2−
		Responsiveness: construct approach (hypothesis testing)	Inadequate	3+/1−
	Boykin et al.[35]	Internal consistency	Inadequate	−
		Hypothesis testing for construct validity: convergent validity	Adequate	7+/5−
	Macchiarola et al.[36]	Internal consistency	Very good	?
		Reliability	Very good	+
		Measurement error	Very good	?
		Hypothesis testing for construct validity: convergent validity	Doubtful	1+/1−
		Responsiveness: construct approach (hypothesis testing) before and after intervention	Adequate	?
	Schafer et al.[37]	Hypothesis testing for construct validity: convergent validity	Very good	?
	Kocher et al.[38]	Internal consistency	Doubtful	?
		Reliability	Adequate	+
		Hypothesis testing for construct validity: convergent validity	Adequate	11+
		Hypothesis testing for construct validity: discriminative validity	Doubtful	1+
		Responsiveness: construct approach (hypothesis testing) before and after intervention	Doubtful	1+
	Jacobsen et al.[39]	Internal consistency	Very good	?
		Reliability	Doubtful	+
		Measurement error	Doubtful	?
		Responsiveness: construct approach (hypothesis testing) before and after intervention	Doubtful	+
	Schmitt et al.[40]	Internal consistency	Very good	?
		Hypothesis testing for construct validity: convergent validity	Adequate	3+
	Herrera et al.[41]	Internal consistency	Doubtful	?
		Reliability	Doubtful	−
Pedi-SVK	Marot et al.[42]	Hypothesis testing for construct validity: convergent validity	Inadequate	1+
		Hypothesis testing for construct validity: discriminative validity	Inadequate	1+
		Responsiveness: construct approach (hypothesis testing) before and after intervention	Inadequate	?
PED-QL	Kothari et al.[43]	Hypothesis testing for construct validity: convergent validity	Adequate	6+/2−
		Hypothesis testing for construct validity: discriminative validity	Doubtful	7+/1−
PODCI	Min Lee et al.[44]	Structural validity	Doubtful	?
PROMIS	Adjei et al.[45]	Internal consistency	Very good	?
	Schafer et al.[37]	Reliability	Adequate	?
		Measurement error	Adequate	?
		Hypothesis testing for construct validity: convergent validity	Very good	2+
		Hypothesis testing for construct validity: convergent validity	Very good	?

(Continued)

Table 3. (Continued)

PROM	References	Measurement property	Methodological quality	Rating
	Masrouha et al.[24]	Hypothesis testing for construct validity: convergent validity	Doubtful	?
	Griffiths et al.[26]	Internal consistency	Very good	?
		Reliability	Doubtful	+
		Hypothesis testing for construct validity: convergent validity	Adequate	2+
		Hypothesis testing for construct validity: discriminative validity	Doubtful	2+
	Schmitt et al.[40]	Internal consistency	Very good	?
		Hypothesis testing for construct validity: convergent validity	Adequate	3+
OxAFQ-c	Burger et al.[46]	Internal consistency	Very good	−
		Reliability	Doubtful	−
		Hypothesis testing for construct validity: convergent validity	Inadequate	6+/1−
		Responsiveness: construct approach (hypothesis testing) comparison between subgroups	Inadequate	?
	Morris et al.[47]	PROMS development quality of elicitation study	Very good	
	Hajebrahimi et al.[48]	Internal consistency	Very good	+
		Reliability	Doubtful	+
		Hypothesis testing for construct validity: convergent validity	Adequate	4+/2−
	Kothari et al.[43]	Hypothesis testing for construct validity: convergent validity	Adequate	6+/2−
		Hypothesis testing for construct validity: discriminative validity	Doubtful	7+/1−
	Martinelli et al.[49]	Internal consistency	Very good	+
		Reliability	Adequate	+
		Measurement error	Adequate	?
		Hypothesis testing for construct validity: convergent validity	Adequate	2+
		Responsiveness: construct approach (hypothesis testing) before and after intervention	Doubtful	8+
	Morris et al.[50]	Internal consistency	Very good	+
		Reliability	Very good	+
		Measurement error	Very good	?
		Responsiveness: construct approach (hypothesis testing) before and after intervention	Very good	4+
	Morris et al.[51]	Structural validity	Adequate	?
		Internal consistency	Very good	?
		Reliability	Doubtful	?
		Hypothesis testing for construct validity: convergent validity	Adequate	1+
		Hypothesis testing for construct validity: discriminative validity	Doubtful	1+
	Hee Cho et al.[52]	Structural validity	Doubtful	?
		Internal consistency	Very good	+
		Hypothesis testing for construct validity: convergent validity	Doubtful	6+
	Alotaibi et al.[53]	Internal consistency	Very good	+
		Reliability	Adequate	+
		Measurement error	Adequate	?
		Hypothesis testing for construct validity: convergent validity	Very good	7+

PROM: patient-reported outcome measurement; ACL-QOL: Anterior Cruciate Ligament Quality of Life; ACL: anterior cruciate ligament; BPII2.0: Banff Patellofemoral Instability Instrument 2.0, DSI: disease-specific instrument; GOAL-LD: Gait Outcome Assessment List for Lower-limb Differences; KOOS: Knee Osteoathritis Outcome Score; LIMB-QOL: lower LIMB Quality of Life; PEDI-IKDC: Pediatric Evaluation of Disability Inventory Internal Knee Documentation Committee; PediSKV: Pediatric Simple Knee Value; PED-QL: PEDiatric Quality of Life inventory; PROMIS: Patient-Reported Outcomes Measurement Information System; PODCI: Pediatric Outcomes Data Collection Instrument; OxAFQ-c: Oxford Ankle Foot Questionnaire for children.

Table 4. Synthesized evidence.

PROM (reference)	Measurement property	Summarized results	Overall rating	Quality of evidence
ACL-QoL[21]	Internal consistency	Cronbach's alpha = 0.93–0.97	?	
	Reliability	ICC = 0.997	+	Moderate
	Measurement error	SEM = 0.75	?	
	Responsiveness	4 out of 4 hypothesis confirmed. Cohen d = 0.56	±	Low
Barnhoft[22]	Structural validity	No results were reported	?	
	Reliability	Weighted kappa = 0.88–1.00	+	Moderate
	Construct validity	1 out of 2 hypothesis confirmed	±	Low
BPII2.0[23]	Internal consistency	Cronbach's alpha = 0.95	?	
	Reliability	ICC = 0.94	+	Moderate
	Construct validity	No hypothesis defined	?	
DSI[24]	Construct validity	No hypothesis defined	?	
GOAL-LD[25,26]	Internal consistency	Cronbach's alpha = 0.85–0.94	?	
	Reliability	ICC = 0.85	+	High
	Construct validity	4 out of 4 hypothesis confirmed.	+	High
KOOS-Child (original Swedish)[30,31]	Internal consistency	Cronbach's alpha = 0.8–0.9	+	High
	Reliability	ICC = 0.78–0.91	+	Moderate
	Measurement error	SEM = 5.3–8.1, SDCind = 14.6–22.6, SDCcgr = 1.7–2.7	?	
	Construct validity	7 out of 7 hypothesis confirmed	+	Moderate
	Responsiveness	5 out of 5 hypothesis confirmed	+	Moderate
KOOS-Child (Dutch version)[27]	Reliability	ICC = 0.7–1.0	+	Moderate
	Measurement error	SEM = 8.9–16.9, SDC = 24.7–46.9	?	
	Construct validity	13 out of 20 hypothesis confirmed	±	Moderate
	Responsiveness	14 out of 20 hypothesis confirmed	±	Moderate
KOOS-Child (French version)[28]	Internal consistency	Cronbach's alpha = 0.622–0.959	?	
	Construct validity	11 out of 15 hypothesis confirmed	±	Moderate
KOOS-Child (Greek version)[29]	Structural validity	Model fit not described	−	Very low
	Internal consistency	Cronbach's alpha = 0.81–0.96	+	
	Reliability	ICC/Kappa not reported	−	Low
	Construct validity	No hypothesis defined	−	Low
	Responsiveness	AUC = 0.99–1.00	+	Moderate
KOOS-Child (German version)[32]	Internal consistency	Cronbach's alpha = 0.77–0.90	+	
	Reliability	Spearman's correlation = 0.85–0.94	?	
	Responsiveness	No hypothesis defined	?	
KOOS-Child (Danish version)[33]	Structural validity	For all subscales, the unidimensional CFA model did not fit the data.	−	High
	Measurement invariance	No important DIF for group factors	+	Moderate
	Reliability	ICC or weighted kappa not reported	?	Low
	Responsiveness	No hypothesis defined	?	Low
Pedi-IKDC[35,37,38,40]	Internal consistency	Cronbach's alpha = 0.765–0.91	+	
	Reliability	ICC = 0.86–0.95	+	Moderate
	Construct validity	22 out of 27 hypothesis confirmed	±	Low
	Responsiveness	1 out of 1 hypothesis confirmed	+	Low
Pedi-IKDC (Dutch version)[27]	Reliability	ICC = 0.8–0.9	+	Moderate
	Measurement error	SEM = 8.6, SDC = 23.8	?	
	Construct validity	6 out of 8 hypothesis confirmed	±	Low
	Responsiveness	3 out of 4 hypothesis confirmed	±	Low
Pedi-IKDC (Italian version)[36]	Internal consistency	Cronbach's alpha = 0.92	+	
	Reliability	ICC = 0.96	+	Moderate
	Measurement error	SEM = 4.4, SDC = 12.3	?	

(Continued)

Table 4. (Continued)

PROM (reference)	Measurement property	Summarized results	Overall rating	Quality of evidence
	Construct validity	1 out 2 hypothesis confirmed	±	Very low
	Responsiveness	No hypothesis defined	?	
Pedi-IKDC (Danish version)[39]	Internal consistency	Cronbach's alpha = 0.90	?	
	Reliability	ICC = 0.90	+	Moderate
	Measurement error	SEM = 4.1, SDC = 11.3	?	
	Responsiveness	ROC = 0.7	+	Low
Pedi-IKDC (Spanish version)[41]	Internal consistency	Cronbach's alpha = 0.824	?	
	Reliability	ICC = 0.649	−	Low
Pedi-SVK[42]	Construct validity	2 out of 2 hypothesis confirmed	+	Moderate
	Responsiveness	No hypothesis defined	?	
Ped-QoL[40,43]	Internal consistency	Cronbach's alpha = 0.911	?	
	Construct validity	16 out of 19 hypothesis confirmed	±	Low
PODCI[44]	Structural validity	2 of 12 items did not fit into unidimensionality scale ($t \geq 1.96$)	?	
PROMIS[24,37,45]	Internal consistency	No Cronbach's alpha reported	?	
	Reliability	No weighted kappa reported	−	Very low
	Measurement error	No standard measurement error reported	−	Very low
	Construct validity	2 out of 2 hypothesis confirmed	+	Low
OxAFQ-c (original version)[43,47,50,51,55]	Structural validity	No model fit reported.	−	Very low
	Internal consistency	Cronbach's alpha = 0.78–0.92	?	
	Reliability	ICC = 0.76–0.95	+	Moderate
	Measurement error	MDC = 6–8	?	
	Construct validity	15 out of 18 hypothesis confirmed	±	Low
	Responsiveness	4 out of 4 hypothesis confirmed	+	Moderate
OxAFQ-c (Dutch version)[46]	Internal consistency	Cronbach's alpha = 0.602–0.826	?	
	Reliability	ICC = 0.538–0.884	−	Very low
	Construct validity	6 out 7 hypothesis	±	Moderate
	Responsiveness	No hypothesis defined	?	
OxAFQ-c (Turkish version)[48]	Internal consistency	Cronbach's alpha = 0.92–0.96	?	
	Reliability	ICC = 0.93–0.96	+	Moderate
	Construct validity	4 out of 6 hypothesis confirmed	±	Low
OxAFQ-c (Italian version)[49]	Internal consistency	Cronbach's alpha in all domain ≥ 0.7	?	
	Reliability	ICC = 0.87–0.99	+	Moderate
	Measurement error	SDD = 7.0–17.0	?	
	Construct validity	2 out of 2 hypothesis confirmed	+	Low
	Responsiveness	8 out of 8 hypothesis confirmed	+	Low
OxAFQ-c (Korean version)[52]	Structural validity	No model fit reported	−	Very low
	Internal consistency	Cronbach's alpha = 0.765–0.901	?	
	Construct validity	6 out of 6 hypothesis confirmed	+	Moderate
OxAFQ-c (Arabic version)[53]	Internal consistency	Cronbach's alpha = 0.80–0.89	?	
	Reliability	ICC = 0.87–0.94	+	Moderate
	Measurement error	SEM = 6.1–10.2, MDC = 16.9–28.2	+	Low
	Construct validity	7 out of 7 hypothesis confirmed	+	Low

PROM: patient-reported outcome measurement; ACL-QoL: Anterior Cruciate Ligament Quality of Life; ICC: intraclass correlation coefficient; SEM: standard error of measurement; BPII2.0: Banff Patellofemoral Instability Instrument 2.0, DSI: disease-specific instrument; GOAL-LD: Gait Outcome Assessment List for Lower-limb Differences; KOOS: Knee Osteoathritis Outcome Score; SDCind: smallest detectable change index group; SDCcgr: smallest detectable change control group; SDC: smallest detectable change; AUC: area under curve; CFA: confirmatory factor analysis; DIF: differential item functioning; PEDI-IKDC: Pediatric Evaluation of Disability Inventory Internal Knee Documentation Committee; PediSKV: Pediatric Simple Knee Value; PED-QoL: PEDiatric Quality of Life inventory; PODCI: Pediatric Outcomes Data Collection Instrument; PROMIS: Patient-Reported Outcomes Measurement Information System; OxAFQ-c: Oxford Ankle Foot Questionnaire for children; SDD: smallest detectable difference; MDC: minimal detectable change.

Two studies[34,47] reported on these parameters and were rated as "very good."

Structural validity: The structural validity was evaluated for six PROMs.[22,29,33,44,51,52] Three studies reported on the structural validity of a cultural adaption of the original PROMs.[29,44,52] All studies [22,29,33,44,51,52] failed to report on the important aspects of the item-related theory or Rasch analyses and did not describe the model fit. The results of the PROMs did not meet the COSMIN criteria for good measurement properties.

Internal consistency: To provide a correct analysis on the quality of the internal consistency within a PROM, the instrument must show at least low-quality evidence of its structural validity. Accordingly, the results of the analyses on internal consistency for 19 PROMs were reported as "indeterminate" due to insufficient evidence on the structural validity studies of these PROMs. The original version of the KOOS-Child questionnaire[30] and the original version of the OxAFQ-c[47] were the only two studies that both showed an adequate quality of evidence for structural validity and were therefore rated.

Reliability: Twenty-one studies[21–23,26,27,29,31–33,36,38,39,41,45,46,48,49–51,53] reported on the test–retest reliability of the instrument. Four studies reported no intraclass correlation coefficient (ICC) values or ICC values ranging below <0.7, thus demonstrating insufficient evidence on the quality of the test–retest reliability of the instrument.[29,41,45,46]

Measurement error: Eleven studies reported the measurement error of their respective PROM.[21,27,31,33,36,39,45,49,50,53] None of the PROMs included in these analyses had previously published information on the minimal important change. Therefore, all included data on measurement error were rated as "indeterminate."

Discussion

This is the first systematic review to provide an extensive overview of the available evidence on the psychometric properties of PROMs used to evaluate pediatric patients with an impairment of the lower extremity. This systematic review is a continuation of the joined effort to provide a comprehensive overview of the evidence on the psychometric properties of PROMs used for both upper and lower extremities.[9,10] With the use of the updated COSMIN methodology, to ensure a high-quality assessment, a total of 6919 publications were screened, resulting in 13 PROMs reporting a total of 112 psychometric properties. The results of this study increase the knowledge on proper use of PROMs through an evidence-based review and extensive recommendations.

Given the data currently available and our corresponding analyses for both upper and lower extremities, which together constitute the current evidence on psychometric properties of pediatric orthopedic PROMs, a lack of sufficient quality evidence becomes unmistakable. Consequently, our extensive overview demonstrates that the use of the instruments included in this review and currently used in pediatric orthopedics for evaluating lower extremity impairment is not sufficiently supported. At least a low-quality evidence on content validity and internal consistency, combined with at least adequate methodological quality of the study itself, is essential to justify proper usage of a PROM in current clinical practice and for research purposes.[11] Therefore, more studies should have provided sufficient evidence to support the structural validity of the instruments and have provided more evidence on content validity and internal consistency.

The most commonly researched PROMs, and commonly used in current practice, are the KOOS-Child, OxAFQ-c, and Pedi-IKDC. Other PROMs (ACL-QoL, Barnhoft, BPII 2.0, DSI, Pedi-SVK, Ped-Qol, PODCI, PROMIS) have shown at least adequate evidence on methodological quality of their studied psychometric properties; however, these PROMs have no evidence on the quality of content validity and PROM development. During our analyses, four PROMs were supported with evidence on PROM development and/or content validity. The OxAFQ-c, LIMB-Q, and GOAL-LD[25,34,47] provided adequate quality of evidence on content validity, the most important psychometric property when determining whether or not the instrument has the ability to detect how well an instrument covers all relevant parts of the construct it aims to measure. The KOOS-Child provided very good evidence on the respective PROM development but no evidence concerning content validity. Our analyses showed none of the PROMs were tested for the same population and disease in two different individual studies. Because of the strengths of the studies on the KOOS-Child, OxAFQ-c, and GOAL-LD, we could formulate a very cautious positive advice on the use of these PROM. The PEDI-IKCD has no supportive evidence on the development of the PROM; however, studies provided quality evidence on the psychometric properties, we could also cautiously advice using the PEDI-IKCD.

For clinicians and researchers using the instrument, it is imperative to know whether the instrument is fitted to test the study population at hand. The use of each instrument should therefore be validated for a specific patient population and it should be clearly stated for what disease or disability the PROM has been validated. The patient and disease-specific characteristics of the included study populations were, in general, not described in extensive detail; the severity of the disease was mostly left out. The applicability of an instrument to be tested on a specific population is therefore lessened, especially combined with the overall very low quality of provided evidence for the suspected population in which the instrument is tested in the original study.

Table 5. Summary of the PROMs which have shown enough evidence to be used in daily clinical practice and for research purposes.

PROM (reference to first article)	Testable construct(s)	Target population	Mode of administration	Original language (and translations)	Summary of all measurement properties found within this review per PROM
GOAL-LD[25]	Lower extremity function	Children 5–18 years (with cerebral palsy)	Self-report and parent proxy	English	Reliability Internal consistency PROM development quality of pilot study Hypothesis testing for construct validity: convergent validity Hypothesis testing for construct validity: discriminative validity
KOOS-Child[30]	Functional assessment	Children 10–18 years with knee injury	Self-report with parental help	Swedish (Dutch, French, Greek, English, German, Danish)	Reliability Responsiveness Structural validity Measurement error Internal consistency PROMS development quality of pilot study Hypothesis testing for construct validity: convergent validity Hypothesis testing for construct validity: discriminative validity Responsiveness: construct approach (hypothesis testing) before and after intervention Responsiveness: construct approach (hypothesis testing) comparison between subgroups
PEDI IKDC[54]	Functional assessment	Children 9-18 years with knee injuries	Self-report and parent proxy	English (Dutch, Italian, Danish, Spanish)	Reliability Internal consistency Measurement Error Hypothesis testing for construct validity: convergent validity Hypothesis testing for construct validity: discriminative validity Responsiveness: construct approach (hypothesis testing) Responsiveness: construct approach (hypothesis testing) before and after intervention
OxAFQ-c[47]	Functional assessment	Children 5–17 years with foot ankle problems	Self-report and parent proxy	English (Dutch, Italian, Danish, Turkish, Korean, Arabic)	Reliability Structural validity Internal consistency Measurement error PROMS development quality of elicitation study Hypothesis testing for construct validity: convergent validity Hypothesis testing for construct validity: discriminative validity Responsiveness: construct approach (hypothesis testing) comparison between subgroups Responsiveness: construct approach (hypothesis testing) before and after intervention

PROM: patient-reported outcome measures; GOAL-LD: Gait Outcome Assessment List for Lower-limb Differences; PEDI: Pediatric Evaluation of Disability Inventory Internal Knee Documentation Committee.

The COSMIN methodology makes a review on psychometric properties of PROMs a standardized process. This systematic approach strengthens the objective results gathered. However, using the COSMIN checklist requires subjective input and judgment by the reviewer. To address this potential source of selection and information bias, two (and sometimes three) reviewers independently extracted, gathered, and evaluated the data according to the COSMIN guidelines after which these results were shared, pooled, and discussed. Throughout the evaluation on the methodological quality of the studies, the "worst score counts" principles was applied in multiple instances. The incidental subjective input from the reviewers combined with a "worst score counts" principles could lead to an underestimate of the true quality and quantity of the evidence available. This is especially important in pediatric orthopedic PROMS, where research is scarce and not widely known or available.

In conclusion, a comprehensive overview was given of PROMs for pediatric patients with lower extremity impairment. We cautiously advice on the use of the KOOS-Child, OxAFQ-c, and GOAL-LD PROM because of their relevant evidence on PROMs development and content validity with supportive evidence on the psychometric properties. The Pedi-IKCD could also, cautiously, be used in practice due to good quality of evidence on the psychometric properties; however, no studies support the content validity or PROM development of this PROM. Noticeably, the absence of content validity–"only" studies on the PROMs accentuates the knowledge gap in proper PROMs usage, as no study has provided any evidence that the PROMs reflect the construct they intend to measure. These results are similar for PROMs used for pediatric patients with upper extremity impairments. Future efforts must be made to develop high-quality and well-validated patient-reported outcomes measurement questionnaire for the pediatric orthopedic patient. A summarized overview of the PROMs which can, cautiously, be used in daily clinical practice and research can be found in Table 5.

Authors' note

All listed authors have contributed to the design and analysis of the decision rule, commented on the drafts, and approved the final manuscript and agree with being accountable for all of its aspects and with its submission to *Journal of Children's Orthopaedics* all in accordance with the ICMJE guidelines.

Acknowledgements

A special thanks to Paulien H. Wiersma MSc, faculty liaison medical sciences (University Library Utrecht, UMC Utrecht), for helping to create our search strings.

Author contributions

In the process of creating the manuscript, the five authors fulfilled the following roles and tasks:

1. TFFS — Main author and first reviewer; design of the review, constructing and executing the search strategy, screening publications, analysis and interpretation of data, conception, and writing of the manuscript.
2. JPRK — Second reviewer; constructing and executing the search strategy, screening publications, analysis, and interpretation of data.
3. INS — Second supervisor; providing critical revisions and help with interpretation of data.
4. DE — Third supervisor; provided orthopedic and scientific expertise and critical revisions to the study protocol.
5. C.J.A.v.B. — First supervisor; conception and design of the review, provided orthopedic and scientific expertise and critical revisions to the study protocol.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethics approval and consent to participate

There are no human participants in this article and informed consent is not applicable.

Consent for publication

Not applicable.

ORCID iDs

Tim FF Saris https://orcid.org/0000-0002-2366-4497

Inger N Sierevelt https://orcid.org/0000-0003-0924-9358

Availability of data and materials

All data analyzed during this study are included in this article and its supplementary files.

Supplemental material

Supplemental material for this article is available online.

References

1. Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health* 2007; 10(Suppl. 2): S125–S137.
2. Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013; 346: f167.
3. Weldring T and Smith SMS. Article commentary: patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). *Health Serv Insight* 2013; 2013: 6.

4. Arguelles GR, Shin M, Lebrun DG, et al. The majority of patient-reported outcome measures in pediatric orthopaedic research are used without validation. *J Pediatr Orthop* 2021; 41(1): e74–e79.
5. Hedström EM, Svensson O, Bergström U, et al. Epidemiology of fractures in children and adolescents: increased incidence over the past decade: a population-based study from northern Sweden. *Acta Orthop* 2010; 81: 148–153.
6. Naranje SM, Erali RA, Warner WC Jr, et al. Epidemiology of pediatric fractures presenting to emergency departments in the United States. *J Pediatr Orthop* 2016; 36(4): e45–e48.
7. Truong WH, Price MJ, Agarwal KN, et al. Utilization of a wide array of nonvalidated outcome scales in pediatric orthopaedic publications: can't we all measure the same thing. *J Pediatr Orthop* 2019; 39(2): e153–e158.
8. Phillips L, Carsen S, Vasireddi A, et al. Use of patient-reported outcome measures in pediatric orthopaedic literature. *J Pediatr Orthop* 2018; 38(8): 393–397.
9. Kalle JPR, Saris TFF, Sierevelt IN, et al. Quality of patient- and proxy-reported outcomes for children with impairment of the upper extremity: a systematic review using the COSMIN methodology. *J Patient Rep Outcomes* 2022; 6: 58.
10. Saris T, Kalle R, Sierevelt I, et al. Quality of patient-reported and proxy-reported outcomes for children with impairment of the lower extremity: systematic review protocol. *BMJ Paediatr Open* 2022; 6(1): e001486.
11. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018; 27(5): 1147–1157.
12. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018; 27(5): 1171–1179.
13. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018; 27(5): 1159–1170.
14. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021; 372: n71.
15. Leclercq E, Leeflang MM, van Dalen EC, et al. Validation of search filters for identifying pediatric studies in PubMed. *J Pediatr* 2013; 162(3): 629–634.
16. Terwee CB, Jansma EP, Riphagen II, et al. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009; 18(8): 1115–1123.
17. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; 63(7): 737–745.
18. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; 19(4): 539–549.
19. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012; 21(4): 651–657.
20. Mokkink LB, Terwee CB, Gibbons E, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based standards for the selection of health status measurement instruments) checklist. *BMC Med Res Methodol* 2010; 10: 82.
21. Marien M, Lafave MR, Hiemstra LA, et al. Validity, responsiveness, and reliability of the ACL-QOL in an adolescent population. *J Pediatr Orthopaed* 2021; 41: e917–e922.
22. Herngren B, Stenmarker M and Enskär K. Barnhöft: a hip specific 6-item questionnaire for children. *J Patient Rep Outcomes* 2017; 1(1): 16.
23. Lafave MR, Hiemstra LA, Parikh SN, et al. Validity and reliability of the Banff patellofemoral instability instrument 2.0 in an adolescent population. *J Pediatr Orthop* 2020; 40(2): e103–e108.
24. Masrouha KZ, Moses MJ, Sala DA, et al. The validity of patient-reported outcome measurement information system (PROMIS) parent proxy instruments to assess function in children with talipes equinovarus. *J Pediatr Orthop* 2019; 39(10): e787–e790.
25. Dermott JA, Wright FV, Salbach NM, et al. Development of the gait outcomes assessment list for lower-limb differences (GOAL-LD) questionnaire: a child and parent reported outcome measure. *Health Qual Life Outcomes* 2021; 19: 139.
26. Griffiths AL, Donnan LT, Iobst CA, et al. The gait outcomes assessment list for children with lower limb difference (GOAL-LD): assessment of reliability and validity. *J Pediatr Orthopaed* 2021; 41: 450–456.
27. van der Velden CA, van der Steen MC, Leenders J, et al. Pedi-IKDC or KOOS-child: which questionnaire should be used in children with knee disorders? *BMC Musculoskelet Disord* 2019; 20: 240.
28. Rioux Trottier E, Beauséjour M, Lamer S, et al. Validation of the French version of the KOOS-child questionnaire. *Knee Surg Sports Traumatol Arthrosc* 2019; 27(7): 2361–2367.
29. Moutzouri M, Tsoumpos P, Bania T, et al. "Greek KOOS-Child: a valid, disease specific, diagnostically accurate and responsive PROM in children with knee-related pathology." *Knee Surg Sports Traumatol Arthrosc* 2021; 29: 1841–1849.
30. Örtqvist M, Roos EM, Broström EW, et al. Development of the knee injury and osteoarthritis outcome score for children (KOOS-child): comprehensibility and content validity. *Acta Orthop* 2012; 83(6): 666–673.
31. Örtqvist M, Iversen MD, Janarv PM, et al. Psychometric properties of the knee injury and osteoarthritis outcome score for children (KOOS-Child) in children with knee disorders. *Br J Sports Med* 2014; 48(19): 1437–1446.
32. Neuhaus C, Camathias C, Mumme M, et al. The German version of the KOOS-child questionnaire (Knee injury and osteoarthritis outcome score for children) shows a good to excellent internal consistency and a high test–retest reliability in children with knee problems. *Knee Surg Sport Traumatol Arthrosc* 2022; 31: 1354–1360.
33. Hansen CF, Østergaard Madsen M, Warming S, et al. KOOS-child exhibits inadequate structural validity in a cohort of

paediatric patients with ACL deficiency. *Br J Sports Med* 2022; 56(22): 1284–1291.
34. Chhina H, Klassen AF, Kopec JA, et al. What matters to children with lower limb deformities: an international qualitative study guiding the development of a new patient-reported outcome measure. *J Patient Rep Outcomes* 2021; 5: 30.
35. Boykin RE, McFeely ED, Shearer D, et al. Correlation between the child health questionnaire and the international knee documentation committee score in pediatric and adolescent patients with an anterior cruciate ligament tear. *J Pediatr Orthop* 2013; 33(2): 216–220.
36. Macchiarola L, Grassi A, Di Paolo S, et al. The Italian cross-cultural adaptations of the paediatric international knee documentation committee score and the hospital for special surgery paediatric functional activity brief scale are reliable instruments in paediatric population. *Knee Surg Sports Traumatol Arthrosc* 2020; 28(8): 2657–2662.
37. Schafer KA, Minaie A and Nepple JJ. Outcome metrics in pediatric sports medicine: do PROMIS computer-adaptive testing metrics correlate with pedi-IKDC. *J Pediatr Orthop* 2020; 40(7): e616–e620.
38. Kocher MS, Smith JT, Iversen MD, et al. Reliability, validity, and responsiveness of a modified international knee documentation committee subjective knee form (Pedi-IKDC) in children with knee disorders. *Am J Sports Med* 2011; 39(5): 933–939.
39. Jacobsen JS, Knudsen P, Fynbo C, et al. Reproducibility and responsiveness of a Danish Pedi-IKDC subjective knee form for children with knee disorders. *Scand J Med Sci Sports* 2016; 26(12): 1408–1414.
40. Schmitt LC, Paterno MV and Huang S. Validity and internal consistency of the international knee documentation committee subjective knee evaluation form in children and adolescents. *Am J Sport Med* 2010; 38: 2443–2447.
41. Herrera Rodríguez JS, Ponce de León MC, Castañeda JF, et al. Transcultural validation and adaptation of the Pedi-IKDC scale for the functional assessment of children that undergo knee surgery. *Rev Esp Cir Ortop Traumatol* 2022; 66: 500–503.
42. Marot V, Vilette H, Dalmas Y, et al. Pediatric simple knee value: a simple patient-reported outcome measure for the knee. *J Child Orthop* 2021; 15: 76–80.
43. Kothari A, Stebbins J, Zavatsky AB, et al. Health-related quality of life in children with flexible flatfeet: a cross-sectional study. *J Child Orthop* 2014; 8(6): 489–496.
44. Lee KM, Chung CY, Park MS, et al. Level of improvement determined by PODCI is related to parental satisfaction after single-event multilevel surgery in children with cerebral palsy. *J Pediatr Orthop* 2010; 30(4): 396–402.
45. Adjei J, Schachne JM, Green DW, et al. Correlation between the PROMIS pediatric mobility instrument and the hospital for special surgery pediatric functional activity brief scale (HSS Pedi-FABS). *HSS J* 2020; 16(Suppl. 2): 311–315.
46. Burger E, Selles R, van Nieuwkasteele S, et al. The Dutch version of the Oxford ankle and foot questionnaire for children: useful for evaluation of pediatric foot problems in groups. *Foot Ankle Surg* 2019; 25(2): 204–210.
47. Morris C, Liabo K, Wright P, et al. Development of the Oxford ankle foot questionnaire: finding out how children are affected by foot and ankle problems. *Child Care Health Dev* 2007; 33(5): 559–568.
48. Hajebrahimi F, Tarakci D, Arman N, et al. Cross-cultural adaptation, validity and reliability of Turkish version of Oxford ankle foot questionnaire for children with congenital talipes equinovarus. *Foot Ankle Surg* 2021; 27(4): 439–442.
49. Martinelli N, Romeo G, Bonifacini C, et al. Validation of the Italian version of the oxford ankle foot questionnaire for children. *Qual Life Res* 2016; 25(1): 117–123.
50. Morris C, Doll H, Davies N, et al. The Oxford ankle foot questionnaire for children: responsiveness and longitudinal validity. *Qual Life Res* 2009; 18(10): 1367–1376.
51. Morris C, Doll HA, Wainwright A, et al. The Oxford ankle foot questionnaire for children scaling, reliability and validity. *J Bone Joint Surg Br* 2008; 90(11): 1451–1456.
52. Cho SH, Chung CY, Park MS, et al. Transcultural adaptation and validation of a Korean version of the oxford ankle foot questionnaire for children. *Health Qual Life Outcomes* 2020; 18: 118.
53. Alotaibi SS, Al-walah MA, Alhusaini AA, et al. Cross-cultural adaptation and validation of the Arabic version of oxford ankle foot questionnaire for children. *Disabil Rehabil* 2023; 45: 889–895.
54. Iversen MD, Lee B, Connell P, et al. Validity and comprehensibility of the international knee documentation committee subjective knee evaluation form in children. *Scand J Med Sci Sports* 2010; 20(1): e87–e95.
55. Morris C, Doll H, Wainwright A, et al. The oxford ankle foot questionnaire for children: review of development and potential applications. *Prosthet Orthot Int* 2010; 34(3): 238–244.