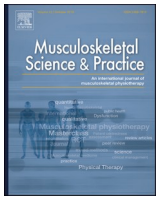
Musculoskeletal Science and Practice 82 (2026) 103511

Contents lists available at ScienceDirect

Musculoskeletal Science and Practice

journal homepage: www.elsevier.com/locate/msksp



Review article

Identifying reliable, valid and feasible outcome measures for adults aged 50 years or older with hip or knee osteoarthritis participating in supervised exercise programs: a scoping review

Nikole Watson [a,b,*], Rosmary Martinez Rueda [c], Hayley Legg [d], Catherine Boden [e], Brenna Bath [a]

[a] School of Rehabilitation Science, University of Saskatchewan, Saskatoon, Canada
[b] Craven SPORT Services, Saskatoon, Saskatchewan, Canada
[c] Department of Community Health and Epidemiology, University of Saskatchewan, Saskatoon, Canada
[d] School of Education and Applied Sciences, University of Gloucestershire, United Kingdom
[e] Leslie and Irene Dube Health Sciences Library, University of Saskatchewan, Saskatoon, Canada

ARTICLE INFO

Keywords:
Psychometric
Clinimetric
Measurement properties
Outcome assessment
Joints
Rehabilitation
Physiotherapy

ABSTRACT

Objective: To identify reliable, valid, and feasible outcome measures that could be used to assess outcomes of participating in supervised exercise programs for adults aged 50 years or older with hip or knee osteoarthritis (OA).

Methods: A scoping review was conducted in accordance with Joanna Briggs Institute and PRISMA-ScR guidelines. Six databases were searched on July 18, 2024. Eligible studies were systematic, literature, or integrative reviews evaluating at least one psychometric property (reliability, validity, or feasibility) of outcome measures for adults aged ≥50 years with hip or knee OA, including those awaiting or having undergone total joint arthroplasty.

Results: Sixteen reviews were included, identifying 102 outcome measures (35 self-report, 67 performance-based). Nine reviews used the COSMIN methodology, while others applied alternative or narrative frameworks. Commonly supported measures included the WOMAC, KOOS, HOOS, LEFS, 6-Minute Walk Test, 40-m fast-paced walk, and Timed Up and Go, which showed strong reliability (ICC ≥0.80) and construct validity. Feasibility was discussed narratively in 13 reviews; none used formal criteria. Only three mentioned exercise, and none evaluated measures in supervised programs.

Conclusion: Multiple outcome measures demonstrated acceptable reliability and validity for adults with hip or knee OA, but feasibility data were limited, and no tools were validated in the context of supervised exercise. Consequently, no measure fulfilled all three psychometric criteria for this setting. Future research should include structured feasibility assessments and evaluate measures within exercise-based interventions.

1. Introduction

Osteoarthritis (OA) is a degenerative joint disease affecting over 4 million Canadians (Vos et al., 2012; Bone and Joint Canada, 2024). The consequences of OA can be debilitating and negatively impact an individual's health status, activities of daily living, and quality of life (Papalia et al., 2020). Over 70% of individuals living with OA are over 55 years of age, and although aging does not play a causal role in the development of OA, it is expected that prevalence will continue to rise with ageing populations (Long et al., 2022). International best practice guidelines for hip and knee OA strongly recommend education and exercise therapy as key conservative management approaches (Kolasinski et al., 2019; Osteoarthritis Action Alliance, 2024). Supervised exercise programs that include disease-specific education have been shown to improve outcomes (Kolasinski et al., 2019). Therefore, it is important to identify suitable metrics used to assess the outcomes of adults aged 50 years or older with hip or knee OA who participate in exercise, as this will help reflect real-world populations affected by this disease and support evidence-based care in clinical settings.

The use of standardized metrics in research and clinical practice can

* Corresponding author. School of Rehabilitation Science, University of Saskatchewan, 104 Clinic Place, Saskatoon, SK, S7N 2Z4.
E-mail address: niw231@usask.ca (N. Watson).

https://doi.org/10.1016/j.msksp.2026.103511
Received 23 August 2025; Received in revised form 29 January 2026; Accepted 1 February 2026
Available online 2 February 2026

help determine pre-intervention baselines, monitor progress, and establish client outcomes. However, previous literature has demonstrated poor utilization of standardized tools in clinical trials and practice with significant variation in the selection of measures for hip and knee OA. Physical therapists play a key role in OA management by providing clients with disease-specific education, self-management strategies, and targeted exercise programs (Zhang et al., 2008; Teo et al., 2019; Skou and Roos, 2017; Woolacott et al., 2012; Regino et al., 2020) –(Zhang et al., 2008; Teo et al., 2019; Skou and Roos, 2017; Woolacott et al., 2012 Physical therapists recognize the benefits of evidence-based practice, but face challenges in implementation, including difficulties interpreting measures, time constraints limiting use, limited applicability to patients, and a lack of guidance from these measures for care planning (Jette et al., 2003, 2009; Verheyden and Meyer, 2016)–(Jette et al., 2003, 2009; Verheyden and Meyer, 2016; Sørensen et al., 2019).

Although standardized metrics are crucial for evidence-based practice, their limited uptake in routine clinical care underscores the need for more consistent and practical integration of tools with adequate clinimetric properties. Nevertheless, while several outcome measures demonstrate strong reliability and validity in OA populations, formal feasibility evaluations are lacking, and none have evaluated these measures specifically within the context of supervised exercise programs. This lack of context-specific validation limits the translation of psychometrically sound tools into real-world rehabilitation settings, where both feasibility and exercise-specific applicability are critical for sustained clinical use.

The objective of this scoping review was to identify outcome measures with adequate psychometric properties that are used to assess the outcomes of participating in supervised exercise programs for adults aged 50 years or older with hip or knee OA. For that purpose, we developed a scoping review to map the existing literature and identify the psychometric properties of the outcome measures used in OA populations (Watson et al., 2024). The information gathered will inform evidence-based practice and provide clinicians with the necessary tools to select appropriate metrics for assessing outcomes in patients aged 50 years or older with hip or knee OA.

2. Methods

2.1. Protocol

The protocol outlining the scoping review was published on July 16, 2024 as a pre-print through figshare (Watson et al., 2024).

2.2. Search strategy

We carried out a scoping review, defined as ‘a form of knowledge synthesis that maps key concepts, evidence types and gaps in the field by systematically searching, selecting, and synthesising existing knowledge’ (Colquhoun et al., 2014). A comprehensive search followed the Population–Concept–Context (PCC) strategy based on the Joanna Briggs Institute (JBI) guidelines for scoping reviews (Peters et al., 2022), which incorporates earlier scoping review guidance developed and refined by previous authors (Colquhoun et al., 2014; Arksey and O'Malley, 2005; Levac et al., 2010). The search was refined with a research librarian (CB) to select keywords and controlled vocabulary. Eligibility used a modified Population–Instrument–Construct–Outcome–Study design (PICOS) framework for measurement instruments. Population: adults ≥50 years with hip or knee osteoarthritis, including those awaiting or having undergone total joint arthroplasty; Instrument: any self-report measure or performance-based test; Construct: reliability, validity, or feasibility, as defined by the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) taxonomy (Mokkink et al., 2010); Outcome: quantitative or narrative evidence for ≥ one of those psychometric properties; and Study design: synthesised evidence (systematic reviews, integrative reviews, meta-analyses, clinical practice guidelines) or relevant grey literature. No comparator was required.

Self-report measures were standardised questionnaires completed by participants to rate pain, function or quality of life, thereby capturing subjective experiences that objective tests may overlook (Zhang et al., 2008). Performance-based tests provide measurable data on physical function that may not be fully reflected by a patient's perception, offering a more accurate and comprehensive assessment of the individual's abilities (Dobson et al., 2012; Chen et al., 2022; Terwee et al., 2006)–(Dobson et al., 2012; Chen et al., 2022; Terwee et al., 2006; Losina et al., 2020). Grey literature was included because JBI scoping-review guidance (Peters et al., 2022) calls for comprehensive coverage and several OA instruments were first described or psychometrically tested outside peer-reviewed journals (Bellamy, 2005; Wylde et al., 2005; Roos and Lohmander, 2003; Nilsdotter et al., 2003).

Search strategies were developed for Medical Literature Analysis and Retrieval System Online (MEDLINE) via Ovid and adapted for Excerpta Medica Database (EMBASE) via Ovid, the Cumulative Index to Nursing and Allied Health Literature (CINAHL), the Physiotherapy Evidence Database (PEDro), Epistemonikos and the Turning Research into Practice (TRIP) database. All searches were run on July 18, 2024; the full MEDLINE strategy appears in Appendix A.

2.3. Study selection

Studies were selected if they were related to hip or knee osteoarthritis, included adults ≥50 years of age, mentioned the validity, reliability, or feasibility of the metrics, and were written in English and published in full text. Only English-language publications were included due to feasibility constraints related to translation resources; as a result, relevant studies published in other languages may have been inadvertently excluded. While total joint arthroplasty (TJA) was not an explicit search term, studies evaluating these populations were included within our defined population scope if the outcome measures were primarily designed for or tested within hip and knee OA contexts. This approach reflects the clinical reality that these populations are often analyzed concurrently in psychometric syntheses for degenerative joint disease. Studies were excluded if they did not meet the above criteria or were diagnostic in nature. Two trial screening processes were completed by four independent reviewers prior to the title and abstract screening. The articles retrieved from the search had title and abstracts screened by two independent researchers using the data management software Zotero (Zotero, 2024). A full-text screen was conducted by two independent reviewers who utilized Microsoft Forms to screen the remaining articles based on the inclusion criteria. Any discrepancies regarding the inclusion status were resolved by a third reviewer at both stages.

2.4. Data extraction

A data extraction form was developed using Microsoft Forms. Two reviewers piloted the form on two full-text articles to ensure consistency, with adjustments made through consensus. Data extraction was performed by a single reviewer due to time constraints. To enhance rigor, the data extraction form was piloted by multiple reviewers, and earlier stages of the review (search strategy, title/abstract screening, full-text screening, and tool piloting) involved dual-reviewer processes. These steps were implemented to reduce bias and improve consistency across data handling. Extracted variables included author(s); study aim; setting and location; participant characteristics (age and joint[s] affected); outcome measures for adults ≥50 years with hip or knee OA; outcomes assessed; psychometric properties of each measure (reliability, validity, feasibility); and the presence of exercise- or education-based interventions, with or without supervision.

All outcome measures explicitly reported in the included studies were evaluated if they 1) were suitable for hip or knee OA or TJA and 2) applied to adults aged ≥50 years. Measures were grouped as self-report

or performance-based, and their inclusion in the synthesis was based solely on their identification in the reviewed studies and alignment with the above criteria.

2.5. Data analysis

Descriptive synthesis was used to collate and summarise all psychometric evidence. For every included study we first identified the appraisal framework (e.g., COSMIN, OMERACT, bespoke checklist, narrative-only) that guided its quality assessment of measurement properties. Frequency analyses were undertaken to quantify 1) the number of publications that investigated each psychometric property (reliability, validity, and feasibility) and 2) the number of independent evaluations conducted for each outcome measure across distinct author groups. Self-reported and performance-based measures were tabulated separately. Within each table, outcome instruments were labelled reliable, valid, or feasible when those domains were explicitly investigated and supported by the evidence presented in the source papers. Strength-of-evidence judgements were made qualitatively by examining consistency and magnitude of findings across the relevant psychometric domains for each measure. To explore the influence of methodological quality, we compared findings across reviews that applied the COSMIN methodology versus those that did not. No statistical pooling or meta-analytic procedures were undertaken.

3. Results

3.1. Overview of included studies

The search resulted in 3028 citations. After removing 717 duplicates, 2311 titles/abstracts were screened; 2282 were excluded. Twenty-five full texts were assessed and nine were further excluded, yielding 16 eligible reviews. A Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) flow chart details the study selection (Fig. 1). Published between 2006 and 2024, half of the 16 studies were issued within the last nine years, and all provided evidence relevant to adults ≥50 years with hip or knee OA while drawing on multinational data sets. Most reviews were systematic (n = 11), two incorporated meta-analysis, two were literature reviews, and one was an integrative review. Seven studies focused on the knee, three on the hip, and six on both joints. Additionally, three studies specifically evaluated the measurement properties of outcome instruments for populations awaiting or having undergone TJA. Descriptive characteristics of the sources are found in Appendix B.

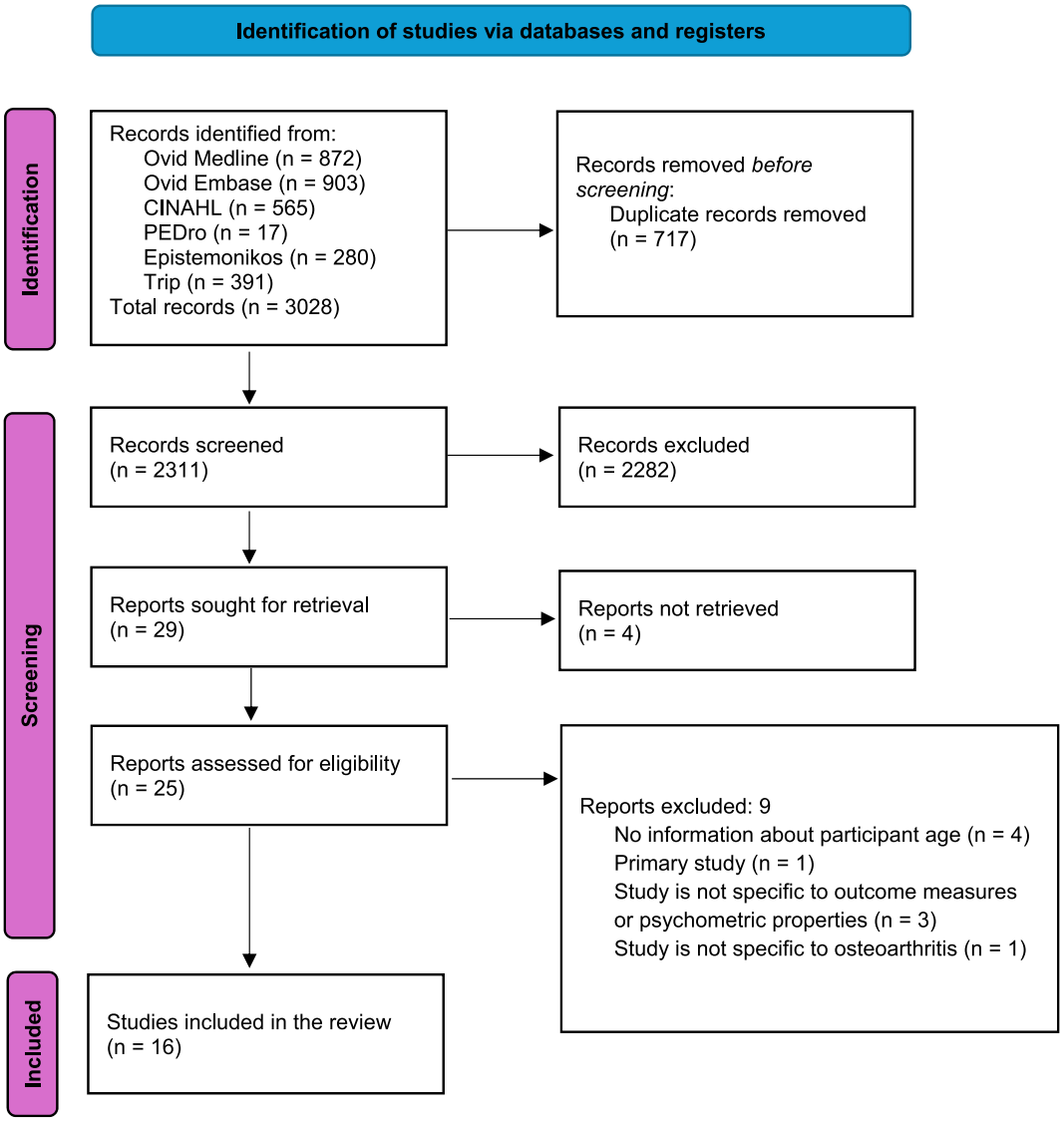


Fig. 1. PRISMA-ScR flow chart outlining the study selection.

3.2. Outcome measures identified

Across the 16 reviews, 102 unique outcome measures were extracted – 35 self-report and 67 performance-based measures. Frequently evaluated self-report tools were WOMAC, KOOS, HOOS, and OKS, whereas common performance tests included the 6MWT, TUG, and 40-m walk. Separate comprehensive inventories for self-report and performance-based measures are presented in Tables 1 and 2 respectively.

3.3. Psychometric evidence across studies

Among the 16 reviews, seven applied the COSMIN methodology to evaluate measurement properties (Dobson et al., 2012; Chen et al., 2022; Braaksma et al., 2020; Collins et al., 2016; French et al., 2020; Peer and Lane, 2013)–(Dobson et al., 2012; Chen et al., 2022; Braaksma et al., 2020; Collins et al., 2016; French et al., 2020; Peer and Lane, 2013; Moore and Barker, 2017), while two used QAPAQ (with COSMIN

Table 1
Psychometric properties and frequency of self-report measures (Moore and Barker, 2017) reported by author group.

Outcome Measure (alphabetical)	Authors	Frequency	Reliability	Validity	Feasibility
Activities of Daily Living Scale (ADLS)	Howe et al., 2012	1	✓	✓	D
Activity Rating Scale (ARS)	Smith et al., 2019;	2	✓	x	D
	Terwee et al., 2011		✓	x	D
American Academy of Orthopedic Surgeons Hip Score (AAOS-HS)	Thorborg et al., 2010	1	✓	✓	NR
AAOS Lower Limb Core Scale (AAOS-LLCS)	Kivlan and Martin, 2013	1	✓	✓	NR
Arthritis Impact Measurement Scales (AIMS/AIMS2)	Kivlan and Martin, 2013;	2	✓	✓	NR
	Samuel and Kanimozhi, 2019		✓	✓	NR
Assessment of Quality of Life (AQoL)	Howe et al., 2012	1	±	±	D
Baecke Questionnaire	Smith et al., 2019;	2	✓	±	D
	Terwee et al., 2011		✓	±	D
Comprehensive Osteoarthritis Test (COAT)	Samuel and Kanimozhi, 2019	1	✓	NR	NR
Daily Activity Questionnaire (DAQ)	Terwee et al., 2011	1	✓	✓	D
Hip Disability & Osteoarthritis Outcome Score (HOOS)	Kivlan and Martin, 2013;	2	✓	✓	NR
	Thorborg et al., 2010;		✓	✓	NR
HOOS Physical Function Short Form (HOOS-PS)	Braaksma et al., 2020	1	✓	✓	D
Human Activity Profile (HAP)	;Howe et al., 2012	3	✓	✓	D
	Smith et al., 2019;		±	±	D
	Terwee et al., 2011		x	±	D
Ibadan Knee OA Outcome Measure (IKHOAM)	Samuel and Kanimozhi, 2019	1	✓	±	NR
International Physical Activity Questionnaire – Short Form (IPAQ-SF)	Smith et al., 2019;	2	± x	x	D
	Terwee et al., 2011			x	D
Joint-Specific Multidimensional Assessment of Pain (J-MAP)	Howe et al., 2012	1	✓	✓	D
Knee Injury & Osteoarthritis Outcome Score (KOOS)	Collins et al., 2016;Howe et al., 2012	4	✓	✓	D
	;		✓	✓	D
	Peer and Lane, 2013;		±	✓	D
	Samuel and Kanimozhi, 2019		✓	✓	NR
KOOS Physical Function Short Form (KOOS-PS)	Braaksma et al., 2020;	3	✓	x	D
	Collins et al., 2016;		✓	±	NR
	Peer and Lane, 2013		✓	✓	NR
Knee OA Fears & Beliefs Questionnaire (KOFBeQ)	Samuel and Kanimozhi, 2019	1	✓	NR	NR
Knee Society Scoring System (KSSS)	Howe et al., 2012;	2	NR	✓	D
	Samuel and Kanimozhi, 2019		±	±	NR
Lequesne Algofunctional Index for OA of the Knee (LAIKOA)	Samuel and Kanimozhi, 2019	1	✓	✓	NR
Lequesne Index of Severity – Hip (LISOH)	Kivlan and Martin, 2013;	2	±	±	NR
	Thorborg et al., 2010		±	±	NR
Lequesne Index of Severity – Knee (LISOK)	Howe et al., 2012	1	NR	NR	D
Lower Extremity Activity Scale (LEAS)	Terwee et al., 2011	1	✓	✓	D
Lower Extremity Functional Scale (LEFS)	Howe et al., 2012	1	✓	✓	D
Measure of Intermittent & Constant OA Pain (ICOAP)	Howe et al., 2012	1	✓	✓	D
Osteoarthritis Knee & Hip QoL (OAKHQOL)	Kivlan and Martin, 2013	1	✓	✓	NR
Oxford Hip Score (OHS)	Kivlan and Martin, 2013;	2	✓	✓	NR
	Thorborg et al., 2010		✓	✓	NR
Oxford Knee Score (OKS)	Howe et al., 2012;	2	✓	✓	D
	Samuel and Kanimozhi, 2019		✓	✓	NR
Patient Specific Index (PASI)	Thorborg et al., 2010	1	✓	✓	NR
Physical Activity Scale for the Elderly (PASE)	Smith et al., 2019	1	±	x	D
Short Form-36 (SF-36)	Howe et al., 2012;	2	±	±	D
	Samuel and Kanimozhi, 2019		✓	✓	NR
Tegner Activity Scale (TAS)	Smith et al., 2019;	2	✓	x	D
	Terwee et al., 2011		±	±	D
Tegner Lysholm Knee Score (TLKSS)	Samuel and Kanimozhi, 2019	1	✓	✓	NR
University of California Los Angeles Activity Scale (UCLA)	Smith et al., 2019;	2	✓	±	D
	Terwee et al., 2011		✓	✓	D
Western Ontario & McMaster Universities Arthritis Index (WOMAC)	Howe et al., 2012;	4	✓	✓	D
	Kivlan and Martin, 2013;		✓	✓	NR
	Samuel and Kanimozhi, 2019;		✓	✓	NR
	Thorborg et al., 2010		±	±	NR

Legend.
✓ = adequate evidence (positive ratings in ≥1 reliability domain or validity domain) where reliability ICC ≥0.70.
± = mixed or indeterminate evidence (one domain adequate, another inadequate or missing).
x = inadequate evidence.
D = descriptive in nature; no formal scoring.
NR = not reported.

Table 2
Psychometric properties and frequency of performance-based measures (67) reported by author group.

Category	Measure	Authors	Frequency	Reliability	Validity	Feasibility
Walk tests	30 sec fast-paced walk	Chen et al., 2022	1	✓	✓	NR
	2 min normal-paced walk	Chen et al., 2022	1	✓	✓	NR
	2 min fast-paced walk (2MWT)	Chen et al., 2022	1	✓	✓	NR
	5 min walking field	Terwee et al., 2006	1	✓	±	D
	6 min walk (6MWT)	Chen et al., 2022;	4	✓	✓	NR
		Dobson et al., 2012;		✓	✓	D
		Howe et al., 2012		✓	✓	D
		Terwee et al., 2006		NR	±	D
	30 min fast-paced walk	Chen et al., 2022	1	✓	✓	NR
	8 ft self-paced walk	Dobson et al., 2012;	2	✓	NR	D
		Howe et al., 2012		✓	NR	D
	50 ft fast-paced walk	Chen et al., 2022;	3	✓	✓	D
		Dobson et al., 2012;		✓	NR	D
		Howe et al., 2012		✓	NR	D
	3 m backward fast-paced walk	Chen et al., 2022	1	✓	x	NR
	3 m self-paced walk	Terwee et al., 2006	1	±	NR	D
	4 m fast-paced walk	Chen et al., 2022	1	✓	✓	NR
	4 m self-paced walk	Chen et al., 2022	1	✓	✓	NR
	5 m multi-paced walk	Dobson et al., 2012;	2	NR	NR	D
		Howe et al., 2012		NR	✓	D
	10 m self-paced walk	Chen et al., 2022	1	✓	x	NR
	13 m self-paced walk	Chen et al., 2022;	4	✓	x	NR
		Dobson et al., 2012;		✓	NR	D
		Howe et al., 2012		✓	NR	D
		Terwee et al., 2006		±	NR	D
	13 m self-paced walk (carpet)	Terwee et al., 2006	1	✓	±	D
	20 m self-paced walk	Chen et al., 2022	1	✓	✓	NR
	40 m fast-paced walk	Chen et al., 2022;	2	✓	X	NR
		Dobson et al., 2012		✓	NR	D
	40 m self-paced walk	Chen et al., 2022;	3	✓	✓	NR
		Dobson et al., 2012;		✓	✓	D
		Howe et al., 2012		✓	✓	D
	40 m walk (unspecified pace)	Chen et al., 2022	1	✓	x	NR
	80 m fast-paced walk	Dobson et al., 2012	1	NR	NR	D
Chair stand tests	30 sec chair stand	Chen et al., 2022;	2	✓	x	NR
		Dobson et al., 2012		✓	NR	D
	Chair rise ×3 reps	Chen et al., 2022	1	✓	✓	NR
	Chair rise ×5 reps	Chen et al., 2022;	2	✓	x	NR
		Dobson et al., 2012		NR	NR	D
Stair/step tests	4-step stair ascend	Terwee et al., 2006	1	NR	NR	D
	15 sec step up/down repetitions	Chen et al., 2022	1	✓	✓	NR
	30 sec stair climb (12 steps)	Terwee et al., 2006	1	NR	NR	D
	2 min step	Chen et al., 2022	1	✓	✓	NR
	Four square step test (FSST)	Moore and Barker, 2017	1	✓	±	D
	L-test	Chen et al., 2022	1	✓	NR	NR
	Single step test (SST)	Chen et al., 2022	1	✓	✓	NR
	Up/down 4 stairs	Dobson et al., 2012;	2	✓	NR	D
		Howe et al., 2012		✓	NR	D
	Up/down 9 stairs	Chen et al., 2022;	3	✓	x	NR
		Dobson et al., 2012;		✓	NR	D
		Howe et al., 2012		✓	NR	D
	Up/down 10 stairs	Chen et al., 2022	1	✓	x	NR
	Up/down 12 stairs	Dobson et al., 2012;	2	NR	✓	D
		Howe et al., 2012		NR	✓	D
	Up/down stairs (unspecified)	Chen et al., 2022	1	✓	x	NR
Other tests	Accelerometry (5m)	Terwee et al., 2006	1	NR	NR	D
	Aggregated Locomotor Function (ALF)	Chen et al., 2022;	4	✓	✓	NR
		Dobson et al., 2012;		✓	✓	D
		Howe et al., 2012		✓	✓	D
		Terwee et al., 2006		✓	±	D
	Community Balance and Mobility Scale (CB&M)	French et al., 2020;	2	✓	✓	D
		Samuel and Kanimozhi, 2019		✓	✓	NR
	DynaPort KneeTest	Terwee et al., 2006	1	NR	±	D
	Footprint analysis (50 ft)	Terwee et al., 2006	1	±	±	D
	Force-plate measures (e.g., single leg stance centre of pressure, medial-lateral/ anterior-posterior sway)	French et al., 2020	1	±	±	D
	Functional Assessment System (FAS)	Dobson et al., 2012;	2	✓	✓	D
		Terwee et al., 2006		✓	x	D
	Gait Analysis 1 (10 m)	Terwee et al., 2006	1	NR	±	D
	Gait Analysis 2 (8 m)	Terwee et al., 2006	1	±	±	D
	Gait Analysis 3 (12 m)	Terwee et al., 2006	1	NR	±	D
	Gait Analysis 4 (5 m runs)	Terwee et al., 2006	1	NR	±	D
	Gait Analysis (3-Dimensional)	Ornetti et al., 2010	1	x	x	NR

(*continued on next page*)

Table 2 (*continued*)

Category	Measure	Authors	Frequency	Reliability	Validity	Feasibility
	Get Up and Go (GUG)	Chen et al., 2022;	3	✓	✓	NR
		Dobson et al., 2012;		✓	x	D
		Terwee et al., 2006		✓	✓	D
	Goniometer	Howe et al., 2012	1	✓	✓	D
	Isokinetic dynamometry	Myers and Bohannon, 2020	1	✓	✓	NR
	Iowa Level of Assistance Scale (ILAS)	Chen et al., 2022;	2	✓	x	NR
		Terwee et al., 2006		✓	±	D
	Lin Battery (8 ft walk, up/down 4 stairs, chair rise x5)	Dobson et al., 2012;	2	✓	±	D
		Terwee et al., 2006		±	±	D
	Locométrix (40 m)	Terwee et al., 2006	1	NR	±	D
	Madsen test (50 m + 30-step)	Terwee et al., 2006	1	NR	NR	D
	Marks test (13 m + 4-step)	Terwee et al., 2006	1	±	NR	D
	Numact Monitor	Terwee et al., 2006	1	NR	±	D
	Pedometer	Terwee et al., 2006;	2	NR	±	D
		Terwee et al., 2011		x	✓	D

± = mixed or indeterminate evidence (one domain adequate, another inadequate or missing).
x = inadequate evidence.
D = descriptive in nature; no formal scoring.
NR = not reported.

overlap) (Smith et al., 2019; Terwee et al., 2011), and seven relied on alternative frameworks or narrative appraisal (Terwee et al., 2006; Howe et al., 2012; Kivlan and Martin, 2013; Myers and Bohannon, 2020; Ornetti et al., 2010; Samuel and Kanimozhi, 2019; Thorborg et al., 2010) as demonstrated in Table 3. Outcome measures recommended in COSMIN-guided reviews consistently included the WOMAC, KOOS, HOOS, LEFS, TUG, 6-MWT, 40-m fast-paced walk, and 30-s chair-stand. Non-COSMIN reviews frequently reported and supported the same core measures, particularly WOMAC, HOOS, KOOS, and performance-based walk and chair-stand tests. As such, stratification revealed considerable overlap in recommended outcome measures between COSMIN and non-COSMIN reviews, suggesting that the choice of appraisal framework did not substantially alter the most utilized tools. However, reviews using the COSMIN methodology tended to report psychometric data in a more structured and transparent manner. Formal feasibility assessments were absent, even among COSMIN-guided reviews, and only thirteen of

Table 3
Appraisal frameworks and psychometric domains reported across included reviews of outcome measures for hip and knee osteoarthritis.

Author (year)	Appraisal framework	Reliability – domains reported	Validity – domains reported	Feasibility – assessment approach
Braaksma et al., 2020	COSMIN (2018 checklist) + GRADE	α, ICC, SEM/LoA/SDC	Content, structural, construct, responsiveness	Partially – length, licence, completion time
Chen et al., 2022	COSMIN (2018)	Test–retest ICCs, α	Construct (hypothesis testing)	–
Collins et al., 2016	COSMIN (2010) + meta-analysis, GRADE	α, pooled ICC, SEM/SDC	Content, structural, construct, cross-cultural, responsiveness	Partially – items, time, missing-item rates
Dobson et al., 2012	COSMIN (2010) + Terwee synthesis	Intra-/inter-rater ICC for 16 tests	Construct, criterion, structural	–
French et al., 2020	COSMIN (2018) + Terwee synthesis	Test–retest ICC, SEM/MDC	Construct, criterion, structural, responsiveness	Partially – time, equipment, ceiling/floor
Howe et al., 2012	OMERACT filter (Truth + Discrimination)	ICC, κ, α ("discrimination")	Content, criterion, construct ("truth")	Partially – feasibility noted during selection
Kivlan and Martin, 2013	Narrative appraisal (no tool)	α, test–retest ICC	Content, construct	Partially – items, time, scoring burden
Moore and Barker, 2017	QUADAS-2 + COSMIN (2010)	Inter-/intra-/test–retest ICC	Concurrent, construct, predictive (AUC)	Partially – set-up, staff, cost, floor effects
Myers and Bohannon, 2020	Narrative appraisal (no tool)	Test–retest ICC, MDC	Content, construct	–
Ornetti et al., 2010	OMERACT filter	Test–retest ICC >0.80	Construct, discriminant	Partially – cost, time, safety
Peer and Lane, 2013	COSMIN (2010)	α, test–retest ICC, SEM/SDC/LoA	Content, construct (a-priori hypotheses)	Partially – completion/scoring time, missing data
Samuel and Kanimozhi, 2019	Narrative appraisal (no tool)	α, ICC	Content, construct	Partially – items, mode, time
Smith et al., 2019	Modified COSMIN (2010) + QAPAQ	α, ICC	Content, construct, criterion	Partially – ease-of-use table
Terwee et al., 2006	Bespoke Terwee checklist	Test–retest/inter-rater ICC	Construct (pre-set hypotheses)	Partially – time, space, burden
Terwee et al., 2011	COSMIN (2010) + QAPAQ	α, ICC, κ	Content, construct, criterion	Partially – QAPAQ attributes
Thorborg et al., 2010	Terwee quality criteria (2007)	ICC, κ, SEM/SDC	Content, construct (≥75 % hypotheses)	Partially – items, time, floor/ceiling

Frequency summary (N = 16): Frameworks: COSMIN = 9; COSMIN + QAPAQ = 2; OMERACT = 2; Terwee criteria = 2; Narrative appraisal = 4; QUADAS-2 = 1 (some reviews used >1 framework)- Reliability: 16/16 reported ≥1 reliability statistic.- Validity: 16/16 reported ≥1 validity domain; most common were Construct (15/16), Content (13/16), Structural (6/16), Responsiveness (4/16), Criterion (5/16), Cross-cultural (1/16).- Feasibility: 0/16 used a formal feasibility rubric; 13/16 reported feasibility narratively; 3/16 did not report feasibility at all.
Abbreviations: COSMIN – COnsensus-based Standards for the selection of health Measurement INstruments; GRADE – Grading of Recommendations, Assessment, Development and Evaluation; QAPAQ – Quality Appraisal for Physical Activity Questionnaires; OMERACT – Outcome Measures in Rheumatology; QUADAS-2 – Quality Assessment of Diagnostic Accuracy Studies-2; α – Cronbach's alpha; ICC – Intraclass Correlation Coefficient; κ – Cohen's kappa; SEM – Standard Error of Measurement; SDC – Smallest Detectable Change; LoA – Limits of Agreement; MDC – Minimal Detectable Change; AUC – Area Under the Curve.
Legend– = property not evaluated.Partially = Descriptive in nature; property discussed narratively without a formal scoring system or threshold.

the reviews offered descriptive feasibility comments.

3.4. Results of individual sources

Tables 1 and 2 summarise the adequacy of each psychometric domain for each identified measures, while Table 3 presents the appraisal framework utilized in each study; below we highlight how these data answer the review question.

Reliability. All 16 reviews extracted at least one reliability index. Pooled Intraclass Correlation Coefficients (ICCs) for multi-item patient-reported scales (WOMAC, KOOS, HOOS, LEFS) and performance tests (6MWT, 40-m walk, ALF, TUG) consistently exceeded 0.80, indicating good–excellent reproducibility. Single-item activity scales (e.g. ARS, Tegner) showed fair reliability and large measurement error in older or post-arthroplasty samples.

Validity. Construct or criterion validity was examined in all studies. Strongest support was found for WOMAC, KOOS, HOOS and LEFS (hypothesis-testing or factor-analytic confirmation ≥75%); moderate–weak or inconsistent evidence was reported for HOOS-PS, KOOS-PS, PASE and many balance or postural-control tests. For performance measures, convergent validity with disease severity or gait speed was high for the 40-m fast-paced walk, 6MWT and chair-rise tests, but variable for stair-climb and force-plate tasks.

Feasibility. Thirteen studies provided descriptive information while three gave no feasibility details. Instruments judged easiest to administer included the UCLA and Tegner scales (≤1 min), WOMAC (≤5 min, free), and walk tests requiring minimal equipment. Devices such as force plates and 3D gait analysis were flagged as resource-intensive and less practical for routine use.

3.5. Synthesis of findings

Taken together, the evidence base supports the WOMAC, KOOS, HOOS, TUG, 6-Minute Walk Test, and 40-m walk as the most psychometrically robust and feasible measures for older adults with hip or knee OA; these conclusions are based on high-methodological-quality reviews that applied validated appraisal frameworks (COSMIN or OMERACT). However, no instrument satisfied reliability, validity, and feasibility at a uniformly high standard, underscoring the need for further work, particularly the systematic, criterion-based evaluation of feasibility. Collectively, the findings highlight several well-supported outcome measures for older adults with hip or knee OA, while also exposing major gaps in feasibility reporting and exercise-specific validation. These gaps frame key priorities for interpretation and future research.

4. Discussion

4.1. Summary of evidence

Across 16 secondary reviews, this scoping review catalogued 102 outcome instruments (35 self-report scales and 67 performance-based tests) relevant to adults ≥50 years with hip or knee OA. All sources assessed reliability and validity, using a range of frameworks including COSMIN and narrative appraisal approaches, yet none applied a formal feasibility rubric. Consequently, no single measure satisfied all three psychometric domains. High-quality evidence converged on the WOMAC, KOOS, HOOS, and performance-based tests such as the TUG, 40-m walk, and 6MWT as the most robust and clinically workable tools for this population.

While these findings confirm that several widely used measures have strong psychometric support, the lack of feasibility evidence and absence of validation in exercise contexts limit their clinical applicability for use in this population. Only three studies referenced exercise, and none evaluated outcome measures within supervised programs. This highlights a disconnect between research and implementation, where psychometrically strong tools remain untested in the setting of supervised exercise programs, where outcome monitoring is critical for adults ≥50 years with hip or knee OA.

4.2. Interpretation and comparison with previous work

Our findings accord with recent COSMIN-based syntheses, which confirm excellent test-retest agreement (ICC ≥0.80) for multi-item patient-reported scales (eg., WOMAC, HOOS, KOOS, LEFS) and identify persistent weaknesses in feasibility reporting (Prinsen et al., 2018; Lundgren-Nilsson et al., 2018). The psychometric strength of multi-item scales like the HOOS and KOOS is further supported by reviews that included patients awaiting or having undergone TJA. Specifically, the HOOS demonstrated strong evidence for use in those undergoing total hip arthroplasty, while the KOOS was found to be a feasible and valid tool for monitoring patients through total knee arthroplasty (Braaksma et al., 2020; Kivlan and Martin, 2013; Thorborg et al., 2010). In contrast with earlier narrative summaries that emphasized biomechanical laboratory measures, the present review highlights consistent psychometric support for simple, clinic-based walk and chair-stand tests (eg., TUG, 6MWT, 40-m fast-paced walk) which are readily deployable in physiotherapy practice (Choursiya et al., 2024; Alghadir et al., 2015; Ho-Henriksson et al., 2025). In relation to our findings, several single-item physical-activity indices (eg., ARS, Tegner) have previously demonstrated fair reliability, limited validity, and pronounced floor effects in older or post-arthroplasty populations, again questioning their clinical utility in these populations (Swanenburg et al., 2014; Mørup-Petersen et al., 2021; Naal et al., 2009). Importantly, no included reviews assessed these outcome measures within supervised exercise programs, a key implementation setting, underscoring an ongoing gap in feasibility and clinical applicability. This disconnect between measurement strength and real-world applicability reinforces the need for outcome evaluations that account for both psychometric rigour and practical use in OA rehabilitation.

4.3. Clinical implications

For clinicians monitoring outcomes in adults ≥50 years with hip or knee osteoarthritis, the present synthesis suggests the following. Choose multi-item self-report scales (e.g., KOOS, HOOS, WOMAC) when pain and function are primary targets, recognising that robust feasibility data remain inconsistent. Accurate assessment of outcomes in hip and knee OA is essential for guiding effective care. As emphasized by Osteoarthritis Research Society International (OARSI) guidelines (Zhang et al., 2008), self-report measures are key to evaluating the effects of interventions. These tools capture patients' subjective experiences that objective tests may overlook and support care that aligns with patients lived experiences. Secondly, adopt short, pragmatic performance tests (TUG, 40-m fast-paced walk, 30-s chair-stand, 6MWT) which combine acceptable psychometric performance with minimal equipment demands, making them suitable for routine progress monitoring. Performance-based measures offer objective data on physical function that may not be captured by self-report tools, enabling a more accurate and comprehensive assessment of an individual's functional capacity and supporting informed clinical decision-making (Losina et al., 2020).

Taken together, clinicians can realistically prioritise WOMAC, KOOS, HOOS, TUG, 6MWT, and 40-m fast-paced walk in current practice, as these measures are consistently supported by strong reliability and validity evidence and are feasible to administer in typical physiotherapy settings. However, because formal feasibility data are lacking and none of these measures have been validated specifically within supervised exercise programs, instrument choice should also consider contextual factors such as clinic space, staff availability, and patient mobility. Until further context-specific validation becomes available, these tools remain the most practical evidence-informed options for monitoring outcomes in older adults with hip and knee OA.

4.4. Strengths and limitations

Strengths include adherence to PRISMA-ScR methodology, duplicate screening, and a comprehensive mapping of both self-report and performance-based measures for hip and knee OA. Limitations include the reliance on secondary evidence, which introduces the potential for inherited bias and error, may not reflect the most up-to-date primary research, and encompasses variability in the methodological quality of included reviews. Grey literature and non-English sources were not included, which may have led to the exclusion of relevant instruments. Additionally, the inclusion of literature encompassing TJA populations alongside conservative OA management may introduce population heterogeneity. Although many outcome measures are used interchangeably across these groups, their functional demands and exercise requirements in a supervised setting can differ significantly, which may impact the clinical applicability of the findings for post-surgical versus non-surgical patients. Furthermore, data extraction was conducted by one reviewer, and while single-reviewer extraction may increase risk of oversight, we mitigated this by using piloted tools developed collaboratively and informed by consensus from prior dual-reviewer stages. Finally, feasibility-related conclusions are limited by the absence of objective criteria, including administration-time thresholds or user-acceptability metrics, in the included reviews.

5. Conclusion

Reliable and valid tools to measure symptoms and function in individuals with hip or knee osteoarthritis, including those awaiting or having undergone TJA, are widely available; however, formal feasibility evidence remains notably lacking. Until future research integrates structured feasibility assessments, specifically within the context of supervised exercise, clinicians will need to weigh psychometric strength against practical constraints such as time, space, and patient mobility when selecting outcome measures. To advance toward a core set of clinically relevant tools, future research should prioritise systematic feasibility evaluations using appropriate frameworks (e.g., COSMIN) and validate outcome measures specifically within exercise-based interventions. In the meantime, clinicians are advised to prioritise validated tools such as WOMAC, KOOS, HOOS, TUG, 6MWT, and the 40-m fast-paced walk, while considering local contextual factors, as these represent the most practical evidence-informed options currently available.

Appendix A

Search Strategy

OVID Medline.
Search conducted on July 18, 2024.

Search	Filter	Query	Records retrieved
1	Population	(Osteoarthritis/OR osteoarthriti$.tw OR osteoarthros$.tw OR (degenerative adj (joint or arthriti$)).tw OR (arthritis adj1 noninflammatory).tw OR (arthrosis or arthroses)) AND (exp Hip/OR exp Hip Joints/OR exp Knee/OR exp Knee Joints/)	23,113
2	Study type	exp clinical pathway/OR exp clinical protocol/OR clinical protocols/OR exp guideline/OR guidelines as topic/OR exp practice guideline/OR practice guidelines as topic/OR Clinical Decision Rules/OR (guideline or practice guideline).pt OR (practice parameter* or best practice*).ti,ab,kf OR (standards or guideline or guidelines).ti,kf OR ((practice or treatment* or clinical) adj guideline*).ab OR (CPG or CPGs).ti OR ((critical or clinical or practice) adj2 (path or paths or pathway or pathways or protocol*)).ti,ab,kf OR recommendat*.ti,kf. OR guideline recommendation*.ab OR (systematic review or meta-analysis).pt OR meta-analysis/OR systematic review/OR systematic reviews as topic/OR meta-analysis as topic/OR "meta analysis (topic)"/OR "systematic review (topic)"/OR ((systematic* adj3 (review* or overview*)) or (methodologic* adj3 (review* or overview*))).ti,ab,kf OR (meta-analy* or metaanaly* or meta analy* or systematic review*).mp,hw	1,138,368
3	Metrics	exp Psychometrics/OR psychometr*.ti,ab OR exp "Outcome Assessment (Health Care)"/OR outcome assessment.ti,ab OR exp Health Status Indicators/OR reproducib*.ti,ab OR (reliab* or unreliab* or valid* or coefficient or homogeneity or homogeneous or "internal	10,349,775

(*continued on next page*)

CRediT authorship contribution statement

Nikole Watson: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Conceptualization. **Rosmary Martinez Rueda:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing. **Hayley Legg:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – review & editing. **Catherine Boden:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing. **Brenna Bath:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Ethics statement

Ethical approval was not required for this study as it is a scoping review of previously published literature and did not involve human participants, identifiable data, or animals.

Funding

Funding support was provided to Nikole Watson through Mitacs Inc. for a Business Strategy Internship. Mitacs had no involvement in the design, conduct, analysis, interpretation, or reporting of this scoping review.

Declaration of interest statement

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

Acknowledgements

Funding was provided through a Business Strategy Internship awarded to Nikole Watson via Mitacs (Canada), in partnership with the University of Saskatchewan (Canada) and Craven SPORT services (Canada). Mitacs had no involvement in any aspect of the scoping review, including its design, conduct, analysis, interpretation, or reporting.

(*continued*)

Search	Filter	Query	Records retrieved
		consistency").ti,ab OR (feasib* or acceptab*).ti,ab OR (cronbach* and (alpha or alphas)).ti,ab OR (item and (correlation* or selection* or reduction*)).ti,ab OR (agreement or precision or imprecision or "precise values" or test-retest).ti,ab OR (test and retest).ti,ab OR (reliab* and (test or retest)).ti,ab OR (stability or interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intraobserver or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interassay or interassay or intraassay or intra-assay or interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant or kappa or kappa's or kappas or repeatab*).ti,ab OR ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).ti,ab OR (generaliza* or generalisa* or concordance).ti,ab OR (intraclass and correlation*).ti,ab OR (discriminative or "known group" or factor analysis or factor analyses or dimension* or subscale*).ti,ab OR (multitrait and scaling and (analysis or analyses)).ti,ab OR (item discriminant or interscale correlation* or error or errors or "individual variability").ti,ab OR (variability and (analysis or values)).ti,ab OR (uncertainty and (measurement or measuring)).ti,ab OR ("standard error of measurement" or sensitiv* or responsive*).ti,ab OR ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).ti,ab OR (small* and (real or detectable) and (change or difference)).ti,ab OR (meaningful change or "ceiling effect" or "floor effect").ti,ab OR exp validation study/OR (pain or mobility or stiffness or "range of motion" or strength or "muscle strength" or balance or walk* or gait or "weight-bearing" or "weight bearing" or "physical function" or "quality of life" or confidence or experience or perception).ti,ab. OR Quality of life/	
4		1 AND 2 AND 3	872

Search Strategy Note regarding Population Scope: While the boolean search string focused on primary osteoarthritis terms (e.g., Osteoarthritis, Degenerative Joint Disease) to capture the broad clinical context, the review's population scope explicitly included studies evaluating these metrics in total joint arthroplasty cohorts. This is consistent with the clinical application of OA-specific metrics (e.g., HOOS, KOOS) in perioperative surgical management.

Appendix B. Descriptive characteristics of the included studies

Author	Evidence	Setting	Joint (s)	Outcomes Assessed	Measures	Exercise	Findings
Howe et al., 2012	Systematic review	Community-based clinical practice	Knee	Mobility, pain, physical function	ADLs, ALF, AQoL, Goniometer, HAP, ICOAP, J-MAP, KSS, KOOS, LEFS, LISOK, OKS, SF-36, Stair climb, TUG, Walk, WOMAC	–	Suitable for use in MSK knee populations: AAOS, goniometer, KOOS, LEFS, WOMAC.
Ornetti et al., 2010	Systematic review	–	Hip Knee	Mobility, gait, walking knee/ hip ranges	3D gait analysis	–	Insufficient data concerning the validity and reliability of kinematic parameters as a measure for OA.
Smith et al., 2019	Systematic review	Public, community, clinical	Hip Knee Feet Hands	Physical activity	ARS, Baecke, HAP, IPAQ-SF, PASE, UCLA, Tegner	Leisure-time physical activity, sports, gym, walking. No structured exercise	Lack of evidence of adequate measurement properties. Not clear which is the most appropriate for OA. Most self-report instruments demonstrated adequate test-retest reliability, quality ranged from poor – excellent
Myers and Bohannon, 2020	Integrative review	–	Knee	Isokinetic strength	Isokinetic dynamometry	Strength, aerobic, balance, power, isokinetic exercise, stretching, hydro	Isokinetic dynamometry is valid and reliable for assessing knee flexor/extensor strength in knee OA and useful for identifying impairments and tracking intervention effects.
Collins et al., 2016	Systematic review Meta-analysis	Public, clinical	Knee	KOOS domains	KOOS	–	KOOS can be used in older patients with knee conditions with adequate internal consistency, test-retest reliability, and construct validity. KOOS-PS requires further evaluation.
Dobson et al., 2012	Systematic review	–	Hip Knee	Walking, chair stand, stairs, multi-activity	ALF, Chair stand, FAS, GUG, Lin, PAR, Stair climb, Steultjens, Stratford, TUG, Walk	–	Best evidence for walk/STS tests: 40m self-paced (hip), 50 ft fast-paced (hip/knee), 30s chair stand + TUG (hip/knee). Limited evidence for stair tests. Best evidence for multi-activity tests: PAR, Stratford, FAS. GUG + Steultjen's not recommended for hip/knee OA.
Chen et al., 2022	Systematic review	In clinic, on-site	Knee	Walking, stairs, chair stand, balance, single-step, multi-activity	ALF, Chair stand, GUG, ILAS, L-test, PAR, PerF, SEBT, SPPB, Stair climb, Step, Timed assessment, TUG, Walk	Exercise was mentioned in some studies, but type was not specified	Best evidence for physical function tools for knee OA: Walk 40 m fast paced, 6MWT, TUG, and chair rise over 30s, PerF and PAR.
Kivlan and Martin, 2013	Literature review	–	Hip	Instrument domains	AAOS-LLCS, AIMS2, HOOS, LISOH, OAKHQOL, OHS, WOMAC	–	The HOOS has the strongest evidence for OA or those undergoing THA - easy to

(*continued on next page*)

(*continued*)

Author	Evidence	Setting	Joint (s)	Outcomes Assessed	Measures	Exercise	Findings
							complete and calculate but time intensive.
Terwee et al., 2006	Systematic review	–	Hip Knee	Physical function, gait, stairs, chair stand, aerobic fitness	Accelerometry, ALF, ALF Madsen, ALF Marks, DynaPort KneeTest, FAS, Footprint analysis, Gait analysis, GUG, ILAS, Lin, Locométrix, Numact monitor, PAR, Pedometer, Stair climb, Steultjens, Walk	–	Cannot identify which of the 26 tests would be the most useful as consensus is lacking. Data is lacking on reproducibility and validity.
French et al., 2020	Systematic review	Clinical, laboratory setting	Knee	Balance, mobility, postural control, stability, single-leg stance	CB&M, Force plates, POMA, SEBT	–	Due to poor-quality studies, psychometric properties of several tools remain unclear. SLS is not suitable for knee OA. CB&M is the preferred tool based on available psychometric data.
Braaksma et al., 2020	Systematic review Meta-analysis	–	Hip Knee	Instrument domains	HOOS-PS KOOS-PS	–	Both measures lacked content and construct validity, with HOOS-PS and KOOS-PS potentially failing to accurately reflect physical function in individuals with hip/knee issues or undergoing TJA.
Peer and Lane, 2013	Systematic review	–	Knee	Instrument domains	KOOS KOOS-PS	–	The sport/rec subscale demonstrated weak-to-moderate reliability and weak construct validity. There is some evidence that the KOOS and KOOS-PS are feasible for research and clinical use.
Moore and Barker, 2017	Systematic review	Clinical	Hip	Step test	FSST	–	FSST is comparable to TUG and may help assess fall risk in arthritis populations. In hip OA, FSST showed strong inter-rater reliability, but weaker test-retest reliability compared to the step test.
Thorborg et al., 2010	Systematic review	–	Hip	Instrument domains	AAOS-HS, HOOS, LISOH, OHS, PASI, WOMAC	–	HOOS has adequate psychometric properties when assessing hip OA patients undergoing conservative treatment or THR.
Samuel and Kanimozhi, 2019	Literature review	–	Knee	Instrument domains	AIMS/AIMS2, CB&M, COAT, IKHOAM, LISOK, KOFBeQ, KOOS, KSSS, OKS, SF-36, TLKSS, WOMAC	–	OKS and WOMAC have excellent reliability and good validity.
Terwee et al., 2011	Systematic review	Clinical, community	Hip Knee	Walking Instrument domains	ARS, Baecke, DAQ, HAP, IPAQ-SF, LEAS, Pedometer, TAS, UCLA	–	The studies were of poor to moderate quality and no instrument received positive ratings for all properties. Inconclusive evidence.

During the preparation of this work the author NW used ChatGPT to adjust the word count of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of this publication.

References

Alghadir, A., Anwer, S., Brismée, J.M., 2015. The reliability and minimal detectable change of timed up and Go test in individuals with grade 1 – 3 knee osteoarthritis. BMC Muscoskelet. Disord. 16 (1), 174.

Arksey, H., O'Malley, L., 2005. Scoping studies: towards a methodological framework. Int. J. Soc. Res. Methodol. 8 (1), 19–32.

Bellamy, N., 2005. The WOMAC knee and hip osteoarthritis indices: development, validation, globalization and influence on the development of the AUSCAN hand osteoarthritis indices. Clin. Exp. Rheumatol. 23 (5 Suppl. 39), S148–S153.

Braaksma, C., Wolterbeek, N., Veen, M.R., Prinsen, C.A.C., Ostelo, R.W.J.G., 2020. Systematic review and meta-analysis of measurement properties of the hip disability and osteoarthritis outcome score - Physical function shortform (HOOS-PS) and the knee injury and osteoarthritis outcome score - Physical function shortform (KOOS-PS). Osteoarthr. Cartil. 28 (12), 1525–1538.

Chen, H., Wang, C., Wu, J., Wang, M., Wang, S., Wang, X., et al., 2022. Measurement properties of performance-based measures to assess physical function in knee osteoarthritis: a systematic review. Clin. Rehabil. 36 (11), 1489–1511.

Choursiya, P., Veqar, Z., Khan, Z., Tanwar, T., Iram, I., Aldabbas, M., 2024. Gait analysis technologies for measurement of biomechanical parameters of knee osteoarthritis. SN Compr Clin Med. 6 (1), 6.

Collins, N.J., Prinsen, C.A.C., Christensen, R., Bartels, E.M., Terwee, C.B., Roos, E.M., 2016. Knee injury and osteoarthritis outcome score (KOOS): systematic review and meta-analysis of measurement properties. Osteoarthr. Cartil. 24 (8), 1317–1329.

Colquhoun, H.L., Levac, D., O'Brien, K.K., Straus, S., Tricco, A.C., Perrier, L., et al., 2014. Scoping reviews: time for clarity in definition, methods, and reporting. J. Clin. Epidemiol. 67 (12), 1291–1294.

Dobson, F., Hinman, R.S., Hall, M., Terwee, C.B., Roos, E.M., Bennell, K.L., 2012. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthr. Cartil./OARS, Osteoarthritis Research Society 20 (12), 1548–1562.

French, H.P., Hager, C.K., Venience, A., Fagan, R., Meldrum, D., 2020. Psychometric properties and domains of postural control tests for individuals with knee osteoarthritis: a systematic review. International journal of rehabilitation research Internationale Zeitschrift fur Rehabilitationsforschung Revue internationale de recherches de readaptation 43 (2), 102–115.

Ho-Henriksson, C.M., Thorstensson, C.A., Nordeman, L., 2025. Self-assessment using 30-second chair stand test for patients with knee osteoarthritis – an intra- and inter-rater reliability study. European Journal of Physiotherapy 27 (2), 65–71.

Howe, T.E., Dawson, L.J., Syme, G., Duncan, L., Reid, J., 2012. Evaluation of outcome measures for use in clinical practice for adults with musculoskeletal conditions of the knee: a systematic review. Man. Ther. 17 (2), 100–118.

Jette, D.U., Bacon, K., Batty, C., Carlson, M., Ferland, A., Hemingway, R.D., et al., 2003. Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists. Phys. Ther. 83 (9), 786–805.
Jette, D.U., Halbert, J., Iverson, C., Miceli, E., Shah, P., 2009. Use of standardized outcome measures in physical therapist practice: perceptions and applications. Phys. Ther. 89 (2), 125–135.
Kivlan, B.R., Martin, R., 2013. Outcome instruments for the hip: a guide to implementation. Orthopaedic Physical Therapy Practice 25 (3), 162–169.
Kolasinski, S.L., Neogi, T., Hochberg, M.C., Oatis, C., Guyatt, G., Block, J., et al., 2019. American college of rheumatology/arthritis foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Rheumatol. 72 (2), 220–233, 2020 Feb.
Levac, D., Colquhoun, H., O'Brien, K.K., 2010. Scoping studies: advancing the methodology. Implementation Sci 5 (69).
Long, H., Liu, Q., Yin, H., Wang, K., Diao, N., Zhang, Y., et al., 2022. Prevalence trends of site-specific osteoarthritis from 1990 to 2019: findings from the global burden of disease study 2019. Arthritis Rheumatol. 74 (7), 1172–1183.
Losina, E., MacFarlane, L., Song, S., Selzer, F., Bronsther, C., Huizinga, J., et al., 2020. Objective performance tests assess aspects of function not captured by self-report in knee osteoarthritis. Osteoarthr. Cartil. 28, S453.
Lundgren-Nilsson, Å., Dencker, A., Palstam, A., Person, G., Horton, M.C., Escorpizo, R., et al., 2018. Patient-reported outcome measures in osteoarthritis: a systematic search and review of their use and psychometric properties. RMD Open 4 (2), e000715.
Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., et al., 2010. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J. Clin. Epidemiol. 63 (7), 737–745.
Moore, M., Barker, K., 2017. The validity and reliability of the four square step test in different adult populations: a systematic review. Syst. Rev. 6 (1), 187.
Mørup-Petersen, A., Skou, S.T., Holm, C.E., Holm, P.M., Varnum, C., Krogsgaard, M.R., et al., 2021. Measurement properties of UCLA activity scale for hip and knee arthroplasty patients and translation and cultural adaptation into Danish. Acta Orthop. 92 (6), 681–688.
Myers, B.J., Bohannon, R., 2020. Isokinetic testing of muscle strength in older adults with knee osteoarthritis: an integrative review. Isokinet. Exerc. Sci. 28 (3), 269–290.
Naal, F.D., Impellizzeri, F.M., Leunig, M., 2009. Which is the best activity rating scale for patients undergoing total joint arthroplasty? Clin. Orthop. Relat. Res. 467 (4), 958–965.
Nilsdotter, A.K., Lohmander, L.S., Klässbo, M., Roos, E.M., 2003. Hip disability and osteoarthritis outcome score (HOOS)–validity and responsiveness in total hip replacement. BMC Muscoskelet. Disord. 4, 10.
Ornetti, P., Maillefert, J.F., Laroche, D., Morisset, C., Dougados, M., Gossec, L., 2010. Gait analysis as a quantifiable outcome measure in hip or knee osteoarthritis: a systematic review. Jt. Bone Spine : revue du rhumatisme 77 (5), 421–425.
Papalia, R., Campi, S., Vorini, F., Zampogna, B., Vasta, S., Papalia, G., et al., 2020. The role of physical activity and rehabilitation following hip and knee arthroplasty in the elderly. JCM 9 (5), 1401.
Peer, M.A., Lane, J., 2013. The knee injury and osteoarthritis outcome score (KOOS): a review of its psychometric properties in people undergoing total knee arthroplasty. J. Orthop. Sports Phys. Ther. 43 (1), 20–28.
Peters, M.D.J., Godfrey, C., McInerney, P., Khalil, H., Larsen, P., Marnie, C., et al., 2022. Best practice guidance and reporting items for the development of scoping review protocols. JBI Evidence Synthesis 20 (4), 953–968.
Prinsen, C.A.C., Mokkink, L.B., Bouter, L.M., Alonso, J., Patrick, D.L., De Vet, H.C.W., et al., 2018. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual. Life Res. 27 (5), 1147–1157.
Roos, E.M., Lohmander, L.S., 2003. The knee injury and osteoarthritis outcome score (KOOS): from joint injury to osteoarthritis. Health Qual. Life Outcome 1, 64.
Samuel, A.J., Kanimozhi, D., 2019. Outcome measures used in patient with knee osteoarthritis: with special importance on functional outcome measures. Int. J. Health Sci. 13 (1), 52–60.
Skou, S.T., Roos, E.M., 2017. Good life with osteoArthritis in Denmark (GLA:D™): evidence-based education and supervised neuromuscular exercise delivered by certified physiotherapists nationwide. BMC Muscoskelet. Disord. 18 (1), 72.
Smith, R.D., Dziedzic, K.S., Quicke, J.G., Holden, M.A., McHugh, G.A., Healey, E.L., 2019. Identification and evaluation of self-reported physical activity instruments in adults with osteoarthritis: a systematic review. Arthritis Care Res. 71 (2), 237–251.
Sørensen, N.L., Hammeken, L.H., Thomsen, J.L., Ehlers, L.H., 2019. Implementing patient-reported outcomes in clinical decision-making within knee and hip osteoarthritis: an explorative review. BMC Muscoskelet. Disord. 20 (1), 230.
Bone and Joint Canada, 2024. Osteoarthritis. Available from: https://boneandjointcanada.com/osteoarthritis/.
Osteoarthritis Action Alliance, 2024. Comorbidities and Co-Occurring Symptoms. Available from: https://oaaction.unc.edu/oa-module/comorbidities-and-co-occuring-symptoms/.
Regino, J., Versales, C.R., Catalan, J.K., Bajado, E.M., De Jesus, M.R., Hernandez, O., et al., 2020. Clinical audit on outcome measures used in examination of knee osteoarthritis in selected hospitals and clinics in the Philippines. PJAHS 4 (1). Available from: https://pjahs.ust.edu.ph/wp-content/uploads/2020/08/8-Regino2020.pdf.
Teo, P.L., Hinman, R.S., Egerton, T., Dziedzic, K.S., Bennell, K.L., 2019. Identifying and prioritizing clinical guideline recommendations Most relevant to physical therapy practice for hip And/Or knee osteoarthritis. J. Orthop. Sports Phys. Ther. 49 (7), 501–512.
Terwee, C.B.M.L., Steultjens, M.P.M., Dekker, J., 2006. Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. Rheumatology 45, 890–902.
Terwee, C.B.B.W., Elsland, SLv, Vet, HCWd, Dekker, J., 2011. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. Osteoarthritis and cartilage/OARS. Osteoarthritis Research Society 19, 620–633.
Thorborg, K., Roos, E.M., Bartels, E.M., Petersen, J., Holmich, P., 2010. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. BJSM online 44 (16), 1186–1196.
Verheyden, G., Meyer, S., 2016. 'To Measure is to Know.' advancing the use of outcome measures in the physiotherapy profession. Physiother. Res. Int. 21 (1), 1–3.
Vos, T., Flaxman, A.D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., et al., 2012. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the global burden of disease study 2010. Lancet 380 (9859), 2163–2196.
Watson, N., Martinez Rueda, R., Bath, B., Legg, H., Boden, C., 2024. Identifying reliable, valid and feasible outcome measures for adults aged 50 years or older with hip or knee osteoarthritis participating in supervised exercise programs. A Scoping Review Protocol, 45377 Bytes.
Woolacott, N.F., Corbett, M.S., Rice, S.J.C., 2012. The use and reporting of WOMAC in the assessment of the benefit of physical therapies for the pain of osteoarthritis of the knee: findings from a systematic review of clinical trials. Rheumatology 51 (8), 1440–1446.
Wylde, V., Learmonth, I.D., Cavendish, V.J., 2005. The Oxford hip score: the patient's perspective. Health Qual. Life Outcome 3 (1), 66.
Zhang, W., Moskowitz, R.W., Nuki, G., Abramson, S., Altman, R.D., Arden, N., et al., 2008. OARSI recommendations for the management of hip and knee osteoarthritis, part II: OARSI evidence-based, expert consensus guidelines. Osteoarthr. Cartil. 16 (2), 137–162.
Zotero, 2024. Roy Rosenzweig Center for History and New Media, George Mason University: Corporation for Digital Scholarship. Available from: https://www.zotero.org/.
Swanenburg, J., Koch, P., N., Wirth, B., 2014. Function and activity in patients with knee arthroplasty: validity and reliability of a German version of the Lysholm Score and the Tegner Activity Scale. Swiss Med Wkly 144 (w13976). Available from: https://smw.ch/index.php/smw/article/view/1875.