Knee-related patient-reported outcome measures for young adults: A scoping meta-review

Nicole Bausch[1], Emma Eyre[2], Gemma Pearce[2], Shea Palmer[3]

[1] Centre for Care Excellence, University Hospitals Coventry and Warwickshire NHS Trust, Coventry, United Kingdom
[2] Centre for Healthcare and Communities, Coventry University, Coventry, United Kingdom
[3] School of Healthcare Sciences, Cardiff University, Cardiff, United Kingdom

Correspondence
Nicole Bausch, Centre for Care Excellence, University Hospitals Coventry and Warwickshire NHS Trust, Clifford Bridge Road, Coventry, CV2 2DX, United Kingdom
Email: Nicole.Bausch@uhcw.nhs.uk
ORCID iD: https://orcid.org/0000-0002-5787-170X

Acknowledgments/Special Thanks: Dr Helen Parsons, Warwick Clinical Trials Unit, University of Warwick, Coventry, United Kingdom

Funding information
This work was supported by a HEE/NIHR ICA Pre-doctoral Clinical and Practitioner Academic Fellowship award for the lead author (NB) (NIHR302724)

Conflict of Interest
None

Background

There is a lack of patient-reported outcome measures (PROMs) research on young adults and knee disorders. This scoping meta-review examined a young adult population and aimed to (1) provide an overview of knee-related PROMs research and (2) evaluate the measurement properties of the five most evaluated knee-related PROMs relevant for individual care and group-level analysis.

Methods

A systematic search of the PubMed and COSMIN databases was conducted on 18 September 2023 and updated on 25 November 2024 to identify systematic reviews of knee-related PROMs in young adults. Data relevant to individual care and group-level analysis of the five most evaluated PROMs were extracted based on the PROM-cycle and analysed guided by COSMIN recommendations.

Results

Fifteen systematic reviews were included, evaluating 80 knee-related PROMs. Ten of the 15 systematic reviews did neither use a tool to synthesise multiple studies nor a PROM evaluation tool. Knee Injury and Osteoarthritis Outcome Score (KOOS) had the strongest evidence to be an appropriate PROM for individual care and Knee Outcome Survey Activities of Daily Living Scale (KOS-ADLS) for group-level analysis in a young adult population. However, none of the five PROMs demonstrated sufficient high-quality evidence across all identified measurement properties.

Conclusion

The scoping meta-review highlights that systemic reviews on measurement properties were reported inconsistently, making it challenging to detangle the extracted data. Therefore, advances in PROMs-specific methods and reporting recommendations should enhance the quality of PROM evidence, allowing readers to appraise relevant evidence and select the most appropriate PROMs for their intended purpose.

1 INTRODUCTION

The knee is the second most common region of complaint (after back pain) in young adults between 25 and 45 years of age. It is the most common lower limb area for all adult age groups to consult their general practitioner in England (Jordan et al., 2010). Young adults experience many knee disorders, including traumatic ligament, osteochondral and meniscal injuries or tendon and bursa overuse pain syndromes (Bunt et al., 2018). In the United Kingdom, extensive work and strategic planning on a national level has taken place to identify patient-reported outcome measures (PROMs) to understand pain and function following anterior cruciate ligament injuries in young people (National Ligament Registry, n.d.). However, using PROMs to understand health outcomes and explore the effectiveness of interventions for the broad spectrum of knee disorders in young adults is less well-researched. There needs to be more understanding about which PROMs are available and should be collected in a young, active population experiencing knee symptoms.

PROMs can be generally categorised as 'disease-specific' when they can be used only with certain knee disorders, e.g. knee osteoarthritis (OA); 'condition-specific' when they can be used to assess symptoms or function in a variety of knee disorders; or 'generic' when they can be used across various medical fields, e.g. orthopaedics, rheumatology or general population samples (Higgins et al., 2023; Meadows, 2011, p.147). All three PROM categories are valuable when considering this unmet need for knee disorders in young adults. Therefore, the all-inclusive term 'knee-related PROMs' is used throughout this review. As age increases beyond 45 years of age, clinical features such as pain, swelling and stiffness can lead to the diagnosis of OA (NICE, 2022), which is the most common reason for a knee replacement.

Therefore, when possible, the search was limited to a population under 45 years of age and excluded condition-specific PROMs evaluating knee OA and replacement from this review to ensure the focus on a young adult population.

PROMs are commonly used in scientific research, such as clinical trials. Increasingly, PROMs are used in routine clinical practice to inform diagnosis, monitor short-term and long-term outcomes (Nelson et al., 2015; Cumberledge, 2020, p.52), facilitate communication between patients and clinicians, or inform patient-centred care (Meadows, 2011; Aaronson et al., 2015). Depending on the purpose and target group, different measurement properties are important when considering using PROMs in clinical or research settings (van der Wees et al., 2019). What needs to be understood is how appropriate knee-related PROMs are for a young, adult population for individual care and group-level analysis.

A preliminary search of the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) database relating to knees was conducted in September 2023. The search identified 65 records concerning self-reported knee questionnaires, unlike other patient-reported methods such as interviews (9 records) or diaries (2 records) on this database, which gave attention to systematic reviews of outcome measurement instruments. As an adequate number of available systematic reviews was crucial for undertaking a scoping meta-review, this current review focuses solely on questionnaires.

This scoping meta-review, therefore, aimed to (1) provide an overview of knee-related PROMs research in a young adult population and (2) evaluate the measurement properties of the five most commonly evaluated knee-related PROMs in a young adult population relevant for individual care and group-level analysis.

2 METHODS

A meta-review of published systematic reviews was conducted to summarise evidence on measurement properties of knee-related questionnaires relevant to a young population and different purposes. Although rigorous attempts had been made to be as comprehensive as possible, practical limitations such as exploring the scope and availability of resources were considered. This meta-review was limited to searching two databases and evaluating the five most common knee-related PROMs and, therefore, a scoping meta-review in nature (Sarrami-Foroushani et al., 2015).

The scoping meta-review was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). A flow diagram illustrates the selection process (Page et al., 2021). Given that the PRISMA reporting guidelines did not focus on providing a rigorous methodology to evaluate measurement properties, the COSMIN Risk of Bias Checklist (Mokkink et al., 2018b) and COSMIN Study Design Checklist (Mokkink et al., 2019), which guided the analysis.

2.1 Search strategy

The databases were selected pragmatically consistent with a scoping approach (Sarrami-Foroushani et al., 2015), and a literature search of the COSMIN and PubMed databases was performed to identify systematic reviews that evaluated measurement properties of knee-related PROMs. The COSMIN search was limited to the search term ’knee’, and the filters ’adults’ and ’patient-reported outcome’ were applied. The PubMed database search combined Medical Subject Headings (MeSH) terms of the keywords ‘knee’ and ‘patient-reported’ combined with the Boolean operator ‘AND’. The PubMed search was narrowed by applying the filter ‘humans’, ‘adult: 19-44 years’ and ‘systematic reviews’ to ensure the focus on the target population. Records published in English without a set timeframe were included.

2.2 Selection criteria

Pre-specified criteria for the inclusion were systematic reviews that evaluated measurement properties of knee-related PROMs, included a young adult population in the validation study of the PROMs, and were validated in English. Systematic reviews focussing on cross-cultural measurement properties were excluded as these were validation studies of non-English PROMs. PROMs specifically evaluating knee outcomes associated with older adults (over 45 years of age, knee OA and knee replacement) and paediatric populations (under 18 years of age) were excluded to ensure a focus on a young adult population. The preliminary review indicated no consensus on a young adult's age range. Therefore, for this scoping meta-review, the searches were limited to the age range between 18 and 44 years of age on PubMed. In the absence of being able to limit the age range to the pre-specified age criteria, the filter ‘adult’ was selected, which the COSMIN database had defined as an age range between 18 to 65 years. However, due to the nature of analysing previously collected data, it was challenging to remain within this set age range and avoid excluding relevant literature; the decision was made to apply the filter for the COSMIN database search. Systematic reviews that were clearly not relevant to the target group of young adults were later excluded at the full-text review stage.

The search and screening of titles, abstracts, and full texts were conducted by the lead researcher (NB). The following pre-selected information was extracted onto a Microsoft Excel spreadsheet: first author's name, year of publication, systematic review, measurement properties, knee-related, young adult population and English language (Supplementary File 1). Consultation and verification of decisions with other authors were conducted at all stages of the review process.

2.3 Data extraction, quality assessment and analysis

To provide an overview of knee-related PROM research in a young adult population, the following information was extracted from the systematic reviews: general information (authors, country and year of publication), analysis method, PROM evaluation tool/evidence grading scale and pre-selected measurement properties. A narrative synthesis was conducted to analyse and summarise the data from multiple systematic reviews (Table 1).

For an in-depth descriptive analysis to explore the measurement properties relevant to individual and group levels, the five most commonly evaluated knee-related PROMs in a young adult population were selected pragmatically to be manageable consisting of a scoping approach and existing publications on the evaluation of measurement properties (Darwich et al., 2020; Sarrami-Foroushani et al., 2015; Shepard et al., 2023). To identify the five most evaluated knee-related PROMs in a young population, the frequency of how often individual PROMs were reported in the systematic reviews was documented (Supplementary File 2). The identification of measurement properties relevant for individual- and group-level analysis was based on the PROM-Cycle (van der Wees et al., 2019). The measurement properties development source, population and responsiveness (individual care), and internal consistency, measurement error and test-retest reliability (group-level analysis) were extracted. The COSMIN Risk of Bias Checklist (Mokkink et al., 2018b) and the COSMIN Study Design Checklist (Mokkink et al., 2019) guided the in-depth descriptive analysis of the six measurement properties. When systematic reviews did not provide any information on the validation sample or development source, the original reports of the development and validation study of the PROMs were retrieved either from the reference list of the included systematic reviews or manually searched to complete this analysis.

3 RESULTS

The literature search of the COSMIN and PubMed databases was performed on 18 September 2023 and updated on 25 November 2024. After removing duplicates, the search identified 139 records (see Figure 1). Following the title and abstract screening, 26 systematic reviews were retrieved and assessed for eligibility. Finally, 15 systematic reviews were included in the final scoping meta-review, and emerging themes relating to a broad overview and evaluation of measurement properties relevant to individual care and group analysis are described below.

3.1 Overview of knee-related PROMs used in a young adult population

The 15 included systematic reviews were published between 2010 and 2024 in six countries, with the United States (n=7) and the United Kingdom (n=3) being the most common. The remainder were published in Brazil, Canada, Denmark, Spain, and Turkey. A total of 80 knee-related PROMs were identified (Supplementary File 2).

The systematic reviews used various methods to analyse measurement properties, including descriptive and/or statistical analysis with/without synthesis of the evaluated studies. 8 out of the 15 systematic reviews used six PROMs evaluation tools to analyse measurement properties, which were the COSMIN Checklist (Terwee et al., 2012), COSMIN Grades of Recommendation, Assessment, Development and Evaluation (Mokkink et al., 2018a), the COSMIN Risk of Bias checklist (Mokkink et al., 2018b), COSMIN Methodology for assessing the content validity of PROMs (Terwee et al., 2018), the Minimal Important Difference Credibility Assessment Tool (Devji et al., 2020) and Level of Evidence (Schellingerhout et al., 2012). These evaluation tools consisted of different evidence grading scales, making comparison challenging. For example, measurement properties graded as: 'excellent', 'good', 'fair', 'poor', 'indeterminate' (Terwee et al., 2012), 'high', 'moderate', low', 'very low' (Mokkink et al., 2018a), or 'strong evidence of positive/negative result', 'moderate evidence of positive/negative result', 'limited evidence of a positive/negative result', 'conflicting evidence', 'unknown due to poor methodological quality' (Schellingerhout et al., 2012) (see Table 1). The authors reported various measurement properties, including validity, reliability, responsiveness, interpretability, floor/ceiling effects, tool development, and feasibility (Table 1). Despite having published a systematic review on measurement properties, 10 of the 15 systematic reviews did neither use a tool to synthesise multiple studies nor a PROM evaluation tool.

3.2 Measurement properties of the top five cited knee-related PROMS in a young adult population

The five PROMS most frequently included in the systematic reviews were the International Knee Documentation Committee Subjective Knee Form (IKDC-SKF), Lysholm Knee Score (LKS), Anterior Knee Pain Scale (AKPS), Knee Outcome Survey Activities of Daily Living Scale - 17 items (KOS-ADLS), and the Knee Injury and Osteoarthritis Outcome Score (KOOS) (Table 2; and Supplementary File 2). These PROMs were developed in Finland (n=1) (Kujula et al., 1993), Sweden (n=2) (Lysholm & Gillquist, 1982; Roos et al., 1998), and the United States (n=2) (Irrgang et al., 1998; Irrgang et al., 2001) and published between 1982 and 2001. All five PROMs were either disease- or condition-specific and are intended to assess outcomes of knee surgery (e.g. anterior cruciate ligament reconstruction) or to evaluate symptoms, function, or quality of life of a variety of knee-related disorders (e.g. patellofemoral pain syndrome or post-traumatic OA).

3.2.1 Measurement properties relevant to individual care

This section describes the findings of the identified five knee-related PROMs in a young adult population, focusing on evaluating development sources, analysed populations, and responsiveness evaluations. All five PROMs were reported to be developed by professionals (e.g., orthopaedic surgeons) or panels. However, little to no information was reported on the involvement of patients or the public. Only KOOS (Roos et al., 1998) reported including an expert panel and patients as part of the PROMs development, which has become a design requirement to assess content validity. Original validation studies included children and older adults; for example, the IKDC (Irrgang et al., 2001) and KOS-ADLS (Irrgang et al., 1998) validation studies included participants ranging from six to 100 and 12 to 76 years of age, respectively.

In general, the systematic reviews included appropriate patient numbers between n=42 and n=4,056 as part of the evaluations of the six measurement properties, which related to a very good design. Although systematic reviews included an average patient age between 22.7 and 48.8 years across the five PROMs (Table 2), the systematic reviews did not report necessary details about the populations' age to be confident that the focus of this meta-review was solely on a young adult population.

Only IKDC-SKF and KOS-ADLS were graded as having excellent and good responsiveness, respectively, by one systematic review (Marinho et al., 2020). However, authors using a different PROM evaluation tool reported that the quality of evidence was poor or low when evaluating responsiveness (Gagnier et al., 2018; Macri et al., 2022).

3.2.2 Measurement properties relevant to group-level analysis

This section describes the measurement properties relevant to group-level analysis: internal consistency, measurement error and test-retest reliability. Similarly to the individual care measurement properties, four different PROMs grading scales were used to synthesise their findings, resulting in conflicting evidence (Table 3). Overall, the authors of this systematic review reported their synthesis of the three measurement properties across the five PROMs as insufficient or indeterminate based on low-quality studies. Only the measurement property test-retest of IKDC, KOS-ADLS and AKPS was reported as sufficient, with only AKPS supported by high-quality evidence.

4 DISCUSSION

This scoping meta-review provides a broad overview and summarises the key findings of 15 systematic reviews that evaluate six measurement properties of knee-related PROMs in a young adult population. It presents measurement properties relevant to individual care and group-level analysis. KOOS included a multi-disciplinary expert panel and patients as part of the PROMs development, and the systematic reviews that analysed KOOS had adequate population numbers. Although the quality of evidence was low for the responsiveness evaluation, it was graded as 'sufficient' by a high-quality systematic review (Hoglund et al., 2023b). Although each of the five PROMs had limitations across the three measurement properties evaluations relevant to the group-level analysis, only AKPS was evaluated based on moderate- and high-quality evidence of internal consistency and test-retest reliability. It needs to be noted that the COSMIN Grades of Recommendation, Assessment, Development and Evaluation (Mokkink et al., 2018b) tool was specifically developed to combine the methodological quality of multiple studies on measurement properties and, therefore, is recommended when synthesising and reporting systematic reviews of PROMs. It is important to highlight that only Hoglund et al. (2023a; 2023b) conducted the synthesis using the recommended COSMIN Grades of Recommendation, Assessment, Development and Evaluation (Mokkink et al., 2018b) tool, and therefore, could be considered as providing the strongest evidence on measurement properties.

Overall, the review found that a broad range of measurement properties were analysed and synthesised using numerous methods, evaluation tools and grading scales, making it challenging to assimilate the findings and highlighting the complexity of PROMs research. The findings of this meta-review highlight advances in PROMs research and methodology when analysing and synthesising measurement properties. Future systematic reviews focusing on synthesising findings of multiple studies on measurement properties should report their findings based on PRISMA-COSMIN for outcome measurement instruments 2024 (Elsman et al., 2024). This new checklist could help authors report their results clearly and thoroughly and facilitate the reproducibility of research. High-quality evidence would also allow readers less familiar with this field to appraise the quality of measurement properties and choose the most appropriate PROMs for their needs.

The average age of the populations was between 22.7 and 48.8 years of age across the five PROMs, indicating the suitability to assess knee-related issues in a young adult population. The systematic reviews reported information inconsistently relevant to this review. In particular, the reporting of the population analysed was inadequate; either the mean age or age range was reported. The inconsistent reporting of age made it difficult to compare the literature. Some analysed populations also included children and older adults. It has been highlighted that children had difficulties comprehending certain concepts of KOOS and IKDC, which were originally validated for adults (Iversen et al., 2019). Therefore, using children in PROMs research might affect the PROM's validity when intended for adults.

Most of the PROMs (and subsequent modifications and revisions) were developed with clinicians such as surgeons or physiotherapists. However, who was included in development panels and committees was not always clear. Only the KOOS development study reported the involvement of patients in the PROM development, which is a COSMIN recommendation when assessing the relevance of a PROM for the target population (Mokkink et al., 2019). Although the original validation studies did not always describe who had been involved in the

PROMs development in detail, content validity can be assessed at a later date, if required. Therefore, future literature reviews could focus on content validity studies to determine these PROMs' true relevance for a young population experiencing knee disorders.

STRENGTHS AND LIMITATIONS

This scoping meta-review is robust as it was informed by COSMIN guidelines and used COSMIN tools such as PROMs-specific databases/filters and checklists (Mokkink et al., 2018b; Mokkink et al., 2019). Although the database searches were less comprehensive than a systematic review, the pragmatic approach for the scoping nature of this paper's aim (Sarrami-Foroushani et al., 2015), was appropriate and gave credibility to the findings.

The exclusion of knee-related PROMs associated with paediatric and older populations attempted to focus on PROMs relevant to young adults. However, it was challenging to make decisions on the inclusion of individual systematic reviews (and associated PROM development and validation studies) that were specific to the age of the target population. Excluding PROMs intended for specific older or younger age groups achieved broad relevance to young adults with knee disorders.

Only disease- and condition-specific PROMs were identified among the five most evaluated knee-related PROMs. This scoping meta-review missed PROMs intended for use in a young population experiencing knee problems across various medical fields.

5 CONCLUSION

In conclusion, KOOS had the most and strongest evidence to be an appropriate PROM when selecting for individual care and KOS-ADLS for group-level analysis in a young adult population experiencing knee problems. However, neither of these five PROMs demonstrated sufficient high-quality evidence in all six measurement properties, highlighting a potential problem if only the most frequently evaluated PROMs were selected for different purposes. Careful attention should be given to why, for whom and in which setting PROMs as well as the quality of evidence on measurement properties when determining which PROM to select. Using advances in PROMs-specific methods and reporting guides will enhance the quality of evidence and allow readers to appraise relevant evidence and select the most appropriate PROMs for their intended purpose.

REFERENCES

Aaronson, N., Elliott, T., Greenhalgh, G., Halyard, M., Hess, R., Miller, D., Reeve, B., Santana, M., & Snyder, C. (2015). User's guide to implementing patient-reported outcomes assessment in clinical practice. *International Society for Quality of Life Research*. Retrieved from https://www.isoqol.org/wp-content/uploads/2019/09/2015UsersGuide-Version2.pdf

Abed, V., Kapp, S., Nichols, M., Shephard, L., Jacobs, C., Conley, C., & Stone, A.V. (2024). Responsiveness of Patient-Reported Outcome Measures After Large Knee Articular Cartilage Transplantation: A Systematic Review and Meta-analysis. *The American journal of sports medicine*, 52(10):2676-2682. doi: 10.1177/03635465231196156.

Abram, S. G. F., Middleton, R., Beard, D. J., Price, A. J., & Hopewell, S. (2017). Patient-reported outcome measures for patients with meniscal tears: a systematic review of measurement properties and evaluation with the COSMIN checklist. *BMJ open*, *7*(10), e017247. https://doi.org/10.1136/bmjopen-2017-017247

Boers, M., Brooks, P., Strand, C. V., & Tugwell, P. (1998). The OMERACT filter for Outcome Measures in Rheumatology. *The Journal of rheumatology*, *25*(2), 198–199.

Bunt, C. W., Jonas, C. E., & Chang, J. G. (2018). Knee Pain in Adults and Adolescents: The Initial Evaluation. *American family physician*, *98*(9), 576–585.

Çelik, D., Çoban, Ö., & Kılıçoğlu, Ö. (2019). Minimal clinically important difference of commonly used hip-, knee-, foot-, and ankle-specific questionnaires: a systematic review. *Journal of clinical epidemiology*, 113, 44–57. https://doi.org/10.1016/j.jclinepi.2019.04.017

Chamorro-Moriana, G., Perez-Cabezas, V., Espuny-Ruiz, F., Torres-Enamorado, D., & Ridao-Fernández, C. (2022). Assessing knee functionality: Systematic review of validated outcome measures. *Annals of physical and rehabilitation medicine*, *65*(6), 101608. https://doi.org/10.1016/j.rehab.2021.101608

Cohen J. (1998). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. New Jersey: Lawrence Erlbaum.

Cumberledge. (2020). *First Do No Harm – The report of the Independent Medicines and Medical Devices Safety Review.* Retrieved from https://www.immdsreview.org.uk/downloads/IMMDSReview_Web.pdf

Darwich, A., Schüttler, V., Obertacke, U., & Jawhar, A. (2020). Outcome Measures to Evaluate Upper and Lower Extremity: Which Scores are Valid? *Zeitschrift fur Orthopadie und Unfallchirurgie*, 158(1):90-103.

Devji, T., Carrasco-Labra, A., Qasim, A., Phillips, M., Johnston, B.C., Devasenapathy, N., Zeraatkar, D., Bhatt, M., Jin, X., Brignardello-Petersen, R., Urquhart, O., Foroutan, F., Schandelmaier, S., Pardo-Hernandez, H., Vernooij, R.W., Huang, H., Rizwan, Y., Siemieniuk, R., Lytvyn, L., Patrick, D.L., Ebrahim, S., Furukawa, T., Nesrallah, G., Schünemann, H.J., Bhandari, M., Thabane, L., & Guyatt, G.H. (2020). Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *The BMJ*, 4;369:m1714, doi: 10.1136/bmj.m1714.

Dietvorst, M., Reijman, M., van Groningen, B., van der Steen, M. C., & Janssen, R. P. A. (2019). PROMs in paediatric knee ligament injury: use the Pedi-IKDC and avoid using adult PROMs. *Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA*, *27*(6), 1965–1973. https://doi.org/10.1007/s00167-017-4687-3

Elsman, E. B. M., Mokkink, L. B., Terwee, C. B., Beaton, D., Gagnier, J. J., Tricco, A. C., Baba, A., Butcher, N. J., Smith, M., Hofstetter, C., Aiyegbusi, O. L., Berardi, A., Farmer, J., Haywood, K. L., Krause, K. R., Markham, S., Mayo-Wilson, E., Mehdipour, A., Ricketts, J., Szatmari, P., Touma, Z., Moher, D., & Offringa, M. (2024). Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *Health and quality of life outcomes*, *22*(1), 48. https://doi.org/10.1186/s12955-024-02256-9

Esculier, J. F., Roy, J. S., & Bouyer, L. J. (2013). Psychometric evidence of self-reported questionnaires for patellofemoral pain syndrome: a systematic review. *Disability and rehabilitation*, *35*(26), 2181–2190. https://doi.org/10.3109/09638288.2013.774061

Gagnier, J. J., Shen, Y., & Huang, H. (2018). Psychometric Properties of Patient-Reported Outcome Measures for Use in Patients with Anterior Cruciate Ligament Injuries: A Systematic Review. *JBJS reviews*, *6*(4), e5. https://doi.org/10.2106/JBJS.RVW.17.00114

Green, A., Liles, C., Rushton, A., & Kyte, D. G. (2014). Measurement properties of patient-reported outcome measures (PROMS) in Patellofemoral Pain Syndrome: a systematic review. *Manual therapy*, *19*(6), 517–526. https://doi.org/10.1016/j.math.2014.05.013

Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., & Welch, V.A. (Eds). (2023). *Cochrane Handbook for Systematic Reviews of Interventions version 6.4*. Cochrane. Retrieved from www.training.cochrane.org/handbook

Hoglund, L. T., Scalzitti, D. A., Bolgla, L. A., Jayaseelan, D. J., & Wainwright, S. F. (2023a). Patient-Reported Outcome Measures for Adults and Adolescents with Patellofemoral Pain: A Systematic Review of Content Validity and Feasibility Using the COSMIN Methodology. *The Journal of orthopaedic and sports physical therapy*, *53*(1), 23–39. https://doi.org/10.2519/jospt.2022.11317

Hoglund L.T., Scalzitti, D.A., Jayaseelan, D.J., Bolgla, L.A., & Wainwright, S.F. (2023b). Patient-Reported Outcome Measures for Adults and Adolescents With Patellofemoral Pain: A Systematic Review of Construct Validity, Reliability, Responsiveness, and Interpretability Using the COSMIN Methodology. *The Journal of orthopaedic and sports physical therapy*, 53(8):460–479. doi: 10.2519/jospt.2023.11730.

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. *Manual therapy*, *17*(2), 100–118. https://doi.org/10.1016/j.math.2011.07.002

Irrgang, J. J., Anderson, A. F., Boland, A. L., Harner, C. D., Kurosaka, M., Neyret, P., Richmond, J. C., & Shelborne, K. D. (2001). Development and validation of the international knee documentation committee subjective knee form. *The American journal of sports medicine*, *29*(5), 600–613. https://doi.org/10.1177/03635465010290051301

Iversen, M. D., Lee, B., Connell, P., Andersen, J., Anderson, A. F., & Kocher, M. S. (2010). Validity and comprehensibility of the international knee documentation committee subjective knee evaluation form in children. *Scandinavian journal of medicine & science in sports*, *20*(1), e87-e95.

Jordan, K. P., Kadam, U. T., Hayward, R., Porcheret, M., Young, C., & Croft, *P.* (2010) Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. *BMC musculoskeletal disorders* 11(144). doi.org/10.1186/1471-2474-11-144

Kujala, U. M., Jaakkola, L. H., Koskinen, S. K., Taimela, S., Hurme, M., & Nelimarkka, O. (1993). Scoring of patellofemoral disorders. *Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association*, *9*(2), 159–163. https://doi.org/10.1016/s0749-8063(05)80366-4

Law, M., & MacDermid, J. (2008). *Evidence-based rehabilitation: a guide to practice.* (2nd ed). Thorofare (NJ): Slack Inc.

Macri, E.M., Young, J.J., Ingelsrud, L.H., Khan, K.M., Terluin, B., Juhl, C.B., Whittaker, J.L., Culvenor, A.G., Crossley, K.M. and Roos, E.M. (2022). Meaningful thresholds for patient-reported outcomes following interventions for anterior cruciate ligament tear or traumatic meniscus injury: a systematic review for the OPTIKNEE consensus. *British Journal of Sports Medicine,* 56(24), 1432-1444.

Magnuson, J. A., Platt, B. N., Zacharias, A. J., Bowers, L. C., Jacobs, C. A., Liu, J. N., & Stone, A. V. (2022). Patient-reported outcome scores following patellar instability surgery-high prevalence does not equal high responsiveness: a systematic review. *Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA*, *30*(4), 1453–1460. https://doi.org/10.1007/s00167-021-06625-5

Marinho, A. P. R., Nunes, G. S., Menezes, E., Benetti, M., & de Noronha, M. (2020). Questionnaires for knee instability assessment in people with anterior cruciate ligament injury: a systematic review of original questionnaires and their translated versions. *Disability and rehabilitation*, *42*(2), 173–182. https://doi.org/10.1080/09638288.2018.1496153

Meadows, K.A. (2011). Patient-reported outcome measures: an overview. *British Journal of Community Nursing*. 16(3):146-51. doi: 10.12968/bjcn.2011.16.3.146. PMID: 21378658.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, *19*(4), 539–549. https://doi.org/10.1007/s11136-010-9606-8

Mokkink, L.B., Prinsen, C.A.C., Patrick, D.L., Alonso, J., Bouter, L.M., de Vet, H.C.W. & Terwee, C.B. (2018a). COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User Manual. Version 1.0. Amsterdam, The Netherlands: COSMIN (Consensus based Standards for the selection of health Measurement INstruments). Retrieved from: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf

Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018b). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation*, *27*(5), 1171–1179.

Mokkink, L.B., Prinsin, C.A.C., Patrick, D.L., Alonso, J., Bouter, L.M., de Vet, H.C.W., & Terwee, C.B. (2019). *COSMIN Study Design checklist for Patient-reported outcome measurement instruments.* Retrieved from https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf#

National Ligament Registry. (n.d.). *National Ligament Registry.* Retrieved from https://www.uknlr.co.uk/

Nelson, E. C., Eftimovska, E., Lind, C., Hager, A., Wasson, J. H., & Lindblad, S. (2015). Patient reported outcome measures in practice. *BMJ (Clinical research ed.)*, *350*, g7818. https://doi.org/10.1136/bmj.g7818

NHS Digital. (n.d.). *Patient Reported Outcome Measures (PROMs).* Retrieved from https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-reported-outcome-measures-proms

NICE. (2022). *Osteoarthritis in over 16s: diagnosis and management NICE guideline*. Retrieved from www.nice.org.uk/guidance/ng226

OCEBM Levels of Evidence Working Group. "The Oxford 2011 Levels of Evidence". Oxford Centre for Evidence-Based Medicine. http://www.cebm.net/index.aspx?o=5653 Available from: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.)*, *372*, n71. https://doi.org/10.1136/bmj.n71

Roos, E.M., Roos, H.P., Lohmander, L.S., Ekdahl, C., & Beynnon, B. D. (1998). Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. *The Journal of orthopaedic and sports physical therapy*, *28*(2), 88–96. https://doi.org/10.2519/jospt.1998.28.2.88

Sarrami-Foroushani, P., Travaglia, J., Debono, D., Clay-Williams, R., & Braithwaite, J. (2015). Scoping meta-review: introducing a new methodology. *Clinical and translational science*, *8*(1), 77–81. https://doi.org/10.1111/cts.12188

Schellingerhout, J.M., Verhagen, A.P., Heymans, M.W., Koes, B.W., de Vet, H.C. & Terwee, C.B. (2012). Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation,* 21(4), 659–670. https://doi.org/10.1007/s11136-011-9965-9

Shephard, L., Abed, V., Nichols, M., Kennedy, A., Khalily, C., Conley, C., Jacobs, C., & Stone, A.V. (2023). International Knee Documentation Committee (IKDC) Is the Most Responsive Patient Reported Outcome Measure After Meniscal Surgery. Arthroscopy, Sports Medicine, and Rehabilitation, 5(3):e859-e865. doi: 10.1016/j.asmr.2023.02.005.

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

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 27(5), 1159–1170. https://doi.org/10.1007/s11136-018-1829-0 van der Wees, P.J., Verkerk, E.W., Verbiest, M.E.A., Zuidgeest, M., Bakker, C., Braspenning, J., de Boer, D., Terwee, C.B., Vajda, I., Beurskens, A., & van Dulmen, S.A. (2019). Development of a framework with tools to support the selection and implementation of patient-reported outcome measures. *Journal of Patient-Reported Outcomes*. 3(1):75. doi: 10.1186/s41687-019-0171-9.

van Tulder, M., Furlan, A., Bombardier, C. & Bouter, L. (2003). Updated Method Guidelines for Systematic Reviews in the Cochrane Collaboration Back Review Group. *Spine*, 28(12), 1290-1299.

Wang, D., Jones, M. H., Khair, M. M., & Miniaci, A. (2010). Patient-reported outcome measures for the knee. *The journal of knee surgery*, *23*(3), 137–151. https://doi.org/10.1055/s-0030-1268691

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M., Sterne, J. A., Bossuyt, P. M., & QUADAS-2 Group (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine*, *155*(8), 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Flow Diagram for selection of systematic reviews evaluating measurement properties relevant to knee-related patient-reported outcome measures (PROMs) in a young adult population.

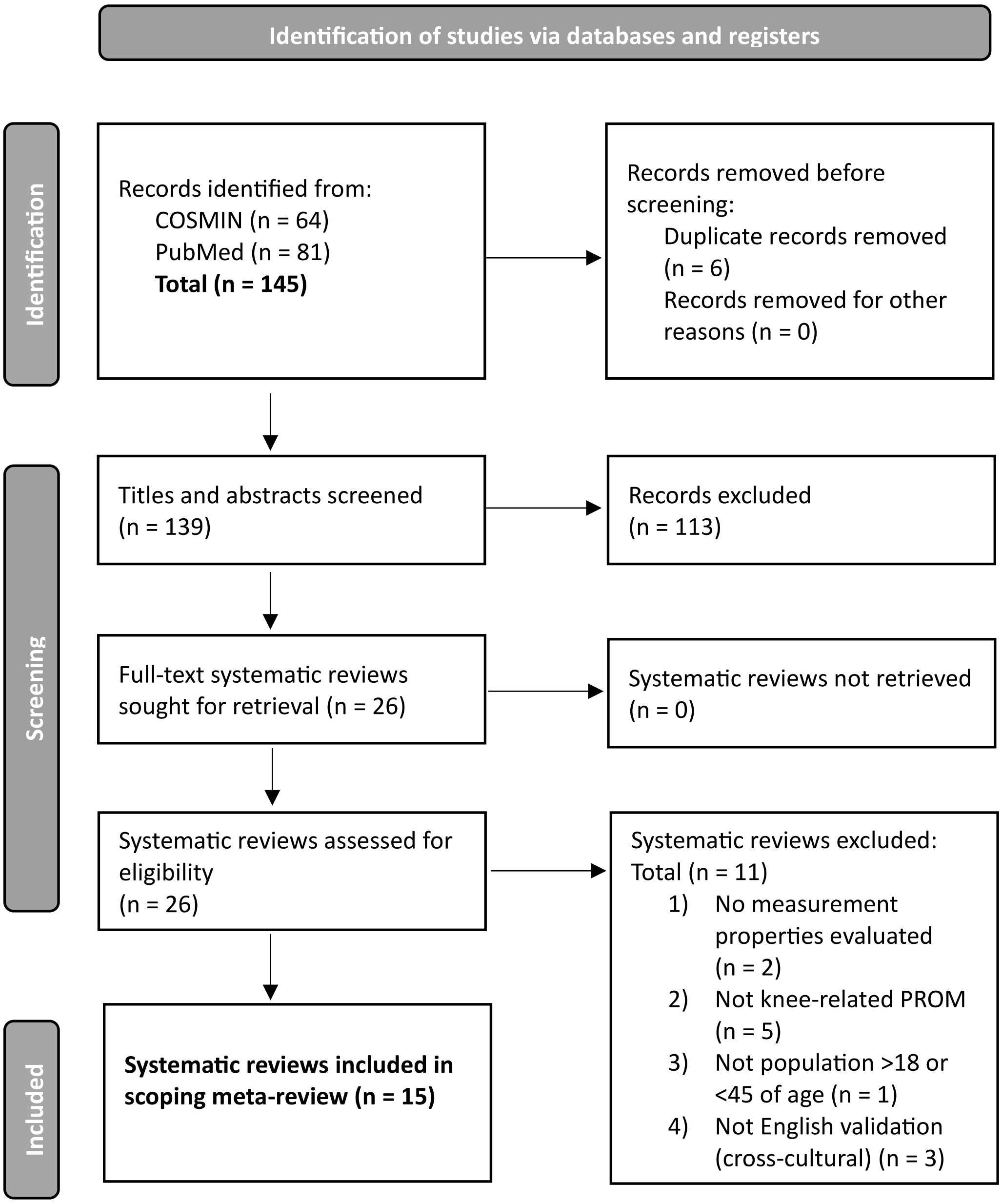

Table 1: Overview of the 15 included systematic reviews on measurement properties of knee-related PROMs in a young adult population

AUTHORS, YEAR OF PUBLICATION/ COUNTRY	ANALYSIS METHOD	PROMs EVALUATION TOOLS	MEASUREMENT PROPERTIES
Abed, Kapp, Nichols, Shephard, Jacobs, Conley and Stone, 2024 United States	Statistical analysis	None	Responsiveness (effect size, relative efficiency)
Abram, Middleton, Beard, Price and Hopewell, 2017 United Kingdom	Analytical synthesis	COSMIN checklists (Terwee et al, 2012[1]), Level of Evidence based on the Cochrane Collaboration Back Review Group by van Tulder, Furlan, Bombardier & Bouter, 2003 (Schellingerhout et al., 2012[6])	Internal consistency, reliability, measurement error, content validity, construct validity, structural validity, hypothesis testing, cross-cultural validity, responsiveness, interpretability, generalisability
Çelik, Çoban and Kılıçoğlu, 2019 Turkey	Descriptive and statistical analysis	None	Minimal clinically important difference using an anchor-based approach
Chamorro-Moriana, Perez-Cabezas, Espuny-Ruiz, Torres-Enamorado and Ridao-Fernández, 2022 Spain	Descriptive and analysis	COSMIN Risk of Bias checklist (Mokkink et al., 2018b[3]), QUADAS-2 (Whiting et al., 2011[7])	Applicability, tool development
Esculier, Roy and Bouyer, 2013 Canada	Descriptive and statistical analysis	None	Construct validity, content validity, factorial validity, floor/ceiling effect, test-retest ICC, standard error of measurement, minimal detectable change, responsiveness (effect size), minimal detectable change, internal consistency, standardised response mean, clinical importance difference, longitudinal validity
Gagnier, Shen and Huang, 2018 United States	Descriptive analysis and synthesis	COSMIN checklist (Terwee et al., 2012[1]), Level of Evidence based on the Cochrane Collaboration Back Review Group by van Tulder, Furlan, Bombardier & Bouter, 2003 (Schellingerhout et al., 2012[6])	Internal consistency, reliability, measurement error, content validity, criterion validity, construct validity, and responsiveness

Green, Liles, Rushton and Kyte, 2014 United Kingdom	Descriptive analysis and synthesis	COSMIN checklist (Terwee et al., 2012[1]), Level of Evidence based on the Cochrane Collaboration Back Review Group by van Tulder, Furlan, Bombardier & Bouter, 2003 (Schellingerhout et al., 2012[6])	Reliability, internal consistency, measurement error, construct validity, criterion validity, cross-cultural validity, responsiveness, interpretability
Hoglund, Scalzitti, Bolgla, Jayaseelan and Wainwright, 2023a United States	Descriptive analysis and synthesis	COSMIN Methodology for assessing the content validity of PROMs (Terwee et al., 2018[4]), COSMIN Grades of Recommendation, Assessment, Development and Evaluation (Mokkink et al., 2018a[2])	Content validity, feasibility
Hoglund, Scalzitti, Bolgla, Jayaseelan and Wainwright, 2023b United States	Descriptive analysis and synthesis	COSMIN Risk of bias checklist (Mokkink et al., 2018b[3]) COSMIN Grades of Recommendation, Assessment, Development and Evaluation (Mokkink et al., 2018a[2])	Construct validity, reliability, responsiveness, interpretability
Howe, Dawson, Syme, Duncan and Reid, 2012 United Kingdom	Descriptive analysis	None	Robustness, reliability, sensitivity
Macri, Young, Ingelsrud, Khan, Terluin, Juhl, Whittaker, Culvenor, Crossley and Roos, 2022 Denmark	Descriptive and statistical analysis	Minimal Important Difference Credibility Assessment Tool (Devji et al., 2020[5]).	Responsiveness (meaningful thresholds)
Magnuson, Platt, Zacharias, Bowers, Jacobs, Liu and Stone, 2022 United States	Statistical analysis	None	Responsiveness (effect size, relative efficiency)
Marinho, Nunes, Menezes, Benetti and de Noronha, 2020 Brazil	Descriptive analysis	COSMIN checklists (Terwee et al, 2012[1])	Internal consistency, construct validity, reproducibility, floor/ceiling effect, responsiveness
Shephard, Abed, Nichols, Kennedy, Khalily, Conley, Jacobs and Stone, 2023 United States	Statistical analysis	None	Responsiveness (effect size, relative efficiency)

Wang, Jones, Khair and Miniaci, 2010 **United States**	Descriptive analysis	None		Reliability, validity (content validity, construct validity, criterion validity), responsiveness

COSMIN: COnsensus-based Standards for the selection of health status Measurement INstruments
OMERACT: Outcome Measures in Rheumatology
MCID: Minimal clinically important difference
QUADAS: Quality Assessment of Diagnostic Accuracy Studies

PROMs evidence grading:
[1]COSMIN checklist (Terwee et al., 2012) 'excellent', 'good', 'fair', 'poor', 'indeterminate'
[2]COSMIN Grades of Recommendation, Assessment, Development and Evaluation (Mokkink et al., 2018a): 'high', 'moderate', low', 'very low'
[3]COSMIN Risk of bias checklist (Mokkink et al., 2018b): 'Very good', 'adequate', 'doubtful', 'inadequate', 'not applicable'
[4]COSMIN Methodology for assessing the content validity of PROMs (Terwee et al., 2018): 'Very good', 'adequate', 'doubtful', 'inadequate', 'not applicable'
[5]Minimal Important Difference Credibility Assessment Tool (Devji et al., 2020): 'yes', 'definitely yes', 'to a great extent', 'no', 'definitely no', 'not so much', 'impossible to tell'
[6]Level of evidence (Schellingerhout et al., 2012): 'strong evidence of positive/negative result', 'moderate evidence of positive/negative result', 'limited evidence of a positive/negative result', 'conflicting evidence', 'unknown due to poor methodological quality'
[7]QUADAS-2 (Whiting et al., 2011): 'low', 'high', 'unclear'

Table 2. Key Characteristics of five most cited PROMs and evaluation of measurement properties relevant to individual care.

PROM / ORIGINAL DEVELOPMENT / INTENDED CONSTRUCT / CATEGORY	ABBREVIATION	PROMs DEVELOPMENT SOURCE	POPULATION ANALYSED (n, mean age, range)	RESPONSIVENESS EVALUATION
International Knee Documentation Committee Subjective Knee Form **Irrgang, Anderson, Boland, Harner, Kurosaka, Neyret, Richmond and Shelborne, 2001, United States** **To assess a variety of knee problems** **Condition-specific**	IKDC-SKF	Previous version of Musculoskeletal Outcomes Data Evaluation and Management Systems questionnaires, Knee Outcome Survey (Irrgang et al., 2001) IKDC members (Wang et al., 2010) Validation sample: n=590 (Irrgang et al., 2001)	N=470, 48.4 years (18 – 81), meniscal surgery (Abram et al., 2017) N=1057, Mean age NR (Range NR), various knee injuries and pathologies (Çelik et al., 2019) N=118 – 2895 (depending on evaluated measurement property), Mean age NR (Range NR), patellofemoral pain syndrome (Esculier et al., 2013) N=2702, 39 years (6 – 100), various knee disorders (Howe et al., 2012) N=1089, 22.7 years (Range NR) patellar instability surgery (Magnuson et al., 2022)	Excellent (Marinho et al., 2020[1]) Indeterminate with poor quality of evidence (Gagnier et al., 2018[2,4]; Green et al., 2014[1,4]) No evidence (Hoglund et al., 2023b[2]) Two out of five items rated low credibility of minimal important change (Macri et al., 2022[3]) Time to follow-up; 3 -24 Months (Abram et al., 2017; Çelik et al., 2019)
Lysholm Knee Score **Lysholm and Gillquist, 1982, Sweden** **To demonstrate the need to evaluate knee ligament surgery** **Disease-specific**	LKS	Modification of the Larson scale, literature review (Lysholm & Gillquist, 1982) Orthopaedic surgeons with sports medicine experience (Wang et al., 2010)	N=790, 41.2 years (Range NR), meniscal surgery (Abram et al., 2017) N=42 – 315 (depending on evaluated measurement property), Mean age NR (Range NR), patellofemoral pain syndrome (Esculier et al., 2013) N=4056, Mean age NR (14 – 88), meniscal, ligament, osteochondral,	Poor (Marinho et al., 2020[1]) Indeterminate with poor quality of evidence (Gagnier et al., 2018[1,4]; Green et al., 2014[1,4]) No evidence (Hoglund et al., 2023b[2]) Time to follow-up: 6 – 79 Months (Abram et al., 2017)

			patellofemoral disorders (Howe et al., 2012) N= 1609, 22.7 years (Range NR) patellar instability surgery (Magnuson et al., 2022)	
Anterior Knee Pain Scale **Kujala, Jaakkola, Koskinen, Taimela, Hurme and Nelimarkka, 1993, Finland** **To score patellofemoral disorders** **Disease-specific**	AKPS	Modification of the Larson scale (Kujala et al., 1993)	N=67 – 204 (depending on evaluated measurement property), Mean age NR (Range NR), patellofemoral pain syndrome (Esculier et al., 2013) N=121, 25.7 years (14 – 40), patellofemoral pain syndrome (Green et al., 2014) N=101, 32.2 years (Range NR), anterior knee pain, patellofemoral pain (Howe et al., 2012) N=3076, 22.7 years (Range NR) patellar instability surgery (Magnuson et al., 2022)	Poor with poor quality of evidence (Green et al., 2014[1,4]) Insufficient with moderate quality of evidence (Hoglund et al., 2023b[2]) Time to follow-up: 6 Weeks (Abram et al., 2017)
Knee Outcome Survey Activities of Daily Living Scale - 17 items **Irrgang, Snyder-Mackler, Wainner, Fu and Harner, 1998, United States** **To assess functional limitations of various knee disorders** **Condition-specific**	KOS-ADLS	Review of existing scales and IKDC guidelines; expert panel consisting of reviewed twelve physiotherapists specialising in musculoskeletal rehabilitation (Irrgang et al., 1998; Wang et al., 2010) Validation sample: n=52, 31.6 years (14 – 66), people receiving physiotherapy for various knee disorders (ligamentous/meniscus injury, patellofemoral pain, OA) or after knee surgery (ligamentous reconstruction,	N=397, 33.3 years (12 – 76), people receiving physiotherapy for various knee disorders (ligamentous/meniscus injury, patellofemoral pain, OA) or after knee surgery (ligamentous reconstruction, meniscectomy, patellofemoral realignment, arthroscopy, high tibial osteotomy, total knee arthroplasty) (Howe et al., 2012, Hoglund et al., 2023) N= 108, Mean years NR (15 – 61), meniscal, ligament, patella-femoral	Good (Marinho et al., 2020[1]) No evidence (Gagnier et al., 2018[1,4]) Sufficient with very low quality of evidence (Hoglund et al., 2023b[2]) Time to follow-up: 1 Week – 6 Months (Howe et al., 2012; Wang et al., 2010)

		meniscectomy, patellofemoral realignment, arthroscopy, high tibial osteotomy, total knee arthroplasty) (Irrgang, 1998)	problems or osteochondritis (Howe et al., 2012)	
Knee Injury and Osteoarthritis Outcome Score **Roos, E.M, Roos, H.O., Lohmander, Ekdahl and Beynnon, 1998, Sweden** **To assess pain; symptoms, such as swelling and restricted range of motion; activities of daily living; sport and recreation function; and knee-related quality of life in young and middle-aged subjects with ACL injury, meniscus injury, or post-traumatic OA** **Condition-specific**	KOOS	Western Ontario and McMaster Universities Arthritis Index, literature review, expert panel including patients (Roos et al., 1998) Expert panel (Wang et al., 2010)	N=786, Mean age NR (Range NR), anterior cruciate ligament injury/surgery, cartilage lesions, and knee injuries (Çelik et al., 2019) N=21, 36 years (18-46), various knee disorders (Howe et al., 2012) N=245, 22.7 years (Range NR) patellar instability surgery (Magnuson et al., 2022)	Poor (Marinho et al., 2020[1]) Indeterminate with poor quality of evidence (Gagnier et al., 2028[1,4]; Green et al., 2014[1,4]) Sufficient with very low quality of evidence (Hoglund et al., 2023b[2]) Time to follow-up: 1 Week – 24 Months (Abram et al., 2017; Çelik et al., 2019)

Abbreviations: NR = Not Reported

PROMs evidence grading:
[1]COSMIN checklist (Terwee et al., 2012) 'excellent', 'good', 'fair', 'poor', 'indeterminate'
[2]COSMIN Grades of Recommendation, Assessment, Development and Evaluation (Mokkink et al., 2018b): 'high', 'moderate', low', 'very low'
[3]Minimal Important Difference Credibility Assessment Tool (Devji et al., 2020): 'yes', 'definitely yes', 'to a great extent', 'no', 'definitely no', 'not so much', 'impossible to tell'
[4]Level of evidence (Schellingerhout et al., 2012): 'strong evidence of positive/negative result', 'moderate evidence of positive/negative result', 'limited evidence of a positive/negative result', 'conflicting evidence', 'unknown due to poor methodological quality'

Table 3. Measurement properties of five most cited PROMs relevant to group-level analysis.

PROM	INTERNAL CONSISTENCY	MEASUREMENT ERROR	TEST-RETEST RELIABILITY
International Knee Documentation Committee Subjective Knee Form	Very good (Chamorro-Moriana et al., 2022[3]) Good (Marinho et al., 2020[1]) Unknown due to poor methodological quality (Abram et al., 2017[1]; Gagnier et al., 2018[1,4]) Indeterminate with very low quality of evidence (Hoglund et al., 2023b[2])	Excellent (Marinho et al., 2020[1]) Limited evidence of a positive result (Gagnier et al., 2018[1,4]) Unknown due to poor methodological quality (Abram et al., 2017[1]) No evidence (Chamorro-Moriana et al., 2022[3]) Indeterminate with very low quality of evidence (Hoglund et al., 2023b[2])	Excellent (Marinho et al., 2020[1]) Sufficient with very low quality of evidence (Hoglund et al., 2023b[2]) Moderate evidence of positive result (Gagnier, et al., 2018[1,4]) Limited evidence of a positive result (Abram et al., 2017[1]) Inadequate (Chamorro-Moriana et al., 2022[3])
Lysholm Knee Score	Poor (Marinho et al., 2020[1]) Doubtful/inadequate (Chamorro-Moriana et al., 2022[3]) Unknown due to poor methodological quality (Abram et al., 2017[1]; Gagnier et al., 2018[1,4]) No evidence (Hoglund et al, 2023b[2])	Poor (Marinho et al., 2020[1]) Inadequate (Chamorro-Moriana et al., 2022[3]) Unknown due to poor methodological quality (Abram et al., 2017[1]; Gagnier et al., 2018[1,4]) No evidence (Hoglund et al., 2023b[2])	Moderate evidence of positive result (Gagnier et al., 2018[1,4]) Limited evidence of a positive result (Abram et al., 2017[1]) Poor (Marinho et al., 2020[1]) Indeterminate with very low quality of evidence (Hoglund et al., 2023b[2]) Inadequate (Chamorro-Moriana et al., 2022[3])

Anterior Knee Pain Scale	Doubtful/inadequate (Chamorro-Moriana et al., 2022[3]) Indeterminate with moderate quality of evidence (Hoglund et al., 2023b[2]) NR (Green et al., 2014[1,4])	Inadequate (Chamorro-Moriana et al., 2022[3]) Insufficient with low quality of evidence (Hoglund et al., 2023b[2]) Unknown due to poor methodological quality (Green et al., 2014[1,4])	Sufficient with high quality of evidence (Hoglund et al., 2023b[2]) Unknown due to poor methodological quality (Green et al., 2014[1,4]) Inadequate (Chamorro-Moriana et al., 2022[3])
Knee Outcome Survey Activities of Daily Living Scale - 17 items	Excellent (Marinho et al., 2020[1]) No evidence (Gagnier et al., 2018[1,4]) Indeterminate with very low quality of evidence (Hoglund et al., 2023b[2]) Inadequate (Chamorro-Moriana et al., 2022[3])	Excellent (Marinho et al., 2020[1]) Unknown due to poor methodological quality (Gagnier et al., 2018[1,4]) Indeterminate with very low quality of evidence (Hoglund et al., 2023b[2]) No evidence (Chamorro-Moriana et al., 2022[3])	Excellent (Marinho et al., 2020[1]) Sufficient with very low quality of evidence (Hoglund et al., 2023b[2]) Unknown due to poor methodological quality (Gagnier et al., 2018[1,4]) Inadequate (Chamorro-Moriana et al., 2022[3])
Knee Injury and Osteoarthritis Outcome Score	Very good (Chamorro-Moriana et al., 2022[3]) Moderate evidence of negative result, (Gagnier et al., 2018[1,4]) Poor (Marinho et al., 2020[2]) No evidence (Hoglund et al., 2023b[2])	Moderate evidence of negative result (Gagnier et al., 2018[1,4]) Poor (Marinho et al., 2020[2]) No evidence (Chamorro-Moriana et al., 2022[3]) NA (Hoglund et al., 2023b[2])	Strong evidence of positive result (Gagnier et al., 2018[1,4]) Inadequate (Chamorro-Moriana et al., 2022[3]) Poor (Marinho et al., 2020[1]) NA (Hoglund et al., 2023b[2])

Abbreviations: NR = Not Reported

PROMs evidence grading:
[1]COSMIN checklist (Terwee et al., 2012) ‘excellent’, ‘good’, ‘fair’, ‘poor’, ‘indeterminate’
[2]COSMIN Grades of Recommendation, Assessment, Development and Evaluation: ‘high’, ‘moderate’, low’, ‘very low’

[3]COSMIN Risk of bias checklist (Mokkink et al., 2018a): ‘Very good’, ‘adequate’, ‘doubtful’, ‘inadequate’, ‘not applicable'
[4]Level of evidence (Schellingerhout et al., 2012): ‘strong evidence of positive/negative result’, ‘moderate evidence of positive/negative result’, ‘limited evidence of a positive/negative result’, ‘conflicting evidence’, ’unknown due to poor methodological quality’

Supplementary File 1: Assessment for Eligibility

Author	Year	Systematic Review	Measurement Properties	Knee-related	Young Adult Population	English Language	Reason for exclusion
Ateef, M.	2022	Yes	Yes	Yes	Yes	No	Cross-cultural
Baborosso, S.	2023	Yes	No	Yes	Yes	No	No measurement property evaluation
Besmens, I.S.	2023	Yes	Yes	No	Yes	Yes	Not knee joint (lower extremity reconstruction)
Copay, A.G.	2018	Yes	Yes	Yes	No	Yes	Not population >18 or <45 of age
Daskalakis, I.	2021	Yes	Yes	Yes	Yes	No	Cross-cultural
Fennelly, O.	2018	Yes	Yes	No	Yes	Yes	Not knee joint (outcomes to evaluate advanced physiotherapy practice)
Filbay, S.R.	2014	Yes	No	Yes	Yes	Yes	No measurement property evaluation
Gallagher, J.	2017	Yes	Yes	No	Yes	Yes	Not knee joint (outcome to evaluate athletic performance)

Le, C.Y.	2021	Yes	Yes	No	Yes	Yes	Not knee joint (outcomes to evaluate quality of life)
Morris, R.	2018	Yes	Yes	No	Yes	Yes	Not knee joint (lower extremity fracture)
Ngwayi, J.R.M.	2022	Yes	Yes	Yes	Yes	No	Cross-cultural

Supplementary File 2: Frequency of knee-related PROMs in published systematic reviews

	Abed et al. 2024	**Abram et al., 2017**	**Çelik et al., 2019**	**Chamorro-Moriana et al.,**	**Esculier et al., 2013**	**Gagnier et al., 2018**	**Green et al., 2014**	**Hoglund et al., 2023a**	**Hoglund et al., 2023b**	**Howe et al., 2012**	**Macri et al., 2022**	**Magnuson et al., 2022**	**Marinho et al., 2020**	**Shephard et al., 2023**	**Wang et al., 2010**	**Frequency**
International Knee Documentation Committee's (Subjective Knee Form)	**x**	**x**	**x**	**x**	**x**	**x**		**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**14**
Lysholm Knee Score	**x**	**x**	**x**	**x**	**x**	**x**		**x**	**x**	**x**		**x**	**x**	**x**	**x**	**13**
Knee Injury and Osteoarthritis Outcome Score	**x**	**x**		**x**		**x**		**x**	**x**	**x**	**x**		**x**	**x**	**x**	**11**
Anterior Knee Pain Scale			**x**	**x**	**x**		**x**	**x**		**x**		**x**			**x**	**8**
Knee Outcome Survey Activities of Daily Living Scale - 17 items				**x**	**x**	**x**		**x**	**x**	**x**			**x**		**x**	**8**
Tegner Activity Level	x		x	x		x				x	x			x	x	8
Cincinnatti Knee Rating Scale	x					x				x	x	x	x		x	7
Functional Index Questionnaire					x		x	x	x	x					x	6
Knee Injury and Osteoarthritis Outcome Score - Activity of Daily Living	x							x	x		x	x		x		6
Knee Injury and Osteoarthritis Outcome Score - Pain Score	x							x	x		x	x		x		6