

Original Article

Knee osteoarthritis and minimal important change for the nine-step stair climb test

Julie B. Pajaniaye^{1, 2}, Eric P. Cheret¹, Cecilie H. Langvad¹, Pætur M. Holm^{3, 4, 5, 6}, Søren T. Skou^{3, 4}, Josefine B. Larsen^{1, 7} & Inger Mechlenburg^{1, 8}

¹) Department of Orthopaedics, Aarhus University Hospital, ²) Department of Dentistry and Oral Health, Aarhus University, ³) The Research and Implementation Unit PROgrez, Department of Physiotherapy and Occupational Therapy, Næstved-Slagelse-Ringsted Hospitals, ⁴) Research Unit for Musculoskeletal Function and Physiotherapy, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, ⁵) Center for Surgery, National Hospital of Faroe Islands, Faroe Islands, ⁶) Faculty of Health Sciences, University of Faroe Islands, Faroe Islands, ⁷) Department of Clinical Medicine, Aarhus University, ⁸) Department of Public Health, Aarhus University, Denmark

Dan Med J 2026;73(3):A03250164. doi: 10.61409/A03250164

ABSTRACT

INTRODUCTION. Stair-climbing tests (SCT) assess stair negotiation in knee osteoarthritis (OA). The minimal important change (MIC) is a measure of interpretability. The MIC of the nine-step SCT for patients with knee OA has not been established. In a secondary analysis from a randomised controlled trial, we aimed to estimate the MIC of the nine-step SCT in patients with knee OA after 12 weeks of neuromuscular exercise.

METHODS. In a previously reported randomised controlled trial, knee OA patients not eligible for knee arthroplasty were randomised to a 12-week neuromuscular exercise intervention with/without additional low-dose strength training. SCT data were collected at baseline and after 12 weeks. Self-reported experience of change in the patient's knee problem was measured with the seven-point anchor question Global Perceived Effect Score (GPE) at 12 weeks. Among the 90 patients included at baseline, 72 had complete follow-up data. In this secondary analysis, an adjusted MIC estimate was calculated using predictive modelling.

RESULTS. The mean SCT time at baseline was 12.6 (± 5.5) seconds, and the mean change in SCT at 12 weeks was 2.6 (± 3.3) seconds. The adjusted MIC value (95% CI) for improvement was 2.2 seconds (95% CI: 1.5; 3.0) for the nine-step SCT.

CONCLUSIONS. The 2.2-second MIC value for the nine-step SCT in knee OA patients may be considered a first indicator of the magnitude of minimal improvement that the average patient would find important.

FUNDING. None.

TRIAL REGISTRATION. ClinicalTrials.gov NCT03215602.

Reduced function and pain are hallmark knee osteoarthritis (OA) symptoms, often affecting the ability to climb stairs [1, 2]. Exercise interventions are effective in reducing pain and improving function [3] and are recommended first-line treatments across guidelines [4]. Stair climb tests (SCT) are inexpensive and easily applicable assessments of functional capacity in knee OA [1, 2] and are available in different modalities [2, 5-7], with the nine-step SCT being one of the performance tests recommended by the Osteoarthritis Research Society International [2].

The minimal important change (MIC) is a measure of interpretability defined as the smallest within-person

change in a score over time that a patient both perceives as a change and assesses as important [8]. An anchor-based approach to MIC estimation combines the functional test with an external patient-reported instrument to assess the importance of change. Several anchor-based methods exist; however, a predictive modelling approach may yield the most precise estimates [9].

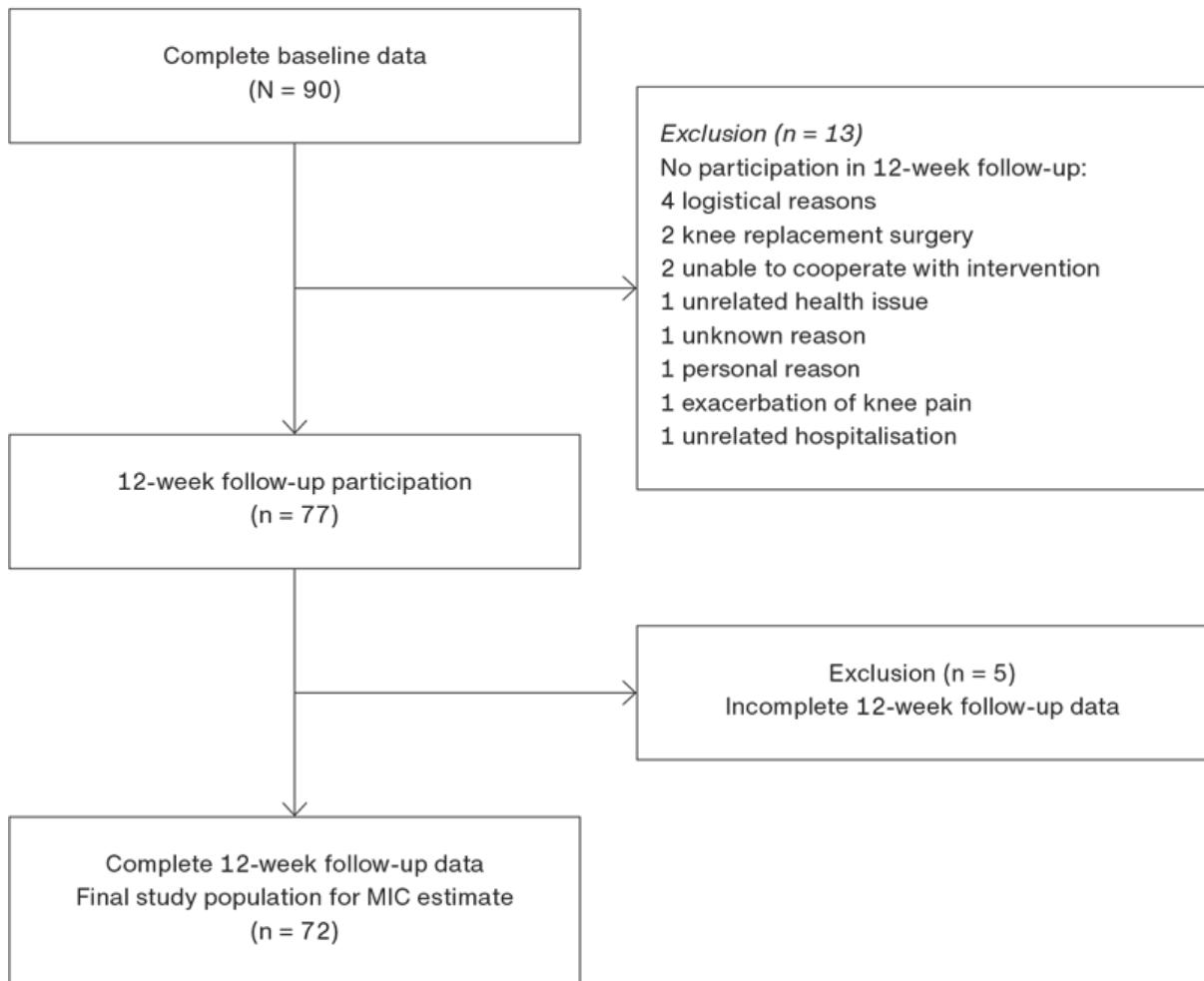
The MIC value has been established for several knee OA outcome measurements [10]. Despite its high clinical feasibility and relevance in trials and clinical practice as a measure that eases the clinical interpretation of treatment results, the MIC for the nine-step SCT has yet to be determined [11]. Firstly, we aimed to estimate the MIC of the nine-step SCT in patients with knee OA who were not eligible for knee arthroplasty following a 12-week exercise intervention with/without additional low-dose strength training using predictive modelling. Secondly, we aimed to estimate the proportion of responders, i.e. patients achieving a change score above the MIC in the same cohort.

METHODS

Study design

This study was planned and reported in accordance with the COSMIN reporting guideline for studies on measurement properties [12] and is a secondary analysis of data from a previously published randomised controlled trial [13]. We used data from a trial conducted in Denmark between July 2017 and October 2018. The primary report of this trial has been published elsewhere [13]. In brief, patients with symptomatic and radiographic (Kellgren-Lawrence grade ≥ 2) knee OA who were ineligible for knee replacement surgery were randomised to 12 weeks of neuromuscular exercise and education (NEMEX-EDU) with/without additional low-dose strength training delivered by trained physiotherapists. NEMEX-EDU was identical to the “Good Life with osteoArthritis in Denmark” (GLA:D) programme [14]. A total of 90 participants were randomised into two groups. Among the 77 who participated in the follow-up, 72 (31 in the intervention group; 41 in the control group) had complete data (Figure 1). Data were collected at baseline and at a 12-week follow-up. The primary outcome of the trial was change in the Activities of Daily Living domain of the Knee Injury and Osteoarthritis Outcome Score (KOOS) [15]. The change in nine-step SCT was one of several secondary outcomes.

FIGURE 1 Study flow chart.



MIC = minimal important change.

Patient characteristics and measurements

Baseline data were collected on sex, age, BMI, pain and all five KOOS subscales. The SCT was performed at baseline and at 12 weeks, supervised by a single trained assessor. The total time in seconds (to the nearest tenth) to descend and ascend nine steps (step height of 20 cm) at a safe pace was recorded by the assessor using a stopwatch. The participant was instructed to descend nine steps to the stairway landing, turn swiftly and climb the same nine steps. Use of handrails for safety was allowed. The test was conducted after the assessor had demonstrated the test.

Anchor question

The anchor questionnaire data were collected at follow-up using a generic Global Perceived Effect (GPE) instrument [16]. The participants rated how much their knee problems had improved from the initial baseline on a seven-point Likert scale. The response options were as follows: 1) Worse, an important deterioration, 2) somewhat worse, enough to be an important deterioration, 3) very small deterioration, not an important deterioration, 4) same, 5) very small improvement, not an important improvement, 6) somewhat better, enough to be an important improvement, and 7) better, an important improvement.

Ethics and data sharing statement

The trial was registered with ClinicalTrials.gov (ID: NCT03215602) and approved by the Danish Scientific Ethical Committee, Region Zealand (SJ-517) and the Danish Data Protection Agency (REG-61-2016). The primary report is published with supplementary material [13]. No additional data are available for this trial.

Statistics

The primary results of the original report found no statistically significant differences in the primary outcome between randomisation groups [13]. In this secondary analysis, this was the starting point for the analysis plan. Data were subjected to descriptive statistical analyses. We estimated frequencies of categorical variables, means and standard deviations of continuous normally distributed variables, and medians with interquartile ranges for non-normally distributed continuous variables. The SCT change score values were plotted across the answers to the GPE question.

The GPE data were dichotomised into either not importantly improved (scores 1-5) or importantly improved (scores 6 and 7) to establish the anchor. Before conducting a MIC analysis, two steps were undertaken to assess its feasibility. First, anchor validity was assessed by calculating a correlation coefficient between the anchor and the SCT change. Low or no correlation between the anchor and the SCT change score may yield imprecise MIC estimates [17]. In accordance with recommendations, a correlation coefficient of ≥ 0.3 was set as satisfactory for proceeding with an MIC analysis [18]. We examined the correlations between the anchor and baseline score, and the anchor and follow-up score, since stronger correlations between these than between anchor and change score may also compromise anchor credibility [17]. Secondly, to assess whether all participants could be considered one cohort irrespective of randomisation group, dependency between randomisation group and the dichotomised GPE question was assessed using a χ^2 test. As both steps yielded satisfactory results, MIC was estimated.

We used the predictive modelling approach [9] to estimate the MIC value. The predictive modelling approach utilises a logistic regression analysis to establish the predictive MIC value (MIC(pred)) and allows for the calculation of a 95% CI. By introducing an interaction term, the method can account for the potential effect modification of the baseline test score on the change in test score [9].

The change in SCT corresponding to a likelihood ratio of 1 was estimated as the MIC(pred) value [9]. The dichotomised anchor was the dependent variable in the analyses. The change in SCT was the independent variable. We adjusted the MIC(pred) for the proportion of improved participants to calculate the adjusted MIC (MIC(adj)) [9, 19] and calculated 95% CIs for MIC(pred) and MIC(adj) using nonparametric bootstrapping with 1,000 iterations.

To assess whether the MIC estimate depended on baseline SCT scores, an interaction term between baseline SCT and SCT change score was included in the logistic regression analysis. Interaction was considered present at $p < 0.05$. The proportion of responders, i.e. patients achieving the MIC, was quantified. Statistical analyses were performed using STATA version 17 (StataCorp LLC, TX, USA) and R version 4.5.1 (R Foundation for Statistical Computing, Vienna, Austria).

Trial registration: ClinicalTrials.gov NCT03215602.

RESULTS

The mean age of the 72 participants with complete follow-up data was 65.5 ± 10.0 years, and 39 (54%) were women (Table 1). Mean SCT at baseline was 12.6 ± 5.5 seconds, improving to 10.0 ± 4.8 seconds at follow-up with

a mean change of 2.6 ± 3.3 seconds. A total of 58% perceived an important improvement (GPE scores 6 or 7) in their knee problem. The distribution of SCT change scores across GPE responses is shown in **Figure 2**. The Spearman correlation between SCT change and the anchor scores was 0.31 ($p = 0.008$), which exceeded the threshold for estimating the MIC. The Spearman correlation between the anchor and the baseline score was -0.05 ($p = 0.66$), and the correlation between the anchor and the follow-up was -0.28 ($p = 0.016$). No dependency between the randomisation group and GPE score was found ($p = 0.66$). The MIC(pred) estimated using the predictive modelling method was 2.3 seconds (95% CI: 1.7- 3.1 seconds). Adjusting for the 58% proportion of importantly improved participants produced a MIC(adj) of 2.2 seconds (95% CI: 1.5; 3.0 seconds). The interaction term between baseline SCT and change in SCT was not statistically significant ($p = 0.32$), indicating that the MIC estimate was independent of baseline values. The proportion of responders was 40% (**Table 2**).

TABLE 1 Baseline characteristics of the study population (N = 72).

Females, n (%)	39 (54)
Age, mean \pm SD, yrs	65.5 \pm 10.0
BMI, median (IQR), kg/m ²	29.0 (7.9)
Site of pain, right knee, n (%)	29 (41)
Contralateral pain, n (%)	46 (65)
<i>VAS pain measures, mean \pm SD</i>	
Within 24 h	50.2 \pm 21.5
Last week	52.9 \pm 19.7
After 30 min. walk	55.1 \pm 25.9
<i>KOOS measures</i>	
Pain, mean \pm SD	47.0 \pm 14.2
Other symptoms, mean \pm SD	53.5 \pm 14.7
Function, daily life, median (IQR)	55.9 (17.7)
Function, sports and recreation, mean \pm SD	19.7 \pm 13.9
Knee-related quality of life, mean \pm SD	33.0 \pm 14.0

KOOS = Knee Injury and Osteoarthritis Outcome Score (0-100);

VAS = visual analogue scale (0-100).

FIGURE 2 Individual change scores in 9-step stair climb test (SCT) by Global Perceived Effect questionnaire categories.

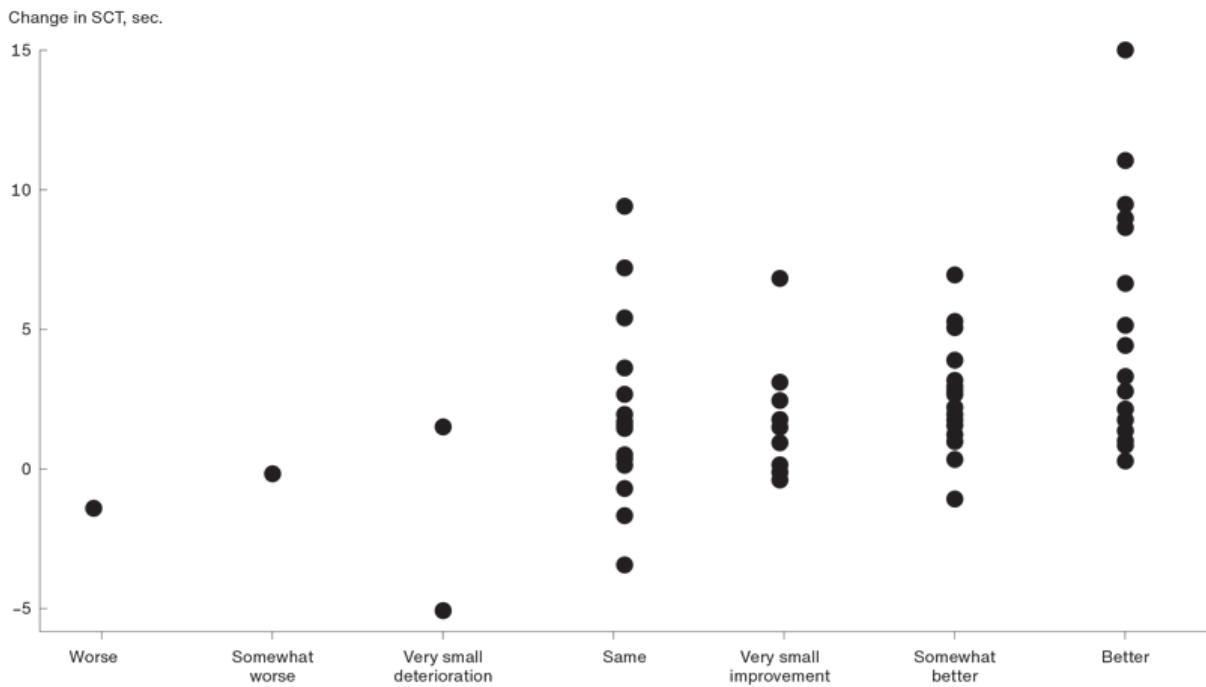


TABLE 2 Distribution of minimal important change responders across anchor question response. The values are n (%) (N = 72).

	Important improvement	No important improvement	Total
Responder ^a	21 (50)	8 (26.7)	29 (40)
Not responder	21 (50)	22 (73.3)	43 (60)

a) Knee osteoarthritis patient reaching the minimal important change threshold of 2.2 sec. of change on the nine-step stair climb test after 12 wks of neuromuscular exercise.

DISCUSSION

This study presents the first estimate of the MIC for the nine-step SCT in mild-to-moderate knee OA patients participating in 12 weeks of exercise, finding a MIC(adj) of 2.2 seconds. This estimate may represent the minimum change in SCT score that an average patient with knee osteoarthritis, who is not eligible for knee arthroplasty and has mild to moderate symptoms, would perceive as clinically important after 12 weeks of exercise. This estimate may serve as a first indicator of a possible MIC value for this patient group undergoing exercise-based interventions and may inform shared decision-making in clinical practice regarding expectations of functional improvement in stair negotiation after 12 weeks of exercise. Furthermore, the MIC estimate may

inform discussions regarding the expected number of responders, i.e. participants reaching the MIC value, in knee OA clinical trials using exercise interventions.

Few other studies have reported MIC estimates for SCTs with different step numbers, using methods ranging from distribution-based to ROC-based approaches.

In a six-step SCT MIC study, pooled data from four intervention studies for participants with knee or hip OA conducted in an Australian population were used to calculate MICs using three methods [6]. The median MIC estimate was 1.37 seconds. Baseline six-step SCT was 9.2 seconds, and follow-up SCT after 12 weeks of intervention was 8.5 seconds.

Our study population may be comparable to the participants in this study regarding functional level and OA symptoms, and the four pooled datasets provide a large study population. Our MIC estimate of 2.2 seconds for the nine-step SCT is larger than that for the six-step test, which may seem reasonable, as one might expect a larger MIC with more steps to complete. However, the differences between the two studies in terms of the underlying conditions of the study population and the applied methodology mean that no such conclusion may be inferred.

Another study estimated a four-step MIC of 3.21 seconds using ROC analysis in a predominantly female West Asian population with mild to moderate knee OA undergoing a four-week training, manual and electro-therapy intervention [7]. Baseline SCT was 16.5 seconds, and the change in SCT was 4.9 seconds. As indicated by the baseline and change SCT values reported in the West Asian study, differences in functional level and osteoarthritis severity between that population and ours seem to limit the expectation that the SCT MIC would increase with greater stair-step counts.

Our findings regarding existing SCT estimates highlight the context-specific nature of MIC. All specificities regarding tests and methodologies require close consideration when assessing the clinical and practical interpretability of our MIC estimate, highlighting some limitations of this study.

Firstly, our sample size was smaller than the currently recommended 100 patients for robust MIC estimation [17]. Also, the anchor question was disease- but not test-specific. This introduces some randomness in responses. Ideally, the anchor question should align with the specific outcome measure. Furthermore, we could not rule out all sources of potential bias. GPE scores may be at risk of recall bias [20]. We found a correlation between the GPE scores and the SCT change that just reached the threshold for MIC estimation, while the correlation between GPE and the follow-up score was nearly as large. Participants with an SCT change below the MIC value were as likely to report an important improvement as participants above it. The distribution of GPE scores (Figure 2) may suggest that participants were more likely to report improvement even with minimal changes, which may limit the validity of our estimate. The original trial found a small but statistically significant difference in SCT time between randomisation groups, pointing to a potential supplementary effect of additional low-dose strength training on SCT performance, which may limit the generalisability of our MIC estimate.

While the study carries some limitations, it also has its strengths. In contrast to previous studies, we were able to establish a 95% MIC CI using predictive modelling and to adjust for the proportion of participants who improved. We used a consistent definition of MIC, avoiding further conceptual and methodological confusion about the interpretability of change. Additionally, the use of the anchor-based method combined with predictive modelling is a key strength of our study.

CONCLUSIONS

MIC estimates represent the smallest change in a score that the average patient perceives as important following

an intervention. We found a MIC(adj) of 2.2 seconds for the nine-step SCT in knee OA patients, not eligible for knee replacement surgery, undergoing a 12-week exercise-based intervention, with 40% being responders. This MIC estimate may serve as a first indicator of the MIC for the nine-step SCT in this population. We invite validation of our findings in larger studies overcoming the present limitations to allow for utilisation of the estimate as the minimal threshold for treatment response in individual patients in intervention studies and clinical practice. Our results may be used to guide discussions on further trial planning and considerations of the expected proportion of responders.

Correspondence Julie B. Pajaniaye. E-mail: jp@dent.au.dk

Accepted 26 November 2025

Published 5 February 2026

Conflicts of interest STS reports financial support from or interest in the European Research Council, the European Union's Horizon, Munksgaard, TrustMe-Ed Nestlé Health Science and GLA:D. JBL reports financial support from or interest in the Association of Danish Physiotherapists, Aarhus University, the Dagmar Marshall Foundation, the L.F. Foght Foundation, the K.A. Rohde Foundation, the Orthopaedic Research Foundation Aarhus, the Frimodt Heineke Foundation, the Danish Shoulder Arthroplasty Registry, Danish Orthopaedic Academy. IM reports financial support from or interest in FADL's Forlag, Danish Hip Arthroplasty Registry. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. These are available together with the article at ugeskriftet.dk/dmj

References can be found with the article at ugeskriftet.dk/dmj

Cite this as Dan Med J 2026;73(3):A03250164

doi 10.61409/A03250164

Open Access under Creative Commons License [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nd/4.0/)

REFERENCES

1. Dobson F, Hinman RS, Hall M, et al. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. *Osteoarthritis Cartilage*. 2012;20(12):1548-1562. <https://doi.org/10.1016/j.joca.2012.08.015>
2. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthritis Cartilage*. 2013;21(8):1042-1052. <https://doi.org/10.1016/j.joca.2013.05.002>
3. Juhl C, Christensen R, Roos EM, et al. Impact of exercise type and dose on pain and disability in knee osteoarthritis: a systematic review and meta-regression analysis of randomized controlled trials. *Arthritis Rheumatol*. 2014;66(3):622-636. <https://doi.org/10.1002/art.38290>
4. Gibbs AJ, Gray B, Wallis JA, et al. Recommendations for the management of hip and knee osteoarthritis: a systematic review of clinical practice guidelines. *Osteoarthritis Cartilage*. 2023;31(10):1280-1292. <https://doi.org/10.1016/j.joca.2023.05.015>
5. Bennell K, Dobson F, Hinman R. Measures of physical performance assessments: Self-Paced Walk Test (SPWT), Stair Climb Test (SCT), Six-Minute Walk Test (6MWT), Chair Stand Test (CST), Timed Up & Go (TUG), Sock Test, Lift and Carry Test (LCT), and Car Task. *Arthritis Care Res (Hoboken)*. 2011;63(suppl 11):S350-S370. <https://doi.org/10.1002/acr.20538>
6. Sharma S, Wilson R, Pryymachenko Y, et al. Reliability, validity, responsiveness, and minimum important change of the Stair Climb Test in adults with hip and knee osteoarthritis. *Arthritis Care Res (Hoboken)*. 2023;75(5):1147-1157. <https://doi.org/10.1002/acr.24821>
7. Mostafaei N, Rashidi F, Negahban H, Ebrahimzadeh MH. Responsiveness and minimal important changes of the OARSI core set of performance-based measures in patients with knee osteoarthritis following physiotherapy intervention. *Physiother*

Theory Pract. 2024;40(5):1028-1039. <https://doi.org/10.1080/09593985.2022.2143253>

- 8. Terwee CB, Peipert JD, Chapman R, et al. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res. 2021;30(10):2729-2754. <https://doi.org/10.1007/s11136-021-02925-y>
- 9. Terluin B, Eekhout I, Terwee CB, de Vet HCW. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. J Clin Epidemiol. 2015;68(12):1388-1396. <https://doi.org/10.1016/j.jclinepi.2015.03.015>
- 10. Silva MDC, Perriman DM, Fearon AM, et al. Minimal important change and difference for knee osteoarthritis outcome measurement tools after non-surgical interventions: a systematic review. BMJ Open. 2023;13(5):e063026. <https://doi.org/10.1136/bmjopen-2022-063026>
- 11. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Interpretability. In: de Vet HCW, Terwee CB, Mokkink LB, Knol DL, eds. Measurement in medicine: a practical guide. Practical guides to biostatistics and epidemiology. Cambridge University Press, 2011:227-274. <https://doi.org/10.1017/CBO9780511996214.009>
- 12. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res. 2021;30(8):2197-2218. <https://doi.org/10.1007/s11136-021-02822-4>
- 13. Holm PM, Schrøder HM, Wernbom M, Skou ST. Low-dose strength training in addition to neuromuscular exercise and education in patients with knee osteoarthritis in secondary care - a randomized controlled trial. Osteoarthritis Cartilage. 2020;28(6):744-754. <https://doi.org/10.1016/j.joca.2020.02.839>
- 14. Skou ST, Roos EM. Good Life with osteoArthritis in Denmark (GLA:D™): evidence-based education and supervised neuromuscular exercise delivered by certified physiotherapists nationwide. BMC Musculoskelet Disord. 2017;18(1):72. <https://doi.org/10.1186/s12891-017-1439-y>
- 15. Roos EM, Lohmander LS. The Knee Injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes. 2003;1:64. <https://doi.org/10.1186/1477-7525-1-64>
- 16. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther. 2009;17(3):163-170. <https://doi.org/10.1179/jmt.2009.17.3.163>
- 17. Devji T, Carrasco-Labra A, Qasim A, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ. 2020;369:m1714. <https://doi.org/10.1136/bmj.m1714>
- 18. Schünemann HJ, Puhan M, Goldstein R, et al. Measurement properties and interpretability of the Chronic Respiratory Disease Questionnaire (CRQ). COPD. 2005;2(1):81-89. <https://doi.org/10.1081/COPD-200050651>
- 19. Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. J Clin Epidemiol. 2017;83:90-100. <https://doi.org/10.1016/j.jclinepi.2016.12.015>
- 20. Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010;63(5):524-534. <https://doi.org/10.1016/j.jclinepi.2009.08.010>