GEMMA M. ORANGE, MSc[1] • DANA A. HINCE, PhD[2] • MERVYN J. TRAVERS, PhD[1] • TASHA R. STANTON, PhD[3,4]
MATTHEW JONES, PhD[5,6] • SAURAB SHARMA, PhD[5,6] • SUMIN KIM, PT[3] • BENEDICT M. WAND, PhD[1] • MYLES C. MURPHY, PhD[1,6]

Physical Function Following Total Knee Arthroplasty for Osteoarthritis: A Longitudinal Systematic Review With Meta-analysis

Total knee arthroplasty (TKA) is a cost- and clinically effective treatment for end-stage knee osteoarthritis (OA).[1,29,47] In Australia, over 54 000 TKAs are performed every year at an annual cost of AU$1.2 billion.[1,2,21] Quantitative methods used to evaluate TKA outcomes typically emphasize surgical success metrics such as joint survivorship[7] and prosthetic alignment,[53,58] whereas people with knee OA nominate decreased pain intensity and improved physical function as their primary indicators of treatment success.[33,49,53,63] The disconnect between how clinicians and patients perceive outcomes[13] can cause dissatisfaction, as many patients are experiencing ongoing functional difficulties postsurgery.[33,62]

OBJECTIVES: To explore the extent of functional improvement following primary total knee arthroplasty for knee osteoarthritis and to compare the trajectories of self-reported and performance-based measures of physical function.

DESIGN: Longitudinal systematic review with meta-analysis

METHODS: We searched 3 electronic databases from January 2005 to February 2023 for longitudinal cohort studies involving adults with knee osteoarthritis undergoing primary total knee arthroplasty. Estimates of self-reported and performance-based physical function were extracted presurgery and up to 5 years postsurgery. Risk of bias was assessed using a 6-item checklist. Self-reported function scores were converted to a 0-100 scale (higher scores indicate worse function). Mixed models provided pooled estimates after excluding low-quality studies.

RESULTS: Out of 230 relevant studies, 72 (n = 19 063) of high quality were included in meta-analyses. Self-reported function significantly improved from presurgery (55.6/100; 95% confidence interval [CI], 53.1 to 58.1) to 3-6 months postsurgery (21.1; 95% CI, 17.9 to 24.3; $P<.001$). A small decline in self-reported function occurred at 6-12 months (31.0; 95% CI, 25.8 to 36.2; $P<.001$), with no further change at 12-24 months (30.9; 95% CI, 23.2 to 38.6; $P = .919$). Performance-based measures exhibited variable trajectories, with most estimates indicating no clinically meaningful improvement following total knee arthroplasty.

CONCLUSION: Total knee arthroplasty resulted in clinically meaningful improvements in self-reported function at 3-6 months postoperatively. There was some deterioration in function after 6 months, and at no other time point did the estimate reach a clinically important change. There was limited evidence of clinically meaningful improvements in performance-based measures of physical function at any time point. *J Orthop Sports Phys Ther 2025;55(1):1-11. Epub 26 November 2024. doi:10.2519/jospt.2024.12570*

KEY WORDS: *arthritis, knee, orthopedics, physical therapy*

When looking into the expected functional improvement following a TKA, there is an absence of high-quality data. There are over 2300 randomized clinical trials on TKA.[51] However, randomized clinical trials are designed to compare the efficacy of different interventions[11]; they are not prognostic. When exploring the prognosis of an intervention, cohort studies represent level I evidence.[42] Thus, a pooled estimate of the prognosis from a systematic review of cohort studies would represent the highest level of evidence.

Longitudinal meta-analysis of clinical outcomes is increasingly common in musculoskeletal health research. As cohort studies represent the highest level of evidence for prognosis,[42] these reviews typically only include cohort studies. For example, the clinical course of low back

[1]School of Health Sciences, The University of Notre Dame Australia, Fremantle, Australia. [2]Institute for Health Research, The University of Notre Dame Australia, Fremantle, Australia. [3]IMPACT in Health, Allied Health & Human Performance Academic Unit, University of South Australia, Adelaide, Australia. [4]Persistent Pain Research Group, Hopwood Centre for Neurobiology, Lifelong Health Theme, South Australian Health and Medical Research Institute (SAHMRI), Adelaide, Australia. [5]School of Health Sciences, Faculty of Medicine & Health, University of New South Wales, Sydney, Australia. [6]Centre for Pain IMPACT, Neuroscience Research Australia (NeuRA), Sydney, Australia. ORCID: Jones, 0000-0002-5534-755X; Orange, 0000-0002-1149-3272. Gemma M. Orange is supported by an Australian Government Research Training Program scholarship. Myles C. Murphy is supported by the Western Australian Future Health Research & Innovation Fund via a Near-miss Award and an Innovation Fellowship. Saurab Sharma is supported by The John J. Bonica Trainee Fellowship from the International Association for the Study of Pain. Gemma M. Orange, Dana A. Hince, Matthew Jones, Saurab Sharma, Sumin Kim, Benedict M. Wand, and Myles C. Murphy do not have any competing interests. Tasha R. Stanton has received payment for lectures on pain and rehabilitation as well as for travel and accommodation to provide lectures and receives author royalties from *Orthopaedic Physical Therapy Practice* for a book on osteoarthritis and pain. Mervyn J. Travers has received payment for lectures on pain and rehabilitation as well as for travel and accommodation to provide lectures. Address correspondence to Gemma M. Orange, School of Health Sciences, The University of Notre Dame Australia, 33 Phillimore St, Fremantle, WA 6160, Australia. E-mail: gemma.orange@nd.edu.au

pain has been established using only cohort studies.[37,60] Although not ideal, when clinical data are more sparse, randomized clinical trials have been included in these analyses.[39]

Longitudinal meta-analyses in TKA, which have focused on pain and self-reported function, used only prospective cohort studies due to the large number of cohort studies for the condition.[47] However, despite function being reported by patients as an indicator of treatment success,[33,49,53,63] no longitudinal synthesis of the literature exists to guide patients and clinicians about self-reported and performance-based functional status following TKA. Better understanding of the expected clinical course of function following TKA and the uncertainties around these estimates could aid and inform shared decision making for patients contemplating TKA and help orientate their expectations of postoperative recovery and outcome.[20,47,54,65]

Thus, the aims of this systematic review with meta-analysis were to investigate the extent of functional improvement following primary TKA for knee OA and to evaluate whether self-reported and performance-based measures of physical function followed similar trajectories.

METHODS

The protocol was prospectively registered (DOI: 10.17605/OSF.IO/6UW3E), and minor deviations are recorded in **SUPPLEMENTAL APPENDIX A**. The review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses[43] and Meta-analysis Of Observational Studies in Epidemiology.[12]

Selection Criteria

Study design We included prospective, longitudinal cohort studies published as full-text articles in English by peer-reviewed journals. To reflect contemporary practice, only studies published since 2005 were considered. We excluded cross-sectional studies, retrospective clinical studies, case series with less than 10 participants, and controlled clinical trials (whether randomized or nonrandomized).

Inclusion of the data from surgical arms in controlled trials may provide information to support the aims of this review. However, they are not the ideal study design to inform prognosis. Cohort studies are better suited for describing prognosis because they involve a representative sample of incident cases.[4,37] Consequently, in the areas of low back pain and knee OA, the practice has evolved from including clinical trial data in prognostic analysis when prospective cohort studies are unavailable to excluding randomized clinical trial data as the literature base grows.[37,47,56,60,61]

To increase confidence in the estimates reported in this review, we only analyzed data from studies rated as more than low quality (ie, unclear- or high-quality trials).[4] Justifications for excluding low-quality studies are manifold and well supported by existing literature. Higgins et al[25,27] underscore the importance of minimizing the inclusion of studies with methodological weakness to ensure the internal validity of systematic reviews. Altman[4] and Jüni et al[31] emphasize that low-quality studies can introduce noise and reduce the precision of effect estimates, thereby compromising the reliability of review findings. Sterne and Egger[50] caution against the inclusion of low-quality studies in synthesis, as they can disproportionately influence effect estimates and introduce bias. Finally, Glasziou et al[22] argue that excluding studies of poor methodological quality incentivizes ethical research practices and contributes to the overall improvement of scientific evidence.

Population Studies providing estimates of function in adults undergoing TKA for a primary diagnosis of knee OA were included. Studies including only participants with a primary diagnosis other than knee OA (eg, rheumatoid arthritis) were excluded. We excluded studies that purposively sampled a particular clinical phenotype (eg, inclusion limited to people with hemophilia undergoing TKA).

Intervention We included studies providing data for primary, unilateral TKA. Studies performing only revision surgery, simultaneous bilateral replacement, or hemiarthroplasty were excluded. Where studies included data on primary, unilateral TKA and other data, for example, hemiarthroplasty, the data were treated as follows: (1) for studies in which data for primary TKA were provided separately, only these data were extracted and analyzed; (2) if separate data were not provided, study authors were contacted to obtain the separate data; (3) if this was unsuccessful, the full data set was extracted and used in analysis provided at least 85% of the cohort had undergone primary TKA for knee OA; and (4) studies that enrolled mixed populations but gave no data on sample proportions were excluded if separate data for people undergoing primary TKA were not able to be obtained from the authors.

Outcome measures We included outcomes assessing the Outcome Measures in Rheumatology Clinical Trials and Osteoarthritis Research Society International core domains of "patient impression of function" as well as "performance-based measures of physical function" and have detailed our selection criteria and extraction hierarchy in **SUPPLEMENTAL APPENDIX B**.

Search Strategy and Screening Process

Electronic searches of 3 electronic databases (MEDLINE [Ovid], EMBASE [Ovid], and CINAHL [EBSCO]) were conducted from January 2005 to February 28, 2023. A combination of medical subject headings and key words for (1) knee OA, (2) TKA, (3) physical function/functional capacity outcomes, and (4) study design were used for the search strategy. Search terms were adapted to individual databases as necessary. The complete search strategy used for MEDLINE can be found in **SUPPLEMENTAL APPENDIX A**. We also searched the reference list of included papers and involved experts to identify missing studies.

Results of searches were imported into Covidence (Veritas Health Innovation, Melbourne, Australia), and duplicates were removed. Two review authors independently assessed the titles and abstracts of all retrieved studies for full-text screening before independently assessing the full-text articles against eligibility criteria. For both processes, any inconsistencies were resolved by discussion between reviewer pairs. For disagreements, a third reviewer was consulted.

Data Extraction

Pairs of reviewer authors independently extracted data from the full text of high- and unclear-quality studies using a standardized and piloted data extraction form, with the complete list of extracted variables detailed in **TABLE 1**.

Study Quality Assessment

Quality of the included studies was rated using a tool adapted from the methodologic criteria suggested by Altman[4] and used in similar reviews.[5,32,37,44] The original tool had 2 items for sampling, 2 items for completeness of follow-up, and 1 item for quality of outcome reporting. After piloting the checklist on several studies investigating total hip arthroplasty, we added an additional item regarding reasons for loss to follow-up from the Quality In Prognosis Studies tool.[26,27] Each of the 6 items was rated as "Yes," "No," or "Unclear." Each study was appraised as high, unclear, or low quality based on the following: high quality = all criteria are "Yes," unclear quality = all criteria are either "Yes" or "Unclear," and low quality = "No" for 1 or more items.

Pairs of authors independently assessed the quality of each included study and resolved any inconsistencies by discussion. In the case of disagreement, a third reviewer ensured consensus. Risk-of-bias assessment was performed prior to data extraction, so quality ratings were not biased by the knowledge of results.

Data Synthesis

All analyses were conducted by a biostatistician (D.A.H.) using Stata version 18 (StataCorp, LLC, College Station, TX). A qualitative synthesis was undertaken where data could not be pooled (see **SUPPLEMENTAL APPENDIX K**). Self-reported measures of function outcomes were linearly rescaled to a common 100-point scale, where "0" indicated best possible function and "100" indicated worst possible function. Further details of our data management can be found in **SUPPLEMENTAL APPENDIX A**.

We defined assessment periods as follows: presurgery, very short term (≥3 months to ≤6 months), short term (>6 months to ≤12 months), medium term (>12 months to ≤24 months), long term (>24 months to ≤36 months), and very long term (>36 months to ≤60 months). We decided not to report on time points greater than 5 years, as estimates this far postsurgery are likely to be confounded by other health-related issues in the cohort of people typically undergoing TKA.

Meta-analysis Pooled means of the time course of self-reported function and performance-based measures of physical function were calculated using separate mixed-effects meta-regression models. Each study mean was weighted by the inverse of the sum of the sampling variance and restricted maximum likelihood estimate of between-study variance using the *meta meregress* function in Stata. A fixed effect for time grouping was included in all models with levels included corresponding to available outcome data. All models included a random study-specific intercept to account for the dependence of repeated observations within studies and a random slope to account for differences in change across time between studies. The random effects were allowed to correlate assuming an unstructured variance-covariance matrix. These models were used to estimate

TABLE 1 Data Extracted From Each Study Included In the Review

Study Characteristics	Variables
Study information	Primary author, country of origin, year of publication, study period
Diagnostic criteria for knee OA	Kellgren-Lawrence, American College of Rheumatology
Study population	Reason for surgery, proportion primary vs revision, proportion bilateral vs unilateral, presurgical rehabilitation (yes/no/unsure)
Surgical details	Surgical approach, prosthesis used, patellar resurfacing (yes/no), number of surgeons, experience level of surgeons
Baseline demographics	Sex, age, height, weight
Baseline clinical characteristics	Pain intensity, functional capacity, radiographic OA severity, comorbidities
Timing of preoperative assessment	≤1 month preoperatively
Postsurgical details	Postoperative analgesia, postsurgical orders, length of inpatient stay, discharge destination, details of postsurgical rehabilitation
Concomitant treatments	Concomitant treatments during the follow-up period (including additional surgery)
Postsurgical complications	Details of any postsurgical complications
Self-reported measures	Self-reported measures of functional capacity and corresponding scores at all follow-up time points
Performance-based measures	Performance-based measures of functional capacity and corresponding scores at all follow-up time points
Follow-up timing	Time of each follow-up

Abbreviation: OA, osteoarthritis.

the pooled (weighted) means and 95% confidence intervals (CIs) for each time point and to test the effect of time for each outcome measure (using restricted maximum likelihood estimation).

The change in each outcome over time (presurgery vs each follow-up time period) was presented as a percentage of presurgery pooled mean and 95% CI at each time point. These were calculated using the weighted pooled estimates from the relevant mixed-effects model described above. This resulted in a negative percentage, indicating improvement. For ease of interpretation, all values were multiplied by –1 so that an increase reflects an improvement in the pooled mean values.

Sensitivity analysis Sensitivity analyses were planned to investigate the impact of high versus unclear study quality, formalized pre- or postoperative rehabilitation, and OA diagnostic criteria used. This was completed if there were sufficient studies (≥10) in the subgroup to do so.

Evaluation of heterogeneity The presence of overall heterogeneity was assessed with the multilevel Cochrane QM residual homogeneity test statistic.[28] The estimated variance of the pooled effect (τ^2) at each time point was used to evaluate the magnitude of heterogeneity at each time point, acknowledging that these estimates borrow information across time points. Forest plots were constructed for visual assessment of the distribution of study observations about the pooled effect, for each time point (**SUPPLEMENTAL APPENDIX C**).

Certainty of the evidence The certainty of the estimate of prognosis at each time point, for which there were 10 or more studies, was summarized using the Grading of Recommendations Assessment, Development, and Evaluation approach, modified for prognosis studies.[30,48] In making the judgment regarding confidence in the estimates of outcome, we considered the following 5 domains: risk of bias, inconsistency, imprecision, indirectness, and publication bias. Description of the decision-making criteria for each domain can be found in **SUPPLEMENTAL APPENDIX D.**

RESULTS

Our searches identified 11001 records after duplicates were removed and 230 manuscripts met the inclusion criteria (**FIGURE 1**). We classified 152 (66%) manuscripts as low quality, and these were excluded from data syntheses. Seventy-eight high- and unclear-quality papers were eligible to contribute data. Key outcome data were unavailable for 6 studies, leaving 72 high- or unclear-quality records for analysis.

Study Quality Assessment

Individual study quality judgments across all items are provided in **SUPPLEMENTAL APPENDIX E.** Fifty-two (23%) studies were rated as unclear quality, and 26 (11%) studies were rated as high quality. Quality of reporting of prognosis (40% studies [n = 61]) and availability of 3-month follow-up data for at least 80% of participants (38% studies [n = 58]) were the biggest contributors to studies being judged as low quality.

Characteristics of High- and Unclear-Quality Studies

The complete baseline characteristics of participants for all high- and unclear-quality studies are presented in **SUPPLEMENTAL APPENDIX F.** The 72 high- and unclear-quality studies represented 66 unique cohorts that enrolled 19063 participants (63% female). The sample size of individual studies ranged from 13 (Chung and Min[15]) to 2393 (Calabro et al[14]). The overall mean (SD, n cohorts) age for the sample was 67.8 years (3.0, n = 66 cohorts), with an overall mean body mass index of 32.6 kg/m^2 (13.0, n = 51 cohorts). Fifty-four cohorts reported self-reported measures only, 5 cohorts reported performance-based measures only, and 7 cohorts reported both.

Clinical Course of Self-reported Function After TKA

Our longitudinal meta-analysis pooled data from 61 cohorts including 143 observations, equating to data from 18 846 participants (for a breakdown of included self-reported outcome measures, see **SUPPLEMENTAL APPENDIX G**).

Pooled weighted mean self-reported function scores for each time category differed across the 5-year follow-up period, $\chi^2(5) = 5360$, $P<.001$ (**FIGURE 2A** and **TABLE 2**). The average self-reported function score was 55.6/100 (95% CI, 53.1 to 58.1) presurgery and improved to 21.1/100 (95% CI, 17.9 to 24.3; moderate-certainty evidence) at very short-term follow-up ($P<.001$). There was a regression in scores from very short-term ($P<.001$) to short-term (31.0/100; 95% CI, 25.8 to 36.2; low-certainty evidence) follow-up, but scores were still better than that presurgery ($P<.001$). At medium-term follow-up, self-reported function was 30.9/100 (95% CI, 23.2 to 38.6; low-certainty evidence), with no evidence of change compared with the previous time point ($P = .919$). There was slight worsening from medium-term scores ($P<.001$) to long-term follow-up (40.0/100; 95% CI, 29.7 to 50.2; certainty not estimated) followed by improvement ($P<.001$) at very long-term follow-up (35.7/100; 95% CI, 22.7 to 48.8; certainty not estimated). These changes in function score across time are also presented as percentage change from presurgery, where an increase in percentage indicates an improvement in function (**TABLE 3** and **FIGURE 2B**).

Our approach to estimating acceptable benefit can be found in **SUPPLEMENTAL APPENDIX H.** Decision thresholds for acceptable benefit were set based, firstly, on the estimates of the smallest worthwhile effect (SWE) for undergoing TKA, as this approach mitigates many of the conceptual and statistical problems inherent in the calculation of minimal clinically important differences (MCIDs).[24] Based on a pooled baseline score of 55.6/100, the point estimate for acceptable benefit was set as 27.2/100.[24] At very short-term follow-up (3-6 months), the outcome estimate exceeded the SWE, and the limit of the CI also exceeded the threshold for acceptable clinical benefit.

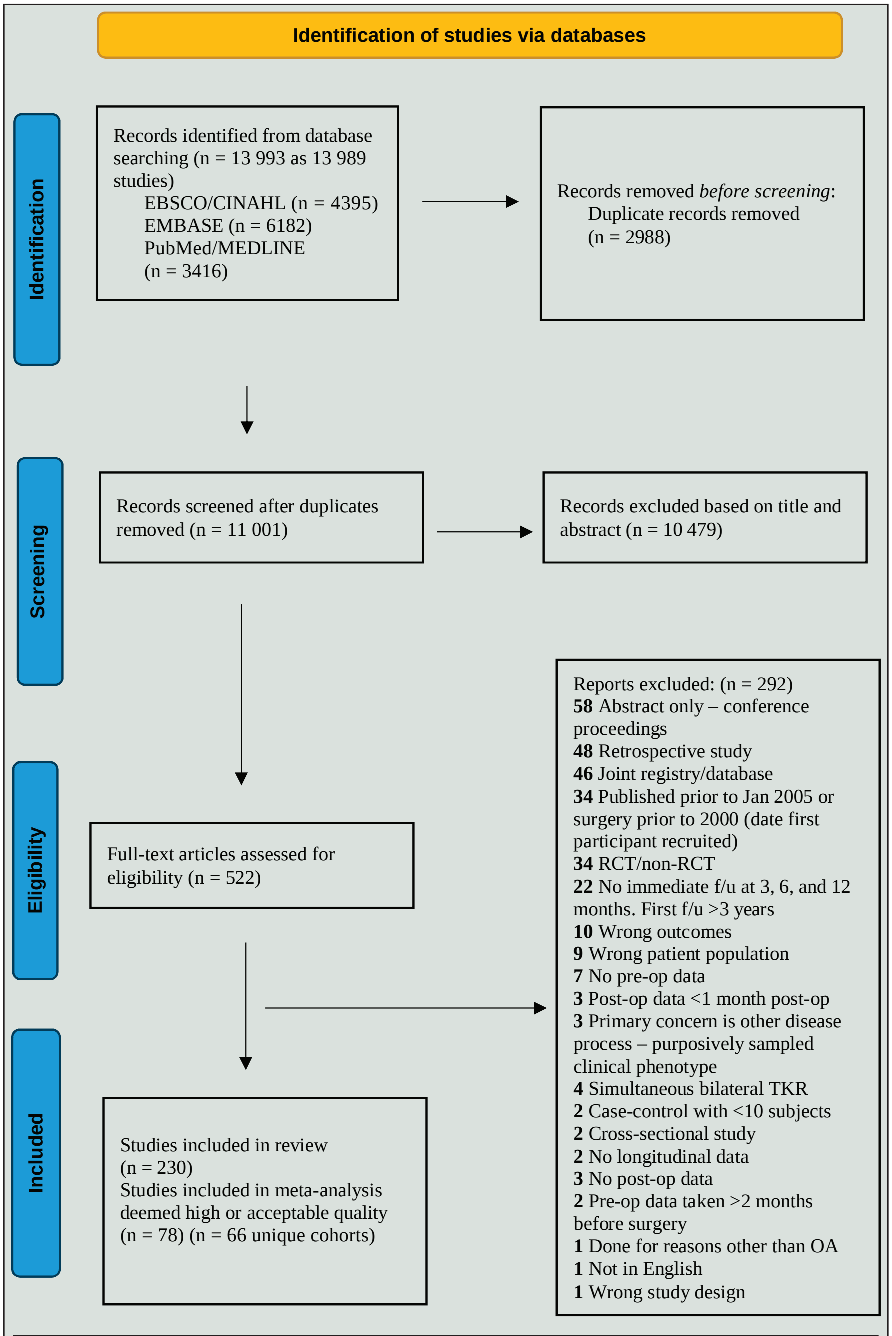

FIGURE 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the review process. Abbreviations: post-op, postoperative; pre-op, preoperative; RCT, randomized controlled trial; TKR, total knee replacement; OA, osteoarthritis.

At no other time point did the point estimate of outcome meet the threshold for acceptable clinical benefit, although the lower bound of the CI exceeded this estimate at all but the long-term follow-up assessment. Our evaluation of heterogeneity and sensitivity analyses are presented in **SUPPLEMENTAL APPENDIX I**.

Clinical Course of Function Assessed Using Performance-Based Measures After TKA Twelve studies reported at least 1 performance-based measure of function following TKA.

Timed up and go Our longitudinal meta-analysis of the timed up and go (TUG) test included 6 studies (n = 432) and 15 observations in total. Pooled weighted means are displayed in **FIGURE 3A**, and percentage change is shown in **FIGURE 3D**. TUG speed improved over the 2-year period, $\chi^2(3) = 105$, $P<.001$, from 11.7 seconds (95% CI, 9.3 to 14.0) presurgery to 9.8 seconds (95% CI, 8.0 to 11.6) at very short-term, 8.5 seconds (95% CI, 6.7 to 10.2) at short-term, and 8.9 seconds (95% CI, 6.5 to 11.2) at medium-term follow-up (**TABLE 3**). While TUG scores at all postsurgical time points differed significantly from the presurgical TUG score, there were no significant differences between sequential postsurgical time points. Only the presurgery versus very short-term follow-up comparison produced evidence for a significant sequential reduction (mean difference, 1.9 seconds; 95% CI, 0.7 to 3.0; $P = .002$). The suggested MCID of a 40% reduction in TUG from baseline[57] would represent a point estimate for acceptable benefit of 7.0 seconds for these data. At no time point did the pooled estimate of TUG speed meet this criterion, although the limit of the 95% CI exceeded the threshold for the medium- and long-term follow-up.

Stair climb test (Stair Climb Capacity) Our longitudinal meta-analysis of the stair climb test (SCT) included 7 studies (n = 435) and 20 observations in total. Pooled weighted means are displayed in **FIGURE 3B**, and percentage change is shown in **FIGURE 3E**. Stair climb capacity changed over time, $\chi^2(3) = 144$, $P<.001$, from 17.8 seconds (95% CI, 12.9 to 22.7) presurgery to 14.3 seconds (95% CI, 10.6 to 18.0) at very short-term, 11.0 seconds (95% CI, 8.1 to 13.8) at short-term, and 13.9 seconds (95% CI, 11.1 to 16.7) at medium-term follow-up (**TABLE 3**). The very short-term follow-up mean was lower than the presurgery mean (mean difference, −3.5 seconds; 95% CI, −5.2 to −1.9; $P<.001$), the short-term follow-up mean was lower

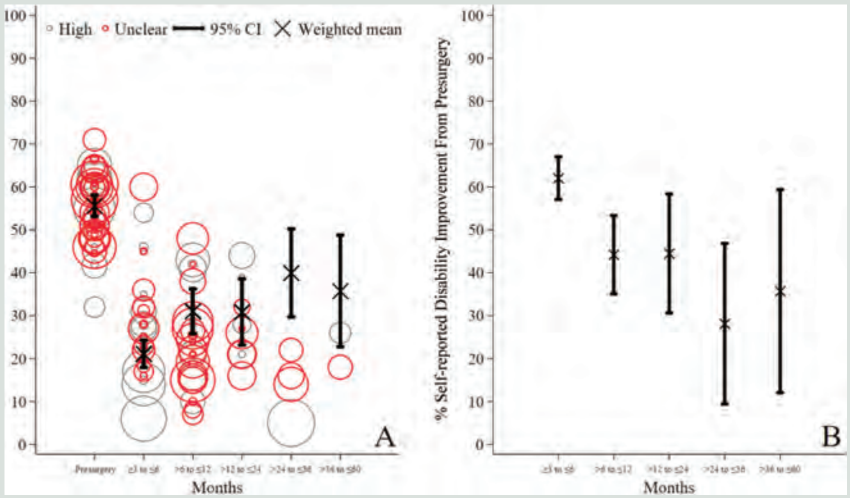

FIGURE 2. (A) The clinical course of self-reported function after total knee arthroplasty. Individual study markers are weighted by the inverse of their sample standard error. (B) Self-reported function as a percentage change from presurgery. These values and 95% confidence intervals (CIs) were calculated from the pooled weighted means (and standard errors), as displayed in panel (A).

than the previous time point value (−3.3 seconds; 95% CI, −5.0 to −1.6; P<.001), and the medium-term follow-up mean was *greater* than the short-term value (an increase of 2.9 seconds; 95% CI, 1.0 to 4.9; P = .002), suggesting some decline in function over the second year of follow-up. A minimal detectable change (MDC) of 5.5 seconds from baseline has been reported[55] and represented a point estimate for acceptable benefit of 12.3 seconds at follow-up for these data. The point estimate of outcome met the threshold for acceptable clinical benefit at short-term follow-up, but not at very short-term or medium-term follow-up.

6-Minute walk test Our longitudinal meta-analysis of the 6-minute walk test (6MWT) included 5 studies (n = 217) and 14 observations. Pooled weighted means are displayed in **FIGURE 3C**, and percentage change is shown in **FIGURE 3F**. The 6MWT distance improved over the 1-year period, $\chi^2(2)$ = 63.7, P<.001, from 369.9 m (95% CI, 293.1 to 446.6) presurgery to 425.6 m (95% CI, 353.1 to 498.0) at very short-term and 450.3 m (95% CI, 380.7 to 519.9) at short-term follow-up (**TABLE 3**). There was an estimated increase of 55.7 m (95% CI, 36.8 to 74.5; P<.001) between presurgery and very short-term follow-up and a further 24.7-m increase (95% CI, 0.7 to 48.7; P = .043) at short-term follow-up. The pooled estimate for 6MWT distance met the MCID criteria at both very short-term (55 m) and short-term (74.3 m) follow-ups, although the lower bound of the 95% CI was below this threshold at both time points.

Overall, there was heterogeneity of study observations across time points (**SUPPLEMENTAL APPENDIX J**).

Sensitivity analysis In the studies that analyzed performance-based measures, no preoperative rehabilitation was reported for any of the groups, but postsurgical rehabilitation was reported in almost all measures (TUG, 5/6 [83%]; SCT, 7/7 [100%]; 6MWT, 5/5 [100%]). Six of these studies were deemed high quality; however, given the small number of studies and small sample sizes, sensitivity analyses were not performed.

Additional functional tests A number of additional functional tests were included that could not proceed to meta-analysis due to the limited number of studies, and a qualitative synthesis for each outcome measure is in **SUPPLEMENTAL APPENDIX K**.

Self-reported versus performance-based functional measures A quantitative analysis comparing the clinical course of

TABLE 2 SELF-REPORTED FUNCTION SCORES, PERCENTAGE CHANGE, AND CERTAINTY OF THE EVIDENCE FOR EACH TIME CATEGORY

Assessment Periods	Self-reported Function Estimated Mean/100 (95% CI)	Self-reported Function Percentage Change From Baseline (95% CI)[a]	Grading of Recommendations Assessment, Development, and Evaluation Judgment
Presurgery	55.6 (53.1 to 58.1)		
3-6 months	21.1 (17.9 to 24.3)	62% (57 to 67)	Moderate certainty
6-12 months	31.0 (25.8 to 36.2)	44.2% (35 to 53.3)	Low certainty
12-24 months	30.9 (23.2 to 38.6)	44.5% (31 to 58.3)	Low certainty
24-36 months	40.0 (29.7 to 50.2)	28% (9.4 to 47)	Not estimated
36-60 months	35.7 (22.7 to 48.8)	36% (12 to 59.3)	Not estimated

Defined assessment periods: presurgery, very short term (≥3 months to ≤6 months), short term (>6 months to ≤12 months), medium term (>12 months to ≤24 months), long term (>24 months to ≤36 months), and very long term (>36 months to ≤60 months).
Abbreviation: CI, confidence interval.
[a]*Positive values indicate improvement.*

TABLE 3 Performance-Based Function Scores and Percentage Change for Each Time Category

Assessment Periods	TUG Estimated Mean (95% CI)	TUG Percentage Change From Baseline (95% CI)[a]	SCT Estimated Mean (95% CI)	SCT Percentage Change From Baseline (95% CI)[a]	6MWT Estimated Mean (95% CI)	6MWT Percentage Change From Baseline (95% CI)[a]
Presurgery	11.7 (9.3 to 14.0)		17.8 (12.9 to 22.70)		369.9 (293.1 to 446.6)	
3-6 months	9.8 (8.0 to 11.6)	16% (7.4 to 24.7)	14.3 (10.6 to 18.0)	18.3% (7.7 to 29)	425.6 (353.1 to 498.0)	18.3% (7.7 to 29)
6-12 months	8.5 (6.7 to 10.2)	24% (8.9 to 39)	11.0 (8.1 to 13.8)	36% (15 to 56.1)	450.3 (380.7 to 519.9)	36% (15 to 56.1)
12-24 months	8.9 (6.5 to 11.2)	19% (−3.3 to 41)	13.9 (11.1 to 16.7)	17.1% (−11.7 to 46)		

Abbreviations: 6MWT, 6-minute walk test; CI, confidence interval; SCT, stair climb test; TUG, timed up and go.
[a]*Positive values indicate improvement.*

self-reported and performance-based measures of physical function was precluded due to limited data for performance-based measures. The percentage changes seen in the performance-based measures appeared somewhat less than was apparent in self-reported measures. Maximal improvement seemed to occur at a later time point for the performance-based measures.

DISCUSSION

We described the clinical course of self-reported and performance-based measures of physical function following primary TKA for knee OA based on data from high- and unclear-quality studies and investigated whether self-reported and performance-based measures of physical function follow similar clinical courses. We included data from 66 unique high- and unclear-quality prospective cohort studies, representing

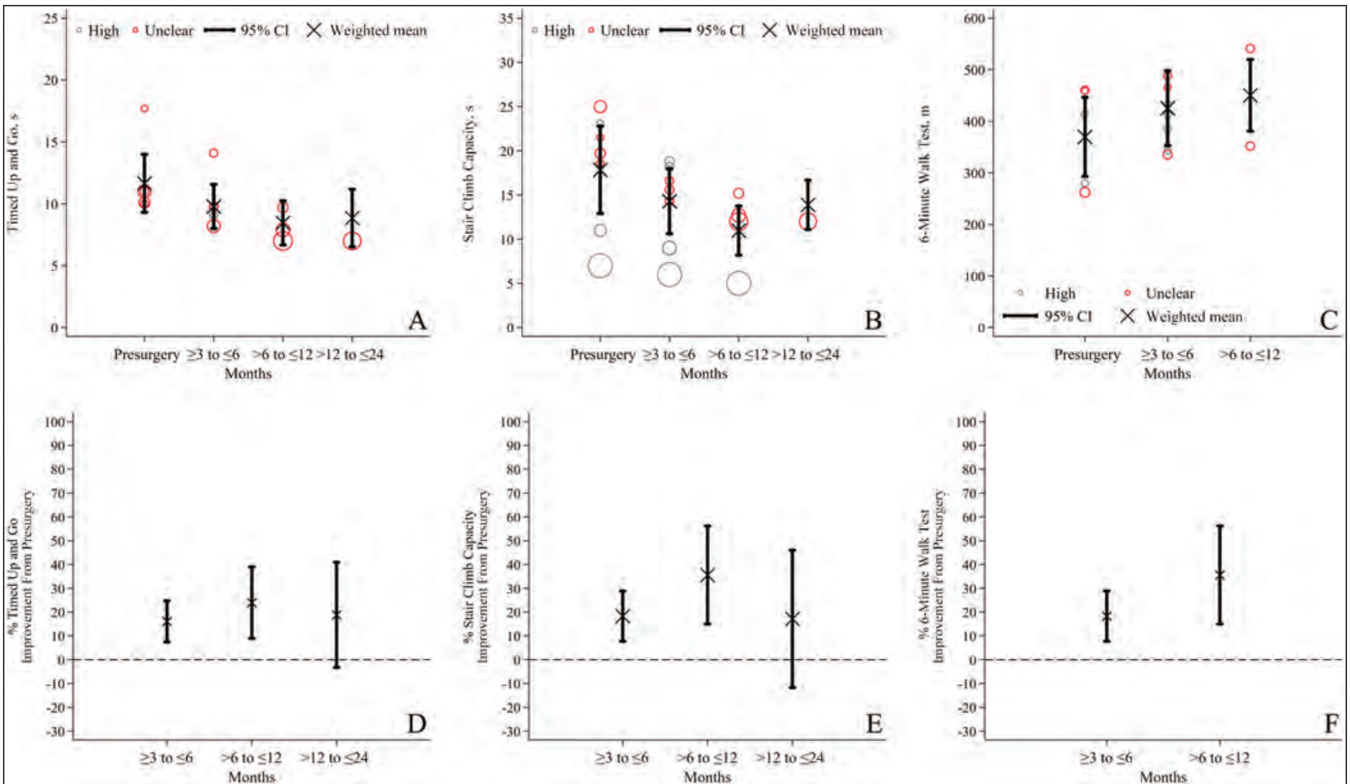

FIGURE 3. (A-C) The clinical course of performance-based measures of function after total knee arthroplasty. Individual study markers are weighted by the inverse of their sample standard error. (D-F) Performance-based measures of function as a percentage change from presurgery. These values and 95% confidence intervals (CIs) were calculated from the pooled weighted means (and standard errors).

19 063 people with knee OA, followed up to a maximum of 5 years.

Clinically meaningful improvement in self-reported function was evident at the 3- to 6-month follow-up based on an estimated SWE of 51% improvement. At no other time point did the estimate of self-reported function meet this criterion for clinically meaningful change. Maximal improvement occurred surprisingly early, and outcomes regressed from this time point and plateaued from the 6- to 12-month follow-up.

Performance-based measures of physical function largely improved post-TKA, although there was a trend for some deterioration in improvement after the 6- to 12-month follow-up. We did not identify an estimate of the SWE for any of the measures used, and decision thresholds for acceptable benefit were based on estimates of MCID or MDC. For approximately two thirds of the assessments, outcome estimates did not meet the criteria for clinically meaningful change, although a small number of studies contributed to these estimates.

Poor agreement between patient-reported and performance-based measures of physical function has been reported previously in people undergoing TKA for knee OA.[38] We were unable to undertake a quantitative comparison of self-reported and performance-based trajectories, although qualitatively, it appears that the percentage change for performance-based measures (**FIGURE 3D-F** and **TABLE 3**) is somewhat less than that for self-reported measures (**FIGURE 2B** and **TABLE 2**). However, the general trend in trajectories was similar. For both patient-reported and performance-based measures, there was improvement at 3-6 months, followed by minimal-to-moderate deterioration in functional improvement beyond the 6- to 12-month follow-up. Sayah et al[47] found a similar trend in their review investigating the clinical course of pain and self-reported function following TKA. Patients experienced marked improvements in both pain and function within the first 12 months postoperatively but little to no improvements between 1 and 10 years postoperatively.[47] This review also reported evidence of slight deterioration in self-reported function in the long term.[47]

We identified 230 recent studies that have evaluated functional status post-TKA. Two thirds of these studies were of low quality and not included in the meta-analyses. The most common problems encountered related to inadequately describing the sample; large loss to follow-up; failure to describe the reasons for missingness; and inadequate reporting of outcome, including clearly describing the timing of data collection. Our review mirrors findings encountered by previous authors.[3,18,45] Researchers planning to evaluate prognosis post-TKA should endeavor to provide high-quality estimates of outcomes that can be easily interpreted by clinicians and consumers by ensuring that these key quality indicators are addressed. Nearly all studies employed consecutive sampling, an important feature in ensuring believability in prognostic studies.

Limitations

Analysis of the data from the surgical arms in controlled trials could provide information to support the aims of this study, but such data were excluded from the review due to problems in both internal and external validity. The lack of data on performance-based measures of physical function from unclear- and high-quality studies meant we were unable to undertake the planned quantitative comparison of trajectories between perceived and measured function.

We applied a single estimate of the SWE derived from the Knee injury and Osteoarthritis Outcome Score[24] across all self-reported measures. This may be deemed a limitation; however, in light of the great variety of thresholds being used to measure response and the lack of standardization in the methodology employed to calculate MCIDs,[9,36] we felt this was a reasonable approach. We were unable to identify an estimate of the SWE for any of the performance-based measures, and the assessment of acceptable benefit for these measures uses scores derived from a variety of nonequivalent methodologies. Finally, there was some between-study variability in the timing of outcome assessment, so our pooled outcome estimates include data collected at slightly different time points.

Clinical Implications

For those contemplating TKA, these data may be useful when deciding between surgery and pursuing other approaches to managing knee OA, particularly if the estimated level of improvement in function is less than what they would deem worthwhile given the cost, risks, and inconvenience of joint replacement surgery. For those choosing TKA, having an evidence-based estimate of postsurgical functional status may reduce reports of dissatisfaction.[10,17] Previous work investigating patient perspectives of pain and function after TKA highlighted the absence of in-depth information about prognosis and functional expectations following surgery.[53] Patients' expectations about postoperative physical function can often remain unfulfilled,[16,41] contributing to dissatisfaction.[23,46,59] Our data can help formulate realistic expectations, which are a feature of satisfaction.

Our review may hold value in alerting clinicians to when patients fall outside of recovery norms, particularly when percentage changes in self-reported function or direct measures of function are less than what was reported here. This may serve as a prompt to instigate activity and lifestyle targeted interventions aimed at increasing physical activity to enhance functional improvements and prevent functional decline.[6,8,52] The lack of clinically meaningful change in most estimates of both self-reported and performance-based measures of function suggests that both clinicians and researchers should urgently consider investigating and implementing ways of enhancing physical function post-TKA.

We have demonstrated that self-reported function will be at the highest levels 6 months post-TKA, and performance-based physical function is unlikely to be improved post-TKA. Specifically, patients should expect early functional improvement, plateauing within the first year after TKA, with minimal improvement (or even deterioration) in physical function thereafter, a pattern mirrored across the TKA literature.[19,34,35,40,64] These data are useful to patients and health professionals to help orientate them to the time course of recovery and expected functional status.

CONCLUSION

For people with knee OA, TKA resulted in clinically meaningful improvements in self-reported function measured at 3-6 months postoperatively. There was some regression in improvement after this, and at no subsequent time point does the estimate of outcome meet our criteria for clinically important change. We found little evidence of clinically meaningful improvement in performance-based measures of physical function at any time point. ◉

KEY POINTS

FINDINGS: Total knee arthroplasty (TKA) results in clinically important improvements in self-reported function at 3-6 months postsurgery. Self-reported function then deteriorates, and at no other time point does the estimate of self-reported function meet our estimate of acceptable benefit. There is limited evidence of clinically meaningful change in performance-based measures of physical function at any time point post-TKA.

IMPLICATIONS: This study provides patients and health care professionals with high-quality data about expected functional recovery post-TKA. These data are important in forming decisions around whether to undergo TKA or not, help formulate expectations in those who decide to proceed with TKA, and should prompt clinicians and researchers to consider ways of enhancing physical function post-TKA.

CAUTION: Certainty of the evidence was moderate to low, and two thirds of the studies were deemed low quality and excluded from evidence synthesis; therefore, there is the possibility that further research may lead to different conclusions. Generalizability to populations other than people with knee OA undertaking a TKA is unclear. Clinicians should discuss these uncertainties with patients to facilitate informed decision making.

STUDY DETAILS

AUTHOR CONTRIBUTIONS: G.M.O., M.C.M., D.A.H., and B.M.W. contributed to study design. G.M.O. and M.J.T. screened titles and abstracts. G.M.O., M.J., S.S., and S.K. screened full-text references, assessed risk of bias of the included studies, and extracted data. M.C.M., D.A.H., and G.M.O. contributed to data analysis. All authors contributed to interpretation of the results and manuscript preparation.

DATA SHARING: Data extracted directly from the included studies are available on request; however, data that were provided directly from the corresponding authors will not be made publicly available and should be requested from the authors.

PATIENT AND PUBLIC INVOLVEMENT: There was no patient involvement.

REFERENCES

1. Ackerman IN, Bohensky MA, Zomer E, et al. The projected burden of primary total knee and hip replacement for osteoarthritis in Australia to the year 2030. *BMC Musculoskelet Disord.* 2019;20:90. https://doi.org/10.1186/s12891-019-2411-9
2. Ackerman IN, Buchbinder R, March L. Global Burden of Disease Study 2019: an opportunity to understand the growing prevalence and impact of hip, knee, hand and other osteoarthritis in Australia. *Intern Med J.* 2023;53:1875-1882. https://doi.org/10.1111/imj.15933
3. Adriani M, Becker R, Milano G, Lachowski K, Prill R. High variation among clinical studies in the assessment of physical function after knee replacement: a systematic review. *Knee Surg Sports Traumatol Arthrosc.* 2023;31:3854-3860. https://doi.org/10.1007/s00167-023-07375-2
4. Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ.* 2001;323:224-228. https://doi.org/10.1136/bmj.323.7306.224
5. Amaro JT, Arliani GG, Astur DC, Debieux P, Kaleka CC, Cohen M. No difference between fixed- and mobile-bearing total knee arthroplasty in activities of daily living and pain: a randomized clinical trial. *Knee Surg Sports Traumatol Arthrosc.* 2017;25:1692-1696. https://doi.org/10.1007/s00167-016-4106-1
6. Arnold JB, Walters JL, Ferrar KE. Does physical activity increase after total hip or knee arthroplasty for osteoarthritis? A systematic review. *J Orthop Sports Phys Ther.* 2016;46:431-442. https://doi.org/10.2519/jospt.2016.6449
7. Australian Orthopaedic Association National Joint Replacement Registry. Hip, Knee & Shoulder Arthroplasty: 2022 Annual Report. North Adelaide, SA: Australian Orthopaedic Association; 2022.
8. Bahl JS, Millar SC, Fraysse F, et al. Changes in 24-hour physical activity patterns and walking gait biomechanics after primary total hip arthroplasty: a 2-year follow-up study. *J Bone Joint Surg Am.* 2021;103:1166-1174. https://doi.org/10.2106/jbjs.20.01679
9. Beiene ZA, Tanghe KK, Kahlenberg CA, AS ML, CH ML, Gausden EB. Defining a successful total knee arthroplasty: a systematic review of metrics of clinically important changes. *Arthroplasty.* 2023;5:25. https://doi.org/10.1186/s42836-023-00178-3
10. Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open.* 2012;2:e000435. https://doi.org/10.1136/bmjopen-2011-000435
11. Bosdriesz JR, Stel VS, van Diepen M, et al. Evidence-based medicine—when observational studies are better than randomized controlled trials. *Nephrology.* 2020;25:737-743. https://doi.org/10.1111/nep.13742
12. Brooke BS, Schwartz TA, Pawlik TM. MOOSE reporting guidelines for meta-analyses of observational studies. *JAMA Surg.* 2021;156:787-788. https://doi.org/10.1001/jamasurg.2021.0522
13. Bullens PHJ, van Loon CJM, de Waal Malefijt MC, Laan RFJM, Veth RPH. Patient satisfaction after total knee arthroplasty: a comparison between subjective and objective outcome assessments. *J Arthroplasty.* 2001;16:740-747. https://doi.org/10.1054/arth.2001.23922
14. Calabro L, Clement ND, MacDonald D, Patton JT, Howie CR, Burnett R. Venous thromboembolism after total knee arthroplasty is associated with a worse functional outcome at one year. *Bone Joint J.* 2021;103:1254-1260. https://doi.org/10.1302/0301-620x.103b7.Bjj-2019-0636.R7
15. Chung JY, Min B-H. Is bicompartmental knee arthroplasty more favourable to knee muscle strength and physical performance compared to total knee arthroplasty? *Knee Surg Sports Traumatol Arthrosc.* 2013;21:2532-2541. https://doi.org/10.1007/s00167-013-2489-9
16. Conner-Spady BL, Bohm E, Loucks L, Dunbar MJ, Marshall DA, Noseworthy TW.

Patient expectations and satisfaction 6 and 12 months following total hip and knee replacement. *Qual Life Res.* 2020;29:705-719. https://doi.org/10.1007/s11136-019-02359-7

17. 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;27:1311-1322. https://doi.org/10.1007/s11136-018-1808-5

18. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. *Osteoarthritis Cartilage.* 2012;20:1548-1562. https://doi.org/10.1016/j.joca.2012.08.015

19. Dowsey MM, Smith AJ, Choong PFM. Latent Class Growth Analysis predicts long term pain and function trajectories in total knee arthroplasty: a study of 689 patients. *Osteoarthritis Cartilage.* 2015;23:2141-2149. https://doi.org/10.1016/j.joca.2015.07.005

20. Evans JT, Walker RW, Evans JP, Blom AW, Sayers A, Whitehouse MR. How long does a knee replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. *Lancet.* 2019;393:655-663. https://doi.org/10.1016/S0140-6736(18)32531-5

21. Finch CF, Kemp JL, Clapperton AJ. The incidence and burden of hospital-treated sports-related injury in people aged 15+ years in Victoria, Australia, 2004–2010: a future epidemic of osteoarthritis? *Osteoarthritis Cartilage.* 2015;23:1138-1143. https://doi.org/10.1016/j.joca.2015.02.165

22. Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ.* 2008;336:1472-1474. https://doi.org/10.1136/bmj.39590.732037.47

23. Gunaratne R, Pratt DN, Banda J, Fick DP, RJK K, Robertson BW. Patient dissatisfaction following total knee arthroplasty: a systematic review of the literature. *J Arthroplasty.* 2017;32:3854-3860. https://doi.org/10.1016/j.arth.2017.07.021

24. Henderson N, Riddle DL. The smallest worthwhile effect is superior to the MCID for estimating acceptable benefits of knee arthroplasty. *J Clin Epidemiol.* 2022;152:201-208. https://doi.org/10.1016/j.jclinepi.2022.10.019

25. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ.* 2011;343:d5928. https://doi.org/10.1136/bmj.d5928

26. Higgins JPT, Altman DG, Sterne JAC, eds. Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. in *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011]. The Cochrane Collaboration; 2011.

27. Higgins JPT, Thomas J, Chandler J, et al, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.3* [updated February 2022]. The Cochrane Collaboration; 2022. www.training.cochrane.org/handbook

28. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002;21:1539-1558. https://doi.org/10.1002/sim.1186

29. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. *Lancet.* 2019;393:1745-1759. https://doi.org/10.1016/S0140-6736(19)30417-9

30. Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ.* 2015;350:h870. https://doi.org/10.1136/bmj.h870

31. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ.* 2001;323:42-46. https://doi.org/10.1136/bmj.323.7303.42

32. Kamper SJ, Rebbeck TJ, Maher CG, JH MA, Sterling M. Course and prognostic factors of whiplash: a systematic review and meta-analysis. *Pain.* 2008;138:617-629. https://doi.org/10.1016/j.pain.2008.02.019

33. Khatri C, Dhaif F, Ellard D, et al. What recovery domains are important following a total knee replacement? A qualitative, interview-based study. *BMJ Open.* 2024;14:e080795. https://doi.org/10.1136/bmjopen-2023-080795

34. Lenguerrand E, Wylde V, Gooberman-Hill R, et al. Trajectories of pain and function after primary hip and knee arthroplasty: the ADAPT cohort study. *PLoS One.* 2016;11:e0149306. https://doi.org/10.1371/journal.pone.0149306

35. Lindberg MF, Rustøen T, Miaskowski C, Rosseland LA, Lerdal A. The relationship between pain with walking and self-rated health 12 months following total knee arthroplasty: a longitudinal study. *BMC Musculoskelet Disord.* 2017;18:1-10. https://doi.org/10.1186/s12891-017-1430-7

36. 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;32:19. https://doi.org/10.1186/s43019-020-00038-3

37. Menezes Costa LDC, Maher CG, Hancock MJ, McAuley JH, Herbert RD, Costa LOP. The prognosis of acute and persistent low-back pain: a meta-analysis. *CMAJ.* 2012;184:E613-E624. https://doi.org/10.1503/cmaj.111271. [published Online First: 20120514]

38. 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;26:728-737. https://doi.org/10.1016/j.arth.2010.06.004

39. Murphy M, Travers M, Gibson W, et al. Rate of improvement of pain and function in mid-portion Achilles tendinopathy with loading protocols: a systematic review and longitudinal meta-analysis. *Sports Med.* 2018;48:1875-1891. https://doi.org/10.1007/s40279-018-0932-2

40. Nilsdotter AK, Toksvig-Larsen S, Roos EM. A 5 year prospective study of patient-relevant outcomes after total knee replacement. *Osteoarthritis Cartilage.* 2009;17:601-606. https://doi.org/10.1016/j.joca.2008.11.007

41. Nilsdotter AK, Toksvig-Larsen S, Roos EM. Knee arthroplasty: are patients' expectations fulfilled? A prospective study of pain and function in 102 patients with 5-year follow-up. *Acta Orthop.* 2009;80:55-61. https://doi.org/10.1080/17453670902805007

42. Ono S, Sebastin SJ, Chung KC. Evidence on prognosis. *Plast Reconstr Surg.* 2013;131:655-665. https://doi.org/10.1097/PRS.0b013e31827c6f90

43. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. https://doi.org/10.1136/bmj.n71

44. Pengel LHM, Herbert RD, Maher CG, Refshauge KM. Acute low back pain: systematic review of its prognosis. *BMJ.* 2003;327:323. https://doi.org/10.1136/bmj.327.7410.323

45. Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: a systematic review. *Arthritis Rheum.* 2008;59:876-883. https://doi.org/10.1002/art.23706

46. Robertsson O, Ranstam J, Lidgren L. Variation in outcome and ranking of hospitals: an analysis from the Swedish Knee Arthroplasty Register. *Acta Orthop.* 2006;77:487-493. https://doi.org/10.1080/17453670610046442

47. Sayah SM, Karunaratne S, Beckenkamp PR, et al. Clinical course of pain and function following total knee arthroplasty: a systematic review and meta-regression. *J Arthroplasty.* 2021;36:3993-4002.E37. https://doi.org/10.1016/j.arth.2021.06.019

48. Schünemann HJ, Oxman AD, Brozek J, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *BMJ Evid Based Med.* 2008;13:162-163. https://doi.org/10.1136/ebm.13.6.162-a

49. Skou ST, Roos EM, Simonsen O, et al. The effects of total knee replacement and non-surgical treatment on pain sensitization and clinical pain. *Eur J Pain.* 2016;20:1612-1621. https://doi.org/10.1002/ejp.878

50. Sterne JAC, Egger M. *Regression Methods to Detect Publication and Other Bias in Meta-Analysis.* Chichester, UK: John Wiley & Sons, ; 2005.

51. Sydney Innovation and Research Symposium. *Placebo Surgical Trials* [video]. YouTube; 2020. https://www.youtube.com/watch?v=jF3d059QBkM

52. Takamura D, Iwata K, Sueyoshi T, Yasuda T, Moriyama H. Relationship between early physical activity after total knee arthroplasty and postoperative physical function: are these related? *Knee Surg Relat Res.* 2021;33:35. https://doi.org/10.1186/s43019-021-00118-y

53. Taylor CEV, Murray CM, Stanton TR. Patient perspectives of pain and function after knee replacement: a systematic review and meta-synthesis of qualitative studies. *Pain Rep.* 2022;7:e1006. https://doi.org/10.1097/pr9.0000000000001006

54. Tolk JJ, Janssen RPA, Haanstra TM, van der Steen MC, Bierma-Zeinstra SMA, Reijman M. The influence of expectation modification in knee arthroplasty on satisfaction of patients: a randomized controlled trial. *Bone Joint J.* 2021;103:619-626. https://doi.org/10.1302/0301-620x.103b4.bjj-2020-0629.r3

55. Tolk JJ, Janssen RPA, Prinsen CAC, et al. The OARSI core set of performance-based measures for knee osteoarthritis is reliable but not valid and responsive. *Knee Surg Sports Traumatol Arthrosc.* 2019;27:2898-2909. https://doi.org/10.1007/s00167-017-4789-y

56. Trieu J, Gould DJ, Schilling C, Spelman T, Dowsey MM, Choong PF. Patient-reported outcomes following total knee replacement in patients <65 years of age—a systematic review and meta-analysis. *J Clin Med.* 2020;9:3150. https://doi.org/10.3390/jcm9103150

57. Unnanuntana A, Ruangsomboon P, Keesukpunt W. Validity and responsiveness of the two-minute walk test for measuring functional recovery after total knee arthroplasty. *J Arthroplasty.* 2018;33:1737-1744. https://doi.org/10.1016/j.arth.2018.01.015

58. Valkering KP, Breugem SJ, van den Bekerom MPJ, Tuinebreijer WE, van Geenen RCI. Effect of rotational alignment on outcome of total knee arthroplasty. *Acta Orthop.* 2015;86:432-439. https://doi.org/10.3109/17453674.2015.1022438

59. Verhaar J. Patient satisfaction after total knee replacement—still a challenge. *Acta Orthop.* 2020;91:241-242. https://doi.org/10.1080/17453674.2020.1763581

60. Wallwork SB, Braithwaite FA, O'Keeffe M, et al. The clinical course of acute, subacute and persistent low back pain: a systematic review and meta-analysis. *CMAJ.* 2024;196:E29-E46. https://doi.org/10.1503/cmaj.230542

61. Woodland N, Takla A, Estee MM, et al. Patient-reported outcomes following total knee replacement in patients aged 65 years and over—a systematic review. *J Clin Med.* 2023;12:1613. https://doi.org/10.3390/jcm12041613

62. Woolhead GM, Donovan JL, Dieppe PA. Outcomes of total knee replacement: a qualitative study. *Rheumatology.* 2005;44:1032-1037. https://doi.org/10.1093/rheumatology/keh674

63. Wylde V, Bruce J, Beswick A, Elvers K, Gooberman-Hill R. Assessment of chronic postsurgical pain after knee replacement: a systematic review. *Arthritis Care Res.* 2013;65:1795-1803. https://doi.org/10.1002/acr.22050

64. Wylde V, Penfold C, Rose A, Blom AW. Variability in long-term pain and function trajectories after total knee replacement: a cohort study. *Orthop Traumatol Surg Res.* 2019;105:1345-1350. https://doi.org/10.1016/j.otsr.2019.08.014

65. Yap YYW, Edwards KL, Soutakbar H, Fernandes GS, Scammell BE. Oxford Knee Score 1 year after TKR for osteoarthritis with reference to a normative population: what can patients expect? *Osteoarthr Cartil Open.* 2021;3:100143. https://doi.org/10.1016/j.ocarto.2021.100143