Proposed Clinical Endpoints Guidance

Knee Osteoarthritis

June 22, 2023

Contents

I. Preamble

Section 1862(l)(1) of the Social Security Act requires that the Secretary of Health and Human Services make available to the public the factors that are considered in making national coverage determinations (NCDs) of whether an item or service is reasonable and necessary. The Centers for Medicare & Medicaid Services' procedures for issuing guidance documents under this authority are set forth in 69 Fed. Reg. 57 325 (September 24, 2004).

NCDs concerning whether a particular item or service is reasonable and necessary under section 1862(a)(1)(A) are based on information including clinical experience, and medical, technical, and scientific evidence.[1] The NCD process also considers public comments. The public is afforded the opportunity to comment on a proposed determination as set forth in section 1862(l). When we make an NCD, we provide a clear statement of the basis for the NCD as well as responses to the comments received from the public.

To encourage innovation and accelerate beneficiary access to new items and services, CMS is proactively publishing this proposed guidance document to provide a framework for more predictable and transparent evidence development.

This proposed guidance represents the Centers for Medicare & Medicaid Services' (CMS's) current thinking on health outcomes for the Treatment of Knee Osteoarthritis. It does not create or confer any rights for or on any person and does not operate to bind CMS or the public. Where warranted by circumstances, CMS may consider an alternative approach if it satisfies the requirements of the applicable statutes and regulations. Individuals interested in discussing an alternative approach are encouraged to contact CMS at CAGInquiries@cms.hhs.gov and reference this guidance.

CMS is seeking public comment on this proposed clinical endpoints guidance for knee osteoarthritis. CMS will review the public comments and respond to the comments in the final document.

Alternatively, written comments may be submitted to the Coverage and Analysis Group, Centers for Medicare and Medicaid Services, mailstop: S3-02-01, 7500 Security Blvd. Baltimore, MD. 21244. Please refer to this guidance document when submitting comments.

To ensure consideration, comments must be received by **August 21, 2023**.

[1] § 1862(a) in the material following (25). ("[I]n making the [national coverage] determination, the Secretary has considered applicable information (including clinical experience and medical, technical, and scientific evidence) with respect to the subject matter of the determination[.]")
https://www.ssa.gov/OP_Home/ssact/title18/1862.htm

II. Introduction

This proposed guidance document identifies health outcomes of interest to CMS when reviewing technologies for the treatment of knee osteoarthritis when considering reasonable and necessary NCDs. Specific technologies were not reviewed, and this guidance is not a National Coverage Analysis or National Coverage Determination.

III. Background

When making national coverage determinations, CMS generally evaluates relevant clinical evidence to determine whether or not the evidence is of sufficient quality to support a finding that an item or service falling within a benefit category is reasonable and necessary for the diagnosis or treatment of an illness or injury or to improve the functioning of a malformed body member under section 1862(a)(1)(A) of the Social Security Act. The overall objective for the critical appraisal of the evidence is to determine to what degree we are confident that the specific assessment questions raised in a National Coverage Analysis can be answered conclusively. In the August 7, 2013, Federal Register (78 FR 48164), we published a notice that describes the processes we use for opening, making or reconsidering national coverage determinations (NCDs).[2]

When conducting National Coverage Analyses for an item or service under the reasonable and necessary statute, CMS generally makes three kinds of assessments: (1) The quality of relevant individual studies; (2) What conclusions can be drawn from the body of the evidence on the direction and magnitude of the intervention's potential harms and benefits; and (3) The generalizability of findings from relevant studies to the Medicare beneficiary population.

Because we are interested in accelerating beneficiary access to new items and services, evidence-based decision making is an important component of establishing national coverage policies. Clinical evidence that supports a determination that an item or service is reasonable and necessary for the Medicare beneficiary population is important because these individuals are often older, with multiple comorbidities, and are often underrepresented or not represented in many clinical studies. In general, CMS looks to the evidence supporting FDA market authorization and the device indications for use for evidence generalizable to the Medicare beneficiary population, data on improvement in health outcomes, and durability of those outcomes. If there are no data on those elements, it is difficult for CMS to make a favorable evidence-based decision on whether the device is reasonable and necessary for the Medicare population.[3]

---

[2]78 FR 48164 (August 7, 2013) *available at* https://www.govinfo.gov/content/pkg/FR-2013-08-07/pdf/2013-19060.pdf

[3]86 FR 62944 (November 15, 2021) *available at* https://www.govinfo.gov/content/pkg/FR-2021-11-15/pdf/2021-24916.pdf

IV. Purpose of Therapeutic Outcomes Reviews

As part of CMS' commitment to improve the transparency of the NCD process, we intend to publish a series of guidance documents that review important clinical outcomes for treatments addressing specific therapeutic areas. These reviews will assist interested parties seeking Medicare coverage for an item or service, such as a drug or device, in understanding the types of evidence CMS expects to review when making national coverage determinations (NCDs) as outlined in 1862(l) of the Social Security Act.

CMS undertakes a number of activities designed to improve the health care provided to Medicare beneficiaries. These activities include coverage policy decisions that determine which services can be covered as "reasonable and necessary" under title XVIII of the Social Security Act. These decisions consider the best available scientific and clinical evidence concerning the benefits and harms of various clinical items and services and apply the highest attainable level of expertise to evaluate such evidence.

CMS may use technology assessments (TAs) to assist in reviewing evidence in the NCD process when determining whether a particular technology is reasonable and necessary or, in some cases, to identify those areas that need further evidence development. A TA can involve the evaluation of a technology's performance characteristics, safety, efficacy, effectiveness, outcomes, and appropriateness. TAs that systematically evaluate available evidence are the most highly regarded assessments. These types of assessments include a range of related activities, such as identifying and prioritizing technologies for assessment, collecting and analyzing data, synthesizing and grading evidence, and disseminating findings and recommendations.[4]

While a TA can evaluate many aspects of a technology, interpretation and critical appraisal of the evidence on patients' health outcomes constitutes its vital component. Accordingly, the performance of a systematic review of the evidence from the medical literature is at the core of every TA, whether undertaken internally or commissioned externally by CMS. Systematic reviews are scientific investigations that synthesize the results of multiple primary investigations on one or more relevant clinical questions. The evidence is then appraised to assess its validity (how credible it is), usefulness (its clinical applicability), and importance (magnitude of effect). To minimize bias, systematic reviews emphasize a comprehensive search of all potentially relevant articles and the use of explicit, reproducible criteria in selecting articles for review. Primary research designs and study characteristics are appraised, data are synthesized, and results are interpreted.[5]

[4] https://www.cms.gov/medicare-coverage-database/view/medicare-coverage-document.aspx?MCDId=7
[5] https://www.cms.gov/MEDICARE-COVERAGE-DATABASE/view/medicare-coverage-document.aspx?MCDId=7#P67_3733

By incorporating methods to assemble, sort through and integrate clinical evidence, the systematic review embedded in a TA represents a rigorous compilation of scientific evidence to answer clinical questions. It enables other parties to understand, replicate, and judge the collection, selection, and analysis of evidence and is particularly useful when the medical research literature is complex or extensive.

Prior TAs have focused on clinical factors pertinent to Medicare patients' health outcomes. For the current review, CMS engaged with a contractor to complete the technology assessment of clinical endpoints in knee osteoarthritis (**Section V**, **Technology Assessment**). This topic was selected because of the high burden of osteoarthritis in the Medicare beneficiary population, because knee replacement surgeries are among the most common surgical interventions provided to Medicare beneficiaries, and because new technologies are being developed that are intended to treat this condition. In **Section VI**, CMS draws conclusions from the technology assessment and makes recommendations. CMS is seeking public comment on this guidance document. In particular, we are interested in public comments on the approach, the methodology, and on the utility of this guidance.

V. Technology Assessment

Overview

Osteoarthritis (OA) of the knee is a common form of arthritis with a broad spectrum of presentation, ranging from an asymptomatic radiological finding to a progressive disease resulting in joint failure. It typically occurs bilaterally, though severity can differ unilaterally. Primary symptoms include pain, limitation of motion, stiffness, tenderness, swelling, joint deformity, and instability. Knee OA is the leading cause of lower-limb disability in adults 50 years and older worldwide (Doherty M & A Abhishek, 2021) and assumes 80% of all OA burden, with women experiencing higher rates of radiographically confirmed symptomatic knee OA than men. Approximately 14 million people in the United States (US) experience symptomatic knee OA, and prevalence is estimated at 7%. The lifetime risk of developing symptomatic knee OA is about 40% for men and 47% for women. (March L & M Cross, 2020)

Management is contingent on levels of pain, functional and participatory impairments, and impacts on quality of life. Nonpharmacologic therapies, topical therapies, or analgesics as needed are the first avenues of care usually administered for mild cases. (Deveza L & K Bennell, 2022) For moderate to severe cases, oral nonsteroidal anti-inflammatory drugs (NSAIDs), duloxetine for patients contraindicated or refractory to NSAIDs, intraarticular glucocorticoid injections, and adjunctive braces or walking devices, as appropriate, are given. If significant symptoms persist after exhausting these options, then surgery is often considered. (Deveza L & K Bennell, 2022) Surgical procedures include total knee replacement (TKR), unicompartmental knee replacement, and knee osteotomy. (Mandl L & G Martin, 2020)

Recent advancements to enhance surgical accuracy have led to an array of new technologies, such as patient-specific instrumentation, sensors, accelerometers, and robotic-assisted surgery. (Batailler C et

al., 2020) As the field evolves to increase the types of innovations available, it becomes imperative to consistently measure outcomes that demonstrate improvements relevant to patients, providers, payers, and other key healthcare stakeholders for evidence-based decision-making. Development and uptake of a core outcome set (COS), a minimum set of outcomes that should be measured and reported in all clinical trials for a given condition (Williamson P et al., 2012) for knee OA is a means to that end. Furthermore, having a COS will promote transparency for innovators seeking to anticipate the evidence needs of healthcare decision-makers.

Objective: This report aims to compile a succinct list of prioritized outcomes and instruments that represent the most relevant, meaningful outcomes that should be used to evaluate medical technologies that treat knee OA.

Methods

Identifying the Literature

Searches were conducted in multiple databases and evidence-based sources to comprehensively capture prioritized outcome measures and instruments related to surgical and other device interventions for knee OA. A systematic search using the terms "knee osteoarthritis" or "osteoarthritis of the knee" and "core measures" or "core outcomes" or "outcome measures" was done in Embase/Science Direct on February 14, 2022, and a corresponding search was completed n PubMed/MEDLINE on February 21, 2022, to retrieve systematic reviews or consensus statements around knee OA outcomes. Search strings are detailed in **Appendix A**, and eligibility criteria are listed in **Table 1**. The project team reviewed all articles at the title and abstract levels. We obtained for full review any articles possibly meeting the inclusion criteria. Individual team members reviewed all retrieved articles for inclusion; discussion and consensus between two team members resolved uncertainty about full-length article inclusion.

Table 1. Eligibility Criteria for PubMed/MEDLINE Screening

Inclusion Criteria	
#1	Paper must report on determining the appropriate outcome measures for management of knee osteoarthritis; inclusion of hip acceptable if knee is also a topic of interest
#2	Paper must have either: a) Implemented a systematic search of the literature to evaluate outcomes b) Used an established process (e.g., Delphi) to arrive at consensus on outcomes
#3	English language publication
#4	Published 14-Feb-2012 and onward
Exclusion Criteria	
#1	Primary randomized controlled trials evaluating treatment
#2	Commentary, opinion, conference abstracts, narrative reviews, protocols for reviews or studies
#3	Non-English language publication
#4	Published prior to 14-Feb-2012
#5	Does not report on knee osteoarthritis

Additional searches were conducted to supplement the findings from PubMed and Embase. A search was done in the Cochrane Library on February 22, 2022, using the term "osteoarthritis of the knee" to

retrieve relevant protocols and systematic reviews, as well a search of all systematic reviews published from 2012 onwards filed under the full list of reviews on the Cochrane Musculoskeletal group's (CMSG) evidence webpage. To ascertain whether added research on knee OA core outcomes was available, two searches were executed in the Core Outcome Measures in Effectiveness Trials (COMET) Initiative's database on February 14, 2022, one for "knee osteoarthritis" and the other for "total joint replacement." Supplementary scans were also completed within the following sources for peer-reviewed or gray literature relevant to the project scope: Agency for Healthcare Research and Quality (AHRQ) comparative effectiveness reviews, Food and Drug Administration (FDA) Patient-Focused Drug Development meeting reports and guidance documents, National Institute on Aging's Osteoarthritis Initiative (OAI), National Institute of Arthritis and Musculoskeletal and Skin Disease's Orthopedic Research Program, Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT), Osteoarthritis Research Society International (OARSI), European Alliance for Health Outcomes Measurement (EULAR), and the International Consortium for Health Outcomes Measurement (ICHOM). Details of these scans are documented in **Appendix A**.

Data Abstraction and Data Management

Included publications were categorized as follows: (1) systematic and comparative effectiveness reviews (SLR/CER), (2) consensus development, and (3) other (i.e., protocols, guidelines, workshop reports). For publications within categories one and three, we abstracted the following study level details: objective, funding source, databases, included study types, and eligibility criteria. If an article was categorized as consensus development, then consensus methodology was also captured. We abstracted data about the characteristics of the included studies, as well as characteristics of the stakeholder groups involved in consensus development if applicable. Finally, for all publications, we abstracted data on the outcomes and associated instruments when available, along with whether they were deemed a prioritized or non-prioritized outcome or instrument.

Data Synthesis

Following abstraction, the data was synthesized to create a prioritized list of outcome domains and instruments that included any abstracted subdomains or items. Outcomes were organized by the four core areas of Life Impact, Pathophysiological Manifestations, Resource Use/Economical Impact, and Death, borrowed from the OMERACT Filter 2.0 guidelines on core outcome measurement set development. (Boers M et al., 2014) An additional "Other" area was used to house outcomes that did not fit under any of the specified core areas. Within each area, outcomes were further organized by whether they were "prioritized" or "non-prioritized" based on the literature. Prioritized outcomes were defined as those designated as either "mandatory", "critical," "core", or "major" by the majority of included publications. In contrast, non-prioritized outcomes were deemed as "important but optional", "non-core," "minor" or received no designation. Outcomes that were classified as "research agenda" warranting further exploration were excluded from the synthesis. Instruments in the synthesis table were organized hierarchically by the number of citations received amongst the included publications, with the most used instruments at the top.

The synthesis table also captures additional information obtained through targeted literature searches on minimal important differences (MID) and its variations (i.e., minimal clinically important differences (MCID), the minimal clinical difference (MCD), minimal clinically important improvement (MCII), minimal important change (MIC), and minimum detectable change (MDC)), as well as validation data. This information was only retrieved if it pertained to knee or knee/hip OA within the outcome domain of assessment. Every attempt was made to find validation and minimal difference studies conducted in the United States or other English-speaking countries where knee OA patients received surgical or other device therapies. Furthermore, instrument-specific characteristics that could potentially contribute to decision-making during multi-stakeholder discussions were captured, such as feasibility (e.g., number of items, time to complete), recall period, reporter, dimensions assessed, the intent of development, and access.

Quality Assessments

To assess the quality of the consensus methodology employed in the consensus development publications, the COMET COS-STAD recommendations were applied. (Kirkham J et al., 2017) These recommendations involve consideration of the following factors: scope specification, stakeholders involved, and consensus process rigor. The methodological quality of included SLRs was first determined using U.S. Preventive Services Task Force (USPSTF) criteria for assessing the internal validity of systematic reviews, (USPSTF, 2017), including the areas of comprehensiveness of sources considered/search strategy used, standard appraisal of included studies, validity of conclusions, and recency and relevance.

Results

Literature

Using the search terms specified in **Appendix A** within the PubMed/MEDLINE and Embase/Science Direct databases, we retrieved 799 records. Of these, 38 were included after title screening, and nine were retained following abstract screening. After full-text articles for the nine records were obtained and screened, five papers met the eligibility criteria for this review. Two of the full-text articles were excluded because they were neither a systematic review nor a consensus statement, and the remaining two articles were excluded because they did not adequately describe a prioritization of outcomes or instruments necessary for the project scope. Notably, we observed that many of the papers discussed knee and hip OA outcomes concurrently, and thus, articles were included if they addressed either solely knee OA outcomes or both knee and hip OA outcomes collectively.

Searches within the Cochrane Library and CMSG about surgical or other device interventions for knee OA or knee/hip OA, published 2012 onwards, yielded two relevant protocols and four systematic reviews. A search within AHRQ using the same criteria retrieved one relevant comparative effectiveness review. Searches targeted at knee or knee/hip OA core outcome sets or consensus statements within the COMET Initiative, OMERACT, OARSI, EULAR, and ICHOM platforms produced ten relevant articles, five of which were duplicates from the PubMed and Embase searches. A webpage was also included from OARSI, which detailed the most important indices to evaluate the algofunctional status of patients

with musculoskeletal diseases. Queries for relevant publications within the FDA and NIH platforms yielded one relevant document, a draft guidance on structural endpoints for the development of drugs, devices, and biological products for OA. A search for summaries from Voice of the Patient or Patient-Focused Drug Development meetings provided an additional record – a report from a Voice of the Patient meeting hosted by the Arthritis Foundation and attended by the FDA. See **Appendix A** for further details on these supplementary scans.

Twenty records were included in this review (**Appendix B**).

Data Abstraction and Data Management

Seven papers were abstracted under the "Systematic and Comparative Effectiveness Review" category. Three papers focused solely on knee OA. (Palmet J et al., 2019; Dulvenvoorden T et al., 2015; Brouwer R et al., 2014) One included knee rheumatoid arthritis, (Hofstede S et al., 2015), one covered other musculoskeletal knee conditions in addition to OA, (Howe T et al., 2012), and the remaining two papers focused on both knee and hip OA. (Konnyu K et al., 2021; Lange T et al., 2017) Five of the reviews were within the context of interventions that included surgery or other devices (Palmet J et al., 2019; Dulvenvoorden T et al., 2015; Brouwer R et al., 2014; Hofstede S et al., 2015; Lange T et al., 2017) while one focused on (p)rehabilitation for TKA or THA, (Konnyu K et al., 2021) and the last did not specify. (Howe T et al., 2012) The number of studies included in each review ranged from 5 to 100, in which the number of patients ranged from 566 to 14,533. Most reviews did not specify a mean age; however, Duivenvoorden et al. (2015) reported a mean of 62 years (range: 48 to 75), and Brouwer et al. (2014) reported a mean of 60 years (range: 42 to 67). Outcomes from the CMSG reviews were designated as either "major" or "minor," and similarly, outcomes were pre-specified as "important/prioritized" in the AHRQ comparative effectiveness review. Lange et al. (2017) explored uptake of the OMERACT-OARSI knee/hip OA COS, reflecting the relative importance researchers placed on these outcomes. Six of the seven review papers discussed instruments that corresponded with their outcomes of interest.

Among the eight "Consensus Development" papers, two addressed knee OA solely (Christensen R et al., 2015; McAlindon T et al., 2015) while the remaining papers focused on both knee and hip OA. Six papers centered around interventions that included surgery or other devices (McAlindon T et al., 2015; Smith T et al., 2019; Hoang A et al., 2017; Singh J et al., 2017; Singh J et al., 2015) and the remaining two papers did not specify an intervention type. (Christensen R et al., 2015; Rolfson O et al., 2016) Of note, two benchmark papers developed sets of "core" or "mandatory" outcome domains that four other included consensus development papers corroborated. Smith et al. (2019) discussed an update of the OMERACT-OARSI COS developed for knee and hip OA clinical trials that McAlindon et al. (2015) supported. Likewise, Singh et al. (2015) introduced an OMERACT-endorsed core set of outcome domains for knee and hip OA patients undergoing total joint replacement, which Hoang et al. (2017) and Singh et al. (2017) substantiated in various stakeholder groups. There was evidence of diverse stakeholder participation in the consensus processes, with patients involved in five of the panels. Four of the processes utilized a Delphi or modified Delphi approach to reach consensus (McAlindon T et al.,

2015; Smith T et al., 2019; Singh J et al., 2015; Rolfson O et al., 2016), whereas others involved a single round of voting or roundtable discussion.

Under "Other Publication Types", we abstracted data from the following records: one webpage containing instrument data, an FDA draft guidance document, a Voice of the Patient report, and two CMSG systematic review protocols. Both review protocols were specific to arthroplasty for the treatment of knee OA. (Jolles B et al., 2013; Singh A & K Vinay, 2013) The remaining records were geared toward OA in general (FDA, 2018; Arthritis Foundation, 2017) or chronic musculoskeletal diseases, including OA. (OSARI, 2022) The authors indicated prioritized outcomes by use of phrasing such as "significant", "major," and "meaningful". Abstracted data for all publications mentioned above can be found in **Appendix B**.

Data Synthesis

Out of the 26 identified outcome domains in the included literature, 11 were classified as prioritized: joint pain, physical function, health-related quality of life (HRQoL), patient satisfaction, function/functional ability, joint structure, stiffness, treatment failure, mortality, serious adverse events (SAEs), and adverse events (AEs) (**Table 2**). Representation from each of the four core areas is present, and recommended instruments per the literature are listed correspondingly for five of the outcome domains. Instruments were considered prioritized if the number of citations met at least 50% of the highest cited instrument's total citations within an outcome domain – for example, under physical function, the Western Ontario and McMaster University OA Index (WOMAC) global score was the highest cited instrument with six citations; thus, any instrument under this domain that had at least three citations was prioritized. Synthesized data can be found in **Appendix C**.

Table 2. Prioritized Outcome Domains and Instruments

Outcome Domains	Instruments
LIFE IMPACT	
Joint pain	VAS, WOMAC pain subscale, Lequesne global score, WOMAC global score
Physical function	WOMAC global score, KOOS, HSS knee score, Lequesne global score, WOMAC function subscale
Health-related quality of life	EQ-5D-5L, SF-12, SF-36
Patient satisfaction	VAS, Likert scale
Function/functional ability	NR
PATHOPHYSIOLOGICAL MANIFESTATIONS	
Joint structure	Radiographic imaging
Stiffness	NR
RESOURCE USE/ECONOMICAL IMPACT	
Treatment failure	NR
DEATH	
Mortality	NR
OTHER	
Serious adverse events	NR
Adverse events	NR

Abbreviations: EQ-5D-5L: EuroQol in 5 dimensions 5 levels; HSS: Hospital for Special Surgery; KOOS: Knee Injury and Osteoarthritis Outcome Score; NR: not reported; SF-12: Short Form 12; SF-36: Short Form 36; VAS: visual analogue scale; WOMAC: Western Ontario and McMaster University Osteoarthritis Index

Five outcome domains were prioritized under life impact: joint pain, physical function, HRQoL, patient satisfaction, and function/functional ability. There was noticeable heterogeneity in the instruments used to measure joint pain, physical function, and HRQoL. The VAS, WOMAC pain subscale, Lequesne global score, and WOMAC global score were the most commonly used instruments to measure joint pain. Each instrument is patient-reported and has been validated within the knee or knee/hip OA patient population as having acceptable reliability and validity, though an MCID for pain could not be found for the Lequesne index. They have a relatively low respondent burden, ranging from a completion time of less than one minute for the VAS to up to ten minutes for the WOMAC global index. The WOMAC and Lequesne scales were developed specifically within the knee and hip OA population, whereas the VAS was not. While the VAS is publicly available, a fee is required for use of the WOMAC. It is unknown whether the Lequesne index is freely available. Importantly, the WOMAC and Lequesne scales are multidimensional assessments of pain whereas the VAS a unidimensional assessment of pain intensity. It has been suggested that a multidimensional measure may better evaluate pain status and responsiveness to interventions since a patient's experience of pain is not a simple construct. (McAlindon T et al., 2015)

The WOMAC global score, Knee Injury and OA Outcome Score (KOOS), Hospital for Special Surgery (HSS) Knee Score, Lequesne global score, and WOMAC function subscale were commonly used to measure physical function. Similar to joint pain, physical function's multidimensional nature was reflected in the types of instruments cited. All instruments were developed in a knee or knee OA-specific population and have been validated within this population with acceptable reliability and validity, although the validity data presented for the KOOS was derived from its short form, the KOOS-PS. MCIDs were found for all but the Lequesne index. It is worth noting here that calculations of minimal differences are highly contextual, a reality reflected in the wide MCID range for the WOMAC global score, which spans from 4.2 for osteotomy to 10.5 to 36.0 for total knee arthroplasty (TKA). Of these recommended instruments, only the KOOS was confirmed to be publicly available. Each instrument is patient-reported, apart from the HSS Knee Score, and possesses a relatively low respondent burden. The HSS Knee Score is clinician reported and consists of a patient interview and physical exam. Although no performance-based measures reached the threshold set for prioritization, incorporating them may have benefits. A study conducted by Mizner et al. (2011) comparing the responsiveness of patient-reported versus performance-based outcome measures after unilateral TKA for end-stage knee OA found that physical performance decreased acutely after surgery per objective measures, but PROs did not concurrently capture these initial functional declines. Acknowledging the importance of objectivity in measuring physical function, OARSI, a leading international organization for scientists and healthcare professionals dedicated to OA research, developed a minimum core set of performance-based tests to assess physical function in knee/hip OA patients. This core set includes the Timed Up and Go (TUG) test, 30-second chair stand test, and 6-minute walk test (6MWT). (Dobson F et al., 2013) Validation data and instrument properties for these tests can also be found in **Appendix C**.

The most cited instruments that measure HRQoL were the EuroQol in 5 dimensions and 5 levels (EQ-5D-5L), 12-item Short Form (SF-12), and 36-item Short Form (SF-36). Though these tools were developed as

patient-reported generic measures for describing health, they have undergone validation studies within the knee or knee/hip OA population and were found to have acceptable reliability and validity. Minimal difference data was available for all three instruments; however, the values presented for the SF-12 separate the scale into its physical and mental component subscales. Researchers have historically advocated for using these component subscales in the Short Forms, stating that there is little evidence to support the relevance of the aggregate score within the knee and hip OA population. (Rannou F et al, 2007) Disaggregated MCIDs were not found for the SF-36. All three prioritized instruments have relatively low to moderate respondent burden. The SF-12 is a condensed version of the SF-36, containing a third of the items. Yet, it assesses the same eight dimensions and highly correlates with the SF-36 in OA patients undergoing knee replacement surgery (Webster K & J Feller, 2016), which might confer upon it an advantage over its longer counterpart. Additionally, all three instruments exist in the public domain, though a licensed version of the SF-36 distributed by Optum has scoring differences compared to the freely available version from RAND.

Patient satisfaction with treatment results and/or procedures can be measured via a patient-reported VAS or Likert scale. Evidence of validation for assessing satisfaction among knee OA patients was not found in the literature for either tool, and no validated, standard question nor point scale was observed. Five- and ten-point scales appeared to be commonly used, ranging from "very/completely/extremely dissatisfied" to very/completely/extremely satisfied." Though no MCID values could be found, a systematic review assessing measures of patient satisfaction with outcomes post-TKR defined satisfaction as a score of ≥7 on a 10-point scale and ≥ four on a 5-point scale VAS. (Klem N et al., 2020) Both scales can be administered without a fee.

The final outcome domain that includes suggested instruments is joint structure, measured by radiographic imaging. Other suggested measurement methods were magnetic resonance imaging (MRI) and sonography, though these needed to meet the threshold set for prioritization. Evaluation of joint structure can include changes in joint structure or observations of disfigurement. Validation was not applicable. Joint structure might be an outcome domain relevant only in specific types of clinical trials, such as those assessing structure-modifying technologies.

Specific instruments for the remaining outcome domains of function/functional ability, stiffness, treatment failure, mortality, SAEs, and AEs were not identified in the included literature. Several domains, such as mortality, treatment failure (e.g., revision surgery, reoperation), SAEs, and AEs, may not warrant a validated instrument. Moreover, stiffness and function/functional ability (e.g., ability to function in society or at work, employability, productivity) may be captured in algofunctional instruments used in other prioritized outcome domains, including the WOMAC, SF-12, and KOOS.

Quality Assessments

The quality of the consensus methodologies implemented across the eight "consensus development" publications was variable. Scope specification was generally acceptable, with all

studies identifying at least three out of the following four factors: health condition, population, intervention, and setting intended for COS use. Robust stakeholder involvement that included knee OA patients as a comparable percentage of the overall panel was apparent in only two studies. (Smith T et al., 2019; Hoang A et al., 2017) All but one study had more than one type of stakeholder group relevant to the COS scope, though the participation of industry representatives who will implement the COS in clinical trials was limited to two studies. (McAlindon T et al., 2015; Smith T et al., 2019) Regarding the consensus process itself, a clearly defined scoring process was provided in four studies. (McAlindon T et al., 2015; Smith T et al., 2019; Hoang A et al., 2017; Singh J et al., 2017) A clear definition of consensus was also provided in four studies. (Smith T et al., 2019; Hoang A et al., 2017; Singh J et al., 2017; Rolfson O et al., 2016) There was not enough information to determine whether the initial list of candidate outcomes adequately incorporated healthcare provider and patient perspectives and avoided ambiguity of language. The quality of the systematic reviews was generally fair to good using the US Preventive Services Task Force (USPSTF) rating system. The main limitations of the reviews rated as fair were due to older search dates. One systematic review was rated as poor, as there were limited search databases, the searches were older, and there was a lack of risk of bias assessment of the included studies. (Lange T et al., 2017) However, for this assessment, the overall quality of the systematic reviews and included literature should not carry the same weight as the quality of the consensus papers. The evidence from the systematic reviews within the context of outcomes prioritization was used merely for retrieval and delineation of major versus minor outcomes and instruments. Hence, the quality of methods and confidence in conclusions drawn by the authors did not affect our findings.

Consensus Assessments

In the current review, professional consensus statements offered relatively homogenous recommendations regarding outcome measures for evaluating efficacy of treatments for knee osteoarthritis. In addition, four Cochrane reviews, one AHRQ comparative effectiveness review, and one FDA Voice of the Patient report were considered. We identified 11 outcome domains, and this report summarizes the instruments used within each domain, ranked by citation volume, and provides reference values for MDIC where they are available.

As seen in **Table 3**, citation volume for outcomes ranged from a low of 10% to a high of 95%. Lack of consensus around how to apply these outcomes consistently is a recognized deficiency of the literature. Greater consensus around how to apply the set would help improve consistency across studies. Having widely accepted measurement timepoints (e.g., every 3 or 6 months), methods of aggregation (e.g., mean), and specific metrics (e.g., change from baseline) will aid in reducing the heterogeneity currently seen in clinical trials and assist with conducting comparative effectiveness reviews.

Table 3. Consensus table

Criteria	Results of evidence synthesis
Professional consensus statements	2 core sets; 4 endorsements of core sets
Stakeholders involved	Patients, clinical experts, researchers, Industry, epis, biostatisticians
Cochrane reviews in time range	4
AHRQ comparative effectiveness reviews	1
FDA Voice of the Patient reports	1
# of records used for outcomes extraction	20
# of identified outcomes (total)	25
# of prioritized outcomes	11
Citation v*olume, n range*	2-19
Citation v*olume, n median*	8
C*itation volume, % range*	10%-95%
Citation v*olume, % median*	40%

Discussion

After a comprehensive review of the literature and other evidence-based sources for relevant, critical outcomes to consider when evaluating different technologies to treat knee OA, a list of eleven prioritized outcome domains and their commonly used instruments was identified. This list is comprised of the outcome domains joint pain, physical function, HRQoL, patient satisfaction, function/functional ability, joint structure, stiffness, treatment failure, mortality, SAEs, and AEs. It is important to note that this report does not designate these outcomes as "core"; instead, it suggests that per the included literature, they are perceived as most critical and perhaps most amenable to being part of a COS.

There are several considerations regarding this prioritized list and its related instruments. Though SAEs and AEs were acknowledged as critical outcomes in the included literature, researchers will sometimes exclude them from a final COS since they are inherently captured in clinical trials to fulfill regulatory requirements. This technique serves the dual purpose of eliminating redundant outcomes and decreasing the size of a COS to promote uptake. Additionally, the prioritized outcomes identified via data synthesis were partially influenced by the COS developed by OMERACT-OARSI for knee/hip OA clinical trials, as well as the OMERACT COS for knee/hip total joint replacement (TJR) clinical trials. The considerable overlap between the reported outcomes list and both COS evidences this. In total, eight papers – six papers from the consensus development category and two from the SLR/CER category, which amounted to 40% of the included publications – discussed or mentioned these two core sets. The consensus processes used to define and ratify both core sets involved diverse groups of stakeholders and structured Delphi methodology. Also, publications not linked to the existing COS appeared to reaffirm their choice of outcomes.

Concerning instruments, the values for minimal differences seen in **Appendix C** were calculated within highly contextual bases and should be interpreted as such. Although these values are relevant to knee or knee/hip OA mainly within the context of specific surgeries, they cannot be generalized to all surgeries or medical devices being evaluated for knee OA. The challenge with MCIDs is that they can be highly variable due to differences in the calculation, patient population, intervention, disease severity, timepoints of analysis, and study setting (e.g., clinical trial versus practice). Moreover, "MCID" is often erroneously used interchangeably with other related terms such as MID, MCD, MDC, and MIC. For example, MCID indicates the smallest clinically meaningful change that might suggest a change in care management. In contrast, MDC is a statistical term that denotes the smallest detectable change considering measurement error. (Maredupaka S et al., 2020) Understanding the distinctions between terms of minimal difference is crucial to utilizing these values appropriately.

Instruments were prioritized in this review based on the number of citations, which suggests general acceptability of use by researchers. Validation within the intended patient population was completed for most prioritized instruments except for the VAS and Likert scales for patient satisfaction, and some are available within the public domain. Not all validation studies reported were conducted in the US or other English-speaking countries, which might limit their applicability to US-based studies. Furthermore,

despite our efforts to record instrument characteristics that could impact judgments around which ones should be used, there may be other characteristics or properties important to stakeholders that were not noted or difficult to find. For example, some participants may care about whether the development of instruments involved substantial patient input. The decision around which instruments to use can be swayed by the types of stakeholder groups and experiences involved in discussions.

VI. CMS Proposed Conclusions

CMS is developing this clinical endpoints guidance series to improve predictability and transparency of our evidence reviews. This review compiles a succinct list of outcomes and instruments that represent the most relevant, meaningful outcomes that may be used to evaluate knee osteoarthritis treatments. It also identifies available published evidence that defines clinically meaningful differences for each endpoint. This guidance is intended as a reference for clinical investigators who are developing knee osteoarthritis studies and for CMS staff who may review studies in this therapeutic area. As this is the first guidance document in the series, the format of these reviews may change over time. CMS is interested in public comment on this guidance document specifically, but also on how this series can deliver the greatest value to CMS and the public.

CMS has reviewed the consensus assessments described in the technology assessment (**Section V**) in this document and has concluded that this guidance can be proposed for public comment without need for a Medicare Evidence Development & Coverage Advisory Committee (MEDCAC).[6]

In general, when making national coverage determinations, CMS evaluates relevant clinical evidence to determine whether or not the evidence is of sufficient quality to support a finding that an item or service falling within a benefit category is reasonable and necessary for the diagnosis or treatment of an illness or injury or to improve the functioning of a malformed body member under section 1862(a)(1)(A) of the Social Security Act. CMS encourages investigators to review our guidance documents on Clinical Evidence Review[7] and Coverage with Evidence Development[8], as applicable, prior to designing clinical studies intended to satisfy the legal standard for national CMS coverage. Additionally, based upon the findings of the technology assessment on clinical endpoints for knee osteoarthritis, CMS can make the following recommendations for clinical studies for technologies in this therapeutic area:

Generalizability to the Medicare population:

CMS may have difficulty drawing conclusions regarding potential benefits and harms associated with an item or service if Medicare eligible beneficiaries are insufficiently represented in clinical studies.

[6] https://www.cms.gov/regulations-and-guidance/guidance/faca/medcac
[7] https://www.cms.gov/medicare-coverage-database/view/medicare-coverage-document.aspx?mcdid=34&docTypeId=1&sortBy=title&bc=16
[8] https://www.cms.gov/medicare-coverage-database/view/medicare-coverage-document.aspx?mcdid=35&docTypeId=1&sortBy=title&bc=16

- CMS recommends that investigators carefully consider the intended recipients of an item or service when designing clinical studies such that the findings may be credibly generalized to relevant Medicare beneficiaries, including important sub-populations.
- CMS recommends that studies avoid poorly justified exclusions based on age or common comorbidities that may limit generalizability of findings.
- CMS recommends that study investigators consider whether the context of care delivery within clinical studies may be reasonably generalized to the context where Medicare beneficiaries are expected to receive care.

Clinical Endpoints & Clinically Meaningful Differences:

Traditionally, CMS relies heavily on health outcomes data to make NCDs. Where there is limited evidence on the health outcomes for individuals in the Medicare beneficiary population, there may not be sufficient evidence to support favorable national Medicare coverage under section 1862(a)(1)(A) of the Act.

- CMS recommends that investigators consider the 11 prioritized outcome domains identified in **Table 2** of this technology assessment when designing knee osteoarthritis clinical studies.
- CMS does not recommend any specific clinical endpoint or instrument that was identified in this review, but generally recommends that clinical studies include a range of outcomes that reflect multiple attributes of an item or service within a clinical study.
- When chosing among the available clinical endpoint options, CMS recommends that clinical studies prioritize validated endpoints / instruments and those with well-established/published minimal clinically important differences (MCID) values because study findings are more readily interpreted.

Duration of follow-up:

CMS may have difficulty reaching conclusions regarding potential risks and harms associated with an item or service if studies lack sufficient follow-up to demonstrate the durability of improved health outcomes.

- For osteoarthritis, CMS generally recommends that clinical studies include follow-up of at least one year for functional and/or patient reported outcomes.
- CMS generally recommends that clinical studies include follow-up of at last two years to establish the durability of implanted devices, though shorter or longer follow-up may be reasonable in some circumstances.

VII. Appendix A. Search Strategies

Set #	Strategy	Search Yield
PubMed/MEDLINE		
21-Feb-2022		
#7	#5 AND #6 Filters: in the last 10 years, Humans, English	**329**
#6	("comet initiative"[Title/Abstract] OR "omeract"[Title/Abstract] OR "oarsi"[Title/Abstract]) in the last 10 years, Humans, English	876
#5	#1 AND #2 Filters: in the last 10 years, Humans, English	706
#4	#1 AND #2 Filters: English, Humans	1,051
#3	#1 AND #2	1,233
#2	(((((core measure[Title/Abstract]) OR (core measures[Title/Abstract])) OR (core outcome[Title/Abstract])) OR (core outcomes[Title/Abstract])) OR (outcome measure[Title/Abstract])) OR (outcome measures[Title/Abstract]) OR (core domain[Title/Abstract])	245,597
#1	(knee osteoarthritis[Title/Abstract]) OR (osteoarthritis of the knee[Title/Abstract])	16,135
Embase/Science Direct		
14-Feb-2022		
#7	#3 AND #4 AND #5 AND #6	**527**
#6	2012:py OR 2013:py OR 2014:py OR 2015:py OR 2016:py OR 2017:py OR 2018:py OR 2019:py OR 2020:py OR 2021:py OR 2022:py	15,934,450
#5	'human' AND 'english'	25,424,914
#4	'comet' OR 'omeract' OR 'oarsi'	34,103
#3	#1 AND #2	8,257
#2	core AND ('outcome'/exp OR outcome) OR (core AND ('outcomes'/exp OR outcomes)) OR (core AND measures) OR 'consensus'/exp OR (('consensus'/exp OR consensus) AND ('development'/exp OR development)) OR (('practice'/exp OR practice) AND ('guideline'/exp OR guideline)) OR (('outcomes'/exp OR outcomes) AND ('research'/exp OR research)) OR consensus OR 'consensus development' OR 'practice guideline' OR 'outcomes research'/exp OR 'outcomes research' OR (('outcome'/exp OR outcome) AND ('assessment'/exp OR assessment))	2,275,845
#1	('knee osteoarthritis'/exp OR 'knee osteoarthritis' OR 'osteoarthritis of the knee') NOT ('elbow' OR 'hip' OR 'ankle' OR (spine AND 'disease') OR (low AND back AND 'pain') OR (spinal AND cord AND 'disease') OR 'lumbar' OR 'wrist' OR 'femoral' OR 'acl' OR 'abdominal' OR 'weight loss' OR 'mandib*' OR 'tmj' OR 'carpal tunnel' OR 'hand' OR 'head' OR 'neck' OR 'cervical spine' OR 'shoulder' OR 'foot' OR 'arm' OR 'abdominal surgery' OR 'spine surgery' OR 'finger')	26,827
Cochrane Library		
22-Feb-2022		
#1	Searched "osteoarthritis of the knee" (https://www.cochrane.org/search/site/osteoarthritis%20of%20the%20knee); 14 protocols and 52 reviews retrieved	66
#2	Hand searched all reviews listed under the "Full list" of reviews on the Cochrane Musculoskeletal Group webpage (https://musculoskeletal.cochrane.org/evidence)	218
Agency for Healthcare Research and Quality (AHRQ)		
20-Feb-2022		
	Searched "comparative effectiveness reviews" and "major joint replacement" and "knee osteoarthritis" (https://www.ahrq.gov/)	116
Food and Drug Administration (FDA)		
18-Feb-2022		
#1	Hand searched all records on "FDA-led Patient-Focused Drug Development (PFDD) Public Meetings" webpage (https://www.fda.gov/industry/prescription-drug-user-fee-amendments/fda-led-patient-focused-drug-development-pfdd-public-meetings)	30
#2	Hand searched all records on "Externally-led Patient-Focused Drug Development Meetings" webpage (https://www.fda.gov/industry/prescription-drug-user-fee-amendments/externally-led-patient-focused-drug-development-meetings)	0

#3	Searched “Osteoarthritis” in FDA Guidance Document Search portal (https://www.fda.gov/regulatory-information/search-fda-guidance-documents)	1
#4	Searched “FDA voice of the patient reports osteoarthritis” via Google search engine and retrieved a webpage for “OA Voice of the Patient | Arthritis Foundation” (https://www.arthritis.org/science/events-publications/oa-vop). Full report from the meeting was retrieved, dated Mar 2017.	
National Institutes of Health (NIH)		
21-Feb-2022		
#1	Hand searched all publications listed on the “Publications Using OAI Data by Year” webpage for the Osteoarthritis Initiative - National Institute on Aging (https://nda.nih.gov/oai/publications#Articles)	557
#2	Searched through National Institute of Arthritis and Musculoskeletal and Skin Diseases webpage (https://www.niams.nih.gov/grants-funding/supported-scientific-areas/orthopaedic-research-program)	0
Core Outcome Measures in Effectiveness Trials (COMET) Initiative		
14-Feb-2022		
#1	Searched “knee osteoarthritis” in database (https://www.comet-initiative.org/)	5
#2	Searched “total joint replacement” in database (https://www.comet-initiative.org/)	6
Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT)		
22-Feb-2022		
#1	Searched “knee osteoarthritis” in webpage (https://omeract.org/) and was led to the Hip & Knee Osteoarthritis Working Group webpage (https://omeract.org/working-groups/hip-knee-osteoarthritis-core-set/)	2
Osteoarthritis Research Society International (OARSI)		
22-Feb-2022		
#1	Searched “knee osteoarthritis” in webpage (https://oarsi.org/)	26
European Alliance of Associations for Rheumatology (EULAR)		
20-Feb-2022		
#1	Searched “knee osteoarthritis” in webpage (https://www.eular.org/index.cfm)	359
International Consortium for Health Outcomes Measurement (ICHOM)		
20-Feb-2022		
#1	Searched “knee osteoarthritis” in ICHOM Connect webpage (https://connect.ichom.org/)	5

VIII. Appendix B. Table of Included Publications

Systematic and Comparative Effectiveness Reviews			
Study Details	**Included Studies**	**Outcomes Identified**	**Instruments/Items Identified**
Reference: Konnyu et al. 2021 **Source:** AHRQ **Objective:** To inform healthcare stakeholders of care for patients who have undergone or will undergo TKA or THA for OA about (p)rehabilitation options **Quality rating of SR: Good** Well-conducted recent review using AHRQ-EPC methodology, searched multiple databases, performed risk of bias and strength of evidence assessments, valid conclusions. <u>Risk of bias of included papers:</u> Moderate to High risk of bias **Funding Source:** AHRQ contract **Databases:** MEDLINE, PsycINFO, EMBASE, CENTRAL, CINAHL, Scopus, clinicaltrials.gov	**Intervention:** Active prehabilitation or rehabilitation for TKA or THA **Comparator:** No prehabilitation prior to TKA or THA, different rehabilitation programs after TKA or THA **Number of included studies:** 83 **Number of Patients:** 14,533 RCTs: 8,397 NRCSs: 6,156 **Diagnosis:** Knee or hip OA **Mean Age, years (range):** Prehab for TKA: NR (63-72) **Female, n (range %):** Prehab for TKA: NR (27-82)	**Identified outcomes:** Performance-based measures: • Mobility of joint function (range of motion)* • Power and tone of muscle (strength)* • Joint stability • Endurance • Gait • Balance PROs: • ADL* • Patient satisfaction with care* • HRQoL • Pain • Injury related to arthroplasty • Time lost from work Healthcare utilization: • Hospital- or surgical clinic-based procedures postoperatively* • Hospital readmission • Postoperative care • Length of stay (postoperative) • Length of postoperative rehabilitation needed • Posthospital disposition *Pre-specified important/priority outcomes	Authors did not specify instruments of interest in the CER; rather, they noted the heterogeneity of outcome measures used. <u>Note</u>: This review is not entirely relevant to our scope due to the intervention of focus, but it includes outcomes relevant to TKR that are prioritized. Outcomes were selected with input from a range of stakeholders.

Included study types: 78 RCTs, 5 adjusted NRCSs **Inclusion Criteria:** Adults 18+ years undergoing elective TKA or THA for primary OA and active habilitation, unilateral TJR, ≥20 patients for NRCSs or per arm for RCTs, cost-effective analyses, published from 01Jan2005 to 03May2021, ≥50% of surgeries occurred after 2005 **Exclusion Criteria:** ≥10% patients underwent partial joint replacement for causes other than primary OA, emergency surgery, revision joint replacement, or bilateral TJR, single arm studies, crossover studies, case reports or series, case-controlled studies	**Mean duration of follow-up, months (range):** Prehab for TKA: NR (NR)		
Reference: Palmer et al. 2019 **Source:** CMSG **Objective:** To assess the benefits and harms of surgical intervention for the management of symptomatic mild to moderate knee OA **Quality rating of SR: Good** Well-conducted recent review using Cochrane methodology, searched	**Intervention:** Surgery (arthroscopy, load-modifying procedures, knee replacement) **Comparator:** Non-surgical interventions (e.g., sham, placebo, exercise/PT, analgesic), injectables, or a different type of surgery than the intervention **Number of included studies:** 5	**Major outcomes:** • Pain • Physical function • Radiographic joint structure changes • HRQoL • Short-term SAEs • Re-operation rate or revision of TKR • Withdrawals due to AEs **Minor outcomes:** NA	**Pain (hierarchy, highest to lowest):** • Overall pain • Pain on walking • WOMAC pain subscale • Pain on activities other than walking • WOMAC global scale • Lequesne OA index global score • Other algogfunctional scale • PGA **Physical function (hierarchy, highest to lowest):** • Global disability score • Walking disability • WOMAC disability subscore

multiple databases, performed risk of bias and strength of evidence assessments, valid conclusions. Risk of bias of included papers: Moderate to High risk of bias **Funding Source:** NR **Databases:** CENTRAL, MEDLINE, EMBASE, clinicaltrials.gov, WHO ICTRP **Included study types:** RCTs, quasi-RCTs **Inclusion Criteria:** Adults with diagnosis of mild to moderate knee OA, full text, abstract only, unpublished data, publications from database inception to 24 May, 2018 **Exclusion Criteria:** Trials with asymptomatic individuals, end-stage OA (full-thickness cartilage loss >1 cm), bone deformity (Kellgren-Lawrence grade 4, degenerate meniscal tears but no radiographic or MRI evidence of OA, history of trauma, inflammatory arthropathy, metabolic bone disease, RA	**Number of Patients:** 566 **Diagnosis:** Mild to moderate knee OA, defined as knee pain and radiographic evidence of non-end stage OA (Kellgren-Lawrence grades 1, 2, 3 or equivalent on MRI/arthroscopy) **Mean Age, years (SD):** NR **Female, n (%):** NR **Mean duration of follow-up, months (range):** NR	Note: Authors did not provide reasons for choice of outcomes and hierarchy of instruments.	• Composite disability scores other than WOMAC • Disability other than walking • WOMAC global scale • Lequesne OA index global score • Other algofunctional scale **Radiographic joint structure changes:** • Minimum joint-space width • Median joint-space width • Semi-quantitative measurement **HRQoL:** • Generic or overall tools • Disease-specific tools • SF-36 mental component score
Reference: Lange et al. 2017 **Source:** PubMed/EMBASE	**Intervention:** Primary TKA (99 studies), primary TKA and revision surgery (1 study)	**Outcome domains, n**: 34 **Outcomes, n**: 379	**Instruments, n**: 111 **Items, n**: 3,265

Objective: To evaluate and critically appraise the use of outcome domains and measurement instruments and to assess their accordance with the OMERACT COS
 	
Quality rating of SR: Poor
 	
Well-conducted but older review, searched two databases up to 2014, no standard appraisal of studies
 	
<u>Risk of bias of included papers:</u> Not assessed; majority of studies (70%) were non-RCTs. ICA judges the evidence base to be at Moderate to High risk of bias.
 	
Funding Source: No specific grant from any funding agency in the public, commercial, or nonprofit sectors
 	
Databases: MEDLINE, EMBASE
 	
Included study types: 30 RCTs, 61 prospective cohort, 9 registry
 	
Inclusion Criteria: ≥50 patients, ≥1 year follow-up, ≥1 intervention/study group, clearly defined outcomes and instruments, published 2007 to 26 Aug, 2014
	Comparator: NA
	
	Number of included studies:
	
	Total: 100
	
	Knee OA only: 91
	
	Knee/hip OA: 9
	
	Knee/other joint OA: 1
	
	Number of Patients: NR
	
	Diagnosis: Knee OA
	
	Mean Age, years (SD): NR
	
	Female, n (%): NR
	
	Mean duration of follow-up, months (range):
	
	RCT: 30 (12-204)
	
	Prospective cohort: 43 (12-240)
	
	Registry: 40 (12-60)
		OMERACT COS domains (% studies):
		
		• Pain (85)
		• Physical function (73)
		• Physical functioning subdomains
		  ○ Walking, stairs (84)
		  ○ Walking, flat surface (83)
		  ○ Stability, knee (78)
		• PGA (21)
		• Joint imaging ≥1 year (27)
		• QoL (NR)
		
		Other outcome domains commonly investigated, i.e., ≥40% studies (% studies):
		
		• Managing household (71)
		• Personal hygiene (69)
		• Range of motion (66)
		• Ability to use public transport (60)
		• Support dependency, instrumental (46)
		• Malalignment (44)
		• Mental health (41)
			Instruments assessing OMERACT COS domains, n:
			
			• Pain: 23
			• Physical function: 40
			• PGA: 20
			• Joint imaging ≥1 year: 3
			• QoL: NR
			
			Most frequently used measurement instruments (% studies):
			
			• KSS (11.8)
			• OKS (8.4)
			• WOMAC (7.1)
			• Radiographic measures (6.1)
			• SF-36 (5.5)

Exclusion Criteria: 20% patients received revision surgery or unicondylar knee arthroplasty, TKA performed due to trauma or tumor, non-English/-German publications			
Reference: Duivenvoorden et al. 2015 **Source:** CMSG **Objective:** To assess the benefits and harms of braces and foot/ankle orthoses in the treatment of patients with OA of the knee **Quality rating of SR: Fair** Well-conducted older review (searches up to 2014) using Cochrane methodology, searched multiple databases, performed risk of bias and strength of evidence assessments, valid conclusions. <u>Risk of bias of included papers:</u> Moderate to High risk of bias **Funding Source:** NR **Databases:** CENTRAL, MEDLINE, EMBASE	**Intervention:** Knee brace (valgus, neutral, neoprene sleeve) or foot/ankle orthoses (laterally or medially wedged insole, neutral insole, variable or constant stiffness shoe) **Comparator:** Active control (e.g., education, PT, restricted activity, orthosis, surgery) or no treatment **Number of included studies:** 13 **Number of Patients:** 1,356 **Diagnosis:** Early to severe knee OA (Kellgren & Lawrence grades 1-4) **Mean Age, years (range):** 62 (48-75) **Female, n (%):** NR All patients were female in 2 trials.	**Major outcomes:** • Pain • Function • Stiffness • QoL • Treatment failure (need to undergo surgery) • SAEs • Non-serious AEs (total #) **Minor outcomes:** • Radiographic scores • Compliance • Walking distance <u>Note</u>: Authors did not provide reason for choice of outcomes.	**Pain:** • VAS • WOMAC pain subscore • Lequesne index **Function:** • WOMAC function subscore • HSS knee scores • MACTAR • Lequesne index **HRQoL:** • EQ-5D <u>Note</u>: Authors recommended using WOMAC since it is validated for measurement of OA.

Included study types: RCTs, CCTs **Inclusion Criteria:** Adults (18+ years) with early to severe knee OA treated with a knee brace or orthosis or given no treatment, published between 2007 and Mar2014 **Exclusion Criteria:** NR	**Mean duration of follow-up, months (range):** NR		
Reference: Hofstede et al. 2015 **Source:** CMSG **Objective:** To assess the benefits and harms of mobile bearing compared with fixed bearing cruciate retaining TKA for functional and clinical outcomes in patients with OA or RA **Quality rating of SR: Fair** Well-conducted older review (searches up to 2014) using Cochrane methodology, searched multiple databases, performed risk of bias and strength of evidence assessments, valid conclusions. Risk of bias of included papers: Moderate to High risk of bias **Funding Source:** NR	**Intervention:** Mobile bearing (meniscal or rotational) cruciate retaining TKA **Comparator:** Fixed bearing (polyethylene) cruciate retaining TKA **Number of included studies:** 19 **Number of Patients:** 1,641 OA, n (%): 1,616 (98.5) RA, n (%): 25 (1.5) **Diagnosis:** Knee OA or RA **Mean Age, years (SD):** NR **Female, n (%):** NR	**Major outcomes:** • Knee pain • Clinical and functional scores • HRQoL • Revision surgery • Mortality • Reoperation rate • SAEs (excluding revision surgery, mortality, and reoperation rate) **Minor outcomes:** • Radiolucent lines • Femorotibial alignment • Performance outcomes (flexion, extension, range of motion) Note: Authors stated that the outcome measurements had to be a functional or clinical measure.	**Knee pain:** • VAS • KSS pain subscore • WOMAC pain subscore • HSS pain subscore • OKS pain subscore **Clinical and functional:** • WOMAC • KOOS • OKS • HSS • Bristol Knee Score • IKDC • KSS functional and clinical subscores • KSS total score **HRQoL:** • SF-36 • SF-12

Databases: CENTRAL, PubMed, EMBASE, CINAHL, Web of Science, clinicaltrials.gov, Multiregister, Current Controlled Trials, WHO ICTRP **Included study types:** RCTs **Inclusion Criteria:** Patients who have had TKA for OA or RA, follow-up of ≥6 months, peer-reviewed publication, published from database inception to 11 Feb, 2014 or 27 Feb, 2014 **Exclusion Criteria:** TKA after prior patellectomy and osteotomy	**Mean duration of follow-up, months (range):** NR		
Reference: Brouwer et al. 2014 **Source:** CMSG **Objective:** To assess the benefits and harms of osteotomy for treating patients with knee OA **Quality rating of SR: Fair** Well-conducted older review (searches up to 2013) using Cochrane methodology, searched multiple databases, performed risk of bias and	**Intervention:** Osteotomy **Comparator:** Inactive interventions (i.e., SoC including PT), active non-operative interventions (e.g., insoles, braces, injections), or operative interventions (different osteotomy techniques, TKA, etc.) **Number of included studies:** 21 **Number of Patients:** 1,065	**Major outcomes:** • Treatment failure rate (incidence of TKA, time to revision) • Pain • Function • HRQoL • SAEs • Neurovascular complications • Mortality • Reoperation rate **Minor outcomes:** • Performance (surgery time, hospital stay, time to healing, postoperative correction achievement, inferior limb length)	**Pain:** • VAS • WOMAC pain subscore **Function:** • KOOS • WOMAC • Lysholm • HSS knee score • KSS • BOA knee score • JOA knee score • Gait analysis **HRQoL:** • NHP score • EuroQol

strength of evidence assessments, valid conclusions. Risk of bias of included papers: Moderate to High risk of bias **Funding Source:** NR **Databases:** CENTRAL, MEDLINE, EMBASE, clinicaltrials.gov **Included study types:** RCTs, CCTs **Inclusion Criteria:** Adult patients (18+ years) with unicompartmental knee OA, published until Nov, 2013 **Exclusion Criteria:** NR	**Diagnosis:** Unicompartmental knee OA confirmed by radiography or arthroscopy **Mean Age, years (range):** 60 (42-67) **Female, n (%):** NR **Mean duration of follow-up, months (range):** NR	• AEs • Patient satisfaction • PGA • Joint imaging • Walking distance (indirect measure of function) • Range of motion • Collateral laxity Note: Authors stated that choice of outcome measures was originally based on OMERACT COS. However, the outcome measures were changed to those recommended by CMSG editors.	**PGA:** • Modified Cincinnati Rating System Questionnaire • Wallgren-Tegner **Patient satisfaction:** • VAS
Reference: Howe et al. 2012 **Source:** PubMed/EMBASE **Objective:** To report on the clinimetric properties of outcome measures for use in clinical practice for adults with musculoskeletal conditions of the knee **Quality rating of SR: Fair**	**Intervention:** NA **Comparator:** NA **Number of included studies:** 47 **Number of Patients:** 12,265	This review did not focus on outcome domains or outcomes.	**Instruments, n:** 37 **Instruments demonstrating adequate "truth" and "discrimination" per OMERACT filter, with clinimetric properties assessed in knee or knee/hip OA patients (n studies):** • AAOS Outcomes Instruments* (2) • IKDC* (4) • KOOS* (1) • LEFS (2) (knee/hip) • WOMAC (7)

Well-conducted older review (searches up to 2011), searched multiple databases, outcome measures assessed using the clinimetric properties of the 'Truth' component of the OMERACT
 	
Filter, valid conclusions.
 	
Risk of bias of included papers: Not applicable for this type of outcome review
 	
Funding Source: Scottish Funding Council, Scottish Government, Chief Scientist Office, NHS Education for Scotland, Scottish Government Health Department
 	
Databases: MEDLINE, EMBASE, AMED, CINAHL
 	
Included study types: NR
 	
Inclusion criteria:
 	
For studies: Full text papers where examination of clinimetric properties of joint specific or generic outcome measures, tested in the population of interest, was the primary aim, published from database inception to 20 Feb, 2011
 	
For outcome measures: ≤20 minutes to administer, easily accessible, low cost
	Diagnosis: Musculoskeletal conditions of the knee (i.e., OA, ligament injuries, meniscal lesions, patellofemoral pain)
	
	Mean Age, years (range):
	
	NR (16-100)
	
	Female, n (%): 3,007 (NR) in 25 of the 47 studies
	
	Mean duration of follow-up, months (range): NR
			Note: Consensus on whether instruments demonstrated these characteristics was determined by an expert panel (n=8) comprised of a GP, orthopedic surgeon, bioengineer, OT, and four PTs, who underwent discussions and open voting. To meet OMERACT filter requirements for truth, ≥1 element of validity was reported, and the construct measured against ≥1 instrument. For discrimination, both reliability and sensitivity were reported.
			
			Other instruments identified, with clinimetric properties assessed in knee or knee/hip OA patients (n studies):
			
			• ADLS of the knee outcome survey* (2)
			• AQoL (1)
			• EQ-5D* (1)
			• HAP (1)
			• HSS score (1)
			• ICOAP (1) (knee/hip)
			• IOSK (1) (knee/hip)
			• J-MAP (1)
			• KSS (1)
			• OKS (1)
			• SF-36 (2)
			• Timed physical tests
			  ○ ALF (1)
			  ○ TUG test (2) (knee/hip)
			  ○ Level walking (1)
			  ○ Stair walking (2)
			  ○ Self-paced walking (1)
			
			*Clinimetric properties assessed in a variety of knee conditions that may include knee or hip OA

per patient use, minimal equipment, minimal space to administer **Exclusion Criteria (for studies):** Study population included those with inflammatory arthritis, acute trauma, congenital abnormalities, osteoporosis, post-surgical interventions, or non-musculoskeletal condition as primary complaint, unpublished reports, abstracts, brief and preliminary reports, unpublished data, descriptive studies on instrument or focus group development, use of laboratory or clinically based equipment, radiographic/imaging techniques, diagnostic tests, or studies that did not describe outcome measure in sufficient detail, non-English publications			
Consensus Development			
Study Details	**Stakeholder Groups**	**Outcomes Identified**	**Instruments/Items Identified**
Reference: Smith et al. 2019 **Source:** PubMed/EMBASE, OMERACT **Objective:** To review and update the 1997 OMERACT-OARSI COS developed for knee/hip OA clinical trials **Consensus Design:** Candidate outcomes list generated from review of COMET database evaluated prior to Delphi to determine missing or duplicative	**Total Number of Stakeholders:** **Review Panel, n (%):** 70 (100) • Patients: 35 (50.0) • Health professionals: 34 (48.6) • Researchers: 1 (1.4) **Delphi Panel, n**: 426 Delphi R3, n (%): 119 (100) • Patients: 42 (35.3) • Health professionals: 29 (24.4) ○ Orthopedic surgeons: 2 (1.7) • Researchers: 42 (35.3) • Industry: 6 (5.0)	**Outcome domains, n**: 21 **Items, n:** 37 **Plenary voting for inclusion (% votes):** **"Mandatory" domains and items:** • Pain (100) • Physical function (100) • QoL (90) • PGA of target joint (91) • Joint structure (80)	**Identified Items:** • **Pain**: Overall, at rest, during the night, during the day • **Physical function**: Mobility (walking), patient-reported leg function, personal ADL (washing, dressing, toileting), sports, exercise, physical activity • **QoL**: Patient-reported overall effect of OA on person with OA • **PGA of target joint**: Patient-reported overall improvement of disease • **Joint structure**: Imaging (radiograph, MRI, ultrasound reflecting changes in structure)

outcomes and to pilot Delphi survey. Delphi consisted of 3 rounds of survey voting using a 9-point scale, wherein a rating of 1-3 meant “not that important”, 4-6 “important”, and 7-9 “critically important”. Missing outcomes could be suggested in surveys, and results were distributed between rounds. Outcomes excluded if rated 1-3 by ≥15% or 4-6 by ≤70% of one or more stakeholder groups. Plenary meeting held after R3 for final ratification of “mandatory” (rated "critical" by ≥70% of patients and other stakeholders) and “important but optional” (rated "critical" by ≥70% of either patients or others but not both) domains. **Quality:** Specified research setting, health conditions, and populations covered by COS; did not specify specific interventions; involved a diverse group of stakeholders that included a substantial number of patients with the condition and healthcare professionals; scoring and voting process clearly defined **Funding Source:** NIHR Oxford Biomedical Research Centre, NIHR Leeds Biomedical Research Centre	**Plenary Panel, n:** 102 Consisted of clinicians, patients, patient advocates, researchers, industry, and methodologists. **Diagnosis, Delphi R3, n (%):** Knee OA: 22 (18.5) Knee/hip OA: 25 (21.0) Hip OA: 10 (8.4) Not affected by OA: 62 (52.1) **Total countries represented, n:** **Review Panel:** 3 countries **Delphi Panel:** 25 countries Delphi R3 representation, n stakeholders: UK: 35, Canada: 14, USA: 13, Australia: 36, Spain: 2, Denmark: 3, Netherlands: 2, Germany: 2, Belgium: 1, Iceland: 1, Norway: 1, Italy: 1, France: 2, India: 1, Sweden: 3, Russia: 1, Singapore: 1 **Plenary Panel**: NR **Mean Age, Years (SD):** NR **Female, Delphi R3, n (%):** 73 (61.3)	**“Important but Optional” domains and items:** • Participation (95) • Psychosocial impact (71) • Sleep (81) • Costs (77) **“Research Agenda” domains and items** (82% overall): • CGA of target joint • Flare • Inflammation • Cognitive function • Fatigue • Effect on family/caregiver	• **Participation**: Role function (ability to do work or vocational activities) • **Psychosocial impact:** Control over disease (self-efficacy, understanding of condition), perceived ability to cope with OA, social withdrawal and isolation • **Sleep:** Falling and staying asleep • **Costs:** Healthcare use (costs of pain killer use, hospital admission, consultation with clinicians), time to surgery (TJR) • **CGA of target joint:** Clinician-reported overall improvement of disease

Reference: Hoang et al. 2017 **Source:** COMET Initiative **Objective:** To examine whether hip/knee arthroplasty patients and arthroplasty surgeons achieve consensus the preliminary OMERACT core domain set for use in knee/hip TJR clinical trials **Consensus Design:** Participants completed an online survey that asked them to rate each core and non-core domain on a 9-point scale, wherein 1-3 meant "limited/no importance for patients", 4-6 "important but not critical", and 7-9 "critical". Median (IQR) values were calculated. Complete consensus was defined as both groups (patients and surgeons) rating each core domain as "critical". **Quality:** Specified research setting, health condition, population, and intervention; inclusion of only patients with the condition and surgeons acceptable given goal was to endorse a previously drafted COS, though patients substantially outnumbered surgeons; scoring process and consensus clearly defined **Funding Source:** NR	**Total Number of Stakeholders, n (%):** 1,316 (100) • Patients: 1,295 (98.4) • Surgeons: 21 (1.6) **Diagnosis, n (%):** Knee/hip OA only: 1,071 (82.7) Knee/hip RA only: 34 (2.6) Another type of arthritis or joint condition: 190 (14.7) THA: 819 (63.2) TKA: 476 (36.8) **Total countries represented, n:** NR **Mean Age, Years (SD):** NR Patients: 1,147 (88.6%) were ≥55 years Surgeons: 12 (57.1%) were ≥55 years **Female, n (%):** 743 (56.4) Patients: 742 (57.3) Surgeons: 1 (4.8)	**Core outcome domains, median rating (IQR) patients vs. surgeons:** • Joint pain: 9 (8, 9) vs. 9 (7, 9) • Function or functional ability: 9 (8, 9) vs. 8 (7, 9) • Patient satisfaction: 9 (8, 9) vs. 8 (8, 9) • Revision surgery: 8 (5, 9) vs. 8 (7, 8) • AEs: 8 (7, 9) vs. 7 (6, 9) • Death: 9 (6, 9) vs. 9 (7, 9) **Non-core domains, median rating (IQR) patients vs. surgeons:** • Cost: 7 (5, 8) vs. 6 (5, 6) • Participation in work and social activities: 8 (6, 9) vs. 8 (6, 8) **Recommended "research agenda" domains by patients, n (%):** • Range of motion: NR • Time to recovery and rehabilitation: 131 (60)	Specific instruments were not discussed.

Reference: Singh et al. 2017 **Source:** COMET Initiative **Objective:** To discuss and endorse the preliminary OMERACT TJR core domain set for THA and TKR for endstage arthritis refractory to medical treatment among a diverse group of stakeholders **Consensus Design:** TJR Working Group discussed the preliminary core domain set from OMERACT 2014, made modifications, and identified challenges with domain measurement **Quality:** Specified health condition, population, and intervention; intended setting not as clear; involved a diverse group of stakeholders that included healthcare professionals and patients, though the number of stakeholders was relatively small; diagnoses of patients unclear; scoring process and consensus definition not defined **Funding Source:** NR	**Total Number of Stakeholders, n (%):** 26 (100) • Patients: 3 (11.5) • Orthopedic surgeons: 2 (7.7) • PTs/OTs: 2 (7.7) • Methodologists: 2 (7.7) • Clinicians/researchers: 15 (57.7) **Diagnosis, n (%):** NR **Total countries represented, n:** NR **Mean Age, Years (SD):** NR **Female, n (%):** NR	**"Mandatory" outcome domains:** • Pain • Function • Patient satisfaction • Revision surgery • AEs • Death **"Important but optional" domains:** • Cost • Participation **"Research agenda" domains:** • Range of motion	Specific instruments were not discussed. However, the working group obtained consensus to assess multidimensional instruments for pain and function/functional ability and use subscales of such instruments as measures of the core domains.
Reference: Singh et al. 2017 **Source:** COMET Initiative	**Total Number of Stakeholders, n:** 87 • Group 1 (leadership of IOS): 18 • Group 2 (members of AAOS-Outcome SIG or ORS): 69	**Core outcome domains, median rating (IQR), group 1 vs. group 2:** • Joint pain: 8 (8, 9) vs. 8 (7, 9) • Function or functional ability (ability to function in society, work, productivity,	Specific instruments were not discussed.

Objective: To achieve international consensus by orthopedic surgeons on the OMERACT core domain set for TJR clinical trials **Consensus Design:** Two independent surveys distributed to different groups of orthopedic surgeons that asked participants to rate the importance of the preliminary core domain set from OMERACT 2014 on a 1-9-point scale, wherein a rating of 1-3 meant "limited importance", 4-6 "important", and 7-9 "critically important". Complete consensus was defined as both groups rating each core domain as "critical". Median (IQR) ratings for each core domain within each group were calculated. **Quality:** Specified research setting, health condition, population, and intervention; involvement of only orthopedic surgeons acceptable given goal was to endorse a previously drafted COS in this particular group; scoring process and consensus clearly defined **Funding Source:** No direct funding	**Diagnosis, n (%):** NA **Total countries represented, n:** NR **Mean Age, Years (SD):** NR Group 1: 67% were ≥55 years Group 2: 64% were ≥55 years **Female, n (%):** Group 1: 0 (0), Group 2: 15 (22)	employability, disability, work disability): 8 (8, 8) vs. 8 (7, 9) • Patient satisfaction (with outcome or procedure): 8 (7, 9) vs. 8 (7, 8) • Revision surgery: 7 (6, 9) vs. 8 (6, 8) • AEs: 7 (5, 8) vs. 7 (6, 9) • Death: 7 (7, 9) vs. 8 (5, 9) **Non-core domains, median rating (IQR):** • Cost: 6 (5, 7) vs. 6 (5, 7) • Patient participation: 6.5 (5, 7) vs. 6 (5, 8)	
Reference: Rolfson et al. 2016 **Source:** COMET Initiative, ICHOM	**Total Number of Stakeholders, n (%):** 22 (100) • Patients: 2 (9.0) • OA healthcare and research experts: 20 (91.0)	**Core outcome categories and domains (% votes):** PROs: • Hip/knee pain (100)	**Identified items/instruments:** • **Pain**: HOOS-PS or KOOS-PS • **Function**: NRS or VAS • **HRQoL**: EQ-5D-3L, SF-12, or VR-12

Objective: To define a minimum standard set of outcome measures and case-mix factors for evaluating, comparing, and improving the clinical care of patient with hip or knee OA, with a focus on outcomes that matter most to patients **Consensus Design:** An international panel reviewed existing literature and practices for assessing outcomes of OA therapies, including surgery. A series of 8 teleconferences incorporating a modified Delphi process using 6 post-teleconference online surveys were held to reach consensus. Included items in the standard set required 67% majority vote. Items with <67% votes were either excluded or revised and presented again for discussion/voting at subsequent teleconference. **Quality:** Specified practice setting, health condition, and population; intervention partially specified; involved a diverse group of stakeholders including patients and healthcare professionals relevant to the condition, though the number of stakeholders was relatively small; patients noticeably outnumbered by the healthcare professionals and diagnoses unknown; consensus definition described; scoring process unclear	○ Included arthroplasty register experts, orthopedic surgeons, primary care physicians (n=3), rheumatologists (n=3), and PTs (n=2) • Response rates for surveys (%): 90, 85, 71, 84, 85, 84 **Diagnosis:** NR **Total countries represented, n:** 10 USA: 9, Australia: 4, UK: 3, Canada: 1, Indonesia: 1, Morocco: 1, New Zealand: 1, Saudi Arabia: 1, Sweden: 1, Netherlands: 1 **Mean Age, Years (SD):** NR **Female, n (%):** NR	• Hip/knee function (88) • HRQoL (100) • Work status (77) • Satisfaction with treatment results (88) Surgical outcomes: • Death (100) • Admissions (88) • Reoperation (94) Disease progression: • Treatment progression (82) • Care utilization (82) Note: This review is not entirely relevant to our scope due to the intended setting (clinical care), but it includes relevant outcomes obtained through consensus.	• **Work status:** Unable to work due to OA, unable to work due to a condition other than OA, not working by choice, seeking employment, part-time, full-time • **Satisfaction with results:** 5-point Likert scale ranging from "very satisfied" to "very unsatisfied" • **Death:** All-cause 30-day mortality • **Admissions**: All-cause 30-day readmissions • **Reoperation**: Any consecutive major or minor open surgery or revision • **Treatment progression**: Treatments undergone in last year for OA-related problems (i.e., information/advice, self-managed care, nonsurgical clinical care, surgery) • **Care utilization:** Health care providers seen in past year for OA-related problems (i.e., health educator/peer support group, dietician, PT or general practitioner, rheumatologist, orthopedic surgeon, alternative health practitioner) Note: Measure selection was based on assessment of domain coverage, psychometric properties, feasibility, and clinical interpretability.

Funding Source: Hoag Orthopedic Institute, Connecticut Joint Replacement Institute, Harvard Pilgrim Health Care			
Reference: Christensen et al. 2015 **Source:** PubMed/EMBASE **Objective:** To evaluate the hierarchy of pain-related continuous outcome measurement instruments recommended for MTA and SLR of knee OA clinical trials **Consensus Design:** A stakeholder panel participated in a workshop discussion to formulate recommendations. **Quality:** Specified research setting, health condition, and population; did not specify intervention; involved a diverse stakeholder panel, though the number of stakeholders was small; did not specify how many patients were involved or their diagnoses; scoring process and consensus definition not defined **Funding Source:** NR	**Total Number of Stakeholders:** 9 Panel consisted of clinical epidemiologists, clinicians, patients, biostatisticians, and the Cochrane Collaboration Editor-in-Chief. **Diagnosis:** NR **Total countries represented, n:** NR **Mean Age, Years (SD):** NR **Female, n (%):** NR	The outcome domain "pain" was the primary focus.	Two hierarchies were supported by panel. Juni et al. (2006) was preferential given support of CMSG., though Juhl et al. (2012) has advantage of using validated WOMAC. **Juni et al. hierarchy (highest to lowest):** • Global pain score • Pain on walking • WOMAC OA index pain subscore • Composite pain scores other than WOMAC • Pain on activities other than walking (e.g., stair climbing) • WOMAC global score • Lequesne OA index global score • Other algofunctional composite scores • PGA • CGA **Juhl et al. hierarchy (highest to lowest):** • WOMAC pain subscale • Pain during activity (VAS) • Pain during walking (VAS) • Global knee pain (VAS) • Pain at rest (VAS) • SF-36 bodily pain subscale • HAQ pain subscale • Lequesne algofunctional index pain subscale • AIMS pain subscale • KSPS • McGill Pain Questionnaire (intensity) • ASES pain subscale • Pain at night (VAS) • Pain during activity (NRS) • Pain on walking (NRS)

			• Number of painful days Note: Panelists noted that any proposed instrument hierarchy needs to be supported by documented psychometric properties rather than simple consensus. This eliminates potential bias due to multiplicity, reporting, or selection bias.
Reference: McAlindon et al. 2015 **Source:** PubMed/EMBASE **Objective:** To update the 1996 OARSI recommendations for the design, conduct, and reporting of clinical trials targeting symptom or structure modification among patients with knee OA **Consensus Design:** A stakeholder panel underwent an iterative process consisting of a series of online discussions to formulate recommendations, followed by individual scoring of the appropriateness of recommendations on a scale from 1-9. A score of 1-3 meant "inappropriate", 4-6 "uncertain", and 7-9 "appropriate". Median score was calculated for each recommendation. **Quality:** Specified research setting, health condition, population, and intervention; did not involve a diverse stakeholder group that included patients	**Total Number of Stakeholders:** 9 Panel consisted of industry and academia representatives. **Diagnosis:** NA **Total countries represented, n:** NR **Mean Age, Years (SD):** NR **Female, n (%):** NR	**Recommended outcome domains** (in support of OMERACT COS)**:** • Pain • Physical function • PGA • HRQoL **Recommended structural outcomes:** • Reduction or reversal of joint space narrowing (median score: 7) • Reducing progression of cartilage damage or reversal of cartilage damage (median score: 7) Note: Panelists noted that PRO accuracy can be influenced by staff/patient expectations, cost/perceived invasiveness of treatment, pain reporting training, pain inflation, and direct/indirect communication. They also recommended disease modification be defined as improvement in OA symptoms and one of the recommended structural outcomes (median score: 7).	**Identified instruments:** **PROs:** • WOMAC (pain, physical function) • KOOS (pain, physical function, QoL) • ICOAP (pain) • SF-36 or SF-12 (QoL) • PROMIS (physical function) • HAQ (pain, physical function) • Improved HAQ (pain, physical function, PGA) • EQ-5D (QoL) • SIP (QoL) • NRS (pain, physical function, PGA) • VAS (pain, physical function, PGA) **Objective measures for physical function** (in support of OARSI recommendations)**:** • 30 s chair stand test • 40 m fast-paced walk test • Stair climb test • TUG test • 6MWT **Measures for structural outcomes:** • Radiography or MRI Note: Panelists recommended that objective PRO measures that are valid, reliable, and responsive to change be used (median score: 9). They also noted that multi-

and providers; stakeholder group was relatively small; scoring process clearly defined; consensus definition unclear **Funding Source:** Printing partially covered by AbbVie, BioClinica, Boston Imaging Core lab, and Flexion			dimensional pain assessment tools (e.g., ICOAP) may more comprehensively evaluate pain status and characterize patients responsive to interventions, versus unidimensional tools (e.g., VAS, NRS).
Reference: Singh et al. 2015 **Source:** COMET Initiative **Objective:** To develop a plan for harmonizing outcomes for endstage knee/hip OA patients undergoing TJR and propose a preliminary core domain set regarding TJR outcomes research for international consensus **Consensus Design:** TJR Working Group remotely reviewed and discussed potential outcome domains and areas for the core domain set within the context of OMERACT Filter 2.0, after which a preliminary set was endorsed during an in-person meeting using a Delphi process. **Quality:** Specified health condition, population, and intervention; did not clearly define intended setting; involved a diverse group of stakeholders that included patients and providers, though size of group is unknown; scoring process	**Total Number of Stakeholders:** NR Stakeholder group consisted of epidemiologists, psychometricians, orthopedic surgeons, rheumatologists, patients (n=2), researchers **Diagnosis, n (%):** NR **Total countries represented, n:** NR **Mean Age, Years (SD):** NR **Female, n (%):** NR	**Core outcome domains:** • Joint pain (pre- and post-TJR) • Function • Patient satisfaction • Revision • AEs • Death **Non-core domains:** • Cost • Participation	SF-36 was briefly discussed for measuring QoL. However, some participants were interested in a joint-specific QoL assessment.

and consensus definition not clearly defined **Funding Source:** NR			
Other Publication Types (e.g., protocols, guidelines, workshop reports)			
Study Details	**Intervention**	**Outcomes Identified**	**Instruments/Items Identified**
Reference: OARSI 2022 **Source:** OARSI **Objective:** To collect the most important indexes used by physicians and researchers to evaluate the algofunctional status of patients with chronic musculoskeletal diseases **Quality rating:** NA **Funding Source:** NA **Databases:** NA **Included study types:** NA **Inclusion Criteria:** NA	**Intervention:** NA **Comparator:** NA **Diagnosis:** Pain from chronic musculoskeletal diseases (e.g., OA) **Follow-up period:** NA	The outcome domain "pain" was the primary focus.	**Specific to knee and/or hip OA:** • ICOAP • WOMAC • Lequesne

Exclusion Criteria: NA			
Reference: U.S. Department of Health and Human Services 2018 **Source:** FDA Draft Guidance for Industry **Objective:** To assist sponsors who are developing drugs, devices, or biological products to treat the underlying pathophysiology and structural progression of OA **Quality rating:** NA **Funding Source:** NA **Databases:** NA **Included study types:** NA **Inclusion Criteria:** NA **Exclusion Criteria:** NA	**Intervention:** Drugs, devices, and biological products **Comparator:** NR **Diagnosis:** OA **Follow-up period:** NA	**Outcomes identified as clinically meaningful to patients:** • Pain (PRO) • Function (PRO) • Prolonged time to end-stage disease <u>Note</u>: No structural endpoints validated to date.	None identified
Reference: Arthritis Foundation 2017 **Source:** Voice of the Patient Meeting	**Intervention:** Medical products for OA **Comparator:** NR	**Identified symptoms that are significant:** • Pain/tenderness • Stiffness	**Pain:** • 10 cm VAS • Likert **Function:**

Objective: To provide key drug development stakeholders the opportunity to hear directly from patients, caregivers, and patient advocates about their disease symptoms and daily impacts of OA that matter most to patients, as well as their perspectives on current approaches to treating symptoms **Quality rating:** NA **Funding Source:** No FDA funding was provided for this meeting. However, FDA presented on their PFDD Initiative as part of the meeting and were present to hear the discussions. This meeting was meant to be a parallel effort to the FDA's PFDD Initiative. **Databases:** NA **Included study types:** NA **Inclusion Criteria:** NA **Exclusion Criteria:** NA	**Diagnosis:** OA **Follow-up period:** NA	• Functional limitations, impaired mobility, and walking/standing limitations • Fatigue • Disfigurement • Bone loss/reduction in bone density • Flexibility • Sleep disturbance • Joint swelling • Numbness **Identified outcomes that are most meaningful:** • Prevention or delay of symptom worsening and disease progression • Reduction of need for medical procedures related to consequences of OA	• WOMAC • Lequesne
Reference: Jolles et al. 2013 **Source:** CMSG	**Intervention:** UKA	**Major outcomes:** • Survival rate of implant – femoral and tibial loosening (aseptic loosening)	**Pain:** • VAS • WOMAC pain subscore

Objective: SLR protocol to assess the benefits and harms of UKA in the treatment of knee OA in adults in terms of decreasing pain, increasing knee function, and postponing the need for TKA **Quality rating:** NA **Funding Source:** NR **Databases:** CENTRAL, MEDLINE, CINAHL, EMBASE, Web of Science, Current Controlled Trials **Included study types:** RCTs, CCTs, CBAs and ITS will be included **Inclusion Criteria:** Adults with unicompartmental knee OA of at least grade 2 (Ahlback radiologic criteria) or grade 4 (Kellegren and Lawrence grading system), published between 1980-2010 **Exclusion Criteria:** NR	**Comparator:** Usual care or any other surgical techniques currently available for treating unicompartmental knee OA (e.g., tibial osteotomy, TKA, mosaicplasty) **Diagnosis:** Unicompartmental knee OA **Follow-up period:** 6 months, 1 year, 5 years, 10 years, 15 years or longer	• Pain • Function • HRQoL • SAEs • Mortality • Failure of treatment rate – time to revision (any complication that needed surgical intervention) **Minor outcomes:** • PGA • Range of motion • Length of hospital stay Note: Authors did not provide reasons for choice of outcomes and instruments.	**Function:** • KSS • WOMAC function subscore **HRQoL:** • SF-36 • EQ-5D
Reference: Singh et al. 2013 **Source:** CMSG	**Intervention:** TKA	**Major outcomes:** • Pain • Function	**Pain:** • VAS • NRS

Objective: SLR protocol to assess the benefits and harms of TKA compared to conservative treatment in patients with knee OA and other non-traumatic diseases with refractory symptoms **Quality rating:** NA **Funding Source:** NR **Databases:** CENTRAL, MEDLINE, CINAHL, EMBASE, ISI Web of Knowledge, WHO CTRP **Included study types:** RCTs, CCTs **Inclusion Criteria:** Adults (16+ years) with knee OA or other non-traumatic knee diseases who are candidates for TKA, published from database inception or 1966 (MEDLINE), 1980 (EMBASE), and 1982 (CINAHL) to present **Exclusion Criteria:** Abstracts	**Comparator:** Conservative treatment (e.g., continued medical therapy, PT, acupuncture, etc.) **Diagnosis:** Knee OA or other non-traumatic knee disease who are candidates for TKA **Follow-up period:** Longest follow-up time after TKA	• HRQoL • Revision rate • Treatment failure (# patients who underwent non-routine secondary surgery for any reason) • SAEs • Death **Minor outcomes:** • Cardiac AEs • Pulmonary AEs • Other complications • Cost • Patient satisfaction • Non-serious AEs related to interventions • Inpatient/outpatient healthcare utilization • Withdrawals (overall and due to AEs) Note: Authors did not provide reasons for choice of outcomes and instruments.	**Function:** • WOMAC • KOOS • OKS **HRQoL:** • SF-36 • SF-12 **Patient satisfaction:** • VAS

Abbreviations: AAOS: American Academy of Orthopedic Surgeons; AAOS-Outcome SIG: AAOS Outcome Special Interest Group; ADLS: Activities of Daily Living Scale; AHRQ: Agency for Healthcare Research and Quality; AIMS: Arthritis Impact Measurement Scales; AKPS: Anterior Knee Pain Scale; ALF: Aggregated Locomotor Function; AQoL: Assessment of QoL; ASES: Arthritis Self-Efficacy Scale; BOA: British Orthopaedic Association; CCT: controlled clinical trial; CER: comparative effectiveness review; CGA: Clinician Global Assessment; CMSG: Cochrane Musculoskeletal Group; COMET: Core Outcome Measures in Effectiveness Trials; COS: core outcome set; EPC: evidence-based practice center; EQ-5D: EuroQol in 5 dimensions; FDA: Food and Drug Administration; GP: general practitioner; HAP: Human Activity Profile; HAQ: Health Assessment Questionnaire; HOOS-PS: Hip Disability and Osteoarthritis Outcome Score short version; HRQoL: health-related QoL; HSS: Hospital for Special

Surgery; ICHOM: International Consortium for Health Outcomes Measurement; ICOAP: Intermittent and Constant Osteoarthritis Pain; ICTRP: International Clinical Trials Registry Platform; IKDC: International Knee Documentation Committee; IOS: International Orthopedic Societies and Surgeons; IOSK: Indices of severity for OA of the knee; IQR: interquartile range; J-MAP: Joint specific Multidimensional Assessment of Pain; JOA: Japanese Orthopedic Association; KOOS: Knee Injury and Osteoarthritis Outcome Score; KOOS-PS: KOOS short version; KSPS: Knee-Specific Pain Scale; KSS: Knee Society Score; LEFS: Lower Extremity Functional Scale; MACTAR: McMaster Toronto Arthritis score; MTA: meta-analysis; NA: not applicable; NHP: Nottingham Health Profile; NIHR: National Institute for Health Research; NR: not reported; NRCS: non-randomized controlled study; NRS: numerical rating scale; OA: osteoarthritis; OARSI: Osteoarthritis Research Society International; OKS: Oxford Knee Score; OMERACT: Outcome Measures in Rheumatoid Arthritis Clinical Trials; ORS: Outcomes Research Group of the Orthopedic Research Society; OT: occupational therapist; PFDD: patient-focused drug development; PGA: Patient Global Assessment; PRO: patient-reported outcome; PT: physiotherapist; QoL: quality of life; RA: rheumatoid arthritis; RCT: randomized controlled trial; R3: Round 3; SAE: serious adverse event; SF-36: Short Form 36; SD: standard deviation; SIP: Sickness Impact Profile; SLR: systematic literature review; SoC: standard of care; THA: total hip arthroplasty; TJR: total joint replacement; TKA: total knee arthroplasty; TKR: total knee replacement; TUG: Timed get Up and Go; UKA: unicompartmental knee arthroplasty; VAS: visual analogue scale; VR-12: Veterans Short Form 12 health survey; WHO: World Health Organization; WOMAC: Western Ontario and McMaster University Osteoarthritis Index; 6MWT: 6-minute walk test

IX. Appendix C. Table of Synthesized Outcomes and Instruments

Outcomes	Instruments (n citations)	MCID/MID for knee/hip OA	Instrument properties (e.g., n items, time, reporter)	Validity/Reliability for knee/hip OA	Other Notes
Core Area: Life Impact					
Prioritized Outcomes (i.e., designated as "mandatory", "critical", "core", or "major" and/or cited by the majority of included publications)					
Domain: Joint pain (pre- and post-procedure) **Subdomains:** NA **Items:** • Overall/global assessment of pain • Pain at rest • Pain during the night • Pain during the day • Pain during walking • Pain on activities other than walking • # of painful days	VAS (7)	**MDC, knee OA** (Alghadir A et al., 2018): 0.08 (SE 0.03) **MCID** (van der Wees P et al., 2017): 20 mm	PRO; 1-item tool that takes <1 minute to complete; 10-cm straight line with "no pain" and "worst possible pain" at either end; recall period is usually the present or last 24 hours	Test-retest reliability and validity for measuring knee OA pain confirmed in 121 knee OA patients in Saudi Arabia (intervention unspecified); (Alghadir A et al., 2018) found most reliable for measuring knee OA pain compared to NRS and VRS	Unidimensional assessment of pain intensity; can be either vertical or horizontal line; publicly available
	WOMAC pain subscale (6)	**MCID, osteotomy for knee OA** (Kim M et al., 2021): 4.2 points **MCID range** (Maredupaka S et al., 2020)**, TKA:** 10.5-36.0	PRO; 5 items assessing pain elicited during ADL	Validated in knee OA patients undergoing osteotomy;(Kim M et al., 2021) demonstrates adequate "truth" and "discrimination" per OMERACT Filter 2.0 (Howe T et al., 2012)	Multidimensional assessment of pain; developed for knee and hip OA for use in clinical trials; not publicly available; response also can be defined as ≥50% improvement in subscale score and absolute change of ≥20 points on scale of 0-100 (Pham T et al., 2003)
	Lequesne global score (4)	MCID NR; score >11-12 points after treatment indicates surgery (Lequesne M, 1997)	PRO; 10 items covering 3 dimensions (pain/discomfort (5 items), walking distance (1 item), ADL (4 items)); 3-4 minutes to complete	Validation done in knee/hip OA patients (Lequesne M, 1997) and in 88 solely symptomatic knee OA patients in France; (Faucher M et al., 2002) test-retest reliability was "good"; construct validity	NA

				could not be confirmed when compared to WOMAC	
	WOMAC global score (4)	**MCID, osteotomy for knee OA** (Kim M et al., 2021)**:** 16.1 **MCID, TKR for OA** (Escobar A et al., 2007): 15	PRO; 24 items covering 3 dimensions (pain (5 items), stiffness (2 items), physical function (17 items)); ~5-10 minutes to complete; recall period is past 48 hours	Validation studies support validity, reliability, and responsiveness in knee/hip OA patients; (Bellamy N et al., 1988; Sun Y et al., 1997) listed MCID derived from validation involving knee OA patients undergoing osteotomy	Developed for knee and hip OA for use in clinical trials; not publicly available
	NRS (3)	**MDC** (Alghadir A et al., 2018): 1.33 (SE 0.48)	PRO; 1-item tool that takes <1 minute to complete; 11-point scale wherein 0 indicates "no pain" and 10 indicates "worse imaginable pain"	Test-retest reliability and validity for measuring knee OA pain confirmed in 121 knee OA patients in Saudi Arabia as compared to VAS and VRS (intervention unspecified); (Alghadir A et al., 2018) relative reliability as measured by ICC was "excellent"	Unidimensional assessment of pain; preferred over VAS by elderly population (Jensen M & P Karoly, 2011)
	ICOAP (3)	**MCID, pain** (Singh J et al., 2014)**:** 18.5	PRO; 11 items assessing 2 domains (constant pain (5 items), intermittent pain (6 items)); (Moreton B et al., 2012) recall period is previous week; <10 minutes to complete	Reliability study done in 81 patients with knee OA in US; (Singh J et al., 2014) reliability/reproducibility found to be moderate and varied with age, gender, and race/ethnicity	Developed specifically to assess OA pain; publicly available
	HAQ pain subscale (2)	NR	PRO; 1 item that asks patients to rate on a scale of 0-100 ("no pain" to "severe pain") how much pain they had in past week	NR	Unidimensional assessment of pain; publicly available

	KOOS/-PS (2)	**MCID, KOOS-PS** (Singh J et al., 2014)**:** 2.2 points **MCID, pain** (Goodman S et al., 2020)**:** 21 points	PRO; KOOS has 42 items assessing 5 domains (pain, other symptoms, ADL, function in sport/recreation, knee related QoL); recall is previous week	KOOS-PS reliability study done in 81 patients with knee OA in US; (Singh J et al., 2014) reliability/reproducibility found to be moderate and varied with age, gender, and race/ethnicity	Developed specifically to address knee and associated problems
	AIMS pain subscale (1)	NR	PRO; 5 items assessing pain; recall period is past month; ~20 minutes to complete	NR	Developed in patients with RA and OA; publicly available
	ASES pain subscale (1)	NR	PRO; 5 items assessing pain on a 10-point Likert scale; (Jonsson T et al., 2019) <5 minutes to complete	NR	Developed in patients with RA and OA
	KSPS (1)	NR; higher scores indicate more severe pain	PRO; 12 items assessing pain intensity, frequency, and distastefulness (Moseley J et al., 2002)	NR	Developed for a clinical trial assessing arthroscopic procedures in 180 knee OA patients
	KSS pain subscore (1)	NR	PRO; 2 items assessing pain on a 10-level scale ("none" to "severe") while walking on level ground and on stairs/inclines (The Knee Society, 2011)	NR	Developed to assess knee and functional abilities before and after TKA due to knee OA (Samuel A & D Kanimozhi, 2019)
	Lequesne pain subscale (1)	NR; lower score indicates less functional impairment	PRO; 5 items assessing pain/discomfort during nocturnal bedrest, waking, standing for 30 minutes, walking, and getting up from sitting without use of arms	NR	NA
	Likert scale (1)	NR	NA	Found to highly correlate and yield similar precision as VAS for discriminating treatments	Likert scales are typically used in in PROs that assess knee OA pain, including KOOS and WOMAC

				during 6-week study of rofecoxib vs. placebo in 219 knee OA patients; (Bolognese J et al., 2003) used 5-point Likert scale ranging from "no pain" to "extreme pain"	
	McGill Pain Questionnaire (1)	NR	PRO; 20 items assessing 4 dimensions of pain (sensory (10 items), affective (5 items), evaluative (1 item), miscellaneous(4 items)); patients select words associated with each dimension (Melzack R, 1975)	Reliability study done for short form version of questionnaire in 70 patients undergoing knee replacement for OA in UK; (Turner K et al., 2017) found to have fair test-retest reliability and good responsiveness with moderate to large effect sizes when compared to ICOAP	Affective subscale found to not have good measurement properties for knee OA patients
	OKS pain subscale (1)	**MDC, knee OA (non-surgical)** (Harris K et al., 2013)**:** 16 **MIC, knee OA (non-surgical)**: 17 **MID, knee OA (non-surgical):** 14	PRO; 5 items	Validation study done in 134 patients undergoing non-operative management for knee OA in England; (Harris K et al., 2013) demonstrated good test-retest reliability and structural validity	OKS pain component subscale confirmed as distinguishable from OKS total score and functional subscale score in 201 patients undergoing TKA in Denmark (Buus A et al., 2021)
	SF-36 bodily pain subscale (1)	**MCID, 6 months to 2 years post-TKR for knee OA** (Escobar A et al., 2007)**:** 17	PRO; 2 items that measure severity of pain and extent of interference with normal activities because of pain	NR	NA
Domain: Physical function	WOMAC global score (6)	**MCID, osteotomy** (Kim	PRO; 17 items that measure ability to perform certain activities such as sitting to	Validation studies support validity, reliability, and responsiveness in knee/hip OA	Developed for knee and hip OA for use in clinical trials; not publicly available

Subdomains: • Range of motion • Joint stability • Flexibility • Endurance • Balance • Gait • Strength • Disability • Mobility • Physical activity/exercise **Items:** • Walking stairs • Walking flat surfaces • Walking distance • Standing limitations • Ability to play sports • Overall/global assessment of disability • Disability other than walking		M et al., 2021)**:** 4.2 points **MCID range, TKA** (Maredupaka S et al., 2020)**:** 10.5-36.0	standing, walking, stairs, putting on socks; ~12 minutes to complete; recall period is 48 hours	patients (Bellamy N et al., 1988; Sun Y et al., 1997)	
	KOOS (5)	**MCID, KOOS-PS** (Singh J et al., 2014)**:** 2.2 points **MCID, function** (Goodman S et al., 2020)**:** 14 points	PRO; 42 items assessing 5 domains (pain, other symptoms, ADL, function in sport/recreation, knee related QoL); ~10 minutes to complete; recall period is previous week	KOOS-PS reliability study done in 81 patients with knee OA in US and found moderate reliability/reproducibility that varied with age, gender, and race/ethnicity; (Singh J et al., 2014) KOOS reliability study done in 4,461 OA patients who underwent primary unilateral knee replacement in HSS registry found strongest correlation at 2 years with SF-12 PCS and strong ceiling effect (Goodman S et al., 2020)	Developed specifically to address knee and associated problems; (Samuel A & D Kanimozhi, 2019) publicly available
	HSS knee score (3)	**MCID:** 8.29 (Singh J et al., 2013)	CRO involving patient interview and physical exam; 11 items covering 7 dimensions (pain, function, range of motion, muscle strength, flexion deformity, instability, subtractions) (Knee Scores in Total Knee Arthroplasty, 2002)	Validity assessment done in patients undergoing primary TKA from Mayo Clinic Total Joint Registry; (Singh J et al., 2013) found to be valid and sensitive	Score <60 considered poor, 60-69 fair, 70-85 good, 86-100 excellent (Wu L et al., 2022) Reference (Singh J et al., 2013) is a conference abstract
	Lequesne global score (3)	MCID NR; score >11-12 points after treatment indicates need for surgery	PRO; 10 items covering 3 dimensions (pain/discomfort (5 items), walking distance (1 item), ADL (4 items)); 3-4 minutes to complete	Validation done in knee/hip OA patients and in 88 solely symptomatic knee OA patients in France; (Faucher M et al., 2002) test-retest reliability was "good"; construct validity	NA

		(Lequesne M et al., 1997)		could not be confirmed when compared to WOMAC	
	WOMAC function subscale (3)	**MCID** (Kim M et al., 2021)**:** 10.1 points **MDC range, knee OA** (Collins N et al., 2011): 13.1-13.3	PRO; 2 items assessing amount of stiffness elicited after staying in certain positions and time of day it is experienced	Validated in knee OA patients undergoing osteotomy (Kim M et al., 2021)	Evidence of insufficient discrimination from pain subscale; (Faucher M et al., 2002) developed for knee and hip OA for use in clinical trials; response also can be defined as ≥50% improvement in subscale score and absolute change of ≥20 points on scale of 0-100 (Pham T et al., 2003)
	IKDC (2)	**MDC:** 9 points	PRO; 18 items assessing 3 domains (symptoms, sports and daily activities, current/prior knee function); (Collins N et al., 2011) recall period ranges from past 4 weeks to the present and prior to injury; ~10 minutes to complete	Validated in 533 patients in US with a variety of knee problems; (Irrgang J et al., 2001) evidence of adequate test-retest reliability and high construct validity when compared to SF-36	Developed to assess knee-specific symptoms, function, and sports activity in a variety of knee conditions; publicly available
	KSS total score (2)	**MCID** (Lizaur-Utrilla A et al., 2020): 7.2 points	Two sections: first is CRO assessing pain, range of motion, flexion deformities, contractures, alignment, and stability; second is PRO assessing patient mobility (walking distance, stairs) and potential walking aids before and after TKA	Validation study done in 345 patients undergoing primary TKA in Canada with 1 year follow-up; (Culliton S et al., 2018) cross-sectional and longitudinal validity supported; internal consistency acceptable for patient-reported section of KSS	Developed to assess knee and functional abilities before and after TKA due to knee OA (Samuel A & D Kanimozhi, 2019)
	NRS (2)	NR	PRO; 1 item	NR for knee OA	NA
	OKS (2)	**MID, clinical trials** (Beard D	PRO; 12 items that assess 2 domains (function (7 items) and	Systematic review of 23 studies assessing validity in patients undergoing knee arthroplasty found good	Developed to document patient pain and function after TKA; publicly available

		et al., 2015): 5 points **MID, cohort studies:** 9 points **MCID, post revision TKA** (Khow Y et al., 2021): 4.9	pain (5 items)); recall period is previous 4 weeks	evidence of reliability, internal consistency, and construct validity (Harris K et al., 2016)	
	VAS (2)	NR	PRO; 1-item tool that takes <1 minute to complete; recall period is usually the present or last 24 hours	NR – validated to assess pain in knee OA patients but not physical function	Unidimensional measure
	BOA knee score (1)	NR	Combines subjective and objective items assessing pain, walking ability, use of walking aids, gait, flexion deformity, maximal flexion, extension lag, valgus/varus angles on stressing, ability to rise out of chair, stair climbing, satisfaction with treatment, general disability	Reliability study conducted in 29 TKA patients in UK found to have greatest reproducibility compared to KSS and OKS (Liow R et al., 2003)	Higher scores indicate better functioning knee
	Bristol knee score (1)	NR	NR	Interobserver correlation assessed in 92 TKR patients in Austria; (Okuda M et al., 2012) found highest interobserver agreement compared to KSS and correlation with HSS	NA
	HAQ (1)	NR	PRO; 20 items assessing 5 dimensions (physical function/disability, pain, drug side effects, health care utilization, mortality; ~20-30	Measurement properties of HAQ Disability Index compared to WOMAC assessed in 271 patients with knee or hip OA in	Multidimensional assessment tool; developed for RA but broadly used in OA patients; publicly available

			minutes to complete; recall period is past week	US; (Bruce B et al., 2004) showed favorable properties	
	JOA knee score (1)	NR	PRO; assesses 4 dimensions (pain on walking, pain on ascending or descending stairs, range of motion, joint effusion)	Validity study done in 85 patients with primary knee OA in Japan; (Okuda M et al., 2012) found adequate construct validity, moderate to high inter/intra-observer reliability, and high internal consistency when correlated with TUG, SF-36, and JKOM	NA
	KSS functional subscore (1)	**MCID (95% CI), TKA for OA 2 years post-surgery** (Lee A et al., 2017)**:** 6.1 (5.1-7.1)	PRO; 20 items assessing 4 dimensions (walking and standing (5 items), standard activities of daily living (6 items), advanced activities (5 items), and discretionary activities (4 items))	Validated in knee/hip OA patients; demonstrates adequate "truth" and "discrimination" per OMERACT Filter 2.0 (Howe T et al., 2012)	NA
	LEFS (1)	**MDC (95% confidence)** Mehta S et al., 2016): Knee/hip OA: 12 Knee OA (0-12 month follow-up period): 18.1 Knee OA (0-6 month): 11.8 **MCID**: Knee OA (0-12 month follow-up): 12.5	PRO; 20 items assessing degree of difficulty of specific functional tasks via 0-4 Likert scale ranging from "extreme difficulty/unable to perform activity" to "no difficulty"	Validated in variety of lower extremity disorders including knee OA; (Mehta S et al., 2016) responsiveness found to be "excellent"; reliability and validity for assessing functional impairment supported when compared to WOMAC; convergent validity and responsiveness superior compared to WOMAC in patients after TKR	Specifically designed to measure function

		Knee OA (0-6 month): 8.5			
	Lysholm knee score (1)	NR	CRO or PRO; (McHugh M et al., 2020) 8 items assessing limp, support, locking, instability, pain, swelling, stair climbing, and squatting	Validated against WOMAC in 1657 patients with knee chondral disorders; (Kocher M et al., 2004) found acceptable test-retest reliability and internal consistency; strong criterion validity	Developed to evaluate knee ligament injuries after surgery; publicly available
	MACTAR (1)	NR	PRO; 42 items assessing 4 dimensions (general health, physical function, social function, emotional function); ~10-15 minutes to complete	Validation study done in 192 knee/hip OA patients in The Netherlands; (Barten D et al., 2012) found moderate construct validity and good responsiveness compared to WOMAC and SF-36	Publicly available
	PROMIS Physical Function (1)	**MID range, knee OA (non-surgical) for short form** (Lee A et al., 2017)**:** 1.9-2.2	PRO; five instruments within PROMIS that are relevant to knee OA with number of items ranging from 4 to 20 (White D & H Master, 2016)	Reliability study done in 204 symptomatic knee OA patients in US; (Driban J et al., 2015) found strong correlation with SF-36 and few floor or ceiling effects	Developed as general measure of health
	Stair climb test (1)	**MDC, 90% confidence** (Almeida G et al., 2010)**:** 1.14 seconds	Performance-based test; measures functional strength, balance, and agility through how long it takes to ascend/descend 8-14 steps (9 steps recommended)	Reliability study done in 43 patients with TKA for 11-step stair climb test in US; (Almeida G et al., 2010) found good inter-rater reliability when compared to other lower extremity performance-based tasks and PROs of physical function	Part of OARSI recommended core set of performance-based tests to assess physical function in knee/hip OA
	TUG test (1)	**MDC, 90% confidence** (Kennedy D et	Performance-based test; measures time taken to rise from	Reliability study done in 81 knee OA patients following TKA in Canada; (Kennedy D et	Part of OARSI recommended minimum core set of performance-based tests to

		al., 2005)**:** 2.49 seconds	a chair, walk 3m, turn, walk back to the chair, then sit down	al., 2005) found adequate test-retest reliability and responsiveness	assess physical function in knee/hip OA
	30m fast-paced walk test (1)	**MDC, 40m test, 90% confidence** (Kennedy D et al., 2005)**:** 4.04 seconds	Performance-based test; measures how long it takes to speed walk 3 x 10m (30m)	Reliability study done in 81 knee OA patients following TKA in Canada for 40m walk test; (Kennedy D et al., 2005) found adequate test-retest reliability and responsiveness	40m fast-paced walk test is part of OARSI recommended core set of performance-based tests to assess physical function in knee/hip OA
	30s chair stand test (1)	**MDC, 90% confidence** (Gill S & H McBurney, 2008)**:** 1.64 stands	Performance-based test; measures maximum number of chair stand repetitions in 30 seconds	Reliability study done in 82 patients awaiting joint replacement for knee/hip in Australia; (Gill S & H McBurney, 2008) found high reliability as correlated with 50 ft Timed Walk	Part of OARSI recommended minimum core set of performance-based tests to assess physical function in knee/hip OA
	6MWT (1)	**MDC, 90% confidence** (Kennedy D et al., 2005)**:** 61.34 meters	Performance-based test; measures maximum distance covered in 6 minutes	Reliability study done in 81 knee OA patients following TKA in Canada; (Kennedy D et al., 2005) found adequate test-retest reliability and responsiveness	Part of OARSI recommended core set of performance-based tests to assess physical function in knee/hip OA
Domain: HRQoL **Subdomains**: NA **Items**: NA	EQ-5D-5L (5)	**MCID** (Bilbao A et al., 2018)**:** 0.32 points **MID, TKR** (Conner-Spady B et al., 2018): 0.20 points	PRO; 6 items assessing 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression); recall period is present; <5 minutes to complete	Reliability study done in 758 patients with knee/hip OA in Spain found minimal floor and ceiling effects and strong correlation with WOMAC pain and physical function scores; (Bilbao A et al., 2018) validation study done in 537 hip/knee OA patients undergoing THR or TKR in Canada found acceptable responsiveness compared with	Developed originally in English and Spanish as a generic instrument for describing and valuing health; (Herdman M et al., 2011) publicly available

				SF-12 and OKS (Conner-Spady B et al., 2018)	
	SF-12 (4)	For TKA, pre- to 12 months post-surgery (Clement N et al., 2019): **MCID, PCS:** 1.8 **MCID, MCS:** 1.5 **MIC, PCS:** 2.7 **MIC, MCS**: -1.4 **MDC, 90% confidence, PCS**: 8.9 **MDC, 90% confidence, MCS:** 13.8	PRO; 12 items assessing 8 dimensions (physical activities, social activities, impact on usual role activities due to physical health issues, impact on usual role activities due to emotional issues, bodily pain, general mental health, vitality, general health perceptions)	Study comparing SF-12 to SF-36 in 407 knee OA patients undergoing knee replacement done in Australia; (Webster K & J Feller, 2016) found PCS and MCS scores were highly correlated between both versions and similar responsive to change	Shortened version of SF-36; publicly available
	SF-36 (4)	**MCID, TKR for knee OA** (Escobar A et al., 2007)**:** 10 points	PRO; 36 items assessing 8 dimensions (physical function (10 items), physical role (4 items), bodily pain (2 items), general health (5 items), vitality (4 items), social functioning (2 items), emotional role (3 items), mental health (5 items))	Overall score discouraged from being used to assess QoL; rather, using separate physical and mental component scores is advised; (Rannou F et al., 2007) separate scores were found to have acceptable convergent and divergent validity in study of 2,540 knee OA patients in France (Rannou F et al, 2007)	Developed as a generic health measure for use in clinical practice, research, and health policy evaluation; publicly available via RAND but has scoring differences from the licensed version from Optum (Laucis N et al., 2015)
	AQoL-4D (1)	NR	PRO, 12 items assessing 4 dimensions (independent living, social relationships, physical senses, psychological wellbeing);	Internal validation conducted using the Rasch measurement model in 196 knee/hip arthritis/OA patients in	Developed in Australia to measure HRQoL

			(Hawthorne G et al., 1997) recall period is over the previous week or 4 weeks, ~1-2 minutes to complete	Australia showing competence for assessing HRQoL for knee/hip joint disease; (Ackerman I et al., 2014) construct validity assessed in 336 knee or knee/other joint OA patients against WOMAC, SF-36, and VAS (Whitfield K et al., 2006)	
	KOOS (1)	**MCID** (Singh J et al., 2014)**:** 8.0	PRO; 42 items assessing 5 dimensions (pain, other symptoms, ADL, function in sport/recreation, knee related QoL)	Study to assess reliability of KOOS QoL subscale performed in 81 knee OA patients in US with 2-week follow-up; (Singh J et al., 2014) found to have minimal to moderate variation in reproducibility and high reliability	NA
	NHP (1)	NR	PRO; 45 items assessing 6 dimensions (sleep, physical mobility, energy, pain, emotional reactions, social isolation)	Correlation study done in Turkey with 140 knee OA patients found it to be sensitive and correlated with clinical status and functional ability as measured by WOMAC and VAS (Yildiz N et al., 2010)	Developed as generic tool to measure perceived health status and impact on daily life; might be at odds with more recent instruments for QoL since it assesses negative aspects of health rather than well-being; not publicly available
	SF-36 mental component score (1)	**MCID range for SF-12 MCS, TKA** (Maradupaka S et al., 2020): -1.4-5.4 points	PRO; 14 items assessing 4 dimensions (vitality, social functioning, role limitations, social well-being)	Validity study done in 2,540 knee OA patients in France found acceptable convergent and divergent validity (Rannou F et al., 2007)	Subscale of the SF-36
	SIP (1)	NR	PRO; 136 items assessing 12 dimensions (sleep/rest, eating, work, home management, recreation/pastimes, mobility, ambulation, body care and	NR for knee OA	Developed as a generic, behaviorally based health status tool to measure the extent to

			movement, social interaction, alertness behavior, emotional behavior, communication); ~20-30 minutes to complete		which health/illness affect daily life and functioning
	VR-12 (1)	NR	PRO; 12 items assessing 8 dimensions (general health, physical functioning, role limitations due to physical and emotional problems, bodily pain, energy-fatigue, social functioning, mental health)	Responsiveness compared to EQ-5D and PROMIS-10 in 50 patients undergoing knee arthroscopy in US; (Oak S et al., 2016) exhibited strong internal and external responsiveness and no statistical differences between other scales	Derived from RAND VR-36 health survey developed with Veterans Health Administration; publicly available
Domain: Patient satisfaction **Subdomains**: NA **Items**: • Satisfaction with treatment outcome/results • Satisfaction with procedure	VAS (2)	NR	PRO; 5-point and 10-point scales commonly used that range from "very or completely dissatisfied" to "very or completely satisfied"	NR for knee OA – correlation study done in 147 patients who underwent THA in The Netherlands found high correlation with pain VAS and Oxford Hip Score	Satisfaction defined as score of ≥7 on 10-point scale and ≥4 on 5-point scale (Klem N et al., 2020)
	Likert scale (1)	NR	PRO; 5-point scale commonly used ranging from "very or extremely dissatisfied" to "very or extremely satisfied"	NR for knee OA	NA
Domain: Function/ Functional ability **Subdomains**: NA **Items**: • Ability to function in society	Specific instruments were not suggested in the included literature. However, stakeholders proposed using subscales of multidimensional PRO instruments.	NA	NA	NA	NA

• Ability to function at work • Work productivity • Employability • Work disability					
Non-prioritized outcomes (i.e., designated as "important but optional", "non-core", "minor", or no designation by the majority of studies)					
Domain: Patient participation **Subdomains**: • Participation in work • Participation in social activities • Role function • Support dependency **Items**: • Work status • Time lost from work • Ability to do vocational activities	Specific instruments were not suggested in the included literature. However, stakeholders proposed using PRO instruments.	NA	NA	NA	NA
Domain: PGA of target joint **Subdomains**: NA **Items**: • Overall improvement of disease	HAQ (1)	**MCII absolute change, knee or hip OA** (Tubach F et al., 2005): -18.3 mm **MCII relative change, knee or hip OA:** -15.2 mm	PRO; has 1 item that pertains to PGA that asks, "Considering all the ways that your arthritis affects you, rate how you are doing on the following scale by placing a vertical mark on the line"; line is 10cm VAS	100mm VAS highly correlated with and similar in precision as 4-point Likert in 6-week OA study of rofecoxib vs. placebo (Bolognese J et al., 2003)	MCII affected by initial degree of severity in symptoms but not by age, disease duration, or sex, as demonstrated in prospective 4-week cohort study with 603 knee OA patients (Tubach F et al., 2005)

	Modified Cincinnati Rating System Questionnaire (1)	NR	PRO; 12 items assessing 4 dimensions (pain, swelling, function, activity level)	NR for knee OA	Developed to assess outcomes following Anterior Cruciate Ligament construction but applicable to a variety of knee conditions
	NRS (1)	NR	PRO; 1 item; commonly used standard question: "Considering all the ways your knee OA affects you, how have you been during the past (time frame)?" (McAlindon T et al., 2015)	No validated standard question or response format for knee OA (McAlindon T et al., 2015)	NA
	VAS (1)	**MCII absolute change, knee or hip OA** (Tubach F et al., 2005): -18.3 mm **MCII relative change, knee or hip OA:** -15.2 mm	PRO; 1-item tool that takes <1 minute to complete; typically a 10-cm straight line; sample prompt: "Considering all the ways your arthritis affects you, mark (X) on the scale for how well you are doing." (Gentelle-Bonnassies S et al., 2000)	100mm VAS highly correlated with and similar in precision as 4-point Likert in 6-week OA study of rofecoxib vs. placebo (Bolognese J et al., 2003)	MCII affected by initial degree of severity in symptoms but not by age, disease duration, or sex, as demonstrated in prospective 4-week cohort study with 603 knee OA patients (Tubach F et al., 2005)
	Wallgren-Tegner activity score (1)	NR	PRO; 1 item that assesses activity; takes ~3 minutes to complete	Validation study done in 100 patients undergoing TKA in US; (Naal F et al., 2009) found adequate test-retest reliability, no floor or ceiling effects, and evidence of construct validity	Developed for use in conjunction with the Lysholm scale in patients with Anterior Cruciate Ligament injury but applicable to a variety of knee conditions; publicly available
Domain: ADL **Subdomains**: • Household management • Personal hygiene	Knee Outcome Survey ADLS (1)	MCID: 10 scale points (SE ±5 scale points)	PRO, 14 items total covering 2 domains (ADL (6 items), function (8 items)), recall period is the present	Validated in a variety of knee conditions that includes OA; (Irrgang J et al., 1998) test-retest reliability over 24-hr period = 0.97; construct validity determined through correlations with Lysholm Knee Scale (r=0.78-0.86) and	Evaluates how knee symptoms affect ability to perform ADL (6 items) and functional tasks (8 items)

Items: • Ability to use public transportation				global rating of knee function (r=0.66-0.75) in a sample of 397 patients	
Domain: Psychosocial impact/Mental health **Subdomains**: • Self-efficacy • Social withdrawal • Social isolation **Items**: • Perceived ability to cope with OA • Posthospital disposition	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA
Domain: Sleep **Subdomains**: NA **Items**: • Falling asleep • Staying asleep	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA
Domain: Fatigue **Subdomains**: NA **Items:** NA	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA
Domain: Disease progression	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA

Subdomains: NA **Items:** • Prolonged time to end-stage disease					
Core Area: Pathophysiological Manifestations					
Prioritized Outcomes (i.e., designated as “mandatory”, “critical”, “core”, or “major” and/or cited by the majority of included publications)					
Domain: Joint structure **Subdomains**: • Changes in structure • Disfigurement **Items**: • Reduction of joint space narrowing • Reversal of joint space narrowing • Decreased progression of cartilage damage • Reversal of cartilage damage	Radiographic imaging (5)	NA	NA	NA	NA
	MRI (2)	NA	NA	NA	NA
	Ultrasound (1)	NA	NA	NA	NA
Domain: Stiffness **Subdomains:** NA **Items:** NA	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA
Non-prioritized outcomes (i.e., designated as “important but optional”, “non-core”, “minor”, or no designation by the majority of studies)					

Domain: Other symptoms **Subdomains:** • Bone loss/bone density • Joint swelling • Numbness **Items:** NA	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA
Domain: Alignment **Subdomains:** NA **Items:** • Malalignment • Femorotibial alignment • Radiolucent lines	Radiographic imaging	NA	NA	NA	NA
Core Area: Resource Use/Economical Impact					
Prioritized Outcomes (i.e., designated as “mandatory”, “critical”, “core”, or “major” and/or cited by the majority of included publications)					
Domain: Treatment failure **Subdomains:** • Revision surgery • Reoperation (any consecutive major or minor open surgery, excluding revision) **Items:**	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA

• Revision rate • Reoperation rate • Time to revision • Time to TKA • Incidence of TKA • # of patients who underwent consecutive non-routine surgery					
Non-prioritized outcomes (i.e., designated as “important but optional”, “non-core”, “minor”, or no designation by the majority of studies)					
Domain: Care utilization **Subdomains:** • Treatment progression • Hospital admission • Postoperative care **Items:** • # of treatments undergone in past year for OA • # of healthcare providers seen in past year for OA • Length of hospital stay • Length of postoperative rehabilitation • Procedures done postoperatively	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA
Domain: Costs **Subdomains:**	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA

• Cost of medication • Cost of hospitalization • Cost of healthcare consults **Items:** NA					
Domain: Surgical performance **Subdomains:** NA **Items:** • Surgery time • Time to healing • Postoperative correction achievement • Inferior limb length	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA
Core Area: Death					
Prioritized Outcomes (i.e., designated as “mandatory”, “critical”, “core”, or “major” and/or cited by the majority of included publications)					
Domain: Mortality **Subdomains:** • All-cause • Cause-specific **Items:** • Mortality at 30 days post-procedure	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA
Other Areas					
Prioritized Outcomes (i.e., designated as “mandatory”, “critical”, “core”, or “major” and/or cited by the majority of included publications)					

Domain: SAEs **Subdomains:** • Short-term SAEs • Long-term SAEs **Items:** NA	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA
Domain: AEs **Subdomains:** NA **Items:** • Injury related to procedure • Cardiac AEs • Pulmonary AEs • Neurovascular complications	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA
Non-prioritized outcomes (i.e., designated as “important but optional”, “non-core”, “minor”, or no designation by the majority of studies)					
Domain: Survival rate of implant **Subdomains:** NA **Items:** NA	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA
Domain: Study withdrawal **Subdomains:** NA	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA

Items: • Withdrawal due to AEs • Overall withdrawal					
Domain: Compliance **Subdomains:** NA **Items:** NA	Specific instruments were not suggested in the included literature.	NA	NA	NA	NA

Abbreviations: ADL: activities of daily living; ADLS: Activities of Daily Living Scale; AE: adverse event; AIMS: Arthritis Impact Measurement Scales; AQoL: Assessment of QoL; ASES: Arthritis Self-Efficacy Scale; BOA: British Orthopaedic Association; CRO: clinician reported outcome; EQ-5D: EuroQol in 5 dimensions; EQ-5D-5L: 5 level version of EQ-5D; HAQ: Health Assessment Questionnaire; HRQoL: health-related QoL; HSS: Hospital for Special Surgery; ICC: intraclass correlation coefficient; ICOAP: Intermittent and Constant Osteoarthritis Pain; IKDC: International Knee Documentation Committee; JKOM: Japanese knee OA measure; JOA: Japanese Orthopedic Association; KOOS: Knee Injury and Osteoarthritis Outcome Score; KOOS-PS: KOOS short version; KSPS: Knee-Specific Pain Scale; KSS: Knee Society Score; LEFS: Lower Extremity Functional Scale; MACTAR: McMaster Toronto Arthritis score; MCD: minimal clinical difference; MCID: minimal clinically important difference; MCII: minimal clinically important improvement; MCS: mental component scale; MDC: minimum detectable change; MIC: minimal important change; MID: minimal important difference; NA: not applicable; NHP: Nottingham Health Profile; NR: not reported; NRS: numerical rating scale; OA: osteoarthritis; OARSI: Osteoarthritis Research Society International; OKS: Oxford Knee Score; OMERACT: Outcome Measures in Rheumatoid Arthritis Clinical Trials; PCS: physical component scale; PGA: Patient Global Assessment; PRO: patient-reported outcome; PROMIS: Patient-Reported Outcomes Measurement Information System; QoL: quality of life; RA: rheumatoid arthritis; RAND: Research and Development Corporation; RCT: randomized controlled trial; R3: Round 3; SAE: serious AE; SF-12: 12-item Short Form; SF-36: 36-item Short Form; SD: standard deviation; SE: standard error; SIP: Sickness Impact Profile; THA: total hip arthroplasty; TKA: total knee arthroplasty; TKR: total knee replacement; TUG: Timed get Up and Go; UK: United Kingdom; US: United States; VAS: visual analogue scale; VRS: verbal rating scale; VR-12: Veterans Short Form 12 item health survey; VR-36: Veterans 36 item health survey; WOMAC: Western Ontario and McMaster University Osteoarthritis Index; 6MWT: 6-minute walk test

X. References

Doherty M, Abhishek A. Clinical manifestations and diagnosis of osteoarthritis UpToDate2021. Available from: https://www.uptodate.com/contents/clinical-manifestations-and-diagnosis-of-osteoarthritis/contributors.

March L, Cross M. Epidemiology and risk factors for osteoarthritis UpToDate2020. Available from: https://www.uptodate.com/contents/epidemiology-and-risk-factors-for-osteoarthritis.

Deveza L, Bennell K. Management of knee osteoarthritis UpToDate2022. Available from: https://www.uptodate.com/contents/management-of-knee-osteoarthritis

Deveza L, Bennell K. Management of moderate to severe knee osteoarthritis UpToDate2022. Available from: https://www.uptodate.com/contents/management-of-moderate-to-severe-knee-osteoarthritis

Mandl L, Martin GM. Overview of surgical therapy of knee and hip osteoarthritis UpToDate2020. Available from: https://www.uptodate.com/contents/overview-of-surgical-therapy-of-knee-and-hip-osteoarthritis

Batailler C, Swan J, Sappey Marinier E, Servien E, Lustig S. New Technologies in Knee Arthroplasty: Current Concepts. J Clin Med. 2020;10(1). Epub 20201225. doi: 10.3390/jcm10010047. PubMed PMID: 33375702; PMCID: PMC7795103.

Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Tugwell P. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012;13:132. Epub 20120806. doi: 10.1186/1745-6215-13-132. PubMed PMID: 22867278; PMCID: PMC3472231.

Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, Conaghan PG, Bingham CO, 3rd, Brooks P, Landewé R, March L, Simon LS, Singh JA, Strand V, Tugwell P. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol. 2014;67(7):745-53. Epub 20140228. doi: 10.1016/j.jclinepi.2013.11.013. PubMed PMID: 24582946.

Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, Williamson PR. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. PLoS Med. 2017;14(11):e1002447. Epub 20171116. doi: 10.1371/journal.pmed.1002447. PubMed PMID: 29145404; PMCID: PMC5689835.

USPSTF. Appendix VI. Criteria for Assessing Internal Validity of Individual Studies Available from: https://uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes/procedure-manual/procedure-manual-appendix-vi-criteria-assessing-internal-validity-individual-studies.

Palmer JS, Monk AP, Hopewell S, Bayliss LE, Jackson W, Beard DJ, Price AJ. Surgical interventions for symptomatic mild to moderate knee osteoarthritis. Cochrane Database Syst Rev. 2019;7(7):Cd012128. Epub 20190719. doi: 10.1002/14651858.CD012128.pub2. PubMed PMID: 31322289; PMCID: PMC6639936.

Duivenvoorden T, Brouwer RW, van Raaij TM, Verhagen AP, Verhaar JA, Bierma-Zeinstra SM. Braces and orthoses for treating osteoarthritis of the knee. Cochrane Database Syst Rev. 2015;2015(3):Cd004020. Epub 20150316. doi: 10.1002/14651858.CD004020.pub3. PubMed PMID: 25773267; PMCID: PMC7173742.

Brouwer RW, Huizinga MR, Duivenvoorden T, van Raaij TM, Verhagen AP, Bierma-Zeinstra SM, Verhaar JA. Osteotomy for treating knee osteoarthritis. Cochrane Database Syst Rev. 2014;2014(12):Cd004019. Epub 20141213. doi: 10.1002/14651858.CD004019.pub4. PubMed PMID: 25503775; PMCID: PMC7173694.

Hofstede SN, Nouta KA, Jacobs W, van Hooff ML, Wymenga AB, Pijls BG, Nelissen RG, Marang-van de Mheen PJ. Mobile bearing vs fixed bearing prostheses for posterior cruciate retaining total knee arthroplasty for postoperative functional status in patients with osteoarthritis and rheumatoid arthritis. Cochrane Database Syst Rev. 2015(2):Cd003130. Epub 20150204. doi: 10.1002/14651858.CD003130.pub3. PubMed PMID: 25650566.

Howe TE, Dawson LJ, Syme G, Duncan L, Reid J. Evaluation of outcome measures for use in clinical practice for adults with musculoskeletal conditions of the knee: A systematic review. Man Ther. 2012;17(2):100-18. doi: 10.1016/j.math.2011.07.002.

Konnyu KJ, Thoma LM, Bhuma MR, Cao W, Adam GP, Mehta S, Aaron RK, Racine-Avila J, Panagiotou OA, Pinto D, Balk EM. AHRQ Comparative Effectiveness Reviews. Prehabilitation and Rehabilitation for Major Joint Replacement. Rockville (MD): Agency for Healthcare Research and Quality (US); 2021.

Lange T, Rataj E, Kopkow C, Lützner J, Günther KP, Schmitt J. Outcome Assessment in Total Knee Arthroplasty: A Systematic Review and Critical Appraisal. J Arthroplasty. 2017;32(2):653-65.e1. doi: 10.1016/j.arth.2016.09.014.

Christensen R, Maxwell LJ, Jüni P, Tovey D, Williamson PR, Boers M, Goel N, Buchbinder R, March L, Terwee CB, Singh JA, Tugwell P. Consensus on the need for a hierarchical list of patient-reported pain outcomes for metaanalyses of knee osteoarthritis trials: An OMERACT objective. J Rheumatol. 2015;42(10):1971-5. doi: 10.3899/jrheum.141384.

McAlindon TE, Driban JB, Henrotin Y, Hunter DJ, Jiang GL, Skou ST, Wang S, Schnitzer T. OARSI Clinical Trials Recommendations: Design, conduct, and reporting of clinical trials for knee osteoarthritis. Osteoarthritis Cartilage. 2015;23(5):747-60. doi: 10.1016/j.joca.2015.03.005.

Smith TO, Hawker GA, Hunter DJ, March LM, Boers M, Shea BJ, Christensen R, Guillemin F, Terwee CB, Williamson PR, Dodd S, Roos EM, Loeser RF, Schnitzer TJ, Kloppenburg M, Neogi T, Ladel CH, Kalsi G, Kaiser U, Buttel TW, Ashford AE, Mobasheri A, Arden NK, Tennant A, Hochberg MC, de Wit M, Tugwell P, Conaghan PG. The OMERACT-OARSI Core Domain Set for Measurement in Clinical Trials of Hip and/or Knee Osteoarthritis. J Rheumatol. 2019;46(8):981-9. Epub 20190115. doi: 10.3899/jrheum.181194. PubMed PMID: 30647185.

Hoang A, Goodman SM, Navarro-Millán IY, Mandl LA, Figgie MP, Bostrom MP, Padgett DE, Sculco PK, McLawhorn AS, Singh JA. Patients and surgeons provide endorsement of core domains for total joint replacement clinical trials. Arthritis Res Ther. 2017;19(1):267. Epub 20171206. doi: 10.1186/s13075-017-1476-9. PubMed PMID: 29208013; PMCID: PMC5718077.

Singh JA, Dohm M, Choong PF. Consensus on draft OMERACT core domains for clinical trials of Total Joint Replacement outcome by orthopaedic surgeons: a report from the International consensus on outcome measures in TJR trials (I-COMiTT) group. BMC Musculoskelet Disord. 2017;18(1):45. Epub 20170126. doi: 10.1186/s12891-017-1409-4. PubMed PMID: 28125985; PMCID: PMC5270232.

Singh JA, Dowsey MM, Dohm M, Goodman SM, Leong AL, Scholte Voshaar M, Choong PF. Achieving Consensus on Total Joint Replacement Trial Outcome Reporting Using the OMERACT Filter: Endorsement of the Final Core Domain Set for Total Hip and Total Knee Replacement Trials for Endstage Arthritis. J Rheumatol. 2017;44(11):1723-6. Epub 20170115. doi: 10.3899/jrheum.161113. PubMed PMID: 28089984.

Singh JA, Dohm M, Sprowson AP, Wall PD, Richards BL, Gossec L, Hawker GA, Riddle DL, Buchbinder R. Outcome Domains and Measures in Total Joint Replacement Clinical Trials: Can We Harmonize Them? An OMERACT Collaborative Initiative. J Rheumatol. 2015;42(12):2496-502. Epub 20150401. doi: 10.3899/jrheum.141201. PubMed PMID: 25834208.

Rolfson O, Wissig S, van Maasakkers L, Stowell C, Ackerman I, Ayers D, Barber T, Benzakour T, Bozic K, Budhiparama N, Caillouette J, Conaghan PG, Dahlberg L, Dunn J, Grady-Benson J, Ibrahim SA, Lewis S, Malchau H, Manzary M, March L, Nassif N, Nelissen R, Smith N, Franklin PD. Defining an International Standard Set of Outcome Measures for Patients With Hip or Knee Osteoarthritis: Consensus of the International Consortium for Health Outcomes Measurement Hip and Knee Osteoarthritis Working Group. Arthritis Care Res (Hoboken). 2016;68(11):1631-9. doi: 10.1002/acr.22868. PubMed PMID: 26881821; PMCID: PMC5129496.

Jolles BM, Eudier AF, Lécureux E. Unicompartmental arthroplasty for knee osteoarthritis. Cochrane Database Syst Rev. 2013(6). doi: 10.1002/14651858.CD010563. PubMed PMID: CD010563.

Singh AK, Vinay K. Surgical treatment of displaced intra-articular calcaneal fractures: is bone grafting necessary? J Orthop Traumatol. 2013;14(4):299-305. Epub 20130514. doi: 10.1007/s10195-013-0246-y. PubMed PMID: 23670493; PMCID: PMC3828503.

FDA. Osteoarthritis: structural endpoints for the development of drugs, devices, and biological products for treatment 2018. Available from: https://www.fda.gov/media/71132/download.

Arthritis Foundation. The Voice of the Patient 2017. Available from: https://www.arthritis.org/getmedia/6f33fa0d-afed-4800-9238-056460c37ae2/OA-Voice-of-the-Patient-Report.pdf

OSARI. Outcome Measures 2022. Available from: https://oarsi.org/research/outcome-measures

Mizner RL, Petterson SC, Clements KE, Zeni JA, Jr., Irrgang JJ, Snyder-Mackler L. Measuring functional improvement after total knee arthroplasty requires both performance-based and patient-report assessments: a longitudinal analysis of outcomes. J Arthroplasty. 2011;26(5):728-37. Epub 20100920. doi: 10.1016/j.arth.2010.06.004. PubMed PMID: 20851566; PMCID: PMC3008304.

Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, Buchbinder R, Snyder-Mackler L, Henrotin Y, Thumboo J, Hansen P, Bennell KL. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage. 2013;21(8):1042-52. Epub 20130513. doi: 10.1016/j.joca.2013.05.002. PubMed PMID: 23680877.

Rannou F, Boutron I, Jardinaud-Lopez M, Meric G, Revel M, Fermanian J, Poiraudeau S. Should aggregate scores of the Medical Outcomes Study 36-item Short Form Health Survey be used to assess quality of life in knee and hip osteoarthritis? A national survey in primary care. Osteoarthritis Cartilage. 2007;15(9):1013-8. Epub 20070326. doi: 10.1016/j.joca.2007.02.012. PubMed PMID: 17387025.

Webster KE, Feller JA. Comparison of the short form-12 (SF-12) health status questionnaire with the SF-36 in patients with knee osteoarthritis who have replacement surgery. Knee Surg Sports Traumatol Arthrosc. 2016;24(8):2620-6. Epub 20160128. doi: 10.1007/s00167-015-3904-1. PubMed PMID: 26821809.

Klem N-R, Kent P, Smith A, Dowsey M, Fary R, Schütze R, O'Sullivan P, Choong P, Bunzli S. Satisfaction after total knee replacement for osteoarthritis is usually high, but what are we measuring? A systematic review. Osteoarthritis and Cartilage Open. 2020;2(1):100032. doi: https://doi.org/10.1016/j.ocarto.2020.100032.

Maredupaka S, Meshram P, Chatte M, Kim WH, Kim TK. Minimal clinically important difference of commonly used patient-reported outcome measures in total knee arthroplasty: review of terminologies, methods and proposed values. Knee Surg Relat Res. 2020;32(1):19. Epub 20200409. doi: 10.1186/s43019-020-00038-3. PubMed PMID: 32660584; PMCID: PMC7219224.

Singh JA, Schleck C, Harmsen WS, Lewallen D. Validation Of The Hospital For Special Surgery Knee Questionnaire: Convergent Validity, Responsiveness and Sensitivity To Change. 2013 ACR/ARHP Annual Meeting; October 25-30, 2013; San Diego, CA2013.

Alghadir AH, Anwer S, Iqbal A, Iqbal ZA. Test-retest reliability, validity, and minimum detectable change of visual analog, numerical rating, and verbal rating scales for measurement of osteoarthritic knee pain. J Pain Res. 2018;11:851-6. Epub 20180426. doi: 10.2147/jpr.S158847. PubMed PMID: 29731662; PMCID: PMC5927184.

van der Wees PJ, Wammes JJ, Akkermans RP, Koetsenruijter J, Westert GP, van Kampen A, Hannink G, de Waal-Malefijt M, Schreurs BW. Patient-reported health outcomes after total hip and knee surgery in a Dutch University

Hospital Setting: results of twenty years clinical registry. BMC Musculoskelet Disord. 2017;18(1):97. Epub 20170303. doi: 10.1186/s12891-017-1455-y. PubMed PMID: 28253923; PMCID: PMC5335788.

Kim MS, Koh IJ, Choi KY, Sung YG, Park DC, Lee HJ, In Y. The Minimal Clinically Important Difference (MCID) for the WOMAC and Factors Related to Achievement of the MCID After Medial Opening Wedge High Tibial Osteotomy for Knee Osteoarthritis. Am J Sports Med. 2021;49(9):2406-15. Epub 20210611. doi: 10.1177/03635465211016853. PubMed PMID: 34115533.

Pham T, Van Der Heijde D, Lassere M, Altman RD, Anderson JJ, Bellamy N, Hochberg M, Simon L, Strand V, Woodworth T, Dougados M. Outcome variables for osteoarthritis clinical trials: The OMERACT-OARSI set of responder criteria. J Rheumatol. 2003;30(7):1648-54. PubMed PMID: 12858473.

Lequesne MG. The algofunctional indices for hip and knee osteoarthritis. J Rheumatol. 1997;24(4):779-81. PubMed PMID: 9101517.

Faucher M, Poiraudeau S, Lefevre-Colau MM, Rannou F, Fermanian J, Revel M. Algo-functional assessment of knee osteoarthritis: comparison of the test-retest reliability and construct validity of the WOMAC and Lequesne indexes. Osteoarthritis Cartilage. 2002;10(8):602-10. doi: 10.1053/joca.2002.0533. PubMed PMID: 12479381.

Escobar A, Quintana JM, Bilbao A, Aróstegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. Osteoarthritis Cartilage. 2007;15(3):273-80. Epub 20061017. doi: 10.1016/j.joca.2006.09.001. PubMed PMID: 17052924.

Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988;15(12):1833-40. PubMed PMID: 3068365.

Sun Y, Stürmer T, Günther KP, Brenner H. Reliability and validity of clinical outcome measurements of osteoarthritis of the hip and knee--a review of the literature. Clin Rheumatol. 1997;16(2):185-98. doi: 10.1007/bf02247849. PubMed PMID: 9093802.

Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. Handbook of pain assessment, 3rd ed. New York, NY, US: The Guilford Press; 2011. p. 19-44.

Singh JA, Luo R, Landon GC, Suarez-Almazor M. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. J Rheumatol. 2014;41(3):509-15. Epub 20140115. doi: 10.3899/jrheum.130609. PubMed PMID: 24429183; PMCID: PMC3943561.

Moreton BJ, Wheeler M, Walsh DA, Lincoln NB. Rasch analysis of the intermittent and constant osteoarthritis pain (ICOAP) scale. Osteoarthritis Cartilage. 2012;20(10):1109-15. Epub 20120705. doi: 10.1016/j.joca.2012.06.011. PubMed PMID: 22771772; PMCID: PMC3526788.

Goodman SM, Mehta BY, Mandl LA, Szymonifka JD, Finik J, Figgie MP, Navarro-Millán IY, Bostrom MP, Parks ML, Padgett DE, McLawhorn AS, Antao VC, Yates AJ, Springer BD, Lyman SL, Singh JA. Validation of the Hip Disability and Osteoarthritis Outcome Score and Knee Injury and Osteoarthritis Outcome Score Pain and Function Subscales for Use in Total Hip Replacement and Total Knee Replacement Clinical Trials. J Arthroplasty. 2020;35(5):1200-7.e4. Epub 20191227. doi: 10.1016/j.arth.2019.12.038. PubMed PMID: 31952945; PMCID: PMC7193650.

Jönsson T, Eek F, Dell'Isola A, Dahlberg LE, Ekvall Hansson E. The Better Management of Patients with Osteoarthritis Program: Outcomes after evidence-based education and exercise delivered nationwide in Sweden. PLoS One. 2019;14(9):e0222657. Epub 20190919. doi: 10.1371/journal.pone.0222657. PubMed PMID: 31536554; PMCID: PMC6752869.

Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, Hollingsworth JC, Ashton CM, Wray NP. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med. 2002;347(2):81-8. doi: 10.1056/NEJMoa013259. PubMed PMID: 12110735.

The Knee Society. The 2011 knee society scoring system licensed user manual 2011. Available from: https://www.kneesociety.org/assets/2011KSS%20Support%20Materials.pdf.

Samuel AJ, Kanimozhi D. Outcome measures used in patient with knee osteoarthritis: With special importance on functional outcome measures. Int J Health Sci (Qassim). 2019;13(1):52-60. PubMed PMID: 30842718; PMCID: PMC6392485.

Bolognese JA, Schnitzer TJ, Ehrich EW. Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. Osteoarthritis Cartilage. 2003;11(7):499-507. doi: 10.1016/s1063-4584(03)00082-7. PubMed PMID: 12814613.

Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain. 1975;1(3):277-99. doi: 10.1016/0304-3959(75)90044-5. PubMed PMID: 1235985.

Turner KV, Moreton BM, Walsh DA, Lincoln NB. Reliability and responsiveness of measures of pain in people with osteoarthritis of the knee: a psychometric evaluation. Disabil Rehabil. 2017;39(8):822-9. Epub 20160330. doi: 10.3109/09638288.2016.1161840. PubMed PMID: 27027698; PMCID: PMC5327871.

Harris KK, Dawson J, Jones LD, Beard DJ, Price AJ. Extending the use of PROMs in the NHS--using the Oxford Knee Score in patients undergoing non-operative management for knee osteoarthritis: a validation study. BMJ Open. 2013;3(8):e003365. Epub 20130821. doi: 10.1136/bmjopen-2013-003365. PubMed PMID: 23965934; PMCID: PMC3753511.

Buus A, Laugesen B, El-Galaly A, Laursen M, Hejlesen OK. The potential of dividing the oxford knee score into subscales for predicting clinically meaningful improvements in pain and function of patients undergoing total knee arthroplasty. Int J Orthop Trauma Nurs. 2021;45:100919. Epub 20211222. doi: 10.1016/j.ijotn.2021.100919. PubMed PMID: 35313190.

Knee Scores in Total Knee Arthroplasty. In: Surgical Techniques in Total Knee Arthroplasty. G.R. S, A.J. T, editors. New York, NY: Springer; 2002.

Wu LP, Mayr HO, Zhang X, Huang YQ, Chen YZ, Li YM. Knee Scores of Patients with Non-Lateral Compartmental Knee Osteoarthritis Undergoing Mobile, Fixed-Bearing Unicompartmental Knee and Total Knee Arthroplasties: A Randomized Controlled Trial. Orthop Surg. 2022;14(1):73-87. Epub 20211206. doi: 10.1111/os.13111. PubMed PMID: 34870364; PMCID: PMC8755879.

Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). Arthritis Care Res (Hoboken). 2011;63 Suppl 11(0 11):S208-28. doi: 10.1002/acr.20632. PubMed PMID: 22588746; PMCID: PMC4336550.

Irrgang JJ, Anderson AF, Boland AL, Harner CD, Kurosaka M, Neyret P, Richmond JC, Shelborne KD. Development and validation of the international knee documentation committee subjective knee form. Am J Sports Med. 2001;29(5):600-13. doi: 10.1177/03635465010290051301. PubMed PMID: 11573919.

Lizaur-Utrilla A, Gonzalez-Parreño S, Martinez-Mendez D, Miralles-Muñoz FA, Lopez-Prats FA. Minimal clinically important differences and substantial clinical benefits for Knee Society Scores. Knee Surg Sports Traumatol Arthrosc. 2020;28(5):1473-8. Epub 20190520. doi: 10.1007/s00167-019-05543-x. PubMed PMID: 31111184.

Culliton SE, Bryant DM, MacDonald SJ, Hibbert KM, Chesworth BM. Validity and Internal Consistency of the New Knee Society Knee Scoring System. Clin Orthop Relat Res. 2018;476(1):77-84. doi: 10.1007/s11999.0000000000000014. PubMed PMID: 29529620.

Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, Price AJ. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. J Clin Epidemiol. 2015;68(1):73-9. Epub 20141031. doi: 10.1016/j.jclinepi.2014.08.009. PubMed PMID: 25441700; PMCID: PMC4270450.

Khow YZ, Liow MHL, Goh GS, Chen JY, Lo NN, Yeo SJ. The oxford knee score minimal clinically important difference for revision total knee arthroplasty. Knee. 2021;32:211-7. Epub 20210909. doi: 10.1016/j.knee.2021.08.020. PubMed PMID: 34509827.

Harris K, Dawson J, Gibbons E, Lim CR, Beard DJ, Fitzpatrick R, Price AJ. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. Patient Relat Outcome Meas. 2016;7:101-8. doi: 10.2147/PROM.S97774. PubMed PMID: 27524925.

Liow RY, Walker K, Wajid MA, Bedi G, Lennox CM. Functional rating for knee arthroplasty: comparison of three scoring systems. Orthopedics. 2003;26(2):143-9. doi: 10.3928/0147-7447-20030201-15. PubMed PMID: 12597217.

Bach CM, Nogler M, Steingruber IE, Ogon M, Wimmer C, Göbel G, Krismer M. Scoring systems in total knee arthroplasty. Clin Orthop Relat Res. 2002(399):184-96. doi: 10.1097/00003086-200206000-00022. PubMed PMID: 12011708.

Bruce B, Fries J. Longitudinal comparison of the Health Assessment Questionnaire (HAQ) and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Arthritis Rheum. 2004;51(5):730-7. doi: 10.1002/art.20695. PubMed PMID: 15478152.

Okuda M, Omokawa S, Okahashi K, Akahane M, Tanaka Y. Validity and reliability of the Japanese Orthopaedic Association score for osteoarthritic knees. J Orthop Sci. 2012;17(6):750-6. Epub 20120807. doi: 10.1007/s00776-012-0274-0. PubMed PMID: 22868702.

Lee AC, Driban JB, Price LL, Harvey WF, Rodday AM, Wang C. Responsiveness and Minimally Important Differences for 4 Patient-Reported Outcomes Measurement Information System Short Forms: Physical Function, Pain Interference, Depression, and Anxiety in Knee Osteoarthritis. J Pain. 2017;18(9):1096-110. Epub 20170510. doi: 10.1016/j.jpain.2017.05.001. PubMed PMID: 28501708; PMCID: PMC5581239.

Mehta SP, Fulton A, Quach C, Thistle M, Toledo C, Evans NA. Measurement Properties of the Lower Extremity Functional Scale: A Systematic Review. J Orthop Sports Phys Ther. 2016;46(3):200-16. Epub 20160126. doi: 10.2519/jospt.2016.6165. PubMed PMID: 26813750.

McHugh M, Droy E, Muscatelli S, Gagnier JJ. Measures of Adult Knee Function. Arthritis Care Res (Hoboken). 2020;72 Suppl 10:219-49. doi: 10.1002/acr.24235. PubMed PMID: 33091264.

Kocher MS, Steadman JR, Briggs KK, Sterett WI, Hawkins RJ. Reliability, validity, and responsiveness of the Lysholm knee scale for various chondral disorders of the knee. J Bone Joint Surg Am. 2004;86(6):1139-45. doi: 10.2106/00004623-200406000-00004. PubMed PMID: 15173285.

Barten DJ, Pisters MF, Takken T, Veenhof C. Validity and responsiveness of the Dutch McMaster Toronto Arthritis Patient Preference Questionnaire (MACTAR) in patients with osteoarthritis of the hip or knee. J Rheumatol. 2012;39(5):1064-73. Epub 20120401. doi: 10.3899/jrheum.110876. PubMed PMID: 22467928.

White DK, Master H. Patient-Reported Measures of Physical Function in Knee Osteoarthritis. Rheum Dis Clin North Am. 2016;42(2):239-52. Epub 2016/03/17. doi: 10.1016/j.rdc.2016.01.005. PubMed PMID: 27133487.

Driban JB, Morgan N, Price LL, Cook KF, Wang C. Patient-Reported Outcomes Measurement Information System (PROMIS) instruments among individuals with symptomatic knee osteoarthritis: a cross-sectional study of floor/ceiling effects and construct validity. BMC Musculoskelet Disord. 2015;16:253. Epub 20150914. doi: 10.1186/s12891-015-0715-y. PubMed PMID: 26369412; PMCID: PMC4570513.

Almeida GJ, Schroeder CA, Gil AB, Fitzgerald GK, Piva SR. Interrater reliability and validity of the stair ascend/descend test in subjects with total knee arthroplasty. Arch Phys Med Rehabil. 2010;91(6):932-8. doi: 10.1016/j.apmr.2010.02.003. PubMed PMID: 20510986; PMCID: PMC2892160.

Kennedy DM, Stratford PW, Wessel J, Gollish JD, Penney D. Assessing stability and change of four performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. BMC Musculoskelet Disord. 2005;6:3-. doi: 10.1186/1471-2474-6-3. PubMed PMID: 15679884.

Gill S, McBurney H. Reliability of performance-based measures in people awaiting joint replacement surgery of the hip or knee. Physiother Res Int. 2008;13(3):141-52. doi: 10.1002/pri.411. PubMed PMID: 18697226.

Bilbao A, García-Pérez L, Arenaza JC, García I, Ariza-Cardiel G, Trujillo-Martín E, Forjaz MJ, Martín-Fernández J. Psychometric properties of the EQ-5D-5L in patients with hip or knee osteoarthritis: reliability, validity and responsiveness. Qual Life Res. 2018;27(11):2897-908. Epub 20180705. doi: 10.1007/s11136-018-1929-x. PubMed PMID: 29978346.

Conner-Spady BL, Marshall DA, Bohm E, Dunbar MJ, Noseworthy TW. Comparing the validity and responsiveness of the EQ-5D-5L to the Oxford hip and knee scores and SF-12 in osteoarthritis patients 1 year following total joint replacement. Qual Life Res. 2018;27(5):1311-22. Epub 20180208. doi: 10.1007/s11136-018-1808-5. PubMed PMID: 29423757.

Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2011;20(10):1727-36. Epub 2011/04/09. doi: 10.1007/s11136-011-9903-x. PubMed PMID: 21479777.

Clement ND, Weir D, Holland J, Gerrand C, Deehan DJ. Meaningful changes in the Short Form 12 physical and mental summary scores after total knee arthroplasty. Knee. 2019;26(4):861-8. Epub 20190517. doi: 10.1016/j.knee.2019.04.018. PubMed PMID: 31109846.

Laucis NC, Hays RD, Bhattacharyya T. Scoring the SF-36 in Orthopaedics: A Brief Guide. The Journal of bone and joint surgery American volume. 2015;97(19):1628-34. doi: 10.2106/JBJS.O.00030. PubMed PMID: 26446970.

Hawthorne G, Richardson J, Osborne R, McNeil H. The Assessment of Quality of Life (AQoL) instrument construction, initial validation & utility scaling 1997. Available from: http://www.aqol.com.au/papers/workingpaper76.pdf

Ackerman IN, Busija L, Tacey MA, Bohensky MA, Ademi Z, Brand CA, Liew D. Performance of the assessment of quality of life measure in people with hip and knee joint disease and implications for research and clinical use. Arthritis Care Res (Hoboken). 2014;66(3):481-8. doi: 10.1002/acr.22129. PubMed PMID: 23983000.

Whitfield K, Buchbinder R, Segal L, Osborne RH. Parsimonious and efficient assessment of health-related quality of life in osteoarthritis research: validation of the Assessment of Quality of Life (AQoL) instrument. Health and quality of life outcomes. 2006;4:19-. doi: 10.1186/1477-7525-4-19. PubMed PMID: 16556304.

Yildiz N, Topuz O, Gungen GO, Deniz S, Alkan H, Ardic F. Health-related quality of life (Nottingham Health Profile) in knee osteoarthritis: correlation with clinical variables and self-reported disability. Rheumatol Int. 2010;30(12):1595-600. Epub 20091013. doi: 10.1007/s00296-009-1195-x. PubMed PMID: 19823830.

Oak SR, Strnad GJ, Bena J, Farrow LD, Parker RD, Jones MH, Spindler KP. Responsiveness Comparison of the EQ-5D, PROMIS Global Health, and VR-12 Questionnaires in Knee Arthroscopy. Orthopaedic journal of sports medicine. 2016;4(12):2325967116674714-. doi: 10.1177/2325967116674714. PubMed PMID: 28210645.

Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, Bombardier C, Felson D, Hochberg M, van der Heijde D, Dougados M. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis. 2005;64(1):29-33. Epub 20040618. doi: 10.1136/ard.2004.022905. PubMed PMID: 15208174; PMCID: PMC1755212.

Gentelle-Bonnassies S, Le Claire P, Mezieres M, Ayral X, Dougados M. Comparison of the responsiveness of symptomatic outcome measures in knee osteoarthritis. Arthritis Care Res. 2000;13(5):280-5.PMID: 14635296.

Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for patients undergoing total joint arthroplasty? Clin Orthop Relat Res. 2009;467(4):958-65. Epub 20080628. doi: 10.1007/s11999-008-0358-5. PubMed PMID: 18587624; PMCID: PMC2650053.

Irrgang JJ, Snyder-Mackler L, Wainner RS, Fu FH, Harner CD. Development of a patient-reported measure of function of the knee. J Bone Joint Surg Am. 1998;80(8):1132-45. doi: 10.2106/00004623-199808000-00006. PubMed PMID: 9730122.